Small-Scale Anisotropies of the Cosmic Microwave Background: Experimental and Theoretical Perspectives

Eric R. Switzer

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF PHYSICS
[Adviser: Lyman Page]

November 2008
Abstract

In this thesis, we consider both theoretical and experimental aspects of the cosmic microwave background (CMB) anisotropy for $\ell > 500$. Part one addresses the process by which the universe first became neutral, its recombination history. The work described here moves closer to achieving the precision needed for upcoming small-scale anisotropy experiments. Part two describes experimental work with the Atacama Cosmology Telescope (ACT), designed to measure these anisotropies, and focuses on its electronics and software, on the site stability, and on calibration and diagnostics.

Cosmological recombination occurs when the universe has cooled sufficiently for neutral atomic species to form. The atomic processes in this era determine the evolution of the free electron abundance, which in turn determines the optical depth to Thomson scattering. The Thomson optical depth drops rapidly (cosmologically) as the electrons are captured. The radiation is then decoupled from the matter, and so travels almost unimpeded to us today as the CMB. Studies of the CMB provide a pristine view of this early stage of the universe (at around 300,000 years old), and the statistics of the CMB anisotropy inform a model of the universe which is precise and consistent with cosmological studies of the more recent universe from optical astronomy.

The recombination history is the largest known uncertainty in the standard cosmological model for CMB anisotropy formation. Here we investigate the formation of neutral helium during cosmological recombination and find that it is significantly different than was previously understood. We show that several new effects produce a modest ($\sim 0.5\%$ at $\ell = 3000$) change in the CMB anisotropy that should be included in the analysis of data from small-scale anisotropy experiments. These small scales contain further information from the primary CMB and from interactions of the CMB with intervening matter as it travels from the recombination era to us today. This information will improve constraints on the spectral slope of primordial perturbations, the baryon density, and possibly dark energy. The methods described in the second half of this thesis are part of the ACT analysis pipeline.
Acknowledgments

The Princeton cosmology community has historically supported a wide range of research, from experimental investigations of the cosmic microwave background to theoretical studies of the earliest moments of the universe. This outlook was established by the “gravity group” of Dicke and others in the 1960s. The work on cosmological recombination in Chapter 2 is an extension of the theory described by P. J. E. Peebles at Princeton in 1968, and the largest correction described in this thesis was also proposed by Peebles in the 1990s. The Atacama Cosmology Telescope, described in the second half of this thesis, is also part of the “gravity group.” The research described here has been possible only because of this diverse community. I am greatly indebted to Lyman Page and Chris Hirata for bringing me on board with the gravity group. I appreciate their guidance, patience, and instruction over the last few years.

The goal of the chapters describing ACT is to document work by the collaboration, and work I did for the collaboration. The construction of ACT required the effort, insight, planning, and patience of many extraordinary people. Chapter 3 describes the ACT telescope systems. Development of these systems has involved nearly everyone in the experimental part of the collaboration. The telescope motion control took great effort to achieve and I would like to specially acknowledge everyone who has spent time uncovering the riddles of “the pendant”: M. Cozza, M. Devlin, R. Dünner, J. Funke, J. Fowler, A. Hincks, J. Klein, and M. Limon. Chapter 3 focuses on the software systems, conceived and coordinated by J. Fowler. J. Fowler and A. Hincks integrated the telescope robotic systems, encoders, synchronization pulses, and raw detector data into a tightly interlocking system that produces the data that are read directly into the mapmaking pipeline. A. Hincks, T. Marriage, and M. Nolta developed the system described in Chapter 3 that schedules all telescope activities. M. Nolta developed the system to track files from when they are acquired to when they are archived at Princeton. M. Amiri, B. Burger, M. Halpern, and M. Hasselfield were the primary developers of the firmware and software for the camera data acquisition systems described here. My role in the software systems-level effort was to develop a message passing system that relays scheduled commands or operator requests to the data acquisition and housekeeping systems, where they are interpreted and executed.

The housekeeping systems (described in Chapter 3) have involved a large number of people, both for the work with the prototype camera and the three-band ACT camera that is now deployed in the field. The work described here was completed in collaboration with A. Aboobaker, M. Devlin, T. Devlin, S. Fletcher, A. Hincks, L. Page, M. Kaul, J. Klein, J. Lau, D. Marsden, B. Netterfield, A. Sederberg, and D. Swetz. We were fortunate to inherit much of the ACT housekeeping from BLAST, in particular through M. Devlin, A. Hincks, J. Klein, and B. Netterfield. I have worked closely with this group to develop the thermal control and cryogenic cycling systems (electronics and software) used both in prototype and final ACT cameras that allow the camera to operate autonomously under remote control.

I would like to thank and acknowledge J. Appel, R. Fisher, J. Fowler, A. Hincks, T. Marriage, M. Niemack, B. Reid, A. Sederberg, J. Sievers, S. Staggs, and D. Swetz for their collaboration with the work described in Chapter 5 describing the calibration of the 145 GHz camera. M. Niemack laid the groundwork for understanding the time constants and efficiency in the lab and in the field. J. Appel, R. Fisher, J. Fowler, A. Hincks, T. Marriage, B. Reid, T. Marriage and I then built up a set of standard calibration tools for the analysis pipeline. A. Hincks has developed an independent pipeline specifically to understand the beams, pointing, and planet amplitudes. The calibration work described here draws on all of these efforts. Finally, mapmaking (which is described in the last chapter of this thesis) has been a large collaborative effort. R. Dünner, R. Fisher, A. Hincks, T. Marriage, J. Sievers and I have worked together to understand the different types of data patholo-
gies, and to find ways that they can be identified and treated. The core map estimation has been led by R. Lupton, T. Marriage, and D. Spergel, and is broken into subgroups for the atmosphere, pointing and calibration. I would like to thank D. Spergel for discussions and methods described in the mapmaking section and appendix. The core analysis work described here concerns calibration.

It has been a pleasure to work with the collaboration as a whole. We’ve dealt with a number of surprises, and even in the hardest, most frustrating times at 5190 m where several difficult systems needed attention, a strong sense of dedication to solving the problem and an adventurous spirit prevailed. In addition to Lyman Page, my advisor, we have a tight-knit group, and I’ve learned much from and enjoyed working with Mark Devlin, Joe Fowler, Mark Halpern, Norm Jarosik, Michele Limon, Jeff Klein, David Spergel and Suzanne Staggs. Aside from work, I’ve had some great officemates, coworkers and mentors in the lab: Asad Aboobaker, John Appel, Adam Hincks, Tom Essinger-Hileman, Ryan Fisher, Lewis Hyatt, Judy Lau, Jeff McMahon, Glen Nixon, and Yue Zhao. I would also like to thank and acknowledge the administrative staff, machinists, and information technologies staff of the physics department. This work has made considerable use of the department computing cluster. I would like to thank C. Champagne especially for his work (and patience) with large lists of obscure components that are now part of the telescope.

The hallway is an unappreciated space in departments. Jeff McMahon and Chris Hirata convinced me (indirectly) to join ACT and work on recombination during conversations we had in the hallway. It’s a great experience not only to interact with people who are passionate about their work, but also to discover new work directions to be passionate about.

I would like to thank Jo Dunkley, Ryan Fisher, Adam Hincks, Chris Hirata, Toby Marriage, Michael Niemack, and Sergio Verdu for critically reading different parts of this thesis. The primary readers, Lyman Page and Suzanne Staggs, provided many comments that were both insightful and detailed, not only in one draft, but through several revisions. I would like to thank Andrew Bazarko, Peter Meyers, and Uroš Seljak for their guidance early in the graduate program and in teaching.

My graduate work was primary conducted from Princeton, but I would like to thank Scott Dodelson (Fermilab), Salman Habib and Katrin Heitmann (Los Alamos), the organizers of the Trieste numerical cosmology workshop (Uroš Seljak) and the Max Planck recombination workshop (J. Chluba, C. Hernandez-Monteagudo, J. A. Rubino-Martin, R. Sunyaev) for guidance and discussions. I would also like to thank the community in San Pedro de Atacama (Chile) and Astro Norte for their hospitality. Much of the work with housekeeping was done at Penn. I’d like to thank Dan Swetz for hosting me a few days there during some critical parts of the setup. I would also like to thank Lyman and Lisa for their hospitality for the last part of the thesis work.

I am greatly indebted to the University of Chicago, in particular to Sean Carroll, Scott Dodelson, Stephan Meyer, Mark Oreglia, and Clifford Lopate for their teaching, advice, and support of independent research during the summer. I also acknowledge the humanities and math programs, in particular the instruction of Dan Grossman in “basic algebra.” I feel fortunate to return to Chicago as a postdoc.

Princeton has also been a very pleasant place to live thanks to some good friends. Eric Dahl, Janice Hester, Chris Hirata, Justin Kinney, Nicole Lemaster, Katie Mack, Amin Nasr, Annika Peter, Antoinette Sutto, Rob Wagner, and Amol Uphadhye are part of the “Kalluri crew,” and while we did have many dinners at Kalluri Corner, they have also been good friends in general for the last few years. Adam Hincks and Philip Eckhoff were great roommates for the year that I lived with them in Hibben. I’ve also made many friends from the entering class and wish them good luck in their endeavors. Above all, I would like to thank Susannah Rutherglen for her support in what seemed at times to be endless work and all the time we’ve had together at Princeton.

I would like to thank my parents, Jay and Barbara Switzer, for all of their support. This thesis is dedicated to them.
Chapter 2 describes cosmological helium recombination and is adapted from three papers with C. Hirata, Switzer and Hirata (2008a); Hirata and Switzer (2008); and Switzer and Hirata (2008b). Chapter 3 was adapted and augmented from an SPIE proceeding describing the ACT software and electronics architecture, see Switzer et al. (2008). The other chapters describing the site stability and calibration are original work specifically for this thesis.
## Contents

Abstract .................................................. iii
Acknowledgments ........................................ iv
List of Figures ........................................... xi

1 Introduction ........................................... 1
  1.1 Introduction ........................................ 1
  1.2 The homogeneous universe and the Planck spectrum .......... 2
  1.3 Departure from homogeneity and isotropy ................... 4
  1.4 The Sunyaev-Zel'dovich effect .......................... 8
  1.5 The Atacama Cosmology Telescope ....................... 9
  1.6 Sensitivity limits for millimeter wave detection .......... 11

2 Recombination .......................................... 13
  2.1 Introduction ......................................... 13
  2.2 A new level code with radiative feedback ................. 18
    2.2.1 Summary of the method ........................... 19
    2.2.2 Radiative feedback in hydrogen .................... 22
  2.3 Hydrogen continuum opacity and helium recombination ..... 23
    2.3.1 The continuous opacity of neutral hydrogen ......... 24
    2.3.2 Feedback from transport between He I lines and continuous opacity ...... 25
    2.3.3 Transport within He I lines and continuous opacity: physical argument ... 29
    2.3.4 The modified escape probability without coherent scattering ......... 30
    2.3.5 An overview of transport with coherent scattering .......... 34
    2.3.6 Atomic occupation rate evolution with coherent scattering .......... 35
    2.3.7 Monte Carlo method .................................. 37
    2.3.8 Incorporation in the level code ..................... 38
  2.4 Additional effects .................................... 40
    2.4.1 Thomson scattering .................................. 40
    2.4.2 Monte Carlo simulation of line transport with Thomson scattering .... 43
    2.4.3 Peculiar velocities .................................. 44
  2.5 Summary of results for helium, discussion ............... 47

3 ACT Systems ............................................ 51
  3.1 Telescope systems overview ............................. 51
  3.2 Telescope systems ..................................... 55
    3.2.1 Time and synchronization ........................... 55
    3.2.2 Camera acquisition and commanding .................. 60
    3.2.3 Housekeeping readout ............................... 61
    3.2.4 Housekeeping: thermal control ...................... 62
    3.2.5 MBAC cycling ........................................ 64
6 Quantifying Sources of Noise and Elements of Mapmaking 137
   6.1 Introduction – converting bits to maps ............................................. 137
   6.2 An overview of mapmaking ............................................................... 138
   6.3 External noise sources and correlations
      6.3.1 Discussion ................................................................................. 139
      6.3.2 Overview of contaminants ............................................................ 140
      6.3.3 Overview of treatments for contaminants ...................................... 140
   6.4 Noise diagnostics – normality .............................................................. 143
   6.5 Noise diagnostics – correlations .......................................................... 143
   6.6 Modeling noise that is correlated in time ............................................. 148
   6.7 Conclusions ...................................................................................... 151

7 Conclusion 153

A Recombination 155
   A.1 The temperature of matter ................................................................. 155
   A.2 The escape probability with continuous opacity and complete redistribution........ 156
   A.3 Transport in Doppler width-dominated lines with electron scattering ................. 158
   A.4 Fluid description of the baryons ......................................................... 161
   A.5 Monte Carlo method: implementation ............................................... 164
   A.6 Convergence of the methods ............................................................. 167
   A.7 Experimental possibilities for detecting the recombination spectral distortion ......... 167
      A.7.1 Introduction ............................................................................... 167
      A.7.2 Spectral contamination at low frequencies ...................................... 170
      A.7.3 Low frequencies – hardware perspective and calibration .................... 171
      A.7.4 High frequency experiments ....................................................... 172
      A.7.5 Conclusions .............................................................................. 173

B ACT Systems 175
   B.1 CCAM heater controller design ......................................................... 175
   B.2 MBAC heater controller design ......................................................... 181
      B.2.1 The heater controller ................................................................. 181
      B.2.2 The current monitor, relay box ...................................................... 182
      B.2.3 The high power drivers and fault detection .................................... 183
   B.3 The pressurized drive systems ........................................................... 187
   B.4 Analysis of the flux lock loop ............................................................. 194
      B.4.1 The MCE $\Sigma - \Delta$ readout ....................................................... 194
   B.5 MUX response analysis ................................................................. 197

C Calibration and Diagnostics 199
   C.1 Effective area in the geometric limit ................................................... 199
   C.2 Planet temperature .......................................................................... 200
      C.2.1 The heliocentric distance .............................................................. 200
      C.2.2 Ellipticity and solid angle ............................................................ 200
      C.2.3 Jupiter ...................................................................................... 202
      C.2.4 Uranus and Neptune ................................................................. 203
      C.2.5 Mars ....................................................................................... 203
      C.2.6 Other calibrators ...................................................................... 205
   C.3 The accuracy of a pulse calibration .................................................... 206
   C.4 Relation between B-splines, the optimal filter, and iterated methods ................. 207
      C.4.1 Iterative solution ........................................................................ 208
List of Figures

1.1 Left: The angular power spectrum of the CMB on large scales (Nolta et al. (2008)). Right: The angular power spectrum of the CMB on small scales (Dunkley et al. (2008)). (The references for other experimental results plotted here are cited in the above work from the WMAP collaboration.) The x-axis is the multipole index $\ell = \pi/\theta$ and the y-axis is the temperature fluctuation power per logarithmic $\ell$ interval.

1.2 The ACT 145 GHz array. The $32 \times 32$ array is the central square inside of the copper carrier. The carrier and the blackened cavity surrounding it are held at 0.3 K by a $^3$He adsorption refrigerator. The array comprises a stack of 32 silicon cards, each of which has 32 folded “pop-up” detectors. (Here “pop-up” refers to the fact that the absorber is perpendicular to the silicon card carrier, and is held up by weak thermal link legs.) Each detector unit is a $\sim 1$ mm $\times$ 1 mm $\times$ 1 $\mu$m square of boron-doped, ion-implanted silicon held up from a silicon carrier card by four $\sim 1$ $\mu$m $\times$ 5 $\mu$m cross-section legs which conduct electrical and optical power from the detector to the bath. These are bonded to SQUID multiplexing and biasing electronics on the silicon carrier card, which are read out and addressed through the ribbons on the left. The array face is covered by a 50$\mu$m silicon AR coupling layer, separated from the detectors by 100$\mu$m, but not shown here. (Photo: M. Niemack)

2.1 The optically thick Ly$\alpha$ “traffic jam.” On the left, de-excitations of one atom produce photons that excite the $1s$ state of a nearby atom in an optically thick gas, and the reverse. Because the universe is expanding, this cycle can be broken if the photons redshift off-resonance before they can excite nearby atoms. This is referred to as Sobolev or radiative escape. Two photon emission from $n = 2$ is optically thin to reabsorption, allowing atoms to relax to the ground state through this pathway, without the analogous traffic jam. In helium $^2P^o - ^1S$ (and other lines with energies $> 13.6$ eV), this traffic jam can also be broken when the photon photoionizes a neutral hydrogen ($H(1s) + \gamma \rightarrow H^+ + e^-$) rather than re-exciting another helium atom. This is described in Sec. 2.3.1 and is one of the major goals of this chapter.

2.2 The life-cycle of hydrogen during recombination in the three-level approximation. The top oval represents a pool of free nuclei and electrons. Electrons are captured onto the nuclei through free-bound (recombination) processes, and liberated from the nuclei through bound-free (photoionization) processes. The excited states $n > 2$ (middle oval) are in equilibrium, so can be treated in a lumped category. This is the essential observations in the three-level approximation. The primary formation rate of atoms in the ground state (bottom oval) is through decays from the $n = 2$ level. In hydrogen this is just $2\gamma$ (from $2s \rightarrow 1s$) and Ly$\alpha$ (from $2p \rightarrow 1s$), while in helium the primary paths are $^2P^o - ^1S$, $^3P^o - ^1S$ and $^2S - ^1S$. (Fig. 2.3 gives the Grotrian diagram for the lowest excited states of helium.)
2.3 Formation of neutral helium: a Grotrian diagram (up to \( n = 3 \)) for He \( \text{I} \). The notation used throughout is the standard term symbol \( n^{2S+1}L_J \) where \( n \) is the principal quantum number of the excited electron, \( S \) is the total spin, \( L \) is the total orbital angular momentum and \( J \) is the total angular momentum. Singlet (\( S = 0 \) parahelium) and triplet (\( S = 1 \) orthohelium) levels and higher-order transitions give a rich system of low-lying transitions. Marked are two-photon transitions (short dashed lines from \( 2^1S_0 \) and \( 3^1S_0 \)), allowed electric dipole transitions (solid lines from \( 2^1P_1^o \) and \( 3^1P_1^o \), like the Lyman series in H \( \text{I} \)), the intercombination lines (long dashed lines from \( 2^3P_1^o \) and \( 3^3P_1^o \)), and quadrupole transitions (from \( 3^1D_2 \)). The dipole transitions are treated in Sec. 2.3.7 using a Monte Carlo method with partial redistribution, forbidden one-photon lines are treated in Sec. 2.3.4 using an analytic method for complete redistribution. (The energy levels are not drawn to scale.)

2.4 Convergence of the iterations to include feedback of non-thermal distortions between lines. These are descending from the difference between no feedback and one iteration (top solid line), between the first iteration and the second, and so on. Note that by the 4th iteration, the effect is roughly \( \Delta x_e < 10^{-4} \), so by going a fifth iteration, any systematic effect is negligible. Note that the integration tolerance taken in the level code is \( 1 \times 10^{-5} \).

2.5 The ionization and relaxation rates (in \( s^{-1} \)) for H \( \text{I} \) during the period of He \( \text{I} \) recombination, assuming each He \( \text{I} \) recombination generates a photon that photoionizes a hydrogen atom. Here we consider two He \( \text{I} \) recombination histories: one in equilibrium and the history derived here (Fig. 2.14). The “H \( \text{I} \) relaxation rate” is \( \eta_{\text{Saha}}^{-1} \) (Eq. 2.22) for H \( \text{I} \) to return to Saha equilibrium if its abundance is perturbed. In either He \( \text{I} \) history, the ionizing radiation from He \( \text{I} \) is not sufficient to push H \( \text{I} \) evolution out of the Saha steady-state because of the large disparity in their rates. The Hubble rate at \( z = 2400 \) was \( \sim 2 \times 10^{-13} \text{ s}^{-1} \).

2.6 The continuum optical depth \( d\tau_c/d\nu = \eta_c \) times the Doppler width of He \( \text{I} \) \( 2^1P^o - 1^1S \) as a function of redshift.

2.7 A comparison of the effect of feedback of a spectral distortion in helium and hydrogen recombination produced by higher-lying states on lower-lying states in the same species after several iterations. \( \Delta x \) is the change in the fractional species abundance (He \( \text{II} \), He \( \text{I} \), and H \( \text{I} \)) with feedback minus without feedback. Here, continuous opacity from hydrogen photoionization between He \( \text{I} \) is included. The effect of this opacity is shown separately in Fig. 2.8. In all cases, feedback retards formation of the neutral species. Here, the uppermost line is the first iteration, moving down with further iterations and better convergence.

2.8 The effect of hydrogen continuum absorption on the feedback between transitions to the ground state in He \( \text{I} \). Feedback slows He \( \text{I} \) recombination, but becomes increasingly less significant as the neutral hydrogen population grows, increasing the continuum opacity. The dashed trace here (with continuous opacity) is analogous to the middle panel of Fig. 2.7.

2.9 Inverse of the differential optical depth \( \eta_c \) from hydrogen photoionization as a function of redshift, compared to the optically thick line-width due to incoherent processes in the \( 2^1P^o - 1^1S \) line. Continuum processes start to become important over scales inside the (incoherent) optically thick part within the line around \( z = 2100 \). Also plotted is the Doppler width of the line, emphasizing that the line is optically thick out into the wings. Continuum processes do not act on scales smaller than the Doppler core until \( z < 1800 \). Here it also clear that the Doppler width is small compared to the optically thick linewidth.
2.10 The radiation phase space density near the intercombination line $^{1}S_{0} \rightarrow ^{2}P_{1}$
resonance with Voigt parameter $a = \Gamma_{\text{line}}/(4\pi \Delta \nu_{D}) = 10^{-5}$ for the Sobolev optical
depth $\tau_{S} = 2.8$ and $N_{C} = 0$, for several sample continuum optical depths $\eta_{C}\Delta \nu_{D}$. 
Here, $\Delta \nu_{D}$ is the Doppler width, so that $\eta_{C}\Delta \nu_{D}$ is the opacity for redshifting through
a Doppler width. Here we have normalized the radiation phase space density by
its equilibrium value with the line ($N/N_{L}$) on the y-axis. The x-axis is the detuning
from line center in Doppler widths, $x = (\nu - \nu_{o})/\nu_{D}$. $N$ does not converge to $N_{L}$
on the red side of the line because the depth $\tau_{S} = 2.8$ is not very thick, so even for
zero continuous opacity the radiation is not fully driven into equilibrium with the line
(only to 94%). Because of the low optical depth and small natural line width, very
little radiation extends more than three Doppler widths above the line, and the effect
of the continuum is significant in relaxing the radiation phase space density near
line-center.

Coherent and incoherent scattering through $^{2}P_{0} - ^{1}S$. If He $^{1}S_{0} \rightarrow ^{2}P_{1}$ is
excited by one photon, and a second photon excites $^{1}S_{0} \rightarrow ^{2}P_{1}$ (left panel), then
if the atom decays back through $^{2}P_{1} \rightarrow ^{1}S_{0}$ (right panel), the outgoing energy will
be completely redistributed across the Voigt profile. We refer to this as “incoherent”
scattering. If, rather than absorbing a second photon and visiting additional levels,
the atom decays $^{2}P_{1} \rightarrow ^{1}S_{0}$ promptly after it was excited through $^{1}S_{0} \rightarrow ^{2}P_{1}$,
then energy will be conserved in the atom’s rest frame and in the gas, the outgoing
photon energy will only be thermally broadened and not completely redistributed
across the Voigt profile. We refer to this as “coherent” scattering. The distinction
here is that the atom cannot “forget” the incoming state energy, as it can in the
multi-photon case where there are additional excursions and energy does not have
to be conserved solely through $^{1}S_{0} ↔ ^{2}P_{1}$. The combination of these redistribution
processes leads to partial redistribution.

2.12 The change in the free electron fraction ($\Delta x_{e}$) due to the additional effects of contin-
uum opacity in the $n^{1}D - ^{1}S$, $n^{3}P_{n} - ^{1}S$, and $n^{1}P_{n} - ^{1}S$ lines. In the solid trace,
we compare a model without feedback and with only $^{2}P_{0} - ^{1}S$ continuum opacity
to a model (also) without feedback and with opacity in the full $n^{1}P_{n} - ^{1}S$ series. In
reality, rates through $n^{1}P_{n} - ^{1}S$ accelerate with continuum opacity included, but in
doing so, they also produce a spectral distortion (in particular, $3^{1}P_{2} - ^{1}S$), which
redshifts and impacts $^{2}P_{0} - ^{1}S$. The long dashed history shows where we dif-
ference these models, both including feedback. Ironically, the effect of accelerating
$n^{1}P_{n} - ^{1}S$ for $n > 2$ is almost cancelled by the effect that the additional spec-
tral distortion impedes $^{2}P_{0} - ^{1}S$. It is not until $z < 2000$ that there is sufficient
continuum opacity to absorb the effect of this extra distortion. The short-dashed
line shows the effect of continuum opacity in the $n^{1}D - ^{1}S$ and $n^{3}P_{n} - ^{1}S$ series
(differencing a base model that includes feedback in both cases). This is becomes
reasonably large $\sim 10^{-4}$. In the dotted trace, we examine the effect of coherent
scattering/partial redistribution from the Monte Carlo results. By neglecting coherent
scattering, the recombination history is slower, but only producing a difference at the
level of $\sim 10^{-4}$. The ultimate accuracy for a recombination code should be $\sim 10^{-3},$
so these higher-order effects can probably be neglected without loss. In particular,
coherent scattering is computationally intensive, but here we have shown that its
contribution is small. This suggests that the analytic approach in Eq. 2.42 would be
sufficient. Recently Rubiño-Martín et al. (2008) have shown that coherent scattering
could be included as a fudge factor to a more analytic approach like Eq. 2.42.
2.13 The modified escape probability from $^4\text{He} \ 2^1 P^o - 1^1 S$ during He I recombination, comparing the results of the standard Sobolev approximation and the modification due to continuous opacity with and without coherent scattering. When we include the continuous opacity developed in Sec. 2.3.3, the probability that a photon escapes the line is more than an order of magnitude larger than in traditional approaches by $z \sim 2000$. Once coherent scattering is introduced, photon diffusion effects increase the escape probability by increasing the span of frequencies traversed by the scattering photon before it escapes. These probabilities are log-interpolated over the grid of $x_{\text{HeI}}$ and $z$ used in the level code. We find that the effect of doubling the grid resolution and with it the smoothness of the interpolated probability is negligible.

2.14 The free electron fraction $x_e$ as a function of redshift $z$ during He I recombination with the inclusion of feedback and continuum opacity in He I lines, compared to a Saha He I recombination history and the standard He I recombination history. (Here we estimate the standard history using the same code, but turn off continuous opacity and feedback, this is labelled “No feedback or continuous opacity” in the key) Note that through the choice to normalize $x_i = n_i/n_{\text{H}}$, (rather then $n_b$) the free electron fraction exceeds 1 during this interval. Continuous opacity starts to become important at $z \sim 2100$ (as suggested by Fig. 2.9), and pushes the He I evolution to a Saha steady-state by $z \sim 1800$. The beginning of H I recombination is visible here starting at $z \sim 1700$, and for later times $x_e$ drops precipitously. The modifications to He I recombination suggested here are twofold: at early times during He I recombination, feedback slows recombination (Sec. 2.2.2), and at later times, continuous opacity (Sec. 2.3.3) accelerates He I recombination relative to standard models.

2.15 The electron scattering kernel at $z = 2500$ spans several THz, allowing photons in the far red side of the line to be scattered to frequencies significantly above the line center frequency. Rather than escaping on the red side of the line, as in ordinary Sobolev escape, the photon can then be absorbed (and likely be re-emitted) by incoherent processes in the line before it can escape, thus reducing its escape probability. Because the radiation phase space density is higher on the red side of the line, a photon is more likely to scatter from below to above the line frequency, thus decreasing the overall escape probability. Once continuum opacity becomes significant in the allowed lines, this can also remove trapped photons.
2.16 Left: The impact of Thomson scattering on escape probabilities in $^2P^o - ^{1}_S$ and $^2P^o - ^{1}_S$ during recombination. A comparison of analytic and Monte Carlo treatments of transport in $^2P^o - ^{1}_S$ which give a modified escape probability due to electron scattering. The analytic description of electron scattering in narrow lines with complete redistribution of Eq. (A.35) of Appendix A.3 agrees well with Monte Carlo results at early times. Once continuous opacity becomes important within the line, the Monte Carlo gives a dramatic increase in the escape probability. Built into the analytic method, however, is the assumption that the line is indefinitely narrow. Thus, the analytic method breaks down $z < 2100$ and approaches the Sobolev theory. (The history without electron scattering is also consistent with the analytic method for complete redistribution and continuous opacity.) Right: The fractional modification to the $^2P^o - ^{1}_S$ escape probability due to electron scattering and $^3He$ scattering. The effect of electron scattering can be split into two regimes: 1) for $z > 2300$ where most of the effect from electron scattering is due to modification to transport far in the wings and the escape probability is decreased, 2) for $z < 2300$, where much of radiation in the red wing is absorbed through $H$ I photoionization, and electron scattering begins to eject more photons from within the line, on average. Thus, for $1600 < z < 2300$ electron scattering slightly increases the escape probability, shown in the right panel as an increase of a few percent above zero for $z \sim 2200$. $^3He$ does not produce any effect at the level of precision of the Monte Carlo run here. These required 3 days across $50 \times 3$ GHz nodes, so improved statistics would require significant computing time.

2.17 Comparing the effect of Thomson scattering, $^3He$ scattering, and feedback in the forbidden and allowed lines. The uppermost curve is the difference between two models with and without Thomson scattering in the $n^3P^o - ^{1}_S$ and $n^1D - ^{1}_S$ where neither has feedback. This retards recombination because, on average, more photons are injected into the optically thick region of the line. (Note that once feedback of the radiative distortion is added, the effect of Thomson scattering is greatly reduced.) Thomson scattering in $^2P^o - ^{1}_S$ decreases the escape rate at early times, but once $H$ I opacity becomes significant, more photons are on average removed from the optically thick part of the line. $^3He$ can be seen to accelerate recombination slightly by assisting photons out of optically thick regions at late times. We note, though, that the typical error induced by resampling the Monte Carlo is of order $10^{-4}$ (see Fig. A.4). To confirm the effect of $^3He$ would take a significantly finer grid of probabilities, but the actual value is immaterial to $He$ I recombination overall if it is this small.

2.18 The cumulative effect on temperature and polarization anisotropies from $H$ I continuous opacity and feedback during $He$ I recombination developed in this work, as calculated by CMBFAST. We note that here we only consider the difference between the reference helium model with and without these effects. Comparison of the full $H$ I and $He$ I history to standard methods such as RECFAST will be the subject of later work.
2.19 Left: The He I recombination history from our multi-level atom code (solid line), compared to the Saha equation (long-dashed) and the commonly used three-level code RECFAST by Seager et al. (1999) (short-dashed). The y-axis gives the free electron fraction relative to $n_{\text{H}}$ and the $x$-axis is the redshift. Both our analysis and that of Seager et al. find that He I recombination is delayed due to the $n = 2$ bottleneck. However we find a slightly faster recombination than Seager et al. due primarily to our inclusion of the intercombination line $2^3P^o - 1^1S$ and the accelerating effect of H I opacity. The latter effect causes our He I recombination to finish at $z \approx 1700$, whereas in RECFAST one-third of the helium is still ionized at that time.

Right: The He II recombination history from the level code developed here. He II recombination is essentially irrelevant for CMB physics, and because of its rates it varies from Saha evolution at the level of $< 0.2\%$. (For example, at $\ell < 3000$ the absolute difference in $C^T_T$ for the full model relative to Saha is $< 3 \times 10^{-5}$.)

3.1 Overview of the ACT structure and site. Left: the ACT site and its components. The workshop container has a hoist shed so that the camera can be assembled and disassembled at the site. The facility is powered by one of two on-site diesel generators, and has on-site diesel storage capacity of $1.5 \times 10^4$ liters (the black tank). It is located at $64^\circ 47' 15''$ W by $22^\circ 57' 31''$ S at 5190 m on Cerro Toco in the Andes east of the Atacama desert of Chile and is a $\sim 1$ hour commute from the small town of San Pedro de Atacama at 2400 m. The moving mass of the telescope is 40 t (the total mass is 52 t) and extends 12 m from the ground. (The outer ground screen is 13 m and reflects far sidelobes to the sky.) The range of motion is through $\pm 220^\circ$ azimuth and $30.5^\circ - 60^\circ$ elevation. It has a maximum azimuth scan speed of $2^\circ /s$, and a maximum acceleration at turnarounds of $2^\circ /s^2$. The elevation is fixed to whatever extent possible and actively controlled to reduce systematic effects, but can re-point to target planets or smaller fields at $0.2^\circ /s$. In practice, we scanned in azimuth at $1^\circ$ per second at $50.5^\circ$ elevation (or $\sim 0.8^\circ /s$ on the sky) in the 2007 season. The smallest diameter of the primary mirror (composed of 71 panels in 8 vertical rows) is 6 m. The secondary mirror is at ambient temperature, is composed of 11 panels, and has a minimum diameter of 2 m. The primary mirror panels are each adjustable through four manual screw adjusters. The secondary mirror panels were aligned at the AMEC facility and are fine-tuned in the field. The overall position can be readjusted by two frame and three mirror linear actuators. (Photo: A. Hincks; drawings, AMEC/Empire Dynamic Systems)

3.2 The ACT receiver cabin at $60^\circ$ elevation (normal observation in the 2007 season was at $50.5^\circ$ elevation). Subsequent plots that show array response orient the detectors as they appear on the sky, where silicon card “columns” are horizontal on the sky and row zero column zero is in the upper right of the projected array on the sky. (Figure: D. Swetz, B. Thornton)
3.3 Left: the millimetric bolometer array camera (MBAC) in its test stand and the University of Pennsylvania. The dewar is \( \sim 1 \text{m}^3 \) and comprises three arrays and their optics. The square 300 K windows have been covered here, but are 4 mm thick ultra high molecular weight polyethylene (UHMWPE) with a Teflon AR coating. Right: The cold optics in the MBAC dewar. The optical path in the three sub-cameras is similar and consists of three AR-coated silicon lenses (1/4-wave Cirlex on float zone silicon, see Lau et al. (2006b); Lau (2007); Fowler et al. (2007)) and a filter stack, moving from IR blocking filters at 300 K to a low-pass stack at 40 K to a 1 K low-pass and a band-definition filter and final lens at 0.3 K. Thus everything behind the bandpass filter is in a 0.3 K cavity. All interior walls are blackened with carbon lampblack and Stycast 2850 FT epoxy.

3.4 Left: the ACT housekeeping electronics rack. From top to bottom: the synchronization serial stamp and trigger generator, the heater control relay and current monitor box, the BLAST analog to digital conversion unit, the ruthenium oxide thermometry bias generator, the housekeeping preamplifier crate, and a line conditioner. The receiver cabin bulkhead is visible at the bottom of the figure through an acrylic sheet. Digital signals from the receiver cabin pass through the bulkhead to the equipment room, and analog channels from the structural housekeeping enter the receiver cabin, where they are read out. Right: The primary mirror (top) and camera hatch (bottom) as viewed from below the right side of the secondary mirror. The illumination pattern onto the primary mirror is not obstructed by the receiver cabin wall. The wall has a mount (visible here as a small circle) for the laser ranger, which is used to survey the panels. After several iterations of measurement and panel adjustment, \( \sim 30 \mu \text{m} \) RMS across the primary mirror has been achieved.

3.5 Overview of the ACT data and control systems, split into telescope (top), on-site (middle), ground station, and North American systems (bottom). Solid lines show data streams while dashed lines show commanding and file information channels. Cylinders represent data storage. We show only one of three detector array acquisition systems ("Array DAS") for simplicity. The ground station RAID array aggregates data from several machines at the site, so we do not show it as being connected to a particular node. To prepare a transport disk with data to ship back to North America, an operator connects the drive to the RAID array, and the data are automatically copied over based on information in the file database. Operators define a schedule each night and upload it to the schedule dispatcher at the site. Here we have only shown a monitor client in the ground station, but the site also has a system monitor terminal.

3.6 Block diagram for the 0.3 K cryogenics. The pulse tube cryo compressor cooler cools the condensation plate of a \(^4\text{He}\) adsorption refrigerator. Once the \(^4\text{He}\) pot has accumulated sufficient liquid, the \(^4\text{He}\) charcoal pump temperature (which was actively controlled at \( \sim 45 \text{K} \)) is rapidly cooled by conducting the heat out through a gas gap heat switch. This pumps the \(^4\text{He}\) liquid, and cools an analogous condensation plate for \(^3\text{He}\). Once enough liquid \(^3\text{He}\) has accumulated, the charcoal pump (which was actively controlled at \( \sim 35 \text{K} \)) cools by conducting through a second gas gap switch. The charcoal then adsorbs the \(^3\text{He}\), and the \(^4\text{He}\) pot reaches the desired bath temperature of \( \sim 0.3 \text{K} \). A typical hold time during operation was 15 hours, and the adsorption refrigerators are cycled once per day, see Fig. 3.8. This closed-cycle system allows ACT to operate in a remote location without additional liquid cryogens. The automated cycle is described in Sec. 3.2.5.
3.7 Outline of the MBAC heater controller systems. Electronics designed for BLAST at the University of Toronto (UofT) manage the housekeeping input/output from a crate in the receiver cabin. These electronics communicate with the housekeeping computer in the equipment room using the BLAST-bus PCI card (BBCPCI), connected through a serial link from the telescope’s receiver cabin. The heater driver card was designed by T. Devlin’s group at Rutgers; Princeton was responsible for the interface between components and the current monitor/relay box (CRBox). One DB37 connector from the UofT diode readout card programs the voltage levels on the heater driver card (located in the analog readout electronics “Penn” crate), and the DB37 connector on the UofT current monitor readout card toggles relays in the CRBox. (The readout cards in the UofT crate are generic, so the heater driver simply uses the digital IO on the card whose analog inputs are used to read out the diode thermometry.) Significant effort has gone into making the housekeeping system and interconnects RF-tight.

3.8 Temperature of key stages during the cryogenic cycle. Vertical lines indicate when the gas-gap heat switches activate. There are two $^4\text{He}$ adsorption refrigerator units: the “optics” unit is responsible for the optics tubes, while the “backing” unit provides the cold stage for $^3\text{He}$ condensation and is activated second here. The topmost plot shows the pump temperatures; the middle plot shows the condensation plate temperatures; the lowest plot shows the liquid pot temperatures, which first reach 0.9 K, after which the $^3\text{He}$ pot drops to 0.3 K.

3.9 MBAC current monitor and relay box (CRBox). The front panel has a DB50 connector for current monitor outputs, a DB37 connector for digital control inputs, and 24 isolated BNCs. Of the 24 BNCs, there are 18 for heater channels on the heater driver card, and 6 auxiliary channels are for additional lines in the dewar. All auxiliary lines are treated differentially, and the 18 heater lines share a common ground which is independent from the ground in the CRBox. The CRBox channels are labeled from right to left, bottom to top, where e.g., the bottom right is channel 1.

3.10 Left: cards in the CRBox. The RF block and regulators from 16.5 V to $\pm12$ V and 5 V are shown in the bottom of the figure. The back panel (the right side of the left figure) has two DB50 ribbons that lead to the dewar and one DB50 from the heater driver card in the housekeeping preamplifier VME crate. The card stack comprises (from top to bottom) a switchboard card, a current monitor card and a relay card. The switchboard is rotated at a right angle relative to the others to permit a clear connection to the two DB50 connectors to the dewar on the back. The BNCs on the front panel (the left side of the left figure) are connected through a DB50 junction to a ribbon to the IDC50 connection on the relay card, and the LEDs are connected through a straight IDC-50 ribbon and junction. The switchboard “top card” permutes heaters in the dewar (which are broken out from the two DB50 connectors on the back) to combinations of the 24 possible heater channels. Right: back of the CRBox, showing the power 4-pin military connector, DB50 connector for input from the heater driver, and two DB50 connectors for outputs destined for the cold break-out to heaters in MBAC.

3.11 The GUI message-passing server window, here displaying users and instruments, and bindings between those users and command destinations. For example, mce1, mce2, and mce3 identify with “mce control” so a command sent to the destination “mce control” will be executed on the three MCEs simultaneously.

3.12 The GUI housekeeping control window. Here clients can control relays, servos and the cycle state of cryogenics at the site from North America. Other windows provide a similar interfaces for commanding telescope motion and data acquisition systems.
4.1 Mean precipitable water vapor (PWV) from the APEX radiometer. This is the thickness of a sheet of liquid-phase water that is equivalent to the quantity of water through zenith pointing. ACT season 1 science observations spanned Nov. 14, 2007 to Dec. 17, 2007. Anomalous PWV around Nov. 21, Dec. 2, and Dec. 15 restricted observations. There is also a diurnal variation of the PWV that is apparent after removing a daily mean. The amount of water vapor in the column as measured by the APEX radiometer typically increases by roughly a precipitable millimeter from the morning until sunset, but the magnitude of the variation depends on other weather trends. The black lines indicate the first season ACT observation window from ∼10 PM to ∼10 AM – note that this starts at the upper black line (10 PM), moves to the top of the plot, then wraps back from 0 (the bottom of the plot) to the lower black line at 10 AM. All other weather plots indicate this interval the same way. Between 10 PM and 10 AM, the PWV typically falls, but may have additional periods of moisture or even large increases from weather systems such as those that produced the anomalously high PWV in the bright bands here.

4.2 Atmospheric effective RJ brightness temperature from the ATM model, Pardo et al. (2001) from Marriage (2006). The main emission here is from the O$_2$ 119 GHz resonance and the 183 GHz water resonance. Here, "Summer Morning" is 1.55 mm, "Winter Morning" is 0.69 mm, "Summer Evening" is 2.15 mm, and "Winter Evening" is 0.94 mm. The median PWV values and figure are from Marriage (2006).

4.3 The drift estimate calibrated into RJ temperature units using the method described in Sec. 5.9.1. This is the most stable part of the morning of Nov. 14, and the PWV was 0.53 mm here, so this represents good observing conditions for the 145 GHz array in the 2007 season. Here we restrict to working detectors in column 15 of the 145 GHz array. (Typical arbitrary offsets that are removed between files are ∼1 K.) The calibrated effective RJ temperature here is consistent with 6.6 K per APEX mm over long time periods suggested by Eq. 4.7. (Because of the dependence of the water distribution on thermal inversions, pointing and geography, an accurate extrapolation from APEX PWV to ACT PWV is non-trivial.)

4.4 The power spectrum (in $\mu K(RJ)/\sqrt{Hz}$) of atmospheric drift, averaged over 10.5 hours. A Hann window is applied here and the Nyquist frequency is 6.25 Hz (the data have been downsampled by a factor of 32 for economy). The small bump at 0.05 Hz is from scan-synchronous magnetic field pickup in the SQUIDs and/or structure in the atmospheric emission that modulates with the scan. The relation between multipole $l$ and audio frequency $f$ is $l \approx f \pi / \theta^{-1} \approx 180 f$. The drift roughly follows the expected $1/f$ pattern. Taking a single-detector NET of ∼1000 $\mu K(CMB) \sqrt{s}$ implies a knee frequency of ∼1.4 Hz. These data were acquired on a good night, representative of roughly half of the season where the single-detector knee frequency was below 2 Hz. The detector noise averaged down over the array is ∼1000 $\mu K(CMB) \sqrt{s}/\sqrt{900}$. Thus, this $1/f$ drift structure presents a significant analysis challenge. The break in the slope at ∼2 Hz could be due to several factors. The first is that at higher frequencies, there is more sub-structure in the atmosphere across the array that is uncorrelated between detectors. This would produce less noise power from the atmosphere relative to being fully correlated across the detectors. The second is that this is based on only the weighted average of working detectors in column 15 (for economy and common airmass), so detector noise can become comparable to the atmospheric fluctuations.
4.5 Data taken November 22, 2008 demonstrating drift in the array driven by the atmosphere. Here, offsets between each 15-minute file segment are removed as described in the text. The power absorbed is inferred from load curves using Eq. 5.37, described in the calibration section. The absolute power offset is taken from Eq. 4.10 (because it currently cannot be determined from the ACT data) but the scaling of PWV to power is determined empirically. Chilean local time is $UT - 4$ hours. For the purpose of showing the relation between PWV and power absorbed, we show an interval where the PWV radiometer was reliable and shows clean structure over short timescales. This period had very high PWV, spanning between 2.35 mm and 3.5 mm because of a weather system, and is cut in the science data analysis. Fig. 4.7 shows nights with PWV of 0.4 – 1 mm. Note that the short timescale fluctuations differ between the APEX radiometer an ACT, but that over hours, the gross trend in the atmosphere is similar in both instruments. See Delgado and Nyman (2001) for a study of directional differences in PWV, which are expected to account for the finer structure here.

4.6 Left: The sensitivity of the effective absorbed power level to 3 K stage temperature drift (referenced by the cryoperm cap) with the 145 GHz camera covered. The colors here indicate the effective power per K change in the 3 K stage. Most detectors give $-1$ pW equivalent power (in pW) per K change. The 3 K drift sensitivity is non-optical, but we have referenced it to a power level here so that it can be easily applied to optical power measurements. The amplitude of the pickup is correlated with the SQUID $V(\phi)$ slope. Right: The SQUID chain slope measured by applying a ramp to the stage 1 SQUID feedback. (This is the number of error digital units per feedback digital unit applied in open-loop mode.) The series array modules are in groups of 8 columns, so that the first module here is anomalous. Because the detectors are read out through a closed flux-lock loop, the slope impacts bandwidth, but does not impact the optical responsivity. Marriage (2006) shows that the power received by the detectors from emission of the optics is negligible, and even dead detectors exhibit the same 3 K stage drift. This drift is therefore consistent with offsets in the SQUID chain. Both the left and the right plot are oriented as the detectors appear on the sky, where “columns” correspond to the silicon card carriers and are horizontal on the sky. In the left panel, dark blue indicates dead detectors, such as column 22. (We have no pW/DAC calibration there, but the SQUID slope may be measurable.)

4.7 Top: A comparison of the PWV inferred from ACT drifts and the APEX radiometer. The ACT traces here are derived from column 15 of the 145 GHz array, so no airmass gradient across columns (which are horizontal on the sky) needs to be included. The ACT data have an arbitrary vertical offset each night, so for each observation interval we find the least-squared best offset to the APEX data. For period such as Nov. 18, 2007, there were insufficient PWV radiometer data to scale the ACT drifts properly. The conversion from APEX PWV to ACT absorbed power (as inferred from the load curves) is 0.43 pW/(APEX mm). Inverting this, we find the PWV from ACT (green/gray solid line). In the blue/darkest solid line, we have subtracted an estimate of the drift component from the 3 K stage temperature drift. The APEX PWV data are the red/lightest dotted trace. Here we scale the PWV down by 12% to estimate the PWV at the ACT site, based on the difference between the APEX radiometer and ACT elevations and a scale height of 1100 m. Bottom: Same traces as top frame showing Dec. 6-7, 2007. On Dec. 7, 2007, the blue curve (which has been corrected for 3 K drift) matches reasonably well with the reported PWV from APEX, while the ACT data from Dec. 6 show a different trend. These differences are not currently understood.
4.8 Mean wind direction from the APEX weather station. During the day, wind normally rises from the dry plain of San Pedro at \( \sim 10 \) m/s. The largest gusts were during mid-day, and exceeded \( \sim 25 \) m/s for several hours. At night, the wind velocity drops to a few m/s (ranging from still to \( \sim 10 \) m/s and greater in gusts) and blows in either from a more northerly direction or from Bolivia/Argentina; see the nights Nov. 21, Dec. 2, Dec 15. Those nights also correspond to higher PWV. 

4.9 Mean ground temperature from the APEX weather station. A strong diurnal component is from solar heating and typically produces \( \Delta T \sim 4 \) K between mid-day and nighttime temperatures. 

4.10 Primary mirror panel temperatures over a typical day during the first season (starting at epoch 1195775243) split into the panel groups from 1 to 8, where panel group 1 is at the bottom of the primary and 8 is at the top. The main conclusions of these data are that 1) there is a gradient (though determining its actual value would require a better absolute temperature calibration), 2) cooling occurs over hours, 3) there is abrupt and non-uniform heating in the morning. Mirror fiducial measurements show that while the temperature is dropping over the night, the structure is stable enough for observation (the structure scales, but the twists and gradients that develop during the day have relaxed) several hours after sunset. Here, Chilean local time is \( UT - 4 \) hours. There is significant beam deformation once the sun rises. After the 2007 season, a sun screen was installed that blocks sunlight on the back of the primary mirror structure, reducing the rate of this rise by a factor of \( \sim 2 \) or better. 

5.1 The function \( \Psi(\hat{n}) \) for the 145 GHz camera measured from observations of Saturn. A. Hincks produced this map using the Cottingham basis spline drift estimation method described in Sec. 6.6 for data over a six minute interval. The total solid angle is 225 nsr. The combination of diffraction from the square (but small) pixel and the optical system produces a round beam. 

5.2 The conversion between \( \delta T_{\text{RJ}} \) in Rayleigh-Jeans units to a CMB anisotropy \( \delta T_{\text{CMB}} \) and the difference between the RJ and CMB (intensity) spectral index which in the Wien limit goes as \( \alpha_{\text{RJ}} - \alpha_{\text{CMB}} \rightarrow -1 + x \) where \( x = h\nu/(k_B T_{\text{CMB}}) \). 

5.3 The detector circuit. An absorber receives power from the sky, its temperature increases, and the TES detector responds by increasing in resistance. The resulting current increase changes the magnetic field produced by the inductor that is sensed by the first stage SQUID (SQ1). A digital control loop (running at 15.15 kHz) locks the output of the SQUID amplifier chain by applying a current to the feedback inductor \( L_{fb} \) that compensates changes in the sky signal. Typical values for normal resistances are 30 m\( \Omega \), the bias points are \( \sim 0.3 \) times the normal resistance, and the shunt resistances are 0.7 m\( \Omega \) (for the 145 GHz array in the 2007 season; see Niemack (2008)). The ratio of the mutual inductance between the input inductor and SQUID to the mutual inductance between feedback inductor and the SQUID is \( \sim 8.5 \) and the typical current through a biased detector is 0.3 mA. 

5.4 Left: Responsivity of the 145 GHz array averaged over the 2007 season as calculated by load curves using Eq. 5.37. These are purely electrical and so do not include changes in optical coupling efficiency across the array. Right: the average operating points of the array in units of \( R_s \). Here the target is 0.3\( R_s \), and is constrained by there being only three independent bias groups for columns 0-15 and 31, 16-23, and 24-30 (column 31 was a different series of detector with lower saturation power and lower bias current; see Niemack (2008)).
5.5 Left: The shift in operating point (as a fraction of $R_n$) as a function of optical loading. The traces marked as “model” are solutions for $\Delta R$ given $\Delta P$ from a DC electrothermal model that includes both Joule heating and the qualitative transition shape in Eq. 5.41. The traces marked as “analytic” are from a DC electrothermal model with only Joule heating. The solutions for $\Delta R$ given $\Delta P$ in the analytic case are summarized in Eq. 5.45. A model with only Joule heating cannot predict the saturation behavior, so we apply a simple saturation condition by truncating the new resistance to $R_n$ if it exceeds $R_n$. This is equivalent to there being sufficient optical power to drive the detectors normal. The qualitative effect of the transition model in Eq. 5.41 is to smooth out the junction between the transition and normal regions. There is no analogous saturation for negative $\Delta P$. In the limit that $R_{sh}$ is small, the Joule power can be made arbitrarily high, so a drop in power can be countered by an increase in Joule power. Finite shunt resistance bounds the total Joule power on the detector and so the electrothermal feedback fails to support decreases in power which lead to $R < R_{sh}$. Right: The change in resistance as a function of original operating point. For example, starting at an operating point of $0.5R_n$ (here 0.5 on the x-axis), a 1 pW decrease in power will decrease the operating point to $\sim 0.4R_n$. Adding enough power has the effect of saturating detectors (here, up to $1R_n$) if they are high enough on the transition. A detector high on the transition will only experience a small change in resistance for a given change in power.

5.6 Left: The fractional change in responsivity for a given change in absorbed power as a function of the initial operating point in units of $R_n$ based on the DC electrothermal model including Joule heating and the transition model in Eq. 5.41 and Eq. 5.42. The slowly-varying term is from finite shunt resistance (when the operating resistance becomes comparable) and the rapid “turn-offs” are due to nonlinear response high on the transition. Take the case of a drop in loading. Here, only if the detector is high on the transition (where responsivity is lower due to the low slope there) will the responsivity drastically improve as lower powers bring the detector on-transition. This is represented by the rapid rise on the right side. On the other hand, for sufficient increases in power, any detector can be forced normal. For example, an increase of 4 pW is sufficient to drive the detectors as low as $0.4R_n$ to a region of decreased responsivity. This also indicates that detectors high on the transition (e.g., greater than $0.7R_n$) will be driven normal in response to a planet $\sim 0.5$ pW. This effect is shown by the rapid turn-offs on the bottom half of the plot as detectors of a given operating point saturate. Right: The fractional change in responsivity estimated for the 145 GHz array using this method for a 1 pW drop in absorbed optical power. In the 2007 season, a less than $\sim 0.7$ mm decrease in PWV was typical, or roughly 0.3 pW. The median shift for 1 pW was $\sim 5\%$, so the median detector should change $\sim 1.5\%$ in a night, while extremal detectors might change by $6\%$ or more.
5.7 Here we estimate the change in responsivity (left) and operating resistance (right) following the initial TES biasing due to the change in loading using the model in Eq. 5.45. The loading is inferred from the atmospheric power model described in Sec. 4.3.3 and the change is with respect to the biasing conditions determined by the IV curve at the start of the night. Here, for each TOD we accumulate the number of detectors that fall into each bin. Because the TODs are 15 minutes long, each entry in the bin corresponds to 0.25 hours of integration time. This gives an effective single-detector integration time at that responsivity shift. We then divide this by the (approximately) 900 working detectors in the array to give the time that the array effectively observes at that responsivity shift. The data here represent $\sim 10^3$ TODs across ACT season 1 (2007). The effect of the PWV cuts is counter-intuitive. The bins with a decrease in sensitivity correspond to nights where the loading increased after the initial biasing; while the typical case (where the PWV falls) is represented by improvements in responsivity as the detectors move lower on the transition. The distribution is slightly bimodal because the PWV either increased or decreased, but typically did not stay constant. By cutting on the PWV at the time the file was taken, we can exclude the case where the PWV increased since the beginning of the night, but not the case where the PWV decreased. The PWV during a given file could be low even if the initial biasing happened at 5 mm. Thus a PWV cut removes many of the “core” observations (where the change in responsivity was $< 2\%$) and observations where the weather gets worse, but is ineffective at removing increases in responsivity from the typical case where the PWV drops. Even in the case without PWV cuts, 50% of detectors change in responsivity between $-0.7\%$ and $1.7\%$. Right: The fractional $\Delta R$ across all detectors across all nights. A larger number of detectors decrease in resistance because the PWV typically decreases from the start of the night.

5.8 Timing diagram for the time-domain multiplexed readout. The fundamental unit is the frame, where all columns within a row are read out. After the row delay, the multiplexer biases (addresses) the SQUIDs in that row, and after a feedback delay, a DAC applies the feedback calculated in the previous step to try to compensate the incoming sky signal on the TES loop inductor. After the system settles, the readout card analog to digital converter acquires $N_s$ samples.

5.9 The standard multibit $\Sigma - \Delta$ converter (top) compared to the multiplexed multibit converter such as used by the multi-channel electronics, the “MCE” (bottom). The integrator and quantizer are swapped to facilitate multiplexing. The filtered output has 3.4 $\mu$K per digital unit, and the digital units in the lock loop’s 14-bit feedback DAC are 4.2 mK. The lock loop samples at 15.15 kHz.

5.10 Magnitude of the transfer function of the feedback loop $f_{n+1} = I x_n + (1 - I) f_n$. In practice, the integral term here depends on the slope of the SQUID output chain, and we do not adaptively pick an integral multiplier in the hardware based on these slopes. Because the bandwidth is limited at roughly $< 122$ Hz by the antialiasing filter, this gives significant latitude for choosing the integral multiplier in hardware. We used a value for the integral multiplier that was half what drove the loop into oscillations and did not find that this impacted either the time constants or the lock loop stability, but further work is needed to understand if there are small effects. The anti-aliasing filter strongly suppresses frequencies above 122 Hz and is shown in Fig. 5.11.
5.11 Response of the digital anti-aliasing filter in the multi-channel readout electronics. Incoming 15.15 kHz data from the detectors are filtered, and then downsampled to the output 398.7 Hz, or every 38th output from the digital filter. The DC gain is 1217.9, and must be accounted for in going from filtered feedback units to DAC device units. The phase response is not flat over the frequencies where the sky signal will be modulated. This must be accounted for by an inverse filter in the analysis. Planets excite response up to $\sim 100$ Hz so the phase and magnitude relation from the filter must be undone to find accurate amplitudes.

5.12 Relative calibration (left) and scatter (right) inferred from diffuse atmospheric loading drift over one hour of stare data on Dec. 10, 2007. The left figure represents the relative response in digital units to a diffuse source. This has been normalized so that the average across the array is unity, and the relative response to the diffuse source varies from $\sim 0.7$ to $\sim 1.2$ times that figure. This represents the relative response to diffuse radiation in digital units, so has structure that depends on the electrical properties of the detectors (such as the responsivity, which is strongly correlated along columns) in addition to the optical coupling. The scatter of the relative calibration inferred from these data was sub-percent for many of the detectors, increasing toward the edges to percent-level. (The quantities here represent the fractional $\sigma$ scatter across several stare segments. Column 22 is dead and column 31 was saturated during much of the 2007 season.)

5.13 Left: The ratio of the relative calibration on Dec. 4 and Dec. 10, 2007. The units are terms of the fractional change of the relative calibration between the two dates, and range is over $\pm 2\%$, while the relative calibration of most detectors changes less than 1%. The vertical lines across common rows are due to correlated noise, which has not been removed in this analysis. Middle: the ratio of responsivities as predicted by load curves at the beginning of the night. The units here are the fractional change in the relative response expected from the load curves taken at the beginning of each night. Right: Applying the correction from load curves to the ratio of relative responsivities determined from the atmosphere. This accounts for and removes some of the structure along common columns. In this case, between these two nights, the relative responsivity was stable to better than a percent for most detectors.

5.14 The residual rms across the array from fits to the diffuse atmosphere. Here, each detector is scaled to match the response of column 15 row 15 at the center of the array. The figure plotted here is the standard deviation of the residual from this scaling divided by the standard deviation of the atmospheric signal estimated in column 15, row 15. All channels have been high-pass filtered (Butterworth, order 1) above 0.05 Hz, and downsampled to 1.6 Hz after being smoothed, so the amplitudes represent the total residual power over those frequencies. These show that the atmospheric pickup is coherent over $\sim 10 \times 10$ blocks of detectors, or inside of radii $\sim 10'$.

Delgado and Nyman (2001) find a thick turbulent layer during the day that does not resolve until late night, so the lack of an aspect ratio may be attributable to a much thicker layer. Another approach is to find the zero lag correlation as a function of detector separation. This is similar to binning this result in radius from column 15, row 15, but across all detectors. Fig. 6.6 shows this, and the conclusion is similar.
5.15 Tally of the planetary sources observed by the 145 GHz camera in ACT season 1. Nearly every night has at least one planet observation. We have found that Uranus and Neptune were weak calibrators compared to Saturn and Mars, but were used to determine telescope pointing. Nights with two Saturn observations were typically in the early morning (~3 AM local) and after sunrise. While the source observations after sunrise were useful to study the optical deformation, they could not be used as calibration sources. Mars was observed in an azimuth scan at transit, so only a common sub-array was illuminated, with the exception of several observations where the scan was intentionally offset.

5.16 Dependence of solid angle on the spatial distribution and size of the source. When the solid angle of a source is comparable to the main beam size, the effective beam width and dilution factor for the planet depend in detail on the spatial distribution of surface brightness. Here we compare the fractional difference of the dilution factor assuming a fully-illuminated disk (neglect e.g., limb darkening) and a Gaussian disk illumination as a function of $\bar{\Omega}_b/\Omega_p$. We choose the Gaussian case in particular because this is the one that corresponds to assuming the main beam solid angle is the inferred solid angle in response to a planet minus the solid angle of the planet. Here, in both the diffuse limit (where $\bar{\Omega}_b/\Omega_p$ is small) and the point source limit (where $\bar{\Omega}_b/\Omega_p$ is large) the effect of the source's spatial distribution is negligible. For intermediate sizes the dilution is significantly different. Even if the beam is 10 times larger than the source solid angle, the dilution is $\sim 5\%$ different whether one assumes a disk or Gaussian source. The spatial distribution of planetary sources is necessary to understand the main beam response, but for calibration it suffices to know only the solid angle of the planet and the solid angle of the main beam inferred from observations of that planet, and no further modeling is needed. For the 145 GHz camera, Saturn has $\bar{\Omega}_b/\Omega_p \sim 30$, while Jupiter can have $\bar{\Omega}_b/\Omega_p$ as low as 6 during some times of the year. Solid angles inferred from Jupiter then depend in detail on the surface brightness (and possibly moon positions).
5.17 The ratio of the estimated amplitudes for planet response between two methods, one with a correction for the time constant and one without. The x-axis gives the 3 dB point of the one-pole response from TES bias step time constant fits that have been scaled to be consistent with optically determined time constants as described in Niemack (2008). The points represent all observations of Saturn across ACT season 1. This is not the ratio of amplitudes from the same estimator. Here, one estimator (called Obsfit) is the 2D Gaussian fit developed by B. Reid, J. Fowler and T. Marriage here without a correction for time constants applied and the second model is the two-step fit developed by A. Hincks which fits a 1-dimensional Gaussian to each scan pass. To find the maximum, the maxima from each pass are then fit to a Gaussian (effectively in the Earth-drift direction). In the Hincks model the time constant correction is applied. In addition to showing that the time constants are an important part of the model, it also shows that the amplitude between the two estimators is consistent to several percent. The outliers are due to erroneous observations that are treated differently between the two estimators. (These are cut or rejected in the calibration analysis) A rule of thumb for the correction between the two models is

$$\Theta(f_{3dB}, i) = 1.2 \exp\left(-\frac{f_{3dB}, i}{20}\right) + 1.02$$

(shown as the dashed line here). This indicates that for detectors with $f_{3dB}, i > 70$ Hz, the impact of a time constant on the amplitude of the response to the planet is small ($< 4\%$), and that the results from the Hincks fit are roughly $2\%$ higher than the Obsfit, which in turn suggests some beam asymmetry (Obsfit here was configured to use a symmetric Gaussian, while the Hincks fit uses a separate slice along the scan and Earth-drift direction).

5.18 Temperatures of the planets in the three ACT bands (“dec” denotes 145 GHz, “null” denotes 220 GHz, “inc” denotes 280 GHz) diluted by the ratio of the planet solid angle by the estimated beam solid angle. We use 225 nsr from A. Hincks’ estimates from first-season data, and then scalings to the other bands from Niemack (2008) for beam widths 1.38’, 1.03’, and 0.88’. (Dilution here refers to the ratio of the planet solid angle divided by the 225 nsr main beam.) The temperatures used here are combinations of several values in the literature. For Jupiter, we use Goldin et al. (1997) (their band 1 for 145 GHz/220 GHz and band 2 for 280 GHz). For Saturn, we use Hill et al. (2008) and the geometric model developed in Appendix C.2 corrected for PH$_3$ in the 280 GHz band by the ratio of bands 2 and 1 in Goldin et al. (1997), who had similar bandwidths, to get a very rough estimate of the effect. Mars temperatures are based on Wright (2007b) corrected by the factor of 0.9 given in Hill et al. (2008). We use the model in Griffin and Orton (1993) for Uranus and Neptune, which may not be accurate for the 220 GHz array because of a CO $J = 2 - 1$ resonance at 231 GHz in Neptune’s atmosphere, which is not included in their smooth model. The solid angle used here is the disk area and ignores limb darkening.

5.19 Collected values of the Saturn disk-averaged brightness from literature. WMAP bands are shown in black and ACT bands are shown in dashed blue. The continuous curves are several models in recent literature. The PH$_3$ $J = 1 - 0$ resonance at 267 GHz will be important for the 280 GHz camera, and has been well-constrained; see Weisstein and Serabyn (1994). The spectrum is relatively flat from the upper ACT band to 145 GHz and 220 GHz.
5.20 Left: the geometric model for Saturn’s emission as a function of ring inclination, Right: the effective disk brightness as a function of ring inclination predicted by the geometric model compared to the high inclination model constrained by the WMAP observations. The geometric model has several regimes: for low inclination, the planetary disk dominates the emission, but as the ring tips, it obscures part of the disk emission. By $\sim 12^\circ$, some of the ring is viewed in emission and as it continues to tip, it obscures less of the planetary disk. Above $22^\circ$ inclination, the rings boost emission. The definition of disk average used here is the effective temperature that the disk at zero inclination would have to be to emit as much as the disk/ring system.

5.21 Left: Solid angles of the planets. The telescope’s beam solid angle (used throughout in this work) is taken to be 225 nsr. Right: The temperature of Saturn as predicted by ring models, relative to observations in Hill et al. (2008). (This is just the temperature in Fig. 5.20 except as a function of time, where the wiggle is from the ring inclination through the year.) The model in Hill et al. (2008) departs significantly from the geometric model described here for low ring inclinations and further work is needed to understand the disk-ring system at low inclination. We believe that the temperatures will be closer to a geometric model, because for low inclination, one must always nearly recover the disk temperature. The MJD for the 2007 season was $\sim 54430$, and the difference between the two Saturn models is $\sim 20\%$. We use the geometric model throughout.

5.22 Left: relative calibration inferred from the average of Saturn observations from the 2007 season divided by the relative calibration from the diffuse source. The planet data were calibrated using the response time transfer method described in Sec. 5.9 to propagate it to the best estimate of the responsivity at the time the diffuse atmosphere measurement was acquired. Here we take only observations with good array coverage (> 750 detectors), PWV less than 2 mm, and observations in the survey elevation of $50.5^\circ$. A component of the variation across the array is expected to be from the variation in the dilution factor caused by beam changes across the array. Nonlinear response to the planet may also impact the ratio, biasing the planet response low compared to the diffuse response. The impact would be expected to be largest in columns 5, 24, and 29 (based on Fig. 5.6) but the total power absorbed is $\sim 0.3$ pW, and would produce variations smaller than those shown here. Right: the raw relative calibration inferred from the Saturn observations. Except for the structure in the left panel at the 15%-level, this is similar to Fig. 5.12.

5.23 Left: Averaged response profile of the calibration bolometer across all detectors for the 2007 season. The profile is not described by a simple thermal time constant because the voltage source has finite impedance (so the power delivered changes with the emitter temperature through its resistance as a function of temperature) and the stage that the emitter is connected to rises from $\sim 0.8$ K to $\sim 1$ K during the pulse. Radiative cooling may also play a role. Right: Response of the array to the calibration emitter, averaged over ACT season 1. We use the relative calibration from atmosphere Sec. 5.5 and the array-wide flux calibration from Saturn (Sec. 5.9) to estimate and effective temperature (in CMB units). The illumination of the array was non-uniform and relatively weak because the emitter is mounted on the Lyot stop edge and attenuated by the bandpass before reaching the array. Navy blue squares indicate dead detectors.
5.24 The array-average of the estimate for the calibration pulse amplitude over the 2007 season through several calibration treatment steps. Only the “no treatment” points have error bars for clarity, but the error bars on the points in the two treatment families are similar. These are the estimate of the error on a single pulse measurement. Here, the amplitude estimator is a simple least-squared fit to the template shown in Fig. 5.23, which becomes less accurate over periods such as between Nov. 17 and 25th, 2007, where the knee frequency for atmospheric drift was consistently above 2 Hz, fast compared to the pulse time of 800 ms. The rebiasing correction uses Eq. 5.37 for the load curve analysis to find the change in responsivity from night to night, while the drift correction is based on Eq. 5.45 and attempts to correct the responsivity based on the power estimated from the PWV radiometer. The rebiasing correction removes most of the response shift here, but calibration pulses cannot currently be used to study nights with large changes in PWV, making it hard to assess the efficacy of responsivity drift correction. The planet response amplitude can be determined accurately, so is one of the best tests for the response regime where atmospheric drift and PWV are large.

5.25 Applying the pair statistic to study the stability of the instrument response and time constants. Upper left: A histogram of the relative responsivity shift between detectors at two random times as estimated by load curves. The 25% and 75% quartiles are at 1% indicating that 50% of detectors detectors will shift more than 1% relative to the array at two randomly chosen times. This is a sharply falling function, and almost all detector will have changed less than 2% relative to the array at two different times. Here, 50% of biasing periods (nights) will have absolute shifts > 4% in responsivity across the array. Upper right: the stability of the time constant measured by bias steps over the season. The time constants used here are from a fit by E. S. and R. Fisher, but are consistent with results used elsewhere from J. Appel, that have been calibrated to (and checked against) optical measurements. There is considerable scatter in both the array-relative and array-wide statistics. Lower left: the stability of the TES operating point over the 2007 season. This indicates that even though the absolute responsivity changes by 5 – 10% between days, the operating resistance is stable to a few percent both in absolute and array-relative changes. Lower right: the stability of the SQUID amplifier chain slopes over the season. These show considerable scatter from variability in the tuning and SQUID behavior. Because of the high loop gain in the flux lock loop, these slopes effectively only modify the bandwidth, and become a concern if their bandwidth falls below the anti-aliasing filter bandwidth out to $f_{\text{3dB}} = 122$ Hz. Further work is needed to understand if this impacts the time constants.
5.26 The calibration from filtered feedback digital units to $\mu$K (CMB) as a function of time for one detector with different levels of calibration treatment applied. The simplest is to assume a flat calibration for all detectors at all times; the next level of treatment is to apply the relative calibration from the diffuse atmosphere developed in Sec. 5.5; to account for changes in biasing, we can use the known bias voltages and load curves each night to develop a time transfer through Eq. 5.37; on top of that, the responsivity will drift due to optical loading, and we can apply the correction from Eq. 5.45 given the inferred change in loading from the APEX PWV radiometer; the last layer is to include optical depth, which is also inferred from the PWV radiometer. Note that responsivity drift and attenuation both apply corrections in the same direction. If the PWV falls, the power falls, the operating point falls in resistance, the responsivity rises, and the correction to counter this rising responsivity decreases. If the PWV falls, then the optical depth falls, this causes the signal from the sky to increase, but to counter this increase, the calibration multiplier must decrease. The responsivity correction is roughly $5\% \cdot 0.43 \cdot \text{PWV}$, or $2\%$ per mm, while the optical depth correction is $\sim 3\%$ per mm. These cannot be combined into a single figure of $\sim 5\%$ per mm because the sense of the change is different. The optical depth depends on the instantaneous PWV, while the responsivity drift depends on the change in loading since the TESs were biased (which provides the responsivity reference for the time transfer). The typical number of detectors that can be calibrated is around 820, but this number is sometimes lower because of failures in the load curve analysis or detectors that are driven close to normal or fall below $R_{\text{sh}}$. 

5.27 Top: The distribution of the ratio of all live detector responses over randomly chosen pairs of times with various stages of the global calibration applied. For a given calibration procedure, this represents the scatter in its prediction for the planet amplitudes. The outermost “no treatment” distribution is the distribution of ratios of the raw digital units. The next distribution includes the variation in planet temperature and solid angle; in from that includes the time transfer standard from load curves; in from that adds the responsivity correction from loading; in from that adds an atmospheric attenuation model. Both the responsivity correction and optical depth model use APEX weather station PWV data. The scatter at this level is comparable to the scatter in the array-relative pair statistic — it is near the level of uncertainty in the planet amplitude measurement. Bottom: The calibration error for the conversion from digital units to CMB temperature as a function of the scaling between PWV and optical depth and optical loading. The model that gives the lowest scatter is $\tau = 0.031 \cdot \text{PWV} + C$ and $P = 0.43 \cdot \text{PWV} + C$, which is consistent with the power inferred from the drift in Sec. 4.1 and the ATM atmospheric model (see Pardo et al. (2001)) in Marriage (2006). These are the scalings for the airmass at 50.5° elevation (PWV is through zenith). (Here $C$ is the dry part of the optical depth and absorbed power, which are 0.012 and 0.236 respectively, but are not relevant to the scatter here because they are constant.)
6.1 A region of data from two detectors in the 145 GHz camera demonstrating long-term drift from changes in the atmosphere, burst noise, and calibration pulses. The y-axis gives the calibrated amplitude of the response in CMB temperature units. The upper trace is the detector in column 15, row 15, and the lower trace is column 6 row 6. The long-term drift is largely common to all detectors, but has higher temporal frequency components that have structure across the array, as shown in Sec. 5.14. The burst noise is due to electrical interference across all of the columns and thus correlates all columns in a given row. The period shown above was flagged as having high burst noise. It is also apparent from the two detectors shown here that the strength of the noise varies across the array, so data cuts must be performed on a detector by detector by time basis. The burst noise is broadband, manifests itself as a series of spikes common to all columns in a row, and is not normally distributed. Calibration pulses (the two spikes here) occur at known times and can be masked out. Scan-synchronous pickup produces correlations along all the rows in a column, but is small and so not discernible in this plot.

6.2 The power spectrum of row 6 column 6 over the season. We refer to this as a “waterfall” plot following audio and radio engineering convention. This is from the raw time-ordered data calibrated into power units using load curves and Eq. 5.37. There is a variable $1/f$ knee, which appears as a jagged line at low frequencies, and tracks weather systems that also produce high $PWV$. The noise increases for frequencies greater than $\sim 25$ Hz because of increased detector noise, then the Butterworth filter rolls off the signal $> 122$ Hz. The line around 105 Hz is believed to be due to a ventilation fan mounted on the bottom of the MCE electronics. It operates at around 88 Hz at sea level. This also shows several periods of anomalous noise. November 14, 2007 had wideband noise at $\sim 100$ Hz and $\sim 37$ Hz. This noise returned on December 1 and cause is not currently known.

6.3 The $K^2$ normality statistic/rms plane for all data from the first half of season 1. Above the 5% significance line in $K^2$, the data have a 5% chance of being drawn from a normal distribution. The rms here is in digital units and is the total power out to the bandwidth defined by the Butterworth filter. There is a clear family of noise which does not produce high rms, but which is significantly non-normal. The rms for these intervals lies roughly between 5000 and 5500 digital units, yet their $K^2$ statistics lie above the 0.3% significance line. Each point is the median $K^2$ across the array for a one-minute chunk of data. Further tests weighing the map quality versus the quantity of cut data will be needed, but here we advocate 0.3% significance, or $K^2 < 11.6$. xxx
6.4 The correlation matrix for data taken at 1970 epoch time 1196986141 in column-dominant ordering. The colors represent the zero-lag correlation between the pair of detectors in that row and column. Column-dominant ordering here means that along the axes of the correlation matrix, the indexing iterates through rows then columns. Lines indicate the division between columns, so that a block delineated by the black lines is the correlation matrix between all the rows in the two columns considered. Only live detectors are included. The top heading “column groups” lists the delineation of the columns. In the 145 GHz array, column 22 failed so this does not appear in the column groups heading, column 31 also has a large number of flagged detectors, so its correlation subarrays have different dimensions. The correlation structure here is due to synchronous pickup of magnetic fields by the SQUIDs. Because the phase of this pickup varies across the columns, this can be either correlation or anti-correlation. The field “stdev” given here is the square root of the median along the diagonal and here is given in units of $\mu$K, but can be power, or digital units depending on the input data. (The value here is high because it is the total power, and includes some residual drift and synchronous pickup.) Here, only the array-wide common mode is removed. If it is left in, it will almost entirely correlate all the detectors. The “quality” field is given by Eq. 6.10 and is just the square root of the average of the correlation coefficient of the off-diagonal terms-squared. The synchronous pickup correlations can be removed effectively using dark detectors.

6.5 The correlation matrix for data taken at 1970 epoch time 1196986141 in row-dominant ordering. The data are the same as Fig. 6.4, but this emphasizes the row-correlation structure from electrical interference, which shows up as light banding off the diagonal. This can also be removed reasonably well using information from dark detectors, improving the quality by a factor of $\sim 3$. (Averaging along columns is much more effective, but also introduces correlations in the signal.) The synchronous pickup correlation amplitude is stable over the season (for a given pointing), but the row-correlated “burst noise” is more intermittent.

6.6 The zero-lag correlation matrix for 1970 epoch time 1196811408 (Dec. 4, 2007 stare data) averaged over pairs of a given separation in detector units. This includes several thousand detector pairs at small separations. If none of the array-wide slows drifts are removed, they dominate the correlation and the structure is less apparent. Here we have applied a one-pole Butterworth high pass filter at $0.3$ Hz to emphasize the structure in the atmosphere that can vary across the array. This is analogous to the residuals shown in Fig. 5.14. Again, this is not a rigorous statistic, but it suggests that part of the atmospheric noise is more coherent for proximal detectors because their beams overlap more through the turbulence of the atmosphere.

A.1 The fractional difference in the matter temperature relative to the radiation temperature, $1 - T_m/T_r$. Note the sign: matter is cooler than radiation at late times because of its different adiabatic index ($5/3$ versus $4/3$). During He I recombination for $1600 < z < 3000$, the fractional difference is $< 10^{-6}$.

A.2 The quantity $q - 1$ (defined in Eq. (A.36)), which quantifies the departure from the Sobolev theory due to electron scattering as a function of the electron scattering differential opacity, for several values of the continuum depth for $2^3 P^o - 1^1 S$ at $z = 2500$, which sets the electron temperature.
A.3 Several scales in the recombination plasma relevant for helium peculiar velocities. The fastest here is the plasma frequency, followed by the electron-proton (and then He II-p) momentum transfer rate. Note that the plasma frequency is significantly faster than both the Thomson rate and the frequency of an acoustic oscillation at the Silk scale. This greatly suppresses the magnitude of the charge separation, see Eq. A.37. In the lower plot, we focus on the region neutral helium evolution ($z < 3500$). The charge exchange momentum transfer rates between He and He$^+$ are much larger than the frequency of baryon acoustic oscillations at the Silk scale during the period of neutral helium recombination. This brings neutral helium into a common fluid with the other charged baryons.

A.4 Comparing several numerical convergence issues in the Monte Carlo-estimated $11 \times 21$ (redshift by $x_{\text{HeI}}$) grid of escape probabilities. Both doubling number of photons in the sample and resampling the Monte Carlo give corrections of order $< 2 \times 10^{-4}$. Grid refinement is a more significant systematic, roughly $< 4 \times 10^{-4}$, and indicates that log-log interpolation on the coarser grid over-predicts the escape velocity. Halving the step size in the level code results in error $|\Delta x_e| < 3 \times 10^{-5}$.

A.5 The spectral distortion to the CMB from hydrogen recombination, from Rubiño-Martín et al. (2006). Transitions between the highly-excited states produce ripples in the CMB spectrum into the radio regime. This appendix describes some of the experimental considerations for detecting this distortion.

A.6 The integration time to reach 10 ppb sensitivity ($1\sigma$) for a several cases from pure photon noise from $N_{\text{Planck}}$. We also take $\Delta\nu$ for the detection to be 5% of $\nu$. The rapid rise accounts for the fact that there are very few photons to average in the Poisson/Wien tail. Also note that additional mode coupling is only beneficial in the wave-noise dominated limit and additional detection efficiency is only beneficial once Poisson noise becomes important.

B.1 Front of the CCAM heater controller box. Each heater driver card has a bypass output and a current monitor output. (BNCs on the left side) BNCs on the right half of the box are for auxiliary inputs and outputs, and heat switch outputs. Above each bypass BNC is an LED which indicates whether that heater driver’s relay is turned on. If the relay is on, then the bypass BNC will read the voltage applied to the heater. If the relay is off, the BNC connects directly to the dewar so external voltage supplies can be connected.

B.2 Inside of the CCAM heater controller box. Functions in the controller are divided into a set of vector cards and backplane. From left to right, these functions are, [1]: optoisolation of the 8 pulse width modulation (PWM) outputs from the University of Toronto crate; [2, 3]: 2 filter cards with 4 Avens 30 Hz 8-pole filters each; [4]: a heat switch controller card; [5,6,7,8,9,10,11]: three high power heater controller cards (for warming the adsorption pumps) two low power cards (for controlling the 300 mK and 600 mK stages) one high power card to drive the 4 K optics temperature, and a third low power card to drive the calibration bolometer. To accommodate heater line allocations in the dewar, the front panel DB50 connector going to the dewar is broken out into lines which can be plugged into the appropriate heater controller sockets. These connections are shown on the bottom right with yellow labels.
B.3 Block diagram of the CCAM heater controller systems, for one heater. There are 8 heaters and 12 (8 heater driver + 4 constant voltage heat switch driver) relays controlled by the University of Toronto (UofT) crate card, originally designed for BLAST. The isolator is a single card that handles all 8 incoming pulse width modulation (PWM) signals. It routes the signal to two filter cards containing 4 30 Hz 8-pole low-pass Bessel filters. Each heater driver is on a separate removable card in the vector backplane and its output is programmed by the voltage from the filter card. The BLAST-bus PCI (BBCPCI) card is in the housekeeping computer in the equipment room.

B.4 Pulse width modulation (PWM) driver, isolator and filter which program the heater driver outputs.

B.5 Relay circuit which limits the output of the heater drivers to the dewar and provides a bypass BNC. The bypass provides access to the driver voltage when the relay is energized, and provides access to the heater channel when the relay is off.

B.6 MBAC heater driver interface timing diagram. Note that both load A and load B are asserted and the selection between A and B occurs with the A/B pin when loading the AD8582 input register.

B.7 Layout and channel correspondence of the relay board.

B.8 Layout and channel correspondence of the current monitors on the board.

B.9 Layout and channel correspondence of the switchboard card in the CRBox. See Fig. figs:topcard for and image of the front of the card. The channels with 4 sockets handle distributing the current to the four shared pump heater CBOB lines. One set of two of these is for the driver output, and the other set of two allows the pumps to be powered by external BNC channels 5 and 6.

B.10 The MBAC low power heater, inverting configuration based on the trimmed OP97FS.

B.11 Block diagram of the MBAC high power heater driver. The driver output has a current monitor and a multiplier which convert the current into what the requested voltage should have been for that current to flow. If this exceeds (in the comparator) the actual voltage that was requested (i.e., there is some margin of current more than expected flowing to the dewar), then the lockout latches up and disables the output on the driver's op-amp booster output. This over-current lockout can be disabled by either pushing a switch on the card, or by sending a reset bit from the controller card.

B.12 The $y = mx + b$ scaler to compare the measured current with expected current draw for the heater based on the programming voltage.

B.13 DB50 connector channel allocation standard used throughout (exceptions are specified). Black indicates the signal lines.

B.14 The switchboard card. Channels originating in the heater driver go in the IDC50 ribbon on the left and are broken out (on the back side of this board) to 2-conductor MTE shrouded headers (Tyco 104257-1) on the front. The two CBOB-bound DB50 connectors are broken out by channel, and terminated in 2-conductor MTE receptacles (Tyco 104450-1) which can be connected to the heater driver headers.

B.15 Front of the current monitor card. The monitor/channel correspondence is shown in Fig. B.2.2.

B.16 Front of the relay card. The relay/channel connection is shown in Fig. B.2.2.

B.17 USB pressurized drives developed by Jeff Klein for ACT (left), and an AT-style pressurized drive originally developed for MINT (right). The AT-drive cable splits into separate lines which go to the controller (40-pin IDC ribbon), power, and a pressure, temperature sensor break-out box. Both drives are thermally sunk to the boxes. The AT-style drive was used exclusively for detector data in the first observing season.
C.1  The heliocentric distance of the planets. This is an essential part of the Mars thermal model, but the effect is not yet part of temperature models for the gas giants and is likely to be much more complex. ......................................................... 201

C.2  The viewing phase of the planets. Mars is viewed over the widest range of phases, and its surface temperature depends directly on the Sun’s illumination. The analogous effects are not understood in the gas giants, where the temperature may be more uniform because of mixing. Saturn’s ring emission is also expected to be a function of fraction we observe that is illuminated by the Sun. ......................................................... 201

C.3  Saturn’s ring inclination over the period spanning WMAP and ACT observations. ... 202

C.4  Models and measurements of the disk temperature of Jupiter. The ACT bands are indicated by dashed blue lines. ......................................................... 203

C.5  Models and measurements of the disk temperature of Uranus. .................. 204

C.6  Models and measurements of the disk temperature of Neptune. .......... 205

C.7  A comparison of the WMAP data and millimetric emission model in the ACT pipeline. This is based on the code from Wright (2007b), corrected by Hill et al. (2008) and the modulation is from the viewing phase and heliocentric distance. ............... 206
Chapter 1

Introduction

1.1 Introduction

The field of physical cosmology originated early last century with the emergence of quantitative tools to frame and test broad questions about the universe. Since then, cosmology has developed into a sophisticated branch of physics. We now have a picture of the universe’s expansion from a hot, nearly homogeneous state to the cooler and locally inhomogeneous state we observe today. As the universe cooled through expansion, light nuclei from hydrogen and helium up to beryllium formed. At a later era known as cosmological recombination, these nuclei captured electrons and formed stable neutral species. The optical depth to Thomson scattering then dropped precipitously and the thermal radiation was able to travel almost freely to us today, carrying with it information about the distribution of matter in this early, hot era.

This cosmic microwave background (CMB) radiation was discovered in 1965 by A. Penzias and R. Wilson at Bell Labs (Penzias and Wilson (1965)). The CMB is conspicuous because it is so uniform across the sky and is consistent with the Planck spectrum to a high level of precision (Fixsen et al. (1996)). It was soon realized that departures from isotropy would trace the structure of the early universe and that departures from the Planck spectrum would test thermodynamic assumptions about the era. These ideas were tested by many experiments and then conclusively by the COBE satellite, which detected a temperature anisotropy at one part in $10^5$, and put strong limits on any departure from the Planck spectrum. Today, CMB experiment is a mature field with sound agreement between the observed and predicted anisotropies, and parameters that describe the universe can be measured at the percent level (see e.g., Komatsu et al. (2008)). There is a strong consensus between conclusions from CMB observations and cosmological studies in optical astronomy. Despite this solid footing, cosmological models require two fundamentally unknown components to be compatible with observations: 1) a “dark matter” component which behaves like matter but does not interact (or does so only weakly) with other matter or light, and 2) a “dark energy” component which has negative pressure and causes the expansion of the universe to accelerate. These two components make up 96% of the energy density in the universe, while matter

---

1By number, helium and lithium comprise a fraction $0.08$ and $10^{-10} - 10^{-9}$ of the universe, respectively. The helium abundance is most often cited by mass as $Y_p \approx 0.25$. Combinations of protons/neutrons/helium nuclei with a helium nucleus are not stable, and reactions combining helium and lithium or beryllium are suppressed by the Coulomb barrier. Multi-body reactions that can occur in stars to produce heavier elements (such as triple-$\alpha$ to carbon) are negligible here because of the low density. See Burles et al. (2001); Olive et al. (2000).

2The name “recombination” is inherited from atomic physics (a free-bound transition), but is not accurate here because neutral matter had not previously been stable. A better term is the “neutralization” or “decoupling” era, followed by “reionization.”

3See Harrison (1970); Peebles and Yu (1970); Sunyaev (1978); Zeldovich (1972); Partridge (1995).
### 1.2 The homogeneous universe and the Planck spectrum

The Friedmann-Robertson-Walker metric describes a homogeneous, isotropic space where

\[ ds^2 = -dt^2 + a^2(t) \left( \frac{dr^2}{1-Kr^2} + r^2d\Omega^2 \right), \tag{1.1} \]

we are familiar with can only account for the remaining 4%. In this sense, physical cosmology has become not only a precise, but also a revolutionary science.

The Wilkinson Microwave Anisotropy Probe (WMAP) (Hinshaw et al. (2008)) provides the best limits on the CMB temperature anisotropy over the sky to \( \sim 0.3^\circ \) resolution. In conjunction with other surveys, these data give sharp limits for the cosmological model parameters (Komatsu et al. (2008)). Fig. 1.1 summarizes recent measurements. Several recent CMB experimental efforts aim to measure the intensity anisotropy on scales much smaller than those measured by WMAP. Small-scale intensity anisotropy measurements are well-motivated because they contain additional information from the primary CMB, from galaxy clusters through the Sunyaev-Zel'dovich (SZ) CMB spectral distortion (see e.g., Carlstrom et al. (2002)), and possibly from other intervening matter through weak lensing of the CMB (see e.g., Lewis and Challinor (2006)). Measurements in this regime are expected to improve constraints on the primordial perturbation power spectrum, the dark energy equation of state, neutrino mass, large scale structure, the kinetic SZ effect, the Ostriker-Vishniac effect, and homogeneous cosmological parameters such as the baryon density (Kosowsky (2003)). This chapter introduces the essential physics of the phenomena described in the rest of the thesis, and focuses on small-scale intensity anisotropies.

![Figure 1.1: Left: The angular power spectrum of the CMB on large scales (Nolta et al. (2008)). Right: The angular power spectrum of the CMB on small scales (Dunkley et al. (2008)).](attachment:image.png)
and \(d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2\) is the distance element on the sphere and \(K\) is the spatial curvature.\(^6\) Applying Einstein’s equation then gives Friedmann’s equations for the evolution of the scale factor as a function of time (defining the Hubble constant \(H(t) = \frac{da}{dt} a^{-1}(t)\)):\(^6\)

\[
H^2(t) = \frac{8\pi G}{3} \rho(t) - \frac{K}{a^2(t)} \tag{1.2}
\]

\[
\frac{d}{dt} \left( \frac{a(t)}{H(t)} \right) = -\frac{4\pi G}{3} [\rho(t) + 3p(t)], \tag{1.3}
\]

\[
\frac{d^2}{dt^2} \left( \frac{a(t)}{H(t)} \right) = -\frac{4\pi G}{3} [\rho(t) + 3p(t)], \tag{1.4}
\]

where \(G\) is Newton’s constant, \(\rho\) is the density,\(^7\) and \(p\) is the pressure. Time dependence is indicated here, but will henceforth be taken implicitly. Units here take \(c = 1\). A photon traveling in this space will redshift so that \(\lambda_{\text{observed}} = a^{-1} \lambda_{\text{emitted}} = (1 + z) \lambda_{\text{emitted}}\), defined this way so that \(z = 0\) today. The matter density simply tracks the volume expansion so that \(\rho_m \propto a^{-3}\), while radiation also decreases in energy through redshifting as the universe expands so that \(\rho_r \propto a^{-4}\). For a fluid with equation of state \(w = p/\rho, \rho \propto a^{-3(1+w)}\). In addition to matter and radiation, we also parameterize dark energy as a component with \(w = -1\), which will be denoted throughout as a subscript \(\Lambda\). It is convenient to divide densities by the critical density \(\rho_c \equiv 3H_0^2/(8\pi G)\) (throughout, nought denotes the value today) to define \(\Omega_i \equiv \rho_i/\rho_c\). Then trading time for redshift,\(^6\)

\[
H(z) = H_0 \sqrt{\Omega_\Lambda + \Omega_K(1 + z)^2 + \Omega_m(1 + z)^3 + \Omega_r(1 + z)^4}, \tag{1.5}
\]

where we have identified \(\Omega_K = 1 - \sum_i \Omega_i = -K/H_0^2 \approx 0\) and \(H_0^{-1} \approx 13.8\) Gyr. The time between two scale factors/redshifts can be integrated either as the coordinate time or the conformal time \(\eta = dt/a(t)\) through

\[
t = \int \frac{da}{aH} = \int \frac{dz}{(1+z)H}, \quad \eta = \int \frac{da}{a^2} = \int \frac{dz}{H}. \tag{1.6}
\]

Before \(z \sim 3200\), radiation dominated the dynamics of the universe because of its steeper dependence on \(a\). After this era, matter dominated, and more recently, dark energy dominates. Thus, the scaling at early times is \(a \propto t^{1/2} \propto \eta\) and in matter domination it is \(a \propto t^{2/3} \propto \eta^2\). For dark energy domination \((w = -1)\) the scale factor expands exponentially, or \(H \propto a \propto t\). The radiation temperature at all times considered here scales as \(T_r(z) = T_CMB(1 + z)\), where \(T_{CMB} = 2.725 \pm 0.002\) at \(2\sigma\); see Fixsen et al. (1996); Mather et al. (1999).\(^8\)

The CMB radiation follows a Planck distribution so closely because of the dense, homogeneous, and ionized conditions of the early universe. The establishment of the Planck distribution is a two-part process. One mechanism must thermalize emitted photons to the Bose-Einstein distribution (conserving photon number), and another must drive the chemical potential of the Bose-Einstein distribution to zero (modifying the photon number). Compton scattering drives photons to the Bose-Einstein distribution, while the radiative processes Bremsstrahlung and double Compton scattering drive the chemical potential of the Bose-Einstein distribution to zero (see Danese and Burigana (1994); Hu and Silk (1993); Lightman (1981)).\(^9\)

\(^6\)Observational evidence indicates that the spatial curvature is nearly zero, or equivalently that the universe is consistent with having critical density; see Dunkley et al. (2008).

\(^7\)This is tied to the energy density, but is conventionally denoted as \(\rho\).

\(^8\)The absolute temperature of the CMB today is known primarily from careful radiometric studies, but some of the earliest temperature measurements were inferred from from equilibria of molecular species. See Losecco et al. (2001); Roth et al. (1993) for reviews. The evolution of the matter temperature is more complicated but necessary to calculate accurate recombination histories, and is described in Appendix A.1.

\(^9\)The collision term (Uehling and Uhlenbeck (1933)) in this regime that describes how a photon with frequency \(\omega\) scatters
1.3 Departure from homogeneity and isotropy

Primordial distortions to the Planck spectrum are small because of the large number of photons per baryon\(^{10}\) and the high interaction rates which re-establish the Planck spectrum for much of the early history of the universe. The last dramatic event was the freeze out of \(e^+ + e^- \leftrightarrow 2\gamma\), when a large number of 511 keV photons were produced at around \(z \sim 10^9\). This radiation thermalized within 2 months (see, e.g., Partridge (1995)).

### 1.3 Departure from homogeneity and isotropy

Until the era of recombination, the universe was optically thick (per Hubble time) to photon-electron scattering. This implies that matter and radiation rapidly exchanged momentum, driving differential flows to zero. As the universe became neutral, the momentum exchange rate dropped and the photons and matter no longer moved in unison. By the end of recombination, the photons were able to travel almost unimpeded to reach us today as the CMB. This section sketches the physical processes which produce the CMB that we see today.\(^{11}\)

There are many possible species and pairs of interactions to consider in the early universe: electrons, ions (protons, doubly and singly ionized helium), neutral species (hydrogen and helium), neutrinos, photons, and the dark matter. The general Boltzmann equation for the evolution of the phase space density of these species in a curved universe would be nearly intractable if there were not three factors that significantly simplify the problem. The main simplification comes about because the universe was homogeneous to one part in \(10^4\) during this era, so perturbations to a uniform phase space density and smooth background space are small and can be calculated in linear theory. The second simplification is that the matter species in the early universe (excluding neutrinos) interact rapidly relative to the Hubble rate and so exchange momentum efficiently. This drives velocity differences between the species to zero, so they move in unison and can be considered as a lumped baryon field.\(^{12}\) The third simplification is that higher fluid moments are.

\(^{10}\)Many photons have to be added to make any difference. At \(z = 1100\), there are \(\sim 2 \times 10^9\) photons per baryon.

\(^{11}\)The full derivations in literature are often involved, so here we have tried to report what is essential to give some intuition for the effects even if some numerical factors cannot be understood outside of a complete calculation. Footnotes explain technical considerations. Hu (2008) gives a comprehensive review of the theory of CMB anisotropies.

\(^{12}\)Momentum exchange from Coulomb scattering between the charged species is rapid, so these species can be lumped together to a good approximation. The neutral-neutral and neutral-ion momentum exchange rate is much lower but is also sufficient to drive a common population; see Hannestad (2001). A local photon dipole can push electrons more easily than protons. This can produce small currents and magnetic fields in the early universe (see Siegel and Fry (2006b,a) for development), but the effect is small (giving fields of order \(10^{-24}\) \(G\)) because of the high momentum transfer rate from Coulomb scattering.
suppressed by the high interaction rates. The radiation distribution is then described to an excellent approximation by its monopole, dipole and quadrupole moments. These factors permit a calculation of the evolution of species, radiation, and metric perturbations with high accuracy.

The essential components of a theory of CMB anisotropies are the continuity equation and an Euler equation to account for momentum exchange between the fluids. The continuity equation is \( \dot{n}_b + \nabla \cdot (n_b \mathbf{v}_b) = 0 \), where subscript \( b \) throughout will denote baryons and the overdot denotes derivatives with respect to comoving time coordinates. We can then convert to \( k \) space where \( \nabla \rightarrow i k \), and define \( \delta_b \equiv \delta n_b / n_b - 1 \) and \( \theta \equiv ik \cdot \mathbf{v} \). The continuity equation is then simply \( \dot{\delta}_b = -\theta_b \). \( \theta_b \) is analogous to a momentum, and its evolution is determined by interactions that exchange momentum with components of the baryon fluid, giving a term analogous to a force law, \( \theta_b = \text{interactions} \).

The momentum exchange between baryons and photons from Thomson scattering is

\[
\dot{\theta}_b = \frac{\hat{\tau}}{3 \rho_b} (\theta - \theta_b). \tag{1.14}
\]

The pre-factor is the momentum exchange rate \( \nu_{b,\gamma} = \hat{\tau}/R \) between baryons and photons, and is the Thomson optical depth \( \hat{\tau} = an_c \sigma_T \) divided by the ratio of the density of baryons to photons, \( R = 3 \rho_b / (4 \rho_\gamma) \) (see Eq. 1.13 for a numerical approximation). Large \( \hat{\tau} \) drives \( \theta_b - \theta \rightarrow 0 \) in the limit of high optical depth to Thomson scattering – in this limit photons and baryons move as a single fluid. The equations governing the evolution of photons are similar, except that here, the effect of momentum exchange between the photons and baryons decouple. (Note that \( \delta_p/\delta \rho = w \).) The sound speed of baryons, and is small compared to \( c_s \), for photons and the baryon-photon plasma by virtue of \( \delta_p/\delta \rho = w \). The \( k^2 \) is due to the fact that density gradients drive velocities (one \( k \)) and that \( \theta \) is the gradient of those velocities (another \( k \)).

\[
\dot{\delta}_b = -\theta_b \quad \dot{\theta}_b = -\frac{\dot{a}}{a} \theta_b + c_s^2 k^2 \delta_b + \text{interactions}, \tag{1.10}
\]

where \(-\frac{\dot{a}}{a} \theta_b \) is due to Hubble expansion and is important especially on large scales, but we will ignore it here to get a qualitative result. The term \( c_s^2 k^2 \delta_b \) is due to photon pressure which is subdominant to photon pressure, but becomes a consideration when the photons and baryons decouple. (Note that \( c_s \) is the sound speed of baryons, and is small compared to \( c_s \) for photons and the baryon-photon plasma by virtue of \( \delta_p/\delta \rho = w \).) The \( k^2 \) is due to the fact that density gradients drive velocities (one \( k \)) and that \( \theta \) is the gradient of those velocities (another \( k \)).

The baryon-photon ratio is given conveniently by (see Hu (2008))

\[
R \approx 0.6 \left( \frac{\Omega_b h^2}{0.02} \right) \left( \frac{a}{10^{-3}} \right). \tag{1.13}
\]
1.3 Departure from homogeneity and isotropy

write

\[ x = (\delta, \theta, \gamma)^T \]
\[ \dot{x} = \begin{pmatrix}
0 & 0 & -4/3 \\
0 & -\dot{\tau}/R & \dot{\tau}/R \\
k^2/4 & \dot{\tau} & -\dot{\tau}
\end{pmatrix} x. \] (1.16)

The eigenvalues of the matrix can be understood as coming from two sources. If \( \dot{\tau} \) is negligible, then one recovers the dispersion relation in a photon gas \( \omega = k/\sqrt{3} \), where the factor of \( \sqrt{3} \) is from the equation of state for radiation, \( w = p/\rho = 1/3 \). The eigenvalues are imaginary, so these represent oscillatory solutions.\(^{17}\) If \( \dot{\tau} \) is large, then the dispersion relation becomes \( \omega = c_s k \) where

\[ c_s^2 = 1/[3(1 + R)]. \] (1.17)

In this case, the photons are coupled to the baryons and the sound speed \( c_s \) is reduced by having to drive additional mass. See Eq. 1.13 for numerical values of \( R \) during this era.

So far, we have described the dynamics of perturbations in Fourier space as wave phenomena. It is also informative to consider the evolution in configuration space, starting with an overdense patch where all the matter is perturbed the same amount (up to corrections of \( 4/3 \) for the relativistic species). These are adiabatic initial conditions. The neutrinos do not interact with any of the species significantly here, so they stream out of the over-density.\(^{18}\) Because the photons exchange momentum so readily with the charged baryons through Thomson scattering at this time, they continue to evolve as one fluid. The baryon-photon perturbation expands as a wave with \( \omega = c_s k \) (much like a wave from a rock thrown into a lake). The comoving distance over which the perturbation has expanded at a time \( \eta \) is the sound horizon

\[ r_s(\eta) = \int_0^\eta d\tilde{\eta} c_s(\tilde{\eta}). \] (1.18)

When neutral species form in the baryon fluid, \( \dot{\tau} \) rapidly drops and the photon perturbations propagate independently of the baryons with dispersion approaching \( \omega = (1/\sqrt{3})k \). The expansion of the baryon perturbation then slows drastically, without the momentum exchange from the expanding photon distribution. Thus, the propagation distance of a baryonic perturbation from the initial disturbance is the sound horizon evaluated at the recombination era, which is about 140 Mpc today. This is the physical space view for a lone perturbation described in Bashinsky and Bertschinger (2002) and this length scale has been observed in the correlation of galaxies (see Percival et al. (2007)).

In Fourier space, for perturbations at a physical scale \( k \), \( \delta \) oscillates roughly as \( \cos(kr_s) \) (for adiabatic initial conditions where there are initial overdensities, but zero velocity) and \( \theta \) oscillates as \( \sin(kr_s) \) (their eigenvalues from Eq. 1.16 are imaginary). The oscillatory behavior of \( \delta \sim \cos(kr_s) \) has an interesting interpretation. For a fixed \( k \) scale on the sky, \( \delta \) will oscillate as \( r_s \) increases. This represents a series of compressions and rarefactions of the baryon-photon gas set up by initial conditions of the perturbation. For high \( k \) (or small length scales) the gas is able to execute a larger number of oscillations before the recombination era. This is the origin of the acoustic peaks in the CMB. To see how this is the case, consider the decomposition of the temperature anisotropy we see today into spherical harmonics:

\[ T(\theta, \phi) = \sum_{l,m} a_{lm} Y_{lm}(\theta, \phi), \] (1.19)

\(^{17}\)If we had retained the baryon pressure term, we would recover waves in the baryon gas at the sound speed.

\(^{18}\)Much later in the universe, neutrinos again participate with the large scale structure of matter. See Abazajian et al. (2005).
and define the angular power spectrum (averaging over $m$ because of statistical isotropy)\textsuperscript{19}

$$C_{\ell}^{\ell} = \langle |a_{\ell m}|^2 \rangle \Rightarrow T_{\ell}^2 = \frac{1}{2\pi} \ell(\ell + 1)C_{\ell}^{\ell\ell}. \tag{1.20}$$

(This normalization is standard; see Fig. 1.1.)

A rough approximation for the sound horizon is to take a sound speed of $1/\sqrt{3}$ (ignoring baryonic effects) so that $r_s \sim \eta_s/\sqrt{3}$, where $\eta_s$ is the comoving distance to recombination. Then, the oscillations $\cos(kr_s)$ are extremal for $k \eta_s = n \pi$, or $k \eta_s \sim n \sqrt{3} \pi / \eta_s$, where $n > 0$ is an integer. Hu (2008) shows that this wavenumber relates to approximately a multipole $\ell$ today\textsuperscript{20} of $\ell \approx \eta k$ as we view it today over the distance $\eta$ in a flat universe. The multipole moment of the extrema at $n$ corresponds to

$$\ell \approx n \sqrt{3} \pi \eta \eta_s^{-1} \approx 180 n, \tag{1.21}$$

where we have used the fact that in a matter dominated universe $\eta \propto a^{1/2}$ and that the recombination era corresponds to roughly $z \approx 1100$. Note that the condition $kr_s = n \pi$ includes compressions and rarefactions, which both appear as increases in the angular power spectrum. This rough figure agrees reasonably well with the experimental and complete theoretical calculations shown in Fig. 1.1. Thus the angular power spectrum shows ringing where the largest scales $\ell \sim 180$ have compressed only once before recombination, and smaller angular scales on the sky have gone through several cycles.\textsuperscript{21}

So far we have only considered only the photon monopole and dipole and found undamped wave solutions. The origin for the higher photon moments is the conversion of inhomogeneities to anisotropies. To see this, consider a temperature inhomogeneity with spatial wavelength $\lambda$. If the photon mean free path is much smaller than $\lambda$, only radiation from a little local patch the size of the mean free path is visible, giving a uniform $Y_{0,0}$ and a gradient over that patch, as $Y_{1,m}$. As the mean free path increases, the local patch will start to develop a $Y_{2,m}$ anisotropy. This transports photons (and momentum) perpendicular to the flow associated with the $Y_{1,m}$ moment and is analogous to viscosity in a gas, where once the mean free path is great enough, random motion in the gas will transport momentum out of a bulk flow.\textsuperscript{22} In addition, as $\tau$ drops, some slip can develop between $\theta_\parallel$ and $\theta_\perp$, and the eigenvalues develop a damping component as momentum of the baryon-photon fluid is translated into a difference in momentum between the baryon and photon fluids. The modification to the dispersion relation to first order in $k/\tau$ (using both polarizations) is

\textsuperscript{19}The $C_\ell$’s are $\chi^2$ distributed with $2\ell + 1$ degrees of freedom so that $C_\ell$ has variance $[2/(2\ell + 1)]C_\ell^2$. This puts a fundamental limit on how well the spectrum can be measured by one observer, and is $1/f_{sky}$ larger for a partial survey with coverage $f_{sky}$. The large number of modes available on small scales gives experiments a significant amount of latitude to observe smaller regions (e.g., from the ground) and still reach high precision. The angular scale on the sky for a given $\ell$ is $\theta \sim \pi/\ell$, and a Gaussian beam with width $\sigma$ transforms to $\exp[-\ell(\ell + 1)/\ell (\ell + 2) \sigma^2]$ in $\ell$ space. This means that the region of $\ell$ that a given telescope can study is set by its diffraction limit.

\textsuperscript{20}A simple calculation is to take $\ell \approx \pi/\theta$ where $\theta = \lambda/\delta = \lambda\eta = 2\pi/(k\eta)$.

\textsuperscript{21}In addition to the density, the temperature anisotropy we view today also depends on the velocity of the region (also called a Doppler term), as there is a fractional change in temperature $\delta T = n \cdot \nu$, that depends on the viewing direction dotted into the velocity. Therefore the term $\theta_\parallel$ which we describes as having a $\sin(kr_s)$ dependence also adds anisotropy power that we observe today that is $90^\circ$ out of phase with the density perturbation. It also has a more complex coupling to the observable CMB because of the angular dependence; see Hu (2008). The temperature anisotropy is dominated by the $\cos(kr_s)$-like density perturbations. The Doppler contribution fills in power between the density perturbation peaks.

\textsuperscript{22}In reality, all the Legendre moments $\ell > 1$ of the Boltzmann equation should be solved to find accurate anisotropy power spectra (see Ma and Bertschinger (1995)). In the notation of Eq. 1.16 it is a large matrix that tracks the evolution of each moment. These are in general suppressed by factors of $\tau$ for each higher $\ell$ moment, so become small.
1.4 The Sunyaev-Zel’dovich effect

Kaiser (1983),

\[ \omega = c_s k \left[ 1 \pm \frac{i k c_s}{2 \dot{\tau}} \left( \frac{16}{15} + \frac{R^2}{R + 1} \right) \right] \]  

(1.22)

where the first term is from the photon quadrupole and the second is from heat conduction. Because of the imaginary part, this additional term damps the acoustic oscillations. The amplitude of a perturbation on scale \( k \) is damped by \( \exp(-k^2/k_D^2) \) when we observe it, where the characteristic damping scale can be found by integrating the damped motion to yield

\[ k_D^{-2} = \int d\eta \frac{1}{\dot{\tau}} \frac{1}{6(1 + R)} \left( \frac{16}{15} + \frac{R^2}{R + 1} \right). \]  

(1.23)

This produces the so-called “damping tail.”

Aside from the numerical factors, the fundamental piece here is \( 1/\dot{\tau} \), which depends on \( a(t) \) and \( N_e(t) \). While the evolution of \( a(t) \) is determined by the homogeneous cosmology, the evolution of \( N_e(t) \) is determined by the evolution of the free electron fraction, which depends in detail on atomic processes as neutral species formed. The largest known uncertainty in theoretical predictions of the primary CMB on small scales is due to the recombination history for \( N_e(t) \). In Chapter 2, we examine just the helium recombination history in detail and find significant corrections to \( N_e(t) \) relative to previous studies.

The process described here is sensitive to cosmological parameters. If there are more baryons, the sound speed decreases (see Eq. 1.17) and the relative importance of pressure (photons) and gravity (baryons) shifts, so the compressions strengthen and the rarefactions weaken. Also, the number density of electrons increases, changing the damping tail. Small-scale anisotropy measurements provide additional sensitivity to the baryon density through these mechanisms. The other relevant cosmological parameter on small scales is the spectral slope of primordial perturbations. This slope in an indicator for inflation or other related processes that occurred in the early universe; see Kinney (2003). The cosmological information from the damping tail is largely degenerate with the recombination history and the window function of the experiment. For this generation of small-scale anisotropy experiments it is therefore important to accurately characterize the beam pattern of the telescope (which determines the window function in \( \ell \)), and to have confidence in the recombination calculation.

1.4 The Sunyaev-Zel’dovich effect

In Chapters 3, 4, 5, and 6 we describe the Atacama Cosmology Telescope (ACT), which has arrays sensitive to 145 GHz, 220 GHz, and 280 GHz radiation. These bands were chosen to coincide with good atmospheric transmission and a spectral distortion of the CMB that is produced as the photons traverse the hot electron gas in galaxy cluster. This distortion was identified by Sunyaev and Zeldovich (see Sunyaev and Zeldovich (1972)) and is caused by the net energy boost of CMB photons as they inverse Compton scatter from the hot electrons. By mapping the...
sky in these frequencies, ACT will produce an catalog of galaxy clusters. This catalog (paired with optical and x-ray studies) may provide additional information about the organization of matter on large scales and the accelerated expansion of the universe by dark energy.

A full calculation of this spectral distortion is involved, but the frequency of the null (which defines the ACT central band at 220 GHz) can be derived quickly. A photon which interacts once in a galaxy cluster will have some probability distribution for scattering from $\nu$ to $\nu'$. This distribution has significant width because of the Doppler effect from scattering off thermal electrons at $\sim 10^7 - 10^8$ K. It also has a non-zero mean, where some small amount of energy is transferred to the photons. It is only the case on average that a photon’s energy is boosted (the secular term). Most of the photons that interact with the cluster simply “rattle around” (a diffusive term). The phase space density of outgoing photons after a single scattering event is just the convolution term). Most of the photons that interact with the cluster simply “rattle around” (a diffusive term).

To zeroth order $e^{-x}$ cancels from both sides. To zeroth order $e^{-x}$ is zero and $x = 4$. The outgoing phase space density is equal to the incoming phase space density at a “null” frequency (in units of $x = h\nu/(k_BT)$):

\[
\text{Distortion Null : } \mathcal{N}(x) = \int_{-\infty}^{\infty} P(x' - x)\mathcal{N}(x')dx'.
\] (1.24)

We can take $\mathcal{N}$ to be the Planck distribution and expand about $\Delta x = x' - x$. Terms in the expansion are the moments $\langle \Delta x^n \rangle$ of $P(\Delta x)$ and are negligible for $n > 2$ here, giving the null condition for the scattering kernel are $\langle \Delta x \rangle = 4xk_BT_e/(mc^2)$ and $\langle \Delta x^2 \rangle = 2x^2k_BT_e/(mc^2)$ (see Sazonov and Sunyaev 2000) and expanded to first order in small $e^{-x}$. Note that this does not depend on the electron temperature, which cancels from both sides. To zeroth order $e^{-x} = 0$ and $x = 4$, while the numerical solution is $x = 3.83$. Using the CMB temperature, this is 217 GHz. For frequencies lower than the null, there is net scattering out, producing a temperature decrement, while for frequencies above the null, there is a net scattering in, producing a temperature increment. The position of the null is robust to the electron temperature (for gas temperatures $< 511$ keV) and independent of redshift. ACT should therefore be able to cleanly separate the SZ component.

### 1.5 The Atacama Cosmology Telescope

The Atacama Cosmology Telescope (ACT) is a 6 m telescope designed to map the intensity (and ultimately the polarization) anisotropy of millimeter radiation with arcminute resolution over $\sim 1000$ square degrees. The primary ACT camera, the millimetric bolometer array camera (MBAC), comprises three arrays that span the SZ distortion’s decrement, null, and increment, centered at 145 GHz, 220 GHz, and 280 GHz, respectively. Each band’s detectors are arrayed into continuing to grow (see Carlstrom et al. 2002), which means that clusters $\sim$ Mpc will typically be larger than one arcminute, which coincides with the ACT resolution.

\[27\text{For reviews and technical aspects of the calculations, see Carlstrom et al. (2002); Sazonov and Sunyaev (2000); Dolgov et al. (2001); Burigana (2007); Burigana et al. (2004); Sunyaev and Zeldovich (1998); Rephaeli (1995a); Birkinshaw (1999) and for cluster cosmology methods and simulations, see Pierpaoli et al. (2005); Sehgal et al. (2005, 2007).}

\[28\text{The Kompaneets equation describes transport using the first two moments, and there are several higher-order treatments; see Itoh et al. (1998); Challinor and Lasenby (1998); Rephaeli (1995b) and Sazonov and Sunyaev (2000) for a calculation based explicitly on the higher scattering moments. The scattering probability distribution can be calculated by solving the kinematics for a single electron-photon scattering and then finding the distribution of those for the thermal struck particle distribution. We solve this in the Thomson limit in Appendix A.3.}

1.5 The Atacama Cosmology Telescope

![Image](236x606 to 415x699)

Figure 1.2: The ACT 145 GHz array. The \(32 \times 32\) array is the central square inside of the copper carrier. The carrier and the blackened cavity surrounding it are held at 0.3 K by a \(^3\text{He}\) adsorption refrigerator. The array comprises a stack of 32 silicon cards, each of which has 32 folded “pop-up” detectors. (Here “pop-up” refers to the fact that the absorber is perpendicular to the silicon card carrier, and is held up by weak thermal link legs.) Each detector unit is a \(\sim 1\,\text{mm} \times 1\,\text{mm} \times 1\,\mu\text{m}\) square of boron-doped, ion-implanted silicon held up from a silicon carrier card by four \(\sim 1\,\mu\text{m} \times 5\,\mu\text{m}\) cross-section legs which conduct electrical and optical power from the detector to the bath. These are bonded to SQUID multiplexing and biasing electronics on the silicon carrier card, which are read out and addressed through the ribbons on the left. The array face is covered by a \(50\,\mu\text{m}\) silicon AR coupling layer, separated from the detectors by \(100\,\mu\text{m}\), but not shown here. (Photo: M. Niemack)

a \(32 \times 32\) square grid of free-space transition edge superconducting (TES) detectors\(^{30}\), cooled to 0.3 K using a \(^3\text{He}\) adsorption refrigerator (from a \(^4\text{He}\) adsorption refrigerator that is cooled by a pulse tube cryorefrigerator.) Measurements across these three bands permit a discrimination between foreground emission sources, galaxy clusters (through the SZ effect), and the primary CMB (see e.g., Kosowsky (2006)). These bands have also been selected to fall within millimeteric transparency windows in the atmosphere between oxygen (119 GHz) and water (183 GHz). These windows can only be exploited in a dry atmosphere at high altitude, where contamination from atmospheric structure and opacity can be reasonably suppressed. To minimize atmospheric contamination, ACT operates from Cerro Toco at 5190 m in the dry Chilean Altiplano.

The 145 GHz array is shown in Fig. 1.2. The absorber is an approximately \(1\,\text{mm}^2\) square of ion-implanted, boron-doped silicon\(^{31}\) and the temperature changes are read through transition-edge superconducting (TES) detector.\(^{32}\)

We scan the telescope azimuth at a constant angular velocity of 1° per second over 10° peak-to-peak with 400 ms turnarounds\(^{33}\) to modulate the celestial signal above the \(1/f\) noise from the atmosphere, while not exceeding the detector time constants.\(^{34}\) Observations take place at 50.5° elevation, so this becomes \(\sim 0.64\,\text{°}\) per second on the sky. The overall scan strategy is to map two to three regions by scanning at fixed elevation and central azimuth as those regions rise and set. The diffraction-limited scale on the sky for the 280 GHz array (where the condition on the sampling rate is most stringent) has \(D\lambda^{-1}/1.22 \sim 4440\,\text{cycles/radian}\), with a temporal frequency

---

\(^{30}\)These bolometers are fabricated at NASA Goddard and the band-defining filter stack is from Cardiff University.

\(^{31}\)The carriers in boron-doped silicon are captured at the operating temperature, so it becomes an insulator. If ions are implanted in the lattice, phonon-assisted tunneling can transport charge between impurity sites; see McCammon (2005).

\(^{32}\)The TES is a molybdenum/gold bilayer that is \(80 \times 75\,\mu\text{m}\), or 0.5% of the absorber area. See Marriage (2006) for a description of the TES parameters. Molybdenum alone has \(T_c \sim 1.1\,\text{K}\). The superconducting proximity effect between the normal gold and superconducting molybdenum gives \(T_c \sim 510\,\text{mK}\) in the first season 145 GHz array. The normal resistance across the bilayer is set by the gold thickness, and was \(R_n \sim 30\,\text{m}\Omega\) for 145 GHz. Additional normal metal (gold) meander bars on top of the bilayer decrease the slope of the superconducting transitions and improve the detector performance. See e.g., Marriage (2006).

\(^{33}\)These are the scan parameters for the 2007 season, which is the subject of analysis in this thesis.

\(^{34}\)For the 145 GHz array, the typical detector response function is described by \(f_{3dB} \approx 90\,\text{Hz}\), with significant scatter. The other two arrays are designed to have typical \(f_{3dB}\) higher in the ratio of their beam sizes, but have not yet been characterized in the field.
\( \omega D \lambda^{-1}/1.22 \sim 50 \text{ Hz} \). Here, \( D \) is the primary mirror diameter and \( \lambda \) is the wavelength. We oversample the beam by a factor of slightly greater than \( \sim 2 \) (and higher in the lower two bands) by filtering the array data to have a \( f_{3dB} = 122 \text{ Hz} \), which is conservatively sampled and stored at 399 Hz. This sets the data rate of the experiment.

ACT observing seasons are naturally separated by the seasons at the site, where moisture from Bolivia makes measurement challenging from late December to March. Science data from the first season of ACT were acquired by the 145 GHz array from November 14 to December 17, 2007, in dedicated observation. This band was chosen because galaxy clusters are most identifiable in the SZ distortion decrement, the atmospheric opacity is lower, and the atmospheric loading and telescope alignment constraints are more relaxed than for the other two bands.

At the time of this writing, ACT has data from one month of observations with the 145 GHz array from 2007, and the full three band 145 GHz, 220 GHz, and 280 GHz camera became operational in August, 2008. It is expected to acquire data until late December, 2008 and then resume in the spring.

### 1.6 Sensitivity limits for millimeter wave detection

The accuracy of a bolometric CMB experiment is limited by the sources of noise in 1) the bolometer itself (phonon noise), 2) the readout (SQUID, Johnson noise, aliasing, readout) and 3) the radiation incident on the detectors.\(^3\)

Here we review phonon and photon noise to motivate the design of the ACT array.

For an absorber with some heat capacity \( C \) connected to a bath through a link with thermal conductance \( G \), there will be fluctuations in the absorber’s temperature from fluctuations in energy as phonons are exchanged with a thermal bath. The variance in the energy for a set of states with a Maxwellian distribution \( \exp(-\beta E) \) (where \( \beta = 1/(k_B T) \)) leads to thermal fluctuation with \( \text{Var}(T) = (k_B T^2)/C \).\(^3\)

Sampling for a time \( \Delta t \) produces a signal to noise for some input power \( P \) of (see Bernstein (2004))

\[
\frac{S}{N} \approx \frac{P \sqrt{\Delta t}}{\sqrt{k_B T^2 G}},
\]

(1.26)

where we have assumed that the incident power produces \( \Delta T = P/G \) in the bolometer. The premium here is to reduce both \( G \) and \( T \), yet, arbitrary reduction of \( G \) also increases the natural time constant \( \tau = C/G \), so to continue to minimize \( G \), one also has to minimize \( C \), which scales as \( T^3 \) at low temperatures due to phonons. In addition, one would like to maximize \( P \) by making the absorber size comparable to the wavelength of the light and be a strong absorber in-band. This implies that the optimal configuration is as cold as possible,\(^3\) has an absorber with dimensions comparable to the wavelength, and is as thin as possible.

In addition to the inherent noise of the detectors, there is also a fundamental limit to sensitivity from photon statistics. The probability \( P(n,T) \) that \( n \) photons occupy a mode in equilibrium at temperature \( T \) is proportional to \( \exp[-nh\nu/(k_B T)] \), and is normalized to unity for the sum over all \( n \). The mean occupation is then

\[
\langle n \rangle = \frac{1}{e^{h\nu/k_BT} - 1},
\]

(1.27)

and the variance is

\[
\sigma^2_N = \langle n^2 \rangle - \langle n \rangle^2 = N + N^2,
\]

(1.28)

---

\(^3\)For information specific to ACT and general reviews, see Lau (2007); Niemack (2008); Marriage (2006); Richards (1994); Hilton and Irwin (2005); McCammon (2005); Lindeman (2000).

\(^3\)The presentation here is only schematic. For more complete treatments of the noise in bolometers, see Mather (1982); Richards (1994) and Hilton and Irwin (2005).

\(^3\)Negative electrothermal feedback will, in general, produce a faster response time.

\(^3\)ACT’s bath temperature is 300 mK and the typical \( T_c \sim 510 \) mK.
1.6 Sensitivity limits for millimeter wave detection

which has a Poisson ($\mathcal{N}$) piece and a classical “wave noise” ($\mathcal{N}^2$) piece. The Poisson component of the variance corresponds to photon counting statistics, while the “wave noise” corresponds to the bandwidth-time uncertainty (see e.g., Rohlf and Wilson (1996)). For a single mode detector, the fractional uncertainty in the power received after some integration time $\tau$ is (Zmuidzinas (2003)),

$$\frac{\sigma_P}{P} = \frac{1}{\mathcal{N}} \sqrt{\frac{\mathcal{N}(1 + \eta \mathcal{N})}{\tau \eta \Delta \nu}},$$  \hspace{1cm} (1.29)

where $\eta$ is the efficiency. In the limit that wave noise (the $\mathcal{N}^2$ term relative to $\mathcal{N}$) dominates, this is just the Dicke equation (see e.g. Rohlf and Wilson (1996)), $\sigma_T/T = 1/\sqrt{\tau \Delta \nu}$. Pure CMB radiation is in the Poisson limit for the ACT bands, but including the atmosphere and loss in the optics gives total effective RJ temperatures\textsuperscript{40} seen by detectors in 145 GHz, 220 GHz, 280 GHz of 13 K, 22 K and 34 K (see Marriage (2006)). This gives $\mathcal{N}$ of 1.9, 2.1, 2.5, respectively for the three bands, which are neither in the Poisson or Dicke regimes. If photon statistics (wave vs. shot noise) are the distinguishing trait between radio and optical telescopes, then ACT is between those regimes.

There are two common units for noise that are used interchangeably. One is the noise per 0.5 second integration time (giving a postdetection bandwidth of 1 Hz), and has units of $\mu$K/$\sqrt{\text{Hz}}$. The other is the noise per 1 second integration time. It has units $\mu$K/$\sqrt{s}$. The former is more popular when describing noise spectra and the latter is more popular when describing sensitivity. We can then get a rough idea for the photon noise for 100% absorption efficiency and 20 GHz bandwidth. This gives $110 \mu$K/$\sqrt{s}$, $190 \mu$K/$\sqrt{s}$, $280 \mu$K/$\sqrt{s}$ for 145 GHz, 220 GHz, and 280 GHz, respectively.\textsuperscript{41}

Photon statistics set the ultimate sensitivity limit for a detector, and typical systems will be limited by the inherent detector noise. With both photon noise and detector noise, it is advantageous to instrument large arrays. For ACT, the detector noise is within 30% of $1000 \mu$K/$\sqrt{s}$. Assuming independence of detector noise across a $32 \times 32$ array gives $31 \mu$K/$\sqrt{s}$, which is at the level of the CMB anisotropy.

\textsuperscript{39}The wave noise ($\mathcal{N}^2$ in the variance) does not depend on efficiency because both the signal and noise scale in the same way with efficiency. In the Poisson limit, efficiency is important. The noise in multi-moded systems is more subtle, and is a consideration for ACT; see Zmuidzinas (2003).

\textsuperscript{40}

$$I_\nu = 2k_B T_{RJ} \nu^2 / c^2 = [(2h \nu^3)/c^2] N \Rightarrow N = k_B T_{RJ} / (h \nu).$$  \hspace{1cm} (1.30)

Note that at low frequencies, the CMB is in the Dicke limit.

\textsuperscript{41}Including an efficiency term, the ratio of the Poisson contribution to the Dicke contribution (for one mode) is $(\mu K^2)_{\text{Dicke}} / (\mu K^2)_{\text{Poisson}} = \mathcal{N} \eta$, so that for efficiencies lower than $\sim 50\%$, Poisson noise will start to become the dominant term, and will increase the noise figures here. For example, in 145 GHz, assuming an efficiency of 50%, the total noise is $\sim 130 \mu$K/$\sqrt{s}$ and the noise terms contribute comparably. The discussion here is only meant to sketch out why the instrument is not limited by photon noise. The full calculation of photon noise is left for future work.
Chapter 2
Recombination

2.1 Introduction

Cosmological recombination occurs when the photon gas in the early universe has cooled sufficiently for bound atoms to form. As the free electrons become locked in the ground states of these atoms, the opacity from Thomson scattering drops, and radiation carrying the signature of thermal inhomogeneities in the recombination plasma begins to stream freely across the universe. This radiation reaches us today as the cosmic microwave background (CMB) radiation. The history of the free electron population is not only essential to the liberation of the CMB – it also shapes the structure of perturbations that we see in the CMB. As described in the introduction, the free electron fraction determines the damping scale for perturbations (Silk damping) and is manifest today by a suppression of power of the temperature anisotropy on small scales.

The fundamental problem in cosmological recombination is to solve consistently for the evolution of the atomic level occupations and the radiation field (which has both a thermal piece, and a non-thermal piece from the radiation of the atoms themselves) in an expanding background. The primordial conditions are simple: the universe is homogeneous gas to an excellent approximation and helium and hydrogen are the only relevant atomic species. The recombination-era universe is dominated (by number) by photons. At $z = 1100$, there are $\sim 260$ protons per $\text{cm}^3$, and $\sim 5 \times 10^{11}$ photons per $\text{cm}^3$. Because of this, recombination to neutral hydrogen cannot occur until the temperature has dropped to the $\sim 1/4 \text{ eV}$ typical during hydrogen recombination, even though the binding energy is 13.6 eV. There are three distinct stages in recombination as different neutral species form: 1) helium nuclei capture one electron; here the binding energy is 54.4 eV and the capture begins $z \sim 7000$ (when the universe is 13.5 kyr old) and has completed by $z \sim 5000$ (25.5 kyr old), 2) helium nuclei capture a second electron; here the binding energy is 24.6 eV and the capture begins $z \sim 3000$ (65 kyr old) and finishes by $z \sim 1600$ (200 kyr old), 3) protons capture the remaining electrons; here the binding energy is 13.6 eV and capture begins at $z \sim 1600$ (200 kyr old) and continue until there is very little residual ionization left by $z \sim 400$ (1.95 million years old).

This chapter is adapted from three papers written with C. Hirata: Paper I Switzer and Hirata (2008a) describes the base recombination model, the influence of the feedback of spectral distortions between lines, and the effect of continuous opacity from H I photoionization in transport phenomena during He I recombination, Paper II Hirata and Switzer (2008) describes absorption of non-thermal radiation in two-photon processes from $n = 2$, non-resonant two-photon decay processes from $n > 2$, and the effect of finite resonance linewidth, Paper III Switzer and Hirata (2008b) describes several additional effects that are negligible to recombination, summarizes the overall magnitude and convergence of effects studied, and gives an error budget in free electron fraction and the anisotropy power in the CMB. Calculations or text by C. Hirata will be indicated with a footnote in the text.

See Sunyaev and Zeldovich (1970); Peebles and Yu (1970); Bond and Efstathiou (1987); Hu et al. (1997); Silk (1968); Hu and White (1997).
2.1 Introduction

The times here are quoted from an assumed initial singularity, and today we observe the recombination era from a universe \( \sim 13.6 \) Gyr old. This chapter addresses primarily effects in the second stage, where the helium nucleus captures a second electron. This process takes \( \sim 130 \) kyr. We will see that even though some of the fastest atomic rates are \( \sim 10^9 \) s\(^{-1}\), recombination as a whole is not actually fast enough (compared to the expansion of the universe) to attain a steady-state ionization fraction. The main new physics described here is that the small Saha\(^3\) abundance of neutral hydrogen during recombination to He I\(^4\) catalyzes a much more rapid recombination, which completes by \( z \sim 1800 \), around \( 40 \) kyr earlier than is assumed in recombination codes used for CMB parameter analysis. This sounds dramatic, but the corrections to the eventual power spectra are small, see Fig. 2.18.

The subtleties in the recombination history emerge from both the high accuracy (\( \sim 10^{-3} \)–\( 10^{-4} \)) required, and the large number of interaction processes. The highly-excited states are kept close to equilibrium\(^5\) by the high rates interconnecting them, and the rate of formation of the ground state in both helium and hydrogen is dominated by the occupation of the \( n = 2 \) states and the rates connecting the \( n = 2 \) states to the ground state. Here, rare processes are important because many of the fast allowed processes have reverse processes that quickly come to equilibrium, blocking overall progress toward recombination through the forward direction. For both helium and hydrogen, the decay channels to the ground state through the allowed transitions are dramatically suppressed relative to the vacuum rates by the optical depth in the gas. This mechanism is described as a “traffic jam” in Fig. 2.1. Indeed, in both systems, the two-photon decay rate from the \( n = 2 \) \( S \) state (singlet in the case of He I, denoted by \( S \)) to the ground state is comparable to the rate in allowed decay channels even though it is forbidden for one photon. This is the so-called “\( n = 2 \) bottleneck” (Seager et al. (2000)). Instead of bringing the ionization state of hydrogen or helium into Saha equilibrium, fast allowed decays simply boost the number of photons in certain regions of phase space to values far greater than the Planck distribution \( N = 1/(e^{h\nu/k_B T} - 1) \) would predict — until some process (such as Sobolev escape\(^6\) in the case of the 21.2 eV line) removes the photons. In general, processes that modify the overall recombination history either break the equilibrium in allowed lines by removing photons, or provide an independent path to the ground state. Examples are opacity from H I photoionization and two-photon decays, respectively. The primary goal of this chapter is to calculate the effect of H I photoionization opacity, and Sec. 2.3.1 describes the essential physics.

The primordial recombination was first investigated theoretically in the 1960s (Peebles (1968); Zeldovich et al. (1968)) using a simple “three-level atom” (TLA) approximation. This is summa-
Recombination

The Trafﬁc Jam

Figure 2.1: The optically thick Lyα “ trafﬁc jam.” On the left, de-excitations of one atom produce photons that excite the 1s state of a nearby atom in an optically thick gas, and the reverse. Because the universe is expanding, this cycle can be broken if the photons redshift off-resonance before they can excite nearby atoms. This is referred to as Sobolev or radiative escape. Two photon emission from n = 2 is optically thin to reabsorption, allowing atoms to relax to the ground state through this pathway, without the analogous trafﬁc jam. In helium 2p−1s (and other lines with energies > 13.6 eV), this trafﬁc jam can also be broken when the photon photoionizes a neutral hydrogen (H(1s) + γ → H+ + e−) rather than re-excite another helium atom. This is described in Sec. 2.3.1 and is one of the major goals of this chapter.

The TLA tracks the abundance of ground state hydrogen atoms (H I 1s), excited hydrogen atoms (assumed to be in Boltzmann equilibrium), and free electrons. The TLA accounts for recombination to and photoionization from excited H I levels, and allows excited atoms to decay to the ground state by Lyα (2p → 1s) or the “forbidden” two-photon (2s → 1s) emission. Subsequently, a substantial literature developed that tested some of the assumptions of the TLA and extended it to include helium recombination.7 Seager et al. (2000) provides the current benchmark precision recombination calculation by simulating 300 levels in H I, 200 levels in He I, 100 levels in He II, interactions with the radiation ﬁeld in the Sobolev approximation, basic hydrogen chemistry, and matter temperature evolution. They found that the three level model with a “fudge factor” inserted to accelerate H I recombination is an accurate approximation to their full multilevel atom solution. Their recombination code, RECFAST (Seager et al. (1999)), is packaged into most of the CMB anisotropy codes in common use, and underlies the cosmological constraints from the CMB, including those recently reported by WMAP.

There has been a recent resurgence in interest in cosmological recombination, both from the viewpoint of developing an accurate history of the era for upcoming precision small-scale CMB studies (such as ACT, Planck, and SPT) and from a renewed interest whether spectral distortion from recombination could be observable, and what could be learned from these distortions.8 9


8Here, atoms emit several photons as they cascade to the ground state. In the recombination era, the only impediment to forming a spectral distortion is the high entropy – many photons have to be added to have a signiﬁcant impact. Recombination originally attracted attention because 2p → 1s radiation would be visible today at ~ 2 THz and would greatly exceed the CMB in its Wein tail. Excitement about this excess has always been damped by the difﬁculty of experiments in this range and the fact that the estimated foreground is orders of magnitude brighter. One consequence of this radiation is that it keeps lithium ionized in the early universe, see Switzer and Hirata (2005), barring using resonant lithium scattering as a cosmological probe. Transitions between higher principal quantum numbers give distortions at much lower frequencies (~ GHz) that may be observable. These also require a physically complete description of cosmological recombination. The prospects for observing these lower frequency distortions are discussed in Appendix A.7.

9The works by Sunyaev and Chluba (2007); Wong et al. (2007); Chluba and Sunyaev (2008a); Lewis (2007); Hirata (2008) give an overview of several modiﬁcations and their relation to cosmological parameter estimates. Many new effects that augment the simple TLA approximation have been recently suggested: matter temperature evolution (Leung et al. (2004); Wong and Scott (2006)), two-photon transitions from high-lying states (Dubrovich and Grachev (2005); Dubrovich (2007); Chluba and Sunyaev (2007b)), the effect of He i intercombination lines (Dubrovich and Grachev (2005); Wong and Scott (2007)), departures of the l sublevels of hydrogen from their statistical population ratios (Chluba et al. (2007)), stimu-
2.1 Introduction

Figure 2.2: The life-cycle of hydrogen during recombination in the three-level approximation. The top oval represents a pool of free nuclei and electrons. Electrons are captured onto the nuclei through free-bound (recombination) processes, and liberated from the nuclei through bound-free (photoionization) processes. The excited states \( n > 2 \) (middle oval) are in equilibrium, so can be treated in a lumped category. This is the essential observation in the three-level approximation. The primary formation rate of atoms in the ground state (bottom oval) is through decays from the \( n = 2 \) level. In hydrogen this is just \( 2\gamma \) (from \( 2s \rightarrow 1s \)) and Ly\( \alpha \) (from \( 2p \rightarrow 1s \)), while in helium the primary paths are \( 2^1P^o - 1^1S \), \( n^3P^o - 1^1S \) and \( 2^1S - 1^1S \). (Fig. 2.3 gives the Grotrian diagram for the lowest excited states of helium.)

Studies in Seager et al. (2000) and Matsuda et al. (1969, 1971) found that there is a helium “\( n = 2 \) bottleneck” analogous to the one in H\( \text{I} \). This delays recombination to He\( \text{I} \). Fig. 2.3 gives an overview of the lowest-lying states of helium that are considered here. Here the \( 2^1P^o - 1^1S \) pathway is optically thick (and so blocked, except for Sobolev escape) and the two-photon \( 2^1S - 1^1S \) and intercombination \( 2^3P^o - 1^1S \) transitions are slow (taken here to be \( 50.94 \text{ s}^{-1} \) and \( 171 \text{ s}^{-1} \), respectively, see Drake (1986); Laughlin (1978)). Rather than relaxing through this bottleneck, atoms are more likely to be ionized again by the Planck radiation. A long-standing suggestion of J. Peebles is that the continuum opacity from hydrogen photoionization could destroy photons in the He\( \text{I} \) \( n^1P^o - 1^1S \) resonances and continuum transitions to the ground state, acting as a catalyst to accelerate He\( \text{I} \) recombination closer the Saha steady-state. Schematically for \( n^1P^o - 1^1S \) photons,

\[
\begin{align*}
\text{He}(n^1P^o) & \rightarrow \text{He}(1^1S) + \gamma, \\
H(1s) + \gamma & \rightarrow H^+ + e^-. \tag{2.2}
\end{align*}
\]

This chain destroys the photons and prevents them from re-exciting other helium atoms and is equally applicable to the \( n^3P^o - 1^1S \) and \( n^1D - 1^1S \) series.

The first description of the interaction between helium recombination and the neutral hydrogen population in literature was Hu et al. (1995), who proposed that because the H\( \text{I} \) photoionization rate exceeded the Hubble rate, He\( \text{I} \) recombination would be driven to the Saha steady-state. This is refuted in Seager et al. (2000), who claim that the photoexcitation rate of \( 2^1P^o - 1^1S \) exceeds the H\( \text{I} \) photoionization rate; this was taken up again in more detail by Kholupenko et al. (2007), who found that the H\( \text{I} \) photoionization rate was significant compared to the total rate through \( 2^1P^o - 1^1S \). Here, we describe the influence of neutral hydrogen in terms of transport scales. Neutral hydrogen presents an optical depth to photoionization from \( 2^1P^o - 1^1S \) (and other photons). This optical depth can make no difference unless it is sufficiently high across frequency scales that appear in

\[\text{lated two-photon transitions (Chluba and Sunyaev (2006); Kholupenko and Ivanchik (2006)), the feedback from higher-lying states (Chluba and Sunyaev (2007a)), radiative transport concerns (Grachev and Dubrovich (2008)) and H\text{I} continuum opacity in helium (Kholupenko et al. (2007)).}\]
Figure 2.3: Formation of neutral helium: a Grotrian diagram (up to $n = 3$) for He$^+$. The notation used throughout is the standard term symbol $n^2 S + 1 L J$ where $n$ is the principal quantum number of the excited electron, $S$ is the total spin, $L$ is the total orbital angular momentum and $J$ is the total angular momentum. Singlet ($S = 0$, parahelium) and triplet ($S = 1$, ortho-helium) levels and higher-order transitions give a rich system of low-lying transitions. Marked are two-photon transitions (short dashed lines from $2^1 S_0$ and $3^1 S_0$), allowed electric dipole transitions (solid lines from $2^1 P_1^0$ and $3^1 P_1^1$, like the Lyman series in H$^+$), the intercombination lines (long dashed lines from $2^3 P_1^1$ and $3^3 P_1^0$), and quadrupole transitions (from $3^1 D_2$). The dipole transitions are treated in Sec. 2.3.7 using a Monte Carlo method with partial redistribution, forbidden one-photon lines are treated in Sec. 2.3.4 using an analytic method for complete redistribution. (The energy levels are not drawn to scale.)
the transport problem. Then, for example, neutral hydrogen opacity can remove photons trapped in the optically thick part of the $2^1P^o - 1^1S$ resonance, allowing it to relax. (The microscopic picture of this is that hydrogen becomes important when it can destroy a photon between the time it is emitted by a helium atom and absorbed by another helium atom, and that this can happen faster than the photon would have otherwise escaped through redshifting.) We find that this transport crossover is at $z \sim 2200$ and that by $z \sim 1800$, He I recombination is driven to the Saha steady-state.

Another effect that we find to be important is the feedback of spectral distortions on lower-lying resonances (see Seager et al. (2000); Chluba and Sunyaev (2007a)). For example, escaped radiation from Ly$\beta$ in H I will redshift onto the Ly$\alpha$ transition and excite atoms; analogous effects occur in helium. We develop an iterative method to study the effect of this feedback that solves for it self-consistently with the level populations in Sec. 2.2.2.

In the ordinary Sobolev approximation, the $2^1P^o - 1^1S$ line shape is not important because the radiation is in equilibrium over the extent of the $2^1P^o - 1^1S$ profile. If continuum opacity provides sufficient optical depth, the it disturbs this equilibrium by destroying photons in the optically thick part of the line. The net effect is that what used to be a flat phase space distribution of photons around the line center is now a population that decays redward of the line as progressively more photons are removed from the field. Then, the integral of the phase space density over the line profile depends on both the frequency dependence of the radiation and on the line profile. This opens many additional possibilities for additional effects that were irrelevant in previous literature that considered a pure Sobolev approximation.

So far, we have described the “core” effects that will be the emphasis of this chapter, but there are several additional effects to consider. These are described in detail in Switzer and Hirata (2008b), but here we restrict to two that, while negligible for the history, are still interesting because they provide more intuition for the physics of the era. A familiar phenomenon from stellar systems is microturbulence, where peculiar motion in a gas that is incoherent over the transport length will give additional line broadening on top of the thermal (Doppler) broadening. In Sec. 2.4.3 we show that the analogy in cosmology is macroturbulence – where there is little peculiar velocity structure (less than thermal velocities) on scales relevant for transport. We also investigate the effect of Thomson scattering on radiative transport within the line in Sec. 2.4.1 and find that it is negligible.

The remainder of this chapter describes a multi-level atom code for helium recombination and several new physical processes during helium recombination. We propose three main corrections to the standard He I recombination model: (1) feedback between allowed lines where a spectral distortion produced by $n + 1$ redshifts from $n$, described in Sec. 2.2.2 and Sec. 2.3.2, (2) additional rates from intercombination processes ($n^3P^o - 1^1S$, see Dubrovich and Grachev (2005)), and (3) treatment of H I continuous opacity in transport within $2^1P^o - 1^1S$ and $2^3P^o - 1^1S$, see Sec. 2.3 and Sec. 2.3.4. We begin with a description of the code in Sec. 2.2, and leave details for Appendix A.

2.2 A new level code with radiative feedback

The standard recombination scenario describes a homogeneous, interacting gas of protons, helium nuclei, electrons, and photons in an expanding background. The underlying physics in the standard problem is known, but the solution is complicated by the interactions between radiation and atomic level occupations, and the variety of atomic rates. Many atomic rates involved

---

10The background dynamics are allowed to include other homogeneously distributed components such as neutrinos, dark matter, and dark energy, but in our treatment we only consider their effect on $H(z)$, neglecting any direct interaction with the baryons and photons. Several extensions to the standard scenario have been proposed but will not be considered here: energy injection from self-annihilating or decaying dark matter, primordial magnetic fields, small-scale inhomogeneities or peculiar velocities, and others (Gopal and Sethi (2005); Battye et al. (2001); Mapelli and Ferrara (2005); Padmanabhan and Finkbeiner (2005); Peebles et al. (2000); Mapelli et al. (2006); Hannestad (2001); Pierpaoli (2004); Avelino et al. (2000); Bonometto and Shouping (1986)).
are time-consuming to calculate, or not well known (for example, collisional and intercombination rates in He I). \footnote{With the development of fast interpolation methods for the recombination history (see Fendt et al. (2008)), there is much less of an onus than there was historically on developing a fast recombination code. Therefore, the main incentive to simplify the description of a process is to gain intuition, and computational speed is just a byproduct.} Many of the important details in the calculations are wrapped up in the notation, so we present the major symbols used here in Table 2.1.

### 2.2.1 Summary of the method

For brevity, we only highlight the physical arguments of the multilevel code, and differences from Seager et al. (2000).\footnote{The base recombination level code with feedback was developed by C. Hirata, and E. S. implemented the transport effects described in this chapter.} The multilevel atom code tracks levels up to a maximum principal quantum number $n_{\text{max}}$, which we take to be 200 for H I (giving 245 levels), 100 for He I (giving 289 levels), and 100 for He II (giving 145 levels).\footnote{Sublevels are resolved up to $n = 10$ and we include quadrupole and intercombination transitions in the He I rates. A detailed discussion of atomic data can be found in Switzer and Hirata (2008a) (Paper I). In Switzer and Hirata (2008b) (Paper III), we investigated the effect of increasing $n_{\text{max}}$, but find that it is negligible.} Unless stated otherwise, we assume a $\Lambda$CDM cosmology with $\Omega_b = 0.04592$, $\Omega_m = 0.27$, $\Omega_r = 8.23 \times 10^{-5}$, zero spatial curvature, massless neutrinos, a Hubble parameter of $h = 0.71$, and present-day radiation temperature $T_r(z = 0) = 2.728$ K.\footnote{Here we have used the older value from Fixsen et al. (1996) instead of Mather et al. (1999) (based on further calibration studies) to be consistent with Seager et al. (2000). Using $T_r = 2.725$ K will change some numerical values slightly, but the physical conclusions hold.} The fiducial fractional helium abundance (by number) is $Y_H = 0.079$.\footnote{The helium abundance is often cited by mass as $Y_p \approx 0.25$. See Burles et al. (2001); Olive et al. (2000).} The Hubble rate in such a cosmology is

\begin{equation}
H(z) = H_0 \sqrt{\Omega_\Lambda + \Omega_m(1 + z)^3 + \Omega_r(1 + z)^4},
\end{equation}

The number density of hydrogen nuclei is given by (Seager et al. (2000, 1999))

\begin{equation}
n_H(z) = 1.123 \times 10^{-5} \frac{\Omega_b h^2}{1 + 3.9715 f_{\text{He}}} (1 + z)^3 \text{cm}^{-3},
\end{equation}

where 3.9715 is the ratio of atomic masses of $^4\text{He}$ and $^1\text{H}$. We will use the photon phase space density $N(\nu)$ to track the radiation spectrum instead of the specific intensity $J_\nu$, because the former is conserved along a trajectory in free space whereas $J_\nu$ decreases as the universe expands. Also, let the fractional abundance of an atomic state $i$ be $x_i = n_i/n_H$.\footnote{Normalizing the states with respect to $n_H$ is convenient but arbitrary. One product of this choice is that the free electron fraction exceeds 1 before hydrogen recombination because of the additional free electrons that will be bound to helium.} The relation between the phase space and intensity is

\begin{equation}
N(\nu) = \frac{c^2}{2h\nu^3} J_\nu.
\end{equation}

Photoionization and recombination contributions to atomic level population dynamics are discussed in Seager et al. (Seager et al. (2000)), and we follow their treatment here. The rate of change of the average occupation of an atomic level is given by a series of bound-bound and bound-free rate equations. The photoionization rate out of some level $i$ has units of $s^{-1}$ and is given by

\begin{equation}
\beta_i = \frac{8\pi}{c^2} \int_{\nu_{\text{th}}}^{\infty} \sigma_{ic}(\nu) \nu^2 N(\nu) d\nu,
\end{equation}

where $N(\nu)$ is the photon phase space density, $\nu_{\text{th}}$ is the photoionization threshold frequency from level $i$, and $\sigma_{ic}$ is the photoionization cross-section from that level, as a function of frequency

$\text{CDM cosmology}$
2.2 A new level code with radiative feedback

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Units</th>
<th>Description</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>1</td>
<td>Voigt unitless width parameter</td>
<td>Eq. (A.54)</td>
</tr>
<tr>
<td>$A_{i \rightarrow j}$</td>
<td>$s^{-1}$</td>
<td>Einstein spontaneous one-photon decay rate for $i$ to $j$</td>
<td></td>
</tr>
<tr>
<td>$\eta_{\text{coh}}$</td>
<td>1</td>
<td>Fraction of line photon absorptions resulting in coherent scattering</td>
<td></td>
</tr>
<tr>
<td>$f_{i}$</td>
<td>1</td>
<td>Fraction of line photon absorptions followed by transition to level $i$</td>
<td></td>
</tr>
<tr>
<td>$\tau_{\text{inc}}$</td>
<td>1</td>
<td>Fraction of line photon absorptions resulting in incoherent processes</td>
<td>$f_{\text{inc}} = 1 - \eta_{\text{coh}}$</td>
</tr>
<tr>
<td>$\tau_{\text{coh}}$</td>
<td>1</td>
<td>Degeneracy of level $i$, not including nuclear spin</td>
<td>$K = \lambda_{\text{line}}^3/(8\pi H(z))$</td>
</tr>
<tr>
<td>$H$</td>
<td>$s^{-1}$</td>
<td>Hubble rate</td>
<td></td>
</tr>
<tr>
<td>$K$</td>
<td>cm$^{-3}$</td>
<td>Peebles $K$-factor (Peebles (1968))</td>
<td></td>
</tr>
<tr>
<td>$n_e$</td>
<td>cm$^{-3}$</td>
<td>Electron density</td>
<td></td>
</tr>
<tr>
<td>$n_{i}$</td>
<td>cm$^{-3}$</td>
<td>Density of atoms in level $i$</td>
<td></td>
</tr>
<tr>
<td>$n_{i}^{\text{HH}}$</td>
<td>cm$^{-3}$</td>
<td>Total density of all hydrogen nuclei</td>
<td></td>
</tr>
<tr>
<td>$N^*$</td>
<td>1</td>
<td>Photon phase space density</td>
<td>Eq. (2.5)</td>
</tr>
<tr>
<td>$N_{\text{C}}$</td>
<td>1</td>
<td>$N^*$ in equilibrium with the continuum opacity</td>
<td>Eq. (2.33)</td>
</tr>
<tr>
<td>$N_{\text{L}}(0)$</td>
<td>1</td>
<td>$N^*$ in equilibrium with the line opacity</td>
<td>Eq. (2.15)</td>
</tr>
<tr>
<td>$N_{\text{L}}(i)$</td>
<td>1</td>
<td>Modification of $N_{\text{L}}$ used with coherent scattering</td>
<td>Eq. (2.50)</td>
</tr>
<tr>
<td>$N_{\text{P}1}$</td>
<td>1</td>
<td>Photon phase space density for blackbody distribution</td>
<td>$N_{\text{P}1} = 1/(e^{\nu_{\text{P}1}/k_B T_R} - 1)$</td>
</tr>
<tr>
<td>$N_{c}$</td>
<td>1</td>
<td>Photon phase space density blue ($+$)/red ($-$)-ward of a line</td>
<td>Eq. (2.18)</td>
</tr>
<tr>
<td>$x$</td>
<td>1</td>
<td>Photoionization rate from level</td>
<td>Eq. (2.37)</td>
</tr>
<tr>
<td>$n_{\text{inc}}$</td>
<td>1</td>
<td>Total number of free particles per hydrogen nucleus</td>
<td>Eq. (2.40)</td>
</tr>
<tr>
<td>$\tau_{\text{inc}}$</td>
<td>1</td>
<td>Escape probability from the line</td>
<td></td>
</tr>
<tr>
<td>$\tau_{\text{MC}}$</td>
<td>1</td>
<td>Prob. of photon in MC being lost by H i absorption or redshifting</td>
<td></td>
</tr>
<tr>
<td>$\xi$</td>
<td>1</td>
<td>Soboilev escape probability</td>
<td>Eq. (2.12)</td>
</tr>
<tr>
<td>$Q$</td>
<td>1</td>
<td>Transition rate from level $i$ to level $j$</td>
<td></td>
</tr>
<tr>
<td>$R_{i,j}$</td>
<td>s$^{-1}$</td>
<td>Heating per hydrogen nucleus per unit time</td>
<td></td>
</tr>
<tr>
<td>$T_{\text{mc}}$</td>
<td>1</td>
<td>Temperature (in K) or radiation (in T$_s$)</td>
<td></td>
</tr>
<tr>
<td>$T_R$</td>
<td>1</td>
<td>Frequency relative to line center in Doppler units</td>
<td>$x = (\nu - \nu_{\text{line}})/\Delta \nu_{\text{D}}$</td>
</tr>
<tr>
<td>$x_e$</td>
<td>1</td>
<td>Abundance of electrons relative to total hydrogen nuclei</td>
<td>$x_e = n_e/n_{i}^{\text{HH}}$</td>
</tr>
<tr>
<td>$x_i$</td>
<td>1</td>
<td>Abundance of the state $i$ relative to total hydrogen nuclei</td>
<td>$x_i = n_i/n_{i}^{\text{HH}}$</td>
</tr>
<tr>
<td>$\alpha_{i}$</td>
<td>cm$^{3}$ s$^{-1}$</td>
<td>Recombination coefficient to level $i$</td>
<td></td>
</tr>
<tr>
<td>$\beta_{i}$</td>
<td>s$^{-1}$</td>
<td>Photoionization rate from level $i$</td>
<td></td>
</tr>
<tr>
<td>$\Gamma_{i}$</td>
<td>s$^{-1}$</td>
<td>Width of level $i$</td>
<td></td>
</tr>
<tr>
<td>$\Gamma_{\text{line}}$</td>
<td>cm$^{3}$ s$^{-1}$</td>
<td>Lorentz heating width of the line</td>
<td></td>
</tr>
<tr>
<td>$\Delta \nu_{\text{D}}$</td>
<td>Hz</td>
<td>Doppler width of line</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{i}$</td>
<td>cm$^{3}$ s$^{-1}$</td>
<td>Differential optical depth from continuous opacity</td>
<td>Eq. (2.20)</td>
</tr>
<tr>
<td>$\nu_{\text{c}}$</td>
<td>Hz$^{-1}$</td>
<td>Spontaneous two-photon decay rate from $i$ to $j$</td>
<td></td>
</tr>
<tr>
<td>$\nu_{\text{line}}$</td>
<td>Hz</td>
<td>Frequency of line center; $\nu_{\text{ul}}$ for specific upper and lower levels</td>
<td></td>
</tr>
<tr>
<td>$\nu_{\text{th},i}$</td>
<td>Hz</td>
<td>Rescaled photon phase space density</td>
<td>Eq. (2.51)</td>
</tr>
<tr>
<td>$\nu_{s}$</td>
<td>Hz</td>
<td>Probability distribution of photon frequency in Monte Carlo</td>
<td></td>
</tr>
<tr>
<td>$\xi_{\text{MC}}$</td>
<td>cm$^{2}$</td>
<td>Photoionization cross section from level $i$</td>
<td></td>
</tr>
<tr>
<td>$\eta_{\text{inc}}$</td>
<td>cm$^{2}$</td>
<td>Thomson cross section</td>
<td></td>
</tr>
<tr>
<td>$\tau_{\text{coh}}$</td>
<td>1</td>
<td>Soboilev optical depth from coherent scattering</td>
<td>$\tau_{\text{coh}} = \eta_{\text{coh}} \tau_{\text{inc}}$</td>
</tr>
<tr>
<td>$\tau_{\text{inc}}$</td>
<td>1</td>
<td>Soboilev optical depth from incoherent processes</td>
<td>$\tau_{\text{inc}} = \eta_{\text{inc}} \tau_{\text{inc}}$</td>
</tr>
<tr>
<td>$\tau_{\text{LL}}$</td>
<td>1</td>
<td>Optical depth from continuous opacity between lines</td>
<td>Eq. (2.23)</td>
</tr>
<tr>
<td>$\Delta i$</td>
<td>1</td>
<td>Total Soboilev optical depth</td>
<td>Eq. (2.11)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>Hz$^{-1}$</td>
<td>Atomic line profile</td>
<td></td>
</tr>
<tr>
<td>$\chi$</td>
<td>1</td>
<td>Photo-atom scattering angle</td>
<td>Eq. (A.56)</td>
</tr>
</tbody>
</table>
Recombination (implicit in subsequent equations). The photorecombination rate coefficient into level $i$ for spontaneous and stimulated processes has units $\text{cm}^3 \cdot \text{s}^{-1}$ and is

$$\alpha_i = \frac{8\pi}{c^2} \left( \frac{n_i}{n_c n_e} \right)_{\text{LTE}} \int_{\nu_{th,i}}^{\infty} \sigma_{i\nu} \nu^2 [1 + N(\nu)] e^{-h\nu/k_BT_m} d\nu,$$  \hspace{1cm} (2.7)

where $T_m$ is the matter temperature and $n_c$ is the number density of the continuum species. The prefactor is the Saha ratio of the occupation of the level $i$ (in local thermal equilibrium, LTE) to the free electron density times the continuum state density,

$$\left( \frac{n_i}{n_c n_e} \right)_{\text{LTE}} = \left( \frac{\hbar^2}{2\pi m_e k_B T_m} \right)^{3/2} \frac{g_i}{2g_c} e^{h\nu_{th,i}/k_BT_m},$$  \hspace{1cm} (2.8)

where $i$ labels a bound state of a species, and $c$ labels the continuum state of the species; $g_i$ and $g_c$ are the state degeneracies. The bound-free rate for a level $i$ is then

$$\frac{dx_i}{dt} = \alpha_i n_e x_c - \beta_i x_i.$$  \hspace{1cm} (2.9)

Except for several transitions in He $I$ where transport is calculated separately to include new effects, single photon bound-bound rates are calculated in the standard Sobolev approximation with complete redistribution,

$$\frac{dx}{dt} = A_{u \rightarrow l} P_S [x_u (1 + N_\nu) - \frac{g_u}{g_l} x_l N_\nu],$$  \hspace{1cm} (2.10)

where $A_{u \rightarrow l}$ is the Einstein rate coefficient connecting an upper bound state $u$ to a lower bound state $l$, and $N_\nu$ is the phase space density of radiation on the blue side\(^{17}\) of the line. (Here $N_\nu$ is the Planck density $N_{Pl}$ plus any spectral distortions from more energetic transitions.) In the Sobolev approximation, the rates are modulated by the probability that a photon will escape from the resonance,\(^{18}\) allowing the average occupation state of the gas to change. The probability is associated with the Sobolev effective optical depth,

$$\tau_S = \frac{A_{u \rightarrow l} c^3}{8\pi H(z) \nu^3_{ul}} n_H \left( x_l g_u / g_l - x_u \right),$$  \hspace{1cm} (2.11)

so that the escape probability is

$$P_S = \frac{1 - e^{-\tau_S}}{\tau_S}.$$  \hspace{1cm} (2.12)

Time derivatives for the occupations and temperatures are converted to redshift derivatives according to

$$\frac{dz}{d\tau} = -(1 + z) H(z).$$  \hspace{1cm} (2.13)

Equations (2.9), (2.10), and (A.3) give a set of stiff equations for the level occupations, as in Seager et al. (Seager et al. (2000)) that can be solved using the semi-implicit Bader-Deuflhard method (Press et al. (1992)). This code is the base upon which we will consider all additional effects.

\(^{17}\)Because the universe is expanding, radiation will redshift onto the line from the blue side, so represents a boundary condition for interaction that occurs subsequently with the line. In a collapsing universe, the red side of the line would define the “input” of the radiative transport through the line.

\(^{18}\)We use “resonance” loosely to refer to the collective property of a gas of atoms, and “transition” to refer to quantum-mechanical transitions of one atom in that gas, depending on whether we want to emphasize the transport or atomic physics.
2.2 A new level code with radiative feedback

2.2.2 Radiative feedback in hydrogen

Photons that escape a resonance constitute a spectral distortion to the Planck spectrum. These distortions, in turn, feedback on the dynamics of recombination. During hydrogen and He II recombination, the optical depth between lines\(^{19}\) is negligible and this distortion can redshift down to excite lower transitions, in addition to the thermal spectrum incident on the blue side of the lines. This process impedes overall recombination to the ground state.\(^{20}\) In this section, we describe our treatment of the effect for hydrogen.

For hydrogen, the distortion is determined by a simple vacuum transport equation for one resonance. In helium, the problem is complicated by continuous opacity from H I photoionization. Here, distortion photons are absorbed by H I in flight, and may not redshift down to the lower level. This is treated in Sec. 2.3.2.

The bound-bound radiative transport equation under complete redistribution of the scattered photons is\(^{21}\)

\[
\frac{\partial N}{\partial \nu} = \tau_S \phi(\nu) [N(\nu) - N_L] \\
\approx \tau_S \phi(\nu) \left[ N(\nu) - \frac{x_u g_l}{x_l g_u} \right],
\]

where \(N_L\) is the radiation phase space density that would be in equilibrium with the line,

\[
N_L = \frac{x_u}{x_l (g_u/g_l)} - x_u \approx \frac{x_u g_l}{x_l g_u}.
\]

This is obtained by setting the de-excitation rate \(A_{u \rightarrow l}(1 + N)\) equal to the excitation rate \(g_u A_{u \rightarrow l} N / g_l\).

Here we have used the approximation that the upper level is significantly less occupied than the lower level, which is valid when the lower level is the ground state since the energy of the first excitation in H I, He I, or He II is many times \(k_B T_r\) (which is a few tenths of an eV to 1 eV during H I and He I recombination, respectively). The transport equation (Eq.2.14) has the solution

\[
N(\nu) = \frac{x_u g_l}{x_l g_u} - C \exp \left[ -\tau_S \int_{\nu}^{\infty} \phi(\nu) d\nu \right],
\]

where \(C\) is the constant of integration. This constant is obtained from the initial condition that the far blue side of the line has some phase space density \(N_+\), which implies

\[
C = \frac{x_u g_l}{x_l g_u} - N_+.
\]

\(^{19}\)This is later defined as \(\tau_{LL}\) and is the optical depth a photon would traverse as it redshifts between the upper and lower line. If this traversal is optically thick, then the feedback of the spectral distortion is suppressed.

\(^{20}\)In the ordinary Sobolev approximation, the redshifting photon escapes to infinity and does not excite any further atoms. Thus we can understand feedback as the statement that while the photon may have escaped a resonance it has not truly escaped the gas. The calculation here shares some similarity to Switzer and Hirata (2005), where we showed that the relic recombination radiation is enough to ionize most of the neutral lithium population in the era following recombination, except here the radiation is fed back onto the recombination dynamics rather than the evolution of a trace species at a later time.

Feedback between levels can be calculated in a number of ways. The most direct way is to simulate a frequency-discretized radiation field along with the atomic levels. In this method, the evolution of the radiation field bins and the evolution of the occupation states of the atoms are solved for simultaneously. Because of the huge range of rate scales in the system and the large number of radiation bins that it is necessary to track, this method quickly becomes difficult to manage and ill-conditioned. For this reason, we use an iterative method to include feedback in the level code. This is both practical to implement, and accurate.

\(^{21}\)In complete redistribution, the photon is re-emitted across the Voigt profile. We will discuss the validity of this assumption in the context of energy conservation in Sec. 2.3.2.
Thus the phase space density \( N_- \) on the red side of the line is

\[
N_- = N_+ + \left( \frac{x_u g_l}{x_l g_u} - N_+ \right) \left( 1 - e^{-\tau_S} \right).
\]  

(2.18)

Notice that as \( \tau_S \) becomes large, then the phase space density on the red side of the line approaches \( N_L \). Because \( \mathcal{N} \) is conserved during the transport between \( \text{H} \, \text{I} \) and \( \text{He} \, \text{II} \) resonances (i.e. there is negligible continuum absorption or emission), the phase space density on the blue side of the \( \text{H} \, \text{I} \) Ly\( n \) resonance (i.e. \( 1s - np \)) is simply the phase space density on the red side of the Ly\( (n + 1) \) resonance at an earlier time:

\[
N_+ (\text{Ly}\, n, z) = N_- [\text{Ly}(n + 1), z'],
\]

\[
z' = \frac{1 - (n + 1)^{-2}}{1 - n^{-2}} (1 + z) - 1,
\]  

(2.19)

where \([1 - (n + 1)^{-2}]/[1 - n^{-2}]\) is the ratio of line frequencies. A similar result holds for \( \text{He} \, \text{II} \). Because of the existence of \( \text{H} \, \text{I} \) continuum opacity during \( \text{He} \, \text{I} \) recombination, Eq.2.18 does not apply to \( \text{He} \, \text{I} \). Sec. 2.3 treats this problem.

The level code calculates the radiation distortions generated by transitions from excited states to the ground state in the first iteration using Eq.2.18, for each redshift step, for each species, assuming thermal radiation is incident on the blue side of the line. In a second iteration, we transport the distortion generated by the \((i + 1)\)th transition to the \(i\)th transition to the ground state of the same species. That is, we only transport the radiation from the next higher-lying ground-excited transition, in the same species. (Inter-species feedback between \( \text{He} \, \text{I} \) and \( \text{H} \, \text{I} \) is discussed in subsequent sections.) The distortion is recalculated for each level in the second pass, and these are transported to the lower levels and applied to a third pass of the level code. The iteration continues in this way. Because the iteration step accounts for much of the total feedback effect, subsequent iterations give progressively smaller corrections, and the procedure converges rapidly. The typical fractional contribution of the fifth iteration is \(|\Delta x_e| \approx 2 \times 10^{-4}\) we stop there. This converges and is shown in Fig. 2.4.

### 2.3 Hydrogen continuum opacity and helium recombination

There are three ways that \( \text{H} \, \text{I} \) continuous opacity could accelerate \( \text{He} \, \text{I} \) recombination: (Seager et al. (2000); Hu et al. (1995))

1. Hydrogen can suppress feedback in the \( \text{He} \, \text{I} \) lines by absorbing spectral distortions produced by more energetic transitions before they redshift down to the next line and excite a helium atom.

2. If the \( \text{H} \, \text{I} \) opacity is large, it could directly absorb \( \text{He} \, \text{I} \) resonance line photons, thus increasing the effective escape probability above the Sobolev value from redshifting alone.

3. \( \text{H} \, \text{I} \) opacity can permit a direct recombination to the ground state by absorbing the photon which would otherwise be nearly guaranteed to ionize another \( \text{He} \, \text{I} \) in its ground state.

We treat mechanism #1 in Secs. 2.3.1 and 2.3.2. The problem of absorption of \( \text{He} \, \text{I} \) line photons by \( \text{H} \, \text{I} \) (mechanism #2) is more complicated. The physical picture is outlined in Sec. 2.3.3, and it is split into two cases. For the helium intercombination and quadrupole lines, there is negligible coherent\(^{22}\) scattering within the line since the upper level has allowed decays (and allowed pathways to other

\(^{22}\)We will develop the distinction between coherent scattering and incoherent processes in Sec. 2.3.2.
2.3 Hydrogen continuum opacity and helium recombination

Figure 2.4: Convergence of the iterations to include feedback of non-thermal distortions between lines. These are descending from the difference between no feedback and one iteration (top solid line), between the first iteration and the second, and so on. Note that by the 4th iteration, the effect is roughly $\Delta x_e < 10^{-4}$, so by going a fifth iteration, any systematic effect is negligible. Note that the integration tolerance taken in the level code is $1 \times 10^{-9}$.

states) whereas re-emission of the photon to the ground state of He i is semiforbidden or forbidden. This case is the simplest to consider and it is treated analytically in Sec. 2.3.4. The other case is that of the allowed He i $n^1P^o - 1^1S$ lines, in which coherent scattering plays a key role alongside incoherent absorption/emission processes and H i opacity in determining the line profile and net decay rate. This situation is treated via Monte Carlo simulation in Sec. 2.3.7. Mechanism #3 produces a negligible effect and is described Switzer and Hirata (2008a). Ultimately, the main effect will be of H i in $2^1P^o - 1^1S$, and while this should formally be treated with partial redistribution, we will see that the case of no coherent redistribution is a reasonable approximation for which the semi-analytic methods developed in Sec. 2.3.4 can be used. The effect of H i opacity is also important within feedback during He i recombination and should be considered an essential component of that analysis.

2.3.1 The continuous opacity of neutral hydrogen

The photons emitted in resonant transitions in He i from excited states to the ground state have energies above the H i photoionization threshold. The opacity from photoionization of H i influences transport both within and between He i lines. In this section, we describe how the continuum opacity is calculated, and Sec. 2.3.4 describes details of transport subject to continuous opacity. Throughout, we use $\eta_c$ to represent the continuum optical depth per unit frequency,

$$\eta_c = \frac{d\tau}{d\nu} = \frac{n_{\text{H}}(z)x_{1s}(z)\sigma_{ci} e}{H(z)\nu},$$

where $\sigma_{ci}$ is the photoionization cross-section of neutral hydrogen, and $x_{1s}$ is the ground state occupation fraction (the excited states are more sparsely occupied and have lower photoionization

\[\text{Here the units are inverse frequency and have the interpretation that } r_c = \Delta \nu \eta_c \text{ is the optical depth to photoionization after redshifting through some } \Delta \nu.\]
cross sections, so we neglect them). Stimulated recombination to H(1s) can be neglected. The continuum optical depth above threshold varies slowly with frequency because hydrogen has a simple (1-electron) photoionization response.

The neutral hydrogen population is well-described by the Saha distribution at early times ($z > 1700$):

$$x_{\text{HI}} \approx x_c x_{\text{HI}0} n_{\text{H}} \left( \frac{\hbar^2}{2 \pi m_e k_B T_m} \right)^{3/2} e^{\chi_H / k_BT_m},$$  \hspace{1cm} (2.21)

where $x_{\text{HI}} = 1 - x_{\text{H}} \approx 1$. (Fig. 2.5 shows the ionization rate and relaxation rate to the Saha steady-state.)

The optical depth to photoionization after redshifting one Doppler width ($\nu_D$, for $^4$He) during the He I recombination era is shown in Fig. 2.6. It is optically thick through a Doppler width after $z \sim 1800$. If the Doppler width were a reasonable transport scale, this would suggest that He I will not impact the recombination history until after $z \sim 1800$. In reality, lines like $2^1P^o - 1^1S$ are optically thick into their wings, so the transport scale is across significantly larger frequencies. This is described in Sec. 2.3.3, where we show that the transport scale cross over is $z \sim 2200$. The Doppler scale is simply an illustrative scale that applies to all $^4$He transitions.

The energy separations between transitions from excited states $i+1$ and $i$ to the ground state in He I are typically much greater than the optically thick line widths of even the broadest $n^1P^o - 1^1S$ series lines. Thus, radiative transport in He I can be thought of as taking place through two phases. In the first, continuum processes influence transport within a line. This modifies the transition rates, and sets the escape probability from a given resonance, which may exceed the Sobolev value. This is described in Sec. 2.3.3. In a second phase, continuum processes influence the transport of radiation between resonances, as described in the next section.

### 2.3.2 Feedback from transport between He I lines and continuous opacity

Sec. 2.2.2 addressed the feedback of a radiation distortion produced by a higher resonance on lower-lying resonances in H I and He II. For H I and He II, this transport is in free space in the approximation that the resonances are spaced more widely than their widths. In He I, the picture is more complicated because the radiative transport is subject to the opacity from the photoionization of neutral hydrogen. There is a sufficient neutral population that, when integrated over the photon’s trajectory, the feedback between levels can be significantly suppressed. This is true especially near the end of the He I recombination and beginning of H I recombination.

The algorithm presented earlier to include feedback iteratively can be easily modified to include feedback suppression between lines: calculate the distortions for all resonances, and in the next step calculate the distortions for all resonances, and in the next step again. This approximation is not entirely valid for H I at late times, however this is not relevant to helium recombination and are deferred to a future paper.
2.3 Hydrogen continuum opacity and helium recombination

Figure 2.5: The ionization and relaxation rates (in $s^{-1}$) for H I during the period of He I recombination, assuming each He I recombination generates a photon that photoionizes a hydrogen atom. Here we consider two He I recombination histories: one in equilibrium and the history derived here (Fig. 2.14). The “H I relaxation rate” is $t_{\text{Saha}}^{-1}$ (Eq. 2.22) for H I to return to Saha equilibrium if its abundance is perturbed. In either He I history, the ionizing radiation from He I is not sufficient to push H I evolution out of the Saha steady-state because of the large disparity in their rates. The Hubble rate at $z = 2400$ was $\sim 2 \times 10^{-13} s^{-1}$.

Figure 2.6: The continuum optical depth $d\tau_c/d\nu = \eta_c$ times the Doppler width of He I $2^1P^0 - 1^1S$ as a function of redshift.
Figure 2.7: A comparison of the effect of feedback of a spectral distortion in helium and hydrogen recombination produced by higher-lying states on lower-lying states in the same species after several iterations. $\Delta x$ is the change in the fractional species abundance (He II, He I, and H I) with feedback minus without feedback. Here, continuous opacity from hydrogen photoionization between He I is included. The effect of this opacity is shown separately in Fig. 2.8. In all cases, feedback retards formation of the neutral species. Here, the uppermost line is the first iteration, moving down with further iterations and better convergence.

iteration, multiply them by a suppression factor before applying them to the lower line. Changing variables to redshift, the total "line-line" depth of the continuum between lines is

$$
\tau_{\text{LL}} = \int_{z_{\text{em}}}^{z_{\text{abs}}} \frac{\mu_{\text{HI}}(z) x_{1s}(z) \sigma c}{H(z)} dz \frac{1}{1+z}
$$

(2.23)

Let $\mathcal{N}$ be the radiation field on the blue side of the lower line assuming there is no line-line optical depth $\tau_{\text{LL}}$. Then the non-thermal distortion produced by the higher state is $\mathcal{N}_{i}$ minus the Planck spectrum (since at times earlier than $z \sim 1600$ H I is in Saha equilibrium, to a good approximation), which is suppressed by the line-line depth. The final radiation field just above the frequency of the lower-lying line is then

$$
\mathcal{N}_{+} = (\mathcal{N}_{i} - \mathcal{N}_{\text{Pl}}) e^{-\tau_{\text{LL}}} + \mathcal{N}_{\text{Pl}},
$$

(2.24)

where the Planck spectrum is $\mathcal{N}_{\text{Pl}} = 1/(e^{h\nu/k_{B}T} - 1)$.

The effect of feedback is shown in Fig. 2.7, and the effect of including $\tau_{\text{LL}}$ is shown in Fig. 2.8.
2.3 Hydrogen continuum opacity and helium recombination

Figure 2.8: The effect of hydrogen continuum absorption on the feedback between transitions to the ground state in He I. Feedback slows He I recombination, but becomes increasingly less significant as the neutral hydrogen population grows, increasing the continuum opacity. The dashed trace here (with continuous opacity) is analogous to the middle panel of Fig. 2.7.

Coherent vs. incoherent processes

Radiative transport within the line (but between physically separate atoms) depends on the photon redistribution mechanisms in the line. The factor that is ultimately the most important is whether a photon escapes the resonance. We will see that this depends on the interaction between the redistribution mechanisms and loss processes (continuous opacity and redshifting). This section briefly outlines some of the physics and conventions necessary to set up a study of continuous opacity in the lines. Optically thick lines in cosmological recombination are usually treated by the Sobolev method, which assumes that photons are emitted and absorbed across the same profile. This is known as “complete redistribution.” We refer to the scattering that gives complete redistribution as “incoherent scattering”, in which the excited atom exchanges one or more photons of other energies with the radiation field before decaying to the lower level. Incoherent scattering follows the Voigt line profile \( \phi(\nu) \) for emission and absorption. The Voigt profile is a convolution of the natural resonance width and the thermal width of the transition.\(^{26}\)

Incoherent scattering is at odds with energy conservation for a direct excitation and decay \( l \rightarrow u \rightarrow l \), where the emitted photon must have the same energy as the incoming photon in the rest frame of the atom. Because of thermal motion in the gas, the photon energy will be redistributed depending on the kinematics of the scattering event and the struck atom’s thermal velocity distribution. This is referred to throughout as “coherent scattering” and conserves energy (in the atom’s rest frame). Coherent scattering complicates the problem because the frequency distribution of photons emitted in the line depends on the existing spectrum of radiation (the emitted photons have some “memory” of the incident distribution because of energy conservation). The mixture of complete redistribution across a Voigt profile and the redistribution from coherent scattering is a “partial redistribution.”

\(^{26}\)A review of line shapes and redistribution can be found in Hubeny (1985); Molisch and Oehry (1998), and in Krolik (1989, 1990). The Voigt profile is the Lorentzian natural line shape convolved by a Doppler Gaussian, and is the applicable profile here because collisional broadening is negligible, so more complex profiles such as Van Vleck-Weisskopf/Ben-Reuven (which are common in atmospheres and dense gases) are not needed.
Whether a photon scatters coherently or incoherently from an atom depends on the fraction of the time the excited state decays directly to the ground state (coherent) as opposed to visiting additional states (incoherent). For allowed transitions, decays directly to the lower level are far more abundant. Here, complete redistribution and the Sobolev approximation are formally incorrect. For forbidden transitions, the fraction of atoms that promptly decay to their lower state is very small relative to the fraction that visit additional states (through photons exchange with the thermal bath). This is because the upper states are connected with much higher rates to other levels than they are to their lower level, unless the upper state has all forbidden transitions out to other levels. Thus, forbidden transitions are very well approximation with complete redistribution.

What saves the ordinary recombination calculation is that the allowed lines are optically thick, so photons redshifting onto the line will rapidly fall into equilibrium, so the dynamics are largely insensitive to the redistribution and profile. Once H I also begins to compete for photons through its non-negligible optical depth, it can deplete photons near the core of the line, and thus radiative transport depends on the profile of the parent He I redistribution mechanism.

2.3.3 Transport within He I lines and continuous opacity: physical argument

The two natural scales in the line transport problem are: (i) the frequency $\eta c^{-1}$ that a photon can be expected to traverse by redshifting before it is absorbed in a hydrogen photoionization event, and (ii) the range of frequencies $\Delta \nu_{\text{line}}$ over which the line is thick to incoherent scattering/absorption. If $\eta c^{-1} \gg \Delta \nu_{\text{line}}$, then hydrogen continuous opacity is negligible as the photon interacts with the helium gas. On the other hand, if $\eta c^{-1} \leq \Delta \nu_{\text{line}}$, then H I can destroy photons that would otherwise have re-excited helium atoms. This accelerates He I recombination. It is important to note that the magnitude of $\eta c$ is irrelevant unless it is compared to a transport scale, such as to find a total depth,

$$\tau C = \eta c \Delta \nu_{\text{line}}.$$ (2.25)

In the case of lines that are optically thick into the damping wings, such as the He I $n^1 P^o - 1^1 S$ lines, $\Delta \nu_{\text{line}}$ may be approximated by integrating the asymptotic line profile $\phi(\nu) \sim \Gamma_{\text{line}}/4\pi^2 \Delta \nu^2$ until the optical depth in the wing becomes unity:

$$\Delta \nu_{\text{line}} = \frac{\Gamma_{\text{line}} \tau_{\text{inc}}}{4\pi^2}. \quad (2.26)$$

Here $\Gamma_{\text{line}}$ is the Lorentz width of the line, $\tau_{\text{inc}}$ is the Sobolev optical depth through the line from incoherent processes, and $\eta c$ is the differential continuum optical depth, Eq. (2.20). Fig. 2.9 compares the optically thick linewidth (from incoherent processes) to the inverse differential optical depth $\eta c^{-1}$ and the Doppler width for the $2^1 P^o - 1^1 S$ transition in $^4$He. The line may be optically thick in the Sobolev sense out to much larger frequency separations due to coherent scattering, but a coherent scattering event results in no net change in the atomic level populations and hence does not directly affect recombination. (It only has an indirect effect by changing the radiation spectrum.)

The general problem of the escape probability including the continuum opacity and coherent scattering as well as incoherent emission/absorption processes is complicated. Therefore we will solve it in two steps. In Sec. 2.3.4 we will solve the problem without coherent scattering. This is a conceptually simple problem – all one has to do is compute the probability of a photon

---

$^{27}$ Even for allowed lines such as the He I $2^1 P^o - 1^1 S$ line where the number of redistributing scatters is by far dominated by coherent scattering, the radiative transport is dominated by incoherent scattering. The reason for this is that incoherent scattering can transport a photon over much larger distances, possibly from the line core to the damping wing in one scatter, or remove the photon from the resonance entirely if the excited state is ionized. In contrast, coherent scattering can only transport over much smaller frequency scales related to the Doppler shift over the thermal velocities.
2.3 Hydrogen continuum opacity and helium recombination

Figure 2.9: Inverse of the differential optical depth \( \eta_c \) from hydrogen photoionization as a function of redshift, compared to the optically thick line-width due to incoherent processes in the \( 2^1P^o - 1^1S \) line. Continuum processes start to become important over scales inside the (incoherent) optically thick part within the line around \( z = 2100 \). Also plotted is the Doppler width of the line, emphasizing that the line is optically thick out into the wings. Continuum processes do not act on scales smaller than the Doppler core until \( z < 1800 \). Here it also clear that the Doppler width is small compared to the optically thick linewidth.

either redshifting out of the line or being absorbed by H I before being re-absorbed by He I – and the machinery for solving it has already been developed for the theory of line transfer in stellar winds (Hummer and Rybicki (1985)). Despite its simplicity, the solution in Sec. 2.3.4 is an accurate description of the intercombination and quadrupole lines because there is negligible coherent scattering in these lines. (The reason is that if an atom absorbs a photon in these lines and reaches a \( n^3P^o \) or \( n^1D \) level, its next step is almost always to undergo one of the allowed decays rather than to emit a photon in the intercombination or quadrupole line and go back to the ground state.)

2.3.4 The modified escape probability without coherent scattering

This section is concerned with calculating the net decay rate in He I lines with continuous opacity (from H I) but no coherent scattering. This is a line radiative transfer problem with “complete redistribution” in the sense that a photon emitted in the line has a frequency distribution given by the intrinsic line profile, so the probability distribution relating frequency in to frequency out, \( p(\nu_{\text{out}} | \nu_{\text{in}}) \) is independent of the incident radiation:

\[
p(\nu_{\text{out}} | \nu_{\text{in}}) = \phi(\nu_{\text{out}}).
\]  

There is negligible coherent scattering in the \( n^3P^o - 1^1S \) and \( n^1D - 1^1S \) lines (see Sec. 2.3.2), and results in this section are readily applied to those lines. The macroscopic picture is that in emission, a line in the gas always has the same shape, regardless of the excitation field. Throughout, \( \phi \) is Voigt-distributed and accounts for the natural resonance width of the line and Doppler broadening from thermal motion, set by the matter temperature.

The fundamental quantity that determines the transition rate between a lower level \( l \) and upper level \( u \) is the radiation phase space density integrated over the atomic absorption profile, \( \bar{N}(\nu_{ul}) \).
Recombination

In the case of complete redistribution through incoherent scattering in the presence of continuous opacity, the radiation phase space density evolves as

$$\dot{N} = \dot{N}_{\text{Hubble}} + \dot{N}_{\text{cont}} + \dot{N}_{\text{inc}}.$$  \hfill (2.28)

In this section, we will develop each of these terms and solve the transport equations for the radiation phase space density over the line and the escape probability. Of these terms, the Hubble redshifting term is,

$$\dot{N}_{\text{Hubble}} = H(z) \nu_{\text{ul}} \frac{\partial N}{\partial \nu}.$$  \hfill (2.29)

We will work in the steady-state approximation, where $\dot{N} = 0$. Then, frequency provides a convenient domain over which to solve for $N$, moving the Hubble term and dividing by $H \nu_{\text{ul}}$.

Breaking up $\dot{N}_{\text{cont}}$ and $\dot{N}_{\text{inc}}$ into emission and absorption pieces, there are four terms that appear in the steady-state equation for the phase space density: absorption and emission by continuum processes and emission and absorption by incoherent processes in the line. We may write these as

$$\frac{\partial N}{\partial \nu} = \left( \frac{\partial N}{\partial \nu} \right)_{\text{cont-abs}} + \left( \frac{\partial N}{\partial \nu} \right)_{\text{inc-abs}} + \left( \frac{\partial N}{\partial \nu} \right)_{\text{cont-em}} + \left( \frac{\partial N}{\partial \nu} \right)_{\text{inc-em}}.$$  \hfill (2.30)

The continuum absorption term has already been determined (Eq. 2.20) and is $\eta_c N(\nu)$. The line absorption term "inc-abs" depends on the line profile and is $\tau_S \phi(\nu) N(\nu)$ since $\tau_S \phi(\nu)$ is the optical depth per unit frequency. The continuum emission term is

$$\left( \frac{\partial N}{\partial \nu} \right)_{\text{cont-em}} = -\eta_c N_C,$$  \hfill (2.31)

where $N_C$ is the phase space density of photons that would be required for the reaction

$$\text{H}(1s) + \gamma \leftrightarrow \text{H}^+ + e^-$$  \hfill (2.32)

to be in equilibrium. Because hydrogen is so nearly in Saha equilibrium, this is just

$$N_C(\nu) = \left( \frac{n_e n_C}{n_i} \right) \left( \frac{n_i}{n_e n_C} \right)^{\text{LTE}} e^{-h\nu/k_B T_{\text{e}}} \approx e^{-h\nu/k_B T_{\text{T}}},$$  \hfill (2.33)

The $n^3P^o - 1^1S$ and $n^1D - 1^1S$ lines are at high energies ($\geq 20.6$ eV), so the photon phase space density $N \ll 1$ (recall that the typical temperature during recombination is $\sim 0.25$ eV at $z = 1100$ so 20.6 eV is in the Wein tail) and we can neglect stimulated emission and other consequences of the photon’s bosonic nature. The line emission term is

$$\left( \frac{\partial N}{\partial \nu} \right)_{\text{inc-em}} = -\tau_S \phi(\nu) N_L,$$  \hfill (2.34)

where $N_L$ is the phase space density that would exist if only the line processes were important. This can also be determined by setting the excitation rate $x_l(g_u/g_l)A_{u\to l}N_L$ equal to the de-excitation rate $x_u A_{u\to l}(1 + N_L)$:

$$N_L = \frac{x_u}{x_l(g_u/g_l) - x_u}.$$  \hfill (2.35)

---

28 see Hirata and Switzer (2008) for a discussion of the time-dependent solution.
2.3 Hydrogen continuum opacity and helium recombination

The transport equation is thus

\[
\frac{\partial N}{\partial \nu} = \eta_c(N - N_C) + \tau_S \phi(\nu)(N - N_L).
\]  (2.36)

In Appendix A.2 we directly integrate the transport equation to find the radiation phase space density around the line. The effect of continuum opacity on radiation near an intercombination line (where complete redistribution valid) is shown in Fig. 2.10. The crucial quantity for atomic rates is the integral of the phase space density across the profile,

\[
\bar{N} = \int_{-\infty}^{\infty} \phi(\nu) \mathcal{N}(\nu) d\nu.
\]  (2.37)

The net downward transition rate from \( u \) to \( l \) is

\[
\dot{x}|_{\text{line}} = A_{u \rightarrow l} \left[ x_u (1 + \bar{N}) - \frac{g_u}{g_l} x_l \bar{N} \right].
\]  (2.38)

For \( x_u \ll x_l \) and \( \bar{N} \ll 1 \), this can be re-expressed in terms of \( N_L \) as

\[
\dot{x}|_{\text{line}} = A_{u \rightarrow l} \left[ \frac{g_u}{g_l} x_l (N_L - \bar{N}) \right],
\]  (2.39)

In the appendix, we solve for \( \bar{N} \) to find

\[
\dot{x}|_{\text{line}} = A_{u \rightarrow l} \left[ \frac{g_u}{g_l} x_l (\bar{N}_L - \bar{N}) \right],
\]  (2.40)

\[
\begin{align*}
\Delta I_L &= I_L - \bar{I}_L \bigg|_{\nu_2 = 0} = -\int_{-\infty}^{\nu_2} \phi(\nu) \int_{\nu_2}^{\nu} \tau_S \phi(\tilde{\nu}) \\
& \times \exp \left\{ -\tau_S \int_{\nu}^{\tilde{\nu}} \phi(y) dy \right\} \left[ e^{-\eta_c(\tilde{\nu} - \nu)} - 1 \right] d\tilde{\nu} d\nu.
\end{align*}
\]  (2.41)

Here \( \nu_2 \to \infty \).

The Voigt profile presents two characteristic integration regions: slowly varying functions in wings, and rapidly-varying functions in the core. We use the Gubner’s series (Gubner (1994)) in the core region and switch to a fourth order asymptotic expansion of the Voigt profile in the wings.\(^{30}\)

\(^{29}\)Formally the optical depth to the boundary condition taken at infinity here is infinite for finite \( \eta_c \), so we have assumed the incident radiation is Planckian \( N_C \) where in actuality \( N_C \) could include spectral distortion from a higher transition. Yet, the width of helium lines is much less than the separation between lines, so if continuous opacity is important enough to impact the escape probability, it has certainly also absorbed the spectral distortion through a high optical depth between lines. For consistency with the rate equation in the level code with feedback (to recover the Sobolev approximation at early times) there is no loss in replacing \( N_C \) with \( N_+ \) as

\[
\dot{x}|_{\text{line}} = P_{\text{esc}} A_{u \rightarrow l} \left( x_u - \frac{g_u}{g_l} x_l N_+ \right).
\]  (2.42)

\(^{30}\)In the boundary between the two regimes, we estimate the error in the asymptotic expansion by using the next higher order and ensure that differences between the two approaches are negligible.
Figure 2.10: The radiation phase space density near the intercombination line $1^1S_0 \rightarrow 2^3P_1$ resonance with Voigt parameter $a = \Gamma_{\text{line}}/(4\pi\Delta\nu_D) = 10^{-5}$ for the Sobolev optical depth $\tau_S = 2.8$ and $N_C = 0$, for several sample continuum optical depths $\eta_c\Delta\nu_D$. Here, $\Delta\nu_D$ is the Doppler width, so that $\eta_c\Delta\nu_D$ is the opacity for redshifting through a Doppler width. Here we have normalized the radiation phase space density by its equilibrium value with the line $(N/N_L)$ on the y-axis. The x-axis is the detuning from line center in Doppler widths, $x = (\nu - \nu_o)/\nu_D$. $N$ does not converge to $N_L$ on the red side of the line because the depth $\tau_S = 2.8$ is not very thick, so even for zero continuous opacity the radiation is not fully driven into equilibrium with the line (only to 94%). Because of the low optical depth and small natural line width, very little radiation extends more than three Doppler widths above the line, and the effect of the continuum is significant in relaxing the radiation phase space density near line-center.
2.3 Hydrogen continuum opacity and helium recombination

Doppler width, exploiting the symmetry of the profile and its integral. These are interpolated using a cubic spline and Eq. (2.42) is integrated using a 61-point Gauss-Kronrod adaptive integration scheme. To apply the numerical results of this integral to the recombination setting, $P_C$ is pre-calculated for a range of $\tau_{inc}$ and $\eta_D\Delta \nu_D$, and then log-interpolated. Additional discussion of the integral can be found in Rubiño-Martín et al. (2008); Switzer and Hirata (2008a). The Monte Carlo methods developed in Sec. 2.3.7 also generate a grid of modified escape probabilities as a function of redshift and the occupation of the He I ground state. These are log-interpolated and applied in to the recombination level code, as described in Sec. 2.3.8.

2.3.5 An overview of transport with coherent scattering

There are three outcomes for an atom in some level $u$. It could decay to the ground level $l$ with the rate $A_{u \rightarrow l}$. It may also decay to another level $a$ with lower energy $E_a < E_u$, with rate $A_{u \rightarrow a}[1 + N(\nu_{ua})]$. A third possibility is that an atom in level $u$ could absorb a photon and transit to a higher level $b$ with rate $A_{b \rightarrow u}(g_b/g_u)N(\nu_{bu})$. The overall width of the level $u$ is then the sum of all the rates out (where the summation over $b$ includes an implied integration over continuum states.)

$$\Gamma_u = A_{u \rightarrow l} + \sum_{a < u, a \neq l} A_{u \rightarrow a}[1 + N(\nu_{ua})]$$

$$+ \sum_{b > u} A_{b \rightarrow u} \frac{g_b}{g_u}N(\nu_{bu}).$$

(2.43)

It is convenient to define the rates

$$R_{ui} = \begin{cases} A_{u \rightarrow i}[1 + N(\nu_{ui})] & E_i < E_u \\ A_{i \rightarrow u}(g_i/g_u)N(\nu_{iu}) & E_i > E_u \end{cases}$$

(2.44)

$$R_{ul} = A_{u \rightarrow l},$$

(2.45)

so that $\Gamma_u = A_{u \rightarrow l} + \sum_i R_{ui}, i \neq l$. The fraction of atoms in $u$ that transit from $u$ to some level $i$ is the $f_i \equiv R_{ui}/\Gamma_u$ (a branching fraction to $i$ from $u$). The radiative rate for one-photon radiative transitions from some level $i \neq l$ back to $u$ is

$$R_{lu} = \begin{cases} A_{u \rightarrow l}(g_u/g_l)N(\nu_{ul}) & E_i < E_u \\ A_{l \rightarrow u}[1 + N(\nu_{lu})] & E_i > E_u \end{cases}. $$

(2.46)

The radiative rate for one-photon excitation from $l$ to $u$ is similar, except that it will be subject to spectral distortions and have its own non-trivial radiation profile (compared to $u \leftrightarrow c$, which are optically thin so do not modify the thermal field). The excitation rate is then proportional to the integral of the phase space density over the line profile ($N$) as

$$R_{lu} = A_{u \rightarrow l} \frac{g_u}{g_l}N.$$

(2.47)

Figure 2.11 summarizes the rates $R_{lu}$, $R_{lu}$, $R_{iu}$, and $R_{lu}$ for the lower excitation states of helium.

The fraction of absorptions $l \rightarrow u$ that are re-emitted coherently is the branching fraction to fall directly back to $l$, $f_{coh} \equiv f_l$. The fraction of $l \rightarrow u$ that absorb another photon an transit to another

---

31 We simplify the problem by 1) neglecting stimulated emission (the line frequencies are in the Wein tail and the distortion is negligible) 2) only considering spectral distortions at $\nu_{ul}$, but not between the upper levels 3) taking only the transitions from the upper state to the ground state to be optically thick, and 4) assuming the cosmological evolution of $x_i$ is slow compared to any atomic rates. Here we describes the fundamental considerations and reproduce the major results. Interested readers should see sections by C. Hirata in Switzer and Hirata (2008a) for a more complete discussion and proofs of several statements here.
Figure 2.11: Coherent and incoherent scattering through $2^1P^o \rightarrow 1^1S$. If He $1^1S_0 \rightarrow 2^1P^o_1$ is excited by one photon, and a second photon excites $1^1S_0 \rightarrow 2^1P^o_1$ (left panel), then if the atom decays back through $2^1P^o_1 \rightarrow 1^1S_0$ (right panel), the outgoing energy will be completely redistributed across the Voigt profile. We refer to this as “incoherent” scattering. If, rather than absorbing a second photon and visiting additional levels, the atom decays $2^1P^o_1 \rightarrow 1^1S_0$ promptly after it was excited through $1^1S_0 \rightarrow 2^1P^o_1$, then energy will be conserved in the atom’s rest frame and in the gas, the outgoing photon energy will only be thermally broadened and not completely redistributed across the Voigt profile. We refer to this as “coherent” scattering. The distinction here is that the atom cannot “forget” the incoming state energy, as it can in the multi-photon case where there are additional excursions and energy does not have to be conserved solely through $1^1S_0 \rightarrow 2^1P^o_1$. The combination of these redistribution processes leads to partial redistribution.

level is then $f_{\text{inc}} \equiv 1 - f_i$; if this excursion ends with a transition back through $u \rightarrow l$, then the emitted photon will have no “memory” of the incoming photon energy that originally excited $l \rightarrow u$, and so will be re-emitted across the Voigt profile (complete redistribution). This allows us to define an optical depth to incoherent scattering $\tau_{\text{inc}} \equiv \tau_{S}(1 - f_i)$ and coherent scattering $\tau_{\text{coh}} \equiv \tau_{S}f_i$.

### 2.3.6 Atomic occupation rate evolution with coherent scattering

The radiation phase space density integrated across the line profile $\bar{N}$ regulates absorption (and stimulated emission) rates and is the most non-trivial part of the calculation for the atomic rates describing the evolution of the occupation fractions. The radiation phase space density is shaped by four processes: redshifting of the photons from Hubble expansion; absorption (and emission) by neutral hydrogen; coherent scattering; and incoherent emission/absorption processes. Schematically, the Boltzmann equation for radiation is then

$$\dot{\bar{N}} = \dot{\bar{N}}_{\text{Hubble}} + \dot{\bar{N}}_{\text{cont}} + \dot{\bar{N}}_{\text{coh}} + \dot{\bar{N}}_{\text{inc}}.$$  \hspace{1cm} (2.48)

It is good approximation and convenient to assume a steady state radiation field where $\dot{\bar{N}} = 0$ (see Hirata and Switzer (2008)). Then the partial differential equation in $\nu$ and $t$ is just an differential equation in $\nu$. In Switzer and Hirata (2008a) we show that this equation in $\nu$ is

$$\frac{\partial \bar{N}}{\partial \nu} = \eta_c[\bar{N}(\nu) - \bar{N}_C]$$

$$+ \tau_{\text{coh}} \left[ \phi(\nu)\bar{N}(\nu) - \int \phi(\nu')\bar{N}(\nu')p(\nu|\nu') \, d\nu' \right]$$

$$+ \tau_{\text{inc}}\phi(\nu)[\bar{N}(\nu) - \bar{N}_L^{(0)}].$$  \hspace{1cm} (2.49)
2.3 Hydrogen continuum opacity and helium recombination

Note that this is nearly identical to Eq. 2.36, except for the new term which contains the scattering kernel for coherent scattering, $p(\nu|\nu')$ and a modified equilibrium with the line

$$N_L^{(\nu)} = \frac{g_i}{g_u \Gamma_u f_{inc}} \sum_{\nu \neq \nu'} x_i R_{iu}. \quad (2.50)$$

In the transport equation, $p(\nu|\nu')$ is diffusive (for a large number of scatterings) and “blurs” the $N(\nu)$ profile across the line. This only impacts transitions that have a significant coherent scattering fraction, such as the allowed lines (the forbidden lines scatter almost incoherently so this effect is negligible there). The effect of this diffusion is usually insignificant in the Sobolev approximation because $N(\nu)$ is in equilibrium with the line over its optically thick width, and this “blurring” does not change the constant level. Once continuum opacity is included, however, the radiation profile is no longer flat, so redistribution can have some impact on $N(\nu)$ over the optically thick width of the line, and this impacts the escape probability. It is convenient to make a change of variables to eliminate $N_C$ and $N_L^{(\nu)}$ from the transport equation, where

$$\xi(\nu) = \frac{N'(\nu) - N_C}{N_L^{(\nu)} - N_C} \implies N'(\nu) = N_C + (N_L^{(\nu)} - N_C)\xi(\nu). \quad (2.51)$$

The goal is to solve the transport equation for $N'$ or $\xi$, and find the escape probability to use in the recombination level code by integrating this $N'$ or $\xi$ over the line profile. Recall that the decay rate (neglecting stimulated emission) is

$$\dot{x}_{|u} = A_{u\rightarrow l} \left( x_u - \frac{g_u}{g_l} x_l N' \right) = \frac{g_u}{g_l} A_{u\rightarrow l} x_l (N_L - N'). \quad (2.52)$$

In Switzer and Hirata (2008a), we make the change of variables from $N'$ to $\xi$ (in Eq. 2.51) and show that the transport with coherent scattering can be encapsulated into an escape probability (just as in the usual Sobolev approximation), where

$$\dot{x}_{|u} = P_{esc} A_{u\rightarrow l} \left( x_u - \frac{g_u}{g_l} x_l N'_\xi \right), \quad (2.53)$$

and $P_{esc}$ is related to the $\xi$ through

$$P_{esc} = \frac{1}{1 - f_{coh}} \left( \frac{1}{1 - \xi} \right) f_{inc} \quad \text{where} \quad \bar{\xi} \equiv \int \xi(\nu)\phi(\nu) \, d\nu. \quad (2.54)$$

The transport equation, Eq. 2.49 is an integro-differential equation that is difficult to solve directly. Our approach will be to solve for it using a Monte Carlo, in the next section. Rather than solve for the radiation field itself and integrate it over the profile, it is more convenient to solve for a probability that can be estimated by the Monte Carlo. In Switzer and Hirata (2008a), we show that $1 - \xi$ can be understood as the escape probability with respect to incoherent scattering, $P_{inc}$, as

$$P_{esc} = \frac{f_{inc} P_{inc}}{1 - f_{coh} P_{inc}}. \quad (2.55)$$

---

32Note that $N_C$ is taken as the input radiation field on the blue side of the line, and that this is the Planck distribution. This is the result of a particular assumption that we have made about the boundary conditions. For some finite continuum opacity, the integral to $\nu \rightarrow +\infty$ has formally infinite optical depth, so the phase space density on the blue side of the line ($N_{\nu}$) will always be driven to its equilibrium value. In actuality, there is feedback of the spectral distortion from a higher-lying line onto this line.

33This is directly analogous to the derivation in Appendix A.2 where we find that $N' = N_C P_{inc} + N_L (1 - P_{inc})$, except that here we have defined $N'(\nu) = N_C + (N_L^{(\nu)} - N_C)\xi$ and use $N_L^{(\nu)}$ rather than $N_L$. Thus $P_{inc} = 1 - \bar{\xi}$. 

---

36
In the next section, the Monte Carlo will solve for the probability $P_{\text{inc}}$ including both continuum opacity and coherent scattering, but for now, we can use the ordinary Sobolev escape probability for escape with respect to incoherent scattering

$$P_{\text{S,inc}} = 1 - e^{-f_{\text{inc}}\tau_{\text{S}}}. \tag{2.56}$$

For forbidden lines where incoherent scattering dominates, $f_{\text{inc}} \sim 1$ and this is the Sobolev probability normally used in level codes. For allowed lines where coherent scattering dominates and the optical depth is high, the escape probability is low and Eq. 2.54 gives

$$P_{\text{esc}} \approx f_{\text{inc}} \frac{1 - e^{-f_{\text{inc}}\tau_{\text{S}}}}{f_{\text{inc}}\tau_{\text{S}}} \approx \frac{1}{\tau_{\text{S}}} \approx P_{\text{S}}. \tag{2.57}$$

That is, the result coincides with the ordinary Sobolev approximation. The only time for which the ordinary Sobolev method is not accurate (still neglecting coherent scattering and continuum opacity) is when the optical depth is low and the scattering is mainly coherent. This is not a common circumstance during recombination, so the traditional Sobolev approach remains accurate. Once continuum opacity is added, the picture is more complicated and the goal of the next section is to understand the escape probability in this case.

### 2.3.7 Monte Carlo method

The basic plan of the Monte Carlo simulation is as follows: we begin by injecting a photon with frequency distribution drawn from the Voigt line profile, $\phi(\nu)$. We simulate its fate by assuming that it redshifts at the rate $\dot{\nu} = -H\nu_{\text{ul}}$; undergoes coherent scattering with probability per unit time $H\nu_{\text{ul}}\tau_{\text{S}}\phi(\nu)$; undergoes continuum absorption with probability per unit time $H\nu_{\text{ul}}\eta_{c}$; and undergoes incoherent absorption with probability per unit time $H\nu_{\text{ul}}\tau_{\text{S}}(1 - f_{l})\phi(\nu)$. The simulation is terminated if the photon is absorbed in the H I continuum, if it redshifts out of the line, or if it undergoes incoherent absorption in He I. Note that within the idealized conditions $\eta_{c} = \text{constant}$ of the simulation, a photon that redshifts out of the line will eventually be absorbed by the continuum so long as $\eta_{c} > 0$ (though in reality the photon would eventually reach other He I lines or redshift to below 13.6 eV). Therefore only the total probability of these two results is meaningful. We thus denote by $P_{\text{MC}}$ the probability that a photon in the Monte Carlo is terminated by redshifting out of the line or by continuum absorption, and let $1 - P_{\text{MC}}$ denote the probability that the photon is terminated by incoherent absorption.

It can be shown that the Monte Carlo method solves the transport problem (see Switzer and Hirata (2008a) and that the escape probability is simply related to the probability estimated in the Monte Carlo as described in the previous section

$$P_{\text{esc}} = \frac{P_{\text{MC}}f_{\text{inc}}}{1 - f_{\text{coh}}P_{\text{MC}}}. \tag{2.58}$$

---

34 A variety of approaches have been taken in the literature for solving line radiative transfer equations including coherent scattering terms. One approach is the diffusive, Fokker-Planck approximation (Hummer and Rybicki (1992)) which replaces the redistribution integral (Eq. 2.49) with a second-order differential operator. This results in a second-order ODE instead of integro-differential equation, which is a substantial improvement for most numerical techniques. The other possibilities are to discretize the transport problem and solve it as a linear algebra problem or to apply a Monte Carlo method; see Hirata (2006); Zheng and Miralda-Escudé (2002); Bonilha et al. (1979); Bernes (1979); Caroff et al. (1972); Auer (1968); Avery and House (1968). The latter two have the advantage of being usable in the Doppler core of the lines, which we expect to be important since for the He I $2^1P_0 - 1^1S$ lines, the width of the line that is optically thick to incoherent processes $\Delta\nu_{\text{line}}$ is only $\sim 30\Delta\nu_{\text{D}}$ during most of He I recombination. Therefore we have not used the Fokker-Planck approach, which we believe is better suited for studying the far damping wings of very optically thick lines such as H I Lyα (The Fokker-Planck operator assumes that many scattering events transport a photon over a region where the line shape varies slowly; yet in the core of the line, single scatterings can transport a photon over the width of the core.) We have chosen the Monte Carlo approach here, mainly because we had a pre-existing code that was capable of handling the problem with minor modifications (Hirata (2006)).
2.3 Hydrogen continuum opacity and helium recombination

2.3.8 Incorporation in the level code

In principle, modifications to transport in the entire $n^1P^o - 1^1S$ series, the quadrupole series ($n^1D - 1^1S$), and intercombination series ($n^3P^o - 1^1S$) of He I could lead to acceleration of He I recombination. To make the problem computationally tractable, the hierarchy of higher order $n$ contributions needs to be cut off where there are diminishing returns. The integral solution to the transport equations with complete redistribution (Sec. 2.3.4), while not accurate for $n^1P^o - 1^1S$ rates, gives a quick estimate of whether, for example, continuum effects become more and more important for higher $n$, or die away. We evaluate the probabilities at $z = 1606$, at the end of He I recombination where the continuum effects are expected to be largest, using the integral Eq. (2.42). We find that impact of continuum opacity on the intercombination lines for $n > 4$ is negligible. For quadrupole transitions ($n^1D - 1^1S$), the corrections are significant out to moderate $n$, (modifying the escape probability by a factor of $\sim 2$ at $n = 8$). In the $n^1P^o - 1^1S$ series, continuum effects lead to significant modifications to the escape probability (by a factor of $\sim 100$ at $n = 9$ toward the end of He I recombination)–the entire $n^1P^o - 1^1S$ series sees significant corrections. (One mitigating factor that makes the calculation practical is that only the $2^1P^o - 1^1S$ line dominates recombination rates.) Based on these estimates, we calculate corrections to transport within the line up to $n = 6$ in the $n^1P^o - 1^1S$ series, up to $n = 4$ in the intercombination lines ($n^3P^o - 1^1S$), and $n = 6$ in the quadrupole lines ($n^1D - 1^1S$). The $n^1P^o - 1^1S$ series is treated via the Monte Carlo technique, while the others are treated here using the integral method of Sec. 2.3.4. (Fig. 2.12 shows that higher allowed transitions are negligible.) Fig. 2.13 shows the escape probabilities subject to H I opacity (with and without coherent scattering) for $2^1P^o - 1^1S$ derived from the Monte Carlo, relative to the ordinary Sobolev results.

Both the Monte Carlo and analytic integral methods of finding the escape probability are prohibitively slow to run in real-time with the level code. In the Monte Carlo method, computing the large number of coherent scatters for one photon trajectory in the He I $n^1P^o - 1^1S$ series is computationally intensive. In the analytic integral method developed for complete redistribution, the variety of scales in the line profile gives, generally, integrands with large dynamic ranges that need to be known to high accuracies. For this reason, we generate tables of the escape probability over a range of parameters and log-interpolate to find the probabilities in the level code.

The escape probability is a function of the coherent, incoherent, and continuum optical depths, and the matter temperature, which sets the Gaussian width for the line,

$$P_{\text{esc}}(\{\tau_{\text{coh}}, \tau_{\text{inc}}, \eta_k, T_m, \Gamma_{\text{line}}\})$$

(2.59)

The wing width is set by $\Gamma_{\text{line}}$. We then switch to a more natural parameter space for the level code where

$$\{\tau_{\text{coh}}, \tau_{\text{inc}}, \eta_k, T_m, \Gamma_{\text{line}}\} \rightarrow \left\{ z, \frac{n_{\text{HI}}}{n_{\text{HI,Saha}}}, x_{\text{HeI}} \right\},$$

(2.60)

since the parameters on the right hand side specify unique values for the transport parameters on the left hand side. We calculate tables for 11 linearly-spaced $z$ values between 1400 and 3000, inclusive, and 21 logarithmically-spaced $x_{\text{HeI}}$ values from $2 \times 10^{-5}$ to 0.08. When we double the fineness of the probability grid over the parameters, the change in free electron density $|\Delta x_e|$ has a maximum of $\sim 5 \times 10^{-4}$ at $z \sim 9000$. The neutral hydrogen population is taken to be the evolution determined by the reference level code. This is nearly Saha until the end of He I recombination, by which time we find that He I has relaxed to equilibrium. (That is, whether you assume the H I population is Saha or evolves through in the full recombination treatment should matter little for the evolution of the He I ground state: even before neutral hydrogen departs significantly from equilibrium, He I is already relaxed into equilibrium, by that point.) This set of probabilities required $\sim 4$ days to evaluate over 50 x 3.2 GHz computer nodes. The convergence criterion is described in Appendix A.5, and shown to converge to 1.25% fractional error in probability with negligible
bias. Doubling the number of Monte Carlo trials led to a maximum change of $|\Delta x_e| < 10^{-4}$, and resampling with a different random generator (Numerical Recipes ran2 instead of ran3 [Press et al. (1992)]) gives a change $|\Delta x_e| < 1.5 \times 10^{-4}$. (More convergence tests are described in Appendix A.6.)

The result of including these in the level code is shown in Fig. 2.14. Here, we can test the diminishing-returns set of modified lines by running a level code with: 1) just $2^1P^o - 1^1S$ modifications, 2) modifications to the $n^1P^o - 1^1S$ series up to $n = 6$, 3) $2^1P^o - 1^1S$ and the 3 intercombination lines, and 4) the $n^1P^o - 1^1S$ series up to $n = 6$, plus the three intercombination lines, shown in Fig. 2.12. It is apparent that most of the effect is due to modification to the $2^1P^o - 1^1S$ escape probability, leading to relaxation of the '$n = 2$' bottleneck and further that coherent scattering is only a small contribution to the recombination history. This greatly improves prospects for including continuum effects in a practical level code through modifications to $2^1P^o - 1^1S$ under the approximations developed in Sec. 2.3.4 (without coherent scattering).
2.4 Additional effects

Figure 2.13: The modified escape probability from $^4\text{He}\ 2^1 P^o \rightarrow 1^1 S$ during He I recombination, comparing the results of the standard Sobolev approximation and the modification due to continuous opacity with and without coherent scattering. When we include the continuous opacity developed in Sec. 2.3.3, the probability that a photon escapes the line is more than an order of magnitude larger than in traditional approaches by $z \sim 2000$. Once coherent scattering is introduced, photon diffusion effects increase the escape probability by increasing the span of frequencies traversed by the scattering photon before it escapes. These probabilities are log-interpolated over the grid of $x_{\text{HeI}}$ and $z$ used in the level code. We find that the effect of doubling the grid resolution and with it the smoothness of the interpolated probability is negligible.

2.4 Additional effects

In the series of papers Switzer and Hirata (2008a); Hirata and Switzer (2008); Switzer and Hirata (2008b), we show that there many other effects to consider which are ultimately negligible to the recombination history. These are summarized in Table 2.2. In this section, we describe two effects that are interesting and give intuition for the physics of the era. These are Thomson scattering (which can transport photons across a wide range of frequencies compared to the atomic transition profiles) and peculiar velocities (which could broaden the atomic line profiles analogously to microturbulence if velocity perturbations are incoherent over radiative transport length scales).

2.4.1 Thomson scattering

During cosmological helium recombination, the bulk of electrons have not recombined so they present an optical depth to Thomson scattering. Because electrons have a large thermal velocity dispersion relative to $^4\text{He}$ atoms, Thomson scattering can transport photons across a wide range of frequencies in a single scattering event. Thus, even if Thomson scattering within the line is rare, it could still impact the escape probabilities and recombination history. $^{35}$ This is important because the phase space density is not flat across all frequencies, and on average, Thomson scattering will transport photons from pools where there are more photons to pools where there are fewer. In

---

$^{35}$Electron scattering should also be folded into radiative transport in H I recombination. This is beyond the scope of discussion here and is deferred to later work. This is a harder problem because of the high optical depth, broad linewidth and partial redistribution in the H I $2p \rightarrow 1s$ system, as well as the more stringent accuracy requirements. The very high optical depth (of order $10^9$) may require a Fokker-Planck or hybrid approach to the problem instead of the Monte Carlo methods applied here.
Figure 2.14: The free electron fraction $x_{e}$ as a function of redshift $z$ during He I recombination with the inclusion of feedback and continuum opacity in He I lines, compared to a Saha He I recombination history and the standard He I recombination history. (Here we estimate the standard history using the same code, but turn off continuous opacity and feedback, this is labelled "No feedback or continuous opacity" in the key) Note that through the choice to normalize $x_i = n_i/n_H$, (rather then $n_h$) the free electron fraction exceeds 1 during this interval. Continuous opacity starts to become important at $z \sim 2100$ (as suggested by Fig. 2.9), and pushes the He I evolution to a Saha steady-state by $z \sim 1800$. The beginning of H I recombination is visible here starting at $z \sim 1700$, and for later times $x_e$ drops precipitously. The modifications to He I recombination suggested here are twofold: at early times during He I recombination, feedback slows recombination (Sec. 2.2.2), and at later times, continuous opacity (Sec. 2.3.3) accelerates He I recombination relative to standard models.
2.4 Additional effects

| Effect | Dir. | $|\Delta x_{e,\text{max}}|$ | ± |
|--------|------|-----------------|----|
| Systematic corrections due to effects: | | | |
| Opacity within lines ($2^2P^o - 1^1S$) | − | $2.5 \times 10^{-2}$ | $\sim 1800$ |
| Feedback between $2^3P^o - 1^1S$ and $2^3P^o - 1^1S$ | + | $1.5 \times 10^{-2}$ | $1800 - 2600$ |
| Continuum opacity modification to feedback | − | $5 \times 10^{-3}$ | $\sim 1800$ |
| $2^3P^o - 1^1S$ inclusion | − | $3 \times 10^{-3}$ | $\sim 1900$ |
| $n^3P^o - 1^1S$, $n \geq 3$ inclusion | ± | $4 \times 10^{-5}$ | $\sim 2000$ |
| $n^1D - 1^1S$ inclusion | ± | $3 \times 10^{-4}$ | $\sim 1900$ |
| Opacity in $n^1P^o - 1^1S$ and $n^3P^o - 1^1S$ (for $n \geq 3$), $n^1D - 1^1S$ | − | $5 \times 10^{-4}$ | $\sim 1900$ |
| Coherent scattering in $2^1P^o - 1^1S$ | − | $2 \times 10^{-4}$ | $\sim 2000$ |
| Distortion, thermal stimulated two-photon effects | + | $4 \times 10^{-5}$ | $2000 - 3000$ |
| Electron scattering | ± | $3 \times 10^{-4}$ | $1800 - 2800$ |
| Uncertainty in the effect’s magnitude: | | | |
| Finite linewidth in He I | ± | $4 \times 10^{-4}$ | $1800 - 3000$ |
| Non-resonant two-photon effects from $n > 2$ | ± | $5 \times 10^{-4}$ | $\sim 2000$ |
| $\pm 50\%$ $2^3P^o - 1^1S$ spontaneous rate | ± | $10^{-3}$ | $\sim 1900$ |
| Uncertainty from the numerical implementation: | | | |
| Modified escape probability grid refinement | + | $5 \times 10^{-4}$ | $\sim 1900$ |
| Monte Carlo resampling | + | $2 \times 10^{-4}$ | $\sim 1900$ |
| Monte Carlo sample size doubling | ± | $10^{-4}$ | $\sim 1800$ |
| Level code: half step size | ± | $3 \times 10^{-5}$ | $\sim 1800$ |
| Convergence in the modified $n^1P^o - 1^1S$ series modified $P_{esc}$ | − | $3 \times 10^{-4}$ | $\sim 1900$ |
| $n_{\text{max}} = 45$ relative to $n_{\text{max}} = 100$ | + | $4 \times 10^{-5}$ | $\sim 2300$ |

Table 2.2: Summary of the magnitude of effects described in Switzer and Hirata (2008a); Hirata and Switzer (2008); Switzer and Hirata (2008b) of which the discussion in this chapter was a subset. From the top category to the bottom, we distinguish the magnitude of an effect (the systematic error if it is not included) from the uncertainty in an effect, and the uncertainty in the implementation. Unsigned upper bounds on $|\Delta x_e|$ are indicated by “±”. Note that these are only meant to give an order-of-magnitude bound on the effect, where more detail is available in the section cited. The effect of forbidden processes includes H I opacity in their transport and feedback. The direction shown is either + (increases $x_e$) or − (decreases $x_e$), and a $\sim$ indicates that only an order-of-magnitude calculation is necessary to find that the effect is negligible.
particular for recombination, the red side of a line has more photons from the distortion produced by the line, and so Thomson scattering will inject more of these, on average to the blue side of the line. Thus a photon that would have escaped no longer can, and most of the time Thomson scattering will delay the line. Thus a photon that would have escaped no longer can, and most of the time Thomson scattering will inject more of the se, on average to the blue side of the line, and so Thomson scattering will illuminate some of the physics of radiative transport during recombination.

Electron scattering would be expected to become important if $\Delta \nu_{\text{line}}$ is flat in frequency and the struck atom’s velocity is not conditioned on the incoming photon frequency. The recoil term $\alpha$ is from differences in the electron kinetic energy and can be dropped. This is nearly identical to the case of coherent line scattering, except that the scattering cross section is flat in frequency and the struck atom’s velocity is not conditioned on the incoming photon frequency.

Thomson scattering can be trivially added to the Monte Carlo by adding an additional scattering process with differential depth $\eta_e$, from Eq. (2.61). As before, we simulate the escape probability over a grid in $\{x_{\text{HeI}}, z\}$ with continuous opacity derived from H I populations in Saha equilibrium (accurate until the end of He I recombination, where it ceases to matter anyway, once He I has almost fully recombined). The grid is taken over 11 linearly-spaced points in redshift spanning $z = 2500$ in Fig. 2.15, and is much larger than the Doppler width of the $^{4}\text{He}$ line, $\sim 50$ GHz.

Electron scattering shares some, but not all of the features of continuous opacity from neutral hydrogen. Like continuous opacity in H I, electron scattering presents a differential optical depth $\eta_e$ approximately flat in frequency and is thermally distributed ($\parallel$ and $\perp$ denote the components parallel to the direction of propagation of the incident photon, and perpendicular to it but in the plane of scattering, respectively), for the photon frequency $\nu_0$, $\alpha = h \nu_0^2/(m_e c^2)$, and $\chi$ is the scattering angle between the direction of propagation of the incoming/outgoing photons. The recoil term $\alpha$ is from differences in the electron kinetic energy and can be dropped. This is nearly identical to the case of coherent line scattering, except that the scattering cross section is flat in frequency and the struck atom’s velocity is not conditioned on the incoming photon frequency.

2.4.2 Monte Carlo simulation of line transport with Thomson scattering

Here we extend the Monte Carlo method developed in Sec. 2.3.7 by including Thomson scattering in addition to atomic processes. Kinematically, the frequency shift in a photon scattering from an electron in the non-relativistic limit is (see Appendix A.5)

$$\Delta \nu = (f_\parallel + \alpha)(1 - \cos \chi) - f_\perp \sin \chi,$$

where $f = \nu_0 V/c$ is thermally distributed ($\parallel$ and $\perp$ denote the components parallel to the direction of propagation of the incident photon, and perpendicular to it but in the plane of scattering, respectively), for the photon frequency $\nu_0$, $\alpha = h \nu_0^2/(m_e c^2)$, and $\chi$ is the scattering angle between the direction of propagation of the incoming/outgoing photons. The recoil term $\alpha$ is from differences in the electron kinetic energy and can be dropped. This is nearly identical to the case of coherent line scattering, except that the scattering cross section is flat in frequency and the struck atom’s velocity is not conditioned on the incoming photon frequency.

Thomson scattering can be trivially added to the Monte Carlo by adding an additional scattering process with differential depth $\eta_e$, from Eq. (2.61). As before, we simulate the escape probability over a grid in $\{x_{\text{HeI}}, z\}$ with continuous opacity derived from H I populations in Saha equilibrium (accurate until the end of He I recombination, where it ceases to matter anyway, once He I has almost fully recombined). The grid is taken over 11 linearly-spaced points in redshift spanning

---

36 Here the units are inverse frequency and have the interpretation that $\tau_e = \Delta \nu \eta_e$ is the optical depth for Thomson scattering after redshifting some $\Delta \nu$. This is a depth to scattering, not loss as is the case for $\eta_e$.

37 This method is computationally expensive. A more analytic alternative is the Fokker-Planck method, but this works best in the regime of frequent small $\Delta \nu$ scattering, rather than infrequent large $\Delta \nu$ scattering as is the case for Thomson scattering.

38 It is of order 200 GHz, small compared to a frequency shift of several THz typical in the exchange due to the Doppler shift. Here, we show that the effect of Thomson scattering is negligible, but if more accuracy is needed (possibly in the case of hydrogen), the recoil term should be included.
2.4 Additional effects

Figure 2.15: The electron scattering kernel at $z = 2500$ spans several THz, allowing photons in the far red side of the line to be scattered to frequencies significantly above the line center frequency. Rather than escaping on the red side of the line, as in ordinary Sobolev escape, the photon can then be absorbed (and likely be re-emitted) by incoherent processes in the line before it can escape, thus reducing its escape probability. Because the radiation phase space density is higher on the red side of the line, a photon is more likely to scatter from below to above the line frequency, thus decreasing the overall escape probability. Once continuum opacity becomes significant in the allowed lines, this can also remove trapped photons.

$z = 1400$ to $z = 3000$, and 21-logarithmically spaced points in $x_{\text{HeI}}$ from $2 \times 10^{-5}$ to 0.08. This is log-interpolated along $x_{\text{HeI}}$ and $P_{\text{esc}}$, and linearly interpolated along $z$ in the level code.

The results of the Monte Carlo calculation of the escape probability for a typical recombination history are shown in Fig. 2.16 (left) for $2^3 P_{o} - 1^1 S$ and in Fig. 2.16 (right) for $2^1 P_{o} - 1^1 S$. Fig. 2.17 shows the implied change in the recombination history. (Fig. 2.16 (left) also has results from an analytic method developed in Sec. A.3 for the escape probabilities in $2^3 P_{o} - 1^1 S$ for $z > 2000$. Fig. 2.16 (right) includes Monte Carlo results for $^3\text{He}$ scattering, developed in Switzer and Hirata (2008b), but these are negligible.) Fig. 2.16 indicates that Thomson scattering leads to a reduced escape probability for $z > 2100$ in $2^3 P_{o} - 1^1 S$ and $2^1 P_{o} - 1^1 S$, as photons that redshift out of the line are Thomson scattered back on onto it, where they are less likely to escape. Once H I opacity becomes important, fewer photons are able to travel far enough to be Thomson scattered, and are, instead, more likely to photoionize an H I atom.

2.4.3 Peculiar velocities

Before matter and radiation decouple, baryonic velocities are $\sim 10^{-5} c$, which are on the same scale as the thermal velocities, $\sim (k_B T_m/m_{\text{He}})^{1/2} \sim 10^{-5} c$. A common approximation in astrophysics is the “microturbulence” limit where the photon mean free path is large compared to the coherence length for velocities in the gas. Then each photon scattering has Doppler shifts from the thermal motion plus an RMS from the bulk flow, effectively increasing the Doppler broadening of the line profile. Because the line profile is important once continuous opacity is included, an interesting question is whether there can be turbulent line broadening during recombination. Here we show that the appropriate limit in cosmology is “macroturbulence”, where the bulk velocity is coherent (barring exotic physics Wandelt (1998); Shaviv (1998)) on transport scales. The relevant
Figure 2.16: Left: The impact of Thomson scattering on escape probabilities in $2^3 P^o - 1^1 S$ and $2^1 P^o - 1^1 S$ during recombination. A comparison of analytic and Monte Carlo treatments of transport in $2^3 P^o - 1^1 S$ which give a modified escape probability due to electron scattering. The analytic description of electron scattering in narrow lines with complete redistribution of Eq. (A.35) of Appendix A.3 agrees well with Monte Carlo results at early times. Once continuous opacity becomes important within the line, the Monte Carlo gives a dramatic increase in the escape probability. Built into the analytic method, however is the assumption that the line is indefinitely narrow. Thus, the analytic method breaks down $z < 2100$ and approaches the Sobolev theory. (The history without electron scattering is also consistent with the analytic method for complete redistribution and continuous opacity.) Right: The fractional modification to the $2^1 P^o - 1^1 S$ escape probability due to electron scattering and $^3$He scattering. The effect of electron scattering can be split into two regimes: 1) for $z > 2300$ where most of the effect from electron scattering is due to modification to transport far in the wings and the escape probability is decreased, 2) for $z < 2300$, where much of radiation in the red wing is absorbed through H I photoionization, and electron scattering begins to eject more photons from within the line, on average. Thus, for $1600 < z < 2300$ electron scattering slightly increases the escape probability, shown in the right panel as an increase of a few percent above zero for $z \sim 2200$. $^3$He does not produce any effect at the level of precision of the Monte Carlo run here. These required 3 days across $50 \times 3$ GHz nodes, so improved statistics would require significant computing time.
2.4 Additional effects

Figure 2.17: Comparing the effect of Thomson scattering, $^3$He scattering, and feedback in the forbidden and allowed lines. The uppermost curve is the difference between two models with and without Thomson scattering in the $n^3P^o - 1^1S$ and $n^1D - 1^1S$ where neither has feedback. This retards recombination because, on average, more photons are injected into the optically thick region of the line. (Note that once feedback of the radiative distortion is added, the effect of Thomson scattering is greatly reduced.) Thomson scattering in $2^1P^o - 1^1S$ decreases the escape rate at early times, but once $\text{H} \, ^1$ opacity becomes significant, more photons are on average removed from the optically thick part of the line. $^3$He can be seen to accelerate recombination slightly by assisting photons out of optically thick regions at late times. We note, though, that the typical error induced by resampling the Monte Carlo is of order $10^{-4}$ (see Fig. A.4). To confirm the effect of $^3$He would take a significantly finer grid of probabilities, but the actual value is immaterial to $\text{He} \, ^1$ recombination overall if it is this small.
Recombination

quantity in the macroturbulence limit is simply the velocity gradient over radiative transport length scales.

The two most important transport scales in recombination are the distance traversed by a photon as it escapes \( 2^1 P^o - 1^1 S \) and \( n^3 P^o - 1^1 S \). These are distinctly different processes because \( 2^1 P^o - 1^1 S \) involves many scattering events, so is diffusive, while \( n^3 P^o - 1^1 S \) typically has \( \tau_S \sim 2 \), dominated by the Doppler core of the line. In Switzer and Hirata (2008b) we find that the diffusion length in \( 2^1 P^o - 1^1 S \) is

\[
L_{\text{diff}} = \frac{1}{2\sqrt{6} \pi^2} \frac{c\Gamma_{\text{line}}\tau_S^3}{aH\nu_{\text{line}}} \sim 3 \text{ kpc}
\]

(2.63)

while the transport scale in \( n^3 P^o - 1^1 S \) is approximately the length traversed in redshifting across the Doppler core, or

\[
L = \frac{c\Delta\nu_D}{aH\nu_{\text{line}}} \sim 2 \text{ kpc},
\]

(2.64)

Both quantities are evaluated at \( z = 2000 \), and are mild functions of redshift. Note that \( L \) and traversal time are not simply related in \( 2^1 P^o - 1^1 S \) because the propagation is not straight, while for \( n^3 P^o - 1^1 S \) \( L \) and the time for traversal are related by the speed of light. These are also much smaller than the Silk length of \( \sim 3 \) Mpc at the time, so Silk damping has strongly suppressed velocity perturbations.\(^{40}\) Fundamentally the reason for this is that atomic scattering (which is responsible for transport) has a much higher optical depth through its lines than the Thomson scattering that supports acoustic modes.\(^{41}\)

The velocity structure function \( S_2(r) \equiv \langle [v(0) - v(r)]^2 \rangle \) can be calculated from the velocity power spectrum produced by the COSMICS (Ma and Bertschinger (1995)) Boltzmann code (applying the same cosmological parameters we use throughout). On scales larger than a few Mpc (above the damping length), the velocities are \( \sim 18 \text{ km s}^{-1} \), while the thermal velocities are \( (k_B T_{m}/m_{\text{He}})^{1/2} = 3.4 \text{ km s}^{-1} \). Yet, evaluating the structure function on scales relevant for radiative transport produces \( S_2^{1/2}(L_{\text{transport}}) < 10 \text{ m s}^{-1} \), which is considerably smaller than the thermal velocity! In this limit, the velocity field simply has a gradient (rather than the effective RMS taken in a microturbulence limit), so effectively changes the Hubble rate in that patch. A distribution of velocity gradients is insufficient to generate an overall difference in recombination because it is zero on average (though produces rms fluctuation from the Hubble rate at \( \text{few} \times 10^{-4} \))—so for each small patch with an effectively smaller Hubble rate, there is a patch with a larger rate. Higher baryon moments may produce a net effect, but are expected to be negligible.

2.5 Summary of results for helium, discussion

The largest new effect presented in this chapter has been the introduction of continuum opacity due to H I photoionization in the He I \( 2^1 P^o - 1^1 S \) transition. We presented two methods of understanding this physics, an integral solution to the radiative transport equations for complete redistribution, and a Monte Carlo method for partial redistribution (to include strong coherent scattering effects). Opacity from H I photoionization dramatically increases the escape probability and accelerates He I recombination once continuum effects become important over transport scales within He I resonances. These effects, in combination with the inclusion of the intercombination

\(^{39}\) C. Hirata was responsible for the diffusion length calculation and E. Switzer was responsible for the fluid length discussion in Appendix A.4.

\(^{40}\) Also note that the horizon size at \( z = 2000 \) is \( \sim 180 \text{ Mpc} \) and so both the diffusion and transport scales are well within the causal horizon.

\(^{41}\) Because the atomic lines have such small width, they themselves do not impact the velocity structure of the gas (see Hernández-Monteagudo et al. (2007)).
2.5 Summary of results for helium, discussion

line as suggested by Dubrovich and Grachev (2005), provide a new picture of He I recombination where the recombination accelerates after $z \sim 2200$ leading to very little overlap with the H I recombination.

Fig. 2.18 shows the cumulative difference in the CMB temperature and polarization anisotropies due to the largest corrections described. We want to emphasize only the modification to the He I recombination history from the effects described in this chapter, so we only compare with a reference model where those effects are absent. The changes shown in Fig. 2.18 would affect a cosmic variance limited experiment at the $\sim 1\sigma$ level at $\ell_{\text{max}} = 1500$, and $\sim 8\sigma$ at $\ell_{\text{max}} = 3000$, but are (thankfully) negligible for WMAP. This means that the effect is in principle at the $\sim 1\sigma$ level for the upcoming Planck satellite Lewis et al. (2006); Wong et al. (2007) or for a high-resolution CMB experiment mapping 1–2% of the sky to $\ell_{\text{max}} = 3000$, although it is possible that systematic uncertainties such as beam modeling and point source removal may prevent one from reaching this accuracy bound. Of the corrections considered, we propose that feedback of non-thermal distortions between the allowed lines, continuum opacity from H I photoionization, and the inclusion of the $^2P_{\Omega} - ^1S$ Dubrovich and Grachev (2005) rate become part of the standard recombination model. A summary of the final He I recombination history developed here can be found in Fig. 2.19 (left). We have not described He II recombination in detail because it is almost a pure Saha history, as shown in Fig. 2.19 (right) and is insignificant for the CMB anisotropy.

The ultimate goal of the recombination calculation is to attain accurate $C_\ell$ power spectra on small scales. To assess the sensitivity of the $C_\ell$'s to the recombination history, we can take the functional derivative with respect to $n_e(z)$. The is evaluated in Switzer and Hirata (2008b) by adding a Gaussian “bump” to the recombination history at a given redshift. This shows, predictably, that the $C_\ell$'s depend most on perturbations to recombination at later times, and the difference is most significant in the damping tail.

We believe the remaining errors on the electron abundance from helium recombination are $|\Delta x_e|_{\text{max}} < 0.003$ in the standard recombination picture described here. (In actuality, the error bound is probably lower because at the time we were dominated by uncertainty in the intercombination rate, which has subsequently been resolved.) The functional derivative gives a 0.2% uncertainty at $C_\ell=3500$. This represents the state of the art for helium recombination, and is sufficient because it is near cosmic variance.

While He I recombination is important cosmologically, it should only be considered as a first step toward a much broader treatment of cosmological recombination as a whole. H I recombination is especially important for the CMB since it affects the free electron abundance near the peak of the CMB visibility function and the evolution of $n_e(z)$ is dominated by hydrogen rather than helium evolution. (Recombination also affects the temperature evolution, 21 cm absorption, and chemistry during the “Dark Ages” between recombination and reionization.) The calculation of H I recombination and its effects on the $C_\ell$'s to sub-percent accuracy are far more difficult than the calculations for He I presented here and it has become an active area of research.

---

42 While we have made every effort to include all effects, we acknowledge that there could be additional processes that escaped our imagination.

43 In H I, radiative transport is much more complex because of the high optical depths, large optically thick line widths, overlapping lines and two-photon processes. Also, the high $n$ states become important at late times, as well as $l$-dependent occupations and collisional processes that mix them.

44 See Dubrovich (2007); Chluba and Sunyaev (2007a,b, 2006); Kholupenko and Ivanchik (2006); Chluba et al. (2007); Hirata (2008) for recent work.
Figure 2.18: The cumulative effect on temperature and polarization anisotropies from H I continuous opacity and feedback during He I recombination developed in this work, as calculated by CMBFAST. We note that here we only consider the difference between the reference helium model with and without these effects. Comparison of the full H I and He I history to standard methods such as RECFAST will be the subject of later work.
Figure 2.19: Left: The He \textsc{i} recombination history from our multi-level atom code (solid line), compared to the Saha equation (long-dashed) and the commonly used three-level code \textsc{Recfast} by Seager et al. (1999) (short-dashed). The y-axis gives the free electron fraction relative to $n_H$ and the $x$-axis is the redshift. Both our analysis and that of Seager et al. find that He \textsc{i} recombination is delayed due to the $n = 2$ bottleneck. However we find a slightly faster recombination than Seager et al. due primarily to our inclusion of the intercombination line $2^3P^0 - 1^1S$ and the accelerating effect of H \textsc{i} opacity. The latter effect causes our He \textsc{i} recombination to finish at $z \approx 1700$, whereas in \textsc{Recfast} one-third of the helium is still ionized at that time. Right: The He \textsc{ii} recombination history from the level code developed here. He \textsc{ii} recombination is essentially irrelevant for CMB physics, and because of its rates it varies from Saha evolution at the level of $< 0.2\%$. (for example, at $\ell < 3000$ the absolute difference in $C^\text{TT}_\ell$ for the full model relative to Saha is $< 3 \times 10^{-5}$.)
Chapter 3

ACT Systems

3.1 Telescope systems overview

ACT hardware is broadly composed of the telescope, its camera and electronic systems, site infrastructure, a ground station, and a data analysis center. In this chapter\(^1\) we describe these systems and their coordination for observations, and subsequent chapters will describe the optical response of the telescope and its detectors.\(^2\) This section introduces the system layout and coordination; Sec. 3.2 describes the telescope systems for science data acquisition, housekeeping, and telescope motion; Sec. 3.3 describes data products and data component merging; Sec. 3.4 describes coordination of the various systems; Sec. 3.5 describes logistics of operating the telescope in Chile.

The telescope in an off-axis Gregorian design with warm (ambient site temperature) 6 m primary and 2 m secondary mirrors, a large (1\(°\)) focal plane (see Fowler et al. (2007)) and clearance for a \(\sim 1\) m\(^3\) cryogenic camera (see Swetz et al. (2008)). To minimize beam deformation and spill during the scan, the telescope has a rigid all-aluminum structure and both an inner (co-rotating) and an outer ground (fixed) screen.

At the ACT site for 2007, there were three permanent shipping containers that served as a workshop, an “equipment room” (a climate-controlled office plus an area for electronics and cryogenic compressors), and storage space. Fig. 3.1 gives an overview of the site in 2007. The telescope has a large receiver cabin under the secondary and primary mirrors which houses MBAC and the housekeeping acquisition electronics. Fig. 3.2 summarizes the orientation of the optics and telescope structure, and Fig. 3.3 shows the MBAC camera and a diagram of the cold optics. MBAC is mounted close to the telescope’s rotation axis (to avoid jarring) and the cabin’s temperature is controlled to improve the stability of the readout systems. It is also large enough for several operators to stand in while installing the camera and electronics, and it provides access to the mirrors for alignment purposes. We will refer to those components located on Cerro Toco as being at the “site,” and those located in San Pedro de Atacama as being in the “ground station.”

Fig. 3.5 gives a schematic of the acquisition system. At the highest level, operations are coor-

\(^1\)This is adapted from a proceedings article for the SPIE, Switzer et al. (2008).

\(^2\)For detailed reviews of the hardware, see Battistelli et al. (2008); Niemack et al. (2008); Fowler et al. (2007); Lau et al. (2006b,a); Lau (2007); Marriage (2006); Niemack (2008) and articles from the 2008 SPIE Astronomical Instrumentation proceedings for further background, Hincks et al. (2008); Swetz et al. (2008); Battistelli et al. (2008); Zhao et al. (2008); Thornton et al. (2008).

\(^3\)The science camera hardware, detectors and electronics were built by a collaboration of teams at Cardiff University, NASA, NIST, Princeton, University of British Columbia, University of Pennsylvania, and the University of Toronto. The telescope was engineered and manufactured by Empire Dynamic Systems Ltd. (formerly AMEC Dynamic Structures Ltd.), and KUKA Robotics (see Hincks et al. (2008)) designed the control systems.
3.1 Telescope systems overview

Figure 3.1: Overview of the ACT structure and site. Left: the ACT site and its components. The workshop container has a hoist shed so that the camera can be assembled and disassembled at the site. The facility is powered by one of two on-site diesel generators, and has on-site diesel storage capacity of $1.5 \times 10^4$ liters (the black tank). It is located at $64^\circ 47' 15''$ W by $22^\circ 57' 31''$ S at 5190 m on Cerro Toco in the Andes east of the Atacama desert of Chile and is a ~ 1 hour commute from the small town of San Pedro de Atacama at 2400 m. The moving mass of the telescope is 40 t (the total mass is 52 t) and extends 12 m from the ground. (The outer ground screen is 13 m and reflects far sidelobes to the sky.) The range of motion is through $\pm 220^\circ$ azimuth and $30.5^\circ - 60^\circ$ elevation. It has a maximum azimuth scan speed of $2^\circ$/s, and a maximum acceleration at turnarounds of $2^\circ$/s$^2$. The elevation is fixed to whatever extent possible and actively controlled to reduce systematic effects, but can re-point to target planets or smaller fields at $0.2^\circ$/s. In practice, we scanned in azimuth at $1^\circ$ per second at $50.5^\circ$ elevation (or $\sim 0.8^\circ$/s on the sky) in the 2007 season. The smallest diameter of the primary mirror (composed of 71 panels in 8 vertical rows) is 6 m. The secondary mirror is at ambient temperature, is composed of 11 panels, and has a minimum diameter of 2 m. The primary mirror panels are each adjustable through four manual screw adjusters. The secondary mirror panels were aligned at the AMEC facility and are fine-tuned in the field. The overall position can be readjusted by two frame and three mirror linear actuators. (Photo: A. Hincks; drawings, AMEC/Empire Dynamic Systems)
Figure 3.2: The ACT receiver cabin at 60° elevation (normal observation in the 2007 season was at 50.5° elevation). Subsequent plots that show array response orient the detectors as they appear on the sky, where silicon card “columns” are horizontal on the sky and row zero column zero is in the upper right of the projected array on the sky. (Figure: D. Swetz, B. Thornton)
3.1 Telescope systems overview

Figure 3.3: Left: the millimetric bolometer array camera (MBAC) in its test stand and the University of Pennsylvania. The dewar is \( \sim m^3 \) and comprises three arrays and their optics. The square 300 K windows have been covered here, but are 4 mm thick ultra high molecular weight polyethylene (UHMWPE) with a Teflon AR coating. Right: The cold optics in the MBAC dewar. The optical path in the three sub-cameras is similar and consists of three AR-coated silicon lenses (1/4-wave Cirlex on float zone silicon, see Lau et al. (2006b); Lau (2007); Fowler et al. (2007)) and a filter stack, moving from IR blocking filters at 300 K to a low-pass stack at 40 K to a 1 K low-pass and a band-definition filter and final lens at 0.3 K. Thus everything behind the bandpass filter is in a 0.3 K cavity. All interior walls are blackened with carbon lampblack and Stycast 2850 FT epoxy.

Synchronized by a scheduler and the data are synchronized by a system-wide serial number “heartbeat.” This ties the encoders, cryogenic housekeeping, and science data to 5 \( \mu s \) precision (after all constant delays have been accounted for). Fig. 3.4 shows the housekeeping rack (left) in the receiver cabin and a view down to the camera from the secondary mirror (right). Figs. 3.6, 3.7, and 3.8 give an overview of the MBAC cryogenics, heaters, cryogenic cycle history, respectively. The synchronization generator and software as well as absolute timing are described in Sec. 3.2.1. All operations of the telescope are coordinated through a common interface server, described in Sec. 3.4. This presents a common interface to an observation scheduler and to telescope operators, and passes messages to the telescope systems.

The housekeeping and science camera acquisition systems (see Sec. 3.2.2 and Sec. 3.2.3) are handled by four separate computers (see Sec. 3.2.7) which write the incoming data to their respective hard drives. These data are then served over the network to real-time data component merger and visualization clients. During normal telescope operation, the housekeeping and science data from the three cameras are merged (Sec. 3.3) into a science data product. These data, as well as the raw data, are transmitted from the site to the ground station over a line-of-sight microwave link and buffered on a RAID storage array as they are copied to external hard drives which are flown back to North America. Both the housekeeping and camera data have associated entries in a file database which is periodically checked to automatically move data from the site computers to the ground station, onto transport disks, and to confirm that the data have been properly received in North America before deleting the data from site computers. The database also has an associated webpage where operators can track the volume and type of data acquired in real time.

The observation schedule is generated on a night-by-night basis and coordinates all operations of ACT: the telescope mechanical warm-up, camera tuning and biasing, acquisition, calibration/beam measurements during the night, and cryogenic cycle timing. We thus have an auto-
3.2 Telescope systems

3.2.1 Time and synchronization

ACT’s timing requirements are twofold. Because the telescope moves during data acquisition, the camera data must be synchronized with the encoders to link the horizon coordinate pointing of the telescope with a given data sample. The ACT data must also be tied to an absolute time reference to find the celestial pointing. The beam-crossing time of a point source through scanning (∼ 10 ms) is much faster than the beam-crossing time of a point source through the Earth’s rotation (several seconds). Thus, the synchronization of the encoders to camera data is more strict than of these data to the absolute (GPS) time.

mated system to process high-level observation requests, execute the measurement, and move data onto transport disks.
3.2 Telescope systems

Figure 3.5: Overview of the ACT data and control systems, split into telescope (top), on-site (middle), ground station, and North American systems (bottom). Solid lines show data streams while dashed lines show commanding and file information channels. Cylinders represent data storage. We show only one of three detector array acquisition systems ("Array DAS") for simplicity. The ground station RAID array aggregates data from several machines at the site, so we do not show it as being connected to a particular node. To prepare a transport disk with data to ship back to North America, an operator connects the drive to the RAID array, and the data are automatically copied over based on information in the file database. Operators define a schedule each night and upload it to the schedule dispatcher at the site. Here we have only shown a monitor client in the ground station, but the site also has a system monitor terminal.
Figure 3.6: Block diagram for the 0.3 K cryogenics. The pulse tube cryo compressor cooler cools the condensation plate of a $^4$He adsorption refrigerator. Once the $^4$He pot has accumulated sufficient liquid, the $^4$He charcoal pump temperature (which was actively controlled at $\sim 45$ K) is rapidly cooled by conducting the heat out through a gas gap heat switch. This pumps the $^4$He liquid, and cools an analogous condensation plate for $^3$He. Once enough liquid $^3$He has accumulated, the charcoal pump (which was actively controlled at $\sim 35$ K) cools by conducting through a second gas gap switch. The charcoal then adsorbs the $^3$He, and the $^3$He pot reaches the desired bath temperature of $\sim 0.3$ K. A typical hold time during operation was 15 hours, and the adsorption refrigerators are cycled once per day, see Fig. 3.8. This closed-cycle system allows ACT to operate in a remote location without additional liquid cryogens. The automated cycle is described in Sec. 3.2.5.

A synchronization box in the telescope’s receiver cabin generates a system-wide trigger and an associated serial stamp. This gives a synchronization reference for the science camera, housekeeping system, and telescope motion encoders. The stamp and trigger are passed to each science camera’s acquisition electronics through a fiber optic, and to the housekeeping computer (in the equipment room) through an RS485 channel in the cable wrap. The synchronization box was designed at the University of British Columbia to be compatible with the camera acquisition electronics, and is used by several experiments. The synchronization box sets the fundamental rate for triggering observations by counting down a 25 MHz clock over 50 cycles per detector row, over 32 rows of detectors plus one row of dark SQUIDs for the array readout, over 38 array reads per sample trigger. This gives a final rate of $\sim 399$ Hz.

The synchronization serial stamp from the RS485 channel is incorporated into the housekeeping data (including the encoders) through the following chain: 1) in the housekeeping computer, the primary housekeeping PCI card receives a 5 MHz biphase signal which encodes serial data stamps at 399 Hz over RS485 from the synchronization box; 2) these trigger CPU interrupts at 399 Hz; 3) a timing driver handles these interrupts by polling the encoders at 399 Hz; 4) the housekeeping PCI card clocks down the serial stamps to 99.7 Hz, which it uses to poll the housekeeping acquisition electronics; 5) the housekeeping software then assembles the housekeeping and encoder time streams, matching serial stamps; 6) these are written to disk in a flat file format, where for every data frame there are, for example, 4 times as many 399 Hz encoder values as there are 99.7 Hz housekeeping data frames. Thus, the encoders and housekeeping channels are

---

4 See Battistelli et al. (2008); Niemack et al. (2008); Battistelli et al. (2008) and Battistelli et al. (2008).
5 This is related to the time domain multiplexer timing in the camera acquisition electronics, which clocks 50 MHz down to $\sim 399$ Hz instead over 100 clock cycles per detector. Timing within the camera acquisition electronics is described in Sec. 3.2.2.
6 From the Balloon-borne Large Aperture Submillimeter Telescope, BLAST. See Pascale et al. (2007).
7 The final merged file has the so-called “dirfile” format, and only the intermediate (unmerged) housekeeping and science cameras use a native flat format. Sec. 3.3.1 describes this in more detail.
3.2 Telescope systems

Figure 3.7: Outline of the MBAC heater controller systems. Electronics designed for BLAST at the University of Toronto (UofT) manage the housekeeping input/output from a crate in the receiver cabin. These electronics communicate with the housekeeping computer in the equipment room using the BLAST-bus PCI card (BBCPCI), connected through a serial link from the telescope’s receiver cabin. The heater driver card was designed by T. Devlin’s group at Rutgers; Princeton was responsible for the interface between components and the current monitor/relay box (CRBox). One DB37 connector from the UofT diode readout card programs the voltage levels on the heater driver card (located in the analog readout electronics “Penn” crate), and the DB37 connector on the UofT current monitor readout card toggles relays in the CRBox. (The readout cards in the UofT crate are generic, so the heater driver simply uses the digital IO on the card whose analog inputs are used to read out the diode thermometry.) Significant effort has gone into making the housekeeping system and interconnects RF-tight.
Figure 3.8: Temperature of key stages during the cryogenic cycle. Vertical lines indicate when the gas-gap heat switches activate. There are two $^4$He adsorption refrigerator units: the “optics” unit is responsible for the optics tubes, while the “backing” unit provides the cold stage for $^3$He condensation and is activated second here. The topmost plot shows the pump temperatures; the middle plot shows the condensation plate temperatures; the lowest plot shows the liquid pot temperatures, which first reach 0.9 K, after which the $^3$He pot drops to 0.3 K.
3.2 Telescope systems

combined into a product that can be merged with the science camera data (see Sec. 3.3.1).  

We use a Meinberg GPS-169 PCI card to discipline the system clock of the housekeeping computer, with a precision of < 1 ms to GPS time, sufficient for absolute timing. The absolute GPS time is attached to the data from each encoder read request, and the GPS position is averaged down to give a geodetic latitude and longitude for astrometry. The housekeeping computer provides a stratum 0 server for the other computers (data acquisition, merging) on the network to synchronize file time stamps that are used to organize and track files.

3.2.2 Camera acquisition and commanding

The ACT camera, MBAC, comprises three bands (145 GHz, 220 GHz, and 280 GHz), each with a $32 \times 32$ free-space coupled square array of TES detectors. Each array is read out by a time-domain SQUID multiplexing circuit. SQUID multiplexers have become the standard for low-impedance, low noise cryogenic detectors (Chervenak et al. (1999)), where wire count and loading on the 0.3 K stage is at a premium. The multiplexer is controlled by a set of Multi-Channel Electronics (MCE) which seek to null the incoming signal (error) from the final stage of the SQUID amplifier chain by applying a feedback current to an inductive coil at the input of the SQUID chain, where the TES circuit is also inductively coupled.

The MCE is designed and produced by the University of British Columbia group, and the cryogenic SQUID circuits are designed and produced by the NIST (Boulder) group. Each of the three bands has an independent MCE which is mounted directly onto the MBAC cryostat and accesses the cold multiplexers through $5 \times 100$ pin MDM connector ports (per MCE). The MCEs, their power supplies, and MBAC all reside in the telescope receiver cabin. The three MCEs are connected to three storage and control computers in the equipment room through 6 (3 TX/RX pairs, 250 Mbps) Tyco PRO BEAM Jr. rugged multimode fiber optic carriers. The signals from each array’s MCE are decoded by a PCI card from Astronomical Research Cameras, Inc. in each of the three acquisition computers.

The base clock rate of the MCE is 50 MHz. This is divided down to 100 clock cycles per detector row by 32 detector rows plus one row of dark SQUIDs. Thus, the native read-rate of the array is 15.15 kHz as it leaves the cryostat. Nyquist inductors at 0.3 K in the $L R$-circuit of the detector loop band-limit the response so the array can be multiplexed with minimal aliased noise while maintaining stability of the loop (see Niemack et al. (2008); Niemack (2008)). For science data, the only bandwidth constraint is to at least Nyquist sample the optical beam, as described in the introduction. Even though the Nyquist sky sampling condition is different for the three arrays, we sample each at 399 Hz. This is sufficient for the 280 GHz array, which has the most stringent requirements. To downsample the 15.15 kHz multiplexing rate to 399 Hz, the MCE applies a 4-pole Butterworth filter with a rolloff $f_{3dB} = 122$ Hz to the feedback stream from each detector. This filter is efficient to implement digitally and has a flat passband. The rolloff then defines our conservative downsampling to 399 Hz, which can be obtained by pulling every 38th sample (at 15.15 kHz) from the filter stack, as synchronized by similar clock counting in the synchronization box, described in Sec. 3.2.1.

Each time an MCE receives a synchronization serial stamp from the synchronization box, it packages the output of the $32 \times 32 + 1 \times 32$ (32 dark SQUIDs) array and sends it over a fiber optic to a PCI card on its acquisition computer, where it is buffered and written to disk. Additional

---

8The synchronization chain was developed by J. Fowler and A. Hincks.
9http://www.astro-cam.com/
10The interaction of the flux lock loop response, the detector circuit bandwidth, and the antialiasing filter is described in more detail in Sec. 5.4.
11Sec. 5.4.2 describes the transfer function of this filter.
information that fully specifies the MCE state in each 15-minute acquisition interval is written to a text "run file."

Each of the three data acquisition computers in the equipment room is responsible for coordinating the operations of its MCE. SQUID tuning (see Battistelli et al. (2008)), detector biasing, and data acquisition are managed by a set of codes in IDL, python, and bash. Messages from the high-level scheduler are passed through the interface server to python clients on each of the acquisition computers that initiate the requests (see Fig. 3.5 and Sec. 3.4.1). The bash scripts for acquisition also each contain a call to a python code which passes the file information to a MySQL database for tracking, as described in Sec. 3.3.1.

For the 2007 season, ACT used a prototype version of software developed for the SCUBA2 experiment for acquiring and storing the MCE data on the acquisition computer. The SCUBA2 software included custom firmware for the PCI acquisition card, a driver compiled against a real-time kernel to provide low latency handling of MCE data, and a master control program based on the DRAMA (see Bailey et al. (1995)) messaging system.

In subsequent seasons, ACT uses the MCE Acquisition Software (MAS), developed at UBC by M. Hasselfield. MAS provides a device driver for the PCI card, a C language API for low and intermediate-level interaction with the hardware, and applications for commanding the MCE from the Linux shell. The MAS driver does not require a real-time kernel, and can provide high data throughput and low single-frame latency without requiring sustained interrupt rates above 10 Hz.

### 3.2.3 Housekeeping readout

Housekeeping comprises all electronics and systems other than the camera and its electronics and the telescope motion control. The primary systems within housekeeping are the cryogenic thermal readouts and controllers, the telescope health readouts, the telescope motion encoders, and auxiliary monitors, such as a flux-gate magnetometer. The telescope’s azimuth and elevation pointing are read by two Heidenhain encoders (27 bit, $0.0097''$ accuracy) through a Heidenhain (model IK220) PCI card in the housekeeping computer over RS485 from the cable wrap.

ACT uses both ruthenium oxide (ROx) and diode thermometers to monitor cryogenic temperatures. The readout system for the diodes measures the DC voltage drop across them. For the ROx sensors, we use an AC-bridged 4-lead measurement driven at 212 Hz for each channel. Both the diodes and ROxs use preamplifier and driver electronics in RF-tight enclosures in the receiver cabin. Data acquisition electronics designed for the BLAST (see Pascale et al. (2007)) experiment digitize the thermometry preamplifier voltages at 10 kHz and apply a 50 Hz anti-aliasing filter so that they can be downsampled to a 99.7 Hz (399/4 Hz) output. This is triggered and read out by the PCI card on the housekeeping computer in the equipment room through an RS485 cable from the telescope. The BLAST electronics comprise modules with 25 ADCs, 3 bytes of digital IO for control, and 4 pulse width modulation (PWM) drivers. These electronics synchronously demodulate the ROx channels and read the diode thermometer voltages. Lookup tables on the housekeeping computer provide conversions to temperature. Fig. 3.4 shows several of these systems in the housekeeping rack in the receiver cabin.

Housekeeping includes data from a variety of auxiliary sources such as a 3-axis magnetometer, dewar pressure monitor, and other system status monitors. A Sensoray 2600 DAQ in the receiver cabin reads auxiliary channels that do not need to be synchronized with the science camera data (such as the primary mirror panel temperatures). It sends these data through the site network at sampling rates from $1 - 20$ Hz. Weather data are available both from an on-site WeatherHawk sta-
3.2 Telescope systems

tion, and from the APEX collaboration. A CCD star camera (developed for BLAST, see Pascale et al. (2007)) can also be mounted to the top of the primary mirror structure as needed to develop a pointing solution, accurate to 5′′, but has not yet been used.

3.2.4 Housekeeping: thermal control

Thermal control in ACT must manage both the cryogenic refrigerator cycling and system stability during operation. Here we give a brief overview of the thermal control systems, and appendix sections B.1 and B.2 describe the hardware and interfaces in more detail.

Cryogenic operation has four stages: the initial cooldown, pumping ³He and ⁴He to reduce the temperatures of the detector stage and optics, observation, and recycling (re-condensing helium in the adsorption refrigerators). Operation after the initial cooldown needs to be autonomous and robust. During data acquisition, heaters maintain the bath temperature through a servo loop. The driver channels can apply power to calibration emitters, or other devices in the dewar.

Commands from the housekeeping software are passed to the BLAST DAS crate digital I/O, which programs output voltages on the heater driver circuits. The outputs of the drivers are each regulated by a relay which determines whether the voltage is applied to heater channels in the dewar. The relay toggles between either 1) routing the internal heater (or voltage) driver to the dewar, or 2) routing a BNC on the front panel of the housekeeping rack to the dewar. The front-panel BNC allows operators to use an external source or function generator for tests and calibration emitter triggering. When it is not energized, a relay prevents the driver voltage from reaching the channel in the dewar. A current monitor tracks the current that ultimately goes to the dewar from either the driver circuit or an external source. The UofT/BLAST housekeeping crate reads these current monitors and incorporates them into the housekeeping stream. Fig. 3.9 describes the current monitor/relay box (CRBox) hardware, Fig. 3.10 shows the card stack (left) and connectors on the back (right) of the CRBox.

The heaters are driven by a specially designed, 18-channel heater board plus 6 external/auxiliary channels, instrumented with current monitors and relays for driver management. The 18 heater drivers are programmable through software, and the 6 additional channels are user-configurable, and can be used for cryogenic tests, calibration emitter drivers, or in case of failure of one of the primary heaters. The 3 digital bytes of digital control on the UofT/BLAST housekeeping acquisition electronics program the driver voltages.

Each stage of temperature control in the camera is managed by a servo that uses a heater and either the diode or ROx on that stage. The respective servos are weakly linked. Each servo then uses a PID loop with finite, decaying integral memory. For the first season, this controlled the 0.3 K stage temperature to a typical stability of < 40 µK. The 0.3 K stage holds the detector array, the bandpass filter, the lens nearest to the array, and the cold enclosure surrounding the array.

---

14http://www.apex-telescope.org/weather/index.html
15The initial cooldown uses a mechanical heat switch to conduct heat out of the cold stages that must be manually switched off once the refrigerator is ready to cycle. This is only performed once, and the system is autonomous after that point. Cryogenic cycles are always initiated by an external request from either an operator or the schedule. Even during autonomous operations, the interface server and the GUI client provide a mechanism for manually controlling any of the cryogenic heaters and setpoints. This is described in Sec. 3.4.1.
16In the 2007 season, this included only the 300 mK detector mount “slab,” while in 2008, this includes the ³He adsorption unit pot (for slow drifts) and the 300 mK arms (for faster transients from motions).
17This was developed locally by A. Hincks and E. S. (interface and cryogenic cycling), and is based on ascp from BLAST.
18The heater drivers actively control temperatures using a continuous controller, but the software can be quickly reconfigured as a discrete, thermostat controller. This can be used to switch a constant voltage from an external source using the relays in the CRBox to regulate, e.g., the charcoal pump temperatures.
19Note that the stability of the thermometer that it actively controlled is not the best measure of the stability of the stage. In Sec. 4.3.4, we argue that the main temperature drift that is important in analysis is the 3 K stage, not the 300 mK stage.
20For subsequent seasons, servo channels are available to control the 4 K SQUID amplifier chain board to limit some
Figure 3.9: MBAC current monitor and relay box (CRBox). The front panel has a DB50 connector for current monitor outputs, a DB37 connector for digital control inputs, and 24 isolated BNCs. Of the 24 BNCs, there are 18 for heater channels on the heater driver card, and 6 auxiliary channels are for additional lines in the dewar. All auxiliary lines are treated differentially, and the 18 heater lines share a common ground which is independent from the ground in the CRBox. The CRBox channels are labeled from right to left, bottom to top, where e.g., the bottom right is channel 1.

Figure 3.10: Left: cards in the CRBox. The RF block and regulators from 16.5 V to ±12 V and 5 V are shown in the bottom of the figure. The back panel (the right side of the left figure) has two DB50 ribbons that lead to the dewar and one DB50 from the heater driver card in the housekeeping preamplifier VME crate. The card stack comprises (from top to bottom) a switchboard card, a current monitor card and a relay card. The switchboard is rotated at a right angle relative to the others to permit a clear connection to the two DB50 connectors to the dewar on the back. The BNCs on the front panel (the left side of the left figure) are connected through a DB50 junction to a ribbon to the IDC50 connection on the relay card, and the LEDs are connected through a straight IDC-50 ribbon and junction. The switchboard “top card” permutes heaters in the dewar (which are broken out from the two DB50 connectors on the back) to combinations of the 24 possible heater channels. Right: back of the CRBox, showing the power 4-pin military connector, DB50 connector for input from the heater driver, and two DB50 connectors for outputs destined for the cold break-out to heaters in MBAC.
3.2 Telescope systems

Table 3.1: Cryogenic cycle parameters for the 2007 season.

<table>
<thead>
<tr>
<th>Threshold/Timer</th>
<th>2007 season value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to let the heat switches cool</td>
<td>600 sec</td>
</tr>
<tr>
<td>Time to leave pumps warm before moving to cooling stage</td>
<td>4000 sec</td>
</tr>
<tr>
<td>Minimum time to wait with the 4 He-1 heat switch on (pumping)</td>
<td>6000 sec</td>
</tr>
<tr>
<td>Minimum time to wait with the 4 He-2 heat switch on (pumping)</td>
<td>6000 sec</td>
</tr>
<tr>
<td>Minimum time to wait with the 3 He-2 heat switch on (pumping)</td>
<td>3600 sec</td>
</tr>
<tr>
<td>Threshold for 4He-1 condensation</td>
<td>4.4 K</td>
</tr>
<tr>
<td>Threshold for 4He-2 condensation</td>
<td>4.2 K</td>
</tr>
<tr>
<td>Threshold for 3He-2 condensation</td>
<td>1.0 K</td>
</tr>
</tbody>
</table>

found that lens temperatures were stable enough that they need not be actively controlled. The remaining servos are coarse and used for cycling the adsorption refrigerators.

3.2.5 MBAC cycling

MBAC has three closed-cycle adsorption refrigerators, described in Fig. 3.6. This comprises two $^4$He refrigerators, where one is for the optics and the other (the “backing” unit) is for cooling a $^3$He refrigerator. The $^3$He refrigerator cools the final stage of the optics and the detectors to 300 mK. These are automatically recycled daily when the sun is up. The recycle proceeds through the chain: 1) heat charcoal adsorption pumps, 2) cool $^4$He-optics stage, 3) cool $^4$He-backing stage, 4) cool $^3$He stage, and move to 5) observation mode (where all heat switches and temperature servos are on).

In the preparation stage, everything is turned off first for 10 minutes. This allows the heat switches (which were hot and conducting from the previous run) to cool. After waiting, the controller turns on the three adsorption pump heater servos and waits a user-programmable amount of time before moving on to the 4He-1 cooling stage.

In each cooling stage, the controller watches the condensation plate temperature of the given cryogenic cooling stage. When it falls below a specified temperature, (indicating that the pot contains the condensed liquid cryogen) the controller turns on the heat switch for that stage. This thermally shorts the pump to the 4 K base plate so that it can cool and pump on the condensed liquid (either 4He or 3He). Once the heat switch is on, the controller waits a minimum amount of time before moving on to the next stage. Table 3.2.5 gives the numerical parameters for the cycle.

3.2.6 Telescope motion

The telescope’s azimuth base is driven by two counter-torqued helical gears to minimize backlash, and the elevation axis (which remains fixed during observation) is driven by two ball screws at the sides of the receiver cabin. The secondary mirror can be repositioned with three mirror and two frame actuators (see Hincks et al. (2008)), read out by linear variable differential transformer (LVDT) sensors. All of the telescope degrees of freedom are driven by a KUKA Robotics controller. The most critical of these axes is the azimuth, where we require a tight tolerance of 400 ms turnarounds at either end of a 10° wide, constant-velocity scan with a maximum velocity of 2° per second (2007 observations used a velocity of 1° per second, while 2008 observations use 1.5° per second). This was achieved with a state-space controller (see Hincks et al. (2008)).

The KUKA robot comprises the motor drivers, an uninterruptible power supply, an embedded computer with a solid-state drive, and a portable “pendant” console. The robot embedded com-
puter has a set of templates in the KUKA-native robot language for executing observations, re- pointing, slewing to home/stow/maintenance positions, and warming up the mechanical systems. Motion parameters for the templates (such as the scan center and speed) are transferred on a DeviceNet\(^{21}\) network between the housekeeping computer and the KUKA robot controller. Once the program and its parameters are loaded, the housekeeping computer sets a run bit and the scan commences. The robot controller can be commanded to stop by setting a stop bit on DeviceNet, which smoothly stops the telescope.\(^{22}\) KUKA’s robot also produces a stream of telescope health information (its internal encoder and resolver readouts, motor currents and temperature) which are broadcast via the user datagram protocol\(^{23}\) and recorded by the housekeeping computer at 50 Hz. The motion control servo loop uses an identical set of 27-bit Heidenhain encoders along the same axes as the ACT housekeeping pointing encoders, but because the KUKA data stream is asynchronous to the housekeeping system, these are only used by ACT as diagnostics.

### 3.2.7 Computing

The ACT computer hardware is described in Tables 3.3 and 3.2. Telescope operation and data acquisition are shared between three science camera acquisition computers, one housekeeping computer, an on-site data merger computer, and a telescope robot (embedded computer).\(^{24}\) Of these, the housekeeping computer has the most varied responsibilities: 1) readout of the primary position encoders, 2) absolute timing discipline, 3) cryogenic housekeeping control and acquisition, 4) command routing, 5) telescope robot messaging, and 6) cryogenic housekeeping and encoder synchronization.

Storage in Chile is buffered from smaller pressurized drives on the mountain to a RAID storage array in San Pedro de Atacama.\(^{25}\) While the data rate of the final, merged product for one array is 70 GB/day, we also keep the un-merged, raw device products for assurance, yielding around 140 GB/day total.\(^{26}\) (Section 3.3 describes data merging and products in detail.) The 4 TB RAID storage unit used in the first season was able to buffer for less than one month before data needed to be shipped to North America. On-site storage was upgraded for the three-array 2008 season with an 11 TB array plus a 6.5 TB compute node.

### 3.3 Data products

#### 3.3.1 Data merging

Each on-site data acquisition computer has a server which can stream data (either in real-time, or in a seek mode) to multiple clients. These clients can be either merger processes or operators that display the data in real-time using kst.\(^{27}\) Each camera observation is keyed to an entry in the MySQL file database through a python hook in the acquisition calls. The incoming camera data are flagged as either un-merged, in the process of merging, or merged. A python front-end of

---

\(^{21}\)Open DeviceNet Vendors Association (ODVA; odva.org)

\(^{22}\)This interface was developed by A. Hincks.

\(^{23}\)The user datagram protocol (UDP) is a connectionless protocol that broadcasts data without the handshaking of the transmission control protocol (TCP).

\(^{24}\)A MySQL database at the site tracks data that are produced by camera, merger and housekeeping systems.

\(^{25}\)This is configured with one hot spare drive – thus if one drive fails it will automatically rebuild onto the hot spare and one additional drive can fail before there is data loss. This is convenient for remote operation because operators at the site may not be able to quickly reach San Pedro de Atacama to rebuild a degraded array. No failures have occurred to-date.

\(^{26}\)Subsequent seasons will only keep and transport the merged product, or approximately 210 GB/day (for three arrays).

\(^{27}\)The visualization software KST [http://kst.kde.org/](http://kst.kde.org/) was developed and managed by B. Netterfield. The data streaming software was developed by J. Fowler and A. Hincks based on a model developed by B. Netterfield, A. Hincks, D. Wiebe for BLAST.

---

65
3.3 Data products

Table 3.2: Description of the storage and peripherals of the ACT site and processing computers

<table>
<thead>
<tr>
<th>System</th>
<th>Location</th>
<th>CPU</th>
<th>HD (GB)</th>
<th>RAM (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housekeeping</td>
<td>Site</td>
<td>2 x 3 GHz</td>
<td>376</td>
<td>1</td>
</tr>
<tr>
<td>Terminal 1</td>
<td>Site</td>
<td>1 x 2.5 GHz</td>
<td>330</td>
<td>2</td>
</tr>
<tr>
<td>Terminal 2</td>
<td>Ground</td>
<td>2 x 3 GHz</td>
<td>230</td>
<td>4</td>
</tr>
<tr>
<td>Gateway, server</td>
<td>Ground</td>
<td>1 x 2.5 GHz</td>
<td>236</td>
<td>1</td>
</tr>
<tr>
<td>MCE acquisition '07</td>
<td>Site</td>
<td>1 x 3.4 GHz</td>
<td>152</td>
<td>0.5</td>
</tr>
<tr>
<td>Data merging '07</td>
<td>Site</td>
<td>2 x 3.0 GHz</td>
<td>752</td>
<td>1.5</td>
</tr>
<tr>
<td>Chile storage '07</td>
<td>Ground</td>
<td>2 x 3.0 GHz</td>
<td>4000</td>
<td>8</td>
</tr>
<tr>
<td>Main storage/analysis node '07</td>
<td>Princeton</td>
<td>4 x 2 x 2 GHz</td>
<td>9400</td>
<td>32</td>
</tr>
<tr>
<td>Data merging '08</td>
<td>Site</td>
<td>2 x 2.5 GHz</td>
<td>2400</td>
<td>3.3</td>
</tr>
<tr>
<td>MCE acquisition '08</td>
<td>Site</td>
<td>1 x 3.0 GHz</td>
<td>400</td>
<td>2</td>
</tr>
<tr>
<td>Chile storage '08</td>
<td>Ground</td>
<td>2 x 3.0 GHz</td>
<td>11000</td>
<td>8</td>
</tr>
<tr>
<td>Chile analysis '08</td>
<td>Ground</td>
<td>2 x 4 x 2.5 GHz</td>
<td>6500</td>
<td>16</td>
</tr>
<tr>
<td>Main analysis node '08</td>
<td>Princeton</td>
<td>92 x 2 x 3 GHz Beowulf</td>
<td>2500</td>
<td>1-4</td>
</tr>
<tr>
<td>Main storage node '08</td>
<td>Princeton</td>
<td>2 x 2 x 2.8 GHz</td>
<td>48000</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 3.3: Basic parameters of the ACT site and processing computers. The list is divided into systems that were common to both season (first section), systems specific to the first season in 2007 (second section), and systems specific to the second season in 2008. The gateway server manages site access from the outside world, a website, and the file database. Terminals on-site and in the ground station display real-time data, and the “Chile analysis 08” node performs basic analysis and mapping tasks in the ground station to give immediate feedback to telescope operators.
the merger queries the database\textsuperscript{28} to determine which files still need to be merged and calls the main merger code (implemented in C). For those data that need to be merged, the merger client connects to the camera and housekeeping data servers and requests the first \(\sim 1\) second of data for that time and seeks to find time stamps in common. It then proceeds through the camera data, associating each record with a record in the housekeeping data with the same stamp.\textsuperscript{29}

The output merged data products are stored in a so-called “dirfile” file format. Here, each directory contains one file for each channel and a format file which specifies the calibrations for the channels. With \(1024\) detector and \(366\) housekeeping files, there is a moderate filesystem overhead, and opening the output requires a large number of open file handles. We did not find these issues to outweigh the convenience the format affords. The dirfile format was advocated by B. Netterfield. It is native to the visualization software \textit{kst}, and avoids large disk strides when reading a single detector at a time – this facilitates fast single-detector analysis tasks and does not impede full-array analysis. The alternative is a flat format, where all channels are packed into a single file, such as the \textit{fits} format standard. We found that this was awkward for the variety of different data types and sampling rates. Dirfile channels can also be compressed individually, simplifying algorithms that reduce the data rate.

Every computer runs a daemon process responsible for automatically transporting data from the telescope to the ground station and then to North America.\textsuperscript{30} Data are automatically relocated from the site computers to RAID storage in the ground station, and then removed from the site computers when the database has identified that they are older than three days and that they have been merged. Data in the ground station are also moved automatically to transfer disks, and then deleted after they have been received in Princeton. We have used both the LaCie 600 GB external drive (USB/striped) and 750 GB and 1 TB commodity internal hard drives (read with a USB/SATA adapter) as they became available. Throughout, md5 checksums confirm that the data were compressed, transferred, and transported properly.

To manage the large data volume, J. Fowler developed a lossless delta compression scheme for each channel to reduce the bit length of each record (assuming each channel’s signal exercises only a fraction of the dynamic range of the data type). The algorithm achieves a compression rate \(\sim 3\) times faster than \textit{gzip}, and an uncompression rate that is similar to \textit{gzip}. For the camera data, a savings of \(\sim 60\%\) is typical for the algorithm, compared to \(\sim 10\%\) for \textit{gzip}. Housekeeping data typically achieve 80\% and 70\% for our algorithm and \textit{gzip}, respectively.\textsuperscript{31}

### 3.4 Coordinating observations and data flow

#### 3.4.1 Message passing

The dominant design considerations for the software commanding system were that it: 1) provide a means to control telescope systems from North America through the Internet; 2) provide both a graphical interface as well as a “robot interface” for scheduled operation; 3) permit systems at the site to communicate with each other if the microwave downlink between the site and the ground station fails; 4) tie the heterogeneous interfaces together into a common interface that is easy to manage; 5) provide centralized logging and accountability for system-wide commands.

\textsuperscript{28}The file management database was developed by M. Nolta and T. Marriage and tracks files from acquisition to analysis.

\textsuperscript{29}Note that the merger rate exceeds the acquisition rate, so the merger client may be met with partial data at the end of the file as it is still being acquired. It will wait for \(\sim 5\) seconds for new data to be written, and if nothing appears, it interprets this as the end of the file – the merger does not have any expectation of the file length.

\textsuperscript{30}Transport uses the standard \texttt{rsync} utility, without compression enabled.

\textsuperscript{31}Huffman coding of the reduced bits led to modest compression gains but doubled the processing time. Lossy compression in CMB experiments and ultimate limits can be found in Maris et al. (2004); Gaztañaga et al. (2001); Gaztanaga et al. (1998). This was experimental in 2007, but is a standard feature of the merged data in 2008.
3.4 Coordinating observations and data flow

Table 3.4: A summary of merged data product channels. Size “m” denotes mixed (some 4-byte numbers and some bitfields). The 399 Hz “encoder+timing” channels comprise two motion encoders, a cpu time in seconds and microseconds, two serial numbers (housekeeping and MCE) and two internal counters from the MCE camera electronics.

<table>
<thead>
<tr>
<th>Group</th>
<th>System</th>
<th>size (bytes)</th>
<th>rate (Hz)</th>
<th>quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>camera data</td>
<td>MCE</td>
<td>4</td>
<td>399</td>
<td>(32\times32)+(1\times32)</td>
</tr>
<tr>
<td>encoders + timing</td>
<td>Sync Box, Housekeeping</td>
<td>4</td>
<td>399</td>
<td>8</td>
</tr>
<tr>
<td>cryogenic thermometry</td>
<td>Housekeeping</td>
<td>4</td>
<td>99.7</td>
<td>48</td>
</tr>
<tr>
<td>telescope mechanical</td>
<td>KUKA, Housekeeping</td>
<td>4</td>
<td>1-99.7</td>
<td>80</td>
</tr>
<tr>
<td>housekeeping drivers</td>
<td>Housekeeping</td>
<td>m</td>
<td>1-99.7</td>
<td>46</td>
</tr>
<tr>
<td>auxiliary</td>
<td>Housekeeping</td>
<td>4</td>
<td>1-99.7</td>
<td>44</td>
</tr>
<tr>
<td>parameters, flags</td>
<td>Various</td>
<td>m</td>
<td>1-399</td>
<td>213</td>
</tr>
<tr>
<td>total rate</td>
<td></td>
<td></td>
<td></td>
<td>54 kb/sec</td>
</tr>
</tbody>
</table>

Figure 3.11: The GUI message-passing server window, here displaying users and instruments, and bindings between those users and command destinations. For example, mce1, mce2, and mce3 identify with “mce control” so a command sent to the destination “mce control” will be executed on the three MCEs simultaneously.

The model that we developed is related to an Internet “chat” client-server system. Operator clients and instruments log in to a central command router on the mountain (the “Interface Server” in Fig. 3.5). Instruments can then declare themselves as “destinations” of known command groups. An operator client can then send a command to a destination, and the server will handle passing that information to any clients or instruments that have identified with that destination. Each client logs in with a unique user name, and all commands or messages are logged with that name and a time stamp. A version-controlled database gives admissible commands, and the GUI for clients builds itself dynamically from this database based on flags that specify the type of entry field (on/off, value, string), default values, and human-readable descriptions of the command or variable. The GUI also provides a server window that displays transactions and allows operators to send a global message (for other operators). This is shown in Fig. 3.11 and Fig. 3.12. We found this system to be convenient especially during engineering periods because several operators can run engineering tests simultaneously from multiple locations and stay mutually aware of the tests.

Users or the scheduler can apply a global lock to prevent interference during an observation. This can be undone, though, if operators need to change the system state during a schedule because of specific failure. Users must have a unique name, so if the connection to the server terminates, the user may have to choose a new login name if their old link has not been reset. Future versions should require the GUI to respond to a ping from the server to determine if they have timed out.
3.4.2 Scheduled observation

Observations with ACT are defined through a schedule file that specifies an execution time for a task with given parameters. There is no decision structure in the scheduler itself. The schedule generator provides flexibility for operators to remake a schedule if conditions or goals for that observation change, though. Scheduled events are translated to specific device calls in the schedule dispatcher. The interface server then passes these to the designated devices (see Fig. 3.5).

3.5 Operational considerations

The failure rate of hard drives rapidly increases at lower pressures as their head flying height decreases, making operation unreliable by 5190 m (see Deng et al. (1994); Strom et al. (2006)). For the first season, we used a mixture of pressurized AT/IDE and USB-style hard drives, while subsequent seasons will use exclusively pressurized SATA or USB interfaces. Appendix B.3 describes the pressure and temperature readout for these drives.

We did not find solid state drives (either flash or RAM) to be a cost-effective solution for the data volumes required. For all computers at the site, the bandwidth of USB 2.0 (480 Mbps) is sufficient for the acquisition tasks and modern BIOS’s and Linux kernels permit system boot directly from an external USB device. All systems that require fast transfers to transport disks and large (11 TB) storage requirements are in the ground station, which is at an altitude of 2400 m, where pressurized drives are no longer needed.

Site electricity is alternately supplied by one of two XJ150 John Deere (Triton Power Generation) diesel generators, with sufficient on-site fuel for 60 days of autonomous operation at ~ 240 liters per day. The generators are rated at 150 kW (at sea level), but typical consumption during

Figure 3.12: The GUI housekeeping control window. Here clients can control relays, servos and the cycle state of cryogenics at the site from North America. Other windows provide a similar interfaces for commanding telescope motion and data acquisition systems.

33The primary reason for this is that our scan strategy involves simple scanning over several patches at constant elevation and central azimuth through the night, with occasional repointings to map a planet. Each mapping region is approximately covered each night, so that a missed patch will be covered subsequently. To control systematic errors, keeping a constant pointing is a premium in favor of repointing to measure strips that were missed during a fault. The scheduler was developed by J. Fowler, A. Hincks, T. Marriage and M. Nolta and comprises a schedule generator and dispatcher. M. Nolta led this effort, and is the primary observation coordinator. E. S. was responsible for the MCE interface, message passing and operator GUI.

34The AT style drive was needed for compatibility with the first season camera DAS, and the 40-pin-capable pressure vessel was inherited from MINT.
3.5 Operational considerations

Full operation is approximately 35 kW. All site systems operate at 120 V and 60 Hz, and have uninterruptible power supplies and several Geist\textsuperscript{35} IP power strips provide the capability to recycle power of some devices remotely.

The site is accessible with a 4 × 4 truck from the Paso Jama road (from Chile to Argentina and Bolivia), and a turn-off on a former mining road to the site. Large shipments can be delivered through a graded road from the ALMA preserve. Travel from the San Pedro de Atacama ground station to the site is roughly one hour via the mining road. While it is not feasible for a human operator to be continuously at the site, operation can be monitored from the ground station and problems can be addressed within an hour.

A Motorola Canopy Power IDU 150/300 provides a data downlink from the telescope site to the ground station\textsuperscript{36} where Entel\textsuperscript{37} provides continuous Internet service. This permits access to the site computers from North America, but the Internet link is only fast enough for engineering data and commanding, but not the main season’s data. Electricity in San Pedro de Atacama is sometimes sporadic, and restricted in the evening, so it must be supplemented by a diesel generator for the ground station electronics in order to keep the link to the Internet live, to display system health to operators in San Pedro de Atacama or North America, and to keep up with buffering of data from the site. This is provided by Astro Norte along with the Internet and office space.

\textsuperscript{35}http://www.geistmfg.com/
\textsuperscript{36}This unit broadcasts at 5 GHz with 63 mW total power. The link capacity for this power is rated to 94 Mbps and is matched to the 100 Mbps ground station network.
\textsuperscript{37}http://www.entel.cl/
Chapter 4

Site Weather and Stability

4.1 Introduction

The weather in Northern Chile is dominated by a subtropical high pressure region over the Pacific Ocean. This makes the Atacama one of the driest regions on Earth. In December, this region migrates southward from changes in solar heating. Circulation in a Bolivian high pressure system can then transport moisture to the site (see Erasmus and van Staden (2001)). In the Andes east of the Atacama, the two main sites for sub-mm and CMB experiments are the Llano Chajnantor (5050 m) and the Pampa la Bola (4750 m). ACT is the only experiment currently situated on Cerro Toco at 5190 m, but is within several kilometers of both other regions.

The atmospheric conditions determine the quality of observations through several mechanisms. These are: optical depth, loading, fluctuation and structure, and surface weather. The most dramatic surface weather phenomena at the site are snow and high winds. Both of these can limit work, and typically occur in the wet season when the telescope is being readied for observations in the dry season. Conditions that prevent work at the site occur infrequently. When they do, the main concern is that the generators remain fueled and operational to maintain the cryogenic systems.

This chapter describes weather information that is relevant for analysis, including water vapor and its relation to loading and optical depth (Sec. 4.2), drift in the atmospheric loading over a night (Sec. 4.3), turbulence (Sec. 4.4), and external temperatures (Sec. 4.5). All data described here are from the 2007 season of ACT.

4.2 Water vapor

Atmospheric optical depth in the ACT bands (denoted throughout as $\tau(\nu)$) has both wet and dry (non-water) components.\(^1\) The 145 GHz array is in the blue wing of the 119 GHz $O_2$ resonance and the red wing of the water 183 GHz resonance (see Danese and Partridge (1989)), while the attenuation across the 220 GHz and 280 GHz bands rises smoothly and is bracketed by the water resonances at 183 GHz and 325 GHz.\(^2\) A direct consequence of the atmospheric attenuation is that the temperature of the atmosphere is converted into an effective emission temperature. This translates to power at the level of $\sim pW$ absorbed by the detectors. We will see in Chapter 5 that this is a consideration for calibration because the changes in atmospheric loading modify the

\(^1\)Extra-terrestrial radiation is attenuated by $e^{-\tau}$. Throughout, “column” will refer to the path of the radiation through the atmosphere.

\(^2\)Ozone also contributes lines, and because it exists mainly the stratosphere, it produces a narrower family of lines; see Pardo et al. (2001).
4.2 Water vapor

operating point and responsivity of the detectors, but that this effect is small. Fluctuations of water vapor in the column also produce drifts, which are described in this section, Sec. 4.4 and Sec. 5.5. These drifts have a strong diurnal “weather” component, and at higher temporal frequencies (less than one hour) a “turbulent” component dominated by fluctuations in the amount of water vapor in the column.

The surface pressure is a tracer of the dry component, and the maximum excursion of the pressure over the first season was $\sim 3\%$. The optical depth contributed by the dry component is $\sim 50\%$ of the total on a typical night. The majority of the optical depth is from water vapor, which can vary widely. Last season reached lows of $\sim 0.3$ mm of precipitable vapor, but also reached values exceeding $6$ mm. (Eq. 4.1 defines what is meant by millimeters of precipitable vapor.) Here we will assume that the dry component is static and focus on understanding how the water vapor component evolves. The effect of the atmosphere is also modulated by the total airmass as $\sec(\theta_{\text{zenith}})$ in a planar atmosphere (which is also accurate for the Earth at pointings we are considering at elevations $> 30^\circ$). All the atmospheric properties described here depend on the telescope pointing.

The distribution of water vapor in the atmosphere depends on the presence of temperature inversions. For periods without strong inversion, the water vapor concentration decays exponentially with altitude and is described by a scale height, while for strong inversion, most vapor is held below a cap which varies in altitude. Both conditions have been observed in radiosonde studies by Giovanelli et al. (2001a) in the region, but low-lying inversion layers are shown there to be prevalent at night. The typical inversion layer is $\sim 5400$ m at night and $\sim 6000$ m during the day, but the altitude, strength (and even number of inversions) has been observed to vary considerably. On average, the water density distribution still follows an exponential scaling. For Chajnantor near Cerro Toco, Giovanelli et al. (2001a) find a water vapor scale height of $1100$ m (with upper and lower quartiles of $840$ m and $1500$ m). The pressure scale height is $\sim 7200$ m and the temperature lapse rate is between $7$ K/km and $9.3$ K/km (see Bustos et al. (2000)).

The standard measure of the amount of water vapor in a column is the precipitable water vapor, PWV. It is defined as the equivalent depth of water in the liquid phase along a column at zenith, or

$$\text{PWV} = \frac{1}{\rho_{\text{liquid}}} \frac{m}{A} = \frac{m_{w}}{\rho_{\text{liquid}}} \int_{0}^{h} n(x) \, dx \to \rho(0) \cdot h_{s} \tag{4.1}$$

where $A$ and $m$ are the area and mass of water in the column, $\rho_{\text{liquid}}$ is the density in the liquid phase, $\rho(0)$ is the density of water at the surface (e.g., $5190$ m), $m_{w}$ is the mass per molecule, $n(x)$ and $\rho(x)$ are the number density and density as a function of altitude $x$ from the ground and $h_{s}$ is the scale height. In the last statement, we take the upper limit $h \to \infty$ and assume an exponential water distribution.

ALMA has developed a series of radiometers at different frequencies to measure the PWV. For real-time tracking they use 183 GHz radiometers to measure the sky temperature relative to a calibrator and convert this to a PWV. These corroborate their 225 GHz radiometer studies. For

---

3For 1 mm of precipitable vapor, the fractions of dry optical depth are $\sim 50\%$, $\sim 20\%$ and $\sim 30\%$ for the 145 GHz, 220 GHz, and 280 GHz arrays, respectively.

4Currently the most reliable indicator of the inversion properties is a radiosonde. Inference of a turbulent layer height should also be possible from correlations in the ACT data and the known overlap of the beams through the atmosphere.

5This represents the e-folding length for water vapor for a given altitude above the surface, which for us is at an elevation of $5190$ m.

6Relative humidity at the surface is the ratio of the partial pressure of water to the partial pressure of saturated air, and depends on temperature. Along with the dewpoint, it is mainly useful for understanding precipitation at the site. To convert this to PWV requires finding the absolute humidity and assuming, for example, an exponential model for water vapor with a known scale height. A radiometer provides a much more direct measurement of the amount of vapor in a column.

7The full ALMA/APEX weather dataset can be accessed through http://www.apex-telescope.org/weather/Historical_weather/. We acknowledge Claudio Agurto for his help.
The PWV reported by the ALMA radiometer studies also matches well with a wide re-analysis of regional radiosondes and weather station and satellite information (Bustos et al. (2000)).

In an exponential water distribution model, the PWV at one altitude \( h_1 \) relative to the PWV at a different altitude \( h_2 \) is \( PWV(h_2) = PWV(h_1)e^{-(h_2-h_1)/h_s} \), thus the difference of 140 m in altitude between the APEX (Chajnantor) and ACT sites provides a \(~ 12\%\) reduction in PWV between the two sites. There are several uncertainties in this simple correction: 1) the water scale height is not constant and because of regular thermal inversions, an exponential model may not be appropriate, 2) the pointing will impact the water vapor along a column during the observation,\(^9\) 3) because of geography the difference in elevation is not necessarily equivalent to the altitude above the surface that the scale height describes. A spare ALMA/APEX radiometer is expected to be available for the ACT site for the second half of the 2008 season and subsequent seasons to perform some of these studies. Until then, we will use the reduction by \(~ 12\%\) with the understanding that it may be inaccurate. Fig. 4.1 shows a compilation of APEX radiometer data over the season.\(^10\)

The scaling of the wet (precipitable water vapor) and dry molecular oxygen optical depth in the three ACT bands was developed in Marriage (2006) using the ATM model (see Pardo et al. (2001)) and rescaled here to 50.5° elevation:

\[
\begin{align*}
\tau_{\text{dec}} &= 0.025 \cdot (\text{PWV mm}) + 0.012 \quad (4.2) \\
\tau_{\text{null}} &= 0.058 \cdot (\text{PWV mm}) + 0.010 \quad (4.3) \\
\tau_{\text{inc}} &= 0.097 \cdot (\text{PWV mm}) + 0.026 \quad (4.4)
\end{align*}
\]

and the effective RJ sky temperatures\(^12\) from this analysis are

\[
\begin{align*}
T_{\text{dec}} &= (6.0 \cdot (\text{PWV mm}) + 3.5) \cdot K \quad (4.5) \\
T_{\text{null}} &= (12.6 \cdot (\text{PWV mm}) + 4.1) \cdot K \quad (4.6) \\
T_{\text{inc}} &= (18.8 \cdot (\text{PWV mm}) + 9.3) \cdot K. \quad (4.7)
\end{align*}
\]

Then, using the estimate for the conversion from emission temperature to power absorbed from Niemack (2008) (and assuming 65% detector efficiency, in accordance with atmospheric drift and planet measurements)

\[
\begin{align*}
P_{\text{dec}} &= (0.43 \cdot (\text{PWV mm}) + 0.25) \cdot \text{pW} \quad (4.8) \\
P_{\text{null}} &= (1.49 \cdot (\text{PWV mm}) + 0.48) \cdot \text{pW} \quad (4.9) \\
P_{\text{inc}} &= (5.63 \cdot (\text{PWV mm}) + 2.80) \cdot \text{pW}. \quad (4.10)
\end{align*}
\]

---

\(^8\)There are many studies in sub-mm wavelengths of the atmosphere in the region. See Giovanelli et al. (2001b,a); Matsushita and Matsuo (2003); Matsushita et al. (1999, 2000); Matsuo et al. (1998a,b); Matsushita et al. (2000), ALMA note 384, and Serabyn et al. (1998); Radford (2002); Archibald et al. (2002) for general surveys.

\(^9\)To one side of the ACT site is the plain of San Pedro \(~ 2500\) m below while to the other is the Chajnantor plain at approximately the same altitude as the site. The actual water vapor in a column depends on the pointing relative to these directions or the wind direction. Delgado and Nyman (2001) study the pointing dependence of the PWV from east to west and find departures at the level of \(~ 5\%\), though ranging from very tight agreement to departures of \(~ 10\%\). Further study is needed to understand the difference in PWV between the rising and setting ACT observations.

\(^10\)The PWV radiometer noise is variable and often has short bursts of anomalous PWV. The median over \(~ 30\) minute time scales gives a cleaner estimate. This median is used throughout.

\(^11\)The optical depths in Marriage (2006) are given at elevation \(~ 45\)° and are modulated by the airmass for the pointing. We use the airmass model in Young (1994), but a simple model \(\tau(\theta_s) = \tau_0 \sec(\theta_s)\) is sufficient for elevations greater than \(~ 20\)° used in practice. Inhomogeneities in water vapor along lines of sight produce additional scatter in the relationship.

\(^12\)Effective temperature in the Rayleigh-Jeans limit for a given intensity \(I_\nu, I_\nu = 2k_B T_{\text{RJ}} \nu^2/c^2\).
4.2 Water vapor

Figure 4.1: Mean precipitable water vapor (PWV) from the APEX radiometer. This is the thickness of a sheet of liquid-phase water that is equivalent to the quantity of water through zenith pointing. ACT season 1 science observations spanned Nov. 14, 2007 to Dec. 17, 2007. Anomalous PWV around Nov. 21, Dec. 2, and Dec. 15 restricted observations. There is also a diurnal variation of the PWV that is apparent after removing a daily mean. The amount of water vapor in the column as measured by the APEX radiometer typically increases by roughly a precipitable millimeter from the morning until sunset, but the magnitude of the variation depends on other weather trends. The black lines indicate the first season ACT observation window from \( \sim 10 \) PM to \( \sim 10 \) AM – note that this starts at the upper black line (10 PM), moves to the top of the plot, then wraps back from 0 (the bottom of the plot) to the lower black line at 10 AM. All other weather plots indicate this interval the same way. Between 10 PM and 10 AM, the PWV typically falls, but may have additional periods of moisture or even large increases from weather systems such as those that produced the anomalously high PWV in the bright bands here.

Here “dec”, “null” and “inc” denote the values for the 145 GHz, 220 GHz, and 280 GHz cameras. Here “dec”, “null” and “inc” denote the values for the 145 GHz, 220 GHz, and 280 GHz cameras.13

Fig. 4.2 shows the spectrum of emission for several typical levels of PWV.

To simplify the analysis, we will assume that atmospheric attenuation can be decomposed into frequency and time-dependent pieces. Then

\[
\exp[-\tau(\nu)] \approx \exp[-\tau(t)]g_A(\nu),
\]

where \( \tau(t) \) is the optical depth at band-center.14

13These are fits to the tabulated value in Marriage (2006) so have uncertainties from the fit that are unrelated to the model, typically \( \sim 5\% \) percent, and the power absorbed depends an assumed efficiency and optical parameters. This has been measured to be 0.43 (see Sec. 4.3.3) from 2007 data, and a subject of future work is to check these predictions for 220 GHz and 280 GHz. The optical depth measured from Saturn observations in Sec. 5.9 is \( \tau = 0.031 \cdot (\text{PWV mm}) + C \).

14In addition to the total attenuation for the three arrays, another consideration is that the attenuation will change across the band, modifying the passband for extra-terrestrial radiation. Between a low PWV (~0.3 mm) night and a high PWV night (several millimeters), the relative contributions of the dry and wet components of the optical depth will vary by a factor of a few. The 145 GHz array is the focus of the work here, and trades off between the decreasing absorption with increasing frequency in the blue wing of the oxygen line and the increasing absorption with increasing frequency in the red wing of the water line. As shown in Fig. 4.2 the emission temperature has a fractional change of \( \sim 20\% \) over the 145 GHz passband (the gray shaded region), going from \( \sim 0.7 \) mm PWV to \( \sim 1.5 \) mm PWV. The total attenuation is \( \sim 4\% \), so as a rough figure, the effective bandpass changes by \( \sim 1\% \) across the band. Analogous conclusions can be drawn for 220 GHz and 280 GHz. The atmospheric passband is relevant to the calibration because it affects the effective frequencies where the planet temperatures and conversion from RJ units to CMB units are evaluated. We have not included it in Chapter 5.
Site Weather and Stability

Figure 4.2: Atmospheric effective RJ brightness temperature from the ATM model, Pardo et al. (2001) from Marriage (2006). The main emission here is from the $\text{O}_2$ 119 GHz resonance and the 183 GHz water resonance. Here, “Summer Morning” is 1.55 mm, “Winter Morning” is 0.69 mm, “Summer Evening” is 2.15 mm, and “Winter Evening” is 0.94 mm. The median PWV values and figure are from Marriage (2006).

4.3 Slow drifts

In this section, we study several sources of drift over long timescales. The dominant drift\textsuperscript{15} is optical and is from changing atmospheric loading. A smaller, but non-negligible component is non-optical and is from temperature drifts of the 3 K stage that holds the SQUID series array. The optical component of the drift can be calibrated into temperature or absorbed power based on the calibration developed in Chapter 5 or load curves. These conversions connect fluctuations in PWV over the night to system drifts, and give an estimate of the spectrum of atmospheric noise. We will use the conversion from PWV changes to power changes in Chapter 5 to study the change in responsivity over the night in 2007 data.

4.3.1 Weather

Ideally, we would measure the atmospheric emission temperature directly from the data over a night, but a given detector can have arbitrary offsets that periodically jump, and the absolute level of

---

\textsuperscript{15}We take drift to mean changes in the absolute level of the detector output, and will indicate when this also produces a responsivity drift. Non-optical “back-end” drift will not change the detector operating point (or responsivity), while optical drift will also produce responsivity drift.
4.3 Slow drifts

any array median has an arbitrary offset from the total power absorbed. To reconstruct this long time-scale drift, we calibrate the raw data into either pW or $\mu K$ using the calibration developed in Chapter 5. (In the case of pW, this is the estimate of absorbed power from the load curve analysis; see Eq. 5.37.) To reduce the data volume we consider either 0.4 Hz or 12.5 Hz downsampling depending on whether we want to see gross trends over the season, or atmospheric structure over shorter times. We estimate the atmospheric drift iteratively over each 15-minute acquisition interval. In a first pass, we find the median time stream from the subset of detectors. In this chapter, the drift is estimated from working detectors in column 15. This is partly for economy of the calculation and partly because the amplitude of the drift is also expected to vary with elevation. The only cut that is applied is to remove detectors that are known to be dead. Taking the median is more robust to jumps and gives a first-pass estimate. In a second pass we take a weighted average, where the weights for each detector are determined from the departure of that detector from the drift determined in the previous step. This gives a new drift time stream estimate. The iteration proceeds by weighting detectors by their departure from the previous average. (This converges quickly, and throughout we use 10 iterations.) We then paste the estimated drift from each 15-minute file together by taking the first left edge to be zero, the left edge of the second 15-minute chunk to be equal to the right edge of the first chunk and so on until a contiguous estimate of the drift over the night is assembled. Fig. 4.5 shows the correspondence between this drift and the PWV and Fig. 4.3 shows the time and frequency domain (Fig 4.4) response to the atmosphere calibrated into effective Rayleigh-Jeans temperature differences. This pasting procedure gives a reasonable understanding of the change in power absorbed by the array and the atmospheric emission temperature over an evening ($\sim 10$ hours). The gross trend of the drift and PWV are similar, but there are variations between the two on time scales less than one hour. We attribute this to the difference between lines of sight of the APEX radiometer and ACT, and to turbulence.

4.3.2 Cryogenic drift

The atmosphere is not the only source of drift, so to understand the loading component the other drift terms should be controlled against or subtracted. The only other drift term identified to-date is the dependence on the 3 K stage cryogenic temperatures from the temperature dependence of the SQUID series array. An aluminum cover blank can be used to block the atmospheric signal and isolate the (internal) 3 K component of the system drift. Fig. 4.6 shows the least-squared best scaling between the detector data and the cryoperm cap temperature (which is a good proxy for the 3 K stage temperature of the SQUID series array). The drift of the 3 K temperatures has very little structure faster than several minutes because of the high heat capacity

---

16 The offset is due to the fact that the SQUID response is periodic, so the absolute magnetic flux from the detector circuit inductor is only known modulo $\phi_0$. Each time a new lock is achieved, it is for feedback that nulls the SQUID response. The magnetic flux will have some finite, fixed value when the loop is locked.

17 Columns are horizontal on the sky, so column 0 and 31 observe different airmasses. For the $\sim 25'$ extent of the array, the airmass varies by 0.9%, so this effect is small.

18 This is biased by the central region of detectors because they have more in common with other detectors (on average, because of the overlap of their beams on the sky), than detectors on the edge of the array. See the discussion in Sec. 5.5.

19 Future work is needed to understand if there are discontinuities in the drift between the rising and setting pointings. The gross trend continues between rising and setting, but we anticipate some discontinuity on shorter time scales. Another direction would be to study the weights that are the product of this iterative scheme. These should correspond to flags for pathological data segments from detectors, but would also contain information about sub-array noise structure and its variability.

20 Further work is needed to quantify atmospheric noise in the science band. This becomes more difficult because of sub-array drift structure and correlations at frequencies $\sim 1$ Hz. Ultimately, the best measure of noise will be to project maps onto the sky for noise in different frequency ranges or with different atmospheric treatments. To calibrate the ACT data, we rely on the APEX radiometer for absolute PWV. Another area for further work is to understand the errors introduced by the difference in water vapor between the APEX and ACT atmospheric columns.

21 These data were taken by D. Swetz and S. Staggs at 1970 epoch time 1194368947.
of the 3 K stage. Typical drift on this stage was a drop of 0.2 K over the night. Multiplying by the response seen on dead detectors, this is equivalent to a change in absorbed optical power of $P \approx (3.6 \times 10^9) \times 0.2 \times (1.7 \times 10^{-7}) \sim 0.1$ pW over the night, making it comparable to the atmospheric loading drift over long times. Some care must be taken in reporting the 3 K drift in units of picowatts because it is not really an optical power, and only provides a point of comparison for atmospheric loading studies. It does not modify the responsivity of the detectors, so, for example, the calibration pulse amplitude described in Sec. 5.7 would be preserved across 3 K drifts, while we will find that it is not preserved across atmospheric loading drifts, because then the actual operating point of the detector changes.

### 4.3.3 Loading from the atmosphere

The main utility of determining the drift over several hours is to understand variations in the absorbed atmospheric power. Scaling the ACT 145 GHz array atmospheric drift estimate to match the APEX radiometer outputs gives $\sim 0.43$ pW per APEX mm of PWV. (This is also consistent with the value for the absorbed power per millimeter of water that we find in Sec. 5.9.) Niemack (2008) converts the effective temperatures from Marriage (2006) into an estimate for the absorbed power. There, all losses except for the absorption efficiency of the detector are modelled out, assuming detector solid angles and areas in the geometric limit. This predicts $P = 0.66(\text{pW/mm})(\text{PWV}) + 0.38$ pW absorbed. In addition to other optical losses, Niemack (2008) finds that an absorption efficiency of 60% for the detectors is consistent with both in-lab cold load measurements and power received from the planets. The measurements of the atmospheric drift shown in Fig. 4.5 are
4.3 Slow drifts

![Graph of atmospheric emission temperature spectrum]

Figure 4.4: The power spectrum (in $\mu K(RJ)/\sqrt{Hz}$) of atmospheric drift, averaged over 10.5 hours. A Hann window is applied here and the Nyquist frequency is 6.25 Hz (the data have been downsampled by a factor of 32 for economy). The small bump at 0.05 Hz is from scan-synchronous magnetic field pickup in the SQUIDs and/or structure in the atmospheric emission that modulates with the scan. The relation between multipole $l$ and audio frequency $f$ is $l \approx \frac{\pi \theta}{2} \approx 180 f$. The drift roughly follows the expected $1/f$ pattern. Taking a single-detector NET of $\sim 1000 \mu K(CMB)\sqrt{s}$ implies a knee frequency of $\sim 1.4$ Hz. These data were acquired on a good night, representative of roughly half of the season where the single-detector knee frequency was below 2 Hz. The detector noise averaged down over the array is $\sim 1000 \mu K(CMB)\sqrt{s}/\sqrt{900}$. Thus, this $1/f$ drift structure presents a significant analysis challenge. The break in the slope at $\sim 2$ Hz could be due to several factors. The first is that at higher frequencies, there is more sub-structure in the atmosphere across the array that is uncorrelated between detectors. This would produce less noise power from the atmosphere relative to being fully correlated across the detectors. The second is that this is based on only the weighted average of working detectors in column 15 (for economy and common airmass), so detector noise can become comparable to the atmospheric fluctuations.

also consistent with efficiency inferred in this way, giving 65%.\textsuperscript{22}

Fig. 4.7 shows the relation applied in reverse to find an inferred PWV from the atmospheric drift estimate on the ACT array. For each night, we offset the ACT drift data to the least-square best fit to the APEX PWV (scaled by 0.88 for the difference in altitude) to estimate the arbitrary offset in the ACT drift power between days. The agreement between the two radiometers over long times is very good when there are sufficient data for comparison. One feature that stands out on some days (see December 7th, or November 25th in Fig. 4.7) is that in the morning, the PWV rapidly rises, but the power inferred from the ACT drift rises earlier and more rapidly than the APEX radiometer data. This could be a sign of additional loading from, for example, the dewar window warming in the morning, or the difference of ACT’s line of sight from APEX and inhomogeneous emission of water trapped in the soil. There are also several nights of low PWV where radiometers produce divergent results; see Fig. 4.7 (right panel).

\textsuperscript{22}This efficiency is not essential in developing a calibration because the entire response of the instrument can be calibrated with planets, so a component-by-component separation is not necessary. However, this decomposition of efficiencies is necessary to understand the absorbed power and device parameters. See Niemack (2008) for development of the instrument model.
Site Weather and Stability

Figure 4.5: Data taken November 22, 2008 demonstrating drift in the array driven by the atmosphere. Here, offsets between each 15-minute file segment are removed as described in the text. The power absorbed is inferred from load curves using Eq. 5.37, described in the calibration section. The absolute power offset is taken from Eq. 4.10 (because it currently cannot be determined from the ACT data) but the scaling of PWV to power is determined empirically. Chilean local time is $UT - 4$ hours. For the purpose of showing the relation between PWV and power absorbed, we show an interval where the PWV radiometer was reliable and shows clean structure over short timescales. This period had very high PWV, spanning between 2.35 mm and 3.5 mm because of a weather system, and is cut in the science data analysis. Fig. 4.7 shows nights with PWV of 0.4 – 1 mm. Note that the short timescale fluctuations differ between the APEX radiometer an ACT, but that over hours, the gross trend in the atmosphere is similar in both instruments. See Delgado and Nyman (2001) for a study of directional differences in PWV, which are expected to account for the finer structure here.

4.3.4 The 0.3 K cryogenic stage

The 300 mK stage had five ROx thermometers during ACT season 1 that are relevant for stability studies: three on silicon cards in the 145 GHz array, one on the copper array holder (“slab”) and one on the $^3$He adsorption refrigerator pot. The temperature of the stage was actively controlled using a heater and the ROx on the slab. The drift here was stabilized to better than 100 $\mu$K excursion over each observation, and typical drift over a night was $\sim \text{few} \times 10$ $\mu$K. Despite the stability of the actively controlled slab temperature, the ROXs on silicon cards 26 and 31 consistently drifted in opposite directions by $\sim 600$ $\mu$K over a typical night. These changes track the ambient telescope temperature (as measured by the primary panel thermometers) much more consistently than they track the loading from the atmosphere, and it is not plausible for the card temperatures to drift in opposite directions.\footnote{This suggests that the thermometry drifts may be due to temperature coefficients in the housekeeping electronics.} In addition to the atmosphere and 3 K, the 0.3 K stage could thus contribute to detector drift, but at this time we do not have a consistent handle on the effect. If the trend is different than atmospheric drift, it must be small because the overall drift is consistent with the atmosphere and 3 K stage temperature drift. If the 0.3 K stage temperature drift tracks the atmospheric loading, then it could amplify the effect of the atmosphere.
4.4 Turbulence

In addition to weather systems that produce fluctuations in the PWV over the period of days, there are also fluctuations over much shorter time-scales from the turbulent structure of the atmosphere.\footnote{See Stull (1988) for general discussion.} Gradients in surface winds can produce a so-called “surface layer,” convection from solar heating can produce a “mixing layer” that terminates on a convective boundary layer, and temperature inversions can have turbulence from the difference in densities and shear.\footnote{Shear in the inversion layer is thought to be the source of turbulence at the South Pole. See Bussmann et al. (2005).} Observations by ALMA suggest that the ACT site may have a mixture of effects; see Delgado and Nyman (2001) and Robson et al. (2001). They observe a thick (3D) turbulence layer that forms in the morning and relaxes late at night (on the basis of the fluctuation spectrum). At night, they find that correlations are consistent with a thin layer. Interferometric studies in Robson et al. (2001) find a turbulent layer at around 500 m (with structure detected in the cross correlation between two receivers spanning from the ground to around 1000 m) above the ALMA site. One interpretation of this is that during the day, thermals rise and interact with a cooler upper layer which falls, producing a thick turbu-

Figure 4.6: Left: The sensitivity of the effective absorbed power level to 3 K stage temperature drift (referenced by the cryoperm cap) with the 145 GHz camera covered. The colors here indicate the effective power per K change in the 3 K stage. Most detectors give $-1$ pW equivalent power (in pW) per K change. The 3 K drift sensitivity is non-optical, but we have referenced it to a power level here so that it can be easily applied to optical power measurements. The amplitude of the pickup is correlated with the SQUID $\nabla (\phi)$ slope. Right: The SQUID chain slope measured by applying a ramp to the stage 1 SQUID feedback. (This is the number of error digital units per feedback digital unit applied in open-loop mode.) The series array modules are in groups of 8 columns, so that the first module here is anomalous. Because the detectors are read out through a closed flux-lock loop, the slope impacts bandwidth, but does not impact the optical responsivity. Marriage (2006) shows that the power received by the detectors from emission of the optics is negligible, and even dead detectors exhibit the same 3 K stage drift. This drift is therefore consistent with offsets in the SQUID chain. Both the left and the right plot are oriented as the detectors appear on the sky, where “columns” correspond to the silicon card carriers and are horizontal on the sky. In the left panel, dark blue indicates dead detectors, such as column 22. (We have no pW/DAC calibration there, but the SQUID slope may be measurable.)
Figure 4.7: Top: A comparison of the PWV inferred from ACT drifts and the APEX radiometer. The ACT traces here are derived from column 15 of the 145 GHz array, so no airmass gradient across columns (which are horizontal on the sky) needs to be included. The ACT data have an arbitrary vertical offset each night, so for each observation interval we find the least-squared best offset to the APEX data. For period such as Nov. 18, 2007, there were insufficient PWV radiometer data to scale the ACT drifts properly. The conversion from APEX PWV to ACT absorbed power (as inferred from the load curves) is 0.43 pW/(APEX mm). Inverting this, we find the PWV from ACT (green-gray solid line). In the blue/darkest solid line, we have subtracted an estimate of the drift component from the 3 K stage temperature drift. The APEX PWV data are the red/lightest dotted trace. Here we scale the PWV down by 12% to estimate the PWV at the ACT site, based on the difference between the APEX radiometer and ACT elevations and a scale height of 1100 m. Bottom: Same traces as top frame showing Dec. 6-7, 2007. On Dec. 7, 2007, the blue curve (which has been corrected for 3 K drift) matches reasonably well with the reported PWV from APEX, while the ACT data from Dec. 6 show a different trend. These differences are not currently understood.
4.4 Turbulence

lent layer. At night, turbulence would then be confined to the inversion layer that forms and from residual turbulence in the mixing layer. A simplifying assumption is that the remaining turbulence is confined to a frozen sheet that moves with the wind speed at that altitude. Fig. 4.8 shows the surface wind direction from the APEX weather station. During the day, the flow is from the plains near San Pedro, while at night, it varies considerably.

Figure 4.8: Mean wind direction from the APEX weather station. During the day, wind normally rises from the dry plain of San Pedro at \( \sim 10 \) m/s. The largest gusts were during mid-day, and exceeded \( \sim 25 \) m/s for several hours. At night, the wind velocity drops to a few m/s (ranging from still to \( \sim 10 \) m/s and greater in gusts) and blows in either from a more northerly direction or from Bolivia/Argentina; see the nights Nov. 21, Dec. 2, Dec 15. Those nights also correspond to higher PWV.

The origin for the turbulence is that large-scale motion injects energy at an “outer scale.” Some component of power from these large modes breaks into smaller eddies, then the power from those smaller eddies breaks into smaller structure still. This process of continues until viscosity can damp the smallest-scale turbulence at some “inner scale.” Because the only physical scales in the problem are the injection (outer) and dissipation (inner) scale, the turbulent power in three dimensions between these scales is scale-free and can be modeled by Kolmogorov-Taylor theory to give power \( \propto k^{-11/3} \). This inner scale is typically several millimeters. Chapter 5 shows that the separation of adjacent detectors on the sky is much larger than the inner scale. Yet, for a layer at 1000 m, the beams of two detectors separated by less than 20 detector lengths will overlap so that the atmospheric fluctuations are coherent over large sections of the array. This is investigated further in Sec. 5.5.

The outer scale is not as strict, and has been taken to be everything from the turbulent layer thickness (tens to hundreds of meters) to several kilometers.\(^{27}\) We expect that detector separations on the array lie within the outer scale and exceed the inner scale so will see turbulence described by Kolmogorov-Taylor theory.

R. Dunner has found that the knee frequency (where the power spectral component of the drift and white noise components are equal) does not correlate strongly with the PWV. This is consistent with the finding in Bussmann et al. (2005) that the PWV is not a good predictor of

\(^{26}\)See Consortini and O’Donnell (1993); Bussmann et al. (2005); Lay and Halverson (2000); Church (1995).

\(^{27}\)The outer scale gives a rough boundary between “weather” and “turbulence.” See, e.g., Lay and Halverson (2000); Church (1995) for further discussion.
atmospheric fluctuations and stability. The knee frequency is the term relevant to understanding contamination from the atmosphere in the final maps, so should be used in parallel with the PWV for data cuts.

4.5 External temperatures

Fig. 4.9 shows temperatures at the APEX site near ACT and Fig. 4.10 shows the panel temperatures over period of a day. We have found that distortions to the primary from solar heating influence the beam radiation pattern and the power received from the planets. In the worst case, the planet response amplitudes are suppressed by nearly a factor of two. In Chapter 5, we will cut any source observations in the morning. A sun screen described in Fig. 4.10 improves the stability of the primary into the morning for seasons subsequent to 2007.

4.6 Conclusions

This section describes the connection between water vapor and atmospheric loading and optical depth. These are both essential to developing a calibration model. A component of the long time-scale drift in the science data stream is also from the 3 K stage temperature and can be removed either through the 3 K sensitivity shown in Fig. 4.6 or by exploiting the fact that the cryogenic drift is slow so that it can be removed by a high pass filter. Here we have used the 3 K correction and in Sec. 5.5 we use the high pass filter. The APEX PWV radiometer is essential to understanding the optical loading and attenuation, and will be used extensively in Chapter 5. The final relevant fact here leading into the development of a calibration is that planet amplitudes after sunrise (in 2007) should not be used for calibration. We have also given an overview of turbulence. The main conclusions are that adjacent detectors will view overlapping regions on a turbulent layer and be in the Kolmogorov-Taylor regime, and that the PWV may be a weak indicator for sub-array or short time-scale temperature fluctuations. The PWV is, however, a good indicator for drifts over periods
4.6 Conclusions

Figure 4.10: Primary mirror panel temperatures over a typical day during the first season (starting at epoch 1195775243) split into the panel groups from 1 to 8, where panel group 1 is at the bottom of the primary and 8 is at the top. The main conclusions of these data are that 1) there is a gradient (though determining its actual value would require a better absolute temperature calibration), 2) cooling occurs over hours, 3) there is abrupt and non-uniform heating in the morning. Mirror fiducial measurements show that while the temperature is dropping over the night, the structure is stable enough for observation (the structure scales, but the twists and gradients that develop during the day have relaxed) several hours after sunset. Here, Chilean local time is UT $- 4$ hours. There is significant beam deformation once the sun rises. After the 2007 season, a sun screen was installed that blocks sunlight on the back of the primary mirror structure, reducing the rate of this rise by a factor of $\sim 2$ or better.

longer than $\sim 15$ minutes. The nature of the turbulence may vary depending on the strength of thermal inversions and the time it takes for daytime turbulence to relax.
Chapter 5

The Instrumental Response to Radiation: Calibration

5.1 Introduction

The chain from a temperature difference in the CMB to a difference in digital units is: 1) $\Delta T_{\text{CMB}}$ produces $\Delta I_\nu$ in intensity, 2) $\Delta I_\nu$ in intensity leads to a change $\Delta P$ in the power received by a detector, 3) $\Delta P$ on the detector produces $\Delta I_{\text{TES}}$ in the current in the TES loop and 4) this produces $\Delta D$ of the digital data recorded to disk. Schematically,

$$\Delta T_{\text{CMB}} \rightarrow \Delta I_\nu \rightarrow \Delta P \rightarrow \Delta I_{\text{TES}} \rightarrow \Delta D.$$  \hspace{1cm} (5.1)

The goal of calibration is to convert a change in detector response $\Delta D$ in digital units (recorded as 32-bit numbers) into a change in the CMB temperature, $\Delta T_{\text{CMB}}$.\(^1\) The model that propagates $\Delta D$ to $\Delta T_{\text{CMB}}$ comprises $32 \times 32$ independent functions of time and sets the inferred power (in $\mu\text{K}^2$) of the anisotropy. This impacts determinations of cosmological parameters (in particular, $\sigma_8$) and the SZ observables. The push to measure these cosmological parameters at the percent level implies that the calibration should also be understood at that level or better.

The general problem of finding the best calibration algorithm is complicated because there are a large number of detectors, observations, and sources of calibration information. Each of these has its own particular property: some detectors are biased lower on the transition, planets were observed through a variety of atmospheric conditions, and some of the calibration information is relative over times, or across the array. Calibration of mm-wave experiments has become a standard procedure, but depends on particulars of the experiments and available calibration tools. Here we extend some of the standard tools, and develop several new models and methods for the $32 \times 32$ 145 GHz array.

One of the most powerful methods that has been developed recently is to calibrate against WMAP, which has achieved a calibration to $\sim 0.2\%$ from the CMB dipole.\(^2\) The spectral index of a planetary source relative to the CMB is also not a concern with this method. There is sufficient

---

\(^1\)For the CMB, this is a linear conversion, but the response can become nonlinear for large changes in power from planetary observations or atmospheric changes. Understanding the nonlinear response is important for calibrating into CMB temperature units because even though this response is linear, drift in the atmospheric loading will change the responsivity over the night, and the amplitude of a response to a planet that is used for a flux calibration will be biased by the nonlinear response.

\(^2\)Recently, ACBAR has achieved absolute map calibrations to $\sim 2\%$ using the WMAP data; see Reichardt et al. (2008). An outline of a possible procedure for ACT is to project the WMAP data into a time domain realization with an ACT simulation, make a map of the synthetic TOD, and use the cross-power to scale the ACT field.
5.1 Introduction

information to calibrate a final map using cross-power with WMAP, but that becomes more difficult if the calibration is broken into smaller domains that include variability across time and across the array. Here we focus on developing independent calibration methods to understand the instrument response over these finer domains. Using these, the response can be flattened and a single scaling to the WMAP results is sufficient for the global calibration. We will also develop a flux calibration using planet observations that will be complementary to the flux calibration from WMAP.

We take calibration to be the zero frequency (DC) detector response to a diffuse source, but planet and CMB observations probe a range of frequencies associated with the array crossing over the anisotropy as the telescope scans. For planets, this time is typically $\sim 10$ ms. Understanding the time-domain response is therefore essential for understanding the DC response of the instrument because of the large corrections to the amplitude of planet responses in slow detectors. Here we use the one-pole time constant model developed in Niemack (2008), which is based on TES-bias steps and has been calibrated to optical chopper response.

The components of the global calibration are 1) planet observations that give roughly one flux calibration per detector per night, 2) observations of the array-wide power received from the atmosphere that give a sharp ($\sim 1\%$) array-relative calibration, 3) an independent measure of the PWV that determines optical loading and the attenuation through the atmosphere, 4) load curves and calibration emitter pulses that tie the instrument response together across different times.

Here we describe the response of the instrument to radiation, working from intensity to power (Sec. 5.2), to electrical signals (Sec. 5.3), to digital information that is written to disk (Sec. 5.4). Sec. 5.3 gives a method for understanding the responsivity drift and nonlinear response of the array to large signals, and describes a method to find the responsivity from load curves, where the TES bias is ramped from the normal to the superconducting regime. Sec. 5.5 then describes the array-relative response inferred from drifts in the atmospheric loading. This levels the response across the array and is analogous to flat-fielding in optical astronomy.

Sec. 5.6.1 then describes the response of the telescope to unresolved sources such as planets. Planets give a flux calibration and an array-relative calibration. Some care must be taken in interpreting these data because the amplitude of the response to a planet depends on the dilution of the unresolved planetary disk by the telescope beam pattern. The beam pattern will vary across the array due to the distribution of detectors across the focal plane and particulars of the optical coupling to each detector. The variation of the dilution means that unresolved sources will have a different relative response across the array than sources that fill the beam (and are, therefore, not subject to dilution).

Sec. 5.7 describes a calibrated emitter that can be used to constrain the response of a detector throughout the season. The absolute power level of the emitter is not known independently and the illumination pattern is non-uniform, so these pulses are useful for testing and developing the model that propagates detector response in time.

The variability in detector response can be treated either with respect to a seasonal average, or a reference period. This is a choice between reporting, for example, that the response has changed $3\%$ relative to Dec. 10, or that the response has changed $3\%$ relative to a seasonal average. The average seems preferable because it is not with respect to an arbitrary time, but it is more difficult to treat rigorously because the calibration information comes from averages over different data sets (planets vs. calibration pulses vs. the atmosphere). The season may also still be in progress, so the notion of a season average would change in time. Because the

---

3Cosmic ray hits provide some information, but do not probe the same behavior of the absorber/TES system as optical response, and could also excite SQUID response.

4“Relative” can be interpreted in several ways here, so we will indicate array-relative (response across the array), and time-relative (same detector across time) separately.

5A standard method for deriving the array-relative response is to dip in elevation. We found that because of drifts in the atmospheric power over the dip that these were not nearly as accurate as other methods (such as load curves, TES bias steps, and the drift method developed in Sec. 4.1.)
seasonal average is ambiguous, we calibrate relative to a reference epoch. It is reasonable to pick an epoch where the PWV was typical of a good night (\( \sim 0.6 \) mm) and there is a solid relative calibration based on response to the diffuse atmosphere. Here, the reference time for the 2007 season coincides with the data for the relative calibration on Dec. 10, 2007. The reference epoch is absolutely arbitrary, but picking a typical night gives some meaning to a night being better or worse than another night.

Sec. 5.8 describes a statistic for the instrument stability and the efficacy of the calibration, and Sec. 5.9 describes the global calibration procedure. In Table 5.1, we have compiled a list of instrumental parameters that are relevant throughout the chapter and in Table 5.2, we have compiled a list of important variables in the chapter. Ultimately, we compress the calibration into a single number for the season that is propagated to all the detectors through relative calibration in time and across the array. This gives a random calibration error of 2.1% at 1\( \sigma \) per detector, though we believe the current systematic error is \( \sim 7\% \).\(^6\)

5.2 Discussion – Relation between incident flux and absorbed power

For a monochromatic point source with frequency \( \nu \) on the sky in the direction \( \hat{n} \), define an instrument response function \( \Psi_{\nu}(\Delta \hat{n}) \) as the ratio of the power received at some offset \( \Delta \hat{n} \) to the power received on center, or

\[
\Psi_{\nu}(\Delta \hat{n}) = \frac{P(\Delta \hat{n})}{P(\Delta \hat{n} = 0)} \bigg|_{\nu},
\]

(5.2)

The solid angle at frequency \( \nu \) is then the integral of \( \Psi_{\nu}(\Delta \hat{n}) \) over \( d\Omega_{\Delta \hat{n}} \). For the case of a Gaussian beam,

\[
\Omega_b(\nu) = \int \Psi_{\nu}(\Delta \hat{n}) d\Omega_{\Delta \hat{n}} = \int \exp \left[ -\frac{1}{2} \frac{\theta^2}{\sigma_b(\nu)^2} \right] 2\pi \theta d\theta = 2\pi \sigma_b^2(\nu),
\]

(5.3)

here assuming that the extent of the beam is small and axially symmetric so that \( d\Omega = \sin \theta d\theta d\phi \rightarrow 2\pi \theta d\theta d\phi \).

In the real system, a bandpass defines the frequency response, \( g_B(\nu) \), there are losses in the atmosphere through an optical depth \( \tau(\nu) \), and an overall efficiency, \( \eta(\nu) \). Define an effective area \( A_e(\nu) \) as the function for which the power absorbed by a detector from any source with surface brightness \( S_B(\hat{n}) \) is\(^7\) (see Page et al. (2003b))

\[
P(\Delta \hat{n}) = \int A_e(\nu) \Psi_{\nu}(R_{\Delta \hat{n}}) S_B(\hat{n}) \eta(\nu) g_B(\nu)e^{-\tau(\nu)} d\Omega d\nu,
\]

(5.4)

where \( R_{\Delta \hat{n}} \) rotates the telescope to a pointing \( \Delta \hat{n} \) relative to \( \hat{n} \), which is the integration variable here. Defining \( A_e(\nu) \) this way removes any subtlety about what area it represents.\(^8\)

The surface brightness can be separated into a spectral and spatial component as \( S_{\nu}(\hat{n}) = \sigma^S(\nu) \Psi^S(\hat{n}) \) under the assumption that the spatial extent of the source does not change with frequency over the receiver’s passband. Here \( \Psi^S \) is normalized in the same way as is \( \Psi \), so that it is unity on-center and \( \sigma^S(\nu) \) carries all the spectral information and the magnitude of the

\(^6\)The sense of this random error is the calibration’s ability to predict system response to a known signal at one time. The correlation of this error between two times separated by \( \delta t \) is the subject of future studies based on the TES bias step and calibration emitter. The random calibration error will then average down over the number of periods over which the calibration error is independent.

\(^7\)The surface brightness has units \( J \cdot s^{-1} \cdot m^{-2} \cdot Hz^{-1} \cdot ster^{-1} = 10^{26} Jy \cdot ster^{-1} \).

\(^8\)Note that \( \eta, \Psi_{\nu}(\Delta \hat{n}), A_e(\nu), \) and \( g_B(\nu) \) can all vary across the array.

---

87
5.2 Discussion – Relation between incident flux and absorbed power

Table 5.1: Physical parameters of the 145 GHz, 220 GHz, and 280 GHz array receivers. These values have been compiled from Cardiff University, studies by J. Fowler, T. Marriage, M. Niemack and studies here; see Marriage (2006) and Niemack (2008). The superscript \( m \), \( l \), and \( d \) denote quantities measured in the field and lab, or derived (based on the hardware design or models) respectively. The band designations 145 GHz, 220 GHz, and 280 GHz will be used throughout to refer to the arrays, but are not the exact band centers. The beam FWHM is from measurements by A. Hincks based on 2007 data and the 220 GHz and 280 GHz values are the model developed in Niemack (2008), rescaled relative to A. Hincks’ value. There is currently an active program to determine the beam FWHM and solid angle for the 145 GHz, 220 GHz, and 280 GHz arrays. The optical efficiency includes the lenses and dewar window, but not the mirror emissivity or band-defining filter stack (which is described by the “peak transmission” value here). The conversion from RJ K to absorbed power was developed in Niemack (2008), and has been rescaled by the efficiency inferred from the absorbed power and calibrated temperature of the atmosphere measured in Sec. 4.3 to estimate the values for 220 GHz and 280 GHz. This scaling is meant only to give a rough estimate for the 2008 season. The Joule power, leg conductance, saturation power and optical time constants are from lab data and design expectations, and are developed in Niemack (2008). The loading from the atmosphere here is the power absorbed and is based on calibrated measurements in 145 GHz described here, which are then used to scale estimates in Niemack (2008) for the dry components and 220 GHz and 280 GHz arrays. The emission temperature and optical depth are as estimated in Marriage (2006), and scaled to 50.5° elevation. PWV is always defined through zenith. The atmospheric loading, temperature and optical depth are split into wet (water) and dry terms. The optical response rolloff frequencies for the 220 GHz and 280 GHz arrays are indicated by \( \dagger \), and are rough estimates based on the device parameters, prior to deployment of the array or optical tests. The device parameters (Joule power, leg conductance, and saturation power) take on a wide range of values and the values here also represent rough estimates.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>145 GHz</th>
<th>220 GHz</th>
<th>280 GHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central frequency (GHz)</td>
<td>148(^{l})</td>
<td>215(^{l})</td>
<td>280(^{d})</td>
</tr>
<tr>
<td>Central frequency (GHz, RJ weight)</td>
<td>149(^{d})</td>
<td>216(^{d})</td>
<td>281(^{d})</td>
</tr>
<tr>
<td>Passband FWHM (GHz)</td>
<td>22.6(^{d})</td>
<td>21.6(^{d})</td>
<td>29.8(^{d})</td>
</tr>
<tr>
<td>RJ to CMB conversion</td>
<td>1.72(^{d})</td>
<td>2.95(^{d})</td>
<td>5.59(^{d})</td>
</tr>
<tr>
<td>Geometric beam FWHM (arcmin)</td>
<td>1.53(^{m})</td>
<td>1.14(^{d})</td>
<td>0.98(^{d})</td>
</tr>
<tr>
<td>Peak transmission</td>
<td>0.78(^{d})</td>
<td>0.63(^{d})</td>
<td>0.66(^{d})</td>
</tr>
<tr>
<td>Efficiency (optics)</td>
<td>0.855(^{d})</td>
<td>0.835(^{d})</td>
<td>0.815(^{d})</td>
</tr>
<tr>
<td>abs. pW per RJ K source incident</td>
<td>0.07(^{m})</td>
<td>0.11(^{d})</td>
<td>0.26(^{d})</td>
</tr>
<tr>
<td>Joule power (pW)</td>
<td>9(^{d})</td>
<td>15(^{d})</td>
<td>17(^{d})</td>
</tr>
<tr>
<td>Thermal leg conductance (pW/K)</td>
<td>90(^{d})</td>
<td>130(^{d})</td>
<td>180(^{d})</td>
</tr>
<tr>
<td>Saturation power (pW)</td>
<td>10(^{d})</td>
<td>18(^{d})</td>
<td>27(^{d})</td>
</tr>
<tr>
<td>Optical resp. rolloff (Hz, ( f_{3dB} ))</td>
<td>90(^{m})</td>
<td>130(^{d})</td>
<td>140(^{d})</td>
</tr>
<tr>
<td>Loading, atmosphere (pW/mm)</td>
<td>0.43(^{m})</td>
<td>1.49(^{d})</td>
<td>5.63(^{d})</td>
</tr>
<tr>
<td>Loading, atmosphere (pW, dry)</td>
<td>0.25(^{d})</td>
<td>0.48(^{d})</td>
<td>2.80(^{d})</td>
</tr>
<tr>
<td>Eff. temperature atmosphere (K/mm)</td>
<td>6.0(^{d})</td>
<td>12.6(^{d})</td>
<td>18.8(^{d})</td>
</tr>
<tr>
<td>Eff. temperature atmosphere (K, dry)</td>
<td>3.5(^{d})</td>
<td>4.1(^{d})</td>
<td>9.3(^{d})</td>
</tr>
<tr>
<td>( \tau ) atmosphere (per mm)</td>
<td>0.025(^{m})</td>
<td>0.058(^{d})</td>
<td>0.097(^{d})</td>
</tr>
<tr>
<td>( \tau ) atmosphere (dry)</td>
<td>0.012(^{d})</td>
<td>0.010(^{d})</td>
<td>0.026(^{d})</td>
</tr>
</tbody>
</table>
Table 5.2: Non-standard variables used in this chapter. We use \( f \) to denote “acoustic” frequencies in the signal such as between 0 – 400 Hz and \( \nu \) to denote optical frequencies and bandwidths.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Units</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Psi_f(\hat{n}) )</td>
<td>1</td>
<td>Frequency-dependent normalized radiation pattern</td>
</tr>
<tr>
<td>( \Psi(\hat{n}) )</td>
<td>1</td>
<td>Band-averaged normalized radiation pattern</td>
</tr>
<tr>
<td>( \tilde{\Psi}(\hat{n}) )</td>
<td>1</td>
<td>Estimated normalized radiation pattern</td>
</tr>
<tr>
<td>( A_f(\nu) ) cm(^2)</td>
<td></td>
<td>Frequency-dependent effective area</td>
</tr>
<tr>
<td>( A_\nu ) cm(^2)</td>
<td></td>
<td>Band-averaged effective area</td>
</tr>
<tr>
<td>( \Omega_b(\nu) ) ster</td>
<td></td>
<td>Main beam solid angle at ( \nu )</td>
</tr>
<tr>
<td>( \Omega_{b,i} ) ster</td>
<td></td>
<td>Band-averaged main beam solid angle for detector ( i )</td>
</tr>
<tr>
<td>( \tilde{\Omega}_{b,i} ) ster</td>
<td></td>
<td>Estimated main beam solid angle</td>
</tr>
<tr>
<td>( P(\hat{n}) ) pW</td>
<td></td>
<td>Power absorbed by detectors at pointing ( \hat{n} )</td>
</tr>
<tr>
<td>( g_A(\nu) ) 1</td>
<td></td>
<td>Effective atmosphere optical passband</td>
</tr>
<tr>
<td>( m(\nu) ) 1</td>
<td></td>
<td>Number of radiation modes received</td>
</tr>
<tr>
<td>( \Phi(\nu) ) 1</td>
<td></td>
<td>The conversion between RJ and CMB units at ( \nu )</td>
</tr>
<tr>
<td>( \Phi ) 1</td>
<td></td>
<td>Band-averaged conversion between RJ and CMB units</td>
</tr>
<tr>
<td>( I(\nu) ) J ( \cdot ) s(^{-1}) ( \cdot ) m(^{-2}) ( \cdot ) Hz(^{-1}) ( \cdot ) ster(^{-1})</td>
<td></td>
<td>Intensity</td>
</tr>
<tr>
<td>( B(\nu) ) J ( \cdot ) s(^{-1}) ( \cdot ) m(^{-2}) ( \cdot ) Hz(^{-1}) ( \cdot ) ster(^{-1})</td>
<td></td>
<td>Intensity from a black-body</td>
</tr>
<tr>
<td>( S^S(\hat{n}) ) J ( \cdot ) s(^{-1}) ( \cdot ) m(^{-2}) ( \cdot ) Hz(^{-1}) ( \cdot ) ster(^{-1})</td>
<td></td>
<td>Source surface brightness = ( \Psi^B(\hat{n})\sigma^S(\nu) )</td>
</tr>
<tr>
<td>( \Psi^S(\hat{n}) ) 1</td>
<td></td>
<td>Normalized radiation of a source on the sky</td>
</tr>
<tr>
<td>( \sigma^S(\nu) ) J ( \cdot ) s(^{-1}) ( \cdot ) m(^{-2}) ( \cdot ) Hz(^{-1}) ( \cdot ) ster(^{-1})</td>
<td></td>
<td>The source intensity</td>
</tr>
<tr>
<td>( \tilde{\sigma}^S ) J ( \cdot ) s(^{-1}) ( \cdot ) m(^{-2}) ( \cdot ) ster(^{-1})</td>
<td></td>
<td>Band-averaged source intensity</td>
</tr>
<tr>
<td>( \eta(\nu) ) 1</td>
<td></td>
<td>Total instrument efficiency</td>
</tr>
<tr>
<td>( g_B(\nu) ) 1</td>
<td></td>
<td>Passband transmission function</td>
</tr>
<tr>
<td>( g_A(\nu) ) 1</td>
<td></td>
<td>Atmospheric effective passband</td>
</tr>
<tr>
<td>( D ) 1</td>
<td></td>
<td>Dilution factor</td>
</tr>
<tr>
<td>( \sigma_b ) arcmin</td>
<td></td>
<td>Equivalent Gaussian beam width</td>
</tr>
<tr>
<td>( \beta ) Hz</td>
<td></td>
<td>Bandwidth-efficiency product</td>
</tr>
</tbody>
</table>

\( D \) 1 Raw data in filtered feedback DAC units
\( R \) pW\(^{-1}\) Conversion from power to DAC units \( D(t) = R(t)P(t) \)
\( \delta_e \) 1 Fractional change in responsivity
\( \delta_i \) 1 The array-relative calibration; detector \( i \), time \( t \)
\( \Theta(f_{3dB,i}) \) 1 Fractional array-relative response change in \( i \) from time \( t \) to \( t_f \)
\( \Theta_i(f_{3dB,i}) \) 1 Optical time constant correction to planet amplitudes
source intensity. (That is, \(\sigma^S(\nu)\) has not been normalized with respect to the intensity and any way.) The superscript \(S\) represents that it is the radiation pattern of the source. This spatial-spectral separation is a reasonable approximation for the planets, where even for the gas giants, the change in the effective disk size over the passband is negligible. With this separation of the surface brightness, the power integrated over frequency can be rewritten as

\[
P(\Delta \hat{n}) = \int \left[ \frac{\int A_e(\nu)\Psi(\nu)(\hat{n})S(\nu)\eta(\nu)g_B(\nu)e^{-\tau(\nu)}d\nu}{\int S(\nu)\eta(\nu)g_B(\nu)e^{-\tau(\nu)}d\nu} \right] \Psi(\hat{n}) \left( \int \sigma^S(\nu)\eta(\nu)g_B(\nu)e^{-\tau(\nu)}d\nu \right) d\Omega_n
\]

identifying the object in the first parentheses as a frequency-averaged area times the frequency-averaged normalized radiation pattern,

\[
\bar{A}_e \Psi(\hat{n}) = \frac{1}{\sigma^S} \int A_e(\nu)\Psi(\nu)(\hat{n})S(\nu)\eta(\nu)g_B(\nu)e^{-\tau(\nu)}d\nu \quad \text{and} \quad \bar{\sigma}^S = \int \sigma^S(\nu)\eta(\nu)g_B(\nu)e^{-\tau(\nu)}d\nu.
\]

Note that the effective \(\bar{A}_e \Psi(\hat{n})\) depends on the spectral distribution of the source, so the beam pattern that observes a thermal planet will be slightly different than a beam pattern that observes the CMB. (Sec. 5.2.1 investigates this effect.) The frequency-averaged telescope solid angle is then defined as the integral of the frequency-averaged response function\(^9\)

\[
\bar{\Omega}_b = \int \bar{\Psi}(\hat{n})d\Omega_n.
\]

The effective throughput of the instrument is then the integral of Eq. 5.6 over solid angle

\[
\bar{A}_e \bar{\Omega}_b = \frac{\int A_e(\nu)\Omega_b(\nu)\sigma^S(\nu)\eta(\nu)g_B(\nu)e^{-\tau(\nu)}d\nu}{\int \sigma^S(\nu)\eta(\nu)g_B(\nu)e^{-\tau(\nu)}d\nu}.
\]

The goal is to measure the CMB anisotropy structure, which has some spatial component \(\Psi^S(\hat{n})\) of radiation from the sky. When the telescope scans over this spatial component, the output of the detectors is the power received from the particular pattern \(\Psi^S(\hat{n})\) at \(\hat{n}\) convolved by the optical response, which includes both time constants and the beam pattern. The response on scales comparable to the beam size are eventually treated by a window function in \(\ell\) on the power spectrum, and time scales faster than the optical time constants are treated by inverting the assumed one-pole response.

We will take calibration to mean the zero-frequency, large-scale limit of this response. We will refer to objects much smaller than the beam solid angle as point sources, and spatial patterns much larger than the beam solid angle as diffuse sources. This can be quantified by a dilution factor \(D\) which is the integral of the telescope band-weighted radiation pattern \(\bar{\Psi}(\hat{n})\) against the source distribution \(\Psi^S(\hat{n})\) (both of these functions are normalized to unity for \(\hat{n} = 0\)) divided by the band-weighted telescope solid angle, as

\[
D = \frac{1}{\Omega_b} \int \bar{\Psi}(\hat{n})\Psi^S(\hat{n})d\Omega_n.
\]

\(^9\)In practice we find the solid angle by looking at the normalized response, so we effectively break \(\bar{A}_e \Psi(\hat{n})\) by dividing by a factor proportional to \(\bar{A}_e\), which does not depend on \(\hat{n}\). In doing this, note that \(\Psi(\hat{n})\) still depends on the integral against \(\bar{A}_e(\nu)\).
In the diffuse limit $D = 1$. We will examine the point limit in more detail in Sec. 5.6.1 to understand planet observations, but up to some subtleties, the dilution of a source with solid angle $\Omega_o$ is $D \approx \Omega_o/\Omega_b$. In the diffuse limit, $P = A_c \Omega_b \sigma S^o$, and the crucial quantity to convert between intensity and power is the throughput, $A_c \Omega_b$.

5.2.1 The effective throughput as a function of spectral index

Sources with different spectral indices will weight the passband differently, giving different beams and sensitivity. In this section, we derive a rough approximation for the effective throughput for sources with different spectral indices. To do this, we assume that the passband is square in frequency, and find all the power falling on the focal plane. It is important to note that the power absorbed by one detector will have a different dependence on the spectral index, and that this requires a detailed optical coupling model for sub-wavelength pixels. This is beyond the scope here, which is simply to demonstrate that the influence of the spectral index is small but may be considerable for foreground, planet, and SZ studies, where the spectral index is different than the CMB. The frequency dependence of the throughput across the passband is

$$\int_{e}^{e} (1 + \delta)^{3-2} d\delta \cdot \left[ \int_{e}^{e} (1 + \delta)^{3} d\delta \right]^{-1}$$

(5.10)

where the second integral is $(\bar{\delta}^3)^{-1}$ and we have defined $\delta \equiv (\nu - \nu_0)/\nu_0$. We can then find the fractional difference between the effective throughput for a source with spectral index $\beta_1$ and a source with index $\beta_2$. The integrand can be expanded in small $\delta$, anticipating small $\epsilon$. The first non-zero term in the expansion over $\epsilon$ is $O(\epsilon^2)$

$$\frac{\bar{A}_b |_{\beta_1} - \bar{A}_b |_{\beta_2}}{\bar{A}_b |_{\beta_1}} \approx \frac{2}{3} (\beta_1 - \beta_2) \epsilon^2,$$

(5.11)

so that the fractional difference in the effective throughput for a $\pm 10\%$ bandpass is $\sim 0.7\%$ per difference in the spectral index. The spectral index of a thermal source is $\approx 2$ (RJ), while the CMB at some $x = h\nu/(k_B T_c)$ is

$$\beta_{\text{CMB}} = 3 - \frac{\epsilon^x x}{\epsilon^x - 1} \quad \text{so that} \quad \lim_{x \to 0} \beta_{\text{CMB}} = 2.$$ 

(5.12)

The CMB spectral index starts at 2 in the RJ (Rayleigh-Jeans) limit, crosses zero, and then becomes negative, scaling as $3 - x$ for large $x$ in the Wien limit. The difference in spectral index of the thermal source and the CMB is shown in Fig. 5.2; for 280 GHz, the difference in index is $\sim 4$ and the bandwidth is nearly $\pm 15\%$, so the modification to the effective throughput is $\sim 6\%$. A similar calculation for 145 GHz gives $\sim 1\%$. Aside from the difference in throughput between the CMB and RJ sources (such as the planets, to a reasonable approximation), there are several other sources with widely varying spectral indices (see de Oliveira-Costa et al. (2008); Gold et al. (2008)): 1) synchrotron $\beta \sim -2.7$ (and $\sim -3.0$ at high latitudes), 2) free-free emission $\beta \sim -2$, and 3) dust $\beta \sim 2$.

---

\[10\]In an ideal, diffraction-limited optical system, all of the $\nu^{-2}$ dependence is absorbed in $\Omega_b$. In real systems, the illumination of the aperture also changes, giving a frequency-dependent effective area and non-trivial frequency dependence for $\Omega_b$. The full throughput scales as the number of modes times $\nu^{-2}$.
5.2 Discussion – Relation between incident flux and absorbed power

5.2.2 The beam

An azimuthally symmetric illumination pattern of $g(r)$ on a circular aperture produces the normalized monochromatic radiation pattern (see e.g., Rohlfs and Wilson (1996))

$$\Psi^2(\theta) = \frac{P(\theta)}{P(0)} = \left[ \int_0^\infty g(r) J_0(2\pi \sin(\theta) r) r dr / \int_0^\infty g(r) r dr \right]^2 .$$  (5.13)

Here, we track only the angular offset $\theta$ from the beam maximum and measure $r$ in physical radius divided by wavelength. If the aperture is uniformly illuminated ($g(r) = 0$ inside the aperture and zero outside) by a single frequency, this is the Airy function (which depends on the usual $J_1$).

The illumination pattern depends in detail on the optical coupling of detectors in the focal plane. Bolometric focal planes have traditionally been arrays of feedhorns ($\sim 2F\lambda$ spacing), and ACT is the first CMB telescope to use a free-space $0.6F\lambda$ (for 145 GHz) array. A significant difference between feed and free-space systems is that the bolometers have a large acceptance solid angle, so the coupling with the telescope must be set by a cold (Lyot) stop on an optical image of the primary and the other solid angle surrounding the detector must be cold enough to not add noise or spurious signals (see Griffin et al. (2002) or Holland et al. (2002)). In ACT, the detectors sit at the bottom of a black cavity at 0.3 K with a 0.3 K lens followed by a bandpass at its aperture. The aperture illumination from a feedhorn has significant taper, while for the free space detector the illumination is expected to be more uniform. For a Lambertian (diffuse reflector) surface, this is given by $\cos(\theta)$, until the cold stop edge, but in reality this is likely to be a more non-trivial function of $\theta$ because of the efficiency of the optical coupling as a function of angle. In both the horn and free-space detector, the tapered illumination $g(r)$ produces beams with lower sidelobes. The diffraction minima in Eq. 5.13 are washed out further by integrating across the wide bandwidth of the experiment.

Without any diffraction, a small square detector would be mapped to a square on the sky. With diffraction in the optical system through a circular aperture, the beam is the convolution of the square and an Airy disk. In the ACT arrays, the physical pixel size is always comparable to or less than the wavelength, so diffraction associated with the pixel is also important. This optical regime is not fully understood. For recent work on the radiation response of sub-wavelength arrays, see Wollack et al. (2006); Withington et al. (2004, 2005); Chuss et al. (2007); Withington and Saklatvala (2007); Saklatvala et al. (2006). (Fig. 4 of Chuss et al. (2007) gives the intensity pattern of emission from a detector as a function of its size.) The approach described in the literature above is to take the detector to be a black emitter, approximated by a black body behind a square aperture. This gives a pattern (due to correlations in the black body radiation) that is a square for pixels much larger than the wavelength, and roughly a Gaussian for pixels smaller than the wavelength. This response is then bandlimited by the optical aperture. As the pixel becomes smaller than the wavelength, the “squaring” of an Airy disk becomes less severe and the beam on the sky is only broadened and more azimuthally symmetric. If the pixel size is comparable to or less than the wavelength, so diffraction associated with the pixel is also important. This approach is then bandlimited by the optical aperture and free-space detector, the tapered illumination is expected to be more uniform. For a Lambertian (diffuse reflector) surface, this is $\cos(\theta)$, until the cold stop edge, but in reality this is likely to be a more non-trivial function of $\theta$ because of the efficiency of the optical coupling as a function of angle. In both the horn and free-space detector, the tapered illumination $g(r)$ produces beams with lower sidelobes. The diffraction minima in Eq. 5.13 are washed out further by integrating across the wide bandwidth of the experiment.

Without any diffraction, a small square detector would be mapped to a square on the sky. With diffraction in the optical system through a circular aperture, the beam is the convolution of the square and an Airy disk. In the ACT arrays, the physical pixel size is always comparable to or less than the wavelength, so diffraction associated with the pixel is also important. This optical regime is not fully understood. For recent work on the radiation response of sub-wavelength arrays, see Wollack et al. (2006); Withington et al. (2004, 2005); Chuss et al. (2007); Withington and Saklatvala (2007); Saklatvala et al. (2006). (Fig. 4 of Chuss et al. (2007) gives the intensity pattern of emission from a detector as a function of its size.) The approach described in the literature above is to take the detector to be a black emitter, approximated by a black body behind a square aperture. This gives a pattern (due to correlations in the black body radiation) that is a square for pixels much larger than the wavelength, and roughly a Gaussian for pixels smaller than the wavelength. This response is then bandlimited by the optical aperture. As the pixel becomes smaller than the wavelength, the “squaring” of an Airy disk becomes less severe and the beam on the sky is only broadened and more azimuthally symmetric. If the pixel size is comparable to or less than the wavelength, so diffraction associated with the pixel is also important.
the estimates of the beam profile and solid angle are

\[
\tilde{\Psi}(\Delta \hat{n}) = \frac{P(\Delta \hat{n})}{P(0)} \quad \text{and} \quad \tilde{\Omega}_b = \int \tilde{\Psi}(\Delta \hat{n})d\Omega_{\Delta \hat{n}}.
\]

(5.14)

Such an estimate is shown for observations of Saturn in Fig. 5.1. Despite the complications described here (sub-wavelength pixel size, tapered illumination from coupling layer efficiency as a function of angle), this profile matches well with an Airy disk convolved by the pixel size integrated over the passband; see Niemack (2008).

Figure 5.1: The function \(\tilde{\Psi}(\Delta \hat{n})\) for the 145 GHz camera measured from observations of Saturn. A. Hincks produced this map using the Cottingham basis spline drift estimation method described in Sec. 6.6 for data over a six minute interval. The total solid angle is 225 nsr. The combination of diffraction from the square (but small) pixel and the optical system produces a round beam.

5.2.3 CMB temperature anisotropies and absorbed power

The power received from a diffuse source is the throughput times the band-average of the spectral component \(\sigma^S(\nu)\) of the surface brightness, \(P = \tilde{A}_e \tilde{\Omega}_b \sigma^S\). We can make this more concrete by returning to Eq. 5.4 and integrating over \(\hat{n}\):

\[
P = \int A_e(\nu)\Omega_b(\nu)\sigma^S(\nu)\eta(\nu)g_B(\nu)e^{-\tau(\nu)}d\nu.
\]

(5.15)

The frequency-dependent throughput \(A_e(\nu)\Omega_b(\nu)\) can be replaced by the modal relation

\[
A_e(\nu)\Omega_b(\nu) = \frac{m(\nu)}{2}c^2/\nu^2,
\]

(5.16)

where \(m(\nu)\) is the number of (spatial and polarization) radiation modes received.

We have two handles on the optical response in the field: 1) the main beam solid angle and 2) an estimate of the power absorbed by a detector from load curve analysis. We will see that even the latter is subject to uncertainties in electrical parameters. There are several other optical properties that we do not currently understand with the confidence needed for calibration, where the goal is to reach percent-level (or better) predictions. These are: 1) the effective primary area, 2) the empirically-determined aperture illumination function, 3) the detector solid angle and effective
5.2 Discussion – Relation between incident flux and absorbed power

area (because of diffraction) the detector optical coupling and full optical loss model. The optical coupling is complicated because of pixel diffraction and possible angular dependence of the AR coupling layer (a matching layer in front of the detector array which may introduce angle-dependent efficiency), but a model for this behavior can be compared to the observed beam patterns and an empirically determined aperture illumination function. From $A_\nu (v) \Omega_\nu (v) = [m(\nu)/2] c^2 / \nu^2$, we see that our lack of knowledge of the effective area trades off with a lack of knowledge of the number of modes received. These factors are common between CMB and planet observations, up to how they are weighted by the difference between the spectral slopes of planets and the CMB. In this chapter, we focus on deriving the full optical response to the CMB and to planets. Even if not all aspects of the optical response to a planet and the CMB are understood independently to the percent-level (because of the factors described above), we can still use the planets to accurately calibrate the CMB response because of the shared terms in the CMB and planet response. We will therefore leave $m(\nu)$ as a free-function.

We can now apply the approximation described in Sec. 4.2 (Eq. 4.11) letting $\exp [-\tau (t)] \approx \exp [-\tau (t)] g_A (\nu)$,

$$P = e^{-\tau (t)} \int_0^\infty \frac{m(\nu)c^2}{2 \nu^2} \sigma S (\nu) \eta (\nu) g_B (\nu) g_A (\nu) d\nu. \quad (5.17)$$

The intensity incident from a Planck source with temperature $T$ is

$$\sigma S (\nu) \bigg|_{\text{Planck}} = B_\nu (T) = \frac{2 h \nu^3 / c^2}{e^{h \nu/k_B T} - 1}. \quad (5.18)$$

To first order, the change in intensity from a change in the blackbody temperature is

$$\Delta B_\nu (T) = \Delta T_{\text{CMB}} \frac{dB}{dT} \bigg|_{T_{\text{CMB}}} = \Delta T_{\text{CMB}} \frac{2 \nu^2 k_B}{c^2} \Phi_\nu (T_{\text{CMB}}) \quad (5.19)$$

where $\Phi \equiv c^2/(2 \nu^2 k_B) \cdot dB/dT$ is a unitless conversion between temperature and intensity units. In terms of $x = (h \nu)/(k_B T_R) = \nu/(20.84 \text{ GHz} \cdot 2.725)$,

$$\Phi_\nu (T_{\text{CMB}}) = \frac{x^2 e^x}{(e^x - 1)^2}. \quad (5.20)$$

For a change $\Delta T_{\text{CMB}}$ in the CMB temperature, the change in power is

$$\frac{\Delta P}{k_B \Delta T_{\text{CMB}}} = e^{-\tau} \int_0^\infty m(\nu) \Phi_\nu (T_{\text{CMB}}) \eta (\nu) g_B (\nu) g_A (\nu) d\nu. \quad (5.21)$$

To get a rough approximation, take $\Phi$, $m(\nu)$ and $g_A (\nu)$ constant over the passband. Then

$$\Delta P \approx \eta m e^{-\tau} \Phi_\nu (T_{\text{CMB}})(k_B \Delta T_{\text{CMB}}) \Delta \nu. \quad (5.22)$$

Define a band-averaged CMB temperature unit conversion $\bar{\Phi}$ and an efficiency-bandwidth product $\bar{B}$ as

$$\bar{\Phi} \equiv \int_0^\infty \frac{m(\nu) \Phi_\nu (T_{\text{CMB}}) \eta (\nu) g_B (\nu) g_A (\nu) d\nu}{\int_0^\infty m(\nu) \eta (\nu) g_B (\nu) g_A (\nu) d\nu}, \quad \bar{B} \equiv \int_0^\infty m(\nu) \eta (\nu) g_B (\nu) g_A (\nu) d\nu, \quad (5.23)$$

so that

$$\frac{\Delta P}{\Delta T_{\text{CMB}}} = e^{-\tau} \bar{\Phi} k_B \bar{B}. \quad (5.24)$$
The Instrumental Response to Radiation: Calibration

5.3 Detectors – electrical response

In the previous section, we derived an expression for the absorbed power from a temperature anisotropy on the sky. In this section, we develop an understanding of how a change in the absorbed power is converted into an electrical response. In a bolometer without electrothermal feedback, the change in temperature is simply \( \Delta T / G \), where \( \Delta T \) is the change in absorbed power and \( G \) is the leg conductance. In response to some \( \Delta T_{RJ} \) a bolometer will change temperature by

\[
\Delta T_{bol} \sim \eta m (k_B \Delta \nu / G) \Delta T_{RJ} \sim (3.1 \text{ mK}) \eta m \Delta T_{RJ} [\Delta \nu / 20 \text{ GHz}] / [G/90(\text{pW/K})].
\]  

(5.25)

A typical transition width is 2 mK (Niemack (2008)) and a rough figure for \( m \eta \) in these terms from planet observations is \( \sim 30\% \).\(^{15}\) Without additional compensation, the large temperature changes in the atmosphere (which can produce \( \Delta T_{RJ} \) of several Kelvin) will easily exercise the transition. The resistance as a function of temperature is locally linear, but it varies widely over the full transition, so compensation is also important to improve the linearity.

To improve the dynamic range, linearity, and speed of the detectors, we operate with negative electrothermal feedback. The TES biasing and readout circuit used in ACT is shown in Fig. 5.3. The TES is voltage biased in this configuration (so long as \( R_{TES} \gg R_{sh} \)) so that the Joule bias power \( P_J = V_{TES}^2 / R_{TES} \) decreases in response to an increase in \( R_{TES} \), compensating some of the incident optical power. The extent of the cancellation depends on the slope of the transition. For detectors in what we will call the “core” range of operating points between 20\% and 70\% of \( R_n \), there is nearly total cancellation between the optical power and Joule power.

The coupled equations for the electrothermal system at fixed bias current are (see e.g., Hilton

\(^{15}\)A. Hincks, private communication.
5.3 Detectors – electrical response

Figure 5.3: The detector circuit. An absorber receives power from the sky, its temperature increases, and the TES detector responds by increasing in resistance. The resulting current increase changes the magnetic field produced by the inductor that is sensed by the first stage SQUID (SQ1). A digital control loop (running at 15.15 kHz) locks the output of the SQUID amplifier chain by applying a current to the feedback inductor $L_{fb}$ that compensates changes in the sky signal. Typical values for normal resistances are $30 \text{ m}\Omega$, the bias points are $\sim 0.3$ times the normal resistance, and the shunt resistances are $0.7 \text{ m}\Omega$ (for the 145 GHz array in the 2007 season; see Niemack (2008)). The ratio of the mutual inductance between the input inductor and SQUID to the mutual inductance between feedback inductor and the SQUID is $\sim 8.5$ and the typical current through a biased detector is $0.3 \text{ mA}$.

and Irwin (2005))

\[
CT = P_J + P - P_{bath} \\
L\dot{I} = V - I[R_{sh} + R_{TES}(I, T)].
\] (5.26)

Here, the heat flows in from the sky $P$ and the bias power $P_J$ and flows out through the bath, $-P_{bath}$. The current $I$ through the TES loop is driven by the Thevenin-equivalent bias voltage $V = R_{sh}I_b$ and is nonlinear in the TES resistance $R_{TES}$. Voltage changes at frequencies $\sim \text{kHz}$ in the loop are suppressed by the overall inductance $L$ which is $\sim 0.7 \text{ \mu H}$ and is the sum of the Nyquist inductor, the SQUID coupling inductor, and the parasitic inductance. The parasitic and Nyquist inductors are not shown in Fig. 5.3. Parasitic resistance in addition to $R_{sh}$ was found to be negligible for the 145 GHz array (see Niemack (2008)). The Joule power is simply $V^2_{TES}/R_{TES}$ and the power to the bath is $K(T^n - T_{b}^n)$, where $n$ and $K$ describe the heat transport between the absorber and the bath (see Zhao et al. (2008) for values and discussion).

5.3.1 Response to small signals

Observations of the CMB will be in the regime where $\delta P$ is small. The change in Joule power can be expanded to first order in changes $\delta I$ and $\delta T$ as

\[
\delta P_J = \delta(I^2R_{TES}) = (2 + \beta)I_oR_o\delta I + \alpha \frac{P_{th}}{T_o}\delta T,
\] (5.27)

16The atmosphere dominates the drift and is not in the linear response regime for hour-to-hour changes over an evening. These changes modify the operating point over the night but the linear regime still holds over the times scales of the CMB measurement. This derivation is similar to Niemack (2008) and Hilton and Irwin (2005), except that we leave $\alpha$ and $R_{sh}$ in the equations to anticipate responsivity variations dependent on them.
The Instrumental Response to Radiation: Calibration

where \( \alpha \) and \( \beta \) are customarily defined as unitless logarithmic derivatives of the TES response on the transition

\[
\alpha = \frac{I_o}{R_o} \frac{dR_{TES}}{dT} \bigg|_{T_o} \quad \text{and} \quad \beta = \frac{I_o}{R_o} \frac{dR_{TES}}{dI} \bigg|_{T_o} \quad (5.28)
\]

\[
\Rightarrow \quad \delta R_{TES} = \alpha R_o \frac{dT}{T_o} + \beta I_o R_o \frac{dI}{T_o} \quad (5.29)
\]

(Quantities evaluated at the operating point are denoted by a subscript nought throughout. Note that \( \alpha, \beta \) are functions of \( R_o \). For values and discussion, see Zhao et al. (2008).) We can then use these to find the linearized dynamics for the temperature

\[
C \frac{dT}{dt} = \delta P - \delta P_{\text{bath}} \quad (5.30)
\]

\[
= \delta P + (2 + \beta) I_o R_o \delta I - \frac{C}{T_o} \left( 1 - \frac{P_{3,0}}{GT_o} \right) \delta T, \quad (5.31)
\]

where \( \tau = C/G \) is the natural thermal time constant of the system. The linearized electronic part of the equation is then

\[
L \frac{dI}{dt} = - \frac{L}{\tau_{\text{el}}} \left( \alpha P_{3,0} GT_o \right) \frac{dT}{dt} + (2 + \beta) I_o R_o \delta I \quad (5.32)
\]

where we have identified an electrical time constant,

\[
\tau_{\text{el}} = \frac{L}{R_{\text{sh}} + R_o (1 + \beta I_o)} \quad (5.33)
\]

This defines the system response to a change in power. For calibration, the relevant quantity is the responsivity at zero frequency, so we can set the time derivatives to zero and eliminate \( \delta T \) between the electrical and thermal equations. This gives

\[
\frac{\partial P}{\partial I} \bigg|_{\omega=0} = - \frac{C}{T_o} \left( 1 - \frac{P_{3,0}}{GT_o} \right) \frac{LT_o}{\alpha I_o R_o \tau_{\text{el}}} - (2 + \beta) I_o R_o \quad (5.34)
\]

\[
= \Gamma I_o \frac{L}{\tau_{\text{el}}} (2 + \beta) I_o R_o \quad (5.35)
\]

where in the last equality we have taken \( R_{\text{sh}} \) small compared to \( R_o \) to first order and we have defined

\[
\Gamma = 1 - \frac{P_{3,0}}{GT_o} = 1 - \mathcal{L}^{-1} \quad (5.36)
\]

The term \( \frac{P_{3,0}}{GT_o} \) is a loop gain which is commonly denoted as \( \mathcal{L} \). The loop gain becomes low for operating points high on the transition so that \( \Gamma \) becomes large. The responsivity is the inverse of the loop gain.
5.3 Detectors – electrical response

of Eq. 5.34, so that a large $\Gamma$ greatly reduces the response of the detector to changes in optical power in the normal regime.\(^{17}\)

On the transition ($\sim 20\% - 70\%$ of $R_o$) the loop gain is high and Eq. 5.34 reduces to\(^{18}\)

$$\frac{\delta P}{\delta I} \bigg|_{\omega=0, \text{ideal}} = -I_o (R_o - R_{sh}).$$

(5.37)

A loop gain so high that $L^{-1} \rightarrow 0$ is equivalent to electrothermal feedback being completely effective, where changes in power produce changes in the TES resistance at constant temperature. To calculate a rough figure for the loop gain take $P_{3,o}$ of 6 pW to 9 pW with an operating point $\sim 0.5$ K, $G$ of $\sim 90$ pW·K$^{-1}$, and $\alpha$ of 100 to 200; see Zhao et al. (2008) and Niemack (2008). This gives loop gains between 13 and 40, very roughly. Those studies also show $\beta$ that is between 0 (high on the transition) to a factor of a few low on the transition.\(^{19}\)

This zero-frequency responsiveness to power can be determined using load curve data, and Eq. 5.37. The procedure for the load curves\(^{20}\) is: 1) set the detectors normal by applying a current through the TES bias exceeding the critical current 2) drop to a bias in the normal metal regime and ramp it to zero. There is an arbitrary offset in the measured TES current through the SQUID back end, but this can be constrained by requiring that the normal branch extrapolated to zero voltages gives zero current. When the bias power falls below $P_{\text{sat}} - P_{\text{optical}}$, the load curve probes the superconducting transition. Using this, we can infer the voltage/current through the TES at its operating point. With a known shunt resistance value and Eq. 5.37, this gives the responsivity. The average responsivity and bias points across 145 GHz in the 2007 season are shown in Fig. 5.4. Note that unless stated otherwise, responsivity is calculated in the ideal (infinite loop gain) limit, not by Eq. 5.34, which depends on $\alpha$ and $\beta$ (which, themselves, are functions of the operating point). Two other considerations are that 1) the responsivity only describes the relation between power absorbed and current, so does not contain any information about the optical coupling, 2) the shunt resistors are not known to the accuracy we need for the calibration, so absolute values of the responsivity are mainly useful for instrument characterization.

5.3.2 Nonlinear response to large signals

If the incident optical power is the same order of magnitude as the saturation power,\(^{21}\) the response invalidates the expansion for small signals used above. This can occur for bright planets or large drifts in the optical loading. The complete solution for nonlinear response must be found by numerically integrating the electrothermal system. This solution would be needed to understand the response to bright planets, so these objects should be avoided except for pointing studies. The main effect of nonlinear response that is relevant to ACT science goals is that drifts in atmospheric power modify the linear-regime responsivity over the night. Rather than solve the nonlinear-differential equation, in this case it is possible to look at the steady state and find the

\(^{17}\)Note that even though the responsivity approaches zero in the normal regime, nothing catastrophic happens to $R_o$, simply $R_o \rightarrow R_o$. The ideal limit in Eq. 5.37 that we use elsewhere only depends on $R_o$, so cannot be used high on the transition near the normal state.

\(^{18}\)Responsivity is defined as $dI/dP$, but for economy, we write and manipulate $dP/dI$.

\(^{19}\)Note that some of the 220 GHz and 280 GHz camera columns have a larger $R_{sh}/R_o$ ratio, giving different response low on the transition. Another way to estimate the loop gain is through the time constant as a function of Joule power. Here, if we neglect $R_{sh}/R_o$, $f_{MB} \approx (G/(2\pi C)) \left(1 + L/(1 + \beta)\right)$ and using the value found in Niemack (2008), 33 Hz · $L/(1 + \beta) = 6.4$ Hz/pW · $P_J$, and for Joule powers of 6 pW to 9 pW, this gives loop gains that are a factor of a few smaller. This is one of the strongest indications that the ordinary electrothermal circuit gives the incorrect time-domain response. R. Dünnen and Y. Zhao have found that a model with additional weakly linked heat capacities explains the complex impedance and noise data better.

\(^{20}\)See Niemack (2008) for additional considerations.

\(^{21}\)Saturation powers were $\sim 10$ pW, $\sim 18$ pW and $\sim 27$ pW for the 145 GHz, 220 GHz and 280 GHz arrays. Typical Mars and Saturn observations in the 2007 season provided 0.3 pW power on the detector absorber.
change in operating resistance due to a change in power. The steady state of Eq. 5.26 gives the thermal balance

\[ 0 = I^2 R + P_{\text{opt}} - P_{\text{bath}} \quad \text{and} \quad 0 = V - I(R_{\text{sh}} + R). \]  

(Note that \( V \) here is the Thevenin voltage, not \( V_{\text{TES}} \).) We would then like to find the change in resistance produced by a change in power. It is convenient to move everything that depends on the TES resistance \( R \) into a function

\[ f(R) \equiv P_{\text{bath}}(R) - \frac{V^2 R}{(R_{\text{sh}} + R)^2} = P_{\text{opt}}. \]  

At some later time, the power (from e.g., sky loading) will change by \( \Delta P \), then

\[ f(R + \Delta R) \equiv P_{\text{bath}}(R + \Delta R) - \frac{V^2(R + \Delta R)}{(R_{\text{sh}} + (R + \Delta R))^2} = P_{\text{opt}} + \Delta P. \]  

We can then consider \( f(R + \Delta R) - f(R) = \Delta P \) and for a given \( \Delta P \) and starting point on the transition, find \( \Delta R \). The non-trivial part of this nonlinear equation is the bath power as a function of resistance. We will find that there are drastic shifts in responsivity when the detectors are driven into and out of the normal regime. This suggests that rather than trying to find an accurate model for the saturation nonlinearity, we simply try to put a bound on the nonlinear response and use this to cut detectors. When the detector is biased on the transition (less than e.g., 70\% of the normal resistance), we will see that we can ignore \( P_{\text{bath}}(R) \), so that the responsivity change from optical loading only requires knowing the operating point, bias power and the shunt resistance. The calibration software provides a “propagate” option that can account for the Joule term changes.
5.3 Detectors – electrical response

and also provides a matrix of flags for detectors that have saturated or have \( R < R_{\text{sh}} \), based on radiometer data and pointing (for the airmass correction).

A model for the resistance as a function of the Joule power which accounts for the rough behavior of the saturated (superconducting and normal) states is\(^{22}\)

\[
R = \frac{R_n}{2} \left[ \frac{2}{\pi} \tan^{-1}\left(60 \cdot \text{pW}^{-1}\right)\left(P - P_o\right) + 1 \right] \quad P - P_o = -\frac{1}{60 \cdot \text{pW}^{-1}} \tan \left[ \frac{\pi}{2} \left(\frac{2R}{R_n} - 1\right) \right].
\] (5.41)

Here \( R \) is the TES resistance, \( P_o \) is the Joule power at the center of the transition and \( P \) is an offset power. This form is purely phenomenological, and the factor of 60 \( \text{pW}^{-1} \) is the transition slope and is consistent with typical response from load curves. This does not handle the “kink” between the transition region and normal region, so will on average over-predict the nonlinearity high on the transition. This model matches well with the transition inferred from IV curves. The other components of this model are the bias power and the resistances. A typical normal resistance in the 145 GHz camera was 30 m\( \Omega \), while a typical shunt resistor was 0.7 m\( \Omega \). We take the bias power to be 6 pW at 0.5\( R_n \). Results from this are shown in Fig. 5.5.\(^{23}\)

We can then calculate the change in responsivity from this change \( \Delta R \) in the operating point. To do this, note that the power must balance as \( P = P_{\text{bath}} - I^2 R \). The response to power is simply \( dI/dP \), and the non-trivial term here is \( P_{\text{bath}}(I) \). We can find its derivative using the model in Eq. 5.41, giving

\[
\frac{dP_{\text{bath}}}{dI} = -\frac{1}{I^2} \frac{\pi}{60 \cdot \text{pW}^{-1}} \frac{V}{R_n} \sec^2 \left[ \frac{\pi}{2} \left(\frac{2R}{R_n} - 1\right) \right].
\] (5.42)

The term \(-I^2 R\) simply gives \(-R_{\text{sh}}(I_b - 2I) = I(R - R_{\text{sh}})\), which was the high loop gain limit responsivity we found in Eq. 5.37. Denote \( dI/dP \) responsivity at operating resistance \( R \) and power \( P \) as \( s(\omega = 0, R) \). We can then find \([s(\omega = 0, R + \Delta R) - s(\omega = 0, R)]/s(\omega = 0, R)\). This is the fractional change in the responsivity as a function of the change in operating point and is shown in Fig. 5.6.

These examples show that until saturation, the change in responsivity with loading is dominated by the effect of the Joule heating term. In this case,

\[
f(R) \approx -\frac{V^2 R}{(R_{\text{sh}} + R)^2} = P_{\text{opt}} \quad \text{and} \quad \frac{V^2 R}{(R_{\text{sh}} + R)^2} = \frac{V^2 (R + \Delta R)}{(R_{\text{sh}} + R + \Delta R)^2} = \Delta P.
\] (5.43)

We can then move terms that do not depend on \( \Delta R \) to get

\[
\frac{R + \Delta R}{(R_{\text{sh}} + R + \Delta R)^2} = \frac{R}{(R_{\text{sh}} + R)^2} - \frac{\Delta P}{V^2} \equiv \gamma.
\] (5.44)

This is quadratic in \( \Delta R \) so can be found exactly for \( \Delta P \), also giving the fractional responsivity change \( \delta_s \) for \( \Delta P \) as

\[
\Delta R = \frac{-[2\gamma(R_{\text{sh}} + R) - 1] + \sqrt{1 - 4\gamma R_{\text{sh}}}}{2\gamma}
\] (5.45)

\[
\delta_s = \frac{[I_s(R_o - R_{\text{sh}})]^{-1} - [I_s(R_o + \Delta R - R_{\text{sh}})]^{-1}}{[I_s(R_o - R_{\text{sh}})]^{-1}}
\] (5.46)

We can apply this method to find \( \Delta R \) and the change in responsivity over the season using the inferred change in loading from the PWV. Fig. 5.7 shows this stability as a histogram.

\(^{22}\)This form was developed for the argument here and is only meant to be qualitative, but could be made more accurate with, for example, a slope that depends on loading, or a better junction between the normal and transition regions.

\(^{23}\)A direction of future work would be to track the optical loading from the atmosphere and track the operating resistance from TES bias steps, and then overplot these on Fig. 5.5 for \( \Delta R \) given \( \Delta P \).
The Instrumental Response to Radiation: Calibration

Figure 5.5: Left: The shift in operating point (as a fraction of $R_n$) as a function of optical loading. The traces marked as “model” are solutions for $\Delta R$ given $\Delta P$ from a DC electrothermal model that includes both Joule heating and the qualitative transition shape in Eq. 5.41. The traces marked as “analytic” are from a DC electrothermal model with only Joule heating. The solutions for $\Delta R$ given $\Delta P$ in the analytic case are summarized in Eq. 5.45. A model with only Joule heating cannot predict the saturation behavior, so we apply a simple saturation condition by truncating the new resistance to $R_n$ if it exceeds $R_n$. This is equivalent to there being sufficient optical power to drive the detectors normal. The qualitative effect of the transition model in Eq. 5.41 is to smooth out the junction between the transition and normal regions. There is no analogous saturation for negative $\Delta P$. In the limit that $R_{sh}$ is small, the Joule power can be made arbitrarily high, so a drop in power can be countered by an increase in Joule power. Finite shunt resistance bounds the total Joule power on the detector and so the electrothermal feedback fails to support decreases in power which lead to $R < R_{sh}$. Right: The change in resistance as a function of original operating point. For example, starting at an operating point of $0.5R_n$ (here $0.5$ on the x-axis), a 1 pW decrease in power will decrease the operating point to $\sim 0.4R_n$. Adding enough power has the effect of saturating detectors (here, up to $1R_n$) if they are high enough on the transition. A detector high on the transition will only experience a small change in resistance for a given change in power.

This procedure can also be used to flag detectors by identifying operating points that have drifted below the shunt resistance (where electrothermal feedback fails) or detectors that have been driven high enough on the transition that their responsivity is less stable and much lower than healthy detectors. The condition for saturation is that $R > R_n$, and the condition for low-bias instability (where $\Delta R$ has an imaginary part) is $R < R_{sh}$. For operating points with $R < R_{sh}$ the TES will become superconducting.\textsuperscript{24}

5.3.3 Errors in the time transfer standard using load curve responsivity

We will use the ratio of the responsivity estimate from load curves as a time transfer between biasing periods. The main uncertainty in this calculation is due to the shunt resistance, which is known to $\sim 10\%$. Here we give a simple estimate for the error in the time transfer standard from an error in the shunt resistors. High on the transition $> 0.7R_n$ and low on the transition $< 0.05R_n$, our estimate will be modified as the assumptions in Eq. 5.37 become less valid and

\textsuperscript{24}M. Niemack, private communication. Additional evidence suggests that these detectors may also oscillate.
5.3 Detectors – electrical response

Figure 5.6: Left: The fractional change in responsivity for a given change in absorbed power as a function of the initial operating point in units of $R_n$ based on the DC electrothermal model including Joule heating and the transition model in Eq. 5.41 and Eq. 5.42. The slowly-varying term is from finite shunt resistance (when the operating resistance becomes comparable) and the rapid “turn-offs” are due to nonlinear response high on the transition. Take the case of a drop in loading. Here, only if the detector is high on the transition (where responsivity is lower due to the low slope there) will the responsivity drastically improve as lower powers bring the detector on-transition. This is represented by the rapid rise on the right side. On the other hand, for sufficient increases in power, any detector can be forced normal. For example, an increase of 4 pW is sufficient to drive the detectors as low as $0.4 R_n$ to a region of decreased responsivity. This also indicates that detectors high on the transition (e.g., greater than $0.7 R_n$) will be driven normal in response to a planet $\sim 0.5$ pW. This effect is shown by the rapid turn-offs on the bottom half of the plot as detectors of a given operating point saturate. Right: The fractional change in responsivity estimated for the 145 GHz array using this method for a 1 pW drop in absorbed optical power. In the 2007 season, a less than $\sim 0.7$ mm decrease in PWV was typical, or roughly 0.3 pW. The median shift for 1 pW was $\sim 5\%$, so the median detector should change $\sim 1.5\%$ in a night, while extremal detectors might change by 6% or more.

$R_{sh}$ becomes comparable to the TES resistance. The responsivity from Eq. 5.37 can be rewritten as $dI/dP = -(R + R_{sh})/[V(R - R_{sh})]$ so that the ratio of responsivities in this ideal limit at two different times, here labelled with subscripts “1” and “2” for detector $i$ is

$$\frac{r_{1,2}}{dI/dP} = \frac{R_1 + R_{sh}}{V_1(R_1 - R_{sh})} \frac{V_2(R_2 - R_{sh})}{R_2 + R_{sh}} \approx \frac{V_2}{V_1} \frac{1 + x_1}{1 - x_1} \approx \frac{V_2}{V_1} [1 + 2(x_1 - x_2)].$$

(5.47)

where $x_1 = R_{sh}/R_1$, which we take to be small (this result will be modified lower on the transition). Then we can take $R_{sh} \rightarrow R_{sh} + \Delta R_{sh}$ or $x_1 \rightarrow x_1^+ \Delta$ and $r_{1,2} \rightarrow r_{1,2}^+$, to study the responsivity with respect to changes in $R_{sh}$. Both $V_1$ and $V_2$ are $R_{sh}I_{b,1}$ and $R_{sh}I_{b,2}$, so the dependence on $R_{sh}$ in the ratio cancels. Thus the fractional error in the ratio of the responsivities from some $\Delta R_{sh}$ is...
Figure 5.7: Here we estimate the change in responsivity (left) and operating resistance (right) following the initial TES biasing due to the change in loading using the model in Eq. 5.45. The loading is inferred from the atmospheric power model described in Sec. 4.3.3 and the change is with respect to the biasing conditions determined by the IV curve at the start of the night. Here, for each TOD we accumulate the number of detectors that fall into each bin. Because the TODs are 15 minutes long, each entry in the bin corresponds to 0.25 hours of integration time. This gives an effective single-detector integration time at that responsivity shift. We then divide this by the (approximately) 900 working detectors in the array to give the time that the array effectively observes at that responsivity shift. The data here represent $\sim 10^3$ TODs across ACT season 1 (2007). The effect of the PWV cuts is counter-intuitive. The bins with a decrease in sensitivity correspond to nights where the loading increased after the initial biasing; while the typical case (where the PWV falls) is represented by improvements in responsivity as the detectors move lower on the transition. The distribution is slightly bimodal because the PWV either increased or decreased, but typically did not stay constant. By cutting on the PWV at the time the file was taken, we can exclude the case where the PWV increased since the beginning of the night, but not the case where the PWV decreased. The PWV during a given file could be low even if the initial biasing happened at 5 mm. Thus a PWV cut removes many of the “core” observations (where the change in responsivity was $< 2\%$) and observations where the weather gets worse, but is ineffective at removing increases in responsivity from the typical case where the PWV drops. Even in the case without PWV cuts, 50% of detectors change in responsivity between $-0.7\%$ and $1.7\%$. Right: The fractional $\Delta R$ across all detectors across all nights. A larger number of detectors decrease in resistance because the PWV typically decreases from the start of the night.

(working to first order in $x$)

$$\frac{r_{1,2}^\Delta - r_{1,2}}{r_{1,2}} = \frac{[1 + 2(x_1^\Delta - x_2^\Delta)] - [1 + 2(x_1 - x_2)]}{1 + 2(x_1 - x_2)}$$

$$\approx 2[(x_1^\Delta - x_1) - (x_2^\Delta - x_2)] = 2\Delta R_{sh}(R_1^{-1} - R_2^{-1}). \quad (5.49)$$

Then let $R_2 = R_1 + \Delta R_{12}$, $\Delta R_{12}$ is the change in bias points between two nights. Then the fractional error in the ratio of the responsivities is

$$\frac{r_{1,2}^\Delta - r_{1,2}}{r_{1,2}} \approx 2\frac{\Delta R_{sh} \cdot \Delta R_{12}}{R_2^2}. \quad (5.50)$$

This is a statement that if the bias point does not change ($\Delta R_{12} = 0$), then an error in the shunt resistance does not impact the time transfer standard. In Sec. 5.8 we develop a statistic to study the array-wide and array-relative stability of system parameters. Fig. 5.25 of that section shows that the operating resistance that is established at the beginning of the night is consistent to $\sim 1\%$ from
5.4 Electronics – The transfer function of the readout electronics and flux lock loop

night to night, even if it may drift over the night. It is important to note that the ratio of responsivities here is much less sensitive to errors in $R_{sh}$ than the responsivity itself, which is biased directly by the error in $R_{sh}$. What is crucial is that the ratio of the operating point $R/R_{sh}$ is nearly constant because each night the detectors are rebiased to fix the median detector operating point in each biasing group, within measurement error. For a error in the shunt resistance measurement to impact the ratio of responsivities, it would have to be wrong by a factor of a few, while the real errors in $R_{sh}$ are probably $< 20\%$. We therefore believe the shunt resistor estimate error is negligible in the time transfer standard.

5.4 Electronics – The transfer function of the readout electronics and flux lock loop

5.4.1 The flux lock loop

The change in power incident on the detectors produces a change in the current in the TES circuit which modifies the magnetic flux $\Phi$ through the stage one SQUID loop (see Fig. 5.3). The output from the stage one SQUID is coupled into a stage 2 SQUID with all the detectors in a card’s “column.” Both the stage 1 and 2 SQUIDS are held at 0.3 K, and the output of the stage two SQUID is coupled to a SQUID series array (at 3 K) which amplifies the current response to a level that can be read by room-temperature electronics. The array is multiplexed in the time domain by independently addressing each stage one SQUID so that the same row across all the card columns can be read out by 32 series arrays in parallel. Each column within the row is then read at the same instant. Fig. 5.8 shows the timing of the multiplexed read operation.

![Figure 5.8: Timing diagram for the time-domain multiplexed read out.](image)

Based on the output of the series array at time $i$, a digital flux-lock loop then estimates the feedback necessary (at 15.15 kHz) to compensate the sky signal using the feedback inductor $L_{fb}$ at time $i + 1$. The lock loop is implemented in firmware on a field-programmable gate array (FPGA) in the readout electronics mounted to the camera. These electronics are known as the Multi-Channel Electronics (MCE), and were developed at UBC by M. Halpern’s group and the firmware in particular for ACT was developed by B. Burger and M. Amiri. The initial impetus for these
The Instrumental Response to Radiation: Calibration

linear, so closing the loop in this way “stiffens” the departure of the flux from its lock point to move over a narrow range of $V(\Phi)$ where the response is linear. The relation between power incident on the detector and feedback applied is linear and well-approximated by a transfer function.

The flux lock loop uses a proportional-integral-derivative (PID) controller to lock the error signal from the series array by applying a feedback to the coil with the output of a DAC. Here, the series array output is digitized by the ADC to give a 14-bit number $B_{\text{det}}(i)$, "the error". The feedback applied to the coil at step $i + 1$ is the sum of a proportional term $P$ times the error at step $i$, an integral term $I$ times the sum of errors up to step $i$, and a derivative term $D$ times change in error between $i$ and $i - 1$, or

$$B_{\text{fb}}(i + 1) = P \cdot B_{\text{er}}(i) + I \cdot \sum_{j=0}^{i} B_{\text{er}}(j) + D \cdot [B_{\text{er}}(i) - B_{\text{er}}(i - 1)] .$$ (5.51)

The integrator sums over the duration of the acquisition. So long as the loop is locked, it does not need to be reset; if a SQUID flux jump is detected, the loop has separate logic to reset the level. Here $B_{\text{fb}}(i)$ and $B_{\text{er}}(i)$ are the internal representation of the feedback and error in the FPGA, respectively. The stream of $B_{\text{fb}}(i)$ is divided down into a 14-bit number and applied by the feedback DAC, and also passed to the antialiasing filter. This set of filtering and division means that one hardware feedback DAC unit represents 1218 filtered feedback units that we write to disk. Thus when we refer to “digital feedback units” in this chapter, these actually refer to the digital output of the filter and not the actual bits that are set on the DAC, which are 1218 times coarser.

Even though it is implemented as a PID loop, the closest analogy to the flux lock loop is a multibit $\Sigma - \Delta$ converter. This is a common design for analog to digital conversion. One key difference between the ACT time domain multiplexing and standard $\Sigma - \Delta$ conversion is that the integrator and the digitizer are swapped (see Fig. 5.9). The reason for this is that a single ADC handles multiple (multiplexed) channels and it is easier to build in multiple digital integrators than it is to multiplex across analog integrators.

In Appendix B.4 we analyze the lock loop with only an integral term in some detail and show that it can be reduced to

$$B_{\text{fb}}(i + 1) = \tilde{I}_{\text{fb}} q_{s} (B_{\text{det}}(i) - q_{a} (B_{\text{fb}}(i))) + B_{\text{fb}}(i) .$$ (5.52)

Where $\tilde{I}_{\text{fb}} = IN_s C_{fb} DAC_{-er} ADC / 2^s$ is an “effective” integral term which is the integral multiplier applied in the FPGA ($I$) times the number of coadded samples ($N_s$) times the SQUID chain slope $C_{fb} DAC_{-er} ADC$ divided by the $2^s$, a bit shift that converts the internal representation of the feedback to a 14-bit number that is sent to the DAC. The functions $q_{a}$ and $q_{s}$ represent the quantization electronics was SCUBA II.

28Digital to analog converter, AD9744 (14-Bit, 210 MSPS).
29Analog to digital converter, AD6644 (14-Bit, 65 MSPS).
30For an analysis of $\Sigma - \Delta$ converters and related discussion, see Gray (1990, 1989, 1987); Gray and Stockham (1993); Chou and Gray (1991); Chen and Xu (2006), and Delta-Sigma Data Converters (Ed. Norsworthy, R. Schreier and G. Temes)
5.4 Electronics – The transfer function of the readout electronics and flux lock loop

Figure 5.9: The standard multibit $Σ − Δ$ converter (top) compared to the multiplexed multibit converter such as used by the multi-channel electronics, the “MCE” (bottom). The integrator and quantizer are swapped to facilitate multiplexing. The filtered output has $3.4 \, \mu K$ per digital unit, and the digital units in the lock loop’s 14-bit feedback DAC are $4.2 \, mK$. The lock loop samples at $15.15 \, kHz$.

If we ignore the quantization and simplify the notation,

$$f_{i+1} = \tilde{I} x_i + (1 - \tilde{I}) f_i.$$  \hfill (5.53)

This is a recursive low-pass filter form and its transfer function is shown in Fig. 5.10.\(^{32}\) Here $x_i$ is the signal at time $i$ and $f_i$ is the feedback applied to at step $i$. For $I = 1$ in these units, $f_{i+1} = x_i$ and the signal is recovered. The bandwidth of the loop is set by the integral term, so by applying an integral term that is too low, the lock loop can impinge not only on the anti-aliasing step in the ADC and DAC.\(^{31}\) The DAC step is not a quantization operation in the sense of converting a real number to a digital word, but it performs a bit shift division, which is lossy. The quantization of the error signal is finer than the feedback quantization and can be effectively ignored. Also, the noise level is sufficient to dither several feedback bits so that the mean of the output is not biased by the coarseness of the feedback quantization. A quick argument for the quantization noise is to assume it is white across the full flux-locked loop bandwidth. Then, if the feedback DAC LSB is $3.43 \, \mu K \cdot 1218 \approx 4.2 \, mK$, $\sqrt{(4.2 \, mK)^2 / 12} \approx 1.2 \, mK / \sqrt{Hz}$, or 1% of the total noise. The feedback DAC also has differential non-linearity at the typical level of 0.3 LSB, meaning that the step sizes are not uniform in feedback, so even with dithering, the output will be biased. One factor that mitigates this is that several bits are exercised and if the nonlinearity is independently distributed between these levels, then the bias is reduced as $0.3 / \sqrt{N}$ LSB for $N$ feedback DAC levels exercised. If we take the noise bandwidth before the antialiasing filter to be $10kHz$ and a rough noise of $1000 \, \mu K / \sqrt{Hz}$, the rms is $\sim 25$ LSB.

Further study is needed to understand the interaction of the flux lock loop noise, $1/f$ drifts, and quantization errors.

\(^{31}\) The DAC step is not a quantization operation.\(^{32}\) This form for the $Σ − Δ$ can be derived for some error $\epsilon_i$ and feedback $f_i$ as

$$f_{i+1} = \tilde{I} \sum_{j=0}^{i} \epsilon_j \quad (5.54)$$

$$= \tilde{I} \sum_{j=0}^{i} q_s (x_j - q_a (f_j)) \quad (5.55)$$

$$= \tilde{I} q_s (x_i - q_a (f_i)) + f_i, \quad (5.56)$$
filter (~122 Hz) described in Sec. 5.4.2, but also the detector time constants. For values of \( I \) greater than 1, there is moderate enhancement at high frequency (several kHz). At the Nyquist frequency, for \( I = 2 - \epsilon \), the transfer function scales as \( 2/\epsilon \). This is rolled off effectively by the 4-pole anti-aliasing filter – by a factor of \( > 10^6 \) for frequencies that might be affected by a high integral term. Thus, so long as the flux loop is stable, it should not add high frequency noise. The lower bound on the integral term comes from the consideration that the flux loop bandwidth should exceed the antialiasing bandwidth by some margin; in these units, this is 0.05. Between this range \( 2 > I > 0.05 \), the system response is less sensitive to the particular value of the integral term because the multiplexing rate is much higher than the downsampled read rate.

Figure 5.10: Magnitude of the transfer function of the feedback loop \( f_{n+1} = Ix_n + (1 - I)f_n \). In practice, the integral term here depends on the slope of the SQUID output chain, and we do not adaptively pick an integral multiplier in the hardware based on these slopes. Because the bandwidth is limited at roughly < 122 Hz by the antialiasing filter, this gives significant latitude for choosing the integral multiplier in hardware. We used a value for the integral multiplier that was half what drove the loop into oscillations and did not find that this impacted either the time constants or the lock loop stability, but further work is needed to understand if there are small effects. The anti-aliasing filter strongly suppresses frequencies above 122 Hz and is shown in Fig. 5.11.

5.4.2 The antialiasing filter

As described in the systems chapter, Sec. 3.2.2, the 15.15 kHz signal from the multiplexer is downsampled to a more practical data rate after applying an anti-aliasing filter. Based on data rate needed to Nyquist-sample the beam in the 280 GHz array, we chose a 399 Hz data rate with \( f_{\text{MB}} = 122 \) Hz in a cascaded biquad anti-aliasing filter with four poles, or

\[
H(z) = H_1(z)H_2(z) = \left( \frac{1 + 2z^{-1} + z^{-2}}{1 + b_{1,1}z^{-1} + b_{1,2}z^{-2}} \right) \left( \frac{1 + 2z^{-1} + z^{-2}}{1 + b_{2,1}z^{-1} + b_{2,2}z^{-2}} \right).
\]

(5.57)

The effective integral term depends on the SQUID chain slope, which has significant scatter, thus some of the scatter in the time constant in Fig. 5.25 may be due to changes in this slope and an integral term that is too low.
5.5 The diffuse calibrator

Here \( z = e^{i\omega} \) and \( z^{-1} \) acts as an operator which delays by one digital record. The filter coefficients are \( b_{1,1} = -2 \cdot 32092/2^{15} \), \( b_{1,2} = 2 \cdot 15750/2^{15} \), \( b_{2,1} = -2 \cdot 31238/2^{15} \), and \( b_{2,2} = 2 \cdot 14895/2^{15} \). The magnitude and phase response of the filter is shown in Fig. 5.11. The output of the feedback DAC at \( \omega = 0 \) is multiplied by a gain of 1217.9. This is not strictly related to the calibration, because (so long as it is constant) it is automatically included in the flux calibration from the planets. However, to understand load curves or bias steps, (or any data taken without the filter applied) this gain is essential for finding the correct power.

![MCE digital filter response](image)

Figure 5.11: Response of the digital anti-aliasing filter in the multi-channel readout electronics. Incoming 15.15 kHz data from the detectors are filtered, and then downsampled to the output 398.7 Hz, or every 38th output from the digital filter. The DC gain is 1217.9, and must be accounted for in going from filtered feedback units to DAC device units. The phase response is not flat over the frequencies where the sky signal will be modulated. This must be accounted for by an inverse filter in the analysis. Planets excite response up to \( \sim 100 \) Hz so the phase and magnitude relation from the filter must be undone to find accurate amplitudes.

5.5 The diffuse calibrator

The atmosphere can be used to find the array-relative (between detectors) calibration. The principle is that the atmosphere presents a large array-common signal and detectors will respond differently to this signal depending on their sensitivity to diffuse radiation. This can be used to “flat field” response to diffuse radiation. Unlike planet observations, it is not modulated by a dilution factor and it does not require moving to a new pointing. It also has high signal to noise which allows it to achieve \( \sim 1\% \) accuracy in the relative calibration. There are several sources of non-optical noise to control against: 1) the 3 K stage series array drift (Sec. 4.1), 2) scan-synchronous pickup of the magnetic field on the SQUIDs, and 3) electrical interference. Terms 2 and 3 are described in more detail in Chapter 6 along with diagnostic tools. The 3 K drift is slow because 34The filter was designed for the ACT sampling rate by M. Amiri and B. Burger and is implemented in the multi-channel electronics firmware.
of the cryogenic time constant, so is effectively removed by applying a high-pass filter at 0.05 Hz. We avoid scan-synchronous pickup by using stare data, which may also simplify the interpretation of the atmosphere. We also pick data with low electrical noise for these studies.\textsuperscript{35} There is little information in the atmosphere above \( \sim \text{few} \) Hz, so it is economical here to smooth and downsample to a 1.6 Hz data rate in addition to the 0.05 Hz high-pass.\textsuperscript{36} Calibration emitter pulses mix the uniform illumination of the atmosphere with the particular illumination pattern of the calibrator (see Fig. 5.23), so these are masked out by connecting a line to the baseline on either side of the pulse.

Fig. 5.12 illustrates the method with \( \sim 1 \) hour of stare data from Dec. 10th, 2007.\textsuperscript{37} Here we find the amplitude as the least-squared best scaling to the array-common drift (scaling to a single central detector time stream without jumps gives similar results). We then evaluate the scatter by comparing the relative calibration from several different files and find that percent-level and possibly sub-percent relative calibration is achievable. We can then compare the difference in the relative calibration achieved by the atmosphere between two nights, shown in Fig. 5.13. Applying the ratio of the responsivity from load curves between the two nights helps reduce some of the structure in this difference, but the main residual structure apparent here is common to rows and is almost certainly due to row-correlated burst noise (electrical interference) in these two intervals. This is described further in Sec. 6.3.

In addition to the array-common component of the atmosphere, the atmosphere also has components at higher temporal frequencies that are not common across the array. To see this, we find the scaling of all detectors that best match column 15, row 15 and then subtract the scaled signal. Fig. 5.14 shows the residuals. Most of the relative calibration is from the array-common component of the drift power with low temporal/spatial frequencies. Once the detectors are scaled and this slow common term is subtracted off, we interpret the higher-frequency residuals as the sub-array atmospheric structure. The central region is coherent with column 15 row 15 because the beams from those detectors are largely overlapping through the first \( \sim 1000 \) m of atmosphere. At wider separation (\( \sim 12' \)), the beams overlap less and are able to resolve temporal structure in the atmosphere.\textsuperscript{38} One detector length along a column (which is parallel to the horizon) is \( (24 \text{ cm})/\sin(\text{Alt}) \) and along a row is \( (21 \text{ cm})/\sin^2(\text{Alt}) \) on a layer at 1000 m, where the difference in the length has to do with the plate scale in the column vs. row directions. The aspect ratio of column to row

\textsuperscript{35}Electrical noise was found to increase the uncertainty in the calibration from the 1\% level here to several percent. The amplitude of the interference varies widely in time, so care must be taken in deriving a relative calibration from bad periods. This is described in Sec. 6.3, and varies from detector to detector. We have not performed a season-wide relative calibration, but suspect that some periods will have more scatter and require noise treatments to reach percent-level accuracy.

\textsuperscript{36}This procedure could be improved by developing a reliable algorithm to estimate the relative calibration over each 15-minute TOD interval. Making this estimate reliably is difficult because of the variety of conditions it depends upon: electrical noise conditions, variable non-optical signals, the atmospheric power and level of fluctuations across the array, and the effect of scanning. Significant care would have to be taken to ensure that this actually improves the array-relative calibration and does not simply add noise. Here we advocate using one or two well-controlled relative calibrations from stare data, then exploit the fact that the array-relative calibration is stable at the percent-level (or better), and that the small variability can be corrected using Eq. 5.45. The PWV on the two nights studied here was \( \sim 0.5 \) mm. Applying this method as-is to science scan data without any quality checks on that data produced a relative calibration with as much as 10% scatter.

\textsuperscript{37}There was no official program of stare data during the 2007 season, and these were acquired during motion faults.

\textsuperscript{38}An outcome of this is that the array-wide common mode confuses the actual atmospheric structure – averages across the array will be biased by central detectors that have more in common with other detectors in the average. The array-common mode then poorly describes detectors at the edges of the array which see less in common with the others. An array-common mode subtraction will leave significant contamination. One alternative is to find the common mode across sub-arrays whose detectors overlap significantly on the turbulent layer, or to use a kernel which optimally weights detectors by the extent of their overlap on the array to find an atmospheric signal estimate for each detector. If the atmosphere is well-approximated by a single, frozen turbulent layer that moves across the field of view, the best separation is to map it in parallel with the atmospheric sky. Yet, the only assumption in the weighting kernel method is that proximal detectors do not resolve turbulent structure, and should be true given the turbulent layer heights found by interferometric studies, Robson et al. (2001).
5.5 The diffuse calibrator

Figure 5.12: Relative calibration (left) and scatter (right) inferred from diffuse atmospheric loading drift over one hour of stare data on Dec. 10, 2007. The left figure represents the relative response in digital units to a diffuse source. This has been normalized so that the average across the array is unity, and the relative response to the diffuse source varies from ~ 0.7 to ~ 1.2 times that figure. This represents the relative response to diffuse radiation in digital units, so has structure that depends on the electrical properties of the detectors (such as the responsivity, which is strongly correlated along columns) in addition to the optical coupling. The scatter of the relative calibration inferred from these data was sub-percent for many of the detectors, increasing toward the edges to percent-level. (The quantities here represent the fractional 1σ scatter across several stare segments. Column 22 is dead and column 31 was saturated during much of the 2007 season.)

The power spectrum and structure of the atmosphere will be the subject of future work by the ACT collaboration. We can use the geometry here to reach some quick conclusions about the atmospheric layer. If we take the illumination of the sky by one detector to be a uniform disk with diameter 5.8 m, the overlap of two disks will drop from 1 at zero separation to 0 at a disk-center separation of 5.8 m. The drop is linear to a reasonable approximation for uniform disks. Zero overlap corresponds to a separation of roughly > 20 detector units for a screen at 1000 m altitude viewed at 50.5° elevation. This is consistent with the scales in Fig. 5.14. Once the beams no longer overlap, the residual power is an indicator of the power spectrum of the atmosphere. A separation of one column length produces a larger separation than one row length by the aspect ratio \( \sin(\text{Alt}) \). This is also consistent with some of the residual structure in Fig. 5.14, but more work needs to be done.

See Sec. 4.4 and Consortini and O'Donnell (1993); Bussmann et al. (2005); Lay and Halverson (2000); Church (1995).
The Instrumental Response to Radiation: Calibration

Figure 5.13: Left: The ratio of the relative calibration on Dec. 4 and Dec. 10, 2007. The units are terms of the fractional change of the relative calibration between the two dates, and range is over ±2%, while the relative calibration of most detectors changes less than 1%. The vertical lines across common rows are due to correlated noise, which has not been removed in this analysis. Middle: the ratio of responsivities as predicted by load curves at the beginning of the night. The units here are the fractional change in the relative response expected from the load curves taken at the beginning of each night. Right: Applying the correction from load curves to the ratio of relative responsivities determined from the atmosphere. This accounts for and removes some of the structure along common columns. In this case, between these two nights, the relative responsivity was stable to better than a percent for most detectors.

5.6 Observations of the planets

Observations of the planets can be used as a flux calibration because their disk-averaged temperatures are relatively well-known. The distribution of planet observations during the 2007 season of ACT is shown in Fig. 5.15. Because Mars was observed at transit, so only has partial coverage across the array, Saturn serves as our primary flux calibration in the 2007 season and has ~ 1 observation per night. (A second observation was in the morning once the sun was up was not useful for calibration because of deformations in the primary mirror from solar heating.) There were no observations of Jupiter during the reference part of the season once all optics were aligned.

In this section, we describe several considerations for interpreting planet observations. Unlike the diffuse calibration in Sec. 5.5, the planets do not fill the beam, yet they are also not truly point sources. Sec. 5.6.1 describes the dilution factor and considerations for interpreting the solid angle from planet observations. The solid angle is expected to vary across the array, and with it, the dilution factor and planet amplitudes. We estimate this effect by applying the relative calibration from the atmosphere in Sec. 5.5.

5.6.1 Interpreting the solid angle from planet observations

Using the dilution factor from Eq. 5.9 and Eq. 5.5/5.6, the power received when the source is straight-on is

\[ P(0) = \int A_c \sigma^S \Psi(\hat{n}) \Psi^S(\hat{n}) d\Omega_n = \sigma^S D A_c \Omega_b = D \int A_c \sigma^S \Psi(\hat{n}) d\Omega_n. \]  (5.58)

---

40 For ACT, Saturn’s disk was ~ 6 nsr in the 2007 season compared to ~ 225 nsr for the main beam.
Figure 5.14: The residual rms across the array from fits to the diffuse atmosphere. Here, each detector is scaled to match the response of column 15 row 15 at the center of the array. The figure plotted here is the standard deviation of the residual from this scaling divided by the standard deviation of the atmospheric signal estimated in column 15, row 15. All channels have been high-pass filtered (Butterworth, order 1) above 0.05 Hz, and downsampled to 1.6 Hz after being smoothed, so the amplitudes represent the total residual power over those frequencies. These show that the atmospheric pickup is coherent over $\sim 10 \times 10$ blocks of detectors, or inside of radii $\sim 10^\prime$. Left: drift data from stare data on Dec. 10, 2007 (03 UT) acquired at $42^\circ$ elevation. The aspect ratio is consistent with power in thin sheet viewed at the $42^\circ$ elevation. Right: drift data from stare data on Dec. 4 (24 UT) at $52^\circ$ elevation. Delgado and Nyman (2001) find a thick turbulent layer during the day that does not resolve until late night, so the lack of an aspect ratio may be attributable to a much thicker layer. Another approach is to find the zero lag correlation as a function of detector separation. This is similar to binning this result in radius from column 15, row 15, but across all detectors. Fig. 6.6 shows this, and the conclusion is similar.
Figure 5.15: Tally of the planetary sources observed by the 145 GHz camera in ACT season 1. Nearly every night has at least one planet observation. We have found that Uranus and Neptune were weak calibrators compared to Saturn and Mars, but were used to determine telescope pointing. Nights with two Saturn observations were typically in the early morning (~3 AM local) and after sunrise. While the source observations after sunrise were useful to study the optical deformation, they could not be used as calibration sources. Mars was observed in an azimuth scan at transit, so only a common sub-array was illuminated, with the exception of several observations where the scan was intentionally offset.

Then the integral of the power over all angles (relevant for the solid angle) is

$$
\int P(\Delta \hat{n})d\Omega_{\Delta \hat{n}} = \bar{A}_e \sigma^S \int \Psi(R_{\Delta \hat{n}}\hat{n})\Psi^S(\hat{n})d\Omega_{\Delta \hat{n}} = \sigma^S \Omega_{p} \bar{A}_e \Omega_{b},
$$

where the solid angle of the planetary source is $\Omega_{p}$ and we have integrated over $\Delta \hat{n}$ first, which gives $\bar{\Omega}_{b}$, regardless of $\hat{n}$. Thus, the solid angle measured by a planet observation is

$$
\bar{\Omega}_{b} = \frac{1}{P(0)} \int \hat{P}(\Delta \hat{n})d\Omega_{\Delta \hat{n}} = \frac{\Omega_{p} \Omega_{b}}{\int \hat{\Psi}(\hat{n})\Psi^S(\hat{n})d\Omega_{p}}
$$

(5.60)

The wideband solid angle of the telescope $\Omega_{b}$ is different than the solid angle $\bar{\Omega}_{b}$ inferred from a planet through $P(\Delta \hat{n})$ when the solid angle of the planet becomes comparable to the telescope solid angle. The dilution factor can be calculated explicitly to relate $\Omega_{b}$ to $\bar{\Omega}_{b}$. For a Gaussian main beam (width $\sigma_{b}$) and a Gaussian surface brightness (width $\sigma_{s}$),

$$
D = \frac{1}{\Omega_{b}} \int \exp \left\{ -\frac{1}{2} \theta^2 \left( \frac{1}{\sigma_{b}^2} + \frac{1}{\sigma_{s}^2} \right) \right\} 2\pi \theta d\theta
$$

$$
= \frac{2\pi \sigma_{b}^2 \sigma_{s}^2}{\Omega_{p} \sigma_{b}^2 + \sigma_{s}^2} = \frac{\Omega_{p}}{\Omega_{p} + \Omega_{b}}.
$$

(5.61)

Thus, in this case, the measured beam solid angle is simply given by $\bar{\Omega}_{b} = \Omega_{b} + \Omega_{p}$, which is consistent with a Gaussian main beam convolved by some Gaussian source. In this case, the known source solid angle $\Omega_{p}$ can be subtracted from $\bar{\Omega}_{b}$ to give an estimate of the telescope beam solid angle $\bar{\Omega}_{b}$. A real planet will not have a Gaussian emission profile, and the surface brightness will depend on the phase function of the regolith/atmosphere41 and limb darkening. A reasonable

41The phase function is the intensity of emitted radiation at an angle relative to normal, $I(\theta)/I(0)$.
5.6 Observations of the planets

approximation is the fully-illuminated disk, where

\[ D \equiv \frac{1}{\Omega_b} \int_{0}^{\theta_s/2} \exp \left\{ -\frac{1}{2} \frac{\theta^2}{\sigma_b^2} \right\} 2\pi \theta d\theta \]

\[ = 1 - \exp \left\{ - \frac{\Omega_p}{\Omega_b} \right\}. \]  

(5.62)

In all cases, the dilution factor becomes unity when the source is much larger than the beam (diffuse), and \( \Omega_p/\Omega_b \) when the source is much smaller than the beam (giving \( \Omega_b = \Omega_b \)). It is only in the regime where \( \Omega_p \sim \Omega_b \) that the surface brightness and radiation pattern impact how the two mix. See Fig. 5.16 for a comparison of the disk and Gaussian.

Figure 5.16: Dependence of solid angle on the spatial distribution and size of the source. When the solid angle of a source is comparable to the main beam size, the effective beam width and dilution factor for the planet depend in detail on the spatial distribution of surface brightness. Here we compare the fractional difference of the dilution factor assuming a fully-illuminated disk (neglect e.g., limb darkening) and a Gaussian disk illumination as a function of \( \bar{\Omega}_b/\Omega_p \). We choose the Gaussian case in particular because this is the one that corresponds to assuming the main beam solid angle is the inferred solid angle in response to a planet minus the solid angle of the planet. Here, in both the diffuse limit (where \( \bar{\Omega}_b/\Omega_p \) is small) and the point source limit (where \( \bar{\Omega}_b/\Omega_p \) is large) the effect of the source’s spatial distribution is negligible. For intermediate sizes the dilution is significantly different. Even if the beam is 10 times larger than the source solid angle, the dilution is \( \sim 5\% \) different whether one assumes a disk or Gaussian source. The spatial distribution of planetary sources is necessary to understand the main beam response, but for calibration it suffices to know only the solid angle of the planet and the solid angle of the main beam inferred from observations of that planet, and no further modeling is needed. For the 145 GHz camera, Saturn has \( \bar{\Omega}_b/\Omega_p \sim 30 \), while Jupiter can have \( \bar{\Omega}_b/\Omega_p \) as low as 6 during some times of the year. Solid angles inferred from Jupiter then depend in detail on the surface brightness (and possibly moon positions).
5.6.2 The relation between planet temperature and absorbed power

The power when viewing the planet straight-on \(P(0)\) can be used to calibrate the sensitivity to changes \(\Delta T\) in the CMB. For planetary sources, the Rayleigh-Jeans temperature is more convenient than the spectral intensity \(\sigma^S(\nu)\) and they are related by \(\sigma^S(\nu) = 2\nu^2k_BT_{\text{RJ}}(\nu)/c^2\). Then

\[
P(0) = \mathcal{D} \int \tilde{A}_e \tilde{d}^S \tilde{\Psi}(\hat{n})d\Omega_n
= \mathcal{D} \int A_e(\nu) \Psi_e(\hat{n}) \sigma^S(\nu) \eta(\nu) g_B(\nu) e^{-\tau(\nu)} d\nu d\Omega_n
= \mathcal{D} \int_0^\infty m(\nu) k_B T_{\text{RJ}}(\nu) \eta(\nu) g_B(\nu) e^{-\tau(\nu)} d\nu, \tag{5.63}
\]

where we have used Eq. 5.6 in the second equality and in the third equality, we integrate over angles first to give a frequency-dependent solid angle \(\Omega_n(\nu)\). We have also replaced the frequency-dependent throughput as \(A_e(\nu) \Omega_n(\nu) = m(\nu)c^2/(2\nu^2)\) where \(m(\nu)\) is the number of radiation modes received. As in the CMB case, we can apply the approximation for the optical depth described in Sec. 4.1, letting \(\exp[-\tau(\nu, t)] \approx \exp[-\tau(t)] g_A(\nu)\) so that,

\[
P(0) = 2\mathcal{D} e^{-\tau} \int_0^\infty m(\nu) k_B T_{\text{RJ}}(\nu) \eta(\nu) g_B(\nu) g_A(\nu) d\nu. \tag{5.64}
\]

We can use Eq. 5.60 for \(\mathcal{D}\). Note that if the extent of the planet is significant compared to the telescope solid angle, then it is important to use the estimate of the beam solid angle \(\Omega_b\) from that planet, as \(\mathcal{D} = \Omega_p/\Omega_b\). This dilution factor for the power automatically accounts for the finite size of the planet. Another consideration in this integral is that the planet temperature across the passband \(T_{\text{RJ}}(\nu)\) can have structure if there are molecular resonances. Sec. 5.6.3 gives an outline of this and Appendix C.2 describes planet emission in millimeter wavelengths in more detail. A third consideration is that the telescope only measures temperature differences between the planet and its background \((T_{\text{RJ},b}(\nu),\) where subscript \(b\) denotes background), through the power difference

\[
P(0) = e^{-\tau} \frac{\Omega_p}{\Omega_b} \int_0^\infty m(\nu) k_B [T_{\text{RJ}}(\nu) - T_{\text{RJ},b}(\nu)] \eta(\nu) g_B(\nu) g_A(\nu) d\nu. \tag{5.65}
\]

The background temperature out of the galactic plane is roughly the CMB temperature, which is small compared to planet temperatures, so the background can be ignored most of the time.\(^{42}\) The total power is roughly \(k_B T_{\text{RJ}} \Delta \nu\) (where \(\Delta \nu\) is the bandwidth) for each radiation mode attenuated by the atmosphere and efficiency factors so that

\[
P(0) \sim \eta e^{-\tau} \frac{\Omega_p}{\Omega_b} m k_B T_{\text{RJ}} \Delta \nu, \tag{5.66}
\]

when the efficiency and mode number are constant across the band.

Define a band-averaged temperature \(^{43}\)

\[
T_{\text{RJ}} = \frac{\int_0^\infty m(\nu) [T_{\text{RJ}}(\nu) - T_{\text{RJ},b}(\nu)] \eta(\nu) g_B(\nu) g_A(\nu) d\nu}{\int_0^\infty m(\nu) \eta(\nu) g_B(\nu) g_A(\nu) d\nu} \tag{5.67}
\]

\(^{42}\)The CMB temperature is not small compared to the diluted planet temperatures, which are typically a few \(K\), but \(T_{\text{RJ}}(\nu) - T_{\text{RJ},b}(\nu)\) is geometrically the correct factor, which is a small correction to the planet temperature in most cases.

\(^{43}\)The effective frequency is typically close to the bandpass center because a linear function is an odd function across the passband, and the passband about its center is roughly an even function.
5.6 Observations of the planets

so that the power is simply the product of the dilution, optical loss, average temperature and bandwidth-efficiency product (Eq. 5.23)

\[ P(0) = e^{-\frac{\Omega_p}{\Omega_b} k_B T_{RJ} \bar{B}}. \]  

(5.68)

To make this practical, we need apply a few considerations. Let \( R(t) \) be the responsivity that converts this power into digital (feedback DAC) units at some time \( t \), as \( D(t) = R(t)P(t) \). Second, calibration is a statement about the response of the system at zero temporal frequency, while planets excite response across roughly \( \sim 10 \) ms time scales. Because the detectors have a finite time constant, an uncorrected estimate for the planet amplitude will be biased low.\(^{44}\) The extent of the bias depends on how the amplitude is estimated because the time constant preferentially broadens the beam in the scan direction – if the fit assumes the same beam width along the scan and Earth-drift directions\(^{45}\), then it will reach a different answer than if it had assumed that the drift and scan direction could have independent width. Rather than develop a correction to the amplitude estimate, the easiest is to undo the response time in the time domain by an inverse filter and estimate of the time constants. Fig. 5.17 shows the difference of two independent methods where one corrects the optical response and the other does not. Here, the time constants used are from TES bias steps, measured by J. Appel and calibrated to the optical response time constant studies by multiplying by the factor of 1.78 developed in Niemack (2008). Let \( \Theta^{-1}(f_{3dB}) \) describe the correction factor for the amplitudes due to the optical time constants. The final consideration is that the efficiency \( \eta \), optical time constant \( f_{3dB} \) and wideband solid angle \( \Omega_b \) can all vary from detector to detector due to their separation from the AR coupling layer, fabrication details, and position in the focal plane.\(^{46}\) Denote the estimate of the solid angle, efficiency, and suppression due to time constants for a particular detector \( i \) as \( \bar{\Omega}_b, \bar{\eta}, \) and \( \Theta(f_{3dB,i}) \), respectively. With these considerations, the amplitude of the peak response in digital counts at some time \( t_p \) is

\[ D_i \bigg|_{\text{peak}} = e^{-\tau(t_p)} R(t) \Theta(f_{3dB,i}) \bar{B} k_B T_{\text{planet},RJ}(t_p) \left( \frac{\Omega_p}{\Omega_b} \frac{\bar{\Omega}_b}{\bar{\eta} \Omega_{b,i}} \right). \]  

(5.69)

Many of the quantities in the calibration problem have time dependence. The solid angle of the planet \( \bar{\Omega}_p \) is given in ephemerides\(^{47}\) and several detailed models give \( T_{RJ}(t_o) \) for the planet based on the ring inclination, heliocentric distance and viewing phase. (See Appendix C.2.) Independent measurements of the PWV from a dipping radiometer at the APEX site can be used to determine the optical depth \( \tau(t) \). Finally, in Sec. 5.3, we have described how the responsivity can vary with time. Sec. 5.7 uses the calibration emitter to study responsivity variations and their treatment and Sec. 5.9 describes the global calibration.

The most difficult of the factors in Eq. 5.69 to determine empirically is the solid angle of each detector. In Sec. 5.6.5 we estimate this by taking the ratio of planetary and diffuse array-relative response. The findings there are consistent with \( \sim 15\% \) fractional shift in the solid angle across the array from physical optics simulations\(^{48}\), but the structure across the array is more complex.

\(^{44}\)This does not change the actual beam dilution factor, but it does change our estimate of \( \bar{\Omega}_b \), so must be corrected. A treatment of the blurring specific to CMB observations can be found in Niemack (2008) and Hanany et al. (1998).

\(^{45}\)Here we take “Earth-drift direction” to be the direction in the map that corresponds to the Earth’s rotation.

\(^{46}\)In principle, the passband may also vary as detectors integrate over a rays of slightly different angles as they pass through the bandpass filter, which may have an angle-dependent passband. The bandpass filter stack position was designed so that the rays that pass through it are nearly parallel, so this is a less significant consideration than, for example, the variation of the solid angle across the array or the passband as a function of position on the filter.

\(^{47}\)The Planetary Rings Node (http://pds-rings.seti.org/saturn/) ephemeris tool tracks ring inclinations and satellite positions. aephem (http://aephem.sourceforge.net/) is a new precision astrometry and ephemeris package developed by A. Hincks.

\(^{48}\)J. Fowler, private communication.
Throughout, we will assume that the main telescope beam is 225 nsr (nano-steradians) and is constant in time, but not necessarily constant across the array.\textsuperscript{49} Further studies are needed to understand main beam variation over the season and at different elevations. There are telescope deformations in the late morning times from non-uniform solar heating of the primary mirror.\textsuperscript{50} The rigidity of the optics as a function of elevation is not yet understood, but is not a significant concern for first-season data, where the science scans are at fixed elevation and there are several planet observations at those elevations, also. The ratio of a detector solid angle to the array-wide solid angle $\Omega_{b,q}(t)/\Omega_{b}(t)$ should be constant because it should depend most on the cold optics rather than the mirrors.

Three independent fits have been developed for the planet data to estimate the response amplitudes in digital units. One, which we use for most calculations here was developed by A. Hincks and uses a 1D fit to the Gaussian response for each scan through the planet. There is no assurance that one of these slices goes through the beam center, so to estimate an equivalent 2D Gaussian amplitude a second Gaussian is fit to the maxima of the azimuth slices (a slice in the Earth-drift direction). This method includes a correction for the detector time constants (see Fig. 5.17). The second method, developed by B. Reid, J. Fowler, and T. Marriage fits the response to the planet scan in the TOD as the equivalent 2D Gaussian over those segments of the TOD in the time domain that project onto a region around the planet in map space.\textsuperscript{51} A third method developed by A. Hincks is to make a map of the planet in each detector and fit an Airy disk to the main beam region.\textsuperscript{52}

### 5.6.3 Overview of effective emission temperatures of the planets at millimeter wavelengths

Appendix C.2 describes the emission of the planets in some detail, and here we summarize the main conclusions. This chapter uses Saturn as the primary flux calibration, and we describe it in detail next in Sec. 5.6.4. The Jovian planets (Jupiter, Saturn, Uranus, Neptune) are complex objects for calibration studies. Their emission depends on the properties of the upper atmosphere where rare species such as $\text{PH}_3$ can produce deep, broad absorption features. The atmospheres of the gas giants are also dynamic, causing well-documented changes in the radio emission over the period of years (see van der Tak et al. (1999); Gulkis et al. (1983); Kramer et al. (2008)).\textsuperscript{53}

Mars is a reasonable calibration object because of the tight correspondence between thermal models developed in Wright (2007b); Rudy (1987); Rudy et al. (1987) and recent WMAP observations (Hill et al. (2008)), up to the correction factor described there of 0.9. One exception to this is when there are global dust storms, which impact the RJ temperature in mm-wavelengths at the 10% level; see Gurwell et al. (2005). The emissivity of Mars is known relatively well and varies much less than the gas giants; see Rudy (1987) and Rudy et al. (1987).

The following observations should be avoided for calibration: 1) 220 GHz array observations of Neptune (CO $J = 2 - 1$ line at 231 GHz), 2) 280 GHz array observations of Saturn (PH$_3$ $J = 1 - 0$ at 267 GHz), 3) Jupiter in all three arrays (because of saturation), 4) observations of Mars during a

\textsuperscript{49}Newer estimated from Kavi Moodley suggest $(232 \pm 4)$ nsr at 1σ, but all the values here are derived assuming 225 nsr.

\textsuperscript{50}The planet amplitudes in the morning were nearly $1/2$ their values during the stable part of the night. This was roughly $<1/4$ of the 2007 season’s observation time. The telescope has a large, offset sun screen in subsequent seasons to mitigate this effect.

\textsuperscript{51}This method does not yet include a time constant correction.

\textsuperscript{52}Amplitudes from the map space are more difficult because the coverage of a single 400 Hz detector stream is not sufficient to build a complete sub-beam resolution map. If the beam center is not known and does not overlap with a hit pixel, then the amplitude of, for example, a 2D Airy fit will be biased by what it identifies as the maximum in the pixels that are hit.

\textsuperscript{53}For example, the 90 GHz emission of Uranus decreased $\sim 10\%$ in the last $\sim 20$ years (see Kramer et al. (2008)), so some care must be taken in using older brightness measurements.
Figure 5.17: The ratio of the estimated amplitudes for planet response between two methods, one with a correction for the time constant and one without. The x-axis gives the 3 dB point of the one-pole response from TES bias step time constant fits that have been scaled to be consistent with optically determined time constants as described in Niemack (2008). The points represent all observations of Saturn across ACT season 1. This is not the ratio of amplitudes from the same estimator. Here, one estimator (called Obsfit) is the 2D Gaussian fit developed by B. Reid, J. Fowler and T. Marriage here without a correction for time constants applied and the second model is the two-step fit developed by A. Hincks which fits a 1-dimensional Gaussian to each scan pass. To find the maximum, the maxima from each pass are then fit to a Gaussian (effectively in the Earth-drift direction). In the Hincks model the time constant correction is applied. In addition to showing that the time constants are an important part of the model, it also shows that the amplitude between the two estimators is consistent to several percent. The outliers are due to erroneous observations that are treated differently between the two estimators. (These are cut or rejected in the calibration analysis) A rule of thumb for the correction between the two models is \( \Theta(f_{3dB}, i) = 1.2 \exp(-f_{3dB,i}/20) + 1.02 \) (shown as the dashed line here). This indicates that for detectors with \( f_{3dB,i} > 70 \) Hz, the impact of a time constant on the amplitude of the response to the planet is small (< 4%), and that the results from the Hincks fit are roughly 2% higher than the Obsfit, which in turn suggests some beam asymmetry (Obsfit here was configured to use a symmetric Gaussian, while the Hincks fit uses a separate slice along the scan and Earth-drift direction).

global storm. In the case of the resonances, a good calibration may be possible with an accurate passband and resonance model (such as Weisstein (1996) for \( PH_3 J = 1 - 0 \) and discussion in Marten et al. (2005) for Neptune). Fig. 5.18 shows the diluted planet temperature for estimated ACT main beam solid angles.

### 5.6.4 The Saturn mm-wave emission model

Fig. 5.19 summarizes recent measurements of Saturn as a function of frequency alongside atmospheric models.\(^{54}\) The main feature in the emission of Saturn’s disk that is relevant for ACT is a \( \sim 20 \) GHz-wide \( PH_3 J = 1 - 0 \) resonance at 267 GHz where the temperature drops from a

\(^{54}\)For reviews of Saturn’s disk temperature, see Gautier et al. (1977); Weisstein and Serabyn (1994); Dunn et al. (2005); van der Tak et al. (1999); Dowling et al. (1987); Grossman et al. (1989); Fiasar et al. (2005); Dunn et al. (2002).
The Instrumental Response to Radiation: Calibration

Temperatures of the Planets (Diluted RJ)

Figure 5.18: Temperatures of the planets in the three ACT bands ("dec" denotes 145 GHz, "null" denotes 220 GHz, "inc" denotes 280 GHz) diluted by the ratio of the planet solid angle by the estimated beam solid angle. We use 225 nsr from A. Hincks’ estimates from first-season data, and then scalings to the other bands from Niemack (2008) for beam widths 1.38’, 1.03’, and 0.88’. (Dilution here refers to the ratio of the planet solid angle divided by the 225 nsr main beam.) The temperatures used here are combinations of several values in the literature. For Jupiter, we use Goldin et al. (1997) (their band 1 for 145 GHz/220 GHz and band 2 for 280 GHz). For Saturn, we use Hill et al. (2008) and the geometric model developed in Appendix C.2 corrected for PH3 in the 280 GHz band by the ratio of bands 2 and 1 in Goldin et al. (1997), who had similar bandwidths, to get a very rough estimate of the effect. Mars temperatures are based on Wright (2007b) corrected by the factor of 0.9 given in Hill et al. (2008). We use the model in Griffin and Orton (1993) for Uranus and Neptune, which may not be accurate for the 220 GHz array because of a CO J = 2 – 1 resonance at 231 GHz in Neptune’s atmosphere, which is not included in their smooth model. The solid angle used here is the disk area and ignores limb darkening.

plateau of 140 K to a minimum of less than ~ 100 K (see Weisstein and Serabyn (1994)). This has been well-measured both with a narrowband FTS (Weisstein and Serabyn (1994)) and with broadband instruments (see, e.g., Goldin et al. (1997)).

In addition, both the ring and moon systems are significant emitters. Here we develop a
5.6 Observations of the planets

Figure 5.19: Collected values of the Saturn disk-averaged brightness from literature. WMAP bands are shown in black and ACT bands are shown in dashed blue. The continuous curves are several models in recent literature. The $PH_3 J = 1 - 0$ resonance at 267 GHz will be important for the 280 GHz camera, and has been well-constrained; see Weisstein and Serabyn (1994). The spectrum is relatively flat from the upper ACT band to 145 GHz and 220 GHz.

A simple model of the rings similar to Weisstein (1996) and Epstein et al. (1980, 1984) in which the disk and rings are raytraced as a function of ring inclination to find the effective disk-averaged temperature. WMAP's studies (see Hill et al. (2008)) have been a boon to studies of the ring at high inclination. Here we use this data and a geometric model to extend to low ring inclinations. Subsequent publications by the WMAP collaboration should be sufficient to secure a ring model over all inclinations. Saturn’s A and B rings are thought to dominate the absorption and emission in the mm range. For the A-ring, we use an estimate of 20 K for the temperature and $\tau \approx 0.7$ for the optical depth. These are a compromise between several values in literature. For the B-ring, all relevant at $>10\%$. The complicated frequency structure and time evolution makes the comparison of broadband results across widely spaced times difficult, because they depend in detail on the passband and orientation of the rings at the time of observation.

To accurately model the emission as a function of ring angle, one must account for several complicating effects. The ring temperature is a function of both the opening angle with the sun and the region occluded by the disk of Saturn, and the optical depth and temperature vary radially. The full solution requires a complex treatment for the interaction of radiation from the disk with the rings, such as a transport Monte Carlo.

See Epstein et al. (1980, 1984); Dunn et al. (2005); Pajot et al. (2006); Dunn et al. (2007, 2004, 2002). The optical depth of the rings in the radio and infrared has been reasonably well-characterized, but we believe there are significant uncertainties in the millimeter-wave region considered here. The B-ring is roughly twice the optical depth of the A-ring, and...
The Instrumental Response to Radiation: Calibration

Figure 5.20: Left: the geometric model for Saturn’s emission as a function of ring inclination. Right: the effective disk brightness as a function of ring inclination predicted by the geometric model compared to the high inclination model constrained by the WMAP observations. The geometric model has several regimes: for low inclination, the planetary disk dominates the emission, but as the ring tips, it obscures part of the disk emission. By $\sim 12^\circ$, some of the ring is viewed in emission and as it continues to tip, it obscures less of the planetary disk. Above $22^\circ$ inclination, the rings boost emission. The definition of disk average used here is the effective temperature that the disk at zero inclination would have to be to emit as much as the disk/ring system.

we use a temperature again of 20 K, but an optical depth $\tau \approx 1.8$, based on the same literature. No limb darkening has been accounted for (an estimate from Pajot et al. (2006) is 4%), and we use the ring geometry from Dollfus (1970). We account for flattening of the disk and its inclination relative to the earth, and take the disk temperature to be 150 K, based on measurements in the range of 145 GHz. See Fig. 5.20 for a model disk, and Fig. 5.21 for a comparison with the WMAP data and projections for ACT and planetary solid angles. Synchrotron emission from the radiation belts of the giant gas planets is only significant (relative to the thermal emission) for Jupiter (see Gibson et al. (2005)).

5.6.5 Point source response relative to the diffuse calibration

The amplitude of the planet is modulated by a dilution factor of $\Omega_p/\Omega_{b,i}$, which depends on the solid angle $\Omega_{b,i}$ of detector $i$. Physical optics studies indicate that this variation is $\sim 15\%$, but there could be several factors in addition to the optics which lead to variability in the planet amplitudes across the array. The coupling of radiation to each detector will not be Lambertian, in particular, because of the angular dependence of the AR coupling silicon layer. If this angular dependence changes over the array from column card alignment or warping, it can change the illumination of the aperture, and this will change the solid angle of the beam on the sky. The other factor that could vary across the array is the TES operating point. For points low on the transition ($< 20\%$ of $R_n$), the responsivity will change in response to the planet, flattening their peak response.$^{58}$ Here we determine this empirically by dividing the relative response of the planet by the relative

$^{58}$ A single detector will absorb $\sim 0.3$ pW from an on-center observation of Saturn, and the sensitivity to loading below $20\%$ of $R_n$ exceeds $3\%$ per pW, giving a small but detectable effect. For operating points high on the transition, the added power could produce significant changes in responsivity.

121
5.7 The calibration emitter

A calibration emitter on the 1 K optics stage behind the Lyot stop is used to develop a time transfer standard. This is complementary to the electrical measurements for both the responsivity (Eq. 5.37) and change in responsivity from drifts in optical loading (Eq. 5.45). The calibration emitter has the advantage relative to the electrical methods that it is a direct measurement that does not depend on any device parameters.

The emitter is a reverse bolometer\(^{60}\), pulsed every 500 s for 800 ms during the 2007 season observations. The radiation passes through the bandpass, limiting the signal to noise of the detector response to \(\sim 10\). To reach a 1% relative sensitivity using a square wave pulse across \(N_{\text{det}}\)

\(^{59}\)The solid angle of a disk weighted by an off-axis Lambertian source is given in Prata (2004) but this geometric approach gives smaller solid angles off axis. A better approach is to find the beam of off-axis detectors using methods in Wollack et al. (2006); Withington et al. (2004, 2005); Chuss et al. (2007); Withington and Saklatvala (2007); Saklatvala et al. (2006) or to simulate the optics numerically. This warrants further followup with the optical model, as it could be expected to change the window function in multipole space across the array, leading to a position-dependent bias on the power spectrum, where central regions of the map with fuller array coverage see an average solid angle while edge regions would have more weight from peripheral detectors, leading to a varying mean solid angle. This could be measurable in power spectra from a sub-divided map, and could be an important systematic effect in ultimately understanding the value of the primordial spectral slope.

\(^{60}\)http://www.haller-beeman.com/
The Instrumental Response to Radiation: Calibration

Figure 5.22: Left: relative calibration inferred from the average of Saturn observations from the 2007 season divided by the relative calibration from the diffuse source. The planet data were calibrated using the response time transfer method described in Sec. 5.9 to propagate it to the best estimate of the responsivity at the time the diffuse atmosphere measurement was acquired. Here we take only observations with good array coverage (> 750 detectors), PWV less than 2 mm, and observations in the survey elevation of 50.5°. A component of the variation across the array is expected to be from the variation in the dilution factor caused by beam changes across the array. Nonlinear response to the planet may also impact the ratio, biasing the planet response low compared to the diffuse response. The impact would be expected to be largest in columns 5, 24, and 29 (based on Fig. 5.6) but the total power absorbed is ~ 0.3 pW, and would produce variations smaller than those shown here. Right: the raw relative calibration inferred from the Saturn observations. Except for the structure in the left panel at the 15%-level, this is similar to Fig. 5.12.
5.8 The all-pairs statistic

detectors sampled at 400 Hz, the pulse must be on for (See Appendix C.3)\textsuperscript{61} \textsuperscript{62}
\[ t_o > (100 \text{ sec}) \cdot \text{SNR}^{-1} \cdot N_{\text{det}}^{-1} \] \hfill (5.70)

Based on this, it is not practical to reach 1% in a single detector per pulse, but array averages and the all-pairs (Sec. 5.8) method can extract useful information. Even though the power applied to the emitter, the bandpass, and the geometry are known, the illumination pattern is complex (See Fig. 5.23). We can then compare a single detector at two different times (exploiting a constant pulse amplitude), or the relative calibration (assuming a constant illumination pattern).

The pulse shape averaged over the 2007 season is shown in Fig. 5.23. The shape is dominated by a cryogenic time constant (which is much slower than any detector time constant) in turning on, and the response does not equilibrate to a “fully on” value. We have developed two methods to find a time transfer standard using the emitter. The first method is to simply find the least-squared best scaling between a season-wide template (Fig. 5.23). The second approach is to fit explicitly for the pulse shape. The full fit could account for a change in the pulse shape if, for example, the cryogenic time constant of the 1 K stage changes over the night as the \textsuperscript{4}He adsorption refrigerator runs out. This method also removes a smooth baseline drift and deweights the signal where it identifies bursts, etc., that are not consistent with the template. Despite these added benefits and flexibility, we found that the most robust measure was simply the scaling to a template, so results presented here used that method. Fig. 5.24 shows the stability of the estimated calibration pulse amplitude under different global calibration treatments.

In practice, we find that the accuracy of the pulse height determination depends strongly on the atmospheric noise. The reason for this is that the pulse width defines a Nyquist frequency of 0.6 Hz for sampling the baseline, while the knee frequency for the drift is 1 – 2 Hz on the best nights and rises to ~ 7 Hz on nights with high PWV. Thus a significant amount of the baseline power cannot ever be represented and subtracted out, limiting precision (percent-level) calibration pulse stability measurements to the best nights. The drift is also largely constant over large fractions of the array, meaning that many detectors share the same bias and do not average down. In addition, some periods during the 2007 season have one-sided burst (row-correlation) noise extending to high frequencies, which biases the measurement unless it is treated by a non-normal correlated noise subtraction method. Development of more intelligent amplitude estimators is well-motivated because an empirical handle on calibration is preferable to the more complex electrical model used here.

5.8 The all-pairs statistic

It is difficult to visualize the stability of the instrument or the efficacy of the calibration over all the detectors across all observations. In this section we develop three statistics to describe the scatter across the array that compress the instrument response and calibration across the season into a histogram, or distribution width. There are three domains where it is useful to understand the stability: the absolute stability, the relative stability across the array, and the array-wide stability across times. The absolute measure of stability that we develop here is the distribution of the ratios of a given detector across the array across all times. In practice, this total stability depends on

\textsuperscript{61}Because of 1/f noise, the preference is for higher amplitudes rather than longer durations and the upper limit on the power that can be applied is set by the 1 K cryogenics. The 1 K optics must relax approximately to their base temperature between subsequent pulses. Otherwise, the 1 K temperature would creep over hours as the emitter heat exceeds the capacity of the \textsuperscript{4}He adsorption refrigerator. Because the hold time of this refrigerator did not limit the observations over one night and a change in the 1 K optics produces a negligible change in loading on the detectors, we maximized the power emitted by this cryogenic criterion.

\textsuperscript{62}Note that this scales as $N_{\text{det}}^{-1}$ because the accuracy of the amplitude estimate scales as $N_{\text{det}}^{-1/2}$ and as $t_o^{-1/2}$.
array-wide shifts (the array as a whole might be 10% more sensitive one night) and shifts within the array (a given detector might be 3% more sensitive than usual compared to the other detectors). These will be referred to as the array-wide and array-relative stability, respectively. These give slightly more nuance to understanding the calibration or stability. For example, we may find that one treatment improves the array-wide response with time, but worsens the array-relative stability.

Take the response model for a standardized observation to be a template $T_i$ (where $i$ labels the detector), then the quantity that is actually measured (in digital units, $D_i(t)$) is

$$D_i(t) = T_i C(t) [1 + \delta_i(t)], \quad (5.71)$$

where $C(t)$ is an array-wide shift and $1 + \delta_i(t)$ accounts for a small relative change across the array, which is defined to have a mean over the detectors of zero, $\langle \delta_i(t) \rangle_i \equiv 0$. We can then look at the ratio of the response in detector $i$ between time $t_1$ and time $t_2$, or

$$M_i^{t_1,t_2} = \frac{D_i(t_1)}{D_i(t_2)} = \frac{C(t_1) [1 + \delta_i(t_1)]}{C(t_2) [1 + \delta_i(t_2)]} \approx \frac{C(t_1)}{C(t_2)} [1 + \delta_i(t_1) - \delta_i(t_2)]$$

$$\approx \langle M_i^{t_1,t_2} \rangle_i [1 + \delta_i(t_1) - \delta_i(t_2)] \quad (5.72)$$

where in the last step we have exploited the fact that $\langle \delta_i(t) \rangle_i \equiv 0$ and assumed $\delta_i(t)$ is small.
5.8 The all-pairs statistic

Figure 5.24: The array-average of the estimate for the calibration pulse amplitude over the 2007 season through several calibration treatment steps. Only the “no treatment” points have error bars for clarity, but the error bars on the points in the two treatment families are similar. These are the estimate of the error on a single pulse measurement. Here, the amplitude estimator is a simple least-squared fit to the template shown in Fig. 5.23, which becomes less accurate over periods such as between Nov. 17 and 25th, 2007, where the knee frequency for atmospheric drift was consistently above 2 Hz, fast compared to the pulse time of 800 ms. The rebiasing correction uses Eq. 5.37 for the load curve analysis to find the change in responsivity from night to night, while the drift correction is based on Eq. 5.45 and attempts to correct the responsivity based on the power estimated from the PWV radiometer. The rebiasing correction removes most of the response shift here, but calibration pulses cannot currently be used to study nights with large changes in PWV, making it hard to assess the efficacy of responsivity drift correction. The planet response amplitude can be determined accurately, so is one of the best tests for the response regime where atmospheric drift and PWV are large.

compared to unity. If we define \( \bar{M}_{t_1,t_2} = \langle M_{t_1,t_2} \rangle \), and \( \Delta_{t_1,t_2} = \delta_i(t_1) - \delta_i(t_2) \), then

\[
\Delta_{t_1,t_2} = \frac{M_{t_1,t_2}}{\bar{M}_{t_1,t_2}} - 1.
\] (5.73)

This is the fractional change in a detector’s response relative to the rest of the array between time \( t_1 \) and time \( t_2 \). It is useful to combine these measurements into a single stack across all the detectors, as

\[
\hat{\Delta}_{t_1,t_2} = \cup_i \{ \Delta_{t_1,t_2} \}
\] (5.74)

This set has mean zero, and the width gives a measure of the stability of the relative calibration between those times. This is a ratio distribution so is not normal; a reasonable number to assign to the width is the quartile range. This can be used to conclude, for example, that 50% of detectors changed less than 1% relative to other detectors in the array.

One pair of observations may have low signal to noise or be anomalous somehow, so a better statistic is to compare all pairs of standardized observations. To do this, make a larger set of all
the residual across all pairs of times as

\[ \hat{\Delta} = \bigcup_{t_1, t_2, t_1 \neq t_2} \cup_i \{ \Delta t_1, t_2 \}. \quad (5.75) \]

We can form a similar statistic for the array-wide shifts where

\[ \hat{M} = \bigcup_{t_1, t_2, t_1 \neq t_2} \bar{M} t_1, t_2. \quad (5.76) \]

The width of the values of \( \hat{M} \) indicates the stability of the absolute/array-wide calibration.

As defined, these statistics only give a measure of the stability of a template observation in the raw data. To make this useful for calibration note that what we do when we calibrate is divide by an estimate of \( C(t)[1 + \delta_i(t)] \) from a calibration model, denote this as \( \tilde{C}(t)[1 + \tilde{\delta}_i(t)] \). Thus, the calibrated response \( \tilde{D}_i(t) \) is

\[ \tilde{D}_i(t) = T_i C(t)[1 + \delta_i(t)] / \tilde{C}(t)[1 + \tilde{\delta}_i(t)] \approx T_i Q(t)[1 + \alpha_i(t)]. \quad (5.77) \]

where \( Q(t) = C(t) / \tilde{C}(t) \) and \( \alpha_i(t) = \delta_i(t) - \tilde{\delta}_i(t) \). This has exactly the same form as the stability statistic, so the same conclusions apply. The goal of finding a good calibration is thus to minimize the width of the array-wide statistic \( \hat{M} \) and the array-relative statistic \( \hat{\Delta} \).

There are several quantities (the time constant, SQUID slopes) where the relative change across the array is large. This breaks the assumption here that allows the decomposition in the array-common and array-relative distributions. For these quantities, we can simply study the population of ratios of detectors chosen across random times. The scatter in this population indicates the overall stability of this figure for the detector.

### 5.8.1 Some applications

In Fig. 5.25 we show the all pairs statistic applied to the responsivity calculated from the load curves during the 2007 season. On the basis of load curves the array-wide shifts in responsivity can be 5% - 10% (with higher outliers), but the relative responsivity across the array is constant to a few percent, with only 50% of detectors exceeding 1% departure relative to the array.

### 5.9 The global calibration

There are six effects we account for in the global calibration solution: 1) the variability of the planet solid angle and temperature with time, 2) the change in responsivity from nightly TES rebiasing, 3) the change in responsivity from loading, 4) the change in the optical depth with time, 5) the array-relative response to diffuse sources, and 6) the array-relative response to point sources. These are all essential for a percent-level calibration.

#### 5.9.1 Method

Sec. 5.2.3 developed the infrastructure to find the change in power received for some \( \Delta T \) in the CMB. The goal of this section is to rephrase this method to make it practical for calibrations by rewriting it to emphasize different domains of the calibration and measurable values. As a first
5.9 The global calibration

Figure 5.25: Applying the pair statistic to study the stability of the instrument response and time constants. Upper left: A histogram of the relative responsivity shift between detectors at two random times as estimated by load curves. The 25% and 75% quartiles are at 1% indicating that 50% of detectors will shift more than 1% relative to the array at two randomly chosen times. This is a sharply falling function, and almost all detector will have changed less than 2% relative to the array at two different times. Here, 50% of biasing periods (nights) will have absolute shifts > 4% in responsivity across the array. Upper right: the stability of the time constant measured by bias steps over the season. The time constants used here are from a fit by E. S. and R. Fisher, but are consistent with results used elsewhere from J. Appel, that have been calibrated to (and checked against) optical measurements. There is considerable scatter in both the array-relative and array-wide statistics. Lower left: the stability of the TES operating point over the 2007 season. This indicates that even though the absolute responsivity changes by 5 – 10% between days, the operating resistance is stable to a few percent both in absolute and array-relative changes. Lower right: the stability of the SQUID amplifier chain slopes over the season. These show considerable scatter from variability in the tuning and SQUID behavior. Because of the high loop gain in the flux lock loop, these slopes effectively only modify the bandwidth, and become a concern if their bandwidth falls below the anti-aliasing filter bandwidth out to $f_{3dB} = 122$ Hz. Further work is needed to understand if this impacts the time constants.
The Instrumental Response to Radiation: Calibration

step, return to Eq. 5.24 and multiply by the responsivity to convert to filtered digital feedback units from some detector \(i\) \((\Delta D_i)\) given some \(\Delta T_{\text{CMB}}\),

\[
\Delta D_i = e^{\tau(t)} R_i(t) \bar{B}_i \bar{\Phi} k_B \Delta T_{\text{CMB},i}(t). \tag{5.78}
\]

The quantity \(\bar{B}_i \bar{\Phi} k_B\) is a combination of instrumental quantities that can be measured, but at this stage these are not known at the percent-level needed for a bottom-up calibration. Instead, we can combine them into a single factor \(A\) that gives the conversion between \(\mu\)K and DAC counts. The goal is then to estimate this empirically using planet measurements. The remaining components of \(R_i(t)\) are unitless and represent relative changes in the response detector \(i\) with time.

We can unpack \(R_i(t)\) into several components. Let the array-relative response to diffuse radiation in detector \(i\) at the reference time \(t_o\) be \(G^{\text{diff}}_i\). (For example, this might give detector \(i\) 10\% less sensitivity to radiation at time \(t_o\) relative to the rest of the array.) The relative calibration \(G^{\text{diff}}_i\) is analogous to flat-fielding, and is determined by measurements of the atmosphere at \(t_o\). We will define it to have an average of 1 over the array \((\langle G^{\text{diff}}_i \rangle_i = 1)\) so that multiplying by it does not shift the average responsivity of the array. The relative calibration could be different for each observation, because it depends on how the array was rebiased at the start of the night and on how the detector responsivities drift relative to one another over the night from optical loading. We can therefore propagate the relative calibration to some other time (that night) using \(1 + \delta^{t_o-t}_i\) where \(\delta^{t_o-t}_i\) is small and averaged across the array is defined to be \(\langle \delta^{t_o-t}_i \rangle_i = 0\). \(\delta^{t_o-t}_i\) might indicate that between time \(t_o\) and \(t\), detector \(i\) dropped in sensitivity to radiation by 2\% relative to the other detectors in the array.) The array-wide shifts can be absorbed in another constant \(\bar{\Phi}\).

The response of a detector \(i\) to changes in the diffuse CMB temperature is then

\[
\Delta D_i = AG^{\text{diff}}_i e^{-\tau(t_o)} \Delta T_{\text{CMB},i}(t). \tag{5.79}
\]

In the reference epoch this is

\[
\Delta D_i = AG^{\text{diff}}_i e^{-\tau(t_o)} \Delta T_{\text{CMB},i}(t_o). \tag{5.80}
\]

The response to a planet observed at time \(t_p\) is similar, where

\[
D_i\bigg|_{\text{peak}} = AG^{\text{diff}}_i e^{-\tau(t_p)} G^{\text{diff}}_i e^{-\tau(t_o)} \Theta(f_{\text{3dB},i}) \left( \frac{\bar{\Omega}_b}{\bar{\Omega}_i} - \frac{\bar{\Omega}_b}{\bar{\Omega}_o} \right) T_{\text{planet},RJ}(t_p) \frac{\bar{\Phi}}{\bar{\Phi}}, \tag{5.81}
\]

where \(\bar{\Phi}\) is the conversion from Rayleigh-Jeans temperatures to changes in CMB temperature and the ratio \(\frac{\bar{\Omega}_b}{\bar{\Omega}_o}\) is determined using the planet/diffuse source analysis to find the change in solid angle across the array. \(\Theta(f_{\text{3dB},i})\) describes the correction to planet amplitudes from the time-domain response. We assume the planet observation is in the linear response regime, so that the amplitude is not distorted. For Mars and Saturn in the 2007 season, this was \(\sim 0.3\) pW absorbed power.

The global calibration factor \(A\) (with units of DAC units per Kelvin CMB) inferred from a measurement at \(t_p\) is then

\[
A(t_p) = e^{\tau(t_p)} \Theta(f_{\text{3dB},i}) \left( \frac{\bar{\Omega}_o \bar{\Omega}_b,i}{\bar{\Omega}_b \bar{\Omega}_o} \right) \left( \frac{\bar{\Phi}}{\bar{\Phi}} \right) \left( \frac{D_i\bigg|_{\text{peak}}}{T_{\text{planet},RJ}(t_p)} \right). \tag{5.82}
\]

\(^{63}\) The distinction between array-wide and array-relative is artificial, but it is a convenient way to think about stability as either changes in the whole array with time or as fluctuations within the array.
5.9 The global calibration

Note that the bandwidth-efficiency product $\bar{B}$ is needed to convert incoming intensity to power, but does not appear here. Here we develop a value for $A$ with Saturn observations that: 1) have good array coverage (> 750 detectors) 2) at the survey elevation of 50.5°.\textsuperscript{64}

The procedure for calibration in the pipeline is:

- Estimate the change in optical loading between the initial TES biasing and the observation using the PWV or drift radiometer.

- Use Eq. 5.45 to estimate the change in responsivity from the beginning of the interval. This flags detectors that are either high on the transition ($\sim R_n$) or low on the transition ($R < R_{sh}$) after the loading change. It also generates a matrix of corrections, but these are typically small, of order 1 – 2% for most detectors. An optional flag propagates the response during the observation back into response in the epoch where they were biased, attempting to undo these responsivity changes over the night. (Opacity is treated separately; see the last bullet here.)

- Use Eq. 5.37 to find the ratio of responsivity inferred from load curves at the beginning of the night and at the beginning of the night of the reference observation. Propagate the response from the given observation to response in the reference epoch using the ratio of the responsivity. Fig 5.25 shows that the changes in the relative calibration between these two times will be small, but there could be large array-wide shifts of order 10%.

- With the response propagated into the reference epoch, use the relative calibration inferred from the atmosphere in that epoch to flatten the array-relative response (see Sec. 5.5).

- After flattening the response to the diffuse source, the calibration pulse will have some non-uniform illumination and the planets will have non-uniform response because of the variation in the dilution factor across the array. These can optionally be leveled using the estimated variation across the array from season-wide calibration pulses and planet measurements.

- The response is now flattened and in digital units in the reference epoch. This can then be converted to a flux using the estimate for the flux calibration from the planets (also with respect to that reference epoch), Eq. 5.82.

- Correct for optical depth based on the PWV from the APEX or ACT drift radiometer. This is done by default but is optional, so can be skipped for atmospheric or calibration pulse studies.

Note that in all of this, the reference epoch is irrelevant for the final calibration. It simply gives a common point to tie together the instrumental response. We take the reference epoch to be 1197255523 seconds since January 1, midnight UTC, 1970. (This is when the array-relative calibration was acquired, in the night starting on Dec. 10, 2007.) Fig. 5.26 shows the steps of the calibration for one detector, and Fig. 5.27 shows the distribution of calibrated planet responses with different levels of treatment and scaling between PWV and absorbed power and optical depth.

Some improvements to the calibration would be to: 1) find a method to estimate the shunt resistance in the field, at the percent level, 2) remove the baseline in the calibration pulse data using information about the drift spectrum, 3) observe a larger number of planets from the “science” elevation and not at transit, 4) find the array-relative calibration in each TOD.

The average over the array gives the global calibration to better than 1%, where the error should be dominated by systematic uncertainties in the beam and planet temperature models.

\textsuperscript{64}Data from the “planet fest” (at multiple elevations) is skipped here, but a subject for future study would be to confirm that the loading and optical depth models are also consistent with these data. Disagreement would suggest a problem with either the model, or the mechanical structure as a function of elevation. We have also neglected Mars data because they were acquired at transit, so had partial array coverage.
5.9.2 Calibration errors

The global calibration is subject to two families of errors. Given a known input (such as a step in \( \Delta T \)) throughout the season, apply the calibration described in Sec. 5.9.1 to the raw digital output to find an estimate for \( \Delta T \). Call this calibrated estimate \( \Delta \tilde{T} \). We will refer to random error as the width of the distribution for the estimate \( \Delta \tilde{T} \) and systematic error as the offset of the central value of \( \Delta \tilde{T} \) from the true value \( \Delta T \).

The first goal of this section will be to understand the random error component of the calibration, which has terms from the calibration time transfer and the relative calibration estimate. The approach here will be to estimate the random error in the flux calibration and relate this to the random error in calibrating real survey data.

Suppose the season were repeated an arbitrary number of times with a different realization of the noise but the same signals. In each realization, we would estimate a different calibration constant for a given detector at a given time. Let the main calibration transfer constant be denoted as \( C = C_{t-o-t_p}(1 + \delta_{t-o-t_p}^o)e^{-\tau(t_p)} \) and the relative calibration just be \( G = G_t^o \). For the planets, compress the relative calibration and variation of dilution across the array into into \( S = G_t^o \frac{\Omega_p}{\Omega_t} \frac{\Omega_n}{\Omega_t} \) and let \( \Theta = \Theta(f_{\text{MB}}) \). (We also drop \( \bar{\Phi} \) by converting the RJ planet temperature to CMB temperature units.) Then the digital output in response to a planet (Eq. 5.81) is \( D_p = ACG\Theta ST_p \) and the digital output in response to the CMB (Eq. 5.79) is \( D_{\text{CMB}} = ACGT_{\text{CMB}} \). The difference between these is that the planet amplitude has the correction for finite time constants \( \Theta \) and variability of the dilution factor across the array, \( S \). In the first step of the calibration, we estimate the variation of the dilution factor (here, \( S \)) across the array by taking the average of \( D_p/(ACG\Theta T_p) \) over \( N \) different planet observations. This is the sum of the fractional error-squared of the amplitude measurement (\( D \)), the calibration time transfer (\( C \)) and the time constant correction (\( \Theta \)), so that for independent (and identically distributed) normal variables

\[
\frac{\sigma_S}{S} = \frac{1}{\sqrt{N}} \left( \frac{\sigma_D^2}{D^2} + \frac{\sigma_C^2}{C^2} + \frac{\sigma_{\Theta}^2}{\Theta^2} \right)^{1/2}.
\]

(5.83)

The fractional error-squared over realizations of the calibration constant \( A \) estimated from a single detector and a single planet is similar, where

\[
\frac{\sigma_A}{A} = \left( \frac{\sigma_D^2}{D^2} + \frac{\sigma_C^2}{C^2} + \frac{\sigma_{\Theta}^2}{\Theta^2} + \frac{\sigma_S^2}{S^2} \right)^{1/2}.
\]

(5.84)

We can estimate \( \frac{\sigma_A}{A} \) by finding the distribution of \( A \) across all detectors at all times from planet measurements.\(^{65} \) This gives

\[
\left| \frac{\hat{A}}{\text{All Saturn}} \right| = (3.43 \pm 0.09) \mu \text{K/DAC} \quad \text{at 1}\sigma \text{ per observation per detector},
\]

(5.85)

or, a 2.6\% fractional error (at 1\( \sigma \)) on \( A \) for a given measurement of the planet at one time. There were a total of \( \sim 10^4 \) independent measurements of \( A \) (15 observations with \( \sim 750 \) working detectors), so the flux calibration can be determined very precisely over the season.

Once we fix \( A \) for the season calibration, it does not contribute a random error, and becomes a systematic error. (A 10\% random error in the determination of \( A \) would become a 10\% systematic error in the calibration over the season.) Then the only random error in the CMB temperature

\(^{65} \)We do not have access to multiple realization of the season except by subdivision, but here we can assume that the estimate of \( A \) for each detector at each time is independent and identically distributed. Recall that all detectors should measure the same \( A \) after flat-fielding.
5.9 The global calibration

measurement is from the propagation of the calibration during the season through the factor $C$ and the relative calibration,

$$\frac{\sigma_T}{T} = \left( \frac{\sigma_D^2}{D^2} + \frac{\sigma_C^2}{C^2} + \frac{\sigma_G^2}{G^2} \right)^{1/2}.$$

(5.86)

If we ignore the noise in the CMB measurement to get a figure for the random error from calibration

$$\frac{\sigma_T}{T} \bigg|_{\text{calibration}} = \left( \frac{\sigma_C^2}{C^2} + \frac{\sigma_G^2}{G^2} \right)^{1/2}.$$

(5.87)

We can put an immediate bound on the calibration error owing to the fact that the error in $A$ is the calibration error in quadrature with many other error terms (the time constant correction, determination of the dilution factor, the planet amplitude measurement). Then

$$\frac{\sigma_C}{C} < \frac{\sigma_A}{A} = 2.6\%.$$

(5.88)

To get a better estimate of $\frac{\sigma_C}{C}$, we can estimate the component of $\frac{\sigma_A}{A}$ that is due to errors in the planet amplitude measurement. One way to do this is to assume the illumination pattern is constant, and find the scatter about the illumination pattern for each planet observation. This is directly related to the array-relative pair statistic from Sec.5.8. Doing this, we find that the scatter is 1.8%, so that the remaining error in the measurement of $A$ is 1.9% (giving the 2.6% in quadrature). The relative calibration varies from detector to detector but can also be determined to $\sim 1.0\%$ to 1σ, giving

$$\frac{\sigma_T}{T} \bigg|_{\text{calibration}} = 2.1\% \text{ at 1σ.}$$

(5.89)

With improved relative calibration from the diffuse atmosphere, it should be possible to achieve a relative calibration across the array to 0.5%, giving 2%. This is significant because it also indicates that the optics are reasonably stable. Note, though, that all Saturn observations used here were at roughly the same time of night. The Ruze formula (see Ruze (1966)) gives 0.94% efficiency for 40 μm rms primary mirror alignment errors in 145 GHz. If this were the only source of random error, this would imply that the rms is constant to 6 μm (at 1σ), alternately, it implies that solid angle (which is taken to be constant here) is stable to 5 nsr.

The systematic error in the calibration is likely dominated by the planet temperature and the solid angle estimate and its variability. Given the state of the art of Saturn measurements and models (see Appendix C.2 for a summary), a reasonable error on the disk temperature is $\pm 5\%$. The effect of the ring inclination has been strongly constrained for high inclinations by recent WMAP Hill et al. (2008), but there is some uncertainty in projecting forward to the lower inclinations during the ACT seasons. The total excursion in the Saturn temperature in the geometric model developed here is 13%. The ring must give the disk temperature at zero inclination and be consistent with WMAP at high inclination. For the purposes here, we put a 20% error (at 1σ) on the remaining variability from the ring, or 2.6%, giving 5.6% error overall (at 1σ). We take the error on the solid

---

66Studies by K. Moodley give a solid angle with 4.3 nsr error at 1σ (which is believed to be dominated by the scatter in the measurements rather than a trend). This rms stability also agrees with the level suggested by studies by R. Dünner. This warrants followup in parallel with beam studies to determine if, for example, using the time-dependent solid angle improves the scatter in the calibration. Aside from optical errors, a remaining random error term is the PWV, which enters both in the responsivity drift and the optical depth. The optical depth and drift produce corrections at the level of a few percent, so PWV estimate errors from the APEX data could produce errors at nearly the percent-level; the noise of the radiometer is not constant in time, and varies from being very precise (percent-level) to having large scatter at the level of tens of percent.

67This is conservative given the errors in, e.g., Goldin et al. (1997), but the temperature of Mars also enters and those models have additional uncertainty.
angle to be 2%, based on the scatter in its measurement. This gives 6% estimated error at 1σ for the diluted temperature. In addition, the amplitude estimator may be systematically biased. Two independent methods (developed by A. Hincks in one case, and J. Fowler T. Marriage, and B. Reid in the other case) are systematically ~ 2% different even for fast optical time constants. Suppose we allow 3% systematic error in the amplitude measurement. This gives 7% for the total systematic error at 1σ. A significant part of this is due to the planet temperature estimate. Subsequent WMAP studies of Saturn at low ring inclinations are expected to improve this and may give gains ~ 3% on the systematic error. Ultimately, the systematic error will be suppressed to the percent-level using cross-correlation with WMAP and the dipole, but alternative calibrations from the planets will be informative.

To summarize, the random error estimated here is 2.1% at 1σ, and the systematic error in the flux calibration from Saturn observations is estimated to be 7%. We expect that the relative calibration could be improved to the sub-percent level possibly on a TOD-by-TOD basis and that the calibration emitter could be analyzed more carefully to achieve errors at the percent level. This would allow an empirical constraint on the random calibration error at the percent-level, but is the subject of future work.

Further study is needed to understand systematic errors in the solid angle from, e.g., models with errors in the side-lobes. Carlos Hernandez-Monteagudo, A. Hincks, K. Moodley, D. Swetz and others are currently finalizing the understanding of the beam model for 2007 data. We have not included the spectral index correction described in Sec. 5.2.1, but expect that it represents a ~ 1% systematic error here. This correction also depends on an accurate knowledge of the passband and spectral slope of the source, so will be the subject of future work. Ultimately, the absolute calibration will come from WMAP. Note that solid angle and amplitude errors from the same estimator are correlated.

As noted in the introduction to this chapter, the random calibration error will average down by the number of intervals with independent calibration error over the season. This will be described in future work.
5.9 The global calibration

Figure 5.26: The calibration from filtered feedback digital units to $\mu$K (CMB) as a function of time for one detector with different levels of calibration treatment applied. The simplest is to assume a flat calibration for all detectors at all times; the next level of treatment is to apply the relative calibration from the diffuse atmosphere developed in Sec. 5.5; to account for changes in biasing, we can use the known bias voltages and load curves each night to develop a time transfer through Eq. 5.37; on top of that, the responsivity will drift due to optical loading, and we can apply the correction from Eq. 5.45 given the inferred change in loading from the APEX PWV radiometer; the last layer is to include optical depth, which is also inferred from the PWV radiometer. Note that responsivity drift and attenuation both apply corrections in the same direction. If the PWV falls, the power falls, the operating point falls in resistance, the responsivity rises, and the correction to counter this rising responsivity decreases. If the PWV falls, then the optical depth falls, this causes the signal from the sky to increase, but to counter this increase, the calibration multiplier must decrease. The responsivity correction is roughly $5\% \cdot 0.43 \cdot$ PWV, or 2% per mm, while the optical depth correction is $\sim 3\%$ per mm. These cannot be combined into a single figure of $\sim 5\%$ per mm because the sense of the change is different. The optical depth depends on the instantaneous PWV, while the responsivity drift depends on the change in loading since the TESs were biased (which provides the responsivity reference for the time transfer). The typical number of detectors that can be calibrated is around 820, but this number is sometimes lower because of failures in the load curve analysis or detectors that are driven close to normal or fall below $R_{th}$. 

134
Figure 5.27: Top: The distribution of the ratio of all live detector responses over randomly chosen pairs of times with various stages of the global calibration applied. For a given calibration procedure, this represents the scatter in its prediction for the planet amplitudes. The outermost “no treatment” distribution is the distribution of ratios of the raw digital units. The next distribution in includes the variation in planet temperature and solid angle; in from that includes the time transfer standard from load curves; in from that adds the responsivity correction from loading; in from that adds an atmospheric attenuation model. Both the responsivity correction and optical depth model use APEX weather station PWV data. The scatter at this level is comparable to the scatter in the array-relative pair statistic – it is near the level of uncertainty in the planet amplitude measurement. Bottom: The calibration error for the conversion from digital units to CMB temperature as a function of the scaling between PWV and optical depth and optical loading. The model that gives the lowest scatter is $\tau = 0.031 \cdot \text{PWV} + C$ and $P = 0.43 \cdot \text{PWV} + C$, which is consistent with the power inferred from the drift in Sec. 4.1 and the ATM atmospheric model (see Pardo et al. (2001)) in Marriage (2006). These are the scalings for the airmass at 50.5° elevation (PWV is through zenith). (Here $C$ is the dry part of the optical depth and absorbed power, which are 0.012 and 0.236 respectively, but are not relevant to the scatter here because they are constant.)
5.9 The global calibration
Chapter 6

Quantifying Sources of Noise and Elements of Mapmaking

6.1 Introduction – converting bits to maps

Chapter 5 describes how the telescope converts an intensity anisotropy into digital time-ordered data (the “TOD”). The calibration developed there converts these digital units into CMB temperature units. The goal then is to convert the $32 \times 32 \times 400$ Hz calibrated data streams into an estimate for the temperature anisotropy on the sky.\(^1\) This chapter gives an overview of the map estimation problem (Sec. 6.2) and some noise sources external to the inherent detector noise (Sec. 6.3). Sec. 6.4 and Sec. 6.5 describe two tools to identify non-normally distributed noise and detector-detector correlations. Sec. 6.6 gives a general overview of correlated noise models and treatments. Appendix C.4 supplements this discussion with some relations between optimal filters, iteration, and basis filters. At this time, the analysis of the first season (2007) and the beginning of second season (2008) is underway. We will focus on methods to apply to multi-detector data rather than on the results of the full analysis.

The motivation for describing the map estimation problem first is that noise sources and identification or cleaning tools have to be considered in parallel to the map estimation problem. Detector-detector correlations and non-normal noise are important to identify because the only computationally feasible noise model for map estimation at this time assumes normally distributed data and neglects detector-detector correlations.

In addition to noise families that invalidate the assumptions of the map estimation, there can also be noise that is not well-constrained by finite data. For example, take a map from a single telescope scan throw. If the emission of the atmosphere decreased during this throw, there is no way to distinguish this from decreasing gradient in the temperature anisotropy on the sky. As more data are acquired, the atmospheric drift terms are random, and it becomes less and less probable that they are also consistent with the anisotropy pattern on the sky. This facilitates a noise model that can be used to remove these components. This is described in Sec. 6.6.

\(^1\)There is a large literature that describes map estimation. See e.g., Tristram and Ganga (2007); Ferreira and Jaffe (2000); Patanchon et al. (2007); Prunet et al. (2000); Hivon et al. (2002); Keihanen et al. (2005); Natoli et al. (2002); Poutanen et al. (2004, 2006); Stompor et al. (2002); Doré et al. (2001); Delabrouille (2001); Ferreira (2001); Natoli et al. (2001); Hamilton (2003).
6.2 An overview of mapmaking

The response of a single detector is stored as a time-ordered vector \( \mathbf{d} \) of length \( N_{\text{TOD}} \) samples, and the value of the temperature anisotropy in each pixel in a map of \( N_{\text{pix}} \) pixels is represented by a vector \( \mathbf{x} \) of length \( N_{\text{pix}} \). The pointing matrix \( \mathbf{M} \) (with dimensions \( N_{\text{TOD}} \times N_{\text{pix}} \)) projects the signal model into the time domain as \( \mathbf{Mx} \). Arbitrary system noise as a function of time contributes through \( \mathbf{n} \) to give

\[
\mathbf{d} = \mathbf{Mx} + \mathbf{n},
\]

where \( \mathbf{n} \) has \( N_{\text{TOD}} \) elements. The goal of mapmaking is to find an estimate \( \hat{\mathbf{x}} \) given \( \mathbf{d} \). The Bayesian approach is to find the probability of a model (or map) given the data, here assuming a flat prior probability \( P(\hat{\mathbf{x}}|\mathbf{d}) \) so that

\[
P(\hat{\mathbf{x}}|\mathbf{d}) \propto \exp[-(\mathbf{d} - \mathbf{M}\hat{\mathbf{x}})^T \mathbf{N}^{-1}(\mathbf{d} - \mathbf{M}\hat{\mathbf{x}})/2]
\]

\[
\propto \exp(-\chi^2/2),
\]

where \( \mathbf{N} \) is the noise covariance, \( \langle \mathbf{n}\mathbf{n}^T \rangle \). The maximum likelihood solution for \( \hat{\mathbf{x}} \) is then (see e.g., Tegmark (1997))

\[
\hat{\mathbf{x}}|_{\text{ML}} = (\mathbf{M}^T\mathbf{N}^{-1}\mathbf{M})^{-1}\mathbf{M}^T\mathbf{N}^{-1}\mathbf{d}.
\]

The maximum likelihood estimator is unbiased and is the minimum variance estimator for the parameters (which in the simplest model are just the map pixels) (see e.g., Natoli et al. (2001)). The prefactor \( (\mathbf{M}^T\mathbf{N}^{-1}\mathbf{M})^{-1} \) is the pixel-pixel noise covariance, \( \mathbf{N}_{pp} \). For uncorrelated noise with constant variance \( \sigma^2 \), \( \mathbf{N}^{-1} \) is \( 1/\sigma^2 \) along the diagonal. In this case, the term \( \mathbf{M}^T\mathbf{N}^{-1}\mathbf{d} \) Eq. 6.3 coadds the data that land in a pixel and divides by \( \sigma^2 \). The map inverse covariance \( (\mathbf{M}^T\mathbf{N}^{-1}\mathbf{M})^{-1} \) in this case is \( \sigma^2/N_{\text{obs},i} \), where \( N_{\text{obs},i} \) is the number of observations that were coadded into a map pixel. The overall calculation is simply an average of the data in the map pixel. For white noise with changing variance, \( \mathbf{N}^{-1} \) is still diagonal with \( \sigma_i^2 \) different at each point in time and then \( \hat{\mathbf{x}}|_{\text{ML}} \) is the ordinary weighted average of data that land on a pixel.

In the case of full covariance, evaluating \( \mathbf{N}^{-1} \) is computationally intensive. If the noise is stationary (so depends on \( |t - t'| \)), each row of \( \mathbf{N} \) is the autocorrelation function of the noise and is same as the row above, but shifted to the right. This matrix has Toeplitz form (see Gray (2002)) and is invertible in \( O(n^2) \) by the Levinson-Durbin algorithm (see Delsarte and Genin (1986); Durbin (1960)). To make this practical, assume further that the autocorrelation goes to zero for short time separations compared to the length of the file. Then assume that the rows of \( \mathbf{N} \) wrap in addition to just shift, correlating the beginning and end of the file. (Here, what was at the right edge of the matrix will be at the left edge in the next row down. This is the circulant matrix form.\(^3\)) In this case, \( \mathbf{N}^{-1}\mathbf{d} \) can be calculated in the Fourier domain as \( \mathcal{F}^{-1}(d(\omega)/n(\omega)) \). Application of \( \mathbf{N}^{-1} \) is just a filter which deweights frequencies that are identified as noise.\(^4\)

Computationally, it is convenient to rewrite the \( ML \) problem in the form \( \mathbf{Ax} = \mathbf{b} \) as

\[
\mathbf{M}^T\mathbf{N}^{-1}\mathbf{M}\hat{\mathbf{x}} = \mathbf{M}^T\mathbf{N}^{-1}\mathbf{d}.
\]

The inversion problem for \( \hat{\mathbf{x}} \) is equivalent to a maximization problem over the likelihood. We solve this maximization problem using the preconditioned conjugate gradient method (PCG, see

\(^2\)Project \( \mathbf{n} \) into a map \( \mathbf{x}_i = (\mathbf{M}^T\mathbf{N}^{-1}\mathbf{M})^{-1}\mathbf{M}^T\mathbf{N}^{-1}\mathbf{n} \), then the covariance

\[
\mathbf{N}_{pp} = \langle \mathbf{x}_i\mathbf{x}_i^T \rangle = (\mathbf{M}^T\mathbf{N}^{-1}\mathbf{M})^{-1}\mathbf{M}^T\mathbf{N}^{-1}(\mathbf{n}\mathbf{n}^T)\mathbf{N}^{-1}\mathbf{M}(\mathbf{M}^T\mathbf{N}^{-1}\mathbf{M})^{-1} = (\mathbf{M}^T\mathbf{N}^{-1}\mathbf{M})^{-1}.
\]

\(^3\)This imposes periodicity on the autocorrelation function, but for correlations dying off after \( \Delta t \) small compared to the length of the file, this only impacts the first and last \( \Delta t \) of the data.

\(^4\)At this point, the \( \mathbf{n} \) that \( \mathbf{N} \) is derived from is the noise. If we knew this noise, we could subtract it from the outset. In practice, \( \mathbf{N} \) is always a model or estimate of the noise. Sec. 6.3 describes how this can be estimated in parallel with the map estimation.
Quantifying Sources of Noise and Elements of Mapmaking

Preconditioning is a transformation of the linear problem under $C^{-1}$ so that $C^{-1}Ax = C^{-1}b$. The goal is to choose a $C^{-1}$ that minimizes the disparity between the eigenvalues of $A$, thus speeding convergence. In our problem, $A$ is the inverse pixel-pixel covariance $N_{pp}^{-1} = M^T N^{-1} M$, and a simple operation for the preconditioner which makes the eigenvalues more uniform is to divide by the diagonal of $N_{pp}^{-1}$. This is called the Jacobi or diagonal preconditioner.$^5$ A measure for the convergence of the algorithm at some iteration $n$ is the residual vector $\epsilon$ weighted and squared to form $\Psi$:

$$\epsilon = M^T N^{-1} M \hat{x} - M^T N^{-1} d = \epsilon^T N_{pp}^{-1} \epsilon.$$ 

(6.6)

Weighting the residual by the pixel-pixel covariance emphasizes minimizing the residual in low-noise pixels, but it is computationally intensive. $\Psi = \epsilon^T \epsilon$ gives the same ultimate answer and is more practical to implement as a convergence or minimization condition. The minimization to find $\hat{x}$ has two computational components: 1) find $M^T N_{i-1} q$, where $q$ is an $N_{TOD}$ vector and can either be the detector data $d$ itself, or a synthetic TOD which represents the signal, and 2) find $M \hat{x}$ for some $\hat{x}$. The iterative solution through the PCG replaces a brute force inversion of $M^T N^{-1} M$ with successive applications of $M^T N^{-1}$ (filtering) in the Fourier domain and projections of models onto the time domain (projecting).

6.3 External noise sources and correlations

6.3.1 Discussion

Contaminants can either be treated in the map estimation as a model in addition to the sky model, or as noise covariance. Map estimation is then a self-contained process that takes the calibrated TOD with all of its contaminants and estimates a map from the components of those data that are consistent with anisotropy on the sky. In general, any pre-processing to remove noise before the map estimation can either irrecoverably remove sky signal, or add unwanted correlations to the data. To whatever extent possible, noise treatments should work within the map estimation structure. If they do not, care must be taken to understand how the pre-processing affects the sky signal and correlation structure.

Noise contaminants in the covariance matrix $N$ will be naturally “deweighted” in the map estimation through $N^{-1}$. The TOD is noise-dominated, so as a first pass, $N$ can simply be the covariance of the TOD (or a model which represents the covariance of the TOD). A better estimate can be found by iterating the mapper to update $N_{i-1}$ at iteration $i$ by taking the TOD minus the current best estimate of the sky signal as an estimate for the noise at that iteration, forming $N_i = \langle n_i n_i^T \rangle$, where $n_i = d - Mx_{m,i}$. In the first iteration, some component of the sky signal may be removed. So long as only a small part of the sky signal is lost in the first iteration, then when this first map is projected back into the time domain and subtracted off, it will give a good estimate of the noise-only component of the TOD. With each iteration, the sky signal is purified out of the noise estimate. This allows the map estimation to effectively deweight components of the TOD that are not consistent with a temperature anisotropy on the sky, and to extract the signal component. 

So far, this discussion has applied to a single-detector experiment, but it can be extended to include $N_{det}$ detectors by appending their data to $d$ and their projections into the model vector to $M$. What was a nearly intractable problem for one detector has now become more intractable for a

---

$^5$Another approach which should yield faster convergence is to use the known form of the array-common drift to give a more accurate estimate of $N_{pp}^{-1}$. The general goal is to find a matrix most like $N_{pp}^{-1}$ using prior information which is easily invertible.
6.3 External noise sources and correlations

A 32 × 32 array of detectors because $N$ is now $(N_{\text{det}} \cdot N_{\text{TOD}}) \times (N_{\text{det}} \cdot N_{\text{TOD}})$. Each $N_{\text{TOD}} \times N_{\text{TOD}}$ block in $N$ contains the cross-correlation for a pair of detectors or, in the Fourier domain, the cross-power of a pair of detectors. The main goal of mapmaking from large detector arrays is then to identify the detector-detector and low temporal frequency correlated noise components and model them out. The problem becomes tractable by assuming that there are no detector-detector noise correlations after successful application of the model. The final map is then the sum of the detector maps weighted by their pixel covariance.

6.3.2 Overview of contaminants

Correlations between detectors are dominated by the atmosphere and pickup in the electronics. As described in Sec. 5.4.1, the detectors are read across all columns in a row at one time. Electrical pickup then naturally correlates all the columns within a row. We describe this throughout as “row-correlated noise” or “burst noise”, because of the characteristic temporal structure of the interference. The other family of noise is caused by the modulation of the Earth’s field through the SQUIDs as the telescope scans. A strong component of this is from the series array channels, and so will correlate all the rows within a column. We describe this throughout as “synchronous pickup” or “column-correlated noise.” The atmosphere also produces strong correlations across the array. As described in Sec. 4.2 and Sec. 5.14, variations in the water vapor along a column produce both changes in loading over the night (drift) and correlations across the array. Fig. 6.1 is an example of these sources of noise in a raw digital time stream and Fig. 6.2 gives a “waterfall plot” of the power spectrum in one detector over the season.

In addition to correlated contaminants, the astronomical sky signal is also correlated between detectors. This is because of the spatial structure of the surface brightness and the $0.5F\lambda$ optical condition in 145 GHz, which causes the beams to overlap on the sky. It is therefore essential to develop a model that removes noise correlations due to the atmosphere and electronic pickup while preserving the correlations from the sky signal and does not simply remove all detector-detector correlations by, for example, diagonalizing the detector-detector covariance.

6.3.3 Overview of treatments for contaminants

There are several possibilities for removing correlations from the atmosphere. The most direct is to assume the turbulence is at an altitude $\sim 1000$ m where the views from detectors that are nearby on the array share significant overlap on the turbulence. This is what led to the pattern of residuals in Fig. 5.14, where the structure was very similar inside a radius $\sim 10$ detector lengths. The atmospheric drift for a detector can then be taken as a weighted average of nearby detectors. As shown in Fig. 4.4, the atmosphere has a $1/f^\alpha$ spectrum. The atmospheric subtraction can also exploit this temporal structure by finding the optimal filter for the known atmospheric spectrum. This is described in Sec. 6.6 and developed in more detail in Appendix C.4. These methods do not exploit anything about the turbulent atmosphere itself other than its height and spectrum. In the approximation that the atmospheric turbulence is a thin, frozen sheet that moves across the field of view, a powerful atmospheric subtraction method is to solve for a map of that sheet in parallel with the astronomical sky map. This is not described here and is the subject of ongoing work, but the discussion in Sec. 6.6 is a toy model for some of the issues there. Methods to treat electrical noise and correlations are also an active area of development.

---

6For example, in a typical 10 hour day which defines an analysis period, one detector produces $1.4 \times 10^7$ samples, while the array as a whole produces $1.4 \times 10^{10}$ samples. Both are impractical to invert by any brute force method. See Patanchon et al. (2007) for a description of detector-detector covariance.

7R. Dünner is developing a method to clean correlations from electronic pickup using the principal components from dead detector signals to identify and subtract common electrical pickup modes that are non-optical. The amplitudes for
Figure 6.1: A region of data from two detectors in the 145 GHz camera demonstrating long-term drift from changes in the atmosphere, burst noise, and calibration pulses. The y-axis gives the calibrated amplitude of the response in CMB temperature units. The upper trace is the detector in column 15, row 15, and the lower trace is column 6 row 6. The long-term drift is largely common to all detectors, but has higher temporal frequency components that have structure across the array, as shown in Sec. 5.14. The burst noise is due to electrical interference across all of the columns and thus correlates all columns in a given row. The period shown above was flagged as having high burst noise. It is also apparent from the two detectors shown here that the strength of the noise varies across the array, so data cuts must be performed on a detector by detector by time basis. The burst noise is broadband, manifests itself as a series of spikes common to all columns in a row, and is not normally distributed. Calibration pulses (the two spikes here) occur at known times and can be masked out. Scan-synchronous pickup produces correlations along all the rows in a column, but is small and so not discernible in this plot.
6.3 External noise sources and correlations

Figure 6.2: The power spectrum of row 6 column 6 over the season. We refer to this as a “waterfall” plot following audio and radio engineering convention. This is from the raw time-ordered data calibrated into power units using load curves and Eq. 5.37. There is a variable $1/f$ knee, which appears as a jagged line at low frequencies, and tracks weather systems that also produce high $PWV$. The noise increases for frequencies greater than $\sim 25$ Hz because of increased detector noise, then the Butterworth filter rolls off the signal $> 122$ Hz. The line around $105$ Hz is believed to be due to a ventilation fan mounted on the bottom of the MCE electronics. It operates at around $88$ Hz at sea level. This also shows several periods of anomalous noise. November 14, 2007 had wideband noise at $\sim 100$ Hz and $\sim 37$ Hz. This noise returned on December 1 and cause is not currently known.

these modes can be iterated in the mapper to find the best scaling to the correlated noise components. So long as there is no principal component associated with actual intensity measurements, these will only subtract non-optical interference modes such as the row and column correlations. It will be described in future publications.
6.4 Noise diagnostics – normality

The maximum-likelihood mapmaking method described in Sec. 6.2 starts from the assumption that the noise is described by a covariance. To make this practical, we then assume that the noise is also stationary and correlated over short times. Sec. 6.3 describes a family of noise from electrical interference that is localized in time and highly non-normal (it also correlates all columns within a row) and thus invalidates these assumptions.

Aside from electrical interference, there are also drifts from the atmosphere that are correlated over long times and not normally distributed. Treatments for these are described separately in Sec. 6.6, and here we focus on the family of noise that we believe is electrical interference. To detect the electrical noise component, the tests here are applied to short timescales of data where the best estimate of the long timescale drifts have been removed. The noise appears sporadically, and the goal of this section is to develop a quick test for it over these intervals.

To quantify the normality of the noise, we use the D’Agostino and Pearson statistic\(^8\) (D’Agostino and Belanger (1990)), which combines a standardized measure of skewness \(Z(b_1)\) and kurtosis \(Z(b_2)\) of the sample that are each approximately normally distributed with variance 1 and zero mean for a normally distributed input sample. Here \(\sqrt{b_1} = m_3/m_2^{3/2}\) is the skewness measure and \(b_2 = m_4/m_2^2\) is the kurtosis measure, where \(m_k = \sum (x_i - \bar{x})^k/n\) and \(Z\) is a function that standardizes the skew and kurtosis so that normally distributed data produce a zero-mean unit variance normal statistic. The expressions for \(Z\) are large and given in D’Agostino and Belanger (1990). This standardization means that these statistics are insensitive to the variance and mean of the signal, so detectors or period with high “legitimate” normally-distributed noise do not trigger false data cuts. Skewness and kurtosis are then rejected independently with a 32%, 5%, 0.3% false-positive rate by cutting the data on \(|Z| > 1, 2, 3\), respectively. The skewness and kurtosis can be combined into a so-called “\(K^2\)” statistic, \(K^2 = Z(\sqrt{b_1})^2 + Z(b_2)^2\). This is the sum of two \(N(\mu = 0, \sigma = 1)\) variables squared and is distributed as \(\chi^2\) with two degrees of freedom, \(F_{\chi^2}(x; 2)\). Non-normality is rejected with 32%, 5%, 0.3% false-positive by cutting the data on \(K^2 > 2.28, 6, 11.6\), respectively. This statistic is applied to each detector in the array, but it is also useful for cuts to have an aggregate statistic for the normality, which we take to be the median of the \(K^2\) statistic across the array. This gives a single figure for the data quality at a given time.

We find that there are periods where burst noise is significant, and other times where it is negligible.\(^9\) The \(K^2\) normality vs. the rms plane is useful for quality cuts and is shown in Fig. 6.3. While the rms is a good statistic for general noise and large departures, it is much less sensitive to burst noise or other artifacts.\(^10\)

6.5 Noise diagnostics – correlations

The detector-detector correlation matrix \((N_{\text{det}} \times N_{\text{det}})\) is simpler than the full noise covariance, which is \(N_{\text{det}} \cdot N_{\text{TOD}} \times N_{\text{det}} \cdot N_{\text{TOD}}\), but to make it an automated diagnostic, it is desirable to

---

\(^8\)There are many tools other than the on described here to assess normality. The standard Kolmogorov-Smirnov (KS) statistic measures the maximum deviation of the departure of the empirical cumulative distribution function (CDF) from a reference CDF. Two shortcomings are that it is most sensitive to deviations at the center of the distribution and that the reference CDF itself must be estimated from the data unless it is known otherwise – it does not test for normality so much as a departure from an assumed distribution. The Anderson-Darling test improves sensitivity to the tails of the distribution (Stephens (1974)), and Shapiro-Wilk performs the test for a series with a known covariance (Shapiro and Wilk (1965)).

\(^9\)This noise is consistent with RF or electrical problems. In the 2008 season of data taking, more careful shielding should mitigate these problems. We also have a spectrum analyzer based on the Ettus (http://www.ettus.com/) software radio which should help correlate interference with RF emission. Analog Devices Application Note AN-347 (Alan Rich, 1983) is an excellent resource for shielding and grounding considerations.

\(^10\)Because the burst noise generically appears as correlation along all the columns in a given row, the correlation matrix and quality factor described next in Sec. 6.5 are sensitive to burst noise even when the normality statistic is inconclusive.
compress the full correlation matrix to a smaller number of quantities to understand the total set of correlations. Having a compressed quality figure also helps evaluate treatments for detector-detector noise correlations, both in the hardware setup and data analysis.

In this section we extend a classic statistic known as “total correlation” or “multi-information” (developed in detail in Watanabe (1960)) for weakly correlated, multivariate normal distributions. The information entropy produced by the array is defined for arbitrary joint probability distributions of the detectors, and can be used to measure the departure of real data from statistical assumptions.\textsuperscript{11} Two standard figures are the negentropy and multi-information. Negentropy measures the entropy difference \( J = H(x) - H(x_{\text{normal}}) \) between the full entropy and the entropy assuming normality (where \( H(x_{\text{normal}}) \) is the entropy generated by a variable with the same covariance as \( x \) but none of its higher moments). This gives a measure of the amount of information in the data that is not describable by a normal distribution, so is an extension of skew and kurtosis. This statistic is commonly used in “independent component analysis” (ICA) to separate non-normal components.\textsuperscript{12}

The second popular entropy statistic is the multi-information, which measures the difference between the entropy of all the variables taken independently relative to the entropy of the joint distribution of variables. The intuition for this statistic is that correlations reduce the amount of information needed to specify the output of the array of detectors. In the limit that they are fully

\textsuperscript{11}See Kullback (1968) and Cover and Thomas (1991).

\textsuperscript{12}The ICA could be a powerful tool to identify non-normal components in the data, but is the subject of future work.

Figure 6.3: The \( K^2 \) normality statistic/rms plane for all data from the first half of season 1. Above the 5\% significance line in \( K^2 \), the data have a 5\% chance of being drawn from a normal distribution. The rms here is in digital units and is the total power out to the bandwidth defined by the Butterworth filter. There is a clear family of noise which does not produce high rms, but which is significantly non-normal. The rms for these intervals lies roughly between 5000 and 5500 digital units, yet their \( K^2 \) statistics lie above the 0.3\% significance line. Each point is the median \( K^2 \) across the array for a one-minute chunk of data. Further tests weighing the map quality versus the quantity of cut data will be needed, but here we advocate 0.3\% significance, or \( K^2 < 11.6 \).
correlated, then a one TOD suffices to explain the output of the entire array. The information produced by an array of detectors is given by the joint entropy\(^{13}\)

\[
H(X) = -\sum_x p(x) \log_2 p(x) = -\langle \log_2 p(x) \rangle, \tag{6.7}
\]

where the sum is taken over all possible values of all detectors. Throughout, \(p(x)\) is shorthand for the joint probability \(p_{X_1,\ldots,X_{N_{\text{det}}}}(x_1,\ldots,x_{N_{\text{det}}})\) and \(p(x_i)\) for the marginalized probability, the PDF for a single detector \(i\). In the case that the detectors are uncorrelated \(p(x)\) factors into the PDF for each detector and

\[
H(X) = -\langle \log_2 \prod_{i=1}^{N_{\text{det}}} p(x_i) \rangle = -\sum_{i=1}^{N_{\text{det}}} \langle \log_2 p(x_i) \rangle = \sum_{i=1}^{N_{\text{det}}} H(X_i). \tag{6.8}
\]

In the case of statistical independence, the information produced by the whole array is just the sum of the information produced by each detector. We can then find the difference in the total entropy rate of the array from the entropy rate assuming there are no correlations. This defines the multi-information \(C(X)\), where

\[
C(X) \equiv \sum_{i=1}^{N_{\text{det}}} H(X_i) - H(X). \tag{6.9}
\]

This quantifies how much less information the array produces due to correlations, but it is awkward to use in real applications because it requires finding the empirical joint PDF and single-detector distributions over \(32 \times 32\) variables with \(\sim 10^5\) data points. Appendix C.5 describes an approximation to \(C(X)\) for zero-mean normally distributed data with weak correlations and shows that the multi-information (with a convenient normalization) is

\[
Q^2 = \frac{2}{n(n-1)} \sum_{i,k \text{ upper}, i \neq k} \rho_{ik}^2(X), \tag{6.10}
\]

where \(\rho_{ik}(X)\) is the matrix of correlation coefficients at zero lag (each value in the matrix is between \(-1\) and \(1\), or totally anti-correlated and totally-correlated, respectively) and has the simple interpretation as the average correlation coefficient-squared between all pairs of detectors. Lower values are better. This is fast to calculate and interpret.

In the standard ACT analysis package, we provide a function that produces correlation matrices, a quality factor \(Q\), and the average noise level along the diagonal. This is shown in Fig. 6.5 and Fig. 6.4. There are two equivalent ways to plot the correlation matrix depending on the emphasis for the correlations. In one, which we refer to as “row-dominant”, the axes of the correlation matrix increment by row first then column so that the final plot has blocks of correlations of all the rows between two different rows and is shown in Fig. 6.5. The alternate approach is to order by incrementing in rows first then column so that the final plot has blocks of correlations of all the rows between two different columns. Fig. 6.4 shows the column-dominant matrix.

Different filters can be applied before calculating the correlation coefficients to emphasize different types of correlated noise. For example, by applying a passband at moderate frequencies where the atmosphere has structure across the array (\(\sim 0.5\) Hz), the correlation matrix and quality factor quantify the atmospheric correlations. An alternate reduction of the correlation matrix to

\footnote{A detector \(i\) will take on values \(\{x_j\}\) each with probability \(p_i(x_j)\) in a given sample. Here \(\{x_j\}\) is the set of possible DAC values applied to the flux loop feedback, and the sum of the detector’s discrete PDF \(p_i(x_j)\) over all \(\{x_j\}\) is unity. The entropy is also a bound on compression of the data at \(R H(X)\) bits per second for a sampling rate \(R\). Also see Maris et al. (2004); Gaztaña et al. (2001); Gaztanaga et al. (1998).}
6.5 Noise diagnostics – correlations

Figure 6.4: The correlation matrix for data taken at 1970 epoch time 1196986141 in column-dominant ordering. The colors represent the zero-lag correlation between the pair of detectors in that row and column. Column-dominant ordering here means that along the axes of the correlation matrix, the indexing iterates through rows then columns. Lines indicate the division between columns, so that a block delineated by the black lines is the correlation matrix between all the rows in the two columns considered. Only live detectors are included. The top heading “column groups” lists the delineation of the columns. In the 145 GHz array, column 22 failed so this does not appear in the column groups heading, column 31 also has a large number of flagged detectors, so its correlation subarrays have different dimensions. The correlation structure here is due to synchronous pickup of magnetic fields by the SQUIDs. Because the phase of this pickup varies across the columns, this can be either correlation or anti-correlation. The field “stdev” given here is the square root of the median along the diagonal and here is given in units of $\mu$K, but can be power, or digital units depending on the input data. (The value here is high because it is the total power, and includes some residual drift and synchronous pickup.) Here, only the array-wide common mode is removed. If it is left in, it will almost entirely correlate all the detectors. The “quality” field is given by Eq. 6.10 and is just the square root of the average of the correlation coefficient of the off-diagonal terms-squared. The synchronous pickup correlations can be removed effectively using dark detectors.
Figure 6.5: The correlation matrix for data taken at 1970 epoch time 1196986141 in row-dominant ordering. The data are the same as Fig. 6.4, but this emphasizes the row-correlation structure from electrical interference, which shows up as light banding off the diagonal. This can also be removed reasonably well using information from dark detectors, improving the quality by a factor of $\sim 3$. (Averaging along columns is much more effective, but also introduces correlations in the signal.) The synchronous pickup correlation amplitude is stable over the season (for a given pointing), but the row-correlated “burst noise” is more intermittent.
6.6 Modeling noise that is correlated in time

One of the earliest rigorous approaches to estimate the array-wide drift from the atmosphere (and other sources) in the context of mapmaking was developed by Cottingham and Boughn (see Cottingham (1987)), who estimated the B-spline that when subtracted from the data in the time domain gave the lowest variance in the map. Here the drift $b(t)$ is approximated as the sum of amplitudes times basis functions $\beta_i(t)$ through

$$b(t) = \sum \alpha_i \beta_i(t).$$

(6.14)

For reviews of basis spines for signal processing, see Pang et al. (1994); Ustuner and Ferrari (1992); Unser (1999); Unser et al. (1993a,b); Samadi et al. (2004). For uniform knot spacing, the B-spline bases of degree $n$ are successive
Quantifying Sources of Noise and Elements of Mapmaking

The basis functions have finite support which is set by the knot spacing (and base order), usually taken to be longer than or equal to the time of a scan throw so they are less covariant with spatial modes on the sky. 15

The Cottingham method is an outcome of solving the maximum likelihood problem with an additional correlated noise model term $Bx_b$ as

$$d = Mx_m + Bx_b + n,$$

where $B$ are the basis functions and $x_b$ is a vector of the amplitudes. 17 This model of drift is just an extension of the standard linear model as

$$d = (M \quad B) \begin{pmatrix} x_m \\ x_b \end{pmatrix} + n = \tilde{M}x + n. \quad (6.16)$$

This is the form of any linear method for modelling some non-sky term in the data. This has the general solution

$$\tilde{x}_{|\text{ML}} = (\tilde{M}^T\tilde{N}^{-1}\tilde{M})^{-1}\tilde{M}^T\tilde{N}^{-1}d. \quad (6.17)$$

Appendix C.4 describes solutions to the maximum-likelihood problem with an additional model, but to get a reasonable sense of the issues with a correlated noise model, take $	ilde{N} = 1$ so 18

$$\tilde{x}_{|\text{ML}} = (M^T \bar{M})^{-1}M^T d = \begin{pmatrix} M^T \bar{M} & M^T \bar{B} \\ B^T \bar{M} & B^T \bar{B} \end{pmatrix}^{-1} \begin{pmatrix} M^T d \\ B^T d \end{pmatrix}. \quad (6.18)$$

In the limit that the off-diagonal matrices are small, the solutions for $x_m$ and $x_b$ are decoupled and the additional model can be estimated independently of the sky model. This is the case if the projections functions in $M$ and $B$ are mutually orthogonal (the elements of $M^T \bar{B}$ are the dot products of these functions evaluated at the sample times).

It is useful to think of the sky map $x_m$ model that estimates the 2D Fourier amplitudes of the map instead of the values in each spatial pixel. In this case, the functions in the pointing matrix convolutions of the rectangular step Unser (1999),

$$\beta^0(t) = \begin{cases} 1 & -1/2 < t < 1/2 \\ 1/2 & \left| t \right| = 1/2 \\ 0 & \text{otherwise} \end{cases} \quad (6.11)$$

so that degree $n$ is the convolution,

$$\beta^n(t) = \bigotimes_{0}^{n+1} \beta^0(t) \quad (6.12)$$

Order 0 is a rectangle, order 1 is a triangle, and higher orders are progressively smoother “bumps”. We can sample at $m$ data points per knot to coincide with the data by defining

$$b^n_m(k) = \beta^n(k/m). \quad (6.13)$$

15 We applied a similar method to the analysis of data from a prototype camera, and A. Hincks has developed a full Cottingham-Boughn mapmaker for beam analysis and planet amplitudes. This provides a parallel pipeline to test conclusions from the main maximum-likelihood mapper. Sec. 5.2.2 shows a beam map produced by this pipeline. This is also implemented in the maximum-likelihood mapmaker, but we have found that a more detailed model of structure within the array is needed to fully remove model atmospheric contamination.

16 See Keihänen et al. (2005) for related literature where an additional prior is placed in the basis amplitudes, and other basis functions are considered.

17 Appendix C.4 shows that the maximum likelihood solution for $x_b$ is a generalization of the Boughn-Cottingham method for arbitrary covariance and emphasizes that 1) B-splines enforce an implicit prior on the drift spectrum from the transfer function of the B-spline filter and 2) both the B-spline and optimal filter can be solved iteratively with an approximate covariance.

18 T. Marriage, private communication.
are oscillatory functions with frequencies roughly characteristic of the scale on the sky. If the basis functions for the drift model have some width $\Delta t_i$, then the overlap with an oscillating function of frequency $f$ becomes small when $f > 1/(\Delta t)$. For B-splines, $\Delta t$ is related to the knot spacing and order of the spline. This is just the recognition that the basis model behaves like a high pass filter for frequencies $f > 1/(\Delta t)$. Lower frequencies are absorbed in the drift estimate.

On larger scales where the models can be indeterminate, having more data in independent circumstances (over long times, crosslinked) will shrink the overlap between the sky model projections functions and (temporally isolated) drift model projection functions. If there is only one pass through a region, the large scales are completely indeterminate, while as more independent observations are added, the probability that a particular drift remains consistent with a signal from the sky becomes lower and lower. If there is insufficient data, then it is necessary to impose a prior on the sky model that it have no modes that overlap with the drift model.\textsuperscript{19}

While this discussion has been oriented towards understanding the behavior where $B$ is the set of basis splines, the concerns are still representative of more generic models. For example, to map a frozen turbulence layer on a moving screen there is simply a function $S$ that projects those model parameters, $x$, into the time domain. The conclusions are analogous to the B-spline, where the goal is to acquire data in a wide enough set of circumstances that the projection functions for the sky and atmosphere models overlap less. In the case of the atmosphere, this suggests that revisiting the same astronomical sky at different times (through cross-linking) is advantageous.

The discussion of B-splines also emphasizes that a model for a correlated noise component can be thought of as a filter. Here we would like to comment briefly on a very general way of thinking about a filter for drift components and propose a simple algorithm. The most general way to treat drift is to model an arbitrary time-domain component $s$ as\textsuperscript{20}

$$d = Mx_m + s + n. \quad (6.19)$$

To prevent this model from being degenerate with the sky signal, impose a prior on $s$ (which looks like a penalty in the $\chi^2$)

$$\chi^2 \propto \Delta d^T K^{-1} \Delta d + \hat{s}^T S^{-1} \hat{s}. \quad (6.20)$$

Where $\Delta d = d - Mx_m - \hat{s}$. The maximum likelihood estimate of $\hat{s}$ (where the prior is taken to be the known spectrum of the drift) is the Wiener filter where the “signal” spectrum is the drift spectrum,

$$\hat{s}(\omega_i) = \frac{\tilde{S}(\omega_i)}{\tilde{S}(\omega_i) + K(\omega_i)} \tilde{d}(\omega_i). \quad (6.21)$$

Here, $S(\omega_i)$ is the drift spectrum (prior) and $K(\omega_i)$ is the noise spectrum. (Subtleties of how to find the noise spectrum are described in Appendix C.4.) For slow drifts, this is just a lowpass filter that extracts the drift component from the time-ordered data. This method has the advantage that a prior for the drift spectrum can be chosen to coincide with the measured drift spectrum, making the filter optimal. This was also the intent with the B-spline filter, but there the prior that the drift does not have high frequencies is put in by hand through the basis functions, which cannot represent those frequencies. An optimal filter can be implemented iteratively, and the simplest implementation is to apply the Weiner filter to the data to get a first-pass estimate of the drift, $\hat{s}_0$, and subtract this from the data. Then one can estimate a map $x_{m,0}$, and project this map 0 back into the time domain and subtract it from the original data. This gives an estimate of the noise component of the TOD with a first pass at the signal component subtracted. This noise estimate can then be Wiener filtered to find $\hat{s}_1$, and a map with that drift subtracted, $x_{m,1}$, and so on. Iterating in this way purifies the signal component out of the noise model and the drift is found optimally based on the known spectrum.

\textsuperscript{19}This is an additional exponential in the likelihood which penalizes $\hat{x}_m$ for representing large-scale modes.

\textsuperscript{20}D. Spergel, private communication.
6.7 Conclusions

This chapter gives an overview of the map estimation problem and several data quality tools to check the inputs of the mapmaker. We find that the normality is a good tool to identify electrical interference and readout problems and can be automated to produce data quality cuts over the season. There are several sources of detector-detector correlation, so we also describe a diagnostic for the total correlation of the detectors and a visual tool to understand those correlations. This has been used extensively to study electrical interference and atmospheric correlations. The multi-information figure presented here compresses the full correlation matrix into number that enables automated cuts for bad weather or electrical noise. These tools can be used to identify bad data or to test the efficacy of noise treatment methods. We also give an overview of models for correlated noise and the relation between some standard drift removal techniques. While the contaminants described here present challenges in the data analysis, there are powerful methods to isolate and remove much of the contamination. These are the subject of future work by the collaboration.
Chapter 7

Conclusion

Physical cosmology has evolved from a field with neither observational handles nor theoretical consensus to a field in which calculations accurate to better than one part in $10^3$ are needed to correctly interpret precise experimental data. On the other hand, there are now many phenomena that are very well-constrained by observations, but cannot be explained by existing physical theory. Cosmology has therefore become a “revolutionary science.” The goal of many recent experiments is to answer some basic questions with confidence: is the spectral slope less than unity?, are there primordial gravitational waves or non-Gaussianity?, is the dark energy equation of state constant?

Measurements of small-scale anisotropies in the CMB will add confidence to answers to pivotal questions about the spectral slope, the dark energy equation of state, and possibly non-Gaussianity. These measurements will also improve our understanding of the standard cosmological model. They will sharpen parameter constraints (in particular for $\Omega_b, \sigma_8$) and check the consistency of predictions about gravitational lensing, the diffuse Sunyaev-Zel'dovich power spectrum, and the abundance and power spectrum of point source contaminants, among others.

Here we have tried to express the excitement about small angular-scale CMB intensity anisotropy measurements and to give a broad view of some of the challenges and considerations. We have also elaborated on recent work in the field and described several new results and methods. The main exportable methods described in the experimental section are 1) a method for tracking the optical loading and PWV using the drift, 2) a model for gain as a function for optical loading and optical response of large arrays, 3) calibration and correlation quality factors for large arrays and noise identification, 4) a response model of the time domain flux lock loop, 5) a relative calibration from the atmospheric drift, 6) a mm-wave emission model for Saturn and its rings, and 7) a message-passing system for commanding instruments remotely. Based on the work with helium recombination here, we propose that the standard model be extended to include continuous opacity from neutral hydrogen, radiative feedback between levels, and transport in the intercombination line. These corrections will be relevant for the current generation of small-scale CMB experiments.

In the next few years, ACT is expected to map approximately 1000 square degrees in 145 GHz, 220 GHz, and 280 GHz. These efforts will produce maps that are separable into large scale structure components (from the Sunyaev-Zel'dovich effect), foreground emission, and the primary CMB. From these data, we will learn more about the early universe and its evolution.
Appendix A

Recombination

A.1 The temperature of matter

During recombination, the temperature of radiation and matter can evolve independently. Adiabatic expansion cools the matter, and scattering with photons drives it closer to the radiation temperature. Here we calculate the matter temperature during helium recombination, and find that its departure from the radiation temperature is negligible.

The matter temperature, $T_m$, departs from the radiation temperature due to adiabatic expansion, Compton cooling, free-free, bound-free, and bound-bound processes (Seager et al. (2000)). Here,

$$ \dot{E} = -\dot{p}V + \text{heat exchange processes}. \quad (A.1) $$

We can then solve for the rate of change of the matter temperature due to expansion and general heating ($\Gamma_Q$) and cooling rates ($\Lambda_Q$) to find

$$ \dot{T}_m = \frac{2}{3k_B x_{\text{tot}}} \left( \frac{\Gamma_Q - \Lambda_Q}{n_H} - \frac{3}{2} \dot{x}_{\text{tot}} k_B T_m \right) - 2HT_m. \quad (A.2) $$

Both $\Gamma_Q$ and $\Lambda_Q$ have units of erg cm$^{-3}$ s$^{-1}$. The term $-2HT_m$ accounts for cooling due to adiabatic expansion, while terms in $\dot{x}_{\text{tot}}$ (which are often neglected) include “heat” processes where the number of particles changes, and with it, the number of kinetic degrees of freedom. We break the temperature evolution to include heat/particle exchange from Compton scattering (“es”), and free-free, bound-free, and bound-bound radiative atomic processes as

$$ \dot{T}_m = \frac{2}{3k_B x_{\text{tot}}} (Q_{\text{es}} + Q_{\text{ff}} + Q_{\text{bf}} + Q_{\text{bb}}) - 2HT_m. \quad (A.3) $$

An individual heat exchange process $Q$ has the general form

$$ Q = \frac{\Gamma_Q - \Lambda_Q}{n_H} - \frac{3}{2} \dot{x}_{\text{tot}} k_B T_m, \quad (A.4) $$

where $\Gamma_Q$ and $\Lambda_Q$ are the heating and cooling rates for that process, and $\dot{x}_{\text{tot}}$ accounts for processes that modify the number of kinetic particles in the gas. The Compton (electron scattering) cooling term is (Seager et al. (2000))

$$ Q_{\text{es}} = -4x_e c \sigma_T a_B T_e^4 k_B (T_r - T_m), \quad (A.5) $$
A.2 The escape probability with continuous opacity and complete redistribution

where we have used the radiation constant \( \alpha_R = \frac{\pi^2 k_B^4}{15 c^3 h^3} \). The free-free contribution to \( \dot{T}_m \) is the integral of the free-free opacity over the radiation temperature blackbody distribution (heating) minus the matter temperature blackbody distribution (cooling),

\[
Q_{ff} = \frac{8\pi}{n_H c^2} \int \nu^2 [N_P(\nu, T_r) - N_P(\nu, T_m)] h \nu \alpha_{ff}(\nu) d\nu, \tag{A.6}
\]

where \( N_P \) is the blackbody distribution, and the length absorption coefficient \( \alpha_{ff}(\nu) \) in units of \( \text{cm}^{-1} \) is (Rybicki and Lightman (1986))

\[
\alpha_{ff}(\nu) = \frac{4 e^6}{3 m_e h c} \sqrt{\frac{2\pi}{3 k_B m_e}} T_m^{-1/2} \nu^{-3} n_e n_{\text{HII}}
\times (1 - e^{-h\nu/k_B T_m}) g_{ff}. \tag{A.7}
\]

The thermally averaged Gaunt factor \( g_{ff} \) is given by Sutherland (1998). In principle, there is an additional correction due to free-free radiation from electron-He II and electron-He III collisions, however these other species are an order of magnitude less abundant than \( \text{H II} \), and deviations from \( T_m = T_r \) are negligible for helium recombination, regardless. Therefore we have not included helium free-free radiation in our matter temperature evolution.

The bound-free contribution to \( \dot{T}_m \) is from energy exchanged in photorecombination and photoionization, and due to the change in \( x_{\text{tot}} \),

\[
Q_{bf} = \sum_i \left[ \frac{\Gamma_{Q,i}}{n_H} \frac{3}{2} \beta_i x_i k_B T_m \right] - \left( \frac{\Lambda_{Q,i}}{n_H} \frac{3}{2} \alpha_i n_e x c k_B T_m \right). \tag{A.8}
\]

The energy exchanged per bound-free process includes the heating due to photoionization,

\[
\frac{\Gamma_{Q,i}}{n_H} = x_i \frac{8\pi}{c^2} \int_{\nu_{th,i}}^\infty \sigma_i \nu^2 N(\nu) h(\nu - \nu_{th,i}) d\nu, \tag{A.9}
\]

and the cooling due to recombination,

\[
\frac{\Lambda_{Q,i}}{n_H} = n_e x_c \frac{8\pi}{c^2} \left( \frac{n_i}{n_e n_c} \right)^{\text{LTE}} \times \int_{\nu_{th,i}}^\infty \sigma_i \nu^2 [1 + N(\nu)] h(\nu - \nu_{th,i}) d\nu. \tag{A.10}
\]

The matter temperature evolution is shown in Fig. A.1. The important conclusion is that the fractional temperature difference \( (T_r - T_m)/T_r \) is negligible throughout helium recombination \( (z \geq 1600) \). This agrees with Seager et al. (2000).

A.2 The escape probability with continuous opacity and complete redistribution

The transport equation developed in Sec. 2.3.4 for complete redistribution through the line and continuous opacity is

\[
\frac{\partial N}{\partial \nu} = \eta_e (N - N_C) + \tau_S \phi(\nu) (N - N_L). \tag{A.11}
\]
Figure A.1: The fractional difference in the matter temperature relative to the radiation temperature, $1 - T_m/T_r$. Note the sign: matter is cooler than radiation at late times because of its different adiabatic index ($5/3$ versus $4/3$). During He I recombination for $1600 < z < 3000$, the fractional difference is $< 10^{-6}$.

It has the general solution:

$$N(\nu) = e^{\Phi_1(\nu)} \left\{ C - \int_{\nu_2}^\nu [N_C \eta_c + N_L \tau_S \phi(\tilde{\nu})] e^{-\Phi_1(\tilde{\nu})} d\tilde{\nu} \right\},$$

where

$$\Phi_1(\nu) = \int_{\nu_1}^\nu [\eta_c + \tau_S \phi(\tilde{\nu})] d\tilde{\nu}. \quad (A.13)$$

Here $\nu_1$ is an arbitrary but fixed frequency, and $\nu_2$ is the frequency at which we set the initial condition. It is convenient to expand Eq. (A.12) into the pieces that depend linearly on the constant $C$ and on $N_C$ and $N_L$:

$$N(\nu) = CI_i(\nu) + N_C IC(\nu) + N_L IL(\nu), \quad (A.14)$$

where the individual profiles are

$$I_i(\nu) = e^{\Phi_1(\nu)},$$

$$IC(\nu) = -e^{\Phi_1(\nu)} \int_{\nu_2}^\nu \eta_c e^{-\Phi_1(\tilde{\nu})} d\tilde{\nu}, \quad \text{and}$$

$$IL(\nu) = -e^{\Phi_1(\nu)} \int_{\nu_2}^\nu \tau_S \phi(\tilde{\nu}) e^{-\Phi_1(\tilde{\nu})} d\tilde{\nu}. \quad (A.15)$$

We integrate this phase space density over the profile, and break the integral into three pieces that emphasize the physical processes:

$$\bar{N} = C\bar{I}_i + N_C\bar{I}_C + N_L\bar{I}_L. \quad (A.16)$$

Here the overbar denotes averaging over the line profile, e.g.

$$\bar{N} = \int_{-\infty}^\infty \phi(\nu)N(\nu)d\nu. \quad (A.17)$$
A.3 Transport in Doppler width-dominated lines with electron scattering

Bringing the outer exponent under the integral in the expression of $\bar{I}_L$:

$$\bar{I}_L = -\int_{-\infty}^{\nu_2} \phi(\nu) \int_{\nu_2}^{\nu} \tau_S \phi(\tilde{\nu}) \exp \left\{ -\int_{\nu}^{\tilde{\nu}} \left[ \eta_c + \tau_S \phi(y) \right] dy \right\} d\nu d\tilde{\nu}. \quad (A.18)$$

We will take the starting frequency to be on the far blue side of the line ($\nu_2 \to \infty$), appropriate for expanding media (Hummer and Rybicki (1992)). With this choice, it is convenient to switch the order of integration and absorb the leading minus sign. The $\bar{I}_L$ line integral is related to the original Sobolev problem by setting $\eta_c = 0$,

$$\bar{I}_L \bigg|_{\eta_c=0} = 1 - \frac{1 - e^{-\tau_S}}{\tau_S} = 1 - P_S. \quad (A.19)$$

Simply, $P_S = 1 - \bar{I}_L(\eta_c = 0)$. We follow Hummer and Rybicki (1985) in then calculating the difference in the line integral with and without continuum absorption (which is related to the probability of absorption between incoherent scattering events, $P_C$):

$$\Delta \bar{I}_L = \bar{I}_L - \bar{I}_L \bigg|_{\eta_c=0} = -\int_{-\infty}^{\nu_2} \phi(\nu) \int_{\nu_2}^{\nu} \tau_S \phi(\tilde{\nu}) \exp \left\{ -\tau_S \int_{\nu}^{\tilde{\nu}} \phi(y) dy \right\} \left[ e^{-\eta_c(\tilde{\nu} - \nu)} - 1 \right] d\nu d\tilde{\nu}. \quad (A.20)$$

The overall value of $\tilde{N}$ is required in order to compute excitation and de-excitation rates. With the boundary $\nu_2 \to \infty$, we have $\Phi_1(\nu_2) \to \infty$ and hence $C \to 0$, while $\bar{I}_i$ remains finite at fixed $\nu_1$. Therefore

$$\tilde{N} = N_C \bar{I}_C + N_L \bar{I}_L. \quad (A.21)$$

Now, if we had $N_C = N_L$, the solution to Eq. (2.36) would be simply $N = N_L$ and hence $\tilde{N} = N_L$. Thus $\bar{I}_C + \bar{I}_L = 1$ and hence

$$\tilde{N} = N_C \bar{I}_C + N_L(1 - P_S + \Delta \bar{I}_L) = N_C(P_S - \Delta \bar{I}_L) + N_L(1 - P_S + \Delta \bar{I}_L). \quad (A.22)$$

A.3 Transport in Doppler width-dominated lines with electron scattering

In the limit that the linewidth is small compared to the characteristic redistribution width for electron scattering and the $\text{H~I}$ photoionization opacity within the line ($\eta_C \Delta \nu_{\text{line}}$) is small, we can derive an approximate solution for the modification to the escape probability with complete redistribution. This is a reasonable approximation in the case of the intercombination lines in He I recombination, where the characteristic width is $\approx 10^2 \text{GHz}$ and the electron scattering redistribution width is several THz.

A key ingredient to this analytic approximation is redistribution kernel for Thomson scattering. The electron velocity components $f_\parallel$ and $f_\perp$ are normally distributed with variance $\sigma^2_{\Delta \nu} = \nu_0^2 k_B T_m/(m_e c^2)$. Thus, for a fixed scattering angle $\chi$, the change in frequency is equal to the
Recombination

The sum of two normal distributions centered around zero, and we write the variance of the photon frequency for a given \( \chi \),

\[
\sigma^2_{\nu}(\chi) = \sigma_D^2 [\sin^2(\chi) + (1 - \cos(\chi))^2]
\]

\[
= 2\sigma_D^2 (1 - \cos \chi). \tag{A.23}
\]

The dipole angular redistribution for electron scattering is

\[
P_{\text{dipole}}(\chi) d\chi = \frac{3}{8} (1 + \cos^2 \chi) \sin \chi d\chi. \tag{A.25}
\]

Integration against the angular redistribution probability gives the redistribution kernel

\[
P_{\nu}(\Delta \nu) \bigg|_{\text{dipole}} = \left\langle \frac{1}{\sqrt{2\pi\sigma_{\nu}(\chi)}} \exp \left\{ -\frac{\Delta \nu^2}{2\sigma_{\nu}(\chi)^2} \right\} \right\rangle_{\chi} \tag{A.26}
\]

\[
= \int_{0}^{\pi} \frac{d\zeta}{2\sqrt{2}\pi\zeta} \frac{P(\chi)}{\sigma_{\nu}(\chi)} \exp \left\{ -\frac{\Delta \nu^2}{2\zeta} \right\}
\]

\[
= \int_{0}^{4\sigma_D^2} \frac{d\zeta}{\sqrt{2}\pi\zeta} \frac{3}{16\sigma_D^2} (1 + \cos^2 \chi) \exp \left\{ -\frac{\Delta \nu^2}{2\zeta} \right\}
\]

\[
= \int_{0}^{4\sigma_D^2} \frac{d\zeta}{\sqrt{2}\pi\zeta} \frac{3}{16\sigma_D^2} \left( \frac{\zeta^2}{4\sigma_D^2} - \frac{\zeta}{\sigma_D^2} + 2 \right) \exp \left\{ -\frac{\Delta \nu^2}{2\zeta} \right\}
\]

The isotropic case is simpler,

\[
P_{\nu}(\Delta \nu) \bigg|_{\text{isotropic}} = \int_{0}^{4\sigma_D^2} \frac{d\zeta}{\sqrt{2}\pi\zeta} \frac{1}{4\sigma_D^2} \exp \left\{ -\frac{\Delta \nu^2}{2\zeta} \right\}. \tag{A.27}
\]

Where we have used \( \zeta = \sigma^2(\chi) = 2\sigma_D^2 (1 - \cos \chi) \) in the second line. The kernel exists in the literature Hummer and Mihalas (1967) in terms of Gaussian and error functions. In Fig. 2.15 we compare the kernel (in physical frequencies \( \Delta \nu \)) for dipole scattering with a with a Monte Carlo calculation. For the analytic work of Sec. A.3 (needed to test the Monte Carlo) it is more convenient to work in the Fourier domain, so instead of \( P(\Delta \nu) \) Hummer and Mihalas (1967) we will concentrate on the characteristic function of the photon \( \Delta \nu \) distribution,

\[
\varphi(k) \big|_{\text{dipole}} = \left\langle e^{ik\Delta \nu} \right\rangle_{\Delta \nu} \equiv \int_{-\infty}^{\infty} e^{ik\Delta \nu} P(\Delta \nu) d\Delta \nu \tag{A.28}
\]

\[
= \frac{3}{16\sigma_D^2} \int_{0}^{4\sigma_D^2} \left( \frac{\zeta^2}{4\sigma_D^2} - \frac{\zeta}{\sigma_D^2} + 2 \right) e^{-k^2\zeta/2} d\zeta
\]

\[
= \frac{3}{16\sigma_D^2} \left[ \frac{R_1(k)}{4\sigma_D^2} - \frac{R_2(k)}{\sigma_D^2} + 2R_3(k) \right], \tag{A.29}
\]

where we have absorbed the Gaussian integrals \( R_1, R_2, \) and \( R_3 \) which are explicitly,

\[
R_1(k) = \frac{16}{k^4} \left( 1 - 2\sigma_D^2 k^4 + 2\sigma_D^2 k^2 + 1 \right) e^{-2k^2\sigma_D^2},
\]

\[
R_2(k) = \frac{4}{k^2} \left( 1 - 2\sigma_D^2 k^2 + 1 \right) e^{-2k^2\sigma_D^2},
\]

\[
R_3(k) = \frac{2}{k^2} \left( 1 - e^{-2k^2\sigma_D^2} \right). \tag{A.30}
\]
A.3 Transport in Doppler width-dominated lines with electron scattering

\[ \omega(k) \bigg|_{\text{isotropic}} = \frac{1}{4\sigma_D^2} \int_0^{4\sigma_D^2} e^{-k^2\zeta/2}d\zeta \]

\[ = \frac{1}{4\sigma_D^2} \frac{2}{k^2} \left( 1 - e^{-2k^2\sigma_D^2} \right) \]

Note that for an isotropic distribution, the result can be expressed very concisely as \( \omega(k) = (4\sigma_D^2)^{-1} R_3(k) \). Note that these require caution numerically for small \( k \) because of the near-cancellation of several terms. This can be remedied by using a high-order expansion in small \( k \) (the first order cancels and leaves divergent terms in \( k \to 0 \)). We also note that while the relativistic electron scattering kernel is in general asymmetric about \( \Delta\nu = 0 \), in the Thomson scattering approximation considered here the kernel is symmetric, leading to all-real \( \omega(k) \).

To find the escape probability using an analytic method, one generally solves for the difference between the phase space density of radiation in equilibrium with the line and the actual radiation continuous opacity and electron scattering is

\[ \frac{\partial N}{\partial (\Delta\nu)} = \eta_C [N(\Delta\nu) - N_C] + \tau_s \phi(\nu)[N(\Delta\nu) - N_L] \]

\[ + \eta_c \left[ N(\Delta\nu) - \int p(\Delta\nu - \Delta\nu')N(\Delta\nu')d\Delta\nu' \right]. \]  \hspace{1cm} (A.33)

In the approximation that \( \phi(\Delta\nu) \to \delta(\Delta\nu) \), transport has the solution (see Switzer and Hirata (2008b))

\[ N_L - \bar{N} \approx \frac{1}{\tau_S} (N_L - N_C) \frac{2}{q + \coth(\tau_S/2)}, \]  \hspace{1cm} (A.34)

giving the modified escape probability

\[ P_{\text{esc}} = \frac{2}{\tau_S[q + \coth(\tau_S/2)]}, \]  \hspace{1cm} (A.35)

Where,

\[ q = 2 \frac{\int_0^\infty k^2 + \{\eta_c + \eta_e[1 - \omega(k)]\}^2 d\nu}{\pi} \int_0^\infty \eta_c + \eta_e[1 - \omega(k)] d\nu. \]  \hspace{1cm} (A.36)

Taking the limit that \( \eta_e \to 0 \), one can see that \( q \to 1 \), regardless of the continuum opacity, \( \eta_c \). Through the use of hyperbolic function identities the \( q = 1 \) case can be shown to give the Sobolev result, \( P_{\text{esc}} = (1 - e^{-\tau_S})/\tau_S \). This is by construction of the approximation – for an indefinitely thin line, finite continuous opacity cannot affect transport “within” the line. We evaluate Eq. (A.36) for \( q \) using a 15-point Gauss-Kronrod rule, using the closed form solution for \( \omega(k) \). The value \( q \) is shown as a function of \( \eta_c \) for several choices of the continuum differential optical depth in Fig. A.2.

Note that the analytic method derived here is only applicable when continuous opacity does not act within the optically thick width of the line. In principle, if the intercombination lines remain narrow compared to \( \eta_C^{-1} \) until \( z < 1800 \), then \( \text{He I} \) will have nearly finished recombination and the analytic method of Eq. (A.35) would be appropriate (even though it is technically incorrect after that point, it is irrelevant). There is significant departure from the assumptions built into Eq. (A.35) starting at \( z \sim 2000 \). Thus in the level code we use an interpolated grid of probabilities from the Monte Carlo, and the method here should only lend confidence to the Monte Carlo result for \( z > 2000 \), where electron scattering matters most.
Figure A.2: The quantity $q - 1$ (defined in Eq. (A.36)), which quantifies the departure from the Sobolev theory due to electron scattering as a function of the electron scattering differential opacity, for several values of the continuum depth for $2^3 P_0 - 1^1 S$ at $z = 2500$, which sets the electron temperature.

### A.4 Fluid description of the baryons

In this appendix, we investigate whether the treatment of the baryons as a single fluid is adequate for the investigation of the peculiar velocity field. In particular, show that a single fluid treatment accurately describes Silk damping. Silk damping then proceeds as in the usual picture at all scales ($k_D \sim 0.37 \text{ Mpc}^{-1} < k < (k_{fs} \sim 2 \text{ Mpc}^{-1})$ at $z = 2000$ (with smaller scales evolving in this range at earlier times Gopal and Sethi (2005)). Below the photon free-streaming scale (the mean free path of photons), only electromagnetic interactions can influence the baryon velocity moment. We will consider the possibility of charge separation on these scales and show that it is a negligible contribution to the velocity structure.

There are five constituents of the baryonic plasma during the recombination epoch: the electrons, ions (H$^+$, He$^+$, He$^{2+}$), and neutral species (He). (During helium recombination there is very little neutral H, and its velocity structure is not relevant because its only significant role is to provide continuum opacity.) These species acquire and exchange momentum through several processes (Hannestad (2001)): 1) radiation pressure, which is only significant for the electrons, by mass, 2) electric fields due to charge separation, 3) collisions, and 4) photoionization/recombination, which can switch particles between the He and He$^+$ constituents.

Siegel and Fry (2006b) have considered electric fields due to charge separation and find that they efficiently prevent the electron and ion densities from departing significantly from each other. In particular, during recombination ($z \sim \text{few} \times 10^3$) they find [See Siegel and Fry (2006b), Eq. (24)]

$$\theta_q \equiv \theta_i - \theta_e = -\frac{\sigma_T m_p c \rho_\gamma}{3\pi e^2 \rho_\nu} \frac{d}{dt}(\theta_\gamma - \theta_b);$$

(A.37)

in their analysis the ions ($i$) consisted entirely of protons but the addition of some He II or He III should cause no qualitative changes. (It should result in the ion expansion $\theta_i$ being replaced by the charge-weighted expansion of all the positive ions, $\sum_j q_j \theta_j n_j/n_e$, where $q_j$ is ion charge and $n_j/n_e$ is a number density ratio.) This equation results from a balance of the radiation pressure on the electrons, which is the driving term in their separation from the ions, with the electrostatic force that seeks to eliminate bulk charges. Here $\theta$ is the peculiar expansion, $\sigma_T$ is the Thomson cross section, and $\rho_\gamma/\rho_\nu$ is the ratio of photon to baryon densities. Plugging in the cosmological...
A.4 Fluid description of the baryons

parameters gives

$$\theta_q = -\Delta t_q \frac{d}{dt} (\theta_q - \theta_b).$$  \hspace{1cm} (A.38)

where

$$\Delta t_q = 6 \times 10^{-20} \text{s} \left( \frac{1 + z}{2000} \right).$$  \hspace{1cm} (A.39)

During an acoustic oscillation $\theta_q$ and $\theta_b$ are similar, but both on the order of 10 km s$^{-1}$ or less. Thus so long as $\Delta t_q$ is much shorter than the oscillation period, $\Delta t_q \omega \ll 1$, the electron and ion velocities will be similar.

Next we come to the collisional momentum exchange times. Of interest are the electron-ion collision rate, the collision rates between ions of various species, and the neutral-ion collision rate. The collisional momentum relaxation rate for particle type 1 against particle type 2 (where we will take 1, and 2 to be electron-proton, and He II-proton) for a thermal distribution and small differential drift velocity is given by

$$\nu_{12} = \frac{2}{3\sqrt{2\pi}} n_2 (e^2 Z_1 Z_2)^2 \frac{4\pi}{\mu m_1 v_{\mu}} \ln(\Lambda)$$  \hspace{1cm} (A.40)

where $\Lambda$ is the Coulomb logarithm, $\mu = m_1 m_2/(m_1 + m_2)$ is the reduced mass, and $v_{\mu} = \sqrt{k_B T/\mu}$ Spitzer (1952). For roughly $z > 6000$, the doubly-ionized helium population exceeds the singly-ionized population. The momentum relaxation rate for doubly-ionized helium against protons is four times larger than the He II-proton rate. For simplicity, we will only consider the He II-proton rate in Fig. A.3, for, if it is sufficient to relax He II to the velocity structure of the other baryons, the He III-proton rate must also be sufficient. Note also that the relaxation rate for He II on protons has no dependence on the number density of He II, while the relaxation rate for protons on He II does.

The rates involving neutral He must be considered at $z < 3500$ when He i is present; they are usually slower because they lack the long-range nature of the Coulomb force. The dominant rate is that of resonant charge exchange with He$^+$:

$$\text{He} + \text{He}^+ \rightarrow \text{He}^+ + \text{He}. \hspace{1cm} (A.41)$$

The charge exchange momentum transfer rate is well-approximated by Banks (1966)

$$\nu_{\text{He}^+, \text{He}} = 4.4 \times 10^{-13} n_{\text{He}^+} (2T_m)^{1/2}$$
$$\times \left( 11.6 - 1.04 \log_{10}(2T_m) \right) \text{s}^{-1},$$  \hspace{1cm} (A.42)

where $T_m$ is in Kelvins and $n_{\text{He}^+}$ is in cm$^{-3}$. We have not considered momentum exchange between He and He$^+$ due to photoionization/recombination, or He-proton scattering; if we did then this would only strengthen the conclusion that the momentum exchange rate to the charged fluid components is fast.

The acoustic oscillation frequency at the Silk damping length is given by

$$\omega_D = \frac{ck_D}{\sqrt{3(1 + R)}},$$  \hspace{1cm} (A.43)

where the damping wavenumber $k_D$ is obtained as in Ref. Zaldarriaga and Harari (1995). The damping length $k_D$ runs from $10^{-7} \text{Mpc}$ comoving at $z = 2 \times 10^8$ to 3 Mpc at $z = 1600$. (Before $z = 2 \times 10^7$ the usual computation is not valid because electron-positron pairs increase the opacity.)

We have plotted the momentum exchange rates and compared them to the plasma frequency $\omega_p$ and the acoustic oscillation frequency at the Silk damping length $\omega_D$ in Fig. A.3. The plot runs from $z = 2 \times 10^8$ until $z = 1600$ when helium recombination has completed, for practical purposes. In all cases the relevant collisional rates are many orders of magnitude faster than the acoustic
Figure A.3: Several scales in the recombination plasma relevant for helium peculiar velocities. The fastest here is the plasma frequency, followed by the electron-proton (and then He ii-p) momentum transfer rate. Note that the plasma frequency is significantly faster than both the Thomson rate and the frequency of an acoustic oscillation at the Silk scale. This greatly suppresses the magnitude of the charge separation, see Eq. A.37. In the lower plot, we focus on the region neutral helium evolution ($z < 3500$). The charge exchange momentum transfer rates between He and He$^+$ are much larger than the frequency of baryon acoustic oscillations at the Silk scale during the period of neutral helium recombination. This brings neutral helium into a common fluid with the other charged baryons.
oscillations. This means that for wavenumbers $k^{-1} > 10^{-7}$ Mpc we expect that hydrodynamics is valid, Silk damping will occur, and the exponential suppression of velocity perturbations predicted by the usual treatment is correct.

A more detailed treatment is required in order to understand what happens at scales below $10^{-7}$ Mpc comoving. Our physical expectation is that acoustic oscillations at such scales would also be damped by photon diffusion, with the Silk damping slightly modified by inclusion of positrons, and by the analogous process of neutrino diffusion (since these scales are well within the horizon when neutrinos decouple). Even if this does not happen, the photon mean free path at $z \sim 2 \times 10^8$ is $10^{-10}$ Mpc comoving, so for $k \leq 10^{10}$ Mpc$^{-1}$ the usual Silk damping calculation applies; since these scales are smaller than the Silk damping length they will be exponentially damped. For higher $k$ it is less obvious what happens, but since the neutral He mean free path is much larger than $10^{-10}$ Mpc comoving during He recombination there can be no structure in the neutral He peculiar velocity field at smaller scales.

### A.5 Monte Carlo method: implementation

Monte Carlo radiative transfer methods have been used extensively in the literature (Hirata (2006); Zheng and Miralda-Escudé (2002); Bonilha et al. (1979); Bernes (1979); Caroff et al. (1972); Auer (1968); Avery and House (1968)). We follow an approach similar to Hirata (Hirata (2006)), and use an algorithm to draw atomic velocity distributions that is discussed clearly in Lee (Lee (1977, 1982)). A general atomic scattering process can be represented by the joint probability function of a photon with $\nu_{\text{in}}$ scattering against the atom coherently to produce a photon of $\nu_{\text{out}}$ through the scattering angle $\chi$,

$$P(\nu_{\text{out}}, \chi | \nu_{\text{in}}).$$

In coherent scattering, the photon is emitted with the same frequency it was absorbed in the atom's rest frame, at some atomic transition with energy $\nu_0$. Conservation of energy gives the change in velocity of the atom along directions parallel $\parallel$ and perpendicular $\perp$ to the incident photon direction,

$$\delta v_\parallel = (1 - \cos(\chi)) \left( \frac{h\nu_0}{mc} \right)$$  \hspace{1cm} (A.44)
$$\delta v_\perp = \sin(\chi) \left( \frac{h\nu_0}{mc} \right)$$  \hspace{1cm} (A.45)

The change in velocity implies a frequency shift

$$h\Delta \nu = \Delta E = \frac{1}{2} m(v + \delta v)^2 - \frac{1}{2} m|v|^2$$  \hspace{1cm} (A.46)
$$= m(v \cdot \delta v) + \frac{1}{2} m|\delta v|^2$$  \hspace{1cm} (A.47)

Taking each term separately,

$$\Delta \nu_I = \frac{1}{2} \frac{m}{h} |\delta v|^2 = \frac{h\nu_0^2}{mc^2} (1 - \cos(\chi)) = \alpha (1 - \cos(\chi))$$  \hspace{1cm} (A.48)

Then, if we convert the atom's velocity into a frequency $f = \nu_0 v/c$,

$$\Delta \nu_{I_I} = m(v \cdot \delta v) = [v_\parallel \sin(\chi) + v_\parallel (1 - \cos(\chi))] \frac{h\nu_0}{mc} \frac{m}{h} = \left[ f_\parallel \sin(\chi) + f_\parallel (1 - \cos(\chi)) \right]$$  \hspace{1cm} (A.49)

This gives the total change in frequency in the lab frame,

$$\Delta \nu = (f_\parallel + \alpha)(1 - \cos \chi) - f_\parallel \sin \chi,$$  \hspace{1cm} (A.50)
Recombination

where $f = \nu_0 v/c$, for the transition frequency $\nu_0$, and $\alpha = h\nu_0^2/(m_pc^2)$. Throughout, $\parallel$ labels the component of a vector in the direction from which the photon came, and $\perp$ labels the component perpendicular to this direction and in the plane of scattering. Thus, the change in frequency between the input and output states in the lab frame is uniquely specified through $\chi$, $f_\parallel$ and $f_\perp$. These three quantities are stochastic, where $f_\parallel$ and $f_\perp$ depend on the thermodynamics of the gas, and $\chi$ is determined by the quantum mechanical scattering distribution. The distribution of the angle $\chi$ between outgoing states and incoming states for dipole scattering is given by (over the range $0 \leq \chi \leq \pi$) (Hirata (2006)),

$$ P(\chi) d\chi = \frac{3}{8} \left( 1 + \cos^2 \chi \right) \sin \chi d\chi. \tag{A.51} $$

We can scale the atomic velocity relative to the other velocity scale in the problem, the characteristic thermal velocity, $u = v \sqrt{m/(2k_BT)}$. This gives the convenient expression $f_\parallel = (\Delta \nu_D) u_\parallel = \sqrt{2}\sigma_D u_\parallel$, where $\Delta \nu_D$ is the Doppler width in standard notation, and $\sigma_D^2$ is the variance of the Doppler Gaussian. In the perpendicular direction, the distribution of atomic velocities is thermal, $\propto e^{-u_\perp^2}$. Bayes’ rule gives the distribution of $u_\parallel$ based on known distributions as

$$ P(u_\parallel | x_{in}) = \frac{P(x_{in} | u_\parallel) P(u_\parallel)}{P(x_{in})} \tag{A.52} $$

where $H(a, x)$ is the Voigt profile,

$$ H(a, x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{-y^2}}{a^2 + (x - y)^2} dy, \tag{A.53} $$

$x = (\nu - \nu_0)/\Delta \nu_D$, and $a$ is the Voigt width parameter:

$$ a = \frac{\Gamma_{\text{line}}}{4\pi \Delta \nu_D}. \tag{A.54} $$

Typically $a \ll 1$, indicating that the Lorentzian width of the line is very small compared to the Doppler width. Lee (Lee (1977, 1982)) presents a rejection method to randomly draw from this distribution.

When a photon is emitted by an incoherent process – i.e. it reaches an excited level $u$ either by a radiative transition from another excited level, or by recombination – the outgoing photon has no knowledge of the phase space distribution of pre-existing line photons, so the probability distribution function for an outgoing photon is just the Voigt profile. Because the Voigt profile is the convolution of the Doppler Gaussian distribution and the Cauchy distribution, a Voigt-distributed random number is easily implemented as the sum of random numbers drawn from each distribution.

The distance between scatters for small frequency shifts is given by $\ell \approx c H^{-1}(z) \nu_0^{-1} \Delta \nu$ (for some central frequency $\nu_0$ and shift $\Delta \nu$). Over the travel times associated with the escape of one photon, the universe has expanded very little, and the line depth parameters are effectively constant.

In standard Sobolev theory, where the fate of the photon is either that it escapes or is absorbed in an incoherent process, it is clear what is meant by an escape probability. The transport problem for He I with continuum opacity is not as clear-cut. Coherent, incoherent, continuum and redshift processes are all active, so there are several choices about what “escape” and “scatter” mean. As shown in the main text however, there is one specific number we need to know: given a photon was
A.5 Monte Carlo method: implementation

just emitted in the line through an incoherent process, what is the probability \( P_{MC} \) that it will escape through redshifting or continuum absorption before it is absorbed in an incoherent process? The easiest way to measure \( P_{MC} \) with the Monte Carlo is to inject many photons whose initial frequency distribution is the Voigt profile, and follow them until they escape (i.e. redshift out of the line or get absorbed by H i) or are re-absorbed in an incoherent process. Here the range of frequencies simulated is \( \pm 360 \) THz with a bin size of 0.36 GHz. (Using a smaller span or less resolution leads to escape probabilities that are biased high at early times during He i recombination, because absorption can occur far in the wings at early times when the incoherent depth is high and the continuum depth is small. Doubling the boundary and halving the frequency step size does not improve the results, within these tolerances.) The basic steps in the Monte Carlo are:

1. Draw a photon from the Voigt distribution, representing emission from an incoherent process.

2. Draw the optical depth that the photon traverses before scattering from the exponential distribution, \( P(\delta\tau)d(\delta\tau) = e^{-\delta\tau}d(\delta\tau) \). The next frequency where the photon scatters is given implicitly by

\[
\tilde{\tau} = \int_{\nu_{\text{start}}}^{\nu} \left[ (\tau_{\text{inc}} + \tau_{\text{coh}})\phi(\tilde{\nu}) + \frac{d\tau}{d\nu}(\tilde{\nu}) \right] d\tilde{\nu} = \int_{\nu_{\text{start}}}^{\nu} d\tilde{\nu}\eta(\tilde{\nu}), \tag{A.55}
\]

where we have identified the integrand \( d\tau_{\text{tot}}/d\nu \) as \( \eta(\nu) \). This is implemented numerically by choosing frequency bins \( \Delta\nu = \nu_{i+1} - \nu_i = 0.36 \) GHz and determining whether a photon crosses that bin, or is absorbed. If the bins are chosen to be small enough, then the integrand is nearly linear over the bin,

\[
\eta(\nu)|_i = \left( \frac{\eta_{i+1} - \eta_i}{\Delta\nu} \right) (\nu - \nu_i) + \eta_i \quad \text{for} \quad \nu_i < \nu < \nu_{i+1} \tag{A.56}
\]

Integrating this gives a quadratic expression for the fraction \( x = (\nu - \nu_i)/(\Delta\nu) \) of the bin that the photon traversed:

\[
\frac{\eta_{i+1} - \eta_i}{2\eta_i} x^2 + x = \frac{\delta\tau}{(\Delta\nu)\eta_i} \tag{A.57}
\]

This can either be solved quadratically, or by working to first order in small bins. We will use the latter approach for small bins,

\[
x_0 = \frac{\delta\tau}{(\Delta\nu)\eta_i} \quad \text{and} \quad x = x_0 \left( 1 - x_0 \frac{\eta_{i+1} - \eta_i}{2\eta_i} \right). \tag{A.58}
\]

The fraction of the bin traversed indicates whether the photon scattered in the bin, or leaves. If it escaped the bin, then we move the photon to the start of the next bin and goes to step 2. If it traversed less than the whole bin, we find the frequency in the bin where it scatters \( \nu_{\text{scatter}} \), and go to step 3.

3. Draw a uniform random number between zero and \( \frac{d\tau_{\text{inc}}}{d\nu}(\nu_{\text{scatter}}) \) and determine the type of event (incoherent scatter/absorption, coherent scatter, or H i continuum absorption).

4. If the photon coherently scatters, draw the scattered atom’s velocity and angle between incoming and outgoing states and use Eq. (A.50) to find the photon energy after scattering, and go to step 2. If the photon is incoherently scattered by He i, undergoes photoelectric absorption by H i, or redshifts through a pre-defined simulation boundary, start a new photon in step 1.
Recombination

The Monte Carlo procedure leaves one issue open, namely the convergence criterion. We repeat the Monte Carlo until 6400 photons escape from the line, which should enable $P_{MC}$ to be determined to a fractional error of $6400^{-1/2} = 0.0125$. A possible concern with this procedure (or any other in which the number of photons simulated is not fixed before running the Monte Carlo) is that the result could be biased if the convergence criterion depends on the results of the simulation. We addressed this question by replacing the function that decides whether the photon escapes with a random number generator that returns “escape” a known fraction $P_{MC}$ of the time and “no escape” (incoherent scatter) a fraction $1 - P_{MC}$ of the time. This produces a very fast code that allows us to map the distribution of the estimated escape probability $\hat{P}_{MC}$ as a function of the true $P_{MC}$. Across the relevant range of $P_{MC}$ (down to $10^{-6}$) we find that $\hat{P}_{MC}$ has a bias $\langle \hat{P}_{MC} \rangle / P_{MC} - 1$ whose absolute value is $< 0.1\%$, and a standard deviation $\sigma(\hat{P}_{MC}) / P_{MC} \approx 0.0125$.

A.6 Convergence of the methods

Here we check the convergence and accuracy of the numerical methods employed throughout this series of papers. The most significant among these for the overall rate are the level of refinement of the probability grids estimated in the Monte Carlo and their accuracy (which can be assessed easily by resampling) and the accuracy of the atomic level code numerical solution.

Here we consider three refinement and resampling cases: 1) doubling the number of MC photons in the sample in the $11 \times 21$ (redshift by $x_{H\alpha}$) grid of probabilities, 2) resampling the $11 \times 21$ grid, and, 3) refining to a $21 \times 41$ grid. These are shown in Fig. A.4. We also confirm that the feedback iterations have converged (in the sense of giving negligible differences between subsequent iterations) in Fig. 2.4. We also considered the level code step size. By doubling the number of steps through the recombination history, we change the recombination history by $|\Delta x_e| < 3 \times 10^{-5}$.

In this series of papers, we have ignored levels with $n > n_{\text{max}}$. As the matter and radiation temperatures drop, $n_{\text{max}}$ must increase to account for levels that have possibly fallen out of equilibrium. If the net recombination (capture) rate to these high states becomes significant, then truncation at $n_{\text{max}}$ will generally contribute less to the formation rate of the neutral species. Here we simply (roughly) halve $n_{\text{max}}$ to $n_{\text{max}} = 45$ (from $n_{\text{max}} = 100$), to find that the change in $x_e(z)$ is negligible and of order $|\Delta x_e| < 4 \times 10^{-5}$. The contribution to the formation rate of the ground state from the decay of these highly excited levels is greatly suppressed by the feedback of the spectral distortions they generate. Indeed, given the high optical depth in the $n^1P^0 - 1^1S$ lines, a model neglecting feedback over-estimates the contribution of these highly-excited states to $x_e$ by slightly over an order of magnitude.

A.7 Experimental possibilities for detecting the recombination spectral distortion

A.7.1 Introduction

During cosmological recombination, the overall relaxation through atomic transitions will produce photons that distort the Planck spectrum at the level of $1 \times 10^{-8}$ in intensity (see Fig. A.5). A detection of this distortion would provide strong confirmation of some of the fundamental assumptions during the CMB/recombination era. In principle, it also contains information about $\Omega_b h^2$, $T_{\text{CMB}}$, and the primordial helium abundance, and pre-recombination energy release (see Sunyaev and Chluba (2007); Dubrovich (1975); Kholupenko et al. (2005); Chluba and Sunyaev (2008b); Rubiño-Martín et al. (2008)). These effect, in particular $\Omega_b h^2$ (which is manifest as a shift in level) may not be accessible by experiment unless the effect is very dramatic.
A.7 Experimental possibilities for detecting the recombination spectral distortion

Figure A.4: Comparing several numerical convergence issues in the Monte Carlo-estimated $11 \times 21$ (redshift by $x_{\text{HI}}$) grid of escape probabilities. Both doubling number of photons in the sample and resampling the Monte Carlo give corrections of order $< 2 \times 10^{-4}$. Grid refinement is a more significant systematic, roughly $< 4 \times 10^{-4}$, and indicates that log-log interpolation on the coarser grid over-predicts the escape velocity. Halving the step size in the level code results in error $|\Delta x_e| < 3 \times 10^{-5}$.

A measurement could detect the distortion, and possibly provide independent information about the CMB temperature. An improved constraint on $\Omega_b h^2$ would be much more difficult because it is manifest as a roughly shift in distortion intensity, requiring an absolute measurement.

Figure A.5: The spectral distortion to the CMB from hydrogen recombination, from Rubiño-Martín et al. (2006). Transitions between the highly-excited states produce ripples in the CMB spectrum into the radio regime. This appendix describes some of the experimental considerations for detecting this distortion.

A fundamental limit on the sensitivity is due to the variance of the photon occupation number $\sigma_N^2 = N + N^2$, which has a Poisson ($N$) piece and a classical wave ($N^2$) piece. For a single mode detector, the fractional uncertainty in the power receiver after some integration time is (Zmuidzinas
In general, the presence of several sources for $N$ will drive the photon noise more toward wave noise. Choosing $N = N_{\text{Planck}}$ gives a good idea of the inherent difficulty of the measurement, Fig. A.6. This indicates that very long integration times will be required, so that in addition to calibrating instrumental passbands, stability is a significant concern.

The spectral distortion ripples span a wide range of frequencies from hundreds of MHz (from transitions between the high-lying states) to several THz (ultimately from the Ly$\alpha$ transition). This gives considerable latitude to think about a wide range of possible experiments. In the radio range, the suggested experiment would be an array of horns; in the mm-wave regime, large focal plane arrays of detectors can be constructed, possibly as part of an balloon/space borne FTS spectrometer. The sub-mm detection of a Ly$\alpha$ distortion would be very difficult (even though it greatly exceeds the intensity of the CMB) because of the cosmic infrared background.

The temperature sensitivity is proportional to $A\Omega$, which for a single moded receiver is just $\lambda^2$. Thus, increasing the collecting area only serves to shrink $\Omega$ (increasing the resolution in the diffraction limit), but does not increase sensitivity. A multi-moded detectors can exploit larger collector areas to reach higher sensitivity. Because the wave ($N^2$) component of the photon noise is known to scale down by the number of modes observed (Zmuidzinas (2003)). Such a receiver in this category is a large collector with a free-space bolometer behind a band-defining filter. Practically speaking, the number of modes cannot be very large in the bands we are considering. Thus the options are to use a large focal-plane array (Griffin et al. (2002)), or to make an array of many smaller receivers.\footnote{One hope of isolating the spectral signature would be if it were consistent with the CMB angular structure by taking the cross power of several independently-calibrated anisotropy maps. Yet, the sensitivity must increase orders of magnitude to...}

Figure A.6: The integration time to reach 10 ppb sensitivity ($1\sigma$) for a several cases from pure photon noise from $N_{\text{Planck}}$. We also take $\Delta\nu$ for the detection to be 5% of $\nu$. The rapid rise accounts for the fact that there are very few photons to average in the Poisson/Wien tail. Also note that additional mode coupling is only beneficial in the wave-noise dominated limit and additional detection efficiency is only beneficial once Poisson noise becomes important.
A.7 Experimental possibilities for detecting the recombination spectral distortion

There are three families of foregrounds that are relevant here: 1) direct contamination from molecular/atomic sources that have spectral structure, 2) spectrally smooth sources that have spatial structure, 3) sources with variable emission as a function of redshift. Spectrally smooth sources with some spatial structure are contaminants because the beam shape changes with frequency, thus mixing spatial amplitudes with spectral amplitudes. This could possibly place constraints on the beam size of the telescope in relation to the spatial statistics of the foregrounds. The window from 1.4 GHz to a few GHz is likely to be the best from the point of view of foregrounds and sensitivity (the ripples are as small as $10^{-9}$ at that level). We will therefore mostly describe experiments in that range. Higher frequencies and their foregrounds are discussed in Sec. A.7.4

A.7.2 Spectral contamination at low frequencies

In the radio region, the Rayleigh-Jeans temperatures for signal components are

$$T_{sys}(\nu, \theta) = T_{CMB}(\nu) + T_{inst}(\nu) + T_{atm}(\nu, \theta) + T_{gal}(\nu, \theta) + T_{ground}(\nu, \theta) + T_{ex}(\nu, \theta)$$

(A.60)

for the CMB, instrument, atmosphere, galaxy, ground, and extragalactic sources. We have indicated sources that can have some spatial structure by $\theta$.

Sensitivity to a ground component has been suppressed to 6 mK (de Amici et al. (1991)) in absolute experiments. While the absolute level is low, the spectral structure at the 10 ppb level necessary is poorly understood and would depend on the site, ground screens and beam pattern. At $\sim 2.5$ GHz, the atmosphere’s effective temperature is approximately 1 K (Bersanelli et al. (1995); Danese and Partridge (1989)). This is only dependent on precipitable water vapor (PWV) at the level of 1 mK/mm, and has only a $\sim 3$ mK contribution from the wing of the lowest molecular oxygen resonance at $\sim 60$ GHz. Non-resonant molecular oxygen emission thus dominates and has a well-understood, very flat dependence on frequency (Danese and Partridge (1989)). The attenuation is a similarly slow function of frequency and is at the level of 3%. There are two primary concerns with the atmosphere. One is that it has some time-varying spatial structure through, for example, a turbulent layer. This will mix with the spectral structure through the beam’s frequency dependence. The second is that there could be additional processes (hyperfine or other splittings) at a much lower level that contribute spectral structure that are not normally considered. In the lower atmosphere, any additional sources will be pressure-broadened, but the upper atmosphere could contribute fine spectral structure. Barring these concerns, the atmosphere is likely to spectrally be a very flat emitter.

The dominant astrophysical concerns here are galactic emission (synchrotron/free-free/dust, and analogous extragalactic processes), hyperfine transition radiation, radio recombination lines, and molecular lines. A combination of radio survey results (de Oliveira-Costa et al. (2008); Tello et al. (2007)) estimates that the galactic diffuse emission at 2.3 GHz is 0.15 K in the cleanest regions out of the plane (and 0.24 K at the next region in). The approximate spectral slope for temperature is $(\nu/\nu_o)^\beta$, where $\beta \approx -3$ (de Oliveira-Costa et al. (2008); Jones et al. (2001); Paladini et al. (2005)). (Thus by 7.5 GHz, the galactic component becomes $\sim 10$ mK (Kogut et al. (1990)).) While spectrally smooth, galactic emission has a complex structure that will be mixed with the spectrum by the frequency dependence of the beam. This suggests that an experiment may also need to have a mapping aspect using large radio surveys. Galactic recombination lines coincide map a 10 ppb fractional change in either the dipole or primary anisotropies at different frequencies, so these experiments are orders of magnitude more difficult despite the gain in uniqueness of the signal.

3The PWV can be measured with a dipping radiometer on a water resonance – with pressure, this could inform a model of the resonant atmospheric component, and emission from the atmosphere could be isolated by scanning the experiment itself in elevation.

4These would not need the same sensitivity because the mixing of spatial structure to frequency is suppressed by the fractional change in the beam across frequency.
with HI\textsubscript{II} regions near the plane and have typical amplitudes of a few mK (Paladini et al. (2003)). The extragalactic/redshifted radio recombination background is poorly understood, but is also a concern for high-redshift 21 cm cosmology experiments. Hydrogen hyperfine emission represents a reasonable lower bound for the spectral survey. The galaxy produces a large distortion at 21 cm, and redshifted emission (Wyithe (2008); Cen (2006)) is expected to have both spatial and spectral dependence, at the level of few hundred \(\mu\)K (Pen et al. (2008)). Primordial atoms and molecules are expected to produce much smaller, smooth distortions at these frequencies (Schleicher et al. (2008); Dubrovich (1997)), with redshifted CO emission not contributing until (Righi et al. (2008)) \(\sim 8\) GHz \((J = 1 - 0\) is at 115 GHz). CS and NH\textsubscript{3} could also contribute, with their lowest transitions being at 49 GHz and 23.7 GHz, respectively, but would need further study. Primordial \(\gamma\) are expected to exist at the level of 3 mK and have (amplitude) spatial structure related to the first collapsed objects and energy injection would be expected to produce \(\mu\) distortions. These are spectrally smooth, but there will be spatial variation of \(\gamma\), at least at the level of primordial perturbations. TRIS provides a limit on extragalactic sources of \(\sim 20\) mK (Zannoni et al. (2008)) in this frequency range.

Terrestrial foregrounds are also significant in the GHz range and are difficult to treat in a widebandwidth experiment. Such a distortion experiment should therefore be situated in eg. a remote area or mountain valley where lines of sight can be suppressed. A separate spectrometer with fine spectral resolution and less sensitivity may be needed to understand the level of contamination of the main spectrometer bins. There are several more exotic foregrounds, such as Zodiacal light, objects in the solar system and possibly dust processes in the galaxy, or scattering. These warrant further consideration.

### A.7.3 Low frequencies – hardware perspective and calibration

Receiver temperatures in this frequency range are typically a few \(\times 10\) K, with the HEMT amplifier stage alone contributing \(\sim 5\) K. Through judicious cooling and instrument design, the total temperature could be suppressed, but will always be the dominant driver of total integration time to reach a given sensitivity.

The transfer function of antennas and amplifier chains is notoriously hard to determine and stabilize. Initially, there was a hope that digital receivers would improve the spectral response by reducing the dependence of analog components. While the behavior may be well-defined once in the digital domain, the digitization process and electronics leading up to it have a complex frequency response. A typical state of the art 8-Gbps ADC will have passband slopes nearing 1 dB/GHz and peak to peak ripple of nearly 2 dB, or \(10^7\) times larger than the anticipated signal (Okuda and Iguchi (2008)). In addition, the origin of this ripple depends on device physics, and may have temperature dependence or drift in time at a level much larger than the signal.

The ideal calibrator would both shut off the sky signal and produce a calibration signal in the same signal pathway the sky signal is received. A popular covering in absolute experiments (see

---

\footnote{There could be additional spectral/spatial structure from redshifted \(^3\)He hyperfine emission at 8.6 GHz.}

\footnote{In the GHz range (Fonseca et al. (2006)), the main considerations are GSM (1.7 – 1.9 GHz), UMTS (1.9 – 2.0 GHz, 2.1 – 2.2 GHz) microwave ovens (2.4 GHz), and WiFi (2.4 GHz, 5.8 GHz). Satellite interference in this range is not mentioned in literature, but is probably also a consideration.}

\footnote{A good "off the shelf" amplifier (Miteq/AFS cryogenic) will have a noise temperature of \(\sim 70\) at room temperature and \(\sim 20\) at 77 K, while eg., the 21 cm amplifier for Effelsberg has achieved 5.1 K at 1.7 GHz. Final system temperatures between 12 – 20 K (receiver noise only) (Keller et al. (2006)), 90 K (Testori et al. (2001)), 62 K (Fonseca et al. (2006)), 100 K (Zannoni et al. (2008)) have been reported for systems in these bands.}

\footnote{This can be achieved by covering the receiver horn with a calibrated emitter or using a 45\(^\circ\) chopping mirror to move between sky and reference. Correlation receivers can also switch between a reference load and the sky. This alone is ineffective as a calibrator because the passband from the load may be different than the passband to the sky, so it inherently calibrating the wrong signal pathway, and will give erroneous calibration due to electrical differences between the load/sky signal pathways.}
A.7 Experimental possibilities for detecting the recombination spectral distortion

Kogut et al. (1990); Bensadoun et al. (1993); Bersanelli et al. (1994); Mandolesi et al. (1986); de Amici et al. (1991); Sironi et al. (1991); Kogut et al. (2006)) is Eccosorb CR-112 which is known to have emissivity very close to one. It has been found to have ripple between 10 GHz and 30 GHz at the level of $3 \times 10^{-4}$ (Kogut et al. (2004)), and shows significant wiggles across a GHz band, thus leaving serious doubts about calibrating passbands to the level of $10^{-8}$.  

One possibility suggested in Subrahmanyan (2002) is to configure two receivers as an interferometer which nulls the diffuse (monopole) CMB signal. Another technique proposed there is to tune the local oscillator to permute which frequencies enter various receiver channels. Another possibility suggested by L. Page is to use a polarization-sensitive receiver which either rotates or observes the NCP. The average difference in polarization then goes to zero. This defines a similar “response zero”.

For a relative experiment, the emissivity does not have to be near one to a factor of $10^9$, it simply has to be constant at that level. One possibility is to construct a precision emitter with some mild optical depth and low effective temperature which only has non-resonant emission in the GHz range.

Because the experimental beam is a function of frequency, there is also a high premium on modelling and measuring the beam. This suggests that a simple horn which can be profile and simulated very cleanly would be preferable to a large dish receiver.

A.7.4 High frequency experiments

The advantage of high frequencies experimentally is that large focal plane arrays can economically average down noise, and that the photon noise itself can be driven lower by exploiting the large bandwidth associated with the distortion’s ripples. For these reasons, CMB polarization experiments are able to target sensitivity to anisotropy in the tens of $nK$ per beam. Such experiment typically have no sensitivity to angular scales larger than the telescope’s chop because they measure differences in the detector output rather than the absolute level. A possible approach with this family of detector would then be to sweep the passband definition before the detector and minimize the $1/f$ knee frequency.

Fabry-Perot, FTS, and grating spectrometers with incoherent bolometric (Richards (1994); Zmuidzinas and Richards (2004)) detectors have been constructed for sub-mm with sensitivity near the photon background limit, and heterodyne receivers have become the standard for finding high resolution spectra. The main application thus far have been toward planetary and astrophysical molecular emission (Biver et al. (2005); Serabyn and Weisstein (1996); Weisstein (1996)). The beam can also be filled by the Moon, rejecting the astrophysical signal (Stankevich et al. (1970)). The emissivity of rock such as in the regolith is not well quantified (Olhoeft and Strangway (1975); Naugolnaya and Soboleva (1986); Linsky (1973)). Bright spectral sources such as Cas A could define the spectral response, but would require that they be resolved, requiring a large telescope. With any source, some concern would always remain that a $\sim 10 nK$ ripple might not be understood.

This would define at least the receiver passband behind the mixer, but is problematic because the behavior of the analog part of the system is also likely to change significantly as a function of frequency, so such a technique may measure a different passband entirely.

Another consideration is to develop a precision line emitter where the frequency can be controlled and the amplitude inferred from the emitter properties.

The bare S-band EVLA horn (Srikanth and Ruff (2007)) has a beam which is $> 10$ dB down at $10^\circ$ and falls off roughly with $1$ dB/°.
Recently antenna-coupled TES arrays have been fabricated where the band is defined by a superconducting microstrip “stub bandpass” (Vayonakis et al. (2002); Goldin et al. (2003); Dubrovich (1997)).

While convenient experimentally for the large arrays they afford, high frequency distortion detection experiments are likely to have complex foregrounds and calibration/passband definitions. Above a few \( \times 10 \) GHz, molecular lines can be excited, and the Earth’s atmosphere has a complex structure (Weisstein (1996)). Many primordial molecular and atomic transitions are known to contaminate the CMB at high frequencies (Dubrovich (1997)), and molecular regions in the galaxy have complex structure.

### A.7.5 Conclusions

Based on anticipated foregrounds, calibration difficulties and the signal of the spectral distortion at high frequencies, an experiment in the few-GHz range would likely be preferable. A large number of these systems would be required to average down the signal in reasonable time due to both their typical receiver temperatures and the photon noise limits for the measurement. Even after reaching the desired sensitivity figure, *in situ* the passband calibration may strongly limit the possibility of detecting the recombination signature.

A theoretical study of foregrounds (described in Sec. A.7.2) and spatial/spectral mixing at low frequencies would be interesting for both this and other efforts such as the 21 cm structure surveys. Thus, a study of detection statistics subject to different noise sources would be informative in evaluating the practicality of distortion detection experiments.

I would like to thank C. Hirata, N. Jarosik, L. Page and participants in the Max-Planck workshop, “The Physics of Cosmological Recombination” for discussions of the recombination spectral distortion measurement and R. Sunyaev for suggesting the study.

---

14 For example, the distortion signature can be made periodic by transforming from frequency to principal quantum number. Thus, a spectrometer could produce spectra in an effective \( \tilde{n} \), and the Fourier transform of the spectrum would thus have a peak at this periodicity.
A.7 Experimental possibilities for detecting the recombination spectral distortion
Appendix B

ACT Systems

B.1 CCAM heater controller design

The appendix describes the CCAM (the “column camera”, an ACT prototype which tested the first $8 \times 32$ array) thermal control systems. The CCAM hardware is described in Lau (2007). Fig. B.3 gives an overview of the system and Fig. B.1 and Fig. B.2 shown the front panel and connections inside the box.

The front panel has 6 IO groups: so called “bypass” BNCs, current monitor BNCs, external BNCs, $2 \times DB37$ connectors for digital relay control, one DB37 connector for current monitor outputs, and a DB50 output to the dewar. Table B.1 gives the pinout of the dewar DB50 connector, and the allocation of heater drivers to channels in the dewar. There are a total of 8 possible drivers across the $4 \times 2$ pulse width modulation (PWM) outputs of the housekeeping DAS crate. The PWM is converted into a programming voltage and sent to the voltage driver card.

The circuit for a single PWM driver, isolator, and filter stage is shown in Fig. B.4. The optoisolators are driven differentially by the UofT crate’s high current high speed SN55451BJG driver, thus the ground reference for the driver is decoupled from the heater box ground. The output from the optoisolators and inverters is then filtered by Avens Signal Equipment Corp. AMLP8L30Hz or AMLP8L120Hz and their output voltages program the heater driver voltages. These are 8 pole Bessel filters and give negligible 15 kHz carrier contamination. There are two cards containing four filters each, the first one containing the AMLP8L30Hz, used for the pump heaters and detector servo, and the AMLP8L120Hz used for the 0.6 K baffle, 4K optics, and calibrator driver, which have faster cryogenic time constants. We have found that a servo rate of 10 Hz is adequate for most operations. Also, because the calibrator has only been used for relative responsivity measurements and not detector time constant tests (its internal time constant is not known), the step response of the 120 Hz filter is negligible.

Digital output signals on the UofT card are configurable as optoisolated (PS2501L-4) inputs or outputs. The core of the UofT card’s control and data processing is handled by the Analog Devices SHARC DSP Microcomputer Family ADSP-2106x. This receives high-level commands from the housekeeping machine in the equipment room. The digital IO on these cards is specified in Table B.4.

Aboobaker (2006) describes the heater driver cards, so here we give a brief summary. The programming voltage is not linear in PWM for duty cycles below 10%, so cannot reasonably ramped to zero. The first stage of the heater card subtracts an offset from an AD580 reference. This also permits slightly negative voltages, so an active clamp sets negative programming voltages to zero.

---

1 The disadvantage of adding the optoisolation is that the turn-off time is longer than the turn on time. For low (<10%) and high (>90%) duty cycle, this means that the filtered response becomes non-linear as a function of the duty cycle.
### B.1 CCAM heater controller design

**Figure B.1:** Front of the CCAM heater controller box. Each heater driver card has a bypass output and a current monitor output. (BNCs on the left side) BNCs on the right half of the box are for auxiliary inputs and outputs, and heat switch outputs. Above each bypass BNC is an LED which indicates whether that heater driver’s relay is turned on. If the relay is on, then the bypass BNC will read the voltage applied to the heater. If the relay is off, the BNC connects directly to the dewar so external voltage supplies can be connected.

**Figure B.2:** Inside of the CCAM heater controller box. Functions in the controller are divided into a set of vector cards and backplane. From left to right, these functions are, [1]: optoisolation of the 8 pulse width modulation (PWM) outputs from the University of Toronto crate; [2, 3]: 2 filter cards with 4 Avens 30 Hz 8-pole filters each; [4]: a heat switch controller card; [5,6,7,8,9,10,11]: three high power heater controller cards (for warming the adsorption pumps) two low power cards (for controlling the 300 mK and 600 mK stages) one high power card to drive the 4 K optics temperature, and a third low power card to drive the calibration bolometer. To accommodate heater line allocations in the dewar, the front panel DB50 connector going to the dewar is broken out into lines which can be plugged into the appropriate heater controller sockets. These connections are shown on the bottom right with yellow labels.
Figure B.3: Block diagram of the CCAM heater controller systems, for one heater. There are 8 heaters and 12 (8 heater driver + 4 constant voltage heat switch driver) relays controlled by the University of Toronto (UofT) crate card, originally designed for BLAST. The isolator is a single card that handles all 8 incoming pulse width modulation (PWM) signals. It routes the signal to two filter cards containing 4 30 Hz 8-pole low-pass Bessel filters. Each heater driver is on a separate removable card in the vector backplane and its output is programmed by the voltage from the filter card. The BLAST-bus PCI (BBCPCI) card is in the housekeeping computer in the equipment room.

Figure B.4: Pulse width modulation (PWM) driver, isolator and filter which program the heater driver outputs.
B.1 CCAM heater controller design

before being sent to the driver. Low power channels drive the heaters based on this voltage using a generic op-amp, while the high-power heater drivers are based on the single-chip Apex Microtechnology PA74 power amplifier. The positive supply of the PA75 is provided by a +28 V supply and the negative supply is generated on-board by a 7908, at -8 V. The relay switching circuit is shown in Fig. B.5. ² A significant problem early on in this design was the interconnect between the PWM driver in the UofT card and isolator stage. The PWM driver also had to drive the stray capacitance ³ in the cable connecting the digital electronics to the housekeeping. The solution was to use RG174 (31 pf/ft) for the PWM channels.

Figure B.5: Relay circuit which limits the output of the heater drivers to the dewar and provides a bypass BNC. The bypass provides access to the driver voltage when the relay is energized, and provides access to the heater channel when the relay is off.

The box is standard EIA 19-inch, alodined (conductive) 4U, 14-series enclosure from parmetal.⁴ The front has been modified to include a hinge, which provides access to the cards while the box is in the instrument rack. Inside, a vector plugboard card cage stands off from the box on a styrofoam mount. Each heater driver is implemented on a 4112-4 Vector plugboard with a common pinout; see Table B.2. This allows channels to be swapped out and tested or replaced (or interchanged) easily if they fail. Cross-card wiring is done between plugboard sockets mounted on the back of the card cage.

Power is provided by a 2-U rack with a switched supply and Vicor ripple attenuator modules. Everything but the pump driver is supplied through a 4-pin male circular connector bulkhead (+16.5 V on pin A, GND on pin C and -16.5 V on pin B), and 28 V is supplied through a 3-pin male circular connector bulkhead. These are regulated to ±12 (Lambda LAS1812/LAS1412) in the box, rated at 1.5 A.

Each heater card is equipped with a high-side current monitor based on a INA114 instrumentation amp that measures the drop across a high-side drop resistor.⁵ The slopes of all the heater channels have been calibrated empirically and are listed in Table B.3.

²The voltage drop through the LED is 1.8 V, plus the voltage drop through the saturated transistor give a roughly 10 V drop across the resistor and relay. The Axicom/Tyco V23026A1001B201 relay’s resistance is 370 Ω, giving a 4.8 V, enough to activate the relay. The total current is < 16 mA, while the LEDs are rated to 30 mA (max) and the 1N4004 forward current is well above that.
⁴http://www.par-metal.com
⁵Because these are configured as high side, their inputs have dividers to bring the common mode within the operating range of the op-amps.
Table B.1: Heater driver allocations and parameters. \( a_0 \) and \( a_1 \) are the coefficients to the linear response of the driver output voltage as a function of PWM duty cycle \( D \), \( V = a_1 D + a_0 \). There are two groups of PWM controllers (one for each UofT card) and PnGm denotes PWM number \( n \) on card \( m \). \( V_{\text{max}} \) specifies the maximum output voltage before passing through a series resistor into the dewar, \( R_{\text{tot}} \) is the total resistance of the heater and connections through the dewar, and \( R_{\text{heater}} \) is the resistance of the heater resistor sunk to the stage. Because of heater allocations, only one calibration emitter can be driven at a time in CCAM. The second is a spare that can be switched into the calibration emitter P3G2 driver. The heater controller supports an additional card whose drive is set by P4G2 and relay bit P1 0x08. The heat switch controller card also has a spare 3.7 V supply, set by relay bit P2 0x80. The CCAM CBOB provides 6 heater channels which are not used, these are H9:{[23, 6]}, H11:{[49, 48]}, H13:{[37, 36]}, H15:{[30, 13]}, H18:{[21, 4]}, and H20:{[45, 44]}, where the quantity in braces lists the \{signal, return\} on the dewar DB50 heater connector.

<table>
<thead>
<tr>
<th>Heater line</th>
<th>channel</th>
<th>pins {s, r}</th>
<th>( V_{\text{max}} ) (V)</th>
<th>( R_{\text{tot}} ) (Ω)</th>
<th>( R_{\text{heater}} ) (Ω)</th>
<th>( a_0 ) (V)</th>
<th>( a_1 ) (V)</th>
<th>Relay bit</th>
<th>PWM num</th>
</tr>
</thead>
<tbody>
<tr>
<td>(^3)He Pump FB1</td>
<td>H0</td>
<td>{31, 28}</td>
<td>28</td>
<td>246.2</td>
<td>200</td>
<td>-4</td>
<td>0.32</td>
<td>P1 0x02</td>
<td>P2G1</td>
</tr>
<tr>
<td>(^4)He Pump FB1</td>
<td>H1</td>
<td>{25, 22}</td>
<td>28</td>
<td>257.8</td>
<td>200</td>
<td>-4</td>
<td>0.32</td>
<td>P1 0x01</td>
<td>P1G1</td>
</tr>
<tr>
<td>(^4)He HS FB1</td>
<td>H2</td>
<td>{19, 2}</td>
<td>3.7</td>
<td>10060</td>
<td>10 k</td>
<td>x</td>
<td>x</td>
<td>P1 0x20</td>
<td>HS2</td>
</tr>
<tr>
<td>4 K Optics</td>
<td>H3</td>
<td>{47, 46}</td>
<td>28</td>
<td>192.4</td>
<td>150</td>
<td>-4</td>
<td>0.32</td>
<td>P2 0x02</td>
<td>P2G2</td>
</tr>
<tr>
<td>cal. emitter 1</td>
<td>H4</td>
<td>{41, 40}</td>
<td>3.2</td>
<td>626</td>
<td>(~590)</td>
<td>-0.4</td>
<td>0.036</td>
<td>P2 0x04</td>
<td>P3G2</td>
</tr>
<tr>
<td>300 mK</td>
<td>H5</td>
<td>{35, 34}</td>
<td>3.2</td>
<td>200</td>
<td>100</td>
<td>-0.4</td>
<td>0.036</td>
<td>P1 0x08</td>
<td>P4G1</td>
</tr>
<tr>
<td>(\mathbf{B}) coil</td>
<td>H6</td>
<td>{32, 15}</td>
<td>x</td>
<td>2794</td>
<td>2.7 k</td>
<td>x</td>
<td>x</td>
<td>N/C</td>
<td>AUX BNC</td>
</tr>
<tr>
<td>(^4)He Pump FB2</td>
<td>H7</td>
<td>{29, 12}</td>
<td>28</td>
<td>251.9</td>
<td>200</td>
<td>-4</td>
<td>0.32</td>
<td>P1 0x04</td>
<td>P3G1</td>
</tr>
<tr>
<td>(^4)He HS FB2</td>
<td>H8</td>
<td>{26, 9}</td>
<td>3.7</td>
<td>10070</td>
<td>10 k</td>
<td>x</td>
<td>x</td>
<td>P1 0x40</td>
<td>HS3</td>
</tr>
<tr>
<td>cal. emitter 2</td>
<td>H10</td>
<td>{20, 3}</td>
<td>x</td>
<td>239.3</td>
<td>(~170)</td>
<td>x</td>
<td>x</td>
<td>N/C</td>
<td>N/C</td>
</tr>
<tr>
<td>600 mK</td>
<td>H12</td>
<td>{43, 42}</td>
<td>3.2</td>
<td>148.4</td>
<td>100</td>
<td>-0.4</td>
<td>0.036</td>
<td>P2 0x01</td>
<td>P1G2</td>
</tr>
<tr>
<td>(^3)He HS FB1</td>
<td>H14</td>
<td>{33, 16}</td>
<td>3.7</td>
<td>10060</td>
<td>10k</td>
<td>x</td>
<td>x</td>
<td>P1 0x10</td>
<td>HS1</td>
</tr>
<tr>
<td>N/C</td>
<td>N/C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>P2 0x08</td>
<td>P4G2</td>
</tr>
<tr>
<td>N/C</td>
<td>N/C</td>
<td>3.7</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>P1 0x80</td>
<td>SPARE HS</td>
</tr>
</tbody>
</table>
B.1 CCAM heater controller design

Table B.2: Vector plugboard contacts for the CCAM heater driver cards.

<table>
<thead>
<tr>
<th>Description</th>
<th>Contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>GND</td>
<td>{1, A}</td>
</tr>
<tr>
<td>+12 V regulated</td>
<td>{2, B}</td>
</tr>
<tr>
<td>-12 V regulated</td>
<td>{3, C}</td>
</tr>
<tr>
<td>+28 V (high power driver)</td>
<td>4</td>
</tr>
<tr>
<td>Led</td>
<td>E</td>
</tr>
<tr>
<td>Digital</td>
<td>6</td>
</tr>
<tr>
<td>Digital GND</td>
<td>10</td>
</tr>
<tr>
<td>BNC Bypass</td>
<td>15</td>
</tr>
<tr>
<td>Current Monitor</td>
<td>19</td>
</tr>
<tr>
<td>Current Monitor GND</td>
<td>U</td>
</tr>
<tr>
<td>Output</td>
<td>V</td>
</tr>
<tr>
<td>Output GND</td>
<td>Z</td>
</tr>
</tbody>
</table>

Table B.3: Heater current monitor allocations and calibration where $b_1$ is relates the current monitor output voltage as a function of current, $I(V_{mon}) = b_1V_{mon} + b_0$. The current offset $b_0$ can vary depending on ground configuration and should be calibrated on a case-by-case basis to get accurate absolute currents. Also given are the amplifier gain resistor $R_G$, the high-side measurement drop resistor $R_S$, and the output voltage scaling drop resistor $R_L$. The scaling resistor $R_L$ is in parallel with the heater resistor in the dewar, and scales the power applied. High power cards are not calibrated as carefully because their currents are mostly used as a diagnostic.

<table>
<thead>
<tr>
<th>Heater line</th>
<th>$R_G$ (Ω)</th>
<th>$R_L$ (Ω)</th>
<th>$R_S$ (Ω)</th>
<th>$b_1$ (V/mA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^4$He Pump FB1</td>
<td>845</td>
<td>0</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>$^3$He Pump FB1</td>
<td>845</td>
<td>0</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>$^3$He Pump FB2</td>
<td>845</td>
<td>0</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>300 mK</td>
<td>51.2</td>
<td>995</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>600 mK</td>
<td>51.2</td>
<td>633</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4 K Optics</td>
<td>844</td>
<td>0</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>cal. emitter 1</td>
<td>51.2</td>
<td>200</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table B.4: Allocation of the digital IO DB37 connector on the UofT card. There are three IO bytes split into 2 4-bit sections, 'a' and 'b'. The return for each group in its input/output configuration is listed in the last two columns.

<table>
<thead>
<tr>
<th>Output Name</th>
<th>Bit/number</th>
<th>Pins</th>
<th>in ret</th>
<th>out ret</th>
</tr>
</thead>
<tbody>
<tr>
<td>I/O group 1a</td>
<td>1, 2, 3, 4</td>
<td>{12, 11, 10, 9}</td>
<td>13</td>
<td>31</td>
</tr>
<tr>
<td>I/O group 1b</td>
<td>5, 6, 7, 8</td>
<td>{31, 30, 29, 28}</td>
<td>13</td>
<td>31</td>
</tr>
<tr>
<td>I/O group 2a</td>
<td>1, 2, 3, 4</td>
<td>{26, 25, 24, 23}</td>
<td>27</td>
<td>8</td>
</tr>
<tr>
<td>I/O group 2b</td>
<td>5, 6, 7, 8</td>
<td>{7, 6, 5, 4}</td>
<td>27</td>
<td>8</td>
</tr>
<tr>
<td>I/O group 3a</td>
<td>1, 2, 3, 4</td>
<td>{36, 35, 34, 33}</td>
<td>37</td>
<td>19</td>
</tr>
<tr>
<td>I/O group 3b</td>
<td>5, 6, 7, 8</td>
<td>{18, 17, 16, 15}</td>
<td>37</td>
<td>19</td>
</tr>
<tr>
<td>PWM</td>
<td>1, 2, 3, 4</td>
<td>{1, 20, 21, 2}</td>
<td>3, 22</td>
<td></td>
</tr>
</tbody>
</table>
B.2 MBAC heater controller design

B.2.1 The heater controller

The heater card has a digital section that converts requests from the UofT digital outputs to programming voltages. These voltages are then the inputs of several drivers which are tuned for the desired heater voltage and current requirements. The logic timing diagram is shown in Fig. B.2.1.\(^6\) One design decision has been to always have the firmware send all the values in its registers, regardless of whether the control software has changed those values. Continuously sending the values ensures that the DACs on the driver card are consistent with the values in firmware, making the behavior more robust.

The digital section of the driver comprises buffers and nine 12-bit dual AD8582 DACs. The output of the DAC is scaled from 0 to \(+4.095\) V by the digital inputs. The UofT card sets the digital level value on 12 input bits, and sets 5 address bits. (4+1 to select which DAC to load, and the A or B unit on that DAC) These are committed through a two step process: the UofT card sets the commit bit low and \(10\) \(\mu\)s after the falling edge, the chip select bit is set low for the addressed chip, which loads the value into either the A or B unit's input register. Then \(10\) \(\mu\)s after the chip select is set high again, the load bit is set low. This transfers the value to the DAC register, which then determines the output voltage. These delays are determined by 4 monostable multivibrators on the heater card, and UofT card firmware is only responsible for setting the commit bit low.

![Figure B.6: MBAC heater driver interface timing diagram. Note that both load A and load B are asserted and the selection between A and B occurs with the A/B pin when loading the AD8582 input register.](image)

The heater driver card draws power from the VME backplane through its Din96 connector, and is in an RF-clean environment. Spectrum control Inc. 56-745-005 DB50 connector feedthroughs are used to reject RF (0.8 MHz 3dB point) on all external connectors. The power is filtered with an RF block terminal strip (to clean >MHz RF) and then a kHz-scale filter. The card uses 3 digital bytes from the UofT crate for control, and produces a Din96 which is broken into a DB50 connector on the backplane which is fed through a DB50-DB50 ribbon onto a connector on the back RF block plate on the crate. The address, value, and output allocation of the DB37 digital IO connector is given in Table B.2.3.

\(^6\)The interface of the heater card was developed with A. Hincks, and the heater driver hardware was designed by T. Devlin's group.
B.2 MBAC heater controller design

B.2.2 The current monitor, relay box

The second unit of the heater controller is a current monitor, relay box (CRBox). This is located in a separate 3U rackmount EMI-tight box (see Fig. 3.9). The CRBox has 3 bytes of digitally-addressable relays (which allow the operator to ensure that a heater channel is off) and 24 current monitors which can be used to infer the total power applied to a heater or calibration bolometer and diagnose errors in the current paths in dewar’s heater lines. The RF-cleaned voltage supply of the analog VME crate is broken into a 4-pin circular connector which supplies power to the CRBox (+16.5 V on pin A, GND on pin C and -16.5 V on pin B).

The total current consumption of the CRBox is < 400 mA at 16.5 V input. (100mA for the current monitors and 10mA for each relay that is energized.) The power is regulated in the box down to ±12 V and +5 V by LAS1812/LAS1412 and 7805-series regulators.

The CRBox functions are split across three Eurocard prototype boards. From top to bottom: the switchboard B.2.2, the current monitor board Fig. B.2.2, and the relay board Fig. B.2.2. The relay board receives 18 inputs from the heater driver on a DB50 connector specified in Table B.2.3. Six external lines from BNCs on the front panel are taken as inputs to the 24 relays (18 heater and 6 external).

![Figure B.7: Layout and channel correspondence of the relay board.](image)

![Figure B.8: Layout and channel correspondence of the current monitors on the board.](image)

Current monitors based on the INA114 are configured to read from high-side resistors, or low-side resistors which look at the current returned by the load.\(^7\) All driver output circuits are based

\(^7\)Low-side resistors add resistance to the ground path and fail if current is returned through other paths, while the high-side can detect down-stream faults. The default is to use high-side current monitors, but for channels which overwhelm the input common mode, we use low-side monitors.
Figure B.9: Layout and channel correspondence of the switchboard card in the CRBox. See Fig. figs:topcard for and image of the front of the card. The channels with 4 sockets handle distributing the current to the four shared pump heater CBOB lines. One set of two of these is for the driver output, and the other set of two allows the pumps to be powered by external BNC channels 5 and 6.

All signals in the CRBox are are treated differentially, except for the 18 heater driver channels, which share a ground on the relay board for simplicity.

B.2.3 The high power drivers and fault detection

The 94 mA high power heater for the ¹He charcoal pumps is based on the inverting, trimmed OP07CS followed by a current booster, where the overall gain is -6.6 (the booster is inverting). It is programmed with the output of the DAC circuit, between 0 and 4.095 V. The 94 mA driver is also equipped with an over-current fault detector and lockout, see Fig. B.2.3.

The fault detector tracks the current applied, and if it exceeds the current, plus a margin, relative to the current that should be flowing for that driver voltage and assumed heater resistance. In the case of a current fault, it latches a circuit that dumps the booster current. Thus for any voltage, if the current exceeds a margin of the expected, the driver shuts down, protecting the heater lines in the dewar. Here we analyze the circuit and determine guidelines for modifying the current limiter in the future.

The first unit of the fault detector is a current monitor which is based on an OP07CS configured on the OP97FS, and use the trim circuit shown in Fig. B.2.2. These are either inverting or non-inverting, and in the high current drivers, these source a booster circuit. Table B.2.3 shows the functions, maximum voltages and currents of the individual heater channels.
Figure B.10: The MBAC low power heater, inverting configuration based on the trimmed OP97FS.

Figure B.11: Block diagram of the MBAC high power heater driver. The driver output has a current monitor and a multiplier which convert the current into what the requested voltage should have been for that current to flow. If this exceeds (in the comparator) the actual voltage that was requested (i.e., there is some margin of current more than expected flowing to the dewar), then the lockout latches up and disables the output on the driver's op-amp booster output. This over-current lockout can be disabled by either pushing a switch on the card, or by sending a reset bit from the controller card.

as a differential amplifier which tracks the voltage drop across a 1 Ω resistor in series with the output. The inputs divide the common mode of the 26 V signal down to 10.7 V, which is within the input rails of the OP07CS. The amplifier uses 0.1% tolerance resistors in the drop (read) resistor and divider. The overall gain for $V_C = GCI$ is $G_C = 0.69$ mV/mA for the current readout. The second unit of the fault detector performs $V_E = mV_C + b$, converting the current measurement in volts back to an estimate of what the programming voltage should have been to get the measured current (plus some margin), given the known heater resistance. This circuit is shown in Fig. B.2.3. The voltage at the output is

$$
V_E = V_C R_1 (R_1^{-1} + R_2^{-1} + R_3^{-1}) - V_b R_1 R_3^{-1}
$$

(B.1)

$$
V_D G_D^{-1}
$$

(B.2)

Where the subscript $E$ indicates that this is an estimate of the programming voltage, $V_D$ is the voltage applied to the heater, $V_C$ is the voltage from the current monitor output. Then the desired condition is that $V_E = \text{voltage applied to the heater} = V_D G_D^{-1}$, where $V_D$ is the programming
voltage and $G_D$ is the gain of the driver relative to the programming voltage, 6.6. (Thus, if the programming voltage from the DAC is 4 V, then the heater output is 26.4 V.)

Figure B.12: The $y = mx + b$ scaler to compare the measured current with expected current draw for the heater based on the programming voltage.

If we rewrite this in favor of the current,

$$I = (G_D G_c m)^{-1} V_D - b(m G_c)^{-1}, \quad (B.3)$$

compared to the current one would expect for driver a heater (plus line) with resistance $R_H$,

$$I_E = R_H^{-1} V_D. \quad (B.4)$$

We will assume that the current monitor gain and booster gain will never be changed, but that the heater resistance may change. In this case, one can redesign the overcurrent fault detector for a specific heater resistance as

$$(4.55 \cdot \Omega) R_1 (R_1^{-1} + R_2^{-1} + R_3^{-1}) = R_H, \quad (B.5)$$

Currently, $R_2$ is small compared to $R_1$ and $R_3$, giving the rule of thumb:

$$(4.55 \cdot \Omega) \left( \frac{R_1}{1k\Omega} \right) \approx R_H, \quad (B.6)$$

A Zener diode and a voltage divider provide a $V_b = 94$ mV. This sets the offset, which is just a current margin. The current margin can be designed using the same approximation that $R_2$ is small,

$$\left( \frac{R_3}{1k\Omega} \right) = \frac{V_b}{G_c I_{\text{margin}}}. \quad (B.7)$$

For the current choice $R_1 = 76.8 \, k\Omega$, $R_2 = 1 \, k\Omega$, and $R_3 = 10 \, k\Omega$, the expected heater resistance is 390 $\Omega$, and current margin of $\sim 13$ mA. In general, if the slope $m$ is not tuned for the heater resistance, the expected current and the actual current curves will intersect—if this happens within the active range of the heater, then the heater’s range is reduced. If the heater resistance changes, this scaling should change also.

The third unit, the lockout, is a comparator on the programming voltage and the estimate of the programming voltage for the measured current. If the current margin is exceeded, the comparator output latches the fourth unit. The fourth unit is an op-amp configured as a flip-flop latch which can be reset with a switch on the front panel, or through an address request to bit 16 of the M74HC154M1R 4-16 addressing multiplexer. If the output of the flip-flop is high, it closes a 2n2222 power transistor which sinks the current before going to the dewar. This is controlled in software an reset before each cryogenic cycle.
B.2 MBAC heater controller design

Modifications to the driver card for compatibility with MBAC electronics

The original heat switch design called for 5.62 kΩ over 10 kΩ inverting amplifier with gain -0.562, giving -2.3 V maximum output. We have replaced the 5.64 kΩ with 10 kΩ to give a maximum value of -4.1 V, and 15.6 kΩ to boost the power for the 2008 season. The calibration bolometer was configured for a maximum voltage output of 10 V, while we would like to limit this to < 1 V. The driver uses a non-inverting amplifier, so it is not possible to bring this into a reasonable range. This has solved temporarily by using diode clamps to limit to 1.8 V, and a software limiter. The main detriment that prevented the use of the calibration pulse driver channels was that the pulse time is set by a loop in the control software which is asynchronous with the timers in hardware. This makes the pulse width unstable. In the end, the calibration pulse was driven with a function generator with two modes of operation. In the 2007 season, it would pulse every ~ 10 minutes for 800 ms, and was regulated to the dewar by a relay. In the 2008 season, the function generator is configured as a one-shot, where one external relay is used to trigger the pulse, and three other external relays select which calibration bolometer should pulse. In the scheduler, the procedure is to, 1) set the selected calibration pulse relay high, 2) set the pulse relay high, 3) set the calibration pulse relay low, 4) set the pulse relay low. This is repeated for each calibration pulse in turn.

All input logic buffers on the original board were ACT-family, and a pull-up was added to make these compatible with the open-collector output of the UofT digital output isolators. The heater card does not have any on-board voltage regulation or filtering. We therefore regulate the +7.5 V from the analog VME crate supply down to +5 V through a 7805, which is located in the VME’s RF filter box. The +5 V logic of the board has a maximum supply voltage of +7 V, so this change must be made for the card to operate and to prevent damage. We added a 0.1 uF capacitor to the overcurrent reset to ground so that the reset would trigger. The op-amps and the booster stages were designed for 15 V while the backplane supplies 16.5 V. The voltage difference should be insignificant for the op-amps, except in changing their rails, but the current sources in the booster section of the circuit are voltage dependent (mildly), so can provide slightly more current than specified.

There are several modifications and additions that could be considered in future versions of the heater card: 1) pull ups on the digital inputs, 2) constant-current heater drivers, (then in the case of a short, constant current would be delivered, rather than the maximum current of the constant voltage source) 3) a DAC reset signal on power-up, and 4) increased DAC resolution 5) on-board current monitors, relays, 6) conservative filters (such as the Avens 30 Hz 8-pole used in CCAM) on the outputs of important heater drivers. The DACs could be set to zero on power-up by adding a monostable multivibrator circuit to issue a reset pulse to the 9 DACs. This would ensure the drivers power up with zero voltage.

It is also very convenient to have the switchboard to connect heater drivers to lines in the dewar. Because of this, a good design would be to put the heater card in a 3U box of its own outside the VME crate. In this case, the optimal configuration would be to have a lower card which has the drivers and current monitors connected to an upper card over a ribbon cable. The upper card would have 24 relays and break out the internal heater channels and some number of external channels into 24 headers. Incoming lines from the dewar would then be broken out and plugged into their respective heater channels. Solid state switching would be a reasonable replacement for the relays to longer lifetime.

Tables and figures

Table B.2.3 gives the relation between the UofT digital outputs and the heater programming channels. Table B.2.3 shows the allocation of heater channels and interconnects and Table B.2.3

---

N. Jarosik, private communications.
Table B.5: Pin-bit correspondence for the heater card controller from the DB37 connector on a UofT digital acquisition card. Here \( \text{ab} \) denotes an address bit, \( \text{lb} \) level bit, and \( \text{cb} \) the commit bit.

<table>
<thead>
<tr>
<th>DB37 pin</th>
<th>IO group</th>
<th>Card bit</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1.1</td>
<td>lb.1</td>
</tr>
<tr>
<td>11</td>
<td>1.2</td>
<td>lb.2</td>
</tr>
<tr>
<td>10</td>
<td>1.3</td>
<td>lb.3</td>
</tr>
<tr>
<td>9</td>
<td>1.4</td>
<td>lb.4</td>
</tr>
<tr>
<td>31</td>
<td>1.5</td>
<td>ab.1</td>
</tr>
<tr>
<td>30</td>
<td>1.6</td>
<td>n/c</td>
</tr>
<tr>
<td>29</td>
<td>1.7</td>
<td>n/c</td>
</tr>
<tr>
<td>28</td>
<td>1.8</td>
<td>n/c</td>
</tr>
<tr>
<td>26</td>
<td>2.1</td>
<td>n/c</td>
</tr>
<tr>
<td>25</td>
<td>2.2</td>
<td>n/c</td>
</tr>
<tr>
<td>24</td>
<td>2.3</td>
<td>n/c</td>
</tr>
<tr>
<td>23</td>
<td>2.4</td>
<td>cb</td>
</tr>
<tr>
<td>7</td>
<td>2.5</td>
<td>ab.2</td>
</tr>
<tr>
<td>6</td>
<td>2.6</td>
<td>ab.3</td>
</tr>
<tr>
<td>5</td>
<td>2.7</td>
<td>ab.4</td>
</tr>
<tr>
<td>4</td>
<td>2.8</td>
<td>ab.5</td>
</tr>
<tr>
<td>36</td>
<td>3.1</td>
<td>lb.9</td>
</tr>
<tr>
<td>35</td>
<td>3.2</td>
<td>lb.10</td>
</tr>
<tr>
<td>34</td>
<td>3.3</td>
<td>lb.11</td>
</tr>
<tr>
<td>33</td>
<td>3.4</td>
<td>lb.12</td>
</tr>
<tr>
<td>18</td>
<td>3.5</td>
<td>lb.8</td>
</tr>
<tr>
<td>17</td>
<td>3.6</td>
<td>lb.7</td>
</tr>
<tr>
<td>16</td>
<td>3.7</td>
<td>lb.6</td>
</tr>
<tr>
<td>15</td>
<td>3.8</td>
<td>lb.5</td>
</tr>
</tbody>
</table>

summarizes the physical parameters of the heaters. Table B.2.3 summarizes MBAC’s heater channels and naming conventions and Table B.2.3 describes the allocation of channels in the current monitor/relay box.

Fig. B.2.3 shows the standard allocation of the DB50 connector, Fig. B.2.3 shows the so called “top card” that connects heater channels to channels in the dewar. Fig. B.2.3 shows the current monitor card front, and Fig. B.2.3 shows the relay card.

![DB50 connector channel allocation standard](image)

Figure B.13: DB50 connector channel allocation standard used throughout (exceptions are specified). Black indicates the signal lines.

### B.3 The pressurized drive systems

The reliability of drives is known to decrease at altitude, so for the critical systems we use a system of pressurized drives. These break either SATA or AT-style cables through hermetic
Table B.6: Hardware layout of the heater driver card, connecting output channels to Din96 to op-amp driver to DAC to 4-16 MUX address. This feeds into the pinout for the relay, current monitor box provided in Table B.2.3. Note that the AD8582 DAC has dual outputs, denoted A and B here. The op-amp index on the board is given to simplify finding the channel and its offset trim on the board. Power to the heater card is allocated as (standard for the VME crate): +15 V is on Row1, -15 V is on Row3, +5 is on Row 4, with a common ground on Row2 and Row5. For the high power heaters, +28 V is on Row 32, with separate ground on Row31. Also, note that the description here is of the intended channel in the dewar, but this can be combined arbitrarily in the CRBox to go to any heater channels on the CBOB. Negative voltages correspond to channels with an inverting amplifier.

<table>
<thead>
<tr>
<th>Intended use</th>
<th>DB50 Ch.</th>
<th>Din96</th>
<th>Op-amp</th>
<th>Current</th>
<th>$\Delta V$ (V)</th>
<th>measured range (V)</th>
<th>DAC(chan)</th>
<th>addr(chan)</th>
<th>index</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1$He charcoal-1</td>
<td>Ch1</td>
<td>Row29</td>
<td>U30</td>
<td>16 mA</td>
<td>18.2</td>
<td>0-18.55</td>
<td>U13:A</td>
<td>8:A</td>
<td>17</td>
</tr>
<tr>
<td>$^4$He charcoal-2</td>
<td>Ch2</td>
<td>Row28</td>
<td>U31</td>
<td>16 mA</td>
<td>18.2</td>
<td>0-18.56</td>
<td>U13:B</td>
<td>8:B</td>
<td>18</td>
</tr>
<tr>
<td>$^3$He heat switch-1</td>
<td>Ch3</td>
<td>Row27</td>
<td>U27</td>
<td>224 uA</td>
<td>-4.1</td>
<td>-3.43</td>
<td>U11:A</td>
<td>7:A</td>
<td>15</td>
</tr>
<tr>
<td>$^3$He heat switch-2</td>
<td>Ch4</td>
<td>Row26</td>
<td>U26</td>
<td>224 uA</td>
<td>-4.1</td>
<td>-3.44</td>
<td>U11:B</td>
<td>7:B</td>
<td>16</td>
</tr>
<tr>
<td>$^4$He charcoal-1</td>
<td>Ch5, Ch6, Ch7, Ch8</td>
<td>Rows 25,24,23,22</td>
<td>U43</td>
<td>94 mA</td>
<td>26.6</td>
<td>14.33-27.07</td>
<td>U5:A</td>
<td>6:A</td>
<td>14</td>
</tr>
<tr>
<td>$^4$He charcoal-2</td>
<td>Ch9, Ch10, Ch11, Ch12</td>
<td>Rows 21,20,19,18</td>
<td>U37</td>
<td>94 mA</td>
<td>26.0</td>
<td>14.32-27.07</td>
<td>U5:B</td>
<td>6:B</td>
<td>13</td>
</tr>
<tr>
<td>$^4$He heat switch-1</td>
<td>Ch13</td>
<td>Row17</td>
<td>U25</td>
<td>224 uA</td>
<td>-4.1</td>
<td>-3.42</td>
<td>U15:B</td>
<td>5:B</td>
<td>12</td>
</tr>
<tr>
<td>$^4$He heat switch-2</td>
<td>Ch14</td>
<td>Row16</td>
<td>U24</td>
<td>224 uA</td>
<td>-4.1</td>
<td>-3.83</td>
<td>U15:A</td>
<td>5:A</td>
<td>11</td>
</tr>
<tr>
<td>300 mK-1</td>
<td>Ch15</td>
<td>Row15</td>
<td>U19</td>
<td>63 uA</td>
<td>-3.1</td>
<td>-3.2</td>
<td>U12:B</td>
<td>4:B</td>
<td>10</td>
</tr>
<tr>
<td>300 mK-2</td>
<td>Ch16</td>
<td>Row14</td>
<td>U20</td>
<td>63 uA</td>
<td>-3.1</td>
<td>-3.21</td>
<td>U12:A</td>
<td>4:A</td>
<td>9</td>
</tr>
<tr>
<td>1 K-1</td>
<td>Ch17</td>
<td>Row13</td>
<td>U28</td>
<td>10 mA</td>
<td>11.3</td>
<td>11.56</td>
<td>U6:A</td>
<td>3:A</td>
<td>7</td>
</tr>
<tr>
<td>1 K-2</td>
<td>Ch18</td>
<td>Row12</td>
<td>U29</td>
<td>10 mA</td>
<td>11.3</td>
<td>11.57</td>
<td>U6:B</td>
<td>3:B</td>
<td>8</td>
</tr>
<tr>
<td>detector 1</td>
<td>Ch19</td>
<td>Row11</td>
<td>U18</td>
<td>3 uA</td>
<td>-3.2</td>
<td>-3.21</td>
<td>U14:B</td>
<td>2:B</td>
<td>6</td>
</tr>
<tr>
<td>detector 2</td>
<td>Ch20</td>
<td>Row10</td>
<td>U17</td>
<td>3 uA</td>
<td>-3.2</td>
<td>-3.21</td>
<td>U14:A</td>
<td>2:A</td>
<td>5</td>
</tr>
<tr>
<td>detector 3</td>
<td>Ch21</td>
<td>Row9</td>
<td>U16</td>
<td>3 uA</td>
<td>-3.2</td>
<td>-3.22</td>
<td>U10:A</td>
<td>1:A</td>
<td>3</td>
</tr>
<tr>
<td>calibrator 1</td>
<td>Ch22</td>
<td>Row8</td>
<td>U23</td>
<td>100 uA</td>
<td>9.8</td>
<td>9.94</td>
<td>U10:B</td>
<td>1:B</td>
<td>4</td>
</tr>
<tr>
<td>calibrator 2</td>
<td>Ch23</td>
<td>Row7</td>
<td>U22</td>
<td>100 uA</td>
<td>9.8</td>
<td>dead</td>
<td>U4:A</td>
<td>0:A</td>
<td>1</td>
</tr>
<tr>
<td>calibrator 3</td>
<td>Ch24</td>
<td>Row6</td>
<td>U21</td>
<td>100 uA</td>
<td>9.8</td>
<td>9.95</td>
<td>U4:B</td>
<td>0:B</td>
<td>2</td>
</tr>
</tbody>
</table>
Table B.7: The CBOB index, resistance, and associated current. Note that the $^3$He charcoal-1 heater allocation is used for the 3K optics, the 300 mK-1 pot heater allocation is used for the 1 K lens servo, detector 1 heater is the detector slab servo, and the detector 2 heater is the 300mK arm. This is the MBAC heater configuration sent to Chile.

<table>
<thead>
<tr>
<th>Dewar name</th>
<th>CBOB index</th>
<th>R (Ω)</th>
<th>socket</th>
<th>$\Delta V$ (V)</th>
<th>I (mA)</th>
<th>sensitivity</th>
<th>$R_{\text{gain}}$ (Ω)</th>
<th>$R_{\text{sense}}$ (Ω)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^3$He heat switch-1</td>
<td>1-5.1</td>
<td>10 k</td>
<td>13</td>
<td>-4.1</td>
<td>-0.41</td>
<td>1 V/mA</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>$^4$He heat switch-2</td>
<td>1-5.2</td>
<td>10 k</td>
<td>14</td>
<td>-4.1</td>
<td>-0.41</td>
<td>1 V/mA</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>$^4$He charcoal-1</td>
<td>1-5.3-6</td>
<td>420</td>
<td>11</td>
<td>26.6</td>
<td>63</td>
<td>0.01 V/mA</td>
<td>5.62 k</td>
<td>1</td>
</tr>
<tr>
<td>3 K lens servo heater</td>
<td>1-6.1</td>
<td>1 k</td>
<td>7</td>
<td>18.2</td>
<td>18.2</td>
<td>0.1 V/mA</td>
<td>511</td>
<td>1</td>
</tr>
<tr>
<td>$^3$He heat switch-1</td>
<td>1-6.2</td>
<td>N/C</td>
<td>9</td>
<td>-4.1</td>
<td>N/A</td>
<td>N/A</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>1 K lens heater</td>
<td>1-6.3</td>
<td>50 k</td>
<td>15</td>
<td>-3.1</td>
<td>0.062</td>
<td>10 V/mA</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>$^4$He 1 K heater-2</td>
<td>1-6.4</td>
<td>1 k</td>
<td>18</td>
<td>11.3</td>
<td>11.3</td>
<td>0.1 V/mA</td>
<td>511</td>
<td>1</td>
</tr>
<tr>
<td>$^3$He 300 mK pot-2</td>
<td>1-6.5</td>
<td>0.4 M</td>
<td>16</td>
<td>-3.1</td>
<td>0.0078</td>
<td>100 V/mA</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>$^4$He 1 K heater-1</td>
<td>1-6.6</td>
<td>1 k</td>
<td>17</td>
<td>11.3</td>
<td>11.3</td>
<td>0.1 V/mA</td>
<td>511</td>
<td>1</td>
</tr>
<tr>
<td>$^3$He charcoal-2</td>
<td>2-5.1</td>
<td>1 k</td>
<td>8</td>
<td>18.2</td>
<td>18.2</td>
<td>0.1 V/mA</td>
<td>511</td>
<td>1</td>
</tr>
<tr>
<td>$^4$He heat switch-2</td>
<td>2-5.2</td>
<td>10 k</td>
<td>10</td>
<td>-4.1</td>
<td>-0.41</td>
<td>1 V/mA</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>$^4$He charcoal-2</td>
<td>2-5.3-6</td>
<td>386</td>
<td>12</td>
<td>26.0</td>
<td>67</td>
<td>0.01 V/mA</td>
<td>5.62 k</td>
<td>1</td>
</tr>
<tr>
<td>300 mK detector slab</td>
<td>2-6.1</td>
<td>462 k</td>
<td>19</td>
<td>-3.2</td>
<td>0.0069</td>
<td>100 V/mA</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>300 mK cold bar</td>
<td>2-6.2</td>
<td>50 k</td>
<td>20</td>
<td>-3.2</td>
<td>0.064</td>
<td>10 V/mA</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>detector 3 heater</td>
<td>2-6.3</td>
<td>N/C</td>
<td>21</td>
<td>-3.2</td>
<td>N/A</td>
<td>N/A</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>calibrator 1</td>
<td>2-6.4</td>
<td>200</td>
<td>22</td>
<td>9.8</td>
<td>49*</td>
<td>0.1 V/mA</td>
<td>511</td>
<td>1</td>
</tr>
<tr>
<td>calibrator 2</td>
<td>2-6.5</td>
<td>200</td>
<td>23</td>
<td>9.8</td>
<td>49*</td>
<td>0.1 V/mA</td>
<td>511</td>
<td>1</td>
</tr>
<tr>
<td>calibrator 3</td>
<td>2-6.6</td>
<td>N/C</td>
<td>24</td>
<td>9.8</td>
<td>N/A</td>
<td>0.1 V/mA</td>
<td>511</td>
<td>1</td>
</tr>
</tbody>
</table>
B.3 The pressurized drive systems

Table B.8: MBAC heater-pin correspondence—overbars denote return lines. (In original CBOB design, each channel was meant to allow a four-point measurement, where the first two in the pair carry the current and the second pair measure voltage. In this case, multiple pairs are used to share current and reduce power dissipated in the lines.) MBAC has two CBOB units which are each broken into 6 pairs of 6 four-lead lines. In both units, heaters are on output groups 5 and 6, channels 1-6, denoted here by, e.g. CB2-5.4 for CBOB2 group 5, four-lead channel 4.

| Name             | CBOB port | DB50 pins $\{I_+|V_+\}$ |
|------------------|-----------|--------------------------|
| $^4$He-1 heat switch | CB1-5.1   | {20, 4}[23, 7]           |
| $^4$He-2 heat switch | CB1-5.2   | {26, 10}[29, 13]         |
| $^4$He-1 charcoal-a | CB1-5.3   | {32, 16}[37, 38]         |
| $^4$He-1 charcoal-b | CB1-5.4   | {43, 44}[49, 50]         |
| $^4$He-1 charcoal-c | CB1-5.5   | {19, 3}[22, 6]           |
| $^4$He-1 charcoal-d | CB1-5.6   | {25, 9}[28, 12]          |
| $^3$He-1 charcoal | CB1-6.1   | {31, 15}[35, 36]         |
| $^3$He-1 heat switch | CB1-6.2   | {41, 42}[47, 48]         |
| $^3$He-1 300 mK heater | CB1-6.3   | {18, 2}[21, 5]           |
| $^4$He-2 1 K heater | CB1-6.4   | {24, 8}[27, 11]          |
| $^3$He-2 300 mK heater | CB1-6.5   | {30, 14}[33, 17]         |
| $^4$He-1 1 K heater | CB1-6.6   | {39, 40}[45, 46]         |
| $^3$He-2 charcoal | CB2-5.1   | {20, 4}[23, 7]           |
| $^3$He-2 heat switch | CB2-5.2   | {26, 10}[29, 13]         |
| $^3$He-2 charcoal-a | CB2-5.3   | {32, 16}[37, 38]         |
| $^3$He-2 charcoal-b | CB2-5.4   | {43, 44}[49, 50]         |
| $^3$He-2 charcoal-c | CB2-5.5   | {19, 3}[22, 6]           |
| $^3$He-2 charcoal-d | CB2-5.6   | {25, 9}[28, 12]          |
| detector-1 heater | CB2-6.1   | {31, 15}[35, 36]         |
| detector-2 heater | CB2-6.2   | {41, 42}[47, 48]         |
| detector-3 heater | CB2-6.3   | {18, 2}[21, 5]           |
| calibrator-1 supply | CB2-6.4   | {24, 8}[27, 11]          |
| calibrator-2 supply | CB2-6.5   | {30, 14}[33, 17]         |
| calibrator-3 supply | CB2-6.6   | {39, 40}[45, 46]         |
Table B.9: Heater output allocation and relay, current monitor box channels. All outputs on the Din96 have signal on column \( c \) and return on column \( a \), column \( b \) is unused. Power connections span \( a \), \( b \) and \( c \) columns of the Din96. Pins on the DB50 connector are based on the standard, Fig. B.2.3, where channels are reversed, eg 2 is return and 18 is signal, denoted here by Ch1

<table>
<thead>
<tr>
<th>Name</th>
<th>Heater Din96</th>
<th>Heater DB50</th>
<th>Box Channel</th>
<th>Relay bit</th>
<th>DB37 logic pin</th>
<th>Curr. Mon. output</th>
</tr>
</thead>
<tbody>
<tr>
<td>external BNC 6</td>
<td>1</td>
<td>3.8</td>
<td>15</td>
<td>Ch1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>external BNC 5</td>
<td>2</td>
<td>3.7</td>
<td>16</td>
<td>Ch2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>external BNC 4</td>
<td>3</td>
<td>3.6</td>
<td>17</td>
<td>Ch3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>external BNC 3</td>
<td>4</td>
<td>3.5</td>
<td>18</td>
<td>Ch4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>external BNC 2</td>
<td>5</td>
<td>3.4</td>
<td>33</td>
<td>Ch5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>external BNC 1</td>
<td>6</td>
<td>3.3</td>
<td>34</td>
<td>Ch6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(^{3})He charcoal-1</td>
<td>Row29</td>
<td>Ch1</td>
<td>7</td>
<td>3.2</td>
<td>35</td>
<td>Ch7</td>
</tr>
<tr>
<td>(^{3})He charcoal-2</td>
<td>Row28</td>
<td>Ch2</td>
<td>8</td>
<td>3.1</td>
<td>36</td>
<td>Ch8</td>
</tr>
<tr>
<td>(^{3})He heat switch-1</td>
<td>Row27</td>
<td>Ch3</td>
<td>9</td>
<td>2.8</td>
<td>4</td>
<td>Ch9</td>
</tr>
<tr>
<td>(^{3})He heat switch-2</td>
<td>Row26</td>
<td>Ch4</td>
<td>10</td>
<td>2.7</td>
<td>5</td>
<td>Ch10</td>
</tr>
<tr>
<td>(^{4})He charcoal-1</td>
<td>Rows 25,24,23,22</td>
<td>Ch5, 6, 7, 8</td>
<td>11</td>
<td>2.6</td>
<td>6</td>
<td>Ch11</td>
</tr>
<tr>
<td>(^{4})He charcoal-2</td>
<td>Rows 21,20,19,18</td>
<td>Ch9, 10, 11, 12</td>
<td>12</td>
<td>2.5</td>
<td>7</td>
<td>Ch12</td>
</tr>
<tr>
<td>(^{4})He heat switch-1</td>
<td>Row17</td>
<td>Ch13</td>
<td>13</td>
<td>2.4</td>
<td>23</td>
<td>Ch13</td>
</tr>
<tr>
<td>(^{4})He heat switch-2</td>
<td>Row16</td>
<td>Ch14</td>
<td>14</td>
<td>2.3</td>
<td>24</td>
<td>Ch14</td>
</tr>
<tr>
<td>300 mK-1</td>
<td>Row15</td>
<td>Ch15</td>
<td>15</td>
<td>2.2</td>
<td>25</td>
<td>Ch15</td>
</tr>
<tr>
<td>300 mK-2</td>
<td>Row14</td>
<td>Ch16</td>
<td>16</td>
<td>2.1</td>
<td>26</td>
<td>Ch16</td>
</tr>
<tr>
<td>1K-1</td>
<td>Row13</td>
<td>Ch17</td>
<td>17</td>
<td>1.8</td>
<td>28</td>
<td>Ch17</td>
</tr>
<tr>
<td>1K-2</td>
<td>Row12</td>
<td>Ch18</td>
<td>18</td>
<td>1.7</td>
<td>29</td>
<td>Ch18</td>
</tr>
<tr>
<td>detector 1</td>
<td>Row11</td>
<td>Ch19</td>
<td>19</td>
<td>1.6</td>
<td>30</td>
<td>Ch19</td>
</tr>
<tr>
<td>detector 2</td>
<td>Row10</td>
<td>Ch20</td>
<td>20</td>
<td>1.5</td>
<td>31</td>
<td>Ch20</td>
</tr>
<tr>
<td>detector 3</td>
<td>Row9</td>
<td>Ch21</td>
<td>21</td>
<td>1.4</td>
<td>9</td>
<td>Ch21</td>
</tr>
<tr>
<td>calibrator 1</td>
<td>Row8</td>
<td>Ch22</td>
<td>22</td>
<td>1.3</td>
<td>10</td>
<td>Ch22</td>
</tr>
<tr>
<td>calibrator 2</td>
<td>Row7</td>
<td>Ch23</td>
<td>23</td>
<td>1.2</td>
<td>11</td>
<td>Ch23</td>
</tr>
<tr>
<td>calibrator 3</td>
<td>Row6</td>
<td>Ch24</td>
<td>24</td>
<td>1.1</td>
<td>12</td>
<td>Ch24</td>
</tr>
</tbody>
</table>
B.3 The pressurized drive systems

Figure B.14: The switchboard card. Channels originating in the heater driver go in the IDC50 ribbon on the left and are broken out (on the back side of this board) to 2-conductor MTE shrouded headers (Tyco 104257-1) on the front. The two CBOB-bound DB50 connectors are broken out by channel, and terminated in 2-conductor MTE receptacles (Tyco 104450-1) which can be connected to the heater driver headers.

Figure B.15: Front of the current monitor card. The monitor/channel correspondence is shown in Fig. B.2.2. Connectors on a canister which can be pressurized using a bicycle pump.⁹ Fig. B.3 shows both

⁹The cavity pressure is read out using a SenSym ASCX30AN sensor, and the temperature is read out using an AD590. These are powered and readout through a separate break-out. Their calibration is

\[ T(\frac{S_T}{V}) \approx \left( \frac{S_T}{1V} \right) \cdot 100 \text{ K}, \]  

(B.8)
Figure B.16: Front of the relay card. The relay/channel connection is shown in Fig. B.2.2.

the AT-style and newer SATA/USB pressurized drive canisters.

Figure B.17: USB pressurized drives developed by Jeff Klein for ACT (left), and an AT-style pressurized drive originally developed for MINT (right). The AT-drive cable splits into separate lines which go to the controller (40-pin IDC ribbon), power, and a pressure, temperature sensor break-out box. Both drives are thermally sunk to the boxes. The AT-style drive was used exclusively for detector data in the first observing season.

\[
P(S_P) = \left( \frac{S_P}{2.205 \text{ V}} - 0.0648 \right) \text{ atm.}
\]  

(B.9)

while for pressure for a signal \( S_P \) in Volts,

At STP, this is \( \sim 3 \) V for the temperature readout and \( 2.348 \) V for the pressure readout (which should drop to \( \sim 1.3 \) V at the ACT site altitude). A drive emits \( \sim 30 \) W, so we use thermal compound between the drive and the box and stand the drive off with rubber grommets.
B.4 Analysis of the flux lock loop

B.4.1 The MCE $\Sigma - \Delta$ readout

The relation between the feedback applied and the error signal read out in open loop involves many quantities that can vary as a function of the SQUID tuning, and choice of resistors to apply feedback. In practice, we determine it empirically by generating a ramp on the stage one squid feedback and measuring the slope of the error response where the system is locked. Here we will work through the chain in the case that the error response is linear in the feedback applied.

There are four readout cards, which each have 8 channels and permit a maximum of 32 columns per MCE.

The input from the series array is first clamped to $\pm 0.6$ mV and the array bias is sent down the read line by a buffered 16-bit ADC. The first readout op-amp gives a gain of $\approx 4$ (this stage has a one-pole roll-off at 64 MHz for high frequency stability). The first stage is equipped with an offset DAC to center the series array voltage reaching the second stage, preventing saturation of the chain. The second stage applies a one-pole roll-off at 10 MHz and provides an additional gain of $\approx 4$. Both the first and second stage are low-noise, low-slew AD797AR amplifiers. The next stage (AD848JR) applies another pole at 10 MHz and a factor of six in gain and has a high slew rate. A final stage (AD8138AR) provides another factor of two in gain, placing the typical series stage (AD848JR) applies another pole at 10 MHz and a factor of six in gain and has a high slew rate. A final stage (AD8138AR) provides another factor of two in gain, placing the typical series array signal in the 2 V peak-to-peak range of the AD664AST ADC through a gain of 195. The ADC is 14 bits over $\pm 2.2$ volts, or $\delta V/\delta V_{\text{ADC}} = 0.7 \, \mu V$ from the series array. At this point the error signal is all digital. There are a variety of variables associated with the MCE mux system – these are described in Table B.10.

We now derive some relations connecting various stages. Choose the sign of the flux from the feedback inductor through the stage 1 loop to be the opposite of the flux through the stage 1 loop from the detector array so that

$$\Phi_{\text{tot}} = \Phi_{\text{det}} - \Phi_{\text{fb}}.$$  \hfill (B.10)

Then the error value read by the series array is digitally summed over $N_s$ 50 MHz clock cycles each time a row is addressed in the multiplexer. For ACT, $N_s = 10$. If the signal is bandlimited, the internal error is simply $N_s \cdot \text{ADC}(V_{\text{ADC}})$, which can be expanded in terms of the sky signal and the feedback as

$$B_{\text{er}} = N_s D_{\text{er}} = N_s \cdot \text{ADC}(V_{\text{ADC}}) = N_s \cdot \text{ADC}(C_{\text{SA}} V \rightarrow \text{ADC} V V_{\text{squids}}(\Phi_{\text{tot}}))$$
$$= N_s \cdot \text{ADC}(C_{\text{SA}} V \rightarrow \text{ADC} V V_{\text{squids}}(\Phi_{\text{det}} - \Phi_{\text{fb}}))$$
$$= N_s \cdot \text{ADC}(C_{\text{SA}} V \rightarrow \text{ADC} V V_{\text{squids}}(C_{V_{\text{det}}} - C_{V_{\text{fb}}} + V_{\text{fb, ind}}))$$
$$= N_s \cdot \text{ADC}(C_{\text{SA}} V \rightarrow \text{ADC} V V_{\text{squids}}(C_{V_{\text{det}}} - \Phi_{V_{\text{det}}} - C_{V_{\text{fb}}} + C_{V_{\text{fb, DAC}}}) = V_{\text{fb}, DAC}(D_{\text{fb}}))$$

For the purposes of analysis, it is convenient to translate signal voltages in the TES loop into feedback DAC units so that $D_{\text{det}} = C_{V_{\text{det}}} - C_{V_{\text{fb}}} + C_{V_{\text{fb, DAC}}} = \Phi_{V_{\text{det}}} - C_{V_{\text{fb}}} + C_{V_{\text{fb, DAC}}}$.

$$B_{\text{er}} = N_s \cdot \text{ADC}(C_{\text{SA}} V \rightarrow \text{ADC} V V_{\text{squids}}(C_{V_{\text{det}}} - C_{V_{\text{fb}}} + C_{V_{\text{fb, DAC}}}))$$
$$= N_s \cdot \text{ADC}(C_{\text{SA}} V \rightarrow \text{ADC} V V_{\text{squids}}(C_{V_{\text{det}}} - C_{V_{\text{fb}}} + C_{V_{\text{fb, DAC}}}))$$
$$= N_s \cdot \text{ADC}(C_{\text{SA}} V \rightarrow \text{ADC} V V_{\text{squids}}(C_{V_{\text{det}}} - C_{V_{\text{fb}}} + C_{V_{\text{fb, DAC}}}))$$

Then, expand around the lock point, giving the slope

$$C_{V_{\text{fb}}} D_{\text{fb}} = C_{\text{SA}} V \rightarrow \text{ADC} V C_{V_{\text{det}}} - C_{V_{\text{fb}}} + C_{V_{\text{fb, DAC}}} V_{\text{fb}} V'((\Phi_{\text{tot}})).$$  \hfill (B.13)
Table B.10: Quantities relevant for the MUX model. Note that the conversion from feedback DAC volts to DAC units allows a voltage to be converted to fractional DAC units. Also, $C_{fb\ DAC\rightarrow ADC\ V}$ assumes that SQUID slope has been linearized at its lock point.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Units</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{fb}$</td>
<td>volts</td>
<td>voltage produced by the feedback DAC</td>
</tr>
<tr>
<td>$V_{fb, ind}$</td>
<td>volts</td>
<td>voltage on the feedback inductor</td>
</tr>
<tr>
<td>$C_{V_{fb, DAC}\rightarrow V_{fb}}$</td>
<td>unitless conversion</td>
<td>voltage in feedback inductor per DAC voltage</td>
</tr>
<tr>
<td>$\Phi_{fb}$</td>
<td>ratio: flux to $\Phi_o$</td>
<td>feedback flux through the stage 1 coil</td>
</tr>
<tr>
<td>$C_{\Phi_{fb}}$</td>
<td>flux per volt</td>
<td>stage 1 flux per volt in feedback</td>
</tr>
<tr>
<td>$D_{fb}$</td>
<td>digital</td>
<td>digital feedback applied (14-bits)</td>
</tr>
<tr>
<td>$B_{fb}$</td>
<td>digital</td>
<td>internal representation of the feedback</td>
</tr>
<tr>
<td>$W_{fb}$</td>
<td>function</td>
<td>scaling from $B_{fb}$ to $D_{fb}$</td>
</tr>
<tr>
<td>$DAC$</td>
<td>function</td>
<td>converts $D_{fb}$ to $V_{fb}$</td>
</tr>
<tr>
<td>$C_{V_{fb, DAC}\rightarrow V}$</td>
<td>function</td>
<td>converts $V_{fb}$ to $V_{ADC}$</td>
</tr>
<tr>
<td>$C_{D_{fb, DAC}\rightarrow D_{fb}}$</td>
<td>function</td>
<td>convert from $D_{fb}$ to $B_{er}$</td>
</tr>
<tr>
<td>$N_s$</td>
<td>number</td>
<td>number of coadded samples</td>
</tr>
<tr>
<td>$V_{SA}$</td>
<td>volts</td>
<td>SA output (volts)</td>
</tr>
<tr>
<td>$V_{ADC}$</td>
<td>volts</td>
<td>ADC input (volts)</td>
</tr>
<tr>
<td>$D_{er}$</td>
<td>digital</td>
<td>internal representation of the error</td>
</tr>
<tr>
<td>$B_{er}$</td>
<td>digital</td>
<td>internal error = $N_sD_{er}$</td>
</tr>
<tr>
<td>$ADC$</td>
<td>function</td>
<td>converts $V_{SA}$ to $D_{er}$</td>
</tr>
<tr>
<td>$C_{V_{SA, DAC, ADC} V}$</td>
<td>unitless conversion</td>
<td>convert from $D_{fb}$ to $B_{er}$</td>
</tr>
<tr>
<td>$Q(x)$</td>
<td>function</td>
<td>$Q(x) = \lfloor x \rfloor$</td>
</tr>
<tr>
<td>$C_{V_{det}}$</td>
<td>flux per volt</td>
<td>stage 1 flux per volt in detector loop signal</td>
</tr>
<tr>
<td>$\Phi_{det}$</td>
<td>ratio: flux to $\Phi_o$</td>
<td>TES loop flux through the stage 1 coil</td>
</tr>
<tr>
<td>$\Phi_{tot}$</td>
<td>ratio: flux to $\Phi_o$</td>
<td>total flux in the stage 1 coil</td>
</tr>
<tr>
<td>$V_{squelch}$</td>
<td>function</td>
<td>total stage 1 flux to voltage on ADC</td>
</tr>
<tr>
<td>$V_{det}$</td>
<td>volts</td>
<td>signal voltage in the detector loop</td>
</tr>
<tr>
<td>$D_{det}$</td>
<td>volts</td>
<td>signal voltage in the detector loop in fb DAC units</td>
</tr>
<tr>
<td>$P$</td>
<td>digital</td>
<td>proportional term</td>
</tr>
<tr>
<td>$I$</td>
<td>digital</td>
<td>integral term</td>
</tr>
<tr>
<td>$D$</td>
<td>digital</td>
<td>derivative term</td>
</tr>
</tbody>
</table>
B.4 Analysis of the flux lock loop

such that,

$$B_{er} = N_s \cdot ADC(C_{fb\ DAC\ -\ ADC} V(D_{det} - D_{fb}))$$  \hspace{1cm} (B.14)

Here we have also absorbed an arbitrary offset into $D_{det}$ without loss of generality. During SQUID tuning, we apply a ramp to $D_{fb}$, and read out $B_{er}$. The ramp response is quantized by both the discrete levels of the ADC and of the DAC. It is convenient for analysis to convert the $ADC()$ function to simply a rounding function $Q(x) = \lfloor x \rfloor$, which is a simple quantizer. We then want a conversion from feedback DAC “ticks” to ADC “ticks”, $C_{fb\ DAC\ -\ ADC}$. This can be measured directly from the SQUID tuning procedure $C_{fb\ DAC\ -\ ADC} = (\text{slope of ADC units per DAC tick})/N_s$. With this choice of variables,

$$B_{er} = N_s Q(C_{fb\ DAC\ -\ ADC} D_{det} - D_{fb}))$$  \hspace{1cm} (B.15)

Now note that bit shift to scale the internal representation of feedback to the output,

$$D_{fb} = W_{fb}(B_{fb}) = \lfloor B_{fb}/2^s \rfloor = Q(B_{fb}/2^s)$$  \hspace{1cm} (B.16)

This allows us to write the digital input in terms of the digital output and encapsulate the response of the analog SQUID hardware,

$$B_{er} = N_s Q(C_{fb\ DAC\ -\ ADC} D_{det} - Q(B_{fb}/2^s)))$$  \hspace{1cm} (B.17)

To close this loop the FPGA applies an all-digital PID loop as

$$B_{fb}(i + 1) = P \cdot B_{er}(i) + I \cdot \sum_{j=0}^{i} B_{er}(j) + D \cdot [B_{er}(i) - B_{er}(i - 1)]$$  \hspace{1cm} (B.18)

Taking just an integral term, the full readout system is then

$$B_{fb}(i + 1) = IN_s \cdot \sum_{j=0}^{i} Q(C_{fb\ DAC\ -\ ADC} D_{det}(j) - Q(B_{fb}(j)/2^s)))$$

$$= IN_s Q(C_{fb\ DAC\ -\ ADC} D_{det}(i) - Q(B_{fb}(i)/2^s)))$$

$$+ IN_s \cdot \sum_{j=0}^{i-1} Q(C_{fb\ DAC\ -\ ADC} D_{det}(j) - Q(B_{fb}(j)/2^s)))$$

so that

$$B_{fb}(i + 1) = IN_s Q(C_{fb\ DAC\ -\ ADC} (D_{det}(i) - Q(B_{fb}(i)/2^s))) + B_{fb}(i)$$  \hspace{1cm} (B.19)

It is also convenient to define a rounding function that rounds in bins of width $a$ and tracks a line with slope 1, as

$$R(x,a) = a \lfloor x/a \rfloor.$$  \hspace{1cm} (B.20)

Then we can rearrange terms and put the detector signal in units of the internal feedback unit $B_{det} = 2^s D_{det}$ to give

$$B_{fb}(i + 1) = IN_s Q(C_{fb\ DAC\ -\ ADC} (1/2^s)(2^s D_{det}(i) - 2^s Q(B_{fb}(i)/2^s))) + B_{fb}(i)$$

$$= IN_s Q(C_{fb\ DAC\ -\ ADC} (1/2^s)(B_{det}(i) - R(B_{fb}(i),2^s))) + B_{fb}(i)$$

$$= IN_s Q(C_{fb\ DAC\ -\ ADC} (1/2^s)C_{fb\ DAC\ -\ ADC}^{-1} 2^s$$

$$\times Q(C_{fb\ DAC\ -\ ADC} (1/2^s)(B_{det}(i) - R(B_{fb}(i),2^s))) + B_{fb}(i)$$

$$= IR(B_{det}(i) - R(B_{fb}(i),2^s), C_{fb\ DAC\ -\ ADC}^{-1} 2^s) + B_{fb}(i)$$

196
where \( \tilde{I} = IN_sC_{fb\rightarrow ADC/2} \). Call the rounding function for the series array \( q_s \) (sensor quantization) and the rounding function for the feedback DAC \( q_a \) (actuator quantization). Then,

\[
B_{fb}(i + 1) = \tilde{I}q_s(B_{det}(i) - q_a(B_{fb}(i))) + B_{fb}(i)
\]  

so that if the quantizers can be ignored,

\[
B_{fb}(i + 1) = \tilde{I}B_{det}(i) + (1 - \tilde{I})B_{fb}(i)
\]  

This has the form of as recursive filter. Note that a pure integral term is sufficient, as

\[
\tilde{I} = IN_sC_{fb\rightarrow ADC/2} = 1 \Rightarrow B_{fb}(i + 1) = B_{det}(i)
\]  

This has neglected nonlinearities in the stage 1 SQUID. Further work is required to understand harmonic distortion in this system.

### B.5 MUX response analysis

In the limit that both quantizers in the MUX loop can be ignored, the feedback at step \( n + 1 \) can be written as

\[
f_{n+1} = Ix_n + (1 - I)f_n.
\]  

And the error \( (e_{n+1} = x_{n+1} - f_{n+1}) \) is

\[
e_{n+1} = x_{n+1} - x_n + (1 - I)e_n.
\]  

These equations are in the recursive filter form for data \( x_n \),

\[
f_n = \sum_{k=0} c_kx_{n-k} + \sum_{k=1} d_ky_{n-k},
\]  

whose transfer function is given by

\[
H(z) = \frac{\sum_{k=0} c_kz^{-k}}{1 - \sum_{k=1} d_kz^{-k}}
\]  

Applying this to the feedback linear difference equation (Eq. B.24),

\[
H_f(z) = \frac{Iz^{-1}}{1 - (1 - I)z^{-1}} = \frac{I}{e^{i\omega} - (1 - I)},
\]  

and to the error linear difference equation (Eq. B.25)

\[
H_e(z) = \frac{1 - z^{-1}}{1 - (1 - I)z^{-1}} = \frac{e^{i\omega} - 1}{e^{i\omega} - (1 - I)}.
\]  

In frequency, the feedback transfer function is

\[
|H_f(z)| = \left| \frac{I}{e^{i\omega} - (1 - I)} \right| = \left( \frac{4(1 - I)}{I^2}\sin^2(\omega/2) + 1 \right)^{-1/2}.
\]
B.5 MUX response analysis

Likewise, the error transfer function is (noting that $|e^{i\omega} - 1| = 2\sin(\omega/2)$)

$$|H_e(z)| = \left|\frac{e^{i\omega} - 1}{e^{i\omega} - (1-I)}\right|$$

$$= \frac{2}{I}\sin(\omega/2)\left(\frac{4(1-I)}{I^2} - \sin^2(\omega/2) + 1\right)^{-1/2}.$$  \hfill (B.32)

$$= \frac{2}{I}\sin(\omega/2)\left(\frac{4(1-I)}{I^2} - \sin^2(\omega/2) + 1\right)^{-1/2}.$$  \hfill (B.33)

We can also calculate the phase response of the feedback loop by noting

$$\text{Re}(z) = \frac{1}{2}(z + \bar{z}) \quad \text{Im}(z) = \frac{1}{2}i(\bar{z} - z) \quad \arg(z) = \tan^{-1}(\text{Im}(z)/\text{Re}(z)) \hfill (B.34)$$

We can then rewrite the transfer function,

$$H_f(z) = \frac{I}{e^{i\omega} - (1-I)} = \frac{I(e^{-i\omega} - (1-I))}{|e^{i\omega} - (1-I)|^2} = \Gamma(e^{-i\omega} - (1-I)).$$ \hfill (B.35)

Then taking the real part,

$$\text{Re}[H_f(z)] = \Gamma(\cos(\omega) - 1 + I),$$ \hfill (B.36)

and the imaginary part,

$$\text{Im}[H_f(z)] = -\Gamma\sin(\omega).$$ \hfill (B.37)

Then the phase is given by,

$$\arg[H_f(z)] = \tan^{-1}\left(\frac{\sin(\omega)}{1 - I - \cos(\omega)}\right).$$ \hfill (B.38)

Note that when $I = 1$, then the tangent is inverted to get $\arg[H_f(z)] = -\omega$. 

198
Appendix C

Calibration and Diagnostics

C.1 Effective area in the geometric limit

If the size of a pixel is much greater than the wavelength, then diffraction in the optical system dominates and the power received by a detecting element is simply the integral of the Airy disk over the extent of the detector. From this, one can calculate the probability that a photon from a point source is absorbed by one detector in the focal plane. This quantity is related to “aperture efficiency” more familiar from radio astronomy. There are two complementary ways to look at the geometric limit. In one, the point source on the sky produces an aperture Airy beam pattern (integrated over the passband) on the focal plane. The power received by a detector is then the integral of the Airy pattern over the square pixel. In another, the beam pattern on the sky is the aperture Airy beam pattern (integrated over the passband) convolved by a square. The central amplitude of the beam pattern is broadened and reduced by this convolution.

To develop an approximation for the aperture efficiency, take the aperture to be fully illuminated and the detector to be small compared the Airy disk so that the solid angle seen by the detector is just the raytraced geometric solid angle, \( \Omega_d = \pi \sin(\theta_h)^2 = \pi/(4F^2) \approx 0.97 \) ster (for ACT). The throughput can be calculated easily if we measure the physical detector length \( \ell \) units of \( F\lambda \), \( \ell = xF\lambda \). Then

\[
A_d\Omega_d = \frac{\pi(xF\lambda)^2}{4F^2} = \frac{\pi}{4}x^2\lambda^2.
\]

Pixels in the 145 GHz camera have physical size \( \sim 0.5 F\lambda \), so the throughput onto one detector in this limit is \( (\pi/16)\lambda^2 \). Yet, the throughput of the instrument is \( A_p\Omega_b = m\lambda^2 \) where \( \Omega_b \) is the main beam solid angle and \( A_p \) is the effective primary area and \( m \) is the number of modes received, which is believed to be unity. The throughput must also be conserved onto the focal plane, so \( (\pi/4)x^2 \) is the fraction of the total power received by the telescope \( A_p\Omega_b \) that is received by one central detector. This has the form of a modal relation where \( A_d\Omega_d = \tilde{m}\lambda^2 \), but convention is to absorb it as an effective area,

\[
A_d\Omega_d = \frac{\pi}{4}x^2\lambda^2 = \frac{\pi}{4}x^2 A_p\Omega_b = A_e\Omega_b.
\]

Here, \( A_e = (\pi/4)x^2A_p \) is an approximation to the effective area in Eq. 5.4 in the limit of uniform illumination and a pixel in the geometric limit that is small within the Airy disk (or equivalently that gives an Airy disk on the sky). Niemack (2008) extends this to include the fact that the pixel subtends a finite fraction of the Airy disk, so blurs the beam. In the picture here, the fraction \( (\pi/4)x^2 \) is decreased to the integral of a square at the center of the Airy disk. The optical coupling is essential for estimates of the power received from a known source or efficiency measurement,
C.2 Planet temperature

but not necessary for the global calibration, where the only factor that enters is the main beam solid angle and passbands. Because of sub-wavelength detector diffraction effects, the real beam pattern and optical coupling is poorly understood and is the subject of future work.

C.2 Planet temperature

C.2.1 The heliocentric distance

Planets with more eccentric orbits will oscillate in temperature as a function of heliocentric distance in their orbit.\(^1\) Denote the reference measurement quantities (temperature, radius, flux) with a subscript \(r\), and analogous quantities at the test point with a subscript \(t\). The flux from the sun at the test point is \(F_t = F_r (r_r/r_t)^2\). We will assume that there are no internal heat sources, so this balances the outgoing, emitted radiation. The re-radiated radiation is proportional to \(T^4\) so that \(T_t^4/T_r^4 = (r_r/r_t)^2\), or \(T_t = T_r/\sqrt{r_t/r_r}\).\(^2\) The heliocentric distance of the planet at apsides is \((1 \pm \epsilon)a\), where \(a\) is the semi-major axis and \(\epsilon\) is the eccentricity. Consider total excursion in temperature in the orbit from the change in the sun’s flux can be estimated to be (for the reference radius at the semimajor length \(a\))

\[
T_{peri} - T_{apo} \approx T_r \left( \frac{1}{\sqrt{1 - \epsilon}} - \frac{1}{\sqrt{1 + \epsilon}} \right) \quad (C.3)
\]

\[
\approx T_r \epsilon \quad (C.4)
\]

The fractional change in temperature during the sidereal orbit can be expected to vary \(\epsilon\) percent. Ulich (1981) shows that this is a reasonable approximation for the largest temperature changes from Mars. Fig. C.1, C.2, and C.3 show several geometrical factors that are relevant for the planet temperature.

C.2.2 Ellipticity and solid angle

Take the planet to be a spheroidal shell specified by,

\[
\frac{x^2}{b^2} + \frac{y^2}{b^2} + \frac{z^2}{a^2} = 1, \quad (C.5)
\]

defining the eccentricity to be \(\epsilon = 1 - b^2/a^2\), and the inclination angle \(\alpha_{inc}\) as the angle from the \(z\) axis. The projected ellipse has semimajor axis \(a_\perp\) and semiminor axis \(b_\perp\) (see Wade et al. (2000)),

\[
a_\perp = a\sqrt{1 - \epsilon^2 \cos^2(\alpha)} \quad (C.6)
\]

\[
b_\perp = a\sqrt{1 - \epsilon^2} \quad (C.7)
\]

The solid angle for a planet at distance \(d_p\) is

\[
\Omega_{p,\epsilon} = \pi \frac{a_\perp b_\perp}{d_p^2} \quad (C.8)
\]

\[
\approx \pi a^2 \left[ 1 - \frac{1}{2} (1 + \cos^2(\alpha)) \epsilon^2 \right] d_p^{-2}. \quad (C.9)
\]

---

\(^1\)This effect is mainly understood for Mars, but could also be significant for Saturn and its rings, which is significantly more complicated.

\(^2\)This assumes that the albedo is uniform, and ignores the time dependence of the surface temperature and viewing phase for a planet with finite thermal inertia.
Figure C.1: The heliocentric distance of the planets. This is an essential part of the Mars thermal model, but the effect is not yet part of temperature models for the gas giants and is likely to be much more complex.

Figure C.2: The viewing phase of the planets. Mars is viewed over the widest range of phases, and its surface temperature depends directly on the Sun’s illumination. The analogous effects are not understood in the gas giants, where the temperature may be more uniform because of mixing. Saturn's ring emission is also expected to be a function of fraction we observe that is illuminated by the Sun.
C.2 Planet temperature

If we define the solid angle for a planet with $\epsilon = 0$, $\Omega_p = \pi a^2 d_p^{-2}$, then

$$
\delta \Omega = \frac{\Omega_{p,\epsilon} - \Omega_p}{\Omega_p} = -\frac{1}{2} (1 + \cos^2(\alpha)) \epsilon^2
$$

(C.10)

It is common to take the oblateness (or flattening) as $f = (a - b)/a = 1 - \sqrt{1 - \epsilon^2}$, which for small $\epsilon$ gives $f = \epsilon^2/2$. To get a sense of how much the solid angle could vary with inclination angle, consider

$$
\frac{\Omega_{p,\epsilon}(\alpha = 0) - \Omega_{p,\epsilon}(\alpha = \pi/2)}{\Omega_{p,\epsilon}(\alpha = 0)} \approx \frac{1}{2} \epsilon^2 = f
$$

(C.11)

C.2.3 Jupiter

The emission from both Saturn and Jupiter depends on the structure and makeup of their atmospheres. For reviews of millimeter wave models, see Naselsky et al. (2007); de Pater and Massie (1985); Killen and Flasar (1996); Pajot et al. (2006); Pollack (1975); Bezard et al. (1986); Lellouch et al. (1984) and for measurements that describe both Jupiter and Saturn, see Goldin et al. (1997); Bergin et al. (2000); Hildebrand et al. (1985); Ulich et al. (1984); Killen and Flasar (1996); Ulich et al. (1980); Elanov et al. (1971); Rather et al. (1974); Rather et al. (1974). Saturn is described in detail in the body of the thesis, but for Jupiter, we have produced a similar compilation of data, see Fig. C.4. For measurements of Jupiter in particular, see Gibson et al. (2005); de Pater (2004); Wrixon et al. (1971); Griffin et al. (1986); Page et al. (2003a); Simon-Miller et al. (2006); Kunde et al. (2004); de Pater et al. (2001); Davis et al. (1997); Lellouch et al. (2006); Moreno et al. (2001); Matthews et al. (2002).
C.2.4 Uranus and Neptune

Uranus and Neptune are dominated by collision-induced absorption between H$_2$, CH$_4$ and He that gives a smooth emission profile and is well-modelled by the expressions given in Griffin and Orton (1993) (see Marten et al. (2005) for a recent discussion of microwave emission). Neptune has fairly deep CO lines bracketing the ACT bands for $J=1-0$, at 115 GHz, $J=2-1$ at 231 GHz, $J=3-2$ at 346 GHz; see Marten et al. (2005); Guilloteau et al. (1993). There was no signature of $J=1-0$ PH$_3$ at 267 GHz, see Encrenaz et al. (1996). The $J=2-1$ line at 231 GHz corresponds to an absorption from the $\sim 90$ K the continuum model would predict to $\sim 75$ K with a width $\sim 10$ GHz and so influences the temperature measured by the 220 GHz camera. Fig. C.5 and C.6 show compilations of recent data and models.

C.2.5 Mars

Mars is preferable to the gas giants because it is relatively simple and most variability can be constrained to $\sim$ few percent. The emission of the planetary disk is determined by the viewing

---

For additional literature for Uranus and Neptune see Marten et al. (2005); Moseley et al. (1985); Griffin and Orton (1993); Klein and Hofstadter (2006); Kramer et al. (2008); Gulkis et al. (1983); Fazio et al. (1976); Orton et al. (1986); Pearl et al. (1990); van Hemelrijck (1982); Marley and McKay (1999); Hofstadter and Muhleman (1989); Lunine (1995, 1993); Guilloteau et al. (1993)
C.2 Planet temperature

phase of the planet from Earth, and temperature changes due to high orbital eccentricity. The
dependence on the viewing phase is due to the night to day temperature difference, and the
change in temperature depends on the thermal constants of the Martian regolith. Based on these,
there are two standard models from Rudy (1987) and Wright (2007b). The ACT pipeline currently
implements Wright (2007b) with the correction from Hill et al. (2008). Fig. C.7 compares this model
to the WMAP data and gives predictions for ACT observations.

Global dust storms on Mars only occur at perihelion with highly variable intensity and duration,
lasting from weeks to a year. These cause a departure of $\sim 10\%$ (or $\sim 20\ K$) from the thermal
model.\(^4\)\(^5\) The ACT season 1 coincided with a particularly large dust storm in from late June to
late August, but this settled before the period of science observations\(^6\).

In addition to dust storms, the polar ice migrates yearly, and clouds and frosts can change the
integrated temperature; these are thought to be at the percent level in the disk-integrated tempera-
ture (Neumann et al. (2003)) and can be neglected.\(^7\) The Martian atmosphere is 95% CO\(_2\), but the

---

\(^4\)See Fernández (1997); Gurwell et al. (2005); Cantor (2007b); Cantor et al. (2001); Cantor (2007a); Medvedev and
Hartogh (2007); Kuroda et al. (2006); Cantor et al. (2001); Cantor (2007a); Clancy et al. (2000).

\(^5\)Scattering by the Martian dust by sub-mm/mm waves is negligible, because their mean diameter is 2 \(\mu m\) (Tomasko
1999). Measurements in the sub-mm and mm then measure the surface temperature, while IR heating of the surface is
less effective (and at night IR emission is less effective).

\(^6\)See the June and July 2007 report from the Mars Rovers http://marsrover.nasa.gov/mission/vir/index.html,
where the visible optical depth to the surface was $> 4.7$.

\(^7\)See Clancy et al. (1983, 1992); Levin (2004); Boynton et al. (2002); Litvak et al. (2007); Mitrofanov et al. (2007); Rudy
surface pressure is 6 mbar, so presents only a few well-separated and narrow molecular features Clancy et al. (1996); Gurwell et al. (2000) which should be irrelevant for wide-band calibration. One effect that is neglected in the Wright analysis which could be significant is the inhomogeneity of the heat capacity of the soil, (from sands to more rocky volcanic soil) which would lead to hourly variations as the planet rotates.8

### C.2.6 Other calibrators

Aside from the planets, there are several other objects that have been considered as calibrators both inside and outside the solar system; See Ott et al. (1994); Mason et al. (1999); Pardo et al. (2005); Sandell (1994); Linsky (1973); Reichart and Stephens (2000); Cohen et al. (2005); Jaeger (1953). For compilations of planet temperatures, see Greve et al. (1994); Harrison and van Leeuwen (2006); Ulich (1981); Mangum (1993); Mason et al. (1999); Rather et al. (1974) and for general discussion of calibration, see Runyan et al. (2003); Kuo et al. (2004); Reichardt et al. et al. (1987); Liu et al. (2003); Hartogh et al. (2007); Biver et al. (2005); Clancy et al. (2000); Gurwell et al. (2000); Clancy et al. (1996); Muhleman and Clancy (1995). for further discussion of the atmosphere and surface weather.

8See Wright (2007a); Spencer et al. (1989); Neugebauer et al. (1971); Chase (1969); Jakosky and Muhleman (1981); Inada et al. (2007); Sun et al. (2006); Ivanov et al. (1998);Jaeger and Johnson (1953); Jaeger (1953); Putzig and Mellon (2007); Wright (1976, 2007b); Serabyn and Weisstein (1996); Simpson et al. (1981); Christensen et al. (2001); Cuzzi and Muhleman (1972); Muhleman et al. (1991); Epstein (1971); Eltanin et al. (1971); Levine et al. (1977) for thermal surface models and observations.
C.3 The accuracy of a pulse calibration

In this appendix, we estimate the ability of a pulsed calibrator to find the relative change in responsivity across times (a time transfer standard). We will take the response function to be a square, which is on for some time \( t_o \), and take a response with no drift. To find the amplitude, we measure the response with the emitter on, distributed as \( N(\mu_o, \sigma_o^2) \), relative to a baseline with the emitter turned off, distributed as \( N(\mu_o, \sigma_o^2) \). The uncertainty in the amplitude of the pulse (the difference between the baseline and “on” response) is then

\[
\sigma_{Amp}^2 = \sigma_b^2 + \sigma_o^2 \tag{C.12}
\]

If the detector noise is which with some variance \( \sigma^2 \), then the variance of the baseline and “on” response is down by the number of samples acquired, as

\[
\sigma_{Amp}^2 = \sigma^2 \left( \frac{1}{N_b} + \frac{1}{N_o} \right) \tag{C.13}
\]

where \( N_b \) is the number of baseline points and \( N_o \) is the number of “on” points.

To calibrate gains, the relevant term is the ratio of pulse amplitudes a different times. Take times \( i \) and \( j \) with amplitudes \( \text{Amp}_i \) and \( \text{Amp}_j \). The amplitude measured at each calibration pulse, \( i \) and \( j \), is normally distributed, and the ratio of these random variables follows the ratio distribution, which is not Gaussian. If the amplitude of measured at time \( i \) is \( \text{Amp}_i + \delta \text{Amp}_i \) for some small
noise \( \delta \text{Amp}_i \) then the ratio of amplitudes at different times can be approximated as
\[
    r = \frac{\text{Amp}_i + \delta \text{Amp}_i}{\text{Amp}_j + \delta \text{Amp}_j} \\
    \approx \frac{\text{Amp}_i}{\text{Amp}_j} \left( 1 + \frac{\delta \text{Amp}_i}{\text{Amp}_i} - \frac{\delta \text{Amp}_j}{\text{Amp}_j} \right)
\] (C.14)
which is just the ratio plus a noise piece. If the \( \delta \text{Amp}_i \) are taken to be Gaussian and small then
\[
    \sigma_r^2 \approx \frac{4}{\text{SNR} \cdot N_o \cdot N_{\text{det}}} < \epsilon^2.
\] (C.19)
Thus to get relative calibrations \( \epsilon \) to better than 1% for taking \( R \) samples/sec over some time \( t_o \),
\[
    t_o \approx \frac{4}{\epsilon^2 \cdot \text{SNR} \cdot N_{\text{det}} \cdot R} \approx \frac{40000}{\text{SNR} \cdot N_{\text{det}} \cdot R}
\] (C.20)
Sampling at 400 Hz we need the source to be on for at least,
\[
    t_o \sim (100 \text{ sec}) \cdot \text{SNR}^{-1} \cdot N_{\text{det}}^{-1}
\] (C.21)

### C.4 Relation between B-splines, the optimal filter, and iterated methods

A B-spline drift is incorporated into the data model by adding \( Bx_b \), where \( B \) are the basis functions and \( x_b \) is a vector of the amplitudes (equivalent to Eq. 6.14). Then
\[
    \mathbf{d} = \mathbf{Mx}_m + Bx_b + \mathbf{n}.
\] (C.22)
The probability distribution of model parameters \( \hat{x}_m \) and \( \hat{x}_b \) given measurements \( \mathbf{d} \) with a multivariate distribution of covariance \( \mathbf{K} \) is (where the likelihood \( \mathcal{L} = P(\mathbf{d} | \hat{x}_m, \hat{x}_b) \))
\[
    P(\hat{x}_m, \hat{x}_b | \mathbf{d}) \propto \mathcal{L} P(\hat{x}_m) P(\hat{x}_b | \hat{x}_m) \propto \mathcal{L} P(\hat{x}_b) \propto \mathcal{L}
\] (C.23)
\[
    \propto \exp \left[ -\frac{1}{2} \Delta \mathbf{d}^T K^{-1} \Delta \mathbf{d} \right].
\] (C.25)
C.4 Relation between B-splines, the optimal filter, and iterated methods

Where the first equality uses Bayes' rule and the probability chain rule. The second equality assumes a flat prior on the map and that the basis amplitudes are uncorrelated with the map. The final expression assumes a flat prior probability of basis amplitudes, so that the probability only depends on the residual $\Delta d = d - M\hat{x}_m - B\hat{x}_b$. Once a numerical method is in place to find the map parameters with maximum likelihood, it is easy to extend it to also estimate the spline amplitudes.

We can then fix $\hat{x}_b$ and find the $\hat{x}_m$ map that maximizes the likelihood along the corresponding slice in the ($\hat{x}_b$, $\hat{x}_m$) plane. Let $y = d - B\hat{x}_b$ be the data minus the fixed estimate of the spline. Then the maximum-likelihood along this slice at $\hat{x}_b$ is given by

$$\hat{x}_m = (M^T K^{-1} M)^{-1} M^T K^{-1} y = \Pi (d - B\hat{x}_b) \quad \quad \text{(C.26)}$$

where $\Pi = (M^T K^{-1} M)^{-1} M^T K^{-1}$ is the maximum-likelihood projection of data onto the map.

The residual can then be re-written as

$$\Delta d = (1 - \Pi \Pi)(d - B\hat{x}_b). \quad \quad \text{(C.27)}$$

(Note that this does not depend on $\hat{x}_m$, $\hat{x}_b$.) The operator $(1 - \Pi \Pi)$ projects the best estimate of the map out of the time-ordered data. The log-likelihood for the basis amplitudes is then

$$\ln \mathcal{L} \propto -(d - B\hat{x}_b)^T (1 - \Pi \Pi M^T) K^{-1} (1 - \Pi \Pi)(d - B\hat{x}_b) \quad \quad \text{(C.28)}$$

Here, the conjugation\(^9\) of the inverse covariance under $(1 - \Pi \Pi)$ rotates out the best estimate of the sky signal. Note that\(^{10}\)

$$(1 - \Pi \Pi M^T) K^{-1} (1 - \Pi \Pi) = K^{-1} (1 - \Pi \Pi). \quad \quad \text{(C.30)}$$

The form of the likelihood is identical to the optimal map estimate, but with projections onto the B-splines and the modified covariance as

$$\hat{x}_b = (B^T K^{-1} (1 - \Pi \Pi) B)^{-1} B^T K^{-1} (1 - \Pi \Pi) d. \quad \quad \text{(C.31)}$$

In the case that $K = 1$, this is identical to the Cottingham-Boughn method.

C.4.1 Iterative solution

Because the astronomical sky signal is so small, it is convenient to approximate the likelihood by ignoring the sky projection $1 - \Pi \Pi$, as

$$\ln \mathcal{L} \propto -(d - B\hat{x}_b)^T K^{-1} (d - B\hat{x}_b) \quad \quad \text{(C.32)}$$

The estimate for the amplitudes reduces to

$$\hat{x}_b \approx (B^T K^{-1} B)^{-1} B^T K^{-1} d = \Pi_B d \quad \quad \text{(C.33)}$$

This is simply the weighted least-squares B-spline amplitude estimate to the TOD. This can be solved iteratively in parallel with a map estimate as,

$$\hat{x}_{b,0} = \Pi_B d \quad \quad \text{(C.34)}$$

$$\hat{x}_{m,0} = \Pi (d - B\hat{x}_{b,0}) \quad \quad \text{(C.35)}$$

---

\(^9\)That is, $(1 - \Pi \Pi M^T) K^{-1} (1 - \Pi \Pi)$.

\(^{10}\)Consider the most complicated cross-term in the projection of the full covariance onto a covariance with the sky removed,

$$\Pi^T M^T K^{-1} \Pi = ((M^T K^{-1} M)^{-1} M^T K^{-1})^T M^T K^{-1} M (M^T K^{-1} M)^{-1} M^T K^{-1} = \Pi^T M^T K^{-1}. \quad \quad \text{(C.29)}$$
and for $i > 0$,

$$\hat{x}_{b,i} = \Pi_B (d - M \hat{x}_{m,i-1}) \quad (C.36)$$

$$\hat{x}_{m,i} = \Pi (d - B \hat{x}_{b,i}). \quad (C.37)$$

In each iteration, the B-spline in the time domain is just $B \Pi_B$ applied to the current best estimate of the data with the target sky signal removed. This converges to the result in the previous section. To see this, consider

$$\hat{x}_{b,i} = \Pi_B (d - M \hat{x}_{m,i-1}) \quad (C.38)$$

$$= \Pi_B (d - M \Pi_B (d - B \hat{x}_{b,i-1})). \quad (C.39)$$

$$= \Pi_B (1 - M \Pi_B)d + \Pi_B M \Pi_B \hat{x}_{b,i-1} \quad (C.40)$$

$$= u + A \hat{x}_{b,i-1} \quad (C.41)$$

The iteration proceeds as

$$\hat{x}_{b,i} = A^i \hat{x}_{b,0} + \sum_{j=0}^{i} A^j u. \quad (C.42)$$

This is a geometric series that converges to

$$\hat{x}_{b,\infty} = \lim_{i \to \infty} \hat{x}_{b,i} = (1 - A)^{-1} u \quad (C.43)$$

$$= (1 - \Pi_B M \Pi_B)^{-1} \Pi_B (1 - M \Pi_B)d. \quad (C.44)$$

From Eq. C.31 in the previous section,

$$\hat{x}_b = (B^T K^{-1}(1 - M \Pi B))^{-1} B^T K^{-1}(1 - M \Pi) d \quad (C.45)$$

$$= (B^T K^{-1}B) \Pi_B (1 - M \Pi B)(B^T K^{-1}B)^{-1} B^T K^{-1}(1 - M \Pi) d \quad (C.46)$$

$$= (\Pi_B (1 - M \Pi B))^{-1} \Pi_B (1 - M \Pi) d \quad (C.47)$$

$$= (1 - \Pi_B M \Pi B)^{-1} \Pi_B (1 - M \Pi) d \quad (C.48)$$

$$= \hat{x}_{b,\infty}. \quad (C.49)$$

Here we have used that $B^T K^{-1} = (B^T K^{-1}B) \Pi_B$, that $B^T K^{-1}B$ is a square matrix, and that $\Pi_B B = 1$.

The speed of the iteration depends on the generalized pointing matrix and covariances through $\Pi_B M \Pi_B$ and its eigenvalues.\(^{11}\) This suggests a simple iterative mapmaker. In a first step, fit and subtract a B-spline from the data and make a map. Then, project the map back onto the time domain and subtract it from the data. Then fit and subtract a B-spline with the first map iteration subtracted, and make a map of this. This process continues until reaching a desired convergence in the map.

**C.4.2 The spline filter**

Convert the spline fitting operation into a filter with some input and output by writing

$$\hat{s} = B \hat{x}_b \approx B (B^T K^{-1}B)^{-1} B^T K^{-1} d. \quad (C.50)$$

Here $\hat{s}$ is a function in the time domain that is a response to $d$, so is exactly a filter. There is a fairly large literature Pang et al. (1994); Ustuner and Ferrari (1992); Unser (1999); Unser et al. (1993a,b);\(^{11}\)Recall that $A^k = P^{-1} D^k P$ where $D$ is the diagonal of eigenvalues if the similarity transformation $P$ exists.
C.4 Relation between B-splines, the optimal filter, and iterated methods

Samadi et al. (2004) about B-spline filters. To understand the frequency-domain response it is necessary to take $K = 1$, this reduces to the case described in Unser et al. (1993a,b),

$$\hat{s} = B\hat{x}_b \approx B(B^T B)^{-1}B^T d.$$  \hspace{1cm} (C.51)

Unser et al. (1993a) rewrites these matrix operations as resampling and convolution operations which are more familiar from filter literature and finds that the response when viewed as a filter, $F(z)$. Interested readers are referred to this work and tabulated values for the filter in Unser et al. (1993b). The filter $F(z)$ plays the same role as $S/(S + K)$ in the optimal filter, except that $F(z)$ has a complex form that depends on the spline order and knot spacing. This form of the filter given in Unser et al. (1993b) and Unser et al. (1993a) assumes uniform knot spacing, infinite data, and constant, white noise. The response function in general is much more complicated. It is difficult, then, to design a general B-spline filter whose transfer function coincides with the actual drift spectrum. The B-spline filter may be near-optimal in many cases, it can be solved in parallel with the map, making it convenient.

C.4.3 The optimal filter argument

In this section, we extend the idea of the B-spline to a model that is an arbitrary time series $s$ where

$$d = Mx + s + n.$$ \hspace{1cm} (C.52)

For noise with covariance $K$ and drift with covariance $S$,

$$\ln L \propto -\Delta d^T K^{-1} \Delta d - \hat{s}^T S^{-1} \hat{s},$$ \hspace{1cm} (C.53)

where $\Delta d = d - M\hat{x}_m - \hat{s}$. Applying the same argument as for the B-spline but now with a prior,

$$\ln L \propto -(d - \hat{s})^T (1 - \Pi^T M^T) K^{-1} (1 - M\Pi)(d - \hat{s}) - \hat{s}^T S^{-1} \hat{s}$$ \hspace{1cm} (C.54)

$$\propto -(d - \hat{s})^T \bar{K}^{-1} (d - \hat{s}) - \hat{s}^T S^{-1} \hat{s}$$ \hspace{1cm} (C.55)

where $\bar{K}^{-1} = (1 - \Pi^T M^T) K^{-1} (1 - M\Pi)$ is the covariance with the sky signal rotated out.

The maximum likelihood estimate of $\hat{s}$ is the Wiener filter written in the Fourier domain for an arbitrarily long, stationary series, or

$$\hat{s}(\omega_i) = \frac{\tilde{S}(\omega_i)}{\tilde{S}(\omega_i) + \tilde{K}(\omega_i)} \tilde{d}(\omega_i).$$ \hspace{1cm} (C.56)

For frequencies where the drift has high signal to noise, the drift is equal to the data. Once the suggested prior of the drift falls below the noise, the filter rolls off the response.

Because $K$ is much more difficult estimate than $\bar{K}$, it is useful to think about maximizing an approximate likelihood for the drift assuming the sky has a small contribution to the covariance $K \approx \bar{K}$ through

$$\ln L \propto -(d - \hat{s})^T K^{-1} (d - \hat{s}) - \hat{s}^T S^{-1} \hat{s}.$$ \hspace{1cm} (C.57)

This gives a simpler form of the optimal filter which depends on the noise covariance directly as

$$\hat{s}(\omega_i) = \frac{\tilde{S}(\omega_i)}{\tilde{S}(\omega_i) + \tilde{K}(\omega_i)} \tilde{d}(\omega_i).$$ \hspace{1cm} (C.58)

\footnote{D. Spergel proposed this method for ACT drift subtraction.}

210
Calibration and Diagnostics

and has an equivalent linear operation in the time domain, \( \hat{s} = \Pi_O d \). Rather than solving for both the drift and the map at once, iterate using an approximate form of the drift as

\[
\hat{s}_0 = \Pi_O d \tag{C.59}
\]

\[
\hat{x}_{m,0} = \Pi(d - \hat{s}_0) \tag{C.60}
\]

and for \( i > 0 \),

\[
\hat{s}_i = \Pi_O (d - M\hat{x}_{m,i-1}) \tag{C.61}
\]

\[
\hat{x}_{m,i} = \Pi(d - \hat{s}_i). \tag{C.62}
\]

This can be implemented simply in an iterative mapmaker. In a first step, filter the data using Eq. C.58 to estimate the drift. Subtract this drift estimate from the raw instrument data and make a map. Then, project this map back onto the time domain and subtract it from the data and filter this using Eq. C.58 to find the next iteration on the drift. The next, subtract this drift iteration from the data, and make a map of this. This process can be continued until a desired convergence in the map is reached.

C.5 The redundancy of weakly-correlated normal variables

The multi-information is an information-theoretic measure for redundancy and is given by the Kullback-Leibler divergence between the full probability distribution and the distribution assuming the variables are uncorrelated,

\[
C(X) = \int \left( \prod_{i=1}^{N_{\text{det}}} dx_i \right) p(x) \ln \left( \frac{p(x)}{\prod_{i=1}^{N_{\text{det}}} p(x_i)} \right). \tag{C.63}
\]

(Here we have traded the sum in entropy sum for an integral.)

This can be simplified considerably for multivariate normal distribution with mean \( \mu_x \) and covariance \( K \),

\[
p(x) = \frac{1}{\sqrt{2\pi \det K_x}} \exp \left[ -\frac{1}{2}(x - \mu_x)K_x^{-1}(x - \mu_x) \right]. \tag{C.64}
\]

(We will assume that the mean has been removed from the data – e.g. by a common mode or high pass filter.) The divergence of two random variables \( X \sim N(\mu_x = 0, K_x) \) and \( Y \sim N(\mu_y = 0, K_y) \) is given by

\[
D_{\text{kl}}(X||Y) = \int \left( \prod_{i=1}^{N_{\text{det}}} dx_i \right) p(x) \ln \left( \frac{p(x)}{p(y)} \right) = \frac{1}{2} \ln \left( \frac{\det K_y}{\det K_x} \right) - \frac{N_{\text{det}}}{2} + \frac{1}{2} \text{tr}(K_xK_y^{-1}). \tag{C.65}
\]

This reduces to the calculation for the multi-information when the \( K_y \) is taken to be the diagonal of the \( K_x \) covariance matrix, \( K_y = \text{diag}(K_x) \). When \( K_y \) is the diagonal of \( K_x \), then the product of \( K_x \) with the inverse of \( K_y \) is 1 along the diagonal (dividing by itself), and the trace of this matrix is just \( N_{\text{det}} \), so that

\[
C(X) = \frac{1}{2} \ln \left( \frac{\det K_y}{\det K_x} \right) = \frac{1}{2} \sum_{i=1}^{N_{\text{det}}} \ln \left( \frac{\sigma_i^2}{\lambda_i} \right). \tag{C.66}
\]

In the second equality, we have used the fact that \( \det K_y \) is the product of the variances of each detector, and \( \det K_x \) is the product of the eigenvalues of the covariance matrix \( \{\lambda_i\} \). We can simplify this calculation considerably for the case of small correlations. Let the full set of correlations be \( K + \delta K \), where \( K \) is the diagonal of the covariance matrix and \( \delta K \) is its off-diagonal
C.5 The redundancy of weakly-correlated normal variables

Elements. The eigenvectors of $K$ are just the unit vectors (because $K$ is diagonal) and the eigenvalues are the variances of each detector, $\sigma_i^2$. The off-diagonal elements $\delta K$ will perturb these, taking $\lambda_i = \sigma_i^2 \mapsto \lambda_i + \delta \lambda_i = \sigma_i^2 + \delta \lambda_i$. The multi-information can then be expanded for small $\delta \lambda_i$,

$$C(X) = \frac{1}{2} \sum_{i=1}^{N_{\text{det}}} \ln \left( \frac{\sigma_i^2}{\sigma_i^2 + \delta \lambda_i} \right) \approx -\frac{1}{2} \sum_{i=1}^{N_{\text{det}}} \delta \lambda_i. \quad (C.67)$$

Note that the multi-information must be positive, and this becomes manifest in Eq. C.71. The $\delta \lambda_i$ are given by ordinary perturbation theory because the full covariance is a symmetric matrix,

$$\delta \lambda_i = \delta K_{i,i} + \sum_{k \neq i} \delta K_{i,k} \frac{\sigma_k^2}{\sigma_i^2 - \sigma_k^2}. \quad (C.68)$$

Because the eigenvectors of the unperturbed $K$ are the unit vectors, a matrix element such as $v_i^T(\delta K)v_k$ just picks out the entries of that matrix, $\delta K_{ik}$. The first-order term is zero because $\delta K$ has no elements along the diagonal. The multi-information is then

$$C(X) = -\frac{1}{2} \sum_{i=1}^{N_{\text{det}}} \delta \lambda_i \frac{\sigma_i^2}{\sigma_i^2 + \delta \lambda_i} = -\frac{1}{2} \sum_{i=1}^{N_{\text{det}}} \left( \sum_{k \neq i} \frac{\delta K_{i,k}^2}{\sigma_i^2(\sigma_i^2 - \sigma_k^2)} \right). \quad (C.69)$$

Considering the perturbation matrix $(\delta K_{i,k}^2)/(\sigma_i^2(\sigma_i^2 - \sigma_k^2))$, the sum is over all off-diagonal elements. For each off-diagonal contribution of $ik$ in the upper triangle to the sum, there is a contribution from $ki$ in the lower triangle. We can thus just sum over the upper triangle

$$C(X) = -\frac{1}{2} \sum_{k,i}^{\text{upper}} \left( \frac{\delta K_{i,k}^2}{\sigma_i^2(\sigma_i^2 - \sigma_k^2)} + \frac{\delta K_{i,k}^2}{\sigma_k^2(\sigma_k^2 - \sigma_i^2)} \right) = \frac{1}{2} \sum_{i,k}^{\text{upper}} \frac{\delta K_{i,k}^2}{\sigma_i^2 \sigma_k^2}. \quad (C.70)$$

The last term is just the cross-correlation matrix ($\rho$) entry squared, and we can convert from nats to bits

$$C_2(X) = \frac{1}{2 \ln 2} \sum_{i,k}^{\text{upper}} \rho_{ik}^2(X). \quad (C.71)$$

The quality factor described in the text has the normalization against the number of possibly unique off-diagonal terms

$$Q^2 = \frac{2}{n(n-1)} \sum_{i,k}^{\text{upper}} \rho_{ik}^2(X). \quad (C.72)$$

After running normality tests, we can either compute $\rho$ directly, or use the average of the cross-power from Fourier transforms that have already been calculated. This result is not known to exist in literature.
Bibliography


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


Laughlin, C. (1978). $^2P^0_{1/2}-^1S^0_0$ and $^3P^0_{1/2}-^1S^0_0$ intercombination-line transitions in helium-like ions. Journal of Physics B Atomic Molecular Physics, 11:L391–L394.
BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


235


