Radio Spectroscopy Applied to a Search for Highly Redshifted Protogalactic Structure

A thesis presented by

Jonathan Weintroub

to

The Division of Engineering and Applied Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the subject of

Applied Physics

Harvard University

Cambridge, Massachusetts

May 1998

Abstract

An experiment to detect proto-galactic structure at high redshift is described. We attempt to observe this in the 1420 MHz line from neutral atomic hydrogen in the redshift range $4.7 \leq z \leq 5.5$ ($219 \leq \nu \leq 251$ MHz). The expected signal is very weak and we need a large aperture and a very long integration time for a detection to be feasible. This is achieved by building a dedicated and specialized instrument, which allows the experiment to run continuously, in the background, at the largest single dish reflector in the world—the Arecibo radio-telescope.

The motivation for the experiment is to provide an observational link between the very smooth conditions presumed to exist in the primordial universe, as confirmed by observations of the Cosmic Microwave Background (CMB) at $z \sim 1000$, and the clumpy conditions that we observe locally in the hierarchy of galaxies, clusters and superclusters. The introduction includes a literature survey and an outline of relevant cosmological theory. A discussion of high-redshift astronomy, including past and current searches for protoclusters, is followed by a description of our approach to the problem.

For simplicity a point feed is used to illuminate the spherical reflector at Arecibo. The design of the optics is discussed—based on this analysis and the radiative characteristics of neutral hydrogen gas we estimate that it is feasible to detect a $10^{14} M_{\odot}$ cloud in a long, though not unachievable, time span. A custom designed radioastronomy receiver backs up the feed system. It includes analog front-end amplifiers and filters, local oscillators and mixers, and an eight kilochannel hybrid spectrometer with 10 kHz spectral resolution over the survey band.

We have written software to automatically reduce the large quantity of data gathered. Radio frequency interference (RFI) pervades our system, and analysis problems include the development of robust methods for co-adding data in the presence of time-varying RFI, and matched filtering to recognize candidate celestial signals in noisy data by their signature in time and frequency. Our sensitivity is limited by interference rather than thermal noise. The outcome of the experiment is negative, and we conclude by setting a limit on the mass of protoclusters in our search volume.

Acknowledgments

My advisor, Paul Horowitz, has been a constant source of ideas and encouragement, ever since he talked me into quitting my regular job to head back to graduate school in 1993. I think one has to work directly with Paul to appreciate his unique brand of infectious enthusiasm, and his bottomless supply of energy—this has been an opportunity like no other! I am grateful to the two other longstanding members of my research committee, Patrick Thaddeus and R. Victor Jones, for advice, and for making time for me, both at our required meetings and when I prevailed on them informally. Unfortunately, Professor Jones was unable to attend my final oral, and I am most grateful to Woodward Yang, who stepped in to serve on the committee at very short notice. Cos Papaliolios has been a most supportive, helpful, and friendly person to have around, whether in discussing technical issues, or life in general. He agreed to serve on the committee for the final oral, and I am grateful for his formal input in that forum. I thank William Paul and Robert M. Westervelt who served short stints as my CHD advisors, and Michael McElroy who served on my qualifying oral committee.

The Harvard SETI lab in Cambridge has been an enjoyable, collegial and most educational place to work. It's been a pleasure to get to know Chip Coldwell, Derrick Bass, Jonathan Wolff, Suhail Shah and Eric Wey. In addition to contributing to the general atmosphere, Darren Leigh helped extensively with software development among other things, and Nick Shectman with printed circuit board layout. Derrick Bass designed and built a spectrometer board tester. Most of the spectrometer boards were assembled by Jeremy Rutman. I greatly enjoyed discussions with Tom Hayes, whose relaxed demeanor and original way of looking at the world was always refreshing.

Other Harvard people who contributed in many ways include the ever cheerful Carol Davis, who was always most helpful and generous with her time, and Pauline Mitchell of the DEAS office, who sadly passed away during my time here, and who was supportive in helping me through the application process and the various grad school milestones. I express my thanks to Susan Wieczorek, and the rest of the DEAS academic office. In the machine shop, Pierre Crevoiserat and Stan Cotreau trained me

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Since the discovery of "excess antenna temperature" by Penzias and Wilson in 1965 [1] and its interpretation as the cosmic microwave background (CMB) by Dicke *et al.* [2], mapping of the CMB's intensity has revealed its very smooth character. In addition to giving strong evidence for the universe originating in a condition of high temperature and density—the hot big bang model—the CMB is thought to be a useful tracer of the distribution of the matter at recombination. This is the case because the photons scatter last from the ionized matter just prior to recombination—the "surface of last scattering"—and then travel unimpeded through a post-recombination transparent universe. It seemed from the early maps of the CMB that this matter was very smoothly distributed in volume. This raised an important cosmological question: how did the large scale structure that we observe in the universe today—the hierarchy of galaxies, clusters and super-clusters—evolve from an essentially featureless primordial universe?

It was expected that there would be some low level structure in the CMB, below the sensitivity level of the instruments used to do the mapping. Quadrupole (and higher order) anisotropy in the CMB is the signature of density inhomogeneities in the primordial gas, which grew and collapsed, primarily under gravitational force, to form the structures we see today. To try and detect this anisotropy in the CMB, the measurements were repeated with more sensitive instruments, generally balloon borne, over the years (see, for example, Lubin *et al.* [3]). The experiments successfully

detected *dipole* anisotropy (which is due to our motion with respect to the CMB) at the level of a part in 10^3. As far as the quadrupole term is concerned, all experiments produced non-detections (thus setting upper limits), until 1992, when results from the Cosmic Background Explorer (COBE) satellite revealed quadrupole anisotropies at the level of approximately a part in 10^5 [4].

The observed anisotropies constrain cosmological theory. They are used to derive the spectrum of density fluctuations at recombination $(z \sim 1000)^1$. These are in turn used to initialize computer codes which apply many-body simulation to predict the large scale structure in the universe as a function of redshift. The predictions can be compared to observation where available. But at redshifts $z > 5$ there are very few observations.

Observation is difficult at large redshifts because the signals are weak, and because of problems separating out foreground objects. The objects which were first seen at high redshifts in emission are the quasars and active galactic nuclei (AGN), which are most common at redshifts between about 2 and 3 and can be seen out to a redshift of about 5. The absorption spectrum in the quasar continuum at redshifts corresponding to the foreground of the quasars is a rich source of observational data. Recently, the detection of seven Lyman-α (Ly-α) emitting galaxies at $z = 5.7$ has been claimed [5], as well as a single galaxy at $z = 5.34$ [6]. It is thought that large scale structures formed at redshifts around and beyond $z = 5$, so theories of structure formation would benefit from more observations at these epochs.

The purpose of the experiment described in this thesis is to add some observational data to the history of the universe between recombination and the quasars. As first suggested by Sunyaev and Zel'dovich [7], we are using the hyperfine emission line from neutral atomic hydrogen as a probe of this structure. To allow the long integration times needed to make a sensitive measurement we have set up a "piggy-back" experi-

---

[1]The redshift parameter z is used in observational cosmology to specify a particular epoch because it is easily related to directly observable quantities. It is defined by $\lambda_{\mathrm{obs}}/\lambda_{\mathrm{em}} = 1 + z$, where λ_{obs} is the redshifted (observed) wavelength, and λ_{em} is the (non redshifted) wavelength of the radiation when emitted. z can, in principle, be related to distance and look back time, but these relationships depend on the value assumed for Hubble's constant, and also on the assumed cosmological model.

ment at Arecibo. The catwalk which is used for access to the feed platform intersects the primary reflector's focal surface. It is thus suitable as a strut from which we have hung a fixed feed antenna, facilitating continuous transit observations. A custom dedicated receiver, built around a hybrid spectrometer with 32 MHz instantaneous bandwidth and 10 kHz binwidth, completes the setup. This allows the experiment to run largely independent of scheduled observing, and of the Arecibo upgrade.

1.1   Observations of the Cosmic Microwave Background (CMB)

The differential microwave radiometer (DMR) mounted on the Cosmic Background Explorer (COBE) satellite has dramatically improved our knowledge of the CMB by mapping its anisotropies [4, 8]. The anisotropies, which are superimposed on the overall CMB temperature of 2.724 K, include a dipolar term of 3.343 ± 0.006 mK caused by the motion of the solar system with respect to the rest frame of the CMB (or, equivalently, to the surface of last scattering). If the dipole term is removed, the residual quadrupole amplitude is only 13 ± 4 μK^2. In spite of the low amplitude this measurement has generated great interest, because the ripples in the CMB are thought to be the seeds of cosmic structure. The measurements are used to provide the initial conditions in theoretical models of structure formation.

It is important to remember that the COBE maps have limited angular resolution. The DMR has an inherent resolution of 7°, and the maps are Gaussian smoothed to a resolution of 10°. They are useful in understanding structure formation on similarly large angular scales. Currently experiments are being conducted to improve the angular resolution of the CMB measurements. A collaboration between groups at Jodrell Bank and Cambridge University is taking measurements at 5° scales on a high and dry mountain top in Tenerife, and these measurements are consistent with the

---

[2]The CMB fluctuations are sometimes expressed as the root-mean-square temperature fluctuation over the whole sky, with the dipole term removed; expressed like this, $\Delta T_{rms} = 36 \pm 4$ μK (on 7° angular scales).

COBE data [9]. Also, balloon borne experiments are being conducted at the South Pole. Results from these experiments should add confidence to theoretical models of structure formation on smaller scales.

The other major result from COBE is confirmation that the spectrum fits the Planck spectrum B_ν, expected for a black body at 2.724 K, extremely closely. This gives confidence that the measured anisotropies discussed above are due to primordial fluctuations, rather than to the interaction of the CMB with matter in the foreground, since such interactions could also perturb the spectrum. An example of an interfering interaction is the Sunyaev-Zel'dovich effect, where the CMB photons encounter a gas of hot electrons and interact with it by inverse Compton scattering. The signature of this effect is a region of lower than expected temperature on the Rayleigh-Jeans side of the CMB's black body spectrum. This effect has been detected in the direction of certain rich clusters [10].

Anisotropies in the CMB are an important tracer of the mass distribution at recombination. The relationship between radiation anisotropy and matter inhomogeneity is expressed by the Sachs-Wolfe equation [11]. If the departure from homogeneity causes the expansion of the universe across the Hubble length, measured in two orthogonal directions, to differ by a fractional amount δ_h, the Sachs-Wolfe equation is

$$\frac{\delta T}{T} = \delta_h \tag{1.1}$$

where $\delta T/T$ is the quadrupole CMB anisotropy. Another way to think about the Sachs-Wolfe effect is in turns of gravitational redshift. Inhomogeneities in mass result in variations in gravitational potential. The CMB photons are redshifted as they climb out of the gravitational wells, and this redshift $\delta\nu/\nu$ maps directly to the observed anisotropies in CMB temperature $\delta T/T$ [12].

1.2   Theories of large scale structure formation

Views of large scale structure formation have historically been divided into "top-down" and "bottom-up" scenarios [13]. In the first of these, initial structures form on very large scales. These massive clouds of gas (protoclusters) later fragment into the smaller structures which become galaxies. Conversely, in the bottom-up scenario, individual smaller protogalaxies form first, and later cluster together to form the larger structures. The seeds of all structure are the density inhomogeneities in the primordial matter but it turns out that the way in which these seeds grow to form the large scale structure—at least as far as theory predicts—depends on the initial composition of the dark matter in the universe.

Big-bang nucleosynthesis theory suggests that 10% or less of the matter in the universe is baryonic [10]. The non-baryonic remainder is not luminous, and interacts only gravitationally. No one knows for sure what the dark matter is made of. We have confidence in nucleosynthesis because predictions of the abundances of the lightest elements agree well with those measured in various astronomical environments. Measurements, using the virial theorem, of the rotational dynamics of stars and gas in galaxies, and of galaxies in clusters, suggest that much more mass is present than the light output from these objects indicates. The infall towards large clusters, such as Virgo, also suggest that the universe is largely comprised of dark matter. The dark matter component inferred from dynamical arguments alone could be non-luminous baryonic matter, but based on nucleosynthesis, we expect a non-baryonic component.

The non-baryonic matter may consist of massive particles (for example, weakly interacting massive particles or "WIMPS", predicted entirely theoretically by supersymmetric theories) or low mass particles (leptons, such as neutrinos). In the low mass case it is expected that the particles were moving relativistically at the epoch of matter-radiation equality—when the energy densities of matter and radiation were equal, at a redshift $z_{eq} \sim 23,000$. The temperature of these particles today is less than typical galactic escape velocities [12]. The theories which assume low mass particles are known as the hot dark matter or HDM theories. Conversely, massive particles move more slowly and are associated with cold dark matter (CDM) theories. The

epoch corresponding to z_{eq} is when the gravitational growth of density fluctuations commences. The relativistic particles tend not to cluster, while the "cold" particles start to cluster on small scales at this time. This is the origin of small scale clustering in CDM models, while in HDM models the clustering commences later, and on larger scales.

After the measurements of CMB anisotropy were used to normalize many-body simulations based on HDM and CDM models it became apparent that the CDM models reproduce present observations more accurately. More strictly, CDM models reproduce the density spectrum $d\rho/\rho$ over a greater range of scales than HDM. In particular, in the HDM model, large scale structure evolves too late. There are problems with CDM too: it is not able to predict the observed superclusters and cluster drift velocities. On balance, though, the CDM model is currently favored, and N-body theorists mostly predict a bottom-up formation scenario, with structure forming first on galactic scales, and later clustering of galaxies. However, the nature of the clustering is poorly understood, even using CDM models.

1.3   Probing the high redshift universe

When we observe at large distances, or, equivalently, high redshifts, we see objects not as they are now, but as they were when the light was emitted. It is assumed that, on cosmological scales, the universe is homogeneous and isotropic—so we expect the large scale statistical properties of the structures we observe at high redshifts to be typical of our local universe at earlier times. This means that we can probe the history and evolution of the universe through high redshift observations.

The observation with the highest possible redshift is that of the CMB. Before recombination, Thomson scattering by free electrons rendered the universe opaque. Between the CMB at $z \sim 1000$ and the distant quasars at $z = 5$ there are very few confirmed observations. Below $z = 5$ there is a rich observational literature. In this section I give a brief survey of high redshift observational astronomy, with a bias towards those observations which are relevant to structure formation.

Quasars are tremendously bright objects which are thought to reside in the center of primordial galaxies; they are the brightest, relative to the underlying galaxy, of the family of objects known as active galactic nuclei (AGN). Quasar is an acronym for quasi-stellar radio source [14, 15]—this is an historical misnomer, since many radiate more strongly in optical bands. A quasar is generally an isolated object which may have some fuzz surrounding it with spectral characteristics typical of an elliptical galaxy. (If a galaxy is clearly visible surrounding the active object it is generally referred to, rather vaguely, as an AGN rather than a quasar [16]. There are eleven classes of AGN [17], the quasars being the most relevant to this work). Quasars are thought to be gravitationally powered, and very often out-shine all of the stars (typically numbering 10^{11}) in their host galaxies.

Quasars are important in observational cosmology because, being so bright, they can be seen at high redshifts—as high as $z = 4.89$ [18]—but they are most abundant at redshifts slightly over 2, and, oddly enough, decline sharply in abundance at lower redshifts. Indeed, at $z = 2$ quasars were a thousand times more common than they are today [13]. Note that the abundance is measured relative to that of the co-existing galaxies, and that the distribution has been corrected for the selection biases associated with counting distant objects.

Even when they were most abundant, quasars were still about 100 times less common than galaxies. There are two ways to interpret this; either only selected galaxies have quasars; or most galaxies have them, but their lifetimes are significantly shorter than those of their hosts. The truth may also lie somewhere in between these two extremes. It has been speculated that quasars burn brightly when the host galaxy is young, but most have burned out by the current epoch, with the remnant being a black hole at the galactic center. While the mechanism is not understood, it is thought that quasars are associated in some way with the process of galaxy formation [13]. Their presence at a particular redshift indicates that galaxy-sized objects have formed at that redshift [18].

Quasars at high redshift are also observationally important as probes of the foreground gas—they are bright background continuum sources, conveniently located at high redshift, for absorption line studies. The Lyman-α line (1216 Å), due to an electron
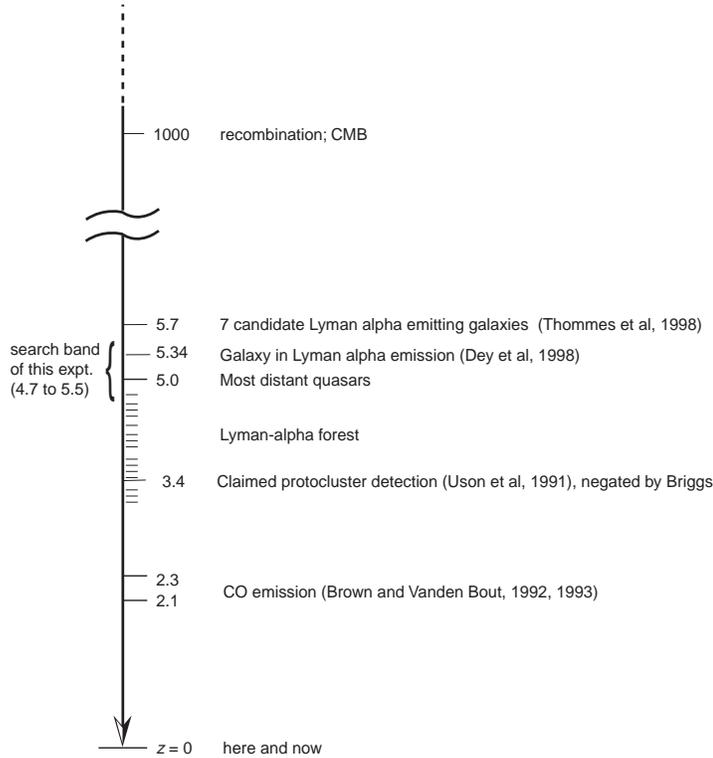
Figure 1.1: High redshift observation timeline.

transition from the first excited to ground state in atomic hydrogen, is very important in the study of the quasars and their foreground spectra. The Lyman continuum break starts at 912 Å (the wavelength just capable of ionizing a neutral hydrogen atom) and is also seen in quasar spectra[3]. A high redshift quasar spectrum typically shows strong Ly-α emission at the quasar redshift, and then a multitude (of the order of hundreds) of narrow absorption lines at intervening redshifts. The absorption spectrum (for $z > 1.8$) is redshifted into the visible, and is known as the Ly-α forest. The forest of narrow lines gets thicker at higher redshifts. In about 10% of cases (lines-of-sight) the absorption spectrum also includes a broader saturated (or damped) line.

The discrete narrow lines suggest that the intergalactic medium (IGM) at $z < 5$ is inhomogeneous. Each narrow line corresponds to a relatively small (sub-galactic scale) hydrogen cloud, while the damped lines correspond to galaxies, or protogalaxies, in the line-of-sight. Associated with the damped lines are absorption lines due to

---

[3]The observed wavelength in both cases is, of course, $\lambda(1 + z)$.

heavy elements in the galactic halo, while the smaller clouds appear to have low heavy-element abundance. It was pointed out by Gunn and Peterson [19], using the observations of Schmidt, that the number density of neutral hydrogen atoms in the IGM would have to be incredibly low, around 10^{-11} cm^{-3}, in order for the light from the quasar not to be completely extinguished by Ly-α absorption. For comparison the number density of hydrogen atoms in interstellar space is about 1 cm^{-3}. The accepted explanation is that the IGM is mostly ionized and that the Ly-α forest is due to small pockets of neutral gas [13]. The source of the ionizing radiation is the ultraviolet emission from quasars, and also possibly from hot stars in starburst galaxies (galaxies in which the vigorous first phase of star formation is taking place).

The lack of a Gunn-Peterson cutoff in the spectra of the most distant quasars indicates that the IGM is mostly ionized by $z \sim 4.7$ [20], and probably at least until $z = 5$. But is the IGM ionized beyond $z = 5$? The question of when, and how, the IGM was re-heated to the point of re-ionization is both an interesting, and an ill understood one. CDM models suggest that the re-heating of the IGM would be unlikely to have started before $z = 10$. There are other models which hypothesize very early re-heating in epochs $z > 10$. Possible sources of energy for re-heating include energy from decaying particles or a very early phase of star formation. However rarity of observations for $z > 5$ severely limits our understanding.

Recent results have pushed measurement of large structures to $z \sim 3$. This adds to our knowledge of the degree of clustering on different mass and size scales as a function of redshift. Cohen *et al.* have re-observed with the Keck telescope certain of the primordial galaxies imaged by the Hubble Space Telescope (HST), and reported redshift clustering [21]. They suggest that the observed structures could be the high redshift counterparts of the largest scale structures ("walls"), observed locally. A technique developed by Steidel and Hamilton and published in 1992 [22] takes advantage of the Lyman continuum break to image very faint high redshift galaxies in the foreground of quasars. The method differences the images obtained with three broad bandpass filters, the bluest of which falls completely below a damped Lyman limit system, such that the background quasar is completely extinguished in the image. The method has recently been applied to the detection of a significant large cluster

of galaxies at a redshift $z \sim 3.1$ [23]. They suggest that this cluster may be the primordial progenitor of an Abell cluster.

Brown and Vanden Bout have searched for molecular line emission from CO and report discoveries at $z = 2.29$ from a gas rich proto-galaxy [24], and at $z = 2.14$ from a damped Ly-α system in the foreground of a quasar [25]. In both cases they detected CO(J =3–2) at 110.1 GHz and CO(J =4–3) at 146.8 GHz. These detections demonstrate that CO emission provides a useful complement for other techniques used to study high redshift structures. Barvainis, Ohta, Omont, and others have detected CO at higher redshifts(see reference [26], and other papers cited therein). The latest detection is of the J =5–4 transition at a redshift of $z = 4.41$, with an earlier detection at a still higher redshift of $z = 4.69$.

Very recently, the discovery of galaxies at $z > 5$ has been claimed. The two optical detections discussed here are based on Ly-α emission. Thommes *et al.* [5] reports the detection of seven candidate galaxies at $z = 5.7$, found during a preliminary phase of the Calar Alto Deep Imaging Survey (CADIS). They used a Fabry-Pérot interferometer and a series of narrowband *veto* filters to discriminate against foreground spectral line emission. In this way, an initial sample of 147 emission line galaxies was whittled down to the seven reported candidates. This detection is not spectroscopically confirmed, and the authors expect that, in view of possible foreground contamination, "two or more of the candidates will turn out to be genuine Ly-α emitting galaxies at $z = 5.7$".

Dey *et al.* [6] report the discovery of a single galaxy at $z = 5.34$. They claim the first galaxy spectroscopically confirmed to be at $z > 5$. Interestingly, they were using the Lyman break technique pioneered by Steidel *et al.*, and the discovery of Ly-α emission was serendipitous. The emission spectrum suggests to the authors that the object is a star-forming galaxy, rather than an AGN.

These two results were reported shortly before this writing, the latter in preprint form. They are dramatic claims, and it would be interesting to see the results of re-observations. The results, if they hold up, bolster directly the existing indirect evidence (for example the presence of metals in excess of primordial abundances in

Ly-α systems) of galaxy formation at $z > 5$.

1.3.1 Structure in neutral HI

The detection of hyperfine emission from neutral hydrogen by Ewen and Purcell in 1951 [27] spurred a great deal of activity in radio astronomy. The line provides a tracer for the structure of the Milky Way and other galaxies, while an abundance of HI indicates star formation activity [28]. But, for the purposes of this experiment, the significance of hydrogen is that it is the most abundant and simplest constituent of the matter in the post-recombination universe, and so is potentially useful as a tracer of primordial structure. Note, here, that ionized hydrogen does not emit in the hyperfine line.

Probing the primordial universe by examining the 21 cm line emission from neutral hydrogen at high redshift was first suggested by Sunyaev and Zel'dovich (1972) [7]. They predicted that structure would form in massive flattened clouds of gas, dubbed Zel'dovich pancakes, and that these may be visible in the hyperfine line in emission. There have been relatively few subsequent searches for protoclusters, and all have yielded negative results (see §1.5). More recently the prevailing theoretical picture of how structure forms has changed from a "top-down" to a "bottom-up" scenario. Nevertheless, observing HI in emission at $z > 5$ remains an interesting objective which could help to constrain theory, and there is considerable interest in the subject. In particular, the Giant Metrewave Radio Telescope (GMRT), an array consisting of 30 steerable 45 m dishes, has detection of high redshift HI emission as one of the primary scientific objectives [29]. The GMRT is located about 80 km north of Pune, India, and is currently being commissioned.

A series of three theoretical papers ("the GMRT papers"), written largely as preparation for the GMRT observations, provide useful background for our Arecibo experiment as well. The first [30], published in 1993, examines the CDM and HDM cosmologies in the light of the COBE anisotropies. Their conclusion, as discussed already in §1.1, is that CDM is a better model. Predictions are made under CDM as to the detectability at three redshifts. At lower redshifts ($z = 3.3$ and 1.3) protoclusters

may be detectable, but such detection is critically dependent on the fraction of mass which is neutral hydrogen. At higher redshifts the predictions are rather pessimistic; around $z = 5.1$ individual proto-condensates will not be detectable, though clusters may be, while at $z = 8.4$ a predicted flux due to protoclusters of a mere 0.02 mJy is not considered to be detectable at all. All of these redshifts are those which the GMRT will be able to observe.

The expected line profile is computed in the second paper (1995 [31]). Peak fluxes of 0.5–1.1 mJy are predicted, with line-widths in the range 0.3–1.8 MHz, originating from a composite large structure consisting of smaller scale clumps. Again the results are critically dependent on the fraction of hydrogen in the protocluster which is neutral. Radio maps showing possible emission from HI clouds at two redshifts are derived from the many-body simulations in the third paper, which was published in 1997 [32]. In this paper a possible expected complex lineshape containing peaks and valleys, and due to the expected bottom-up evolution of the protocluster, is published. Both of these papers are cautiously optimistic about the ability of the GMRT to detect protoclusters at $z = 3.3$, and quite pessimistic about the prospects at higher redshifts. Our own feasibility calculation in §1.6 is simpler, and more optimistic in its conclusions regarding detectability at $z = 5$.

Another relevant theoretical paper from 1997 is by Madau, Meiksen and Rees [33]. They suggest that the IGM at $z > 5$ may be heated by soft X-rays from QSOs, soft X-rays from early galactic halos, or the continuum UV radiation from an early generation of stars. They suggest a beam-differencing experiment to find a patchwork of emission, with emission corresponding to un-ionized areas, and the lack of 21 cm emission corresponding to an ionized IGM. In a more optimistic conclusion than that of the GMRT papers discussed in the previous two paragraphs, they suggest that the GMRT may be able to detect this patchwork in emission at redshifts $z = 5.1$ and 8.4. Note that this view is also somewhat different from the conclusions reached by Scott and Rees discussed above. It is as well to remind ourselves at this point that, observationally speaking, we know very little about the state of the hydrogen—both its distribution, and its ionization—at $z > 5$.

1.3.2 HI in absorption

Scott and Rees have suggested that, depending on the cosmological history of hydrogen spin temperature, it may be more feasible to detect protoclusters in absorption [34]. The background source for this absorption is the CMB, which, at high redshift, is at a higher temperature given by $2.7(1+z)$. For the clouds to be seen in absorption, it is necessary that the spin temperature is less than the CMB temperature at that epoch. They suggest that more work be done on the spin temperature history to determine whether the line will likely be seen in emission or absorption. Our experiment is sensitive to absorption lines, if such exist, and we keep in mind when scrutinizing the data the possibility that the line we are searching for may be seen in absorption.

1.4 Motivation for this experiment

The purpose of this experiment is to try to provide some observational data at epochs between the CMB and the quasars. The structures we are looking for are simply density variations in the primordial gas. If the density variations are large enough, the gas will decouple from the Hubble flow, and self gravitating condensations of gas will form—precursors to the formation of galaxies and clusters.

Simply detecting HI structures at high redshift would be a useful result, and the searches discussed in §1.5 limit themselves to this simple objective. Better would be to measure some of the properties of the structures, such the mass, volume, number density and redshift distributions. These features are summarized in the spectrum of density fluctuations, which is expressed as the normalized density variation $\delta\rho/\rho$ as a function of the scale size of the structure. This spectrum is, of course, also a function of z. It has also been suggested that, if HI clouds are detected, that they be observed at other wavelengths, with a view to detecting heavier elements or simple molecules, such as LiH [35].

1.5 Searches for high redshift HɪI emission

The first search for protoclusters looking for HI in emission was conducted by Davies, Pedlar and Mirabel using the Mk 1A radio telescope at Jodrell Bank in 1977 [36]. Two 2.5 MHz bands were searched, at frequencies corresponding to $z = 3.33$ and $z = 4.92$. The authors report a negative result, and derive upper limits on the mass ($\leq 3 \times 10^{15} M_\odot$) and total number ($\leq 10^6$) of protocluster condensates. These limits were derived using the Sunyaev-Zel'dovich model. The authors note, as an additional caution, that Sunyaev and Zel'dovich, in a subsequent paper published in 1975, seemed to favor $z > 5$ as a more likely region for protocluster formation.

In 1978 and 1979 Bebbington used the 6C telescope, an interferometer located at Cambridge, U.K., to search the sky north of 82° at 151 and 152 MHz [37]. This corresponds to a considerably higher redshift of $z = 8.4$, while the bandwidth is sufficient to probe a 1% redshift range around this value. The result again was negative, with no sources at the 2σ level in the search volume. They conclude that their observations combined with those of Davies *et al.* rule out the pancake model of structure formation.

Hardy and Noreau used the VLA for two searches, reported in 1987 and 1988 [38, 39]. The searches were at slightly different center frequencies and different bandwidths (the center frequency was adjusted to minimize the effect of interfering carriers in the search band, while in the first search half of the search band was unusable due to the inadvertent inclusion of a filter in the analog system). The first search was at 331.85 MHz with a bandwidth of 6.25 MHz, and the second at 331.4 MHz in a bandwidth of 12.5 MHz, or $z \simeq 3.3$ and $\Delta z = 0.08$ and 0.16 respectively. Their search strategy was targeted to searching the neighborhood of quasars. They conclude by their negative result that the redshift they searched is too late in the history of the universe for significant amounts of HI to exist in large scale concentrations.

Wieringa, de Bruyn and Katgert also chose to search around $z = 3.3$ [40]. The observations were made between 1985 and 1989, so, for most of the time, they did not have the benefit of Hardy and Noreau's conclusions regarding interesting redshift ranges. They used the Westerbork Synthesis Radio Telescope operating at 327 MHz,

and searched six separate fields and a total of 32 square degrees. The negative result confirms the conclusions of Hardy and Noreau, that no, or very few, large neutral structures have survived until this point in the history of the universe.

Uson, Bagri and Cornwell published two papers in 1991 both describing searches at the VLA. The first, at $z = 3.3$, reports a negative result [41]; the second, in addition to detecting HI absorption against a radio continuum source, claims the first detection of highly redshifted HI emission from a Zel'dovich pancake at $z = 3.3970$ [42]. Unfortunately, this detection was negated at a high ($> 10\sigma$) level of confidence by Briggs, Sorar and Taramopoulos [43] in a re-observation at Arecibo in 1993. Briggs *et al.* do confirm the absorption feature.

The construction of the Giant Metrewave Radio Telescope (GMRT), currently coming on line in India, [29], has, as a primary scientific motivation, a search for high redshift HI emission. With an effective area comparable to Arecibo, resolution comparable to the VLA, and the ability to search at redshifts $z = 8.3, 5.0, 3.37$ and 1.33, with bandwidths of 6, 10, 32 and 32 MHz respectively, in a relatively RFI-quiet environment, this exciting new instrument should set new limits for sensitivity of similar searches.

The table above compares the observational parameters and summarizes the results of the experiments discussed here. Note that an unusual feature of the experiment discussed in this thesis is that the search volume is greater than that of the other experiments (mostly a consequence of our wide instantaneous bandwidth). Also, by using Arecibo's large aperture, we had hoped to have greater sensitivity—unfortunately our sensitivity was significantly degraded, primarily because of radio frequency interference (RFI). The sensitivity of our experiment will be discussed from a theoretical standpoint in §1.6, and more direct estimates from experimental data will be made in chapter 3.

The GMRT search also interrogates greater volumes and is more sensitive than previous searches. So far results have not been reported.

search	telescope	z_0	Δz	volume $(10^6~\mathrm{Mpc}^3)$	rms noise (mJy)
Davies *et al.*, 1977	Mark 1A	3.33	0.033	0.36	50
	(Jodrell)	4.92	0.062	0.97	15
Bebbington, 1978/9	6C	8.4	0.084	12	5
	(Cambridge)				10
Hardy & Noreau, 1987	VLA	3.28	0.081	1.0	12
Noreau & Hardy, 1988	VLA	3.29	0.16	2.5	12
Wieringa *et al.*, 1985-9	Westerbork	3.35	0.07	1.5	1.1
Uson *et al.*, 1991	VLA	3.4			
this work	Arecibo	5.1	0.8	100	
Swarup *et al.*	GMRT	8.3	0.37	96	3.7
		5.0	0.26	40	1.3
		3.37	0.43	40	0.5
		1.33	0.12	4	0.3

Table 1.1: Observational parameters of previous and current high-z HI searches, adapted from Wieringa *et al.* (1992) [40]. The quoted sensitivity for the Westerbork experiment is the average of that for six fields. GMRT data is taken from Swarup [29], with predicted noise assuming a ten hour integration. Volume calculation scaled to assume $\Omega_0 = 1$ and $H_0 = 100~\mathrm{km\,s^{-1}\,Mpc^{-3}}$.

1.6   Using Arecibo in a background mode

In 1993 a collaboration between Harvard, NAIC and MIT to undertake a background search for highly redshifted neutral hydrogen at Arecibo was arranged. An extract from the proposal submitted then to NAIC for observing time summarizes the results of a preparatory RFI survey undertaken at the observatory in 1990 [44].

In the fall of 1990, Burke, Conner and Lehár carried out observations at Arecibo in order to evaluate the interference environment there, with a view to starting a search for primeval hydrogen at about 200 MHz ($z = 6$)

using existing observatory equipment. The results of this work showed that television stations in Puerto Rico kept on their test patterns when they were not broadcasting, so if observations were to be made in the frequency range 180–212 MHz, private arrangements would be needed to have these stations leave the air during the early morning hours when they were not broadcasting programs. Surprisingly, the station using channel 13 left their carrier on but with no modulation, so from 212–250 MHz, the environment did not appear to be unduly hostile. There was occasional interference in the band 220–250 MHz, but it was not widespread and, except for a few persistent stations, the frequencies were clear most of the time. From 250–280 MHz, there is interference generally present; whether one can use the intervals between the carriers would have to be determined empirically.

The experiment discussed in this thesis has its origins in that 1990 survey trip. In discussions between Dr. Michael M. Davis of the National Ionospheric and Atmospheric Center (NAIC), which operates the Arecibo Observatory, and Professor Bernard F. Burke of MIT, the idea of hanging a feed from the telescope catwalk to do a background search originated. It would also be necessary to build instrumentation to back up the feed system. The idea lay fallow for a few years, until 1993 when it occurred to Dr. Davis that there might be some overlap with Harvard's fast Fourier transform (FFT) spectrometer work for the "Billion Channel Extraterrestrial Assay" (BETA) search for extraterrestrial intelligence (SETI) project [45]. As a result the collaboration was arranged, with our Harvard group taking primary responsibility for designing and building the spectrometer hardware, the MIT group providing guidance in astrophysics and data analysis, and NAIC co-ordinating and supporting on-site activities.

Using order-of-magnitude parameters, the feasibility of searching for highly redshifted HI emission using the Arecibo reflector in a background mode is examined here. We write the expected luminosity of such an HI cloud as,

$$\mathcal{L} = \frac{3}{4} N_H A_{21} h\nu \qquad (1.2)$$

where N_H is the number of HI atoms, A_{21} is the Einstein coefficient for spontaneous emission ($A_{21} = 2.85 \times 10^{-15}$ s^{-1}), h is Planck's constant and ν is the frequency of the emitted radiation. The initial factor of $3/4$ is the ratio of statistical weights for the hyperfine excited and ground states. The cloud is assumed to be optically thin, so that every emitted photon contributes to the luminosity. It is interesting to note that the value of the A_{21} coefficient implies that a single atom of neutral hydrogen will emit a photon spontaneously only once every 11 million years—it is the great abundance of hydrogen in the Universe which makes the line useful, both locally and cosmologically, in spite of the low transition probability.

It is instructive to express the luminosity in units of mass comparable to that one might expect in a protocluster. We set the unit of mass, allowing for about 90% dark matter, to about one tenth the virial mass of local rich clusters, or $M_{14} = 10^{14}$ $M_\odot$. In these units we write,

$$\mathcal{L} = 2.4 \times 10^{39} M_{14} \text{ erg s}^{-1} \qquad (1.3)$$

We assume a simple Euclidean geometry to estimate the total (integrated) flux,

$$S = \frac{\mathcal{L}}{4\pi D_H^2 z^2 (1+z)^2} \qquad (1.4)$$

where $D_H = c/H_0$ is the Hubble distance. We convert the total flux to a flux density by dividing by $\Delta\nu$, the width of the spectral line to obtain,

$$S_\nu = \frac{\mathcal{L}}{4\pi D_H^2 z^2 (1+z)^2 \Delta\nu} \qquad (1.5)$$

We take the velocity dispersion of the cloud, Δv, to be roughly 100 km s^{-1}, corresponding at $z = 5$ to $\Delta\nu$ of about 80 kHz. Substituting numeric values for the quantities in the above equation gives,

$$S_\nu = 3 M_{14} h_{100}^2 \text{ mJy} \tag{1.6}$$

where h_{100} shows the scaling of this estimate if H_0 deviates from the nominal value of $100 \text{ km s}^{-1}\text{Mpc}^{-1}$, that is $h_{100} = H_0/(100 \text{ km s}^{-1}\text{Mpc}^{-1})$.

In order to compute the feasibility of a detection, assuming this estimate for power flux, we also need to know the sensitivity and noise characteristics of the instrument. The calculation in §2.1.1 shows that, by using a point feed[4] suspended over the Arecibo dish we can illuminate about 150 m of the reflector with negligible loss due to spherical aberration. From this estimate, and other work by our collaborator at MIT, Ian Avruch (§2.1.2), we estimate the sensitivity of our receiving system to be 3.8 K/Jy. For a $10^{14} M_\odot$ neutral hydrogen cloud this corresponds to an increase in antenna temperature due to the cloud of $T_A = 11.4$ mK.

The signal competes with system noise due to losses in the feed and cabling to the feed, feed beampattern spillover, receiver noise, and galactic synchrotron background radiation. At these relatively long wavelengths galactic background radiation contributes significantly to the receiver temperature. We use an all-sky survey [46], adjusted based on the expected spectral index, to estimate the contribution at 235 MHz. Of course the contribution varies with sidereal time but has a gentle, persistent null of about 80 K off the Galactic center. A bench measurement of system front end shows a noise temperature of about 110 K. We estimate ground spillover noise to be about 20 K, giving an estimate of total system temperature $T_{\text{sys}} = 210$ K.

Observing for a long time and integrating is beneficial, in that the signal grows faster than the noise according to the Dicke radiometer equation. The signal-to-noise ratio improves with time as $\sqrt{B\tau}$ where B is the bandwidth of the receiver. Assuming we integrate until the signal is n_σ times the noise, we have

---

[4]The original idea was to have two feeds suspended from the catwalk, forming two sky beams separated East–West, and originally two feeds were built and installed. Unfortunately, due to interactions, one of the feeds had to be removed, and the experiment was conducted with a single feed.

$$\frac{T_{sys}}{\sqrt{B\tau}} = \frac{T_A}{n_\sigma} \qquad (1.7)$$

Rearranging, we obtain an explicit expression for the integration time,

$$\tau = \frac{1}{B}\left(\frac{T_{sys}}{T_A}n_\sigma\right)^2 \qquad (1.8)$$

The resolution binwidth of the spectrometer is about 10 kHz. However the reduction of the data includes matched filter processing, so that the effective bandwidth after processing is matched to the expected linewidth. Based on this assumption we set $B = \Delta\nu$ and take $n_\sigma = 4$. This gives an estimate for a 4σ detection of $\tau = 68 \times 10^3$ s. The sky beam width is about 0.75°, corresponding to a drift time of 180 seconds. Thus, given a $10^{14}M_\odot$ cloud contributing a change in antenna temperature ΔT_A, we expect a 4σ detection in about a year of drift scanning.

Several of the parameters which go into this estimate are quite uncertain, and that time to a detection with a specified signal-to-noise ratio is very sensitive to these. The most important of these parameters are the mass of neutral hydrogen in the protocluster, Hubble's constant, and the velocity dispersion in the cloud. The time to detection is sensitive to these according to the proportionality,

$$\tau \propto M_{14}^{-2} \cdot h^{-4} \cdot \left(\frac{\Delta v}{100 \text{ km s}^{-1}}\right) \qquad (1.9)$$

It is also clear that if a detection with a lower confidence limit is acceptable, subject perhaps to a tracking re-observation to confirm or negate it, then the time estimated here is reduced by $(n_\sigma/4)^2$.

The remaining chapters cover the design, construction and installation at Arecibo of the receiving system. This is followed by a description of data analysis procedures, cataloging of reduced data, and results obtained. We conclude by suggesting approaches to the design of improved instruments and of future searches for highly redshifted hyperfine emission.

Chapter 2

Experiment design

In order to run a "piggy-back" experiment, we designed, built, and installed a complete radioastronomy receiving system, except, of course, for the primary reflector (since the Arecibo Observatory has an adequate one). As described in chapter one, we have hung a feed from the catwalk used to access the feed platform, where it intersects the primary reflector's focal surface. With the dedicated receiver, our system facilitates continuous transit (or drift-scan) observations, enabling our experiment to run relatively independently of scheduled observing, and of the Arecibo upgrade. This chapter discusses the design of the receiving, data processing and data logging system.

The original plan was to have two feeds suspended from the catwalk, forming two sky beams separated East–West, and originally, two feeds were built and installed. In addition to an effective 3 dB improvement in signal-to-noise ratio over a single feed, two feeds help to discriminate radio frequency interference (RFI) from true celestial signals. In a drift-scan, celestial signals pass in a predictable way, first through the East beam and then through the West beam, while RFI generally couples simultaneously into both channels. Unfortunately, due to feed-feed interactions, one of the feeds had to be removed early on in the experiment, and we continued using a single feed. We planned to hang the second feed again, with a greater separation to minimize the interactions, but the required separation turned out to be mechanically awkward, and impractical given available resources. In the end all the data presented

here was gathered with a single feed system. The receiving system was built for two feeds, and is documented as a two-channel system here for the benefit of future users—since after the tenure of this experiment, the complete spectrometer will remain at Arecibo for use as a general-purpose instrument.
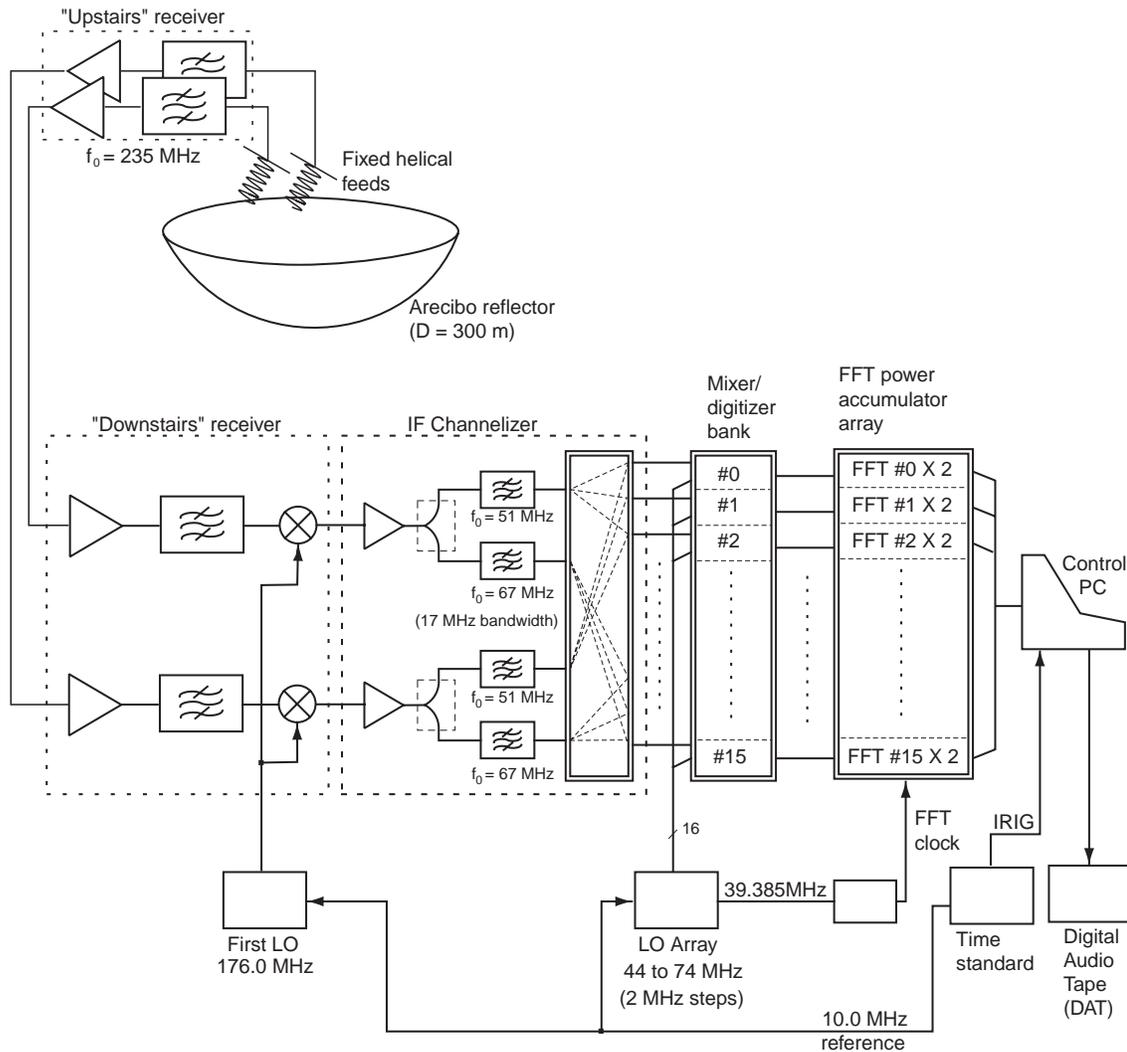
Figure 2.1: The early hydrogen search system block diagram. The details of the various blocks are broken down in later sections.

The natural place to start is a discussion of the optics of a point feed illuminating Arecibo's spherical reflector. The remainder of the system consists of analog receivers, mixers, spectrometer, and a computer for data logging and control. The spectrometer

is a hybrid unit, comprising a filter bank[1] followed by a bank of fast Fourier transform (FFT) spectrometer boards, with built-in power computation and real-time bin-wise integration. For an architectural overview, see the system-level block diagram (figure 2.1). We discuss the motivations behind the system design, and break down the designs of the various system components.

2.1 The optics of Arecibo's spherical reflector

Arecibo's reflector is somewhat unusual—a fixed spherical cap rather than the more usual steerable paraboloid. This design choice was made, when Arecibo was built in the 1950's, because the reflector is so large (300 m in diameter) that it is impractical and expensive to attach it to a steerable mount. Indeed, if large reflector is steerable, gravitational force tends to distort the reflector depending on its orientation, while wind force is a major structural load. Being a fixed reflector, the Arecibo dish avoids the issue of gravitational distortion, while its low profile construction, which takes advantage of a bowl-shaped valley in the limestone karst topography, is crucial given its location in a hurricane zone. Thus the Arecibo reflector itself can afford to be of relatively lightweight construction. It is made of mesh panels suspended by tensioned steel cables.

The sky beam formed by a fixed spherical reflector can be steered, albeit within a limited range, by moving the feed system to illuminate different sections of the reflector. The bore-sight axis of the formed sky-beam is the radial line from the feed position to reflector's center of curvature. However, because of the requirement to point the feeds, Arecibo's feed platform is a rather heavy and complicated triangular framework with a movable "azimuth arm" from which are suspended "carriage houses" with feeds and cooled receivers (see figure 2.2). The carriage houses run on tracks along the azimuth arm. In contrast to the reflector, the feed platform is exposed to the wind. It is suspended roughly 150 m over the dish from three main cables, which are strung over three towers built on the surrounding karsts and anchored into the ground with concrete blocks. As part of the Arecibo upgrade one of the

---

[1]Conceptually accurate; however a more complete description is "mixer-filter-digitizer bank"

carriage houses has been replaced by a Gregorian radome containing secondary and tertiary reflectors and a large complement of feeds and receivers. This is a still heavier structure, whose installation required upgrading the concrete anchors and cables which support the platform. An interesting report comparing the performance and economics of large spherical and parabolic reflectors was published by the North-East Radio Observatory Corporation (NEROC) [47].

Figure 2.2: Showing the position of the catwalk above the Arecibo reflector. Looking closely, one can make out the dual helical feeds suspended from the catwalk. Shortly after this photograph was taken one feed was removed. Notice the substantial structure of the feed platform suspended above the dish and the line-feeds suspended from the azimuth arm. The ground-screen which surrounds the reflector was added in 1993 as part of the upgrade project.

Unfortunately, the spherical reflector focuses to a radial line instead of a point. This manifests as spherical aberration if a conventional prime focus point feed is used to gather the reflected energy. Until the recent upgrade Arecibo has been outfitted with line feeds which correct for this spherical aberration. Line feeds are slotted waveguides, which collect the energy along the line of focus, with appropriate phase shifts

to sum the contributions in-phase. This is an inherently narrowband correction, so that a custom line feed is needed for each frequency, limiting the available operating frequencies. One of the main elements of the current Arecibo upgrade is the installation of an innovative Gregorian secondary and tertiary reflector system. These reflectors are shaped to correct for spherical aberration. Because this correction is derived from the equalization of ray paths, rather than of phase shifts, the correction is broadband. The high frequency limit is determined by the surface accuracy of the primary reflector.

For simplicity and low cost a fixed point feed is used in our experiment. The observed wavelength of the hyperfine line at a redshift of $z = 5$ is 0.21 m$\times(1 + z) \simeq 1.3$ m. At these long wavelengths it is adequate to use a point feed, because, as long as the illuminated aperture is not too big, the sphere deviates only a small fraction of a wavelength from the ideal paraboloid and the gain loss relative to that paraboloid is small. The deviation increases as larger sections of the reflector are illuminated and at some point becomes unacceptable. At this point the benefits of extra physical area are outweighed by the deleterious effects of the phase errors.

This section examines the optics, and finds the limits on aperture size assuming a point feed illumination. An analytical treatment is presented first, which assumes that the feed illuminates the dish "on-axis"; that is, along the radius from feed position to dish. The focal surface of a spherical reflector is a section of the concentric sphere with half the radius of curvature. It is sometimes called the "paraxial" surface, because rays close to the optical axis of an illuminated section (paraxial rays) focus at half the radius of curvature. As will be demonstrated in §2.1.2 it can be shown that for fairly large illuminated sections the best choice of focal point lies slightly closer to the reflector. This is also discussed in some detail in reference [47].

Next, we discuss a more general numerical treatment, which allows for squinting of the feed, so that a greater area of the dish can be illuminated, while reducing the effect of noise contributions due to the feed illuminating the 300 K warm ground (spillover, or vignetting). This model also allows the feed to be located a slight distance below the paraxial surface, since it turns out that this gives the better focus with point feed illumination. The analytic treatment (which is my own work) is in §2.1.1. An outline

25

of the numerical work (due to Ian Avruch of MIT) and pertinent results, are given in
§2.1.2.

2.1.1 Analytic maximization of gain with point feed illumination
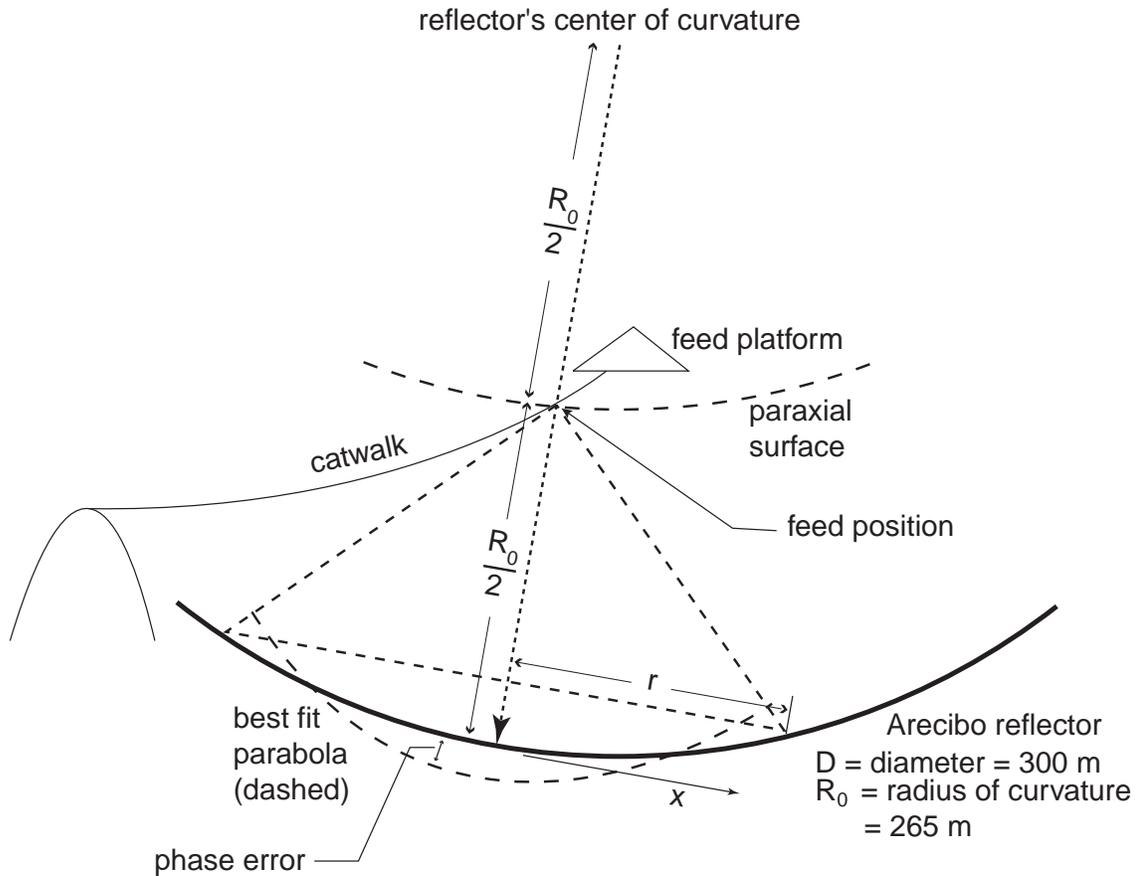
Figure 2.3: Feed positioning geometry for RMS phase error calculation.

For any size aperture we find, by least-squares, the best fit paraboloid and hence its
deviation from the sphere. From this the RMS phase error as a function of the size
of the illuminated sector is computed. The aperture efficiency can then be estimated,
from which is derived the effective area. Maximizing the effective area is equivalent
to maximizing the gain, and also the sensitivity. Note that the calculation of the
RMS phase error involves a numerical integration, so that the analytic calculation

does include a small non-analytic component.

The spherical cap and the best fit paraboloid are defined by their cross-sections, a circular segment and parabola respectively. Let the radius of curvature of the sphere be R_0 and the chordal radius of the illuminated spherical cap be r. R_0 is a constant, for Arecibo it is 265 m. r varies depending on the amount of the reflector illuminated, with its maximum value $r_{\max} = 150$ m.

For $-r \leq x \leq r$, we write the equation of the circular segment,

$$f(x) = R_0 - \sqrt{R_0^2 - x^2} \tag{2.1}$$

and of the best fit parabola,

$$g(x) = Ax^2 + B \tag{2.2}$$

The best choice of the parameters A and B is that which minimizes the area-weighted mean square deviation of $g(x)$ from $f(x)$ with respect to A and B. So we write:

$$\frac{\partial}{\partial A} \left(\int_0^r [f(x) - g(x)]^2 2\pi x \, dx \right) = 0 \tag{2.3}$$

$$\frac{\partial}{\partial B} \left(\int_0^r [f(x) - g(x)]^2 2\pi x \, dx \right) = 0 \tag{2.4}$$

Rearranging, and applying the chain rule for partial derivatives,

$$-4\pi \left(\int_0^r [f(x) - g(x)] \frac{\partial}{\partial A} [g(x)] x \, dx \right) = 0 \tag{2.5}$$

$$-4\pi \left(\int_0^r [f(x) - g(x)] \frac{\partial}{\partial B} [g(x)] x \, dx \right) = 0 \tag{2.6}$$

Noting that $\frac{\partial}{\partial A} g(x) = x^2$ and $\frac{\partial}{\partial B} g(x) = 1$ we write,

$$\int_0^r [f(x) - g(x)]x^3 \, dx = 0 \qquad (2.7)$$

$$\int_0^r [f(x) - g(x)]x \, dx = 0 \qquad (2.8)$$

Substituting for $f(x)$ and $g(x)$ and splitting the integrals,

$$(R_0 - B)\int_0^r x^3 dx - \int_0^r x^3\sqrt{R_0^2 - x^2} \, dx - A\int_0^r x^5 dx = 0 \qquad (2.9)$$

$$(R_0 - B)\int_0^r x \, dx - \int_0^r x\sqrt{R_0^2 - x^2} \, dx - A\int_0^r x^3 dx = 0 \qquad (2.10)$$

Integrating, substituting limits, and rearranging,

$$(\frac{r^6}{6})A + (\frac{r^4}{4})B = \frac{R_0^2(R_0^2 - r^2)^{\frac{3}{2}} - R_0^5}{3} + \frac{R_0 r^4}{4} - \frac{(R_0^2 - r^2)^{\frac{5}{2}} - R_0^5}{5} \qquad (2.11)$$

$$(\frac{r^4}{4})A + (\frac{r^2}{2})B = \frac{R_0 r^2}{2} + \frac{(R_0^2 - r^2)^{\frac{3}{2}} - R_0^3}{3} \qquad (2.12)$$

Setting $\alpha = r/R_0$ and solving yields the result,

$$A = \frac{4}{5}\frac{1}{R_0}\left[\frac{5\alpha^2 - 4 + (1 - \alpha^2)^{\frac{3}{2}}(4 + \alpha^2)}{\alpha^6}\right] \qquad (2.13)$$

$$B = \frac{-R_0}{15\alpha^4}\left[4(1 - \alpha^2)^{\frac{3}{2}}(6 - \alpha^2) - (15\alpha^4 - 40\alpha^2 + 24)\right] \qquad (2.14)$$

Now we have expressions for A and B in terms of r. At any value of x, the absolute error is

$$\epsilon = |f(x) - g(x)| \qquad (2.15)$$

and the mean square deviation of the sphere from the best-fit paraboloid is

28

$$\epsilon_{\text{rms}}^2 = \frac{\int_0^r [f(x) - g(x)]^2 2\pi x \, dx}{\int_0^r 2\pi x \, dx} \tag{2.16}$$

Substituting the expressions for $f(x)$ and $g(x)$ and taking the square root of the integral above yields an expression for the root mean square (RMS) deviation.

$$\epsilon_{\text{rms}} = \sqrt{\frac{2}{r^2} \int_0^r [R_0 - \sqrt{R_0^2 - x^2} - Ax^2 - B]^2 x \, dx} \tag{2.17}$$

The RMS error is often written in units of wavelength, λ. There are rules-of-thumb which are sometimes used to determine if the errors have grown too large to be tolerable; for example $\epsilon_{\text{rms}} \leq \lambda/16$ is often used. In this treatment we use a more quantitative approach. The Ruze equation [48] gives the expected fractional reduction in gain corresponding to a given RMS error.

$$\eta_{\text{surf}} = \frac{G}{G_0} = \exp\left(-\left[\frac{4\pi\epsilon_{\text{rms}}}{\lambda}\right]^2\right) \tag{2.18}$$

This equation is derived for, and is strictly only applicable to, a random Gaussian distribution of phase error across the telescope aperture. However we intuitively expect it to give a reasonable estimate of the gain loss attributed to the rather smoothly varying, and correlated, phase errors due to the differences between sphere and paraboloid. The numerical work in §2.1.2 ties up nicely with the Ruze estimates, giving confidence that this approach is reasonable. The integral in equation 2.17 was computed numerically using the 'integrate' function on an HP-32S pocket calculator. Table 2.1.1 gives the RMS deviations for several values of r, and the associated surface efficiency computed with the Ruze formula. The last column is the effective area, A_e, computed by multiplying the physical illuminated area by the Ruze efficiency.

From table 2.1.1 we see that the illuminated diameter corresponding to the largest effective area, or gain, is about 200 m. At this illumination the phase errors result in about a 25% gain loss relative to a perfect paraboloid, while the effective area is about 23000 m^2. We can compute the expected sensitivity from the effective area as $\Gamma = A_e/2k$, where k is Boltzmann's constant. This gives an estimate for the sensitivity of a 200 m diameter aperture of 8.4 K/Jy.

r	A	B	$g(r)$	$f(r)$	$\epsilon(r)$	$(\epsilon_{\mathrm{rms}})$		η_{surf}	A_e
(m)	(m^{-1})	(mm)	(m)	(m)	(mm)	(mm)	(λ)		10^3m^2
50	0.001904	-7.15	4.75	4.76	7.0	3.2	0.003	0.999	7.9
60	0.001912	-14.97	6.87	6.88	15.2	6.8	0.005	0.996	11.3
70	0.001921	-28.06	9.38	9.41	28.7	12.7	0.010	0.984	15.1
80	0.001932	-48.51	12.31	12.36	49.9	22.0	0.017	0.955	19.2
90	0.001944	-78.93	15.67	15.75	81.9	35.9	0.028	0.884	22.5
100	0.001959	-122.46	19.50	19.60	128.0	56.0	0.044	0.737	23.2
110	0.001975	-182.90	23.70	23.90	194.0	84.1	0.066	0.503	19.1
120	0.001994	-264.90	28.40	28.70	284.0	122.6	0.096	0.233	10.5
130	0.002015	-374.20	33.70	34.10	406.0	174.2	0.136	0.054	2.9
140	0.002038	-517.60	39.40	40.00	571.0	242.9	0.190	0.003	0.19
150	0.002064	-703.50	45.70	46.50	790.0	333.0	0.261	0.000	0.00

Table 2.1: Best fit paraboloid parameters, RMS deviations and effective area as a function of illuminated physical aperture.

This estimate is optimistic for a number of reasons. Firstly, phase errors due to the difference between the sphere and best fit paraboloid are not the only sources of surface error. Secondly, there are many other sources of inefficiency, for example missing primary reflector panels, aperture blockage due to the feed platform, ohmic losses in the reflector and feed, and spillover due to sidelobes in the feed[49]. Also, we have implicitly assumed uniform illumination of the aperture, which gives the best possible *aperture* efficiency, but poor *beam* efficiency. Beam efficiency is the ratio of power in the main lobe of the sky beam to the total power radiated. Maximum beam efficiency is achieved by tapering, or apodizing, the aperture illumination. This is desirable, and more practical, but comes at the expense of aperture efficiency. In fact it can be argued that in this experiment, where we are searching for a very weak signal, we should be optimizing aperture efficiency at the expense of beam efficiency. In practice apodization is difficult to avoid with any practical aperture illumination scheme, since all practical feeds have a built in taper.

It turns out that the biggest impact on achievable sensitivity is due to more mundane

constraints. The feed must be hung from the catwalk where it crosses the focal surface. This point is offset from the symmetry axis of the reflector[2], and the illumination spills over the edge of the dish, or vignettes, before a 200 m cap has been illuminated. Spillover increases the noise in the system because the feed illuminates 300 K warm ground. For on-axis illumination, we are limited by vignetting to a diameter of about 160 m, corresponding to an effective aperture of 19200 m^2, or a sensitivity of about 6.9 K/Jy. Note that this number is still optimistic because it still does not account for other sources of inefficiency discussed above.

The foregoing discussion assumes on-axis illumination. It is tempting to consider squinting the feed inboard, towards the center of the dish. This trades the cost of increasing the phase errors (because the illumination is no longer radial) with the greater physical area which can be illuminated without spillover. In order to properly optimize the optics allowing for squinting it is not sufficient to simply maximize the effective area, or gain. What is needed here is a calculation which maximizes signal-to-noise ratio (SNR) as a function of squint angle and beamwidth. Such a numerical calculation is outlined in §2.1.2.

2.1.2   Accounting for spillover, and allowing squint

Ian Avruch has modeled the optics numerically, allowing for the effect of ground noise due to feed spillover. In this treatment the feed is allowed to squint—the boresight illumination is allowed to stray inboard of the radial line from feed position to dish—thereby reducing ground spillover at the expense of increasing phase errors. In order to account for spillover, it is necessary to include the feed beampattern in the calculation. This implicitly allows for non-uniform illumination of the reflector. A detailed description of the model, and the numerical integrals used to compute quantities such as gain, efficiency and sensitivity can be found in Avruch's thesis [50]. I will give a brief outline here of the methods and formulae used, before presenting

---

[2]Actually, being close to the focal surface would be adequate, and the catwalk is nearly horizontal in the neighborhood of the feed. However the feed must also be clear of the path of rotation of the azimuth arm. There are discrete structural members, spaced at 20 ft. intervals, beneath the catwalk, and the feed is attached to the first of these which is clear of the path of the azimuth arm.

results.

The gain function is computed by numerically integrating, with appropriate phase factors, the field on the aperture at a point in the far-field which lies on the normal to the plane of the aperture. This gives the power flux at that point, which is then normalized by the power flux that would have existed there if the the antenna had radiated isotropically. This integral, with an appropriate phase error term, can be written as,

$$G(\hat{\mathbf{n}}) = \frac{4\pi}{\lambda^2} \frac{\left| \int_S f(\mathbf{x}') e^{-i\mathbf{k}\cdot\mathbf{x}'} e^{i\delta(\mathbf{x}')} dS' \right|^2}{\int_S f^2(\mathbf{x}') dS'} \qquad (2.19)$$

where S is the aperture surface with unit normal $\hat{\mathbf{n}}$, the $\mathbf{x}'$ are all possible space vectors contained in S, $f(\mathbf{x}')$ is the field strength in the aperture, $\delta(\mathbf{x}')$ is the phase error due to the spherical approximation at $\mathbf{x}'$, $\mathbf{x}$ is the point of observation, $\mathbf{R} = \mathbf{x} - \mathbf{x}'$, and $\mathbf{k} = (2\pi/\lambda)\hat{\mathbf{R}}$.

Before using the computer model which applies equation 2.19 to extend the results of §2.1.1, it was used to confirm them. Figure 2.4 shows the results of the computer model when used to determine the optimum illuminated aperture with on-axis uniform illumination, and no vignetting. The contours represent sensitivity as the illuminated radius and focal heights are allowed to vary. The dotted line shows the optimum feed placement, with sensitivity as determined by the analytic methods of §2.1.1. The dotted line passes close by the peak of the sensitivity surface. The maximum sensitivity is 8.8 K/Jy, very similar to the number found in §2.1.1. The good agreement in absolute sensitivity, and in optimum placement, gives some confidence in both the analytical and numerical treatments.

Now the analysis is extended to allow for non-uniform illumination, taking account of the feed power pattern p defined over 4π steradians. In addition to adding ground noise, feed spillover is also a source of aperture inefficiency. It reduces the gain G computed using equation 2.19 to G' given by,

$$G' = G \left(\frac{\int_{\text{dish}} p\, d\Omega}{\int_{4\pi} p\, d\Omega} \right) \qquad (2.20)$$
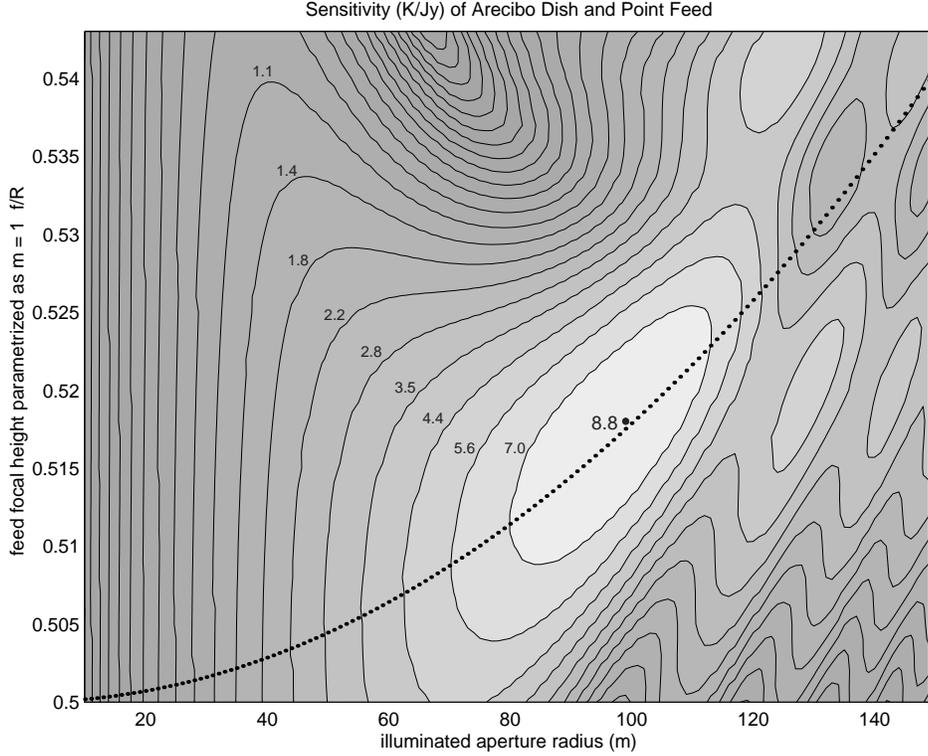
Figure 2.4: Sensitivity of the Arecibo dish, uniformly illuminated, for a variety of illuminated radii and focal heights. The focal height is expressed as $m = 1 - f/R$, where f is the height of the feed above the dish, and $R = R_0 = 265$ m is the radius of curvature of the Arecibo sphere. The peak of the contours is 8.8 K/Jy, in good agreement with the 8.4 K/Jy obtained using the analytic method. The dotted line is the locus of optimum feed placements, with sensitivity calculated using the methods of §2.1.1. [50].

Squinting reduces the spillover noise contribution, but at the expense of an additional phase error term. The additional phase error is accounted for by suitable choice of $\delta(\mathbf{x}')$. However the extent to which the spillover noise contribution degrades overall performance needs to be considered relative to the noise which is *already* in the system. The sensitivity, Γ, is not a suitable figure of merit in this case, since it makes no reference to system noise.

The quantity which is minimized in this treatment is the system equivalent flux density (SEFD). This is the additional flux incident on the aperture which doubles

the system temperature. If S_ν is the incoming flux, and SNR is the signal-to-noise ratio out of the receiver, SEFD= S_ν/SNR. SEFD is useful as a figure of merit which weights the spillover noise appropriately relative to gain degradation. If the sensitivity of the receiver is Γ and the system temperature of the receiver is T_{sys} then

$$\text{SEFD} = \frac{T_{\text{sys}}}{\Gamma} \tag{2.21}$$

The system temperature (T_{sys}) is the sum of contributions from the sky background and the ground (T_A), noise from the front end receiver amplifiers (T_R) and noise due to lossy elements which precede the receiver (T_L). For the purposes of the optimization, the increase in temperature due to a the celestial signal, T_{sig} is included in T_{sys}, although it is negligible compared to the other terms. So the system temperature is

$$T_{\text{sys}} = T_{\text{sig}} + T_A + T_R + T_L \tag{2.22}$$

T_{sig} is related to the incoming flux through the sensitivity, Γ,

$$T_{\text{sig}} = S_\nu \Gamma \tag{2.23}$$

which in turn is related to the effective area A_e or to the gain G as follows,

$$\Gamma = \frac{A_e}{2k} = \frac{\lambda^2}{4\pi} G \tag{2.24}$$

where k is Boltzmann's constant (in unusual units, rather useful for radio astronomy, $k = 1.38 \times 10^3$ Jy m^2/K)

The antenna temperature, T_A is computed by integrating the contributions from sky and ground over all solid angles, weighted by the beampattern of the feed p. The integral is divided into four domains, corresponding to the feed's view of the world. These are: the reflector itself, the ground screen which surrounds the reflector, the ground itself from the edge of the ground screen to the horizon, and the direct view of the sky, via the rear-ward pointing lobes of the feed beampattern. The reflector and

ground screen both reflect the temperature of the sky background, which is taken to be 90 K, while the ground is taken to be 300 K. So the antenna temperature integral becomes,

$$
T_A = \frac{1}{\int_{4\pi} p\, d\Omega} \left[\int_{\text{dish}} T_{\text{sky}}\, p\, d\Omega + \int_{\text{scrn}} T_{\text{sky}}\, p\, d\Omega + \int_{\text{gnd}} T_{\text{gnd}}\, p\, d\Omega + \int_{\text{sky}} T_{\text{sky}}\, p\, d\Omega \right]
\tag{2.25}
$$

The feed beampattern function p is estimated from anechoic chamber measurements of a 9 GHz model (see §2.2.1). An elevation beam-plot obtained from these measurements is rotated to form a solid of revolution. The signal-to-noise ratio is then SNR$= T_{\text{sig}}/T_{\text{sys}}$ The SEFD can then be computed as T_{sys}/Γ.
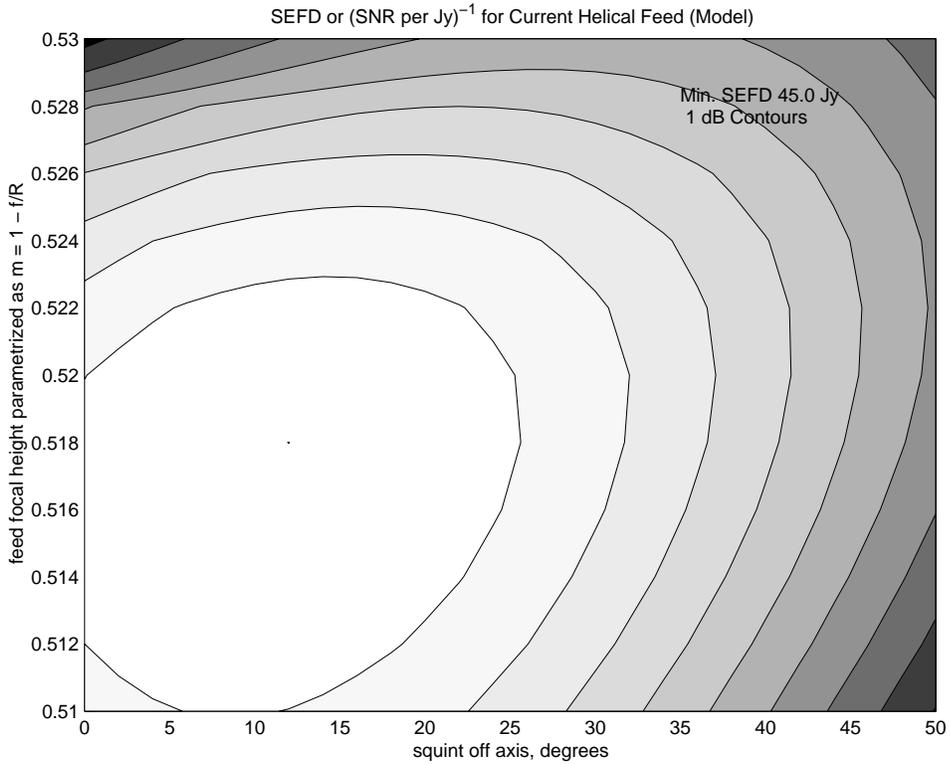
Figure 2.5: SEFD as a function of feed height and squint [50].

Using the machinery described in this section, the SEFD can now be numerically computed. The sensitivity, Γ, comes from equations 2.19, 2.20 and 2.24. The system temperature is estimated using equations 2.25 and 2.22, and these are then used to

derive the SEFD from equation 2.21. Numerical computation of the SEFD over a range of squint angles and focal heights produces figure 2.5 which shows contours of SEFD as a function of feed height and squint. The feed height is parameterized as $m = 1 - f/R$, where f is the height of the feed above the dish, and R is the radius of curvature. The optimal position of the feed, corresponding to minimum SEFD, can be read off this plot. The optimum squint is about $12°$ and the best value of the parameter $m = 0.518$, corresponding to a height above the dish of about 128 m, or about 4 m below the paraxial surface at $R_0/2$. The minimized value of the SEFD is 45.0 Jy.

The system temperature at the optimum SEFD is 170 K. Hence, at the optimal feed position the predicted sensitivity, allowing for vignetting in addition to phase error inefficiency, is 3.8 K/Jy. Although not all of the possible sources of aperture inefficiency have been considered here, the phase error and spillover loss are thought to be the biggest contributers. We use this number as our best estimate of predicted sensitivity.

Note that when this computation was performed, the best available estimates for the various temperatures were used, namely $T_R = 50$ K, $T_L = 10$ K, and $T_{\text{sky}} = 90$ K. Subsequent direct receiver measurements suggest a better estimate of $T_R = 110$ K. The rather gentle minimum in SEFD and the fact that $T_{\text{gnd}} = 300$ K dominates suggest that the feed position would be little changed if the revised estimates were used instead. However the estimated sensitivity of 3.8 K/Jy is purely a function of the feed position and beam pattern, and that we have placed the feed according to the guidelines derived in this section.

The effective aperture corresponding to a sensitivity of 3.8 K/Jy is $A_e = 2k\Gamma = 10488$ m^2. The gain is $G = 4\pi A_e/\lambda^2 = 78 \times 10^3$ or about 49 dB. The solid angle subtended by the beam is $\Omega = 4\pi/G = 1.61 \times 10^{-4}$ steradians. For small angles the beamwidth as a function of solid angle is $\phi = \sqrt{4\Omega/\pi} = 1.43 \times 10^{-2}$ rad or about $0.82°$.

2.1.3 Optimal optics in the presence of RFI

The analysis here assumes thermal noise limited conditions. At the time we worked out these optimization calculations as part of the planning for the experiment, and based on the RFI survey discussed in §1.6, we thought that the system would indeed be thermal noise limited—at least in those regions of the spectrum not corrupted with discrete carriers. As will be discussed in greater detail in later chapters, RFI pervades our system. In the reduced data, even the cleanest regions of the spectrum show RMS fluctuations at a much higher level than would be expected from thermal noise alone.

It is difficult to model how the RFI couples into the system, and it it therefore difficult to predict how that coupling might change with different feed positions and squint angles. It is likely that the RFI either couples directly into the feed, or perhaps scatters off the platform at a variety of angles, into the dish and then by reflection into the feed. It seems unlikely that squinting would reduce the level of the coupling in a predictable way. Based on these arguments it is suggested that the optimal feed positioning strategy in RFI limited conditions is to maximize sensitivity without regard to spillover noise. This is a different strategy to that which is discussed here, and which we actually used. But note that even in an RFI limited analysis, spillover, feed sidelobes, and illumination taper would affect sensitivity, and so squinting would most likely still be required. The gently sloping maximum in the sensitivity surface indicates that the gains from repositioning the feed would be small.

2.2 Feed antenna construction and evaluation

Note that the language used in this section, as is commonly the case when discussing antennas, assumes *reciprocity*. That is, the pattern into which an antenna radiates when used as a transmitter is also the sensitivity pattern for that same antenna used as a receiver. So we use terms like "sensitivity pattern" and "illumination pattern" interchangeably.

Figure 2.6: The dual helical feeds (before one was removed), suspended from the catwalk.

Helical wire antennas, designed and built by the Arecibo electronics department, form the feeds for this system. This antenna design was first investigated by Kraus [51]. The antenna used in this project is an axial mode helix—the radiation mode of helices whose winding circumference is of the order of a wavelength. In this mode the helix behaves as an elegant end-fire array with a single ended feedpoint. With appropriate parameters (for example, pitch angle) the current excitation in successive turns has just the right phase shift to interfere constructively in the axial direction. It is usual to mount the helix on a ground plane, to confine radiation to only one of the axial directions. A good recent review of helix antenna design is given in a chapter by King and Wong in the *Antenna Engineering Handbook* [52].

Our initial plan was to use dual feeds to form two simultaneous sky beams, separated East-West. In addition to providing twice as much time on-source per day, equivalent to a 3 dB improvement in signal to noise ratio, this arrangement also helps to distinguish RFI from celestial signals. The RFI, which couples directly into the feeds, is present in both beams simultaneously, while a celestial signal shows a characteristic

signature, appearing first in the east beam and then, after an appropriate sidereal delay, in the west beam. The dual feeds initially mounted on the catwalk suffered from interactions between the two elements which distorted their beamshapes. One of the units was removed because of this problem. It turned out that eliminating the interactions would have required an impractically large separation between the feeds [53]. The experimental data in this thesis has all been taken with a single helix.

Fortunately, in the single sky beam case a celestial signal still has a signature. Firstly, if it is a point source, it will trace the sky beam shape as it drifts through the beam. An extended source will trace a more complicated shape which can, in principle, be deconvolved into the source shape and the beam shape. (However our ability to do this successfully is limited if the signal to noise ratio is poor.) Secondly, since we survey the same declination strip each day, the "on source" pattern will repeat with a periodicity of a sidereal day. This is almost as strong a constraint as we had in the two beam case. However removing one of the feeds effectively halves the amount of time spent on source each day, thus doubling the time to detection.

The single unit which remains is very well matched to 50 Ω, as the Smith chart measurement in figure 2.7 shows. To achieve this match, the helix uses a matching transformer design due to John Kraus [54].

2.2.1   9 GHz scale models

To improve the performance of the system optics and reduce the susceptibility to RFI, we spent some time in an effort to improve the helical feed design. The main requirement of a candidate replacement feed is low sidelobes to reduce the susceptibility of the system to RFI. In particular, reducing the levels of the "horizon response", that is the sensitivity for all azimuth angles in the plane of the ground disk, is seen as very important—assuming that the RFI couples directly into the sidelobes from terrestrial or maritime based antennas[3]. The other requirement is beamwidth—to illuminate a physical diameter of 160 m we need a full width at half maximum (FWHM) of about

---

[3]Of course, if the RFI scatters off the feed platform, into the dish, and into the main lobe of the feed, there is little that can be done to reduce the antenna sensitivity.
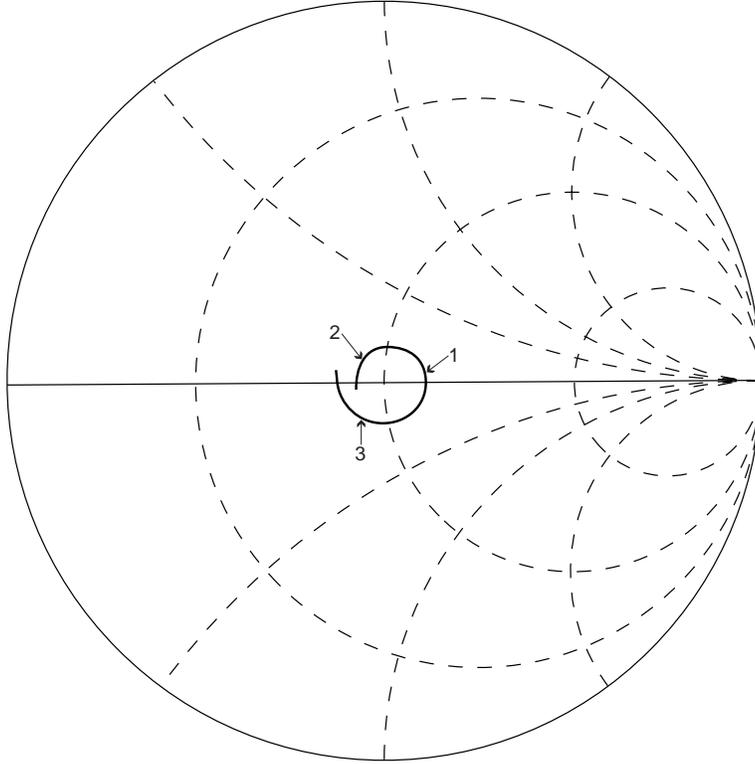
Figure 2.7: Smith chart plot of the input impedance of the helix antenna, measured in place on the catwalk, August 1997. The frequency and impedance at the three markers in order are: 235 MHz, $(63 + i)$ Ω; 219 MHz; $(45 + 6i)$ Ω; 251 MHz, $(45 - 9i)$ Ω.

$47°$.

We built models of candidate helical feeds scaled to 9 GHz ($\lambda = 3.3$ cm) and tested them in a small anechoic chamber at the Center for Astrophysics (CfA) in Cambridge. Our efforts yielded some improvements in both sidelobe levels and horizon response. However we decided that these improvements were too small to justify building a considerably more complicated feed antenna to replace the helix in place. In particular, the improved units have considerably larger and more complex ground planes, which raised concerns about wind loading. The results are nonetheless interesting, and lend some understanding to the performance of the optical system. I briefly summarize the most interesting results here.

We did make use of some of the measurements, those of the beampattern of a 9 GHz model of the full size helical feed which is in use at Arecibo. The measured beam

shape of this model is the best estimate we have for the beampattern of the full size unit, since we do not have access to facilities to make direct measurements. The measurement of the model's beamshape was used in the optics calculations of §2.1.2.

feed	gnd. dia.	plane type	no. turns	helix circ.	pitch angle	taper?
1	1λ	flat	5.5	0.9λ	$13.3°$	no
2	1λ	flat	6	1.1λ	$12.8°$	no
3	1λ	flat	7	1.1λ (mean)	$13.6°$	yes
4	2λ	flat	7	1.1λ (mean)	$13.6°$	yes
5	2λ	1-slot	7	1.1λ (mean)	$13.6°$	yes
6	2λ	2-slot	7	1.1λ (mean)	$13.6°$	yes
7	1.5λ	3-slot	7	1.1λ (mean)	$13.6°$	yes

Table 2.2: Characteristics of the 9 GHz feed models we built and tested.

Table 2.2 summarizes the characteristics of the models we tested. In the next three sections we present selected results. The first is the model of the feed currently in place at Arecibo (line 1 in the table), of a very similar model but with the uniform helix replaced by a tapered helix (3), and one of the best candidates, which has a tapered winding and a 1.5λ slotted groundplane (7). It should be said that this candidate had attractive beampattern characteristics, but, given its 1.5λ triple slotted groundplane, a full scale incarnation was expected to be very impractical and expensive to build and hang. In fact the 2λ groundplane units are even better, but they are completely out of the realm of full size realizability.

A scale model of the existing feed

Figure 2.8 is a photograph of a 9 GHz scale model of the helix antenna currently in place at Arecibo. We call this the "Arecibo model". Figure 2.9 is measurement of the elevation beampattern at two orthogonal azimuth angles. The worst case sidelobe blends into the main lobe, and is about 15 dB down from the peak in one of the azimuth plots. This feed has the characteristics given line 1 of table 2.2.
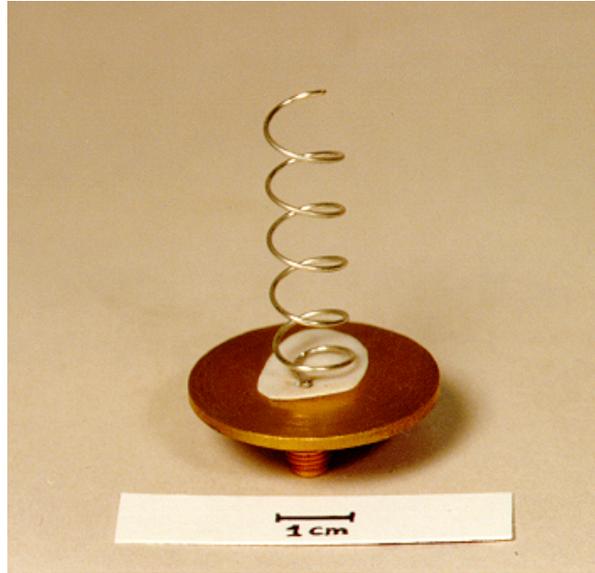
Figure 2.8: Photograph of the Arecibo antenna model.
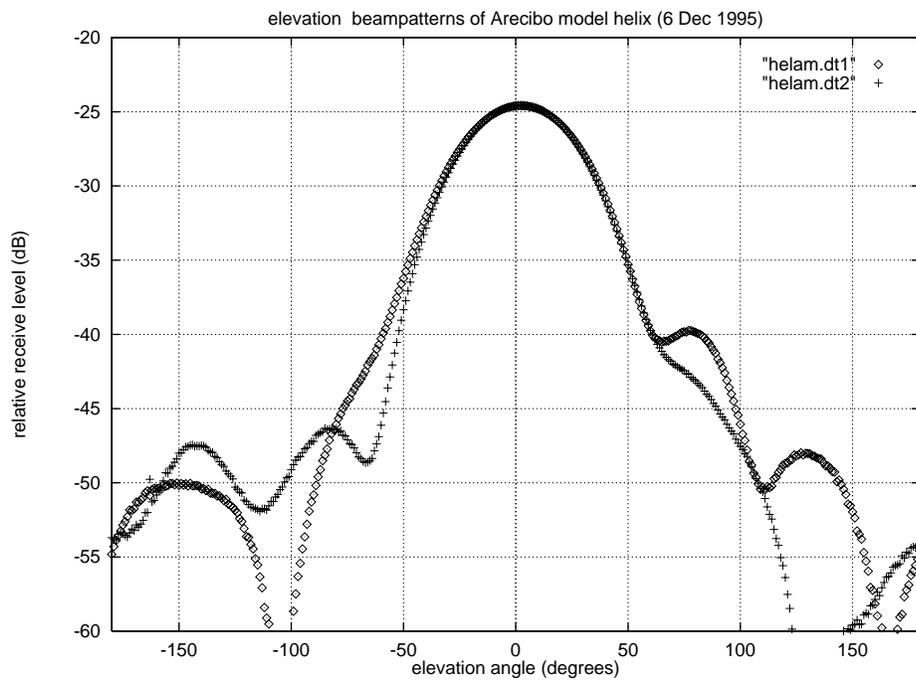
Figure 2.9: Elevation plot at two azimuths: Arecibo model antenna.

Tapered windings

The winding of a tapered helix has a greater circumference near the groundplane and gradually tapers as one moves away from the groundplane in the boresight direction.

42
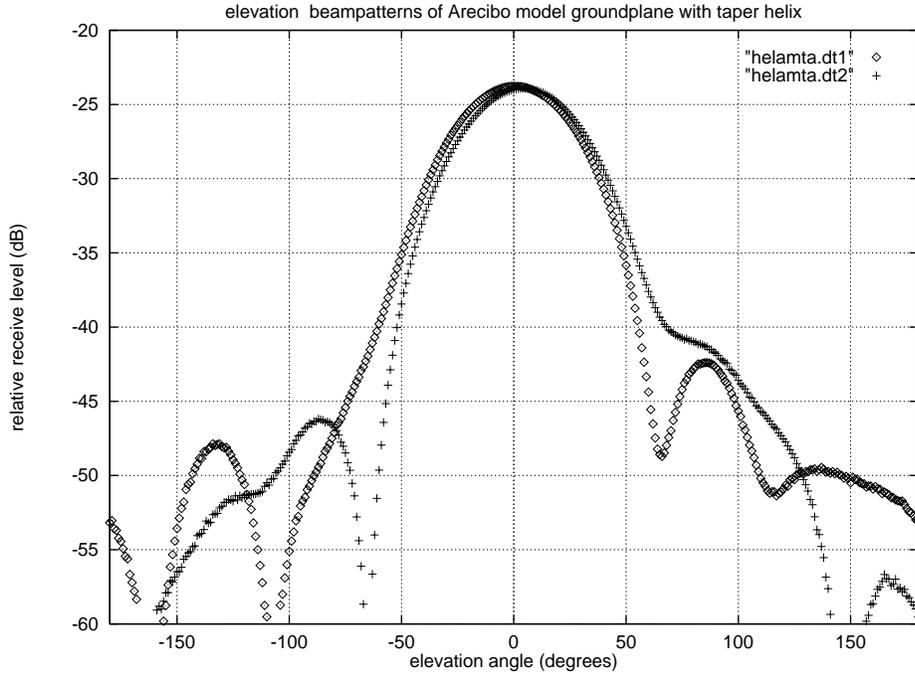
Figure 2.10: Elevation plot at two azimuths; tapered winding on a 1λ flat ground-plane.

The idea of taper is to reduce the impedance mismatch at the end of the helix, thereby reducing the amplitude of reflected current waves propagating in the winding towards the groundplane. Because the reflected current wave sees a winding of opposite helicity, the waves they set up have a circular polarization with the opposite handedness to that of the main wave. By reducing the amplitude of the reflected current wave, the polarization purity of the beam is improved.

Although it is not emphasized in the literature, we also noticed a small reduction in sidelobe levels in published data [52]. Figure 2.10 shows beampattern of a tapered winding mounted on the same size groundplane as the model of the Arecibo unit. Figure 2.10 shows about a 2 dB improvement in the peak to sidelobe ratio for the tapered unit over the uniform.

Figure 2.11: Photograph of a 9 GHz model with a triple slotted "soft" groundplane. Note that this photo shows a shorter helix than that corresponding to the measurements discussed here. The measured model had a seven turn tapered helix.

Soft surface candidates

To reduce horizon response, we tried the "soft surface" groundplanes suggested by Kildall [55] in combination with a helix wire antenna. Soft surface groundplanes have circumferential slots, $\lambda/4$ deep. The groundplane is, of course, a conductor, so the bottom of the slot is a short circuit. Considering the slot to be a parallel plate waveguide, the short at the bottom of the slot transforms to an open circuit at the top. Thus the effect of the slots is to suppress currents in the surface of the groundplane. This has the desirable effect of reducing horizon sensitivity, since the groundplane currents set up coplanar radiation fields.

We fabricated model soft groundplanes by machining them out of brass round bar stock. Figure 2.11 is a photograph of the tested unit. The beamplot in figure 2.12 shows that it has impressively low sidelobes, in worst case more than 20 dB down from the peak.
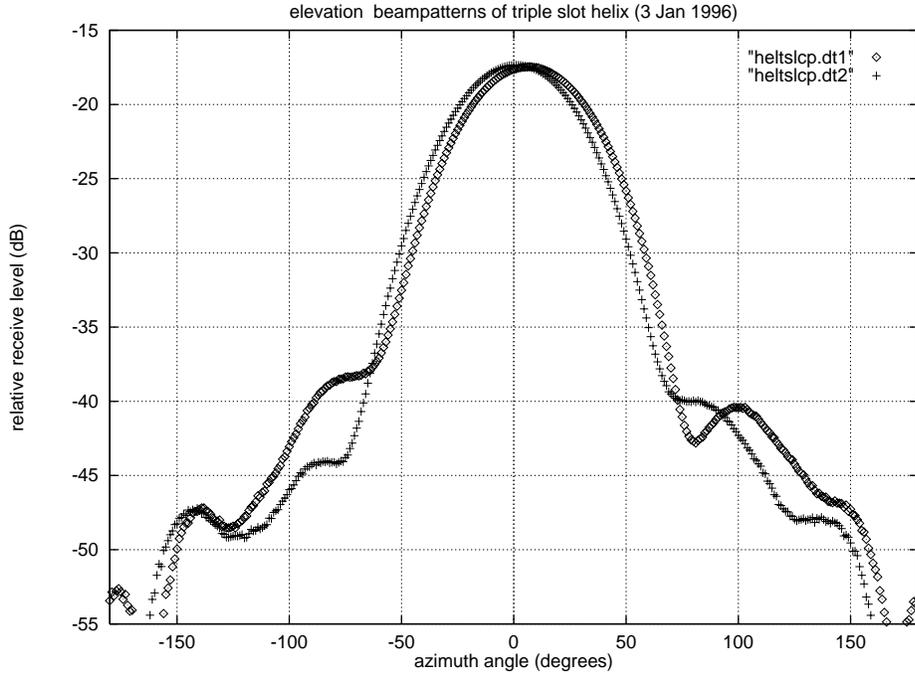
Figure 2.12: Elevation beampattern of a 9 GHz model with both a tapered helix, and a triple slotted "soft" groundplane

Dual polarization measurements

When we first saw the soft groundplane elevation plots, we thought we had achieved a significant improvement in the feed design. Further investigation, measuring beampatterns in both circular polarizations, showed that the improvement is not quite as large as these first measurements imply. The azimuth plots in two polarizations of figures 2.13 and 2.14 show that the soft groundplane appears to couple some of the sidelobe energy into the cross circular polarization. Both helices are left handed windings and are expected to radiate in left circular polarization (LCP). It is well known that for helices only the main lobe is expected to be purely polarized, so it is no surprise that both polarizations in the Arecibo model measurements are at approximately the same level.

But interestingly, the measurements for the slotted groundplane unit show up to about 10 dB of suppression, depending on azimuth angle in the LCP, but the cross-pole (RCP) pattern is about 3 dB *higher* on average than that for the Arecibo model.

45
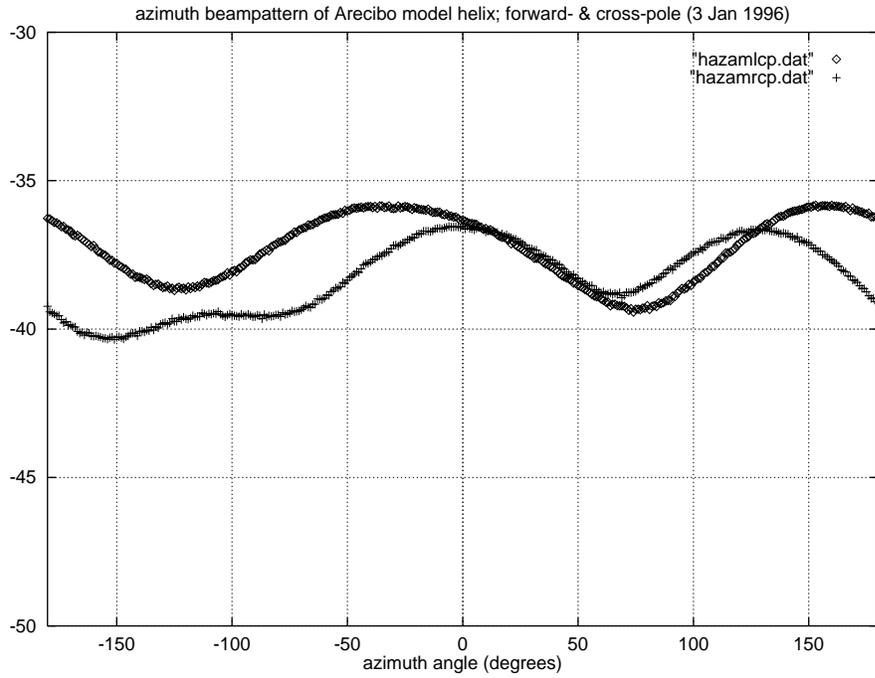
Figure 2.13: Azimuth beampattern of Arecibo model in both circular polarizations.
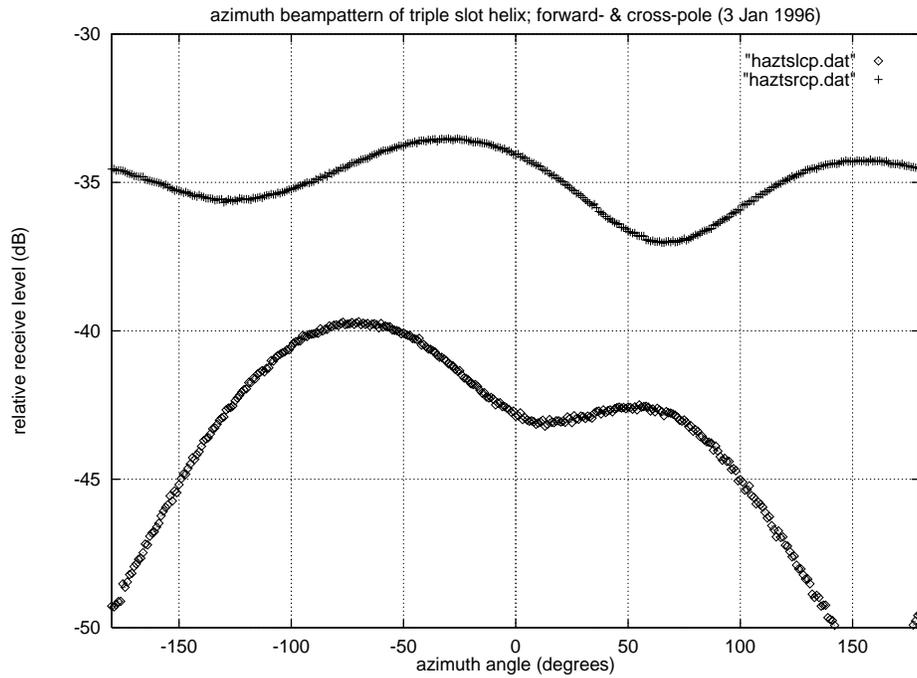
Figure 2.14: Azimuth beampattern of triple slotted helix in both circular polarizations.

Since RFI may appear in either polarization, the benefits that appear to accrue to the slotted groundplane design not be as compelling as they first appear. At this point we elected to continue working with the helices already in place at Arecibo.

2.3 Analog hardware

It is important in a radioastronomy receiving system to place a high-gain low noise stage very close to the front end. The noise figure of the very first gain stage dominates the overall system noise performance, with the effective contribution of subsequent stages reduced by the gain that precedes them. Conversely, any loss which precedes the first gain stage contributes directly to the system's overall noise figure. It is usual in radioastronomy receiving systems to cryogenically cool the front end amplifiers to optimize the noise figure. However, our project is un-supervised; our receivers are located on the catwalk, with very limited space, and far from the power points and other resources in the carriage house receiver rooms; and our budget is also quite limited. Based on these constraints we have had to be content with an uncooled front end amplifier. As we now know, the sensitivity of our experiment is limited by interference, rather than by the other sources of noise we have discussed—so, in retrospect, the compromise made by using uncooled amplifiers has made very little difference.

The analog amplifiers are divided into two boxes, one located upstairs on the catwalk close to the feed. The radio frequency (RF) signal is shipped downstairs, where it enters a second analog box which includes additional amplification stages and well as the first mixer. The first mixer multiplies the RF signal with a low-side local oscillator (LO) at a frequency of 176.0 MHz. The IF analog signal is routed to the IF channelizer where it passes through IF filters and is split 16-ways, before being piped to the mixer-digitizer boards. The signal is mixed to baseband in-phase (I) and quadrature (Q) channels on the mixer digitizer boards, and passes through a final anti-alias filtering stage before the analog signal chain ends at 8-bit analog to digital converters.

Because there is no need to tune the received frequency by adjusting the first LO, it is straightforward to arrange for image rejection. Sharp cutoff bandpass filters before and after both mixers ensure that image frequencies do not mix into the band of interest.

2.3.1   Upstairs receiver box
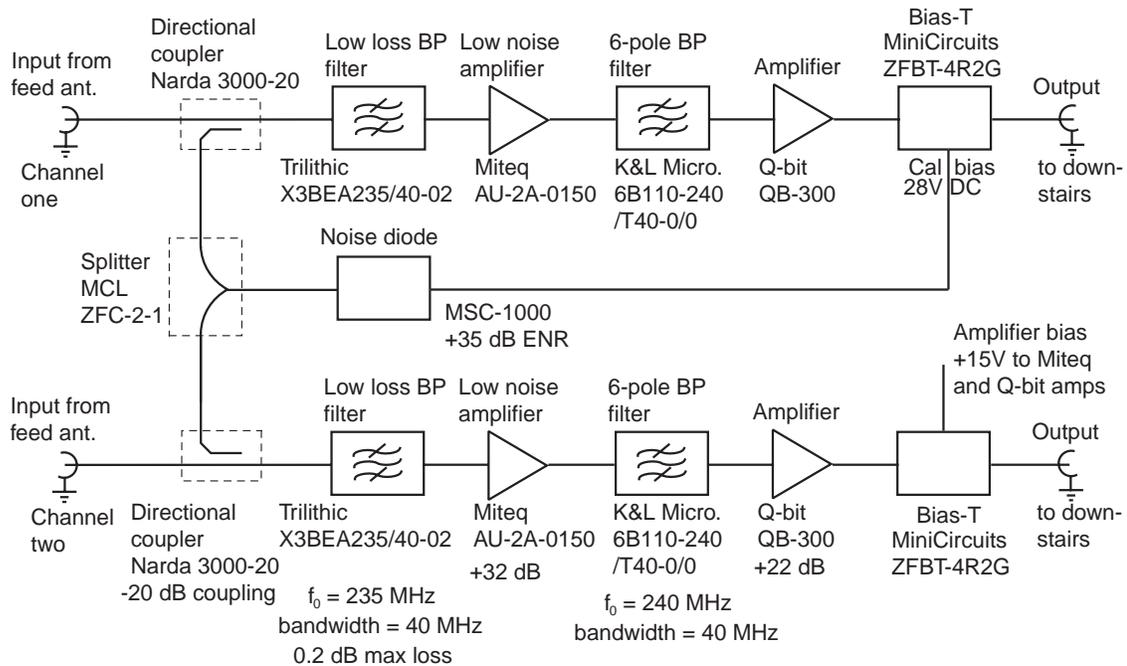
Figure 2.15: Block diagram of the upstairs receiver electronics.

Figure 2.15 is a block diagram of the upstairs receiver electronics. Ideally, the first stage of amplification should be placed directly on the feed antenna, to position the gain as close to the front end as possible so as to minimize loss and thus optimize noise performance. In practice, for serviceability reasons, it has been placed in a weatherproof box on the catwalk as close to the feeds as can be arranged. About 4 m of low loss Heliax[4] cable connects the antenna to the front end. The Heliax cable contributes a small loss of about 0.15 dB, which contributes directly to the noise figure of the system.

_____

[4]Heliax is a registered trademark, but is also generally used to describe very low loss coaxial cable constructed with a helical outer conductor, and an air, or low density foam, dielectric.

As is common practice in radio astronomy, we express the noise in our system as a noise temperature rather than a noise figure. To convert from noise figure, F, in dB, and noise temperature, T_n, in Kelvin, the following equation is used,

$$T_n = T(10^{F/10} - 1) \qquad (2.26)$$

where $T = 300$ K is the ambient temperature, typical for outdoors in Puerto Rico. This is the appropriate temperature to use when computing the noise contribution due to loss in the upstairs front end which located outdoors in the tropics. Note that when using this equation to transform the specified noise figure, F, of an amplifier to its noise temperature, conventionally one uses $T = 290$ K. Using this equation we see that the loss in the Heliax cable contributes about 10 K to the receiver temperature. Note that the contribution of the Heliax cable to the overall system temperature is not included in bench measurements of receiver temperature. A plot of the measured receiver temperature from one of the front end channels, that which was used to produce the data presented in chapter 3, is in figure 2.16(a).

It is necessary to minimize the number of other components placed ahead of the first amplifier, since their insertion loss also contributes to noise figure. After due consideration we did place two low loss components before the first gain stage. A -20 dB directional coupler is used to couple the output of a diode noise standard into the front end of the system. At periodic intervals, controlled by the host computer, the diode fires for a complete integration period, and adds a calibrated amount of noise to the receiver. The aim is to provide a means to calibrate out front end gain fluctuations.

The diode has a nominal excess noise ratio (ENR) of +35 dB. ENR is the power output relative to 290 K. The output of the noise diode is split two ways in a Mini Circuits ZFSC-2-1 power splitter, resulting in 3 dB of intrinsic loss, plus about 0.5 dB excess loss. The noise power then couples into the front end through a -20 dB coupler, so the effective contribution is 11.5 dB ENR or about 4100 K. Figure 2.16(b) shows the receiver temperature with the noise diode on. It is clear that the actual value of the noise diode's contribution is considerably less than the predicted 4100 K, and that
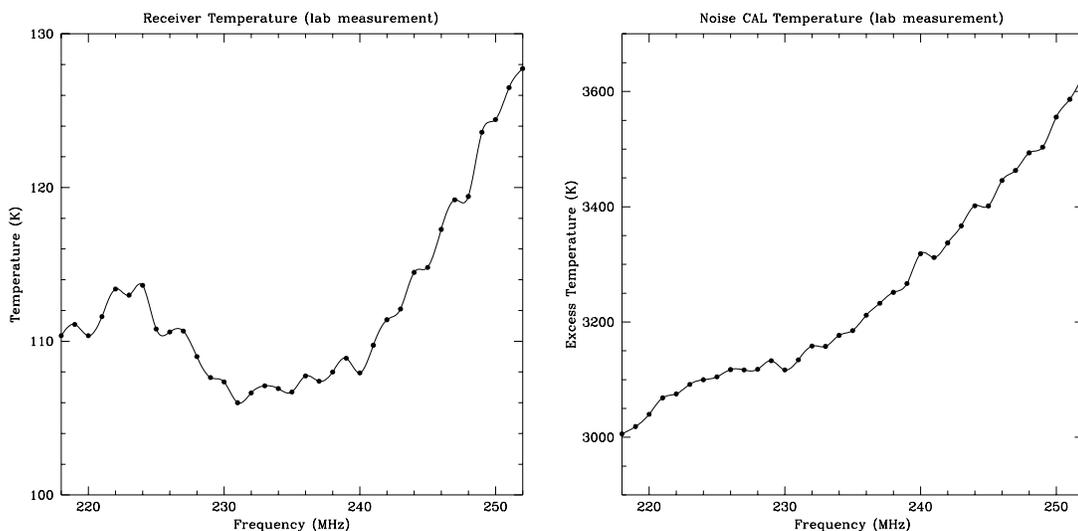
| (a) noise diode off | (b) T_n, noise diode on |

Figure 2.16: Receiver temperature measurements. These measurements were taken by the author on 1st August 1997, using a Hewlett Packard 8970B Noise Figure Meter. Data entry and plots are courtesy of I. Avruch

it varies somewhat with frequency. As discussed later, we noticed also that the diode sometimes fires weakly, and based on this observation coupled with this measurement, we had some doubts as to whether to use the diode data at all in our final analysis[5].

The directional coupler has about 0.1 dB of in-line loss, and also adds a small amount of thermal noise to the system—since when the calibration diode is not firing it still looks like a 50 Ω load at 300 K. This couples into the received signal at the −20 dB level, thus adding about 3 K to the noise temperature, while the 0.1 dB in line loss translates to about 7 K. A low loss 2-stage bandpass filter custom designed by Trilithic for this application, follows. It is necessary to have some filtering before the first amplifier, because in this RFI region the out-of-band RFI could cause non-linearity and hence inter-modulation products, if not excluded. This filter was custom designed to have a very low in-band insertion loss (< 0.2 dB). As a result it has fairly

---

[5]In the end, we reduced the data twice, the first time with diode based calibration, and the second time without.

gentle skirts since the low loss requirement dictates relatively few sections in the design. The filter is centered at 235 MHz and has a 40 MHz bandwidth.

The front end amplifier is a Miteq AU-2A-0150, a low noise bipolar amplifier with a gain of 32 dB. The specified maximum noise figure is 1.3 dB, equivalent to a noise temperature of 100 K. Bench measurements show it to behave a little better than this. After the first stage of amplification the front end is less sensitive to all sources of additional noise, and a six pole bandpass filter, with somewhat more loss is included. The filter is centered at 240 MHz and has a 40 MHz bandwidth (K&L Microwave 6B110-240/T40-0/0). It is followed by a second amplifier, the Q-BIT QB-300, which has a gain of 22 dB and a much more relaxed maximum noise figure of 3.8 dB or about 400 K.

The reduction in the noise contribution by system components after the first gain stage is substantial. To put some figures to this effect, we note that if G_n is the linear gain of a stage, and T_n its noise contribution, that the effective noise temperature T_{eff} is given by a sum with as many terms as there are stages.

$$T_{\text{eff}} = T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1 G_2} + \dots \tag{2.27}$$

In linear terms, the 32 dB gain of the first stage is about 1600, so the noise contribution of the second stage is reduced by this factor. The second filter has about 1 dB of insertion loss, equivalent to a noise contribution of about 80 K if it were in the front end. But because it is reduced by the gain of the Miteq amplifier, it contributes a mere 0.05 K. Likewise, the QB-300 amplifier's 400 K is reduced to an effective 0.25 K by the first stage gain. After the Q-Bit the overall gain is 54 dB, or about 250,000, so amplifier and loss noise contributions are entirely negligible from that point on

The last component in the upstairs receiver chain is a bias-T (Mini-Circuits ZFBT-4R2G). This component is used to decouple the DC power that is used to run the upstairs receiver from the cable linking bottomside to topside. In fact two cables link bottomside to topside. The first carries 15 V to run the amplifiers in the topside receiver box, and the second switched 28 V power to run the front end noise diode. The 28 V power is pulsed under computer control at periodic intervals, and fires the

noise diode.

2.3.2 Cables

There are two cables connecting the upstairs electronics box to the downstairs rack. They both run above the catwalk up to the feed platform. They have different diameters, and, hence cable losses. Cable 1 is $7/8''$ and is $1576'$ long with a measured total loss at 230 MHz of 19 dB. Cable 2 is a $6''$ coaxial waveguide, $1424'$ long, and with a loss at 230 MHz of 10 dB [56].

There are assorted joins and sections of other cable types in the cable chain. For example an RG-8 link is made between the platform and the receiver box approximately $60'$ down the catwalk. Likewise, at the bottomside end assorted RG-58 links are made from the ends of the cables to the downstairs electronics rack in the Arecibo correlator room. A possible side-effect of these splices are standing waves, which would result in a ripple in the baseline of the spectrum. The difference in cable loss between the two channels is compensated by including an extra 10 dB attenuator pad in line with the low loss cable in the downstairs receiver. The cable loss does not contribute significantly to the receiver temperature because of the significant gain which precedes it.

2.3.3 Downstairs receiver

The chain of amplifiers and filters continues in a box which is mounted in the downstairs electronics rack. We refer to this as the downstairs receiver, and figure 2.17 is a block diagram. The first component is a bias-T, which couples DC power onto the RF signal line for transmission to upstairs. Channel 2 is the channel with the 6 inch waveguide as the connection to the upstairs electronics box, and it has a -10 dB pad in line, to compensate for the lower loss of this cable and so balance out the signal levels in the two channels. A Q-Bit QB-300 amplifier provides 20 dB of gain, and is followed by another -10 dB pad and then another 6-pole bandpass filter. Pads are generally placed before filters to prevent possible amplifier instability due to out-
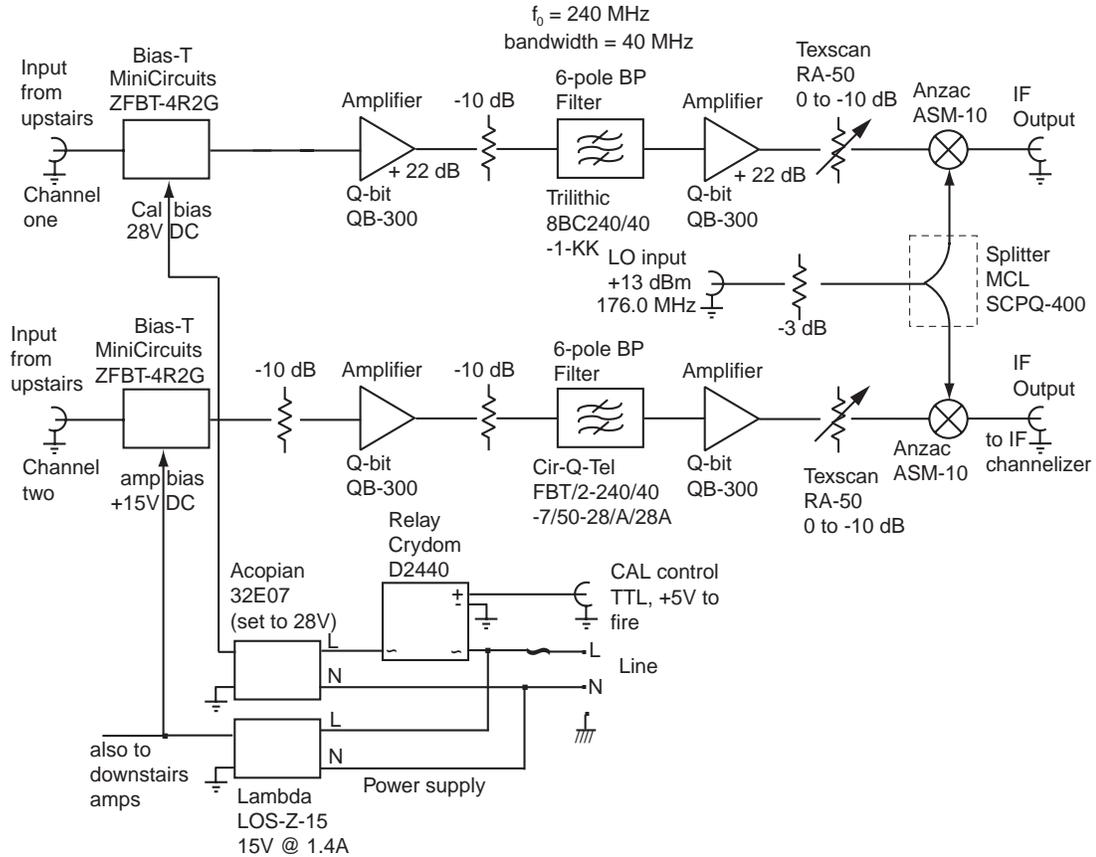
Figure 2.17: Block diagram of the downstairs receiver electronics.

of-band energy being reflected back into the amplifier output. Yet another another QB-300 follows the filter.

A Texscan RA-50 attenuator allows the signal level to be adjusted 0 to -10 dB in 1 dB steps, allowing the two channels to be balanced. The double-balanced mixer (Anzac ASM-10) is positioned at the tail end of the signal chain. The two channels' mixers are fed by the same LO, provided from an external input (see section 2.3.4), and a two-way splitter, the MCL SCPQ-400.

Power is provided to the amplifiers, and via the bias-T to the upstairs electronics from a Lambda LOS-Z-15 15 V, 1.4 A power supply. An Acopian 32E07 provides the pulsed 28 V for the noise diode. It is switched by a Crydom D2440 solid state relay which is driven by an external logic level provided in turn by the master control computer.

2.3.4 First local oscillator
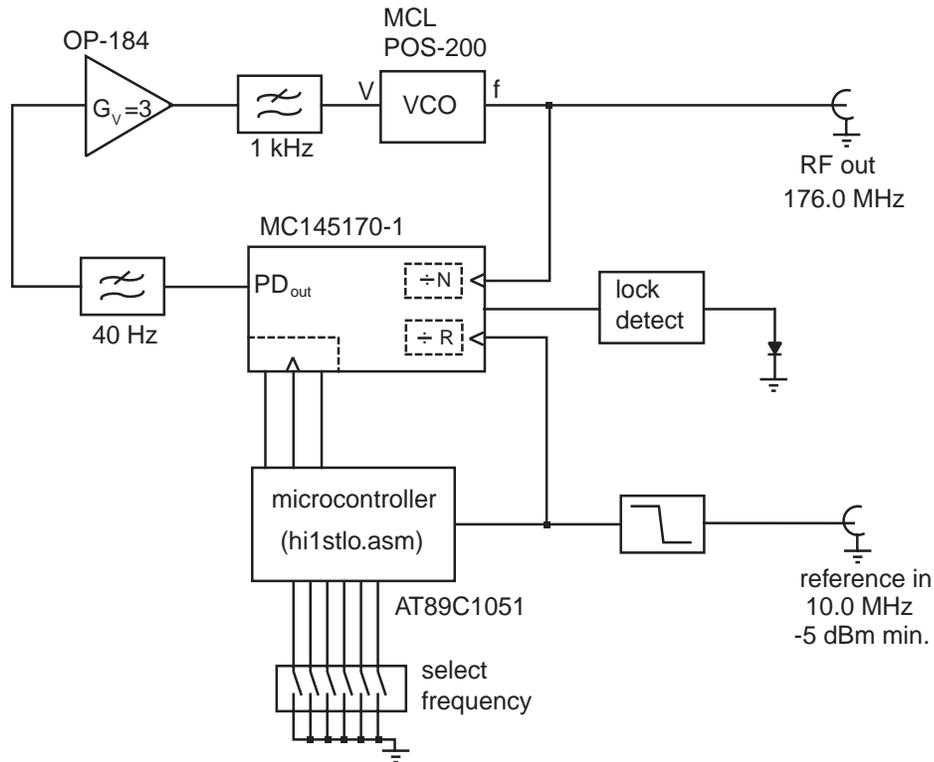
Figure 2.18: Block diagram of 176.0 MHz phase locked first local oscillator.

A custom 176.000 MHz phase locked loop (PLL) oscillator serves as the system's first LO. The design uses the MC145170-1, the -1 indicating that it is a faster grade of the Motorola PLL synthesizer chip used in the IF LO array, good to 185 MHz. The Mini-Circuits POS-200, a hybrid voltage controlled oscillator (VCO) in a printed circuit mountable metal can package, is the second crucial component in the circuit. The OP184 rail-to-rail low noise op amp from Analog Devices is used to boost the loop gain by a factor of three. A low noise amplifier is needed here, because it translates directly to phase noise which degrades the spectral purity. In addition to using an inherently low noise op-amp, we also limit its output bandwidth to 1 kHz to further reduce the total noise power.

The PLL controller chip is digitally programmable, and an AT89C1051 microcontroller is used to initialize it at startup or whenever the oscillator is reset. The values of the controller parameters are $R = 50$, which divides the 10 MHz reference input to

produce the 200 kHz compare frequency, and $N = 880$, which multiplies the compare frequency back up to give a phase locked 176.00 MHz output. The microcontroller also reads a 6-position DIP switch which allows the programmed frequency to be varied about the 176 MHz nominal center in 200 kHz steps, corresponding to unit changes in N. This flexibility is not used in our experiment, but was added in view of the plan to leave the spectrometer at Arecibo after this experiment is concluded.
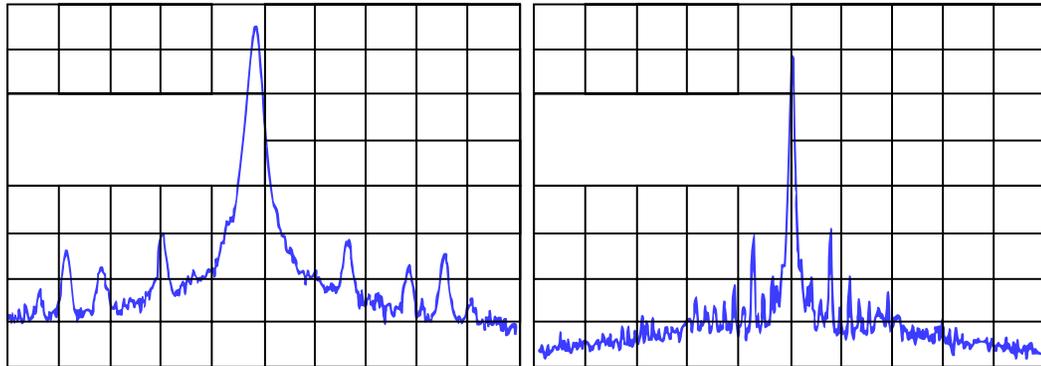
(a) span 1 kHz, resolution 10 Hz   (b) span 10 kHz, resolution 30 Hz
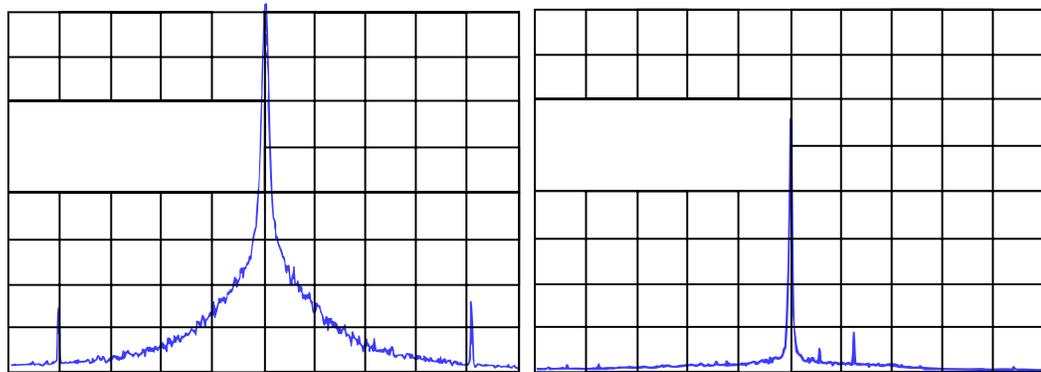
(c) span 500 kHz, resolution 1 kHz   (d) span 32 MHz, resolution 10 kHz

Figure 2.19: PLL first local oscillator output spectra. Vertical scale 10 dB/div.

As with any PLL circuit the choice of loop filter cutoff affects the spectral purity of the output signal. The initial value for the filter coefficient, chosen using a rule-of-thumb, [57], was modified to optimize the spectral characteristics of the output, while, of course, ensuring that the PLL locks up automatically on power-up. Figure

2.19 shows the output spectrum, at various resolutions, after the time constant had been optimized. It's quite clean, with the ratio of the peak power to the highest spur about 40 dB—and that spur, being within only 800 Hz is much less than the 10 kHz resolution of the spectrometer.

The highest spur and its close-in relatives are spaced at 120 Hz, and we suspect they result from power supply noise feeding through into the oscillator output. This intuition was confirmed when we noticed that the spurs were much reduced by carefully routing the power supply line side wiring. Another interesting spur is due to feed-through at the "compare" frequency of the PLL; it is separated from the fundamental by 200 kHz, well within an instrumental binwidth, and is about 65 dB down from the peak.

2.3.5   Intermediate frequency (IF) channelizer

The IF channelizer, mixer/digitizer and local oscillator (LO) array designs are taken, with minor tuning modifications, from designs by Paul Horowitz for the BETA SETI program. These designs are discussed, and performance data for the finished units is documented, in status reports written for that project [59]. A brief description of each design is given here.

The IF channelizer splits the signal into 16 channels for the mixer/digitizer boards. This is done in two stages, the first a two-way power split. Each of these outputs is furnished with a custom IF filter, with characteristics specified for this project. Both filters have a 17 MHz bandwidth but with differing center frequencies of 51 MHz and 67 MHz, referred to as the "low IF" and "high IF" bands respectively. Splitting the IF into two analog bands rejects out of band interference, and thus helps to preserve headroom, reducing the possibility of inter-modulation due to non-linearity in components downstream in the system. Given the forest of RFI in the system this precaution is clearly justified, and it could be argued that four or eight still narrower IF sub-bands would be an even better design. However the custom miniature IF filters used here, from Trilithic, are about 300 dollars each and budget concerns constrained us to two sub-bands.

After the filters the signal is split a further 9-ways in each of the low and high IF bands. The 9-way power splitter is actually a 12-way with 3 of the ports terminated in 50 Ω loads. This gives 16 IF outputs for the mixer-digitizer, and a "monitor" output for each IF band. To preserve the signal levels despite the power splitting, various stages of amplification are provided. We use printed circuit board mount amplifiers from Avantek (now a subsidiary company of Hewlett Packard), the GPD-404 and GPD-405, each providing about 10 dB of gain.

2.3.6 Mixer/digitizer boards

Each analog IF output from the channelizer is fed into one channel of a mixer/digitizer board. Each mixer digitizer is configured to support two IF channels[6], although each of these channels is fed the same LO. The LO array, discussed in §2.3.7, supplies one IF LO frequency to each mixer digitizer board. The mixer/digitizer board has discrete mixers which produce in-phase and quadrature outputs, low noise wideband current-feedback amplifiers and 7-pole low pass anti-alias filters.

The low pass filter bandwidths were modified from those used in the BETA project. Table 2.3 shows the filter components needed to tune the baseband filters for the three applications in which it has been used. Baseband bandwidth for the BETA project is 2.00 MHz, this work uses 2.25 MHz, and the AOFTM pulsar spectrometer (see §2.5) uses 2.50 MHz. An accurate filter characteristic requires close tolerance components. The resistors are 1% precision metal film components. Each capacitor value is made from a pair of mica dielectric 5% tolerance components in parallel; one makes up the bulk of the required value, is individually measured with a bridge, and one of a selection of small value (hence close absolute tolerance) components added as a "shim". The inductors are slug tunable and are trimmed out of circuit with a 1 MHz bridge.

A TRW TMC1175 fast flash analog to digital converter (ADC) then digitizes both

[6]The mixer/digitizer printed circuit board supports three channels, as required for the East, West and terrestrial channels of the BETA project. For the purposes of this experiment, components for only two of the channels are loaded.

component	2.00 MHz	2.25 MHz	2.50 MHz
R1, R2	374 Ω	422 Ω	475 Ω
C1, C4	532 pF	420 pF	340 pF
C2, C3	944 pF	746 pF	604 pF
L1, L3	91.2 μH	91.2 μH	91.2 μH
L2	101 μH	101 μH	101 μH

Table 2.3: Filter component values for the mixer/digitizer boards.

the I and Q baseband channels. The conversion clock for the ADC is generated by the spectrometer board, which is the next component in the signal chain (described in §2.4.1). The mixer/digitizer array can be thought of as a filter bank, which precedes the FFT. In addition to isolating strong interference to a single 2.5 MHz region at baseband, the 2.25 MHz cutoff provides the necessary anti-aliasing filter prior to baseband digitization.

2.3.7   Local oscillator array

The local oscillator (LO) array synthesizes a set of low-noise local oscillator signals of high spectral purity, phase locked to the 10 MHz Arecibo reference. Sixteen discrete output frequencies are spaced at 2.00 MHz intervals in the range 44 MHz to 74 MHz. They are used as the IF LO inputs to the mixer/digitizer bank. By spacing the LOs at 2 MHz intervals the 16 board mixer/digitizer bank covers a total band of 32 MHz, with the 2.25 MHz baseband channels overlapping by 0.125 MHz on each side.

The LO array is a bank of phase locked loop (PLL) synthesizers. They each use a discrete varactor tuned JFET oscillator, with a programmable IC, the MC145170 from Motorola, controlling each PLL. This IC is a lower speed grade version of the PLL controller used in the RF local oscillator circuit, discussed earlier (§2.3.4 (The LO array design was originally for the BETA SETI project [45], and significantly predates the first LO design. In fact, the LO array inspired the first LO design, and we were pleasantly surprised to find that, in the interim, Motorola had released a part rated

adequately to serve in the RF LO.) In addition to choosing the parameters R and N, each PLL has different component values in the JFET oscillator [57], a Colpitts circuit, which oscillates naturally at approximately the correct output frequency, and is locked in exactly by the PLL.

Additionally an output at a nominal 40 MHz is provided for the spectrometer board digital clock. It is routed to the spectrometer PC interface board, where it is squared up into a digital signal, converted to emitter coupled logic (ECL), and is used to provide a phase locked digital clock distributed over the ribbon cable bus to the FFT spectrometer boards. Following a suggestion by Dr. Michael Davis, we adjusted the controller parameters to set the frequency of this output to 39.384615 MHz. This ensures that *exactly* 208 bins fit into a 2.0 MHz sub-band, making it possible to exactly concatenate sub bands to produce the composite 32 MHz total bandwidth. The PLL controller parameters for this output are $R = 65$ giving (for this PLL only) a 153.846154 kHz "compare" frequency, and $N = 256$, which multiplies this compare frequency up to the required 39.384615 MHz.

2.4 Digital hardware

The spectrometer is an array of 16 FFT boards for each antenna feed, each FFT fed by a digitized eight-bit time series originating from one of the 2.25 MHz baseband channels of the mixer-digitizer array. Each quadrature analog signal is digitized at 8-bit integer precision then multiplied by a 256-point Hanning window. An FFT ASIC, the Austek A41102, performs 256-point complex FFTs at the rate of 10^4 spectra per second. The resolution binwidth is approximately 10 kHz. The complex spectra are converted to squared modulus, and many successive spectra are then integrated with accumulated spectra stored temporarily.

At programmable dump intervals (i.e., spectral integration times) the accumulated spectrum is FIFO-buffered to a 16-bit shared parallel ribbon cable bus, where it is gathered into the host PC. The computer assembles the pair of 4K-point accumulated spectra (256 points from each of 16 FFT boards in two channels), adds time stamps

and other status information, and stores them to its disk. Once a day, the data files are transferred to digital audio tape (DAT), and the tapes are shipped periodically to Cambridge. The spectrometer attains 40 MHz total bandwidth, as 4K channels of 10 kHz each, for each of two antenna feeds.

2.4.1 FFT based integrating spectrometer

The basics of continuous and discrete Fourier transforms are reviewed here for the convenience of the reader, and to establish notation. We then compare radio spectroscopy methods, and motivate our use of the FFT. Finally the detailed design of the FFT spectrometer board is discussed.

Fourier transforms in a nutshell

The continuous Fourier transform of a function $h(t)$ is defined by

$$H(f) = \int_{-\infty}^{\infty} h(t) \, e^{2\pi i f t} \, dt \qquad (2.28)$$

The inverse Fourier transform recovers $h(t)$ from $H(f)$

$$h(t) = \int_{-\infty}^{\infty} H(f) \, e^{-2\pi i f t} \, df \qquad (2.29)$$

The discrete Fourier transform (DFT) H_n of a time series $h_k = h(t_k)$, $k = 0, 1, 2...$ formed by sampling the function h at regular intervals t_k is defined by

$$H_n \equiv \sum_{k=0}^{N-1} h_k \, e^{2\pi i k n / N} \qquad (2.30)$$

and it can be shown that the inverse DFT

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} H_n \, e^{-2\pi i k n / N} \qquad (2.31)$$

exactly recovers the set of h_k's from the H_n's. The DFT is a useful approximation to the continuous Fourier transform. The h_k's are samples of $h(t)$, and assuming these samples are taken faster than the Nyquist rate, the H_n's are good estimates of samples of $H(f)$.

The fast Fourier transform (FFT) is an algorithm for exact fast computation of the DFT. Whereas equation 2.30 implies that the number of operations required is of order N^2 ($O(N^2)$), it turns out that the exact transform can be computed in $O(N \log_2 N)$ operations. The FFT algorithm became widely known in the mid-1960's due to Cooley and Tukey, but was discovered earlier by many researchers, including Gauss [58].

A formulation due to Danielson and Lanczos is one of the clearest. It shows that a Fourier transform of length N can be written as the sum of two Fourier transforms of length $N/2$, one containing the even subscripted components of the time series and the other the odd, with a phase factor multiplying one of the transforms. Writing the phase factor $P = e^{2\pi i/N}$ and using H_n^e and H_n^o to represent the even and odd components of H_n respectively, the Danielson-Lanczos Lemma (DLL) is written,

$$H_n = H_n^e + P^n H_n^o \tag{2.32}$$

The FFT can be thought of as recursive application of the Danielson-Lanczos Lemma to the H_n, which are the unknowns at the outset. If N is an integer power of two[7], the eventual outcome is the H_n in terms of DFTs of unit length. The unit length DFTs reduce to the identity operation $H_n = h_k$, and the H_n then drop out of the recursion. A slightly inconvenient byproduct of this algorithm is *bit-reversal*—the output H_n are produced in a permuted order $H_{n'}$ where the n' are given by the n, written in binary, and then read from right to left.

There are other equivalent ways of viewing the FFT, and the Austek chip uses one such alternative approach. Figure 2.20 is a block diagram showing the simplified internal workings of the Austek A41102 [60] configured for an 8-point transform. The top part of the diagram shows how the FFT algorithm is implemented as a series of

---

[7]We restrict our attention to this, by far the most usual, case. See reference [58] for the treatment of $N \neq 2^{\text{integer}}$

61
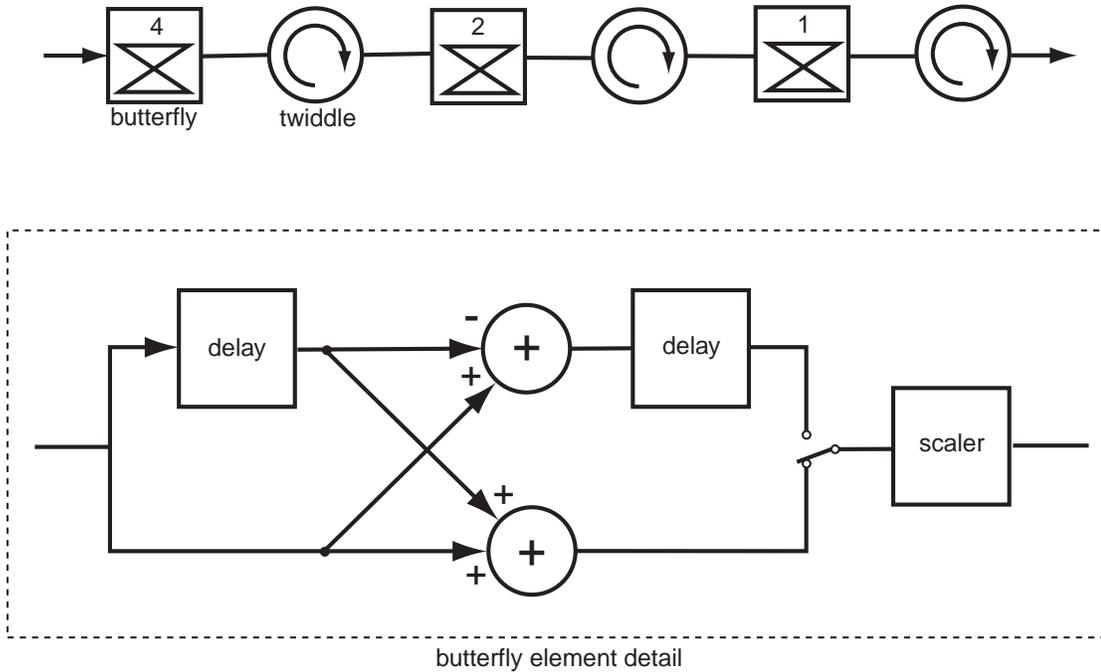
Figure 2.20: Simplified 8-point FFT operational flow [60].

"butterflies" and "twiddle" phase factor multiplies. This makes use of the fact that a long 1-dimensional FFT may be implemented by arranging the data in a rectangular array, doing a DFT on the columns, multiplying by a twiddle factor, and then doing a DFT on the rows. This process is pipelined to yield the DFT of the entire time series.

The twiddle factors are the phase factor multiplies associated with each stage. They transform the independent row and column transforms of the two dimensional array into a one dimensional transform of length equal the number of array elements. Also, because the twiddles are purely phase factors, and don't change the amplitude of the complex numbers they multiply, the multiplies can be simply implemented with delays and summations. The bottom half of figure 2.20 shows a butterfly element which takes advantage of this simple implementation. Note that the first delay in the upper branch of the butterfly, and the associated crossovers and adders, is the twiddle factor implementation. The second delay is not a twiddle, but a pipeline, which stores the second half of the vector until the output multiplexer is switched to the top branch to collect the second half of the vector.

Windowing and spectral resolution

In order to estimate the power spectrum by computing a discrete Fourier transform, we operate on the analog received signal in two distinct and rather radical ways. Firstly we divide time into "windows", about $\tau = 100$ μs long. For mathematical convenience we consider the range of the window as being $-\tau/2 \leq t \leq \tau/2$. Secondly, we sample the signal at discrete times, about once every $\Delta t = 400$ ns per complex valued sample. It is important to understand how these two operations affect the estimate of the power spectrum.

It is clear, by the uncertainty principle, that the resolution of the output spectrum is at best the inverse of the window length, τ, or about 10 kHz. One way to think about this is to consider the sampled time window to be formed from the signal defined for all time, multiplied by rectangular function which is unity during the window time, and zero elsewhere. In the frequency domain, this is equivalent to convolving the spectrum of the signal with the Fourier transform of the rectangle, which has the form of a sinc function.

$$W(f) = \tau \left[\frac{\sin(\pi f \tau)}{\pi f \tau} \right] \tag{2.33}$$

The sinc function has a main lobe, and oscillatory sidelobes, with zeros at $f = 1/\tau$ and its integral multiples. Each spectral component may be thought of as a delta function convolved with the sinc. For spectral components at the center of a frequency bin, corresponding to sinusoids whose period is an exact integral sub-multiple of τ, the zeros lie at the centers of off-signal bins, and all of the spectral energy is confined appropriately to a single bin. This is a very special case, however. In general signals are incommensurate, and the sidelobes contribute significant spectral energy to bins containing no signal. This is called *spectral leakage*.

It is usual to minimize spectral leakage by applying a multiplicative *window* function to the input time series before transforming to frequency. This taper function is analogous to the apodization discussed earlier in reference to the optics of the spherical reflector. Generally window functions taper the time samples to zero towards the edge

of the window. A good window function has a transform with much lower sidelobe energy than the sinc function, but at the expense of a broader main lobe. Thus spectral leakage is reduced. The reduction in spectral leakage comes at the expense of spectral resolution, since the taper inevitably broadens the main lobe. In addition, the taper at the edge of the time window reduces the total energy in the sample of the time function, and thus slightly reduces the sensitivity of the spectral measurement.

Window functions are discussed in detail in a review article by Harris [61]. A rectangular window (that is equivalent to *no* window) is clearly unacceptable; however the differences between the many candidate taper functions discussed in the literature are quite subtle [58]. We use a Hanning window[8], which is a raised cosine function given by,

$$w(t) = 0.5 + 0.5 \cos(\frac{2\pi t}{\tau}) \qquad\qquad \frac{-\tau}{2} < t \le \frac{\tau}{2} \qquad (2.34)$$

The Fourier transform of equation 2.34 is

$$W(f) = \tau \left[0.5 \frac{\sin(\pi \tau f)}{\pi \tau f} + 0.25 \left(\frac{\sin[\pi(\tau f - 1)]}{\pi(\tau f - 1)} + \frac{\sin[\pi(\tau f + 1)]}{\pi(\tau f + 1)} \right) \right] \qquad (2.35)$$

The transform of the Hanning function has first sidelobe levels about 32 dB below the mainlobe, as compared to 13 dB for the rectangular window, with rather rapid -18 dB per octave falloff beyond the first sidelobe. The latter characteristic is crucial in an environment with high levels of RFI, since leakage of RFI spikes is isolated to a small area around the interfering carrier.

Sampling rate and aliasing

The second operation is the sampling of the signal at discrete times $n\Delta t$ (n an integer). Here it is crucial to satisfy the Nyquist criterion, which, for complex valued samples may be stated as $\Delta f \le 1/\Delta t$. Our instrument has a 400 ns sample period, corre-

---

[8]In common usage it is called a Hanning window, but more correctly it's a Hann, or von Hann, window, after Julius von Hann [62].

sponding to a Nyquist frequency of 2.5 MHz. Prior to sampling, the mixer digitizer boards filter the baseband signal to a bandwidth of 2.25 MHz, allowing a 0.25 MHz "guard band" relative to Nyquist. The penalty for not adequately filtering prior to sampling the signal is aliasing—signals outside the FFT bandwidth wrap back into the sampled spectrum, causing errors in the estimates of the spectral components on the edges of the band.

Word growth

In the integer arithmetic of the FFT processor, each butterfly results in an increase in the number of bits needed to represent the signal. For a noise-like signal the growth of the largest point in the data stream is half a bit per butterfly. This follows from the definition of the DFT as given in equation 2.30. For a coherent signal the growth is one bit per butterfly. Incidentally, transforms of impulses don't grow at all. Intuitively we see that the growth for a coherent signal is greater than for noise because the signal power concentrates itself into half the number of bins as each butterfly is processed.

Each butterfly in the Austek chip is followed by a 1-bit scale, which may be optionally enabled or disabled. To exactly mimic equations 2.30 and 2.31, all scales should be off for forward transforms, and on for inverse transforms. The scales then provide the factor of $1/N$ preceding the summation in the inverse transform. The word growth as outlined above assumed all scales were off. If enabled, the scale shifts right 1-bit after the butterfly, an effective division by 2. Careful setting of the scales for a particular signal helps to ensure optimal use of the available dynamic range. Small signals should be allowed to grow in the early stages, with scales being turned on later in the processing. In particular it is crucial that the FFT processor does not overflow or underflow during processing, as this would result in a forest of spurious spectral components.

In the power accumulating spectrometer we compute power after the FFT. This doubles the number of bits in the integer representation of the output. In addition we integrate power for typically 10,000 spectra. Each doubling of the number of spectra integrated requires another bit in the data bus. While the output of the

65

Austek chip is 16-bits, we expand the bus to 32 bits wide to accommodate the power computation and integration.

Autocorrelation spectroscopy

Historically in radio astronomy, spectra of sampled data have typically been computed via the autocorrelation function. Expressed for continuous valued functions, the autocorrelation function is,

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(t)h(t-\tau)\, dt \qquad (2.36)$$

It turns out that the power spectrum and autocorrelation function are exactly a Fourier transform pair. This is the Wiener-Khinchin theorem [63],

$$S(f) = |H(f)|^2 = \int_{-\infty}^{\infty} R(\tau)e^{2\pi i f \tau}\, d\tau \qquad (2.37)$$

A rather amazing result, due to van Vleck and Middleton [64], is also very important in its application to autocorrelation spectroscopy. The autocorrelation function of Gaussian noise can be entirely reconstructed, to within a constant factor which is the total power of the signal, from a clipped representation of that noise. In other words, if one samples a noiselike signal to 1-bit precision and runs the time series through a 1-bit auto-correlator the result is ρ_c, a normalized and predictably distorted version of $R(\tau)$, given by

$$\rho_c = \frac{2}{\pi} \sin^{-1} \rho \qquad (2.38)$$

where $\rho = R(\tau)/\sigma^2$. This result is inverted and scaled by the variance σ^2 to get the true autocorrelation function from the output of a 1-bit correlator.

$$R(\tau) = \sigma^2 \sin(\frac{\pi}{2}\rho_c(\tau)) \qquad (2.39)$$

As a consequence of this result, hardware auto-correlators are typically 1-bit arithmetic devices, and are fairly straightforward to build. Sampling at one bit is equivalent to simply extracting the sign of the input data. Sampling is usually implemented with a fast comparator on the input [63]. Figure 2.21 is a simplified block diagram of an autocorrelation spectrometer. The multipliers, if limited to one-bit precision, can be implemented with simple logic.
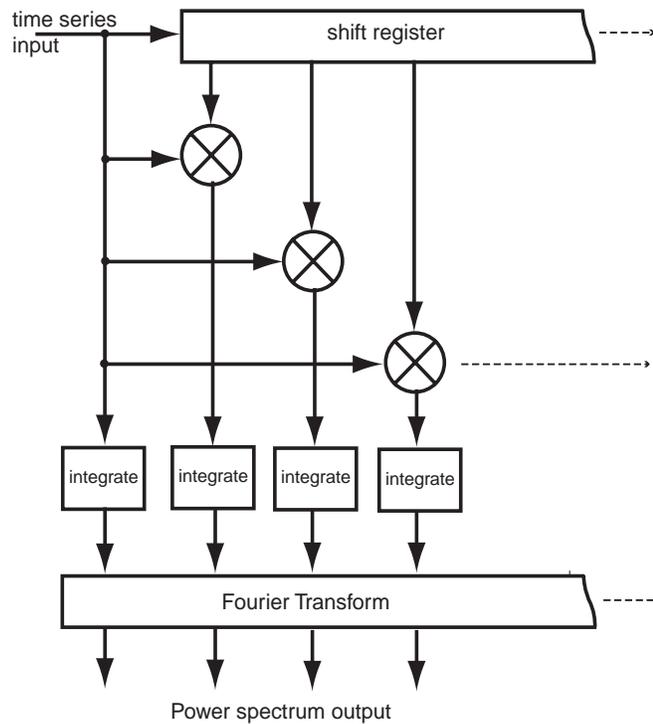
Figure 2.21: Simplified correlation spectrometer operational flow.

The autocorrelator is a continuously pipelined machine; that is, the data are not divided into short windows, as for an FFT. The stages in the delay line are called lags—the longest lag defines the effective window length, which we called τ in the direct DFT case. The correlator is typically run for the integration time of the measurement, with the correlated output at each lag averaged, or integrated. After the integration time is over, the output is corrected based on equation 2.39, and a single FFT, applied to the averaged autocorrelation function, produces the power spectrum. In practice the output is often downloaded to a computer which does both

the distortion correction, and the FFT.

A comparison of FFT and autocorrelation based spectrometers

The practical advantage of the correlator approach is that, since the voltages are already multiplied together to get units of power, integration can be implemented by adding successive products in a local accumulator at each lag. Thus the architecture is highly distributed, with no large interconnecting busses, so the metal interconnect on a correlator chip is much simpler than that of a comparably sized FFT device. The benefit of the FFT, on the other hand, is embodied in the number of multiplies being $O(N \log_2 N)$, while the number of multiplies in an autocorrelator is $O(N^2)$[9]. As demonstrated by van Vleck and Middleton, one-bit arithmetic is adequate in an autocorrelator for noiselike signals. In this case the correlator is simpler than an FFT since the multipliers are easily implemented.

A 1-bit correlator is a very efficient approach to detecting faint signals buried in Gaussian noise. It's important to realize, though, that equation 2.39 only applies to noiselike signals. The RFI environment in radio astronomy is deteriorating in all bands as the radio spectrum is used more and more heavily for commercial purposes, and in the 219–251 MHz band the situation is particularly bad. Accommodating coherent interfering carriers without saturating the processor places stringent requirements on dynamic range. In our system we use 8-bit sampling, and, allowing for word growth we use 16-bit arithmetic in the FFT. So the number of multiplies being $O(N \log_2 N)$ in an FFT becomes an extremely important benefit relative to the N^2 of the autocorrelator. An additional simplification occurs because multiplies in the FFT are with phase factors, and so can be implemented as delays and adds, as shown in figure 2.20. In a correlator, present samples are multiplied with lagged samples. These multiplies are not simply by phase factors so the multiplier has to be a more complicated general purpose implementation. Of course the output of the correlator is already in units of power, while after the FFT it is still necessary to compute power. This involves

---

[9]Of course it is possible to multiplex many data paths through fewer multipliers, but the greater the number of operations, the faster these elements need to be. Like real estate, speed is also expensive.

squaring the in-phase and quadrature components of each point in the spectrum, an additional $2N$ multiplies. Another $2N$ multiplies are required to apply the window taper function to the input time series. However this is a small number of operations compared to $N \log_2 N$ (for moderately large N), and even more so compared to N^2.

There are some more subtle tradeoffs in choosing between FFT and autocorrelation. A truncated data stream of N points has N samples of lag zero but only 1 sample of lag $N-1$. The autocorrelation function does not truncate the data stream, and so has equal numbers of samples of each lag. The natural tapering in the number of lags in a rectangular windowed FFT gives a sidelobe response in the power spectrum which is of form $(\sin(x)/x)^2$, while the autocorrelation function is of form $\sin(x)/x$, which has higher sidelobes. Another way to see this is to note that, in a correlator, the FFT is performed after the data has been squared into power. Thus the unwindowed autocorrelation is more susceptible to spectral leakage—as a result the ACF is often windowed before transforming. In particular, the power spectrum derived from the ACF of a narrowband signal which lasts longer than the longest lag generally shows a lot of spectral leakage.

NAIC has developed a new correlator chip which does 1024 multiplies, 1024 adds, and 1024 data shifts every 10 ns, or 3×10^{11} operations per second, on chip which dissipates less than 2 W. This chip uses 3-level, or ternary, arithmetic. This is also a simple arithmetic scheme, but with greater dynamic range than 1-bit arithmetic. Also, ternary arithmetic extends quite naturally to 9-level, 27-level etc., calculations since it involves no carries from one place-value to the next, so that no communication is needed between the chips. At the end of the integration the results are combined. For 9-level the sample is formed by taking $9 \times C_{11} + 3 \times (C_{10} + C_{01}) + C_{00}$, where the C_{nn} are the correlator chip accumulations. Extending the number of levels is, unfortunately, expensive—the number of chips required is one third the number of levels. Ternary arithmetic is discussed in Donald Knuth's classic "The Art of Computer Programming" [65].

If complex spectra are required, as for the pulsar spectrometer discussed in §2.5, the FFT fills the bill naturally. The data are simply taken prior to computing power. It is relatively simple to add such an operational mode to an FFT design, and we have

indeed done that in our hardware. Autocorrelators can be used on complex data as well, but not as elegantly. One needs the the autocorrelation and cross-correlation of both in-phase and quadrature channels. The complex FFT then gives the complex power spectrum. Thus the use of complex sampling requires twice as much hardware for the same bandwidth and resolution.

Another small advantage of complex FFT outputs is that they have already isolated a particular frequency bin. Thus a small selection of "interesting" bins can be chosen for further analysis and broken down into still higher resolution sub-spectra—this is used to advantage in the "Project Phoenix" SETI program [66].

We are grateful to Dr. Michael Davis of the Arecibo Observatory for numerous discussions and email correspondence on FFT autocorrelation spectroscopy, which contributed significantly to this discussion [66].

Hardware description

Figure 2.22 is a simplified block diagram of the integrating radioastronomy spectrometer; appendix D has a schematic of the board, and E contains detailed timing diagrams.

The spectrometer board has an 8-bit input port that accepts digitized time-series data at 2.5 complex megasamples per second, corresponding to a signal bandwidth of 2.5 MHz. The time series is multiplied by a programmable window function and Fourier transformed by the Austek chip. The FFTs are a maximum of 256 points in length and, when this length, are computed seamlessly at a rate of 10^4 spectra per second. The resolution bandwidth is thus about 9.8 kHz. The complex spectra are converted to squared modulus in a 16x16 multiplier/accumulator (MAC), the 7210 (from Integrated Device Technology) and are binwise integrated with the intermediate result stored temporarily in a 256x32 bit fast SRAM buffer. At programmable dump intervals (i.e. spectral integration times) the accumulated spectrum is routed to a FIFO-buffered output port, where it is gathered asynchronously into a host computer. A handshaking protocol allows this output port to be bussed, so that a system with multiple spectrometers can be assembled if required, all boards serviced by a single
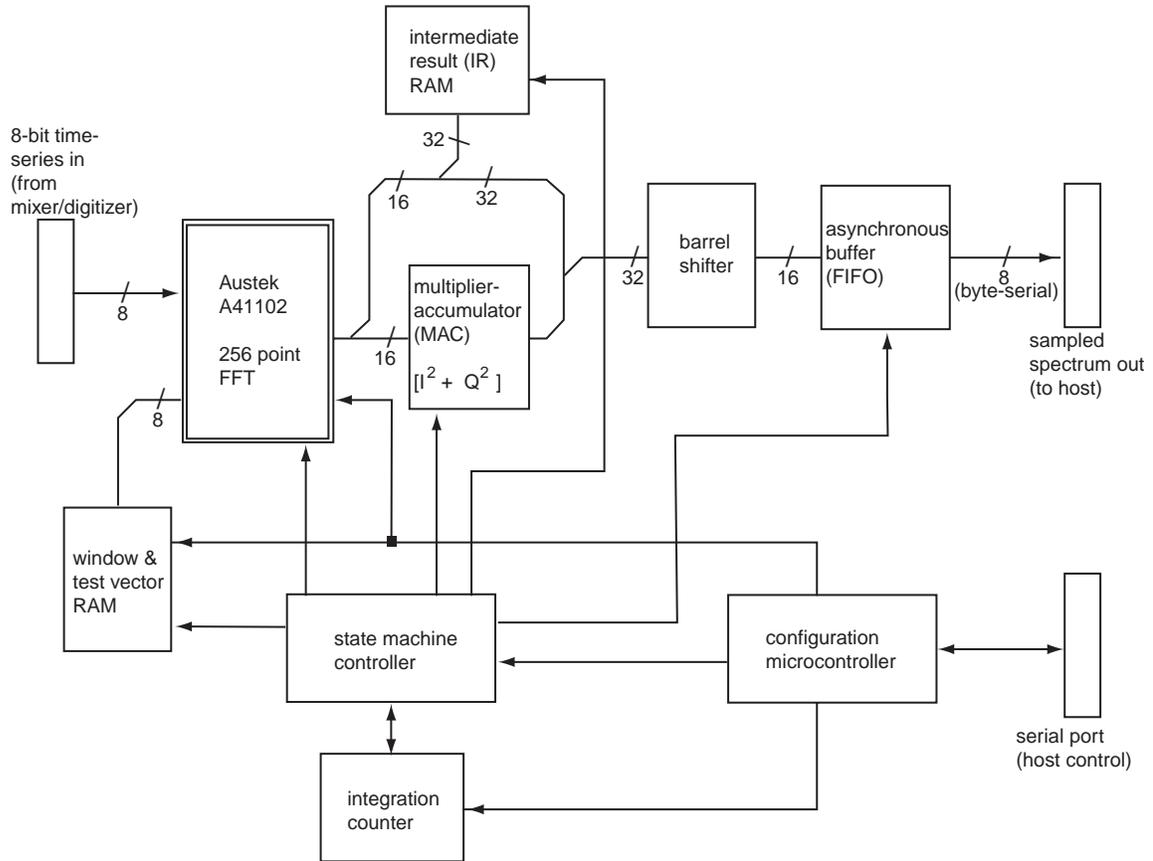
Figure 2.22: Simplified block diagram of the FFT power-accumulating spectrometer board.

host.

Figure 2.23 is a photograph of the complete assembled spectrometer board, and figure 2.24 shows some test data from bench measurements. The benefit of integrating, as predicted by the Dicke radiometer equation, is demonstrated in these plots. We note that the signal becomes dramatically clearer the longer the integration time. We are adding successive spectra, rather than averaging them, so the absolute baseline level may increase with increasing integration time (depending also on the output scaling). However the signal is growing faster than the baseline, while the level of baseline *fluctuations* are decreasing relative to both the signal and the baseline height.

Note that the narrowband feature, a continuous pure sinusoid, is broadened by the Hanning window to two bins wide, while spectral leakage is sufficiently low level to

Input data rate	2.5 complex $MSa\,s^{-1}$
Total bandwidth	2.5 MHz
FFT length	programmable, any power of 2 in range 2–256
Resolution bandwidth	9.8 kHz (assuming a 256-point FFT)
Spectral integration	$I^2 + Q^2$, 32 bit accumulation, (16 bit output)
Integration time	programmable 102.4 ms–26.2 s in 102.4 ms steps
	number of spectra integrated in range 1 to 256 k)
Output data rate (sustained)	$\frac{2.5}{N}$ $MSa\,s^{-1}$
	where N is the number of spectra integrated
Output data rate (peak)	10 $MSa\,s^{-1}$
Output spectrum rate	$\frac{10^4}{N}$ $spectra\,s^{-1}$
	(assuming a 256 point FFT)
Output data width	programmable to 8 or 16 bits
	(through asynchronous 8-bit output port)
Configuration	on-board μcontroller with command interface
	(via RS-422 serial port)
D.C. Power	$+5$ V at 0.7 A and -5.2 V at 0.04 A
Connectors	input data—20-pin dual row header
	output data and control—34-pin dual row header
	serial port to μcontroller—10-pin dual row header
	power—9-pin D-subminiature male
Board dimensions	9.187″ (6U) by 7.000″

Table 2.4: FFT spectrometer board specifications.

be buried in the baseline fluctuations. In the 13 s plots ripple in the baseline becomes apparent; this is the passband ripple in the anti-alias low pass filter on the mixer digitizer board. The skirts on the edges of the spectrum show the shape of the filter's roll off. The ordering of frequencies in each plots is from most negative on the left through DC in the middle to most positive on the right. The dip in the middle of the plot is a measurement of the DC blocking characteristic on the mixer/digitizer.

The spectrometer has an on-board microcontroller that initializes it into a default
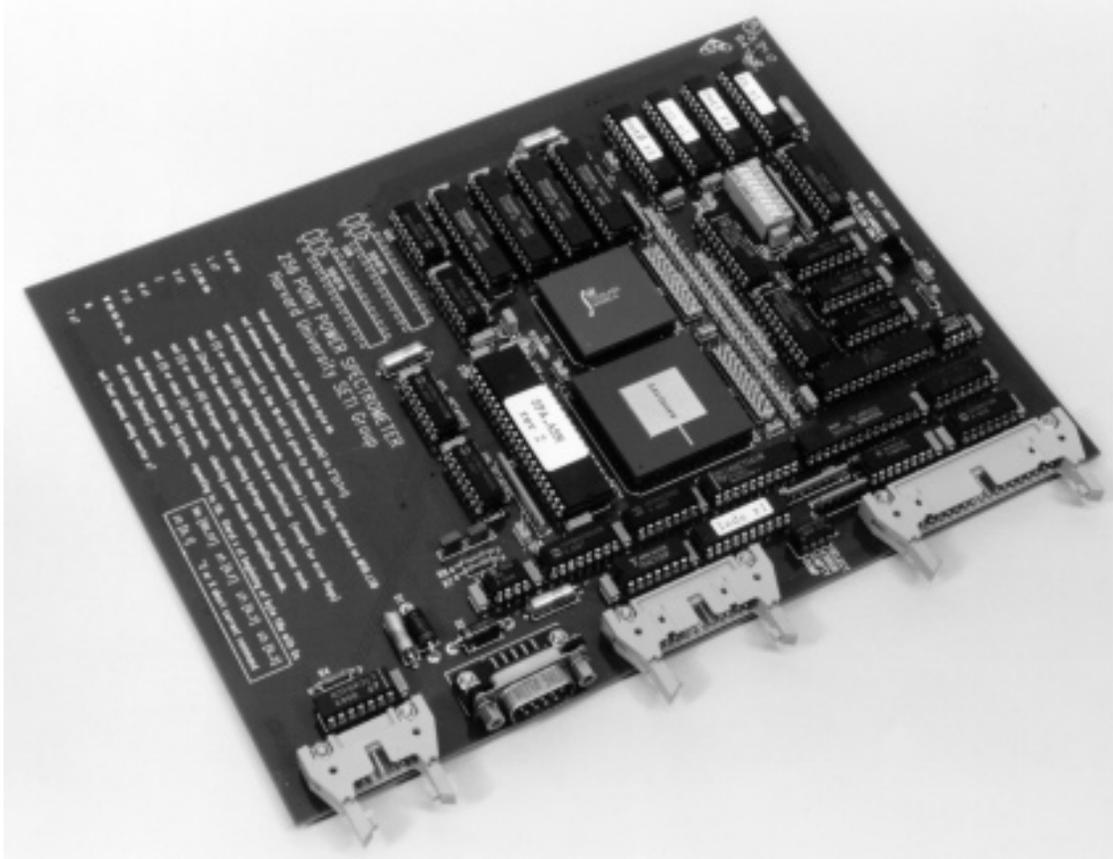
Figure 2.23: Photograph of an assembled FFT spectrometer board. Photo by Paul Horowitz.

configuration. A flexible command protocol, interfaced to the host computer via the microcontroller's serial port, then permits a custom configuration. An arbitrary window function may be loaded. The integration time may be set to as short as 0.1 ms, corresponding to every power spectrum dumped, or as long as 26.2 s, corresponding to 256 kilospectra integrated. The FFT length may be chosen as any power of 2 from 4 to 256 points per board. The start time of an FFT is synchronously triggerable, so that time windows may be interleaved from board to board. If required, the output may be taken as complex amplitudes instead of power, although in this mode integration must be disabled. With two data words per point, data rates in this mode are very high. The spectrometer board's output interface is designed to permit the high data rates needed in complex amplitude mode or with short integration times.
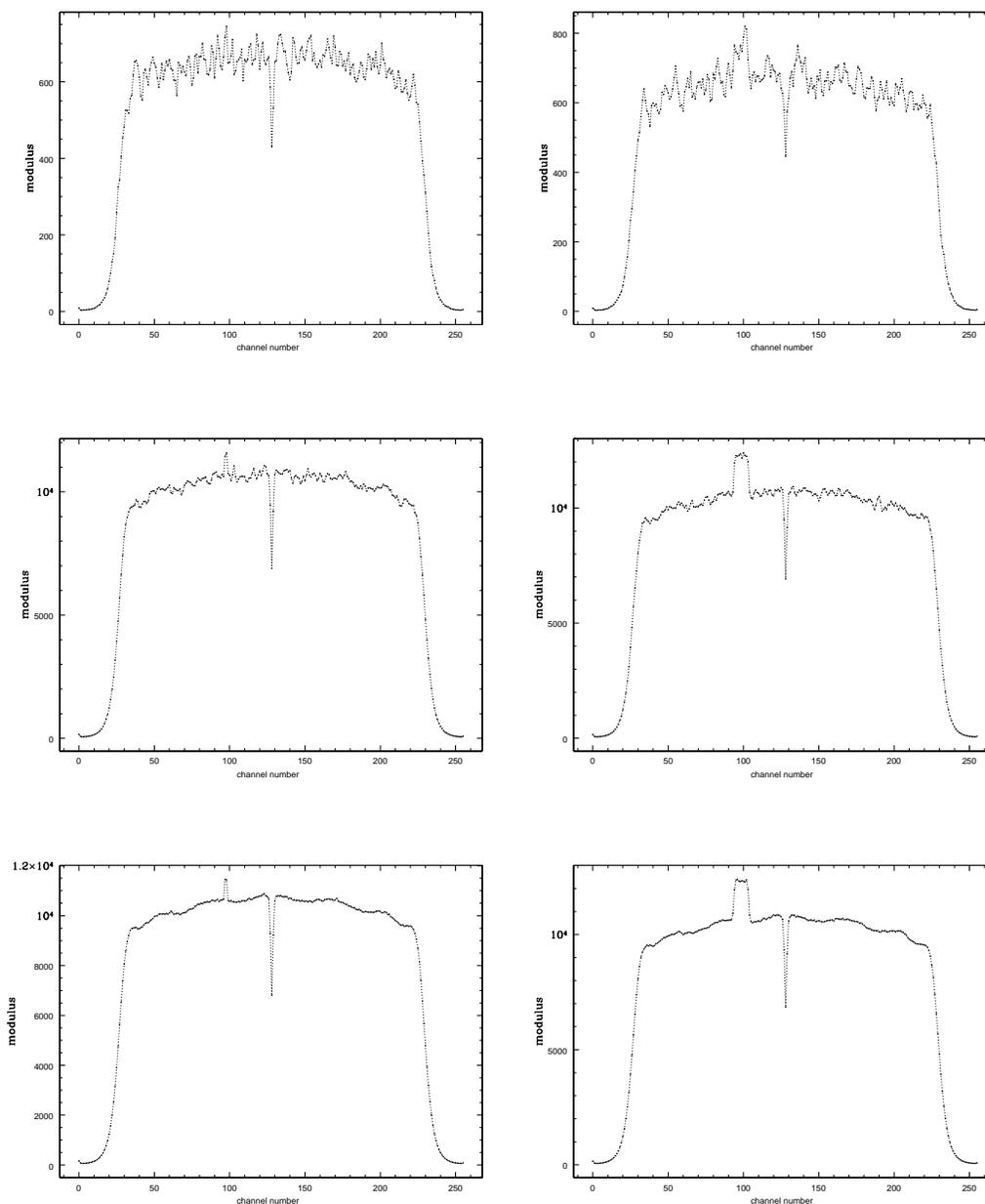
Figure 2.24: Performance of a mixer/digitizer and FFT spectrometer board set on the bench. A test signal is hidden in noise—the feature in the leftmost set of plots is a narrowband carrier, mid-bin, with SNR=-30 dB in 2 MHz; on the right it is broader with SNR=-20 dB. A Hanning window is used. The plots show the improvement in SNR with integration time τ. For the three rows the integration time is 50 ms, 0.8 s and 13 s

2.4.2 Master control computer

An 80486 based personal computer is the final element in the signal chain. The machine has two custom designed hardware interface boards installed in its industry standard architecture (ISA) user card slots. The first of these boards interfaces with multiple spectrometer boards on a ribbon-cable bus. The second custom board, a design due to the Berkeley SETI group [67], interfaces to the serial IRIG time data from the observatory time standard. The computer is also equipped with a commercial Ethernet network card, and a small computer systems interface (SCSI) card.

The spectrometers are configured via the computer's serial port, which is interfaced to the boards on a two-wire bus via an RS-232–RS-422 converter unit. Spectra are transferred over a 34-pin ribbon cable bus to the PC for display, processing and storage. The data are stored temporarily in a file on the computer's hard drive. In a day about 160 MBytes of data are gathered. Data taking is shut down once a day, at 10am AST, and the data file is spooled via a SCSI interface to a 2 GByte digital audio tape (DAT) drive. The drive takes about 13 days to fill, at which point the operator is automatically notified via email, and the tape is changed. About once a month a batch of tapes is mailed to Cambridge for analysis.

2.4.3 Software

The master control computer runs custom C language software[10] which configures the spectrometers and starts the data taking. The program then runs in an executive loop, which collects and displays spectra on the computer monitor for real-time scrutiny, and for quality control purposes. The software drives the spectrometer interface board, using PC-bus output port write cycles to generate the appropriate interface

---

[10]A note on credits: The author and I. Avruch wrote the basic executive loop which programs and interrogates the spectrometers. D. Leigh rewrote much of this code, contributed the user menu and fast graphics for the spectrum display, and wrote some elegant Ethernet-based data storage software (which, unfortunately, we did note use). Avruch enhanced and rewrote the graphics display code, programmed the tape data logging and the logging of status files to the observatory workstations, tied up loose ends, and maintained the software.

signals. 16-bit data is collected from each spectrometer board in turn via the ribbon-cable bus. Status information is read from the status port. The IRIG time is collected from the IRIG interface board, and is used to synchronize the internal PC clock to the observatory time standard.

The software adds two types of headers to the data. An ASCII file header, one per data file, contains information relevant to the whole file, including date and time data taking started, number of boards in the system, settings of the spectrometers, for example Austek scale bits. There is also a comment field, which is a copy of the first 290 characters of a free-format file "comment.txt", containing operator comments, and created with a text editor in the directory in which the program runs. Every 10 seconds a data record containing the spectra from the current dump is written to the file. Attached to the data from each board is a separate binary record header, with time sensitive status information. Most important are the time stamps, which are logged to a precision of 0.1 ms—note, however, that the IRIG interface only updates once per second, adequate accuracy given the 180 s drift time. The record header also contains the board number, a flag indicating if the calibration diode is firing in this record, the number of spectra integrated on the board, and parity and overflow error flags. Appendix F contains a complete description of the file format.

The PC runs DOS, which is not a multitasking operating system. So, once a day, the data taking is shut down, and a script controls the spooling of data from the hard drive to digital audio tape (DAT). This process takes about one half-hour. The script also sends a status message via the Ethernet to one of the workstations on the Arecibo network, which in turn forwards the message to a mailing list consisting of experimenters at Harvard, MIT and Arecibo, and the telescope operators. A crucial function of the status message is to notify the telescope operators when it is time to change the data tape. A warning also appears on the monitor screen. The status message also notes any errors which the self-test functions of the system may have picked up. A common error occurs when, for one reason or another, the IRIG serial line gets disconnected. The status messages are very useful during data analysis, warning us about possible problems with the instrument during data acquisition.

2.5 Features for pulsar work—the AOFTM

We have cooperated closely with Professor James Cordes at Cornell to make the spectrometer useful for the higher speed requirements of pulsar studies. The dump interval can be shortened to as little as a single spectrum ($\sim 100 \ \mu$s). The FFT length can be chosen as a power of 2 from 4 to 256 points per board. To avoid under-sampling the spectrum, it is possible to asynchronously trigger the start of data taking, allowing the time windows to be overlapped, if required. The spectral data points can be output as either squared modulus (power) or complex vectors (thereby preserving phase information). From pairs of boards providing complex spectra in orthogonal polarizations, the four Stokes parameters can be formed, facilitating polarization and Faraday rotation studies. A design study [68] by Professor Cordes gives a more detailed motivation for these features.

Cordes and Koechner [69] have designed an interface which handles the high data rates needed for dynamic pulsar studies. The interface is included in an instrument whose components also include an IF section (including a set of mixer/digitizer boards and an LO-array, designs taken from the Harvard's BETA project [45]), a set of 16 spectrometer boards, the high speed data interface and a workstation. Together these form the Arecibo Observatory Fourier Transform Machine (AOFTM)—a pulsar spectrometer which is one of the instruments built as a part of the Arecibo upgrade. The spectrometer boards are grouped in logical units, four boards per group, each handling a single 2.5 MHz chunk of spectrum, in two polarizations and with two time windows 50% interleaved.

The cause of a major difficulty in pulsar observing are the data rates which must be supported. The worst case is for complex amplitude output, where each data point is two coordinates, spectral integration is not an option. If each number is represented by two bits the data rate is

$$R = \frac{16 \text{ boards} \times 256 \text{ points} \times 2 \text{ coordinates} \times 2 \text{ bits}}{102.4 \ \mu \text{s}} = 160 \, \text{Mbit s}^{-1} = 20 \, \text{MByte s}^{-1}$$

$$(2.40)$$

Rates of this magnitude stretch the limits of currently available data storage technologies, both in terms of sustained data rates, and total storage capacity, for any reasonable observation time. If power (as opposed to complex) spectra are extracted the situation is a little easier to handle, since polarizations are typically summed, and, depending on the time-resolution required for the particular observation, a few spectra can be co-added to throttle back the data rates. It's expected that the AOFTM will mostly run in power spectrum mode, until fast data storage technologies are capable of sustaining the rates the machine can deliver.

2.6 Application to undergraduate teaching

Professor Cordes is also using the spectrometer board, mixer digitizer, and PC interface card designs, to teach a laboratory course, Astronomy 410, at Cornell. They apply the spectrometer, a 5 m antenna and has a 60K uncooled low noise amplifier to 21 cm spectroscopy. Typically his students do drift scans and take data for 24 hours. The beam cuts through the galactic plane twice, once toward the inner galaxy, once toward the anti-center. With those measurements they measure the range of Doppler shifts and solve for the rotation velocity of the galaxy at the solar circle and also make an estimate for the radial size of the HI disk. They achieve greater than 10 dB SNR in the strongest directions with a one second integration. The equipment has been very reliable and useful, according to Professor Cordes.

Figure 2.25: The completed rack of experimental equipment, with the author, just after its installation in the control building at Arecibo. Starting at the top of the rack it shows the downstairs receiver, IF splitter, local oscillator array, mixer/digitizer array, FFT spectrometer array, first local oscillator (before our much smaller and simpler in-house design was completed to replace this general-purpose Hewlett Packard synthesizer), and power supplies. The control PC stands to the left of the rack. The photo is by Dr. Davis.

Chapter 3

Data analysis procedures

We have been collecting data since January 1995, with some down-time due to in-strumental problems, hurricanes, and Arecibo upgrade activities. RFI is a serious problem and appears to have worsened over the course of the experiment. We use robust techniques, which minimize the effects of time variable RFI, to co-add many days worth of observations. The output from this process are maps of signal intensity as a function of frequency and sidereal time.

Non time-variable, or stationary, RFI is not dealt with in the co-adding stage. We apply a modified matched filtering technique to the co-added maps which draws out the desired signal, and mitigates the stationary RFI. Matched filtering is cross-correlating with a template which has the shape of the expected signal. The signal has a shape in time, a result of the source drifting thought the beam, and a shape in frequency, which is a characteristic of the emission from the protocluster. We modify the shape of the correlation template to reject stationary RFI; this slightly degrades the match to the signal, but greatly improves the matched filter output. The development on the matched filter algorithms is described in §3.6.

Candidate signals are further investigated by splitting the data in a given sidereal time range into sub-sets, and creating a map from each. We apply the matched filters to each map and look for the signal to repeat. This indicates the sidereal periodicity we expect in a true celestial signal.

3.1 Data artifacts

The RFI environment in the frequency band under investigation at Arecibo is extremely hostile. We built a custom instrument, including a custom feed and an uncooled front end receiver. Arecibo was undergoing a significant upgrade during the operation of our experiment. Upgrade related activities included welding on the platform, and mechanical changes to the feed platform supports, which caused the platform and the catwalk to move up and down. Sometimes it was necessary to disconnect the system cabling when these mechanical changes were made. So it is not surprising that the data initially obtained with the system included artifacts. This section includes a discussion of such artifacts and their presumed causes.

3.1.1 Radio frequency interference (RFI)

The radio spectrum is a limited resource, while the commercial and military applications of radio are rapidly growing. Especially troubling is the recent trend to space-based use of the spectrum. Examples of recent commercial forays include the development of constellations of low-Earth-orbit satellites used for direct handset-to-satellite communication systems. Iridium, being developed by Motorola, plans 66 satellites by this year, while Teledesic, a private startup, plans 288 by the year 2002. Iridium uses the L-band frequencies 1616–1626.5 MHz for satellite to handset communication. These are uncomfortably close to the scientifically interesting HI and OH spectral lines. Both systems use the Ka-Band for satellite to base station; Iridium uses 19.4–19.6 GHz for the downlink and 29.1–29.3 GHz for the uplink while Teledesic will operate at 18.8–19.3 GHz for the downlink and 28.6–29.1 GHz for the uplink. The frequencies are very close to a molecular line associated with water at 22 GHz, important in maser studies.

Plans for increased use of satellites in personal communication systems are especially troubling long term for radio astronomy. Since the satellites are celestial objects, they appear squarely in the sky beam of radio telescopes. It is necessary, for the long term health of radio astronomy, that the use of scientifically important frequency bands,

certain times of day, and geographic areas close to radio telescopes, be protected as far as possible for scientific purposes. This involves both statutory protection via FCC allocations, and negotiation with commercial concerns. The scientific staff at Arecibo happens to be closely involved in these negotiations, and statutory frequency allocation committees. There is a lot of information on the web pages for Iridium [70] and Teledesic [71], while two recent news articles on these systems and their impact on radio astronomy are cited in the bibliography [72] and [73].

ν (MHz)	Allocation
218 - 219	IVDS (Interactive Video and Data Services)
	Wildlife and ocean buoy tracking and telemetry
219 - 220	Automated Maritime Telecommunications System (AMTS)
	Mississippi River and connecting waterways, Gulf of Mexico
	Amateur point-to-point fixed digital message forwarding systems,
	Wildlife and ocean buoy tracking and telemetry
220 - 221	Land mobile - base stations
221 - 222	Land mobile - mobile stations
222 - 225	Amateur (1.25 meter band)
225 - 243.00	Military
243.00	Survival craft stations, ELTs
243.00 - 328.6	Military

Table 3.1: Frequency allocation chart for the HI experiment band.

Unfortunately the frequency band from 219–251 MHz is not protected in any way, and we have to live with the situation. Table 3.1.1 shows that the band is mostly allocated to military services. It has been suggested [74] that the military has recently transfered more traffic to these bands, after other military bands were reallocated by the Federal Communications Commission (FCC) to other uses. We have noticed that the interference situation seems to have deteriorated over the course of our experiment.

The RFI is mostly narrow-band. Some carriers are fairly steady, but others have time-variable amplitudes, or seem to turn on and off. The low and the high frequency

regions of the spectrum seem to suffer the most from RFI, with the region from 223–
248 MHz somewhat quieter. Later sections will discuss the analysis methods we use
in our attempt to deal with the RFI.

(a) clean   (b) dirty
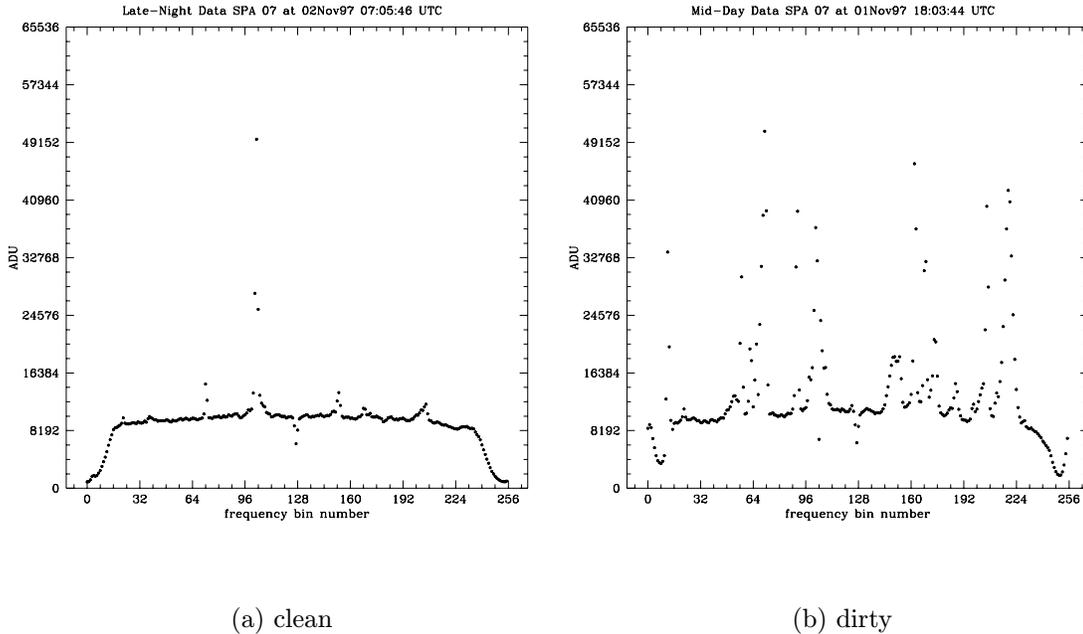
Figure 3.1: Typical "clean" and "dirty" spectra from a single board. The left hand
plot is from the early morning, while the other is from a Saturday in the early after-
noon. Plots courtesy of I. Avruch.

As might be expected, the average level of the RFI shows a diurnal variation. Figure
3.1 shows two spectra from a particular spectrometer board, centered at 232 MHz, at
two times of day. The times in the titles of these plots are universal time (UTC)—to
convert to Atlantic Standard Time (AST) subtract four hours. So the relatively clean
plot is from about 3 am local time, while the plot on the right is from about 2 pm.
By scrutinizing plots similar to these we have determined that the cleaner times of
day are limited to the hours between 1 am and 5 am AST. Outside of this window
the data quality deteriorates markedly. It should also be said that this spectrum is
typical of an exceptionally clean sub band; the data from other boards looks more
like the right hand plot even at 3 am! We have decided to only use data from the
early morning hours between 1 am and 5 am in our analysis.

We have also noticed RFI with a broadband structure. This type of interference lifts the level of the entire baseline for short time periods. Early on in the experiment, a source of broadband RFI was traced to a faulty high voltage power supply for an on-site laser used by Arecibo's atmospheric scientists. The interference abated after this unit was repaired, but we have noticed similar interference subsequently, mostly during work days. Welding on the feed platform by upgrade workers is the presumed cause.

3.1.2 Intermittent baseline ripples
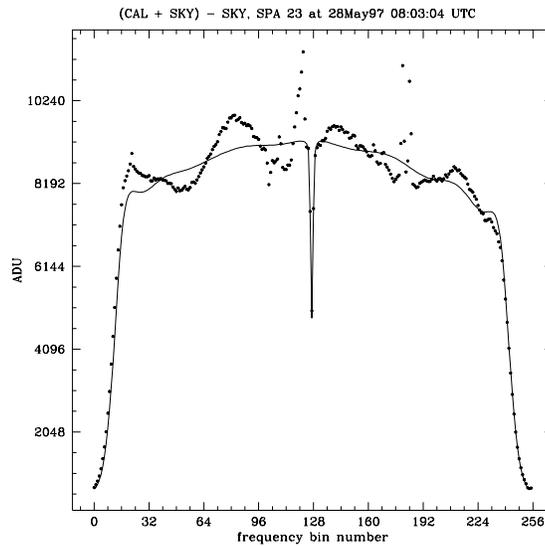
Figure 3.2: Baseline ripples during a cal cycle. Plot courtesy of I. Avruch.

The baseline of the spectrum has systematic structure. Some structure is well understood, such as ripples from the passbands of the various filters (RF, IF and baseband) in the system. However we see other ripples, which, while clearly systematic, are not consistently present; hence they must have another, intermittent, source. Arecibo's feed platform is a large reflecting structure above the dish, and it is natural to suppose that standing waves may be set up by radiation reflecting between feed platform and reflector. The scientific staff at Arecibo have observed standing waves of this sort in other observations [66].

We have noticed that the amplitude of the ripples tends to increase when the calibration diode fires. This suggests that it may be the source of some of the standing wave power, and that perhaps some of the power is leaking out of the feed antenna when the diode fires. However there are many ways to set up standing waves other than by reflections between platform and reflector and excited by the feed. Bench measurements on the receiver indicate that the feed is well isolated from the noise diode by the -20 dB directional coupler in the system front end. The wavelength of the ripples is not consistent with the shortest path between platform and reflector (133 m); it's consistent with a longer path. Since the standing wave mode is a half wavelength the frequency associated with the path length s is $\nu = c/2s$. Hence 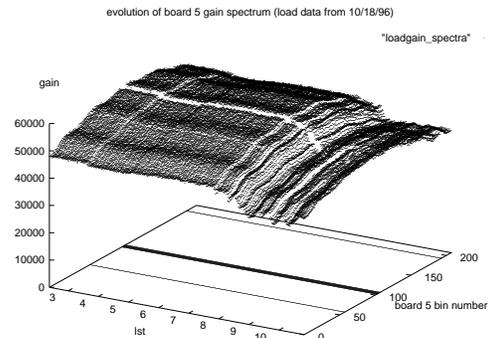a 133 m path length predicts 1.13 MHz for the wavelength[1] of the baseline ripples. Scaling off a spectral plot, we estimate 0.8 MHz for the wavelength of the ripples. This corresponds to about a 188 m path length.

There are some other hypothetical paths: standing waves in the cables to downstairs (about 457 m or 0.33 MHz—not consistent with the observed ripples), or perhaps a leaky connector somewhere along the path radiating back into the helix. It is also possible that radiation from the analog electronics in the downstairs rack is leaking from the unshielded room in which it is housed, and feeding back into the helix. These feedback paths are possible because we ship RF downstairs instead of the more usual practice of mixing upstairs. This kind of feedback is worrysome because of potential instability, but even without oscillation such a feedback path may cause standing waves. Most perplexing, though, is that the baseline ripples are not consistently present, even when the calibration diode fires. The intermittent nature of the ripples, and our limited opportunities to visit the site, especially recently, has limited our ability to investigate these ripples properly. Frankly, we do not properly understand them, but their apparent correlation with the firing of the calibration diode, as well as the inconstant nature of the cal power (discussed in chapter 2), prompted us to shut down the cal on the 20th November 1997.

---

[1]This is the wavelength of ripples in a spectrum, and so is expressed in units of frequency.
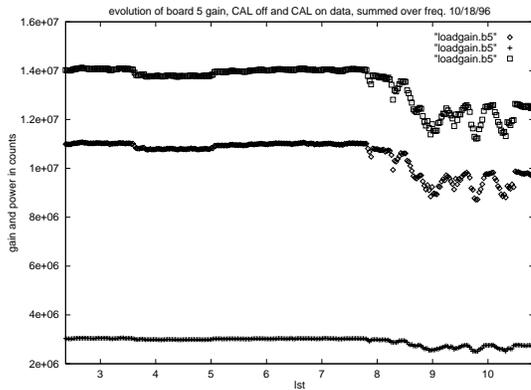
(a) load total power, board 5   (b) load spectrum, board 5

Figure 3.3: Load spectra from board 5, showing broadband fluctuations in gain. Gain is computed by subtracting from the cal-on spectrum the nearest cal-off spectrum.

3.1.3 Gain fluctuations

We have noticed changes in the continuum level out of the receiver over long time scales. Spectra at a given sidereal time but from different days show a 20% peak variation in the baseline level. These changes could be attributed to continuum RFI, or to gain fluctuations. In practice it is difficult to tell broadband RFI from gain fluctuations in sky data. However spectra taken with a load connected to the front end in place of the antenna show significant fluctuations. The receiver seems to be rather stable at night, but when the sun rises the fluctuations become significant—8 h LST corresponds to about 6:30 am AST in this plot. Figure 3.3 shows the fluctuations in the load data in a single board in both total power, and across the spectrum. The fluctuations are duplicated similarly in all the boards, so they are truly broadband.

We installed the calibration diode in mid-1996 so that we would have a means to remove the gain fluctuations. The cal did not perform as expected, and further was suspected of giving rise to artifacts, and so was shut down as discussed above. We have done two data reductions, one with cal (limited to the times when the calibration diode was installed) and one without (which uses more data, since more is available with the calibration diode).

3.2 Chronology of system changes and faults

For the past three and a half years Arecibo has been undergoing a major upgrade, hurricane seasons have come and gone, and more sophisticated computer models for optimizing our feed position have been developed [50]. The table here documents changes and disruptions to the antenna, as well as other major failures and maintenance shutdowns of the front end system. Some of the faults, such as the various times the antenna or the IRIG time standard were inadvertently disconnected, may be thought of as an unavoidable consequence of running a complex experiment at a remote site. We use the table below when reducing data to exclude data from times when there are problems with the experiment.

3.3 Robust methods for co-adding data

We need methods for co-adding data in the presence of strong RFI. Our original intention was to simply average the data in sidereal coordinates, thereby improving signal to noise ratio as predicted by the radiometer equation. However this equation assumes a signal hidden in Gaussian distributed noise. The RFI dominated nature of our experiment means that this distribution is a very poor model. Neither are the statistics of our data stationary. It is well known that the median is an estimator which is robust against outliers. We did a simulation to investigate the use of the median in the presence of RFI-like outliers. We discuss these simulations below.

3.3.1 The average and the median in the presence of outliers

The feasibility calculations in section 1.6 were computed assuming that the data is "co-added" in the sense of averaging. However the median is more robust against the outliers caused by RFI. For Gaussian noise the variance of the median of N points is larger than the variance of the average; that is, the median is noisier than the average. Nonetheless, our simulation shows that when the data are corrupted by RFI the median is the better statistic to use.

date	event	comments
07/10/94	data taking first started	2 feeds, 8 kilochannels
10/24/94	2nd helix canted 30°, disconnected	feed interactions
11/15/94	2nd helix taken down	feed interactions
01/20/95	remaining helix repointed 10° S	gain ↑, noise ↓
	RFI monitor disconnected	
01/23/95	RFI monitor reconnected	
01/31/95	feed position measured	
'soon after'	one catwalk tiedown loosened	upgrade activity
02/08/95	feed moved one hanger inboard	improve gain [50]
02/20/95	catwalk tiedown tightened	upgrade activity
08/95	flatline, front end amplifier failure	replaced on this date
08/27/95	helix removed; replaced by load	hurricane season
03/15/96	helix replaced, restarted	pointing per section 2.1.2
08/96	CAL installed	CAL software upgraded
09/09/96	hurricane Hortense	6 in. coax disconnected
09/26/96	6 in. coax still disconnected	no power to front end
10/03/96	power restored to front end	
10/14/96	channel 1 down	
10/14/96	RFI monitor misconnected to ch 1	date of diagnosis
10/15/96	ch.1 repaired, helix replaced	restarted
12/96	Antenna disconnected, S-band tests	upgrade activity
8/1-6/97	front end shutdowns	noise measurements
8/5/97	new LO array	replaced borrowed unit
8/6/97	Antenna switched to ch. 2	lower noise channel
10/08/97	system shut down	bad DAT tape drive
10/20/97	restarted	DAT replaced
11/4/97	noticed ch 1 flatlined	inactive channel, data ok
11/20/97	CAL shutdown	noticed CAL misfiring
2/7–20/98	no status messages, data OK	network problem

Table 3.2: Chronology of system faults.
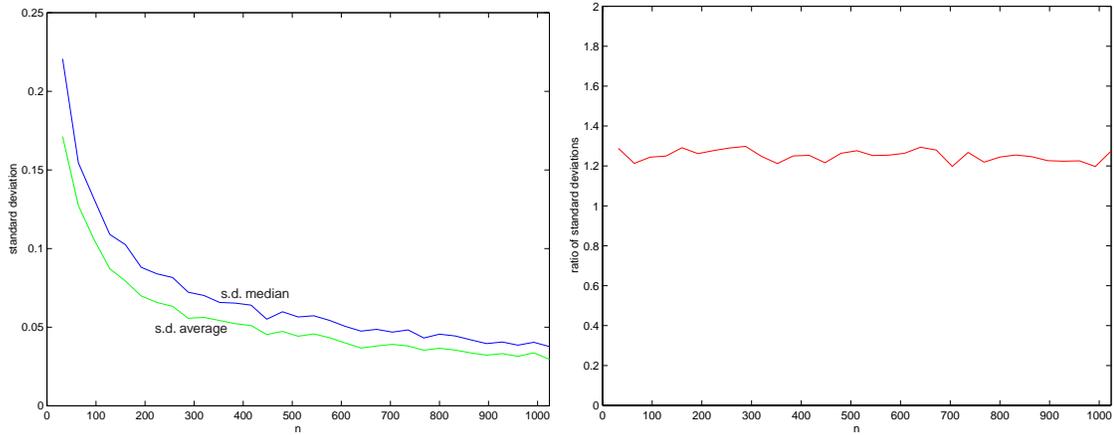
(a) $\sigma_{\mathrm{median}} > \sigma_{\mathrm{average}}$ (b) $\sigma_{\mathrm{median}}/\sigma_{\mathrm{average}}$

Figure 3.4: Simulated sample standard deviation of the average, and of the median as a function of number of data points averaged, and the ratio of the curves. Original data are drawn from a Gaussian distribution with $\sigma = 1.0$.

Consider samples of Gaussian distributed noise of variance σ^2. The variance of the average of N samples is $\sigma^2_{\mathrm{average}} = \sigma^2/N$. Interestingly, the distribution of the median can be expressed analytically, and the corresponding expression for the variance of the median is $\sigma^2_{\mathrm{median}} = \pi\sigma^2/2N$ [75], or in terms of the standard deviation,

$$\sigma_{\mathrm{median}} = \sqrt{\frac{\pi}{2N}}\,\sigma = \frac{1.253}{\sqrt{N}}\,\sigma \tag{3.1}$$

We wrote a Matlab program to numerically demonstrate this result. We first verified that the simulation produces results consistent with equation 3.1. Then we experimented with more realistic distributions simulating data corrupted by RFI. Figure 3.4 show hows the standard deviation for average and median behaves with increasing number of samples N. The underlying distribution is Gaussian with unit standard deviation. The ratio of the two curves is the right hand plot. As predicted by the equation the ratio of the curves is a constant and approximately equal to 1.25.

Having verified the simulation program for the Gaussian case, we simulated data corrupted by RFI by replacing 10% of the Gaussian noise points with exponentially
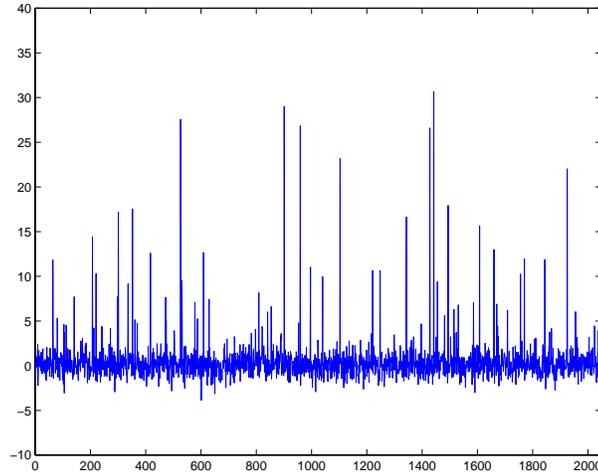
Figure 3.5: Noiselike data corrupted by simulated RFI. Made by generating Gaussian pseudo-random numbers ($\sigma = 1.0$), and replacing 10% of the points by exponentially distributed random outliers.
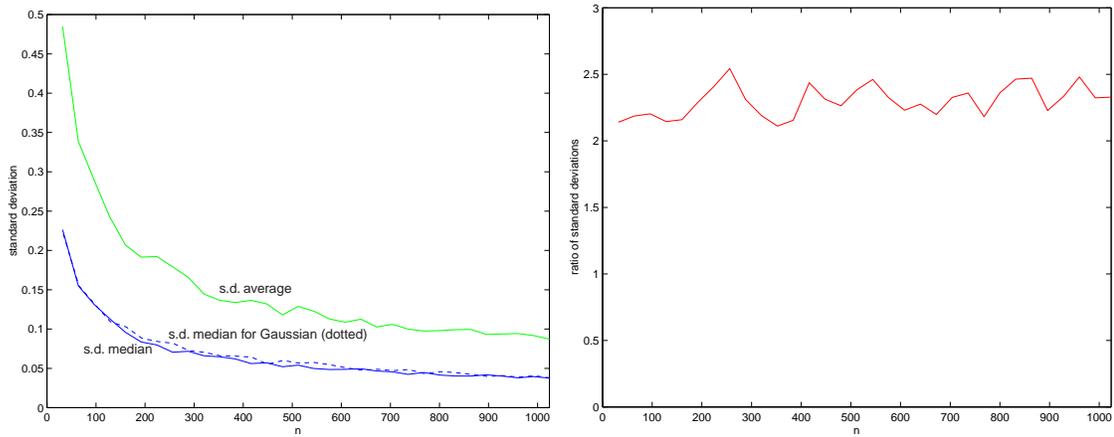
(a) $\sigma_{\text{median}} < \sigma_{\text{average}}$

(b) $\sigma_{\text{average}}/\sigma_{\text{median}}$

Figure 3.6: Simulated sample standard deviation of the average, and of the median, and the ratio of the curves. Original data are Gaussian distributed random variables with $\sigma = 1.0$, corrupted by exponentially distributed outliers. For reference, the curve for Gaussian noise is included (dotted line).

distributed points of much greater variance, uniformly and randomly distributed in time. Figure 3.6 shows what the time series data looks like, qualitatively much like

our RFI corrupted data. Once again we co-add the simulated data using the average and the median. Inspection of the curves shows that the standard deviation of the median changes very little with the addition of the outliers, while the variance of the average doubles. In the presence of time varying outliers it is clear that the median is the better estimator of the desired celestial data. It is also possible that a combination of average and median may combine the resistance to outliers of the median with the better performance of the average for Gaussian noise. We will discuss this further in section 3.5.

3.4 The equivalence between folding and comb filtering

In thinking about methods for filtering out RFI and favoring the signal, we discovered an interesting, and somewhat counterintuitive, result. We wanted to use the known sidereal periodicity of the signal as a constraint to help us pick the signal out of the RFI and other noise, which has no such periodicity. It occurred to us that there may be some profit in trying to filter out the periodic signal in the Fourier domain. Specifically, no matter what the shape of a periodic signal, it can only contain spectral energy at harmonics of its repetition frequency. So prior to folding, it seemed intuitively that it would be beneficial to transform to the Fourier domain, zero out all frequencies except for the signal and its harmonics, transform back to time, and then fold the data. This filters out as much of the RFI energy as possible, while retaining all of the signal energy. In fact, to out surprise, it turns out that this process is exactly equivalent to just bin-wise averaging the data, which we term "folding".

We give here the proof in a special case. Recall the equations for the forward and inverse DFT

$$H_n = \sum_{k=0}^{N-1} h_k \, e^{2\pi i k n / N} \tag{3.2}$$

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} H_n \, e^{-2\pi i k n / N} \tag{3.3}$$

We examine the special case $N = 16$ with the signal repeating 4 times. The filtering operation will preserve only the fundamental and every fourth harmonic in the Fourier domain. So we are left with H_0, H_4, H_8 and H_{12} given by

$$H_0 = \sum_{k=0}^{15} h_k \tag{3.4}$$

$$H_4 = \sum_{k=0}^{15} h_k \, e^{\pi i n / 2} \tag{3.5}$$

$$H_8 = \sum_{k=0}^{15} h_k \, e^{\pi i n} \tag{3.6}$$

$$H_{12} = \sum_{k=0}^{15} h_k \, e^{3\pi i k n / 2} \tag{3.7}$$

It turns out that the phase factors in the above components are all either purely real or purely imaginary for all k; there are no exponential phase factors. Writing the summations out explicitly

$$H_0 = h_0 + h_1 + h_2 + h_3 + h_4 + h_5 + h_6 + \ldots + h_{14} + h_{15} \tag{3.8}$$

$$H_4 = h_0 + ih_1 - h_2 - ih_3 + h_4 + ih_5 - h_6 - \ldots - h_{14} - ih_{15} \tag{3.9}$$

$$H_8 = h_0 - h_1 + h_2 - h_3 + h_4 - h_5 + h_6 - \ldots + h_{14} - h_{15} \tag{3.10}$$

$$H_{12} = h_0 - ih_1 - h_2 + ih_3 + h_4 - ih_5 - h_6 + \ldots - h_{14} + ih_{15} \tag{3.11}$$

Now we compute the inverse transform of the filtered spectrum by direct application of the inverse DFT. Conveniently, most of the terms cancel and the phase factors once again are purely real or imaginary. We indicate the post filtering terms in the time series with a prime. The zeroth term has no phase factors and is readily seen to be the average of every fourth component in the original time series

$$h_0' = \frac{1}{4}(h_0 + h_4 + h_8 + h_{12}) \tag{3.12}$$

And it's almost as straightforward to show that the first component is also a simple average

$$h_1' = \frac{1}{16}(H_0 + H_4 e^{-i\pi/2} + H_8 e^{-i\pi} + H_{12} e^{-3i\pi/2})$$

$$h_1' = \frac{1}{16}(H_0 - iH_4 - H_8 + iH_{12})$$

$$h_1' = \frac{1}{4}(h_1 + h_5 + h_9 + h_{13}) \tag{3.13}$$

and the last two terms follow similarly,

$$h_2' = \frac{1}{4}(h_2 + h_6 + h_{10} + h_{14}) \tag{3.14}$$

$$h_3' = \frac{1}{4}(h_3 + h_7 + h_{11} + h_{15}) \tag{3.15}$$

Clearly these are the same components one would get by folding. The time series must have a periodicity of four samples, since it derives from every fourth harmonic. Hence the pattern repeats, with $h_0' = h_4' = h_8' = h_{12}'$.

It turns out that this result is quite general, independent of the length of the overall series and the period of the desired signal. Folding is exactly equivalent to filtering out only those spectral components corresponding to the the harmonics of the signal

of interest. In this sense folding appears to be an optimum technique for extracting signals with the same periodicity as the folding.

3.5   Co-adding procedure details

The co-adding programs were written in the "C" programming language by Ian Avruch. Avruch developed the algorithms discussed in this section, although the investigations of the co-adding algorithm in §3.3.1 and §3.4 are the work of the author. Initially Avruch was very committed to using the calibration diode to remove gain fluctuations and to absolutely calibrate the output maps. It was only late in 1997 that the difficulties associated with the calibration data became clear to us. It should be noted here that the very high levels of RFI in the maps conspired to hide these difficulties from us. Specifically, it is often difficult in a given spectrum to even pick out the baseline, to get a reliable indication of its level, and to separate spurious ripples from RFI.

Avruch, in fact, did two reductions, the first which uses calibration data, albeit possibly unreliable, to produce maps with calibrated intensity, and a second reduction which ignores the calibration data. I summarize here the details of the algorithms used, including the calibration step. The second, uncalibrated, reduction differs in another way. It combines median filtering to exclude outliers; however a range of inner points are retained and these are then averaged. We felt that this method would give the lowest noise in the output maps, especially in the face of possible gain fluctuations. Avruch has also attempted to derive a calibration for these observations by comparing them to all-sky maps of galactic background. He was quite successful at times when our beam was pointing at the galaxy, but when off the center of the galaxy the RFI again prevented getting a reliable measure of sky temperature.

3.5.1 Gain calibration

In order to adjust each spectrum to the same gain level we need some measure of the gain when that spectrum was taken. We get this from the incremental power added by the calibration diode when it fires every ten minutes. We know from our bench measurements of the front end receiver the characteristic of the calibration diode as a function of frequency, as well as the receiver temperature (see figure 2.16). When the antenna is connected, the system temperature includes ground temperature as well as a contribution from the sky which changes depending on the sidereal time. When the diode fires, its temperature is added to these existing system temperature components. However it is the calibration diode's temperature alone which is a measure of current system gain. In order to separate out the contribution from the diode, we subtract the diode-on spectrum from the average of the two neighboring diode-off spectra.

We also need a measure of the baseline power out of a given board. In the presence of time variable RFI this is a difficult problem, since an RFI carrier turning on increases the average power level in the spectrum, and this could be confused with a gain change. We need some measure of the baseline level in an RFI corrupted spectrum. Once again the median is applied as a robust estimator. Fig 3.7 shows that, if an RFI corrupted spectrum is sorted it appears that the 33% level has the gentlest slope. We take a point at the 33% level to be an RFI robust indicator of the baseline level for a given spectrum. In this way correction factors which correct for the gain fluctuations are computed.

3.5.2 Re-sampling to LMST time ticks

Our integration time is about 10 seconds, and the integration period drifts with respect to sidereal time. Given the three minute drift time of a point source through the beam, we have approximately eighteen integrations per beam, so our beam is heavily oversampled. Every 10 minutes, the calibration diode fires. Data during a cal cycle are used for calibration, but are not co-added with the rest of the data. To simplify the co-adding, the non-calibration data are re-sampled so that each sidereal
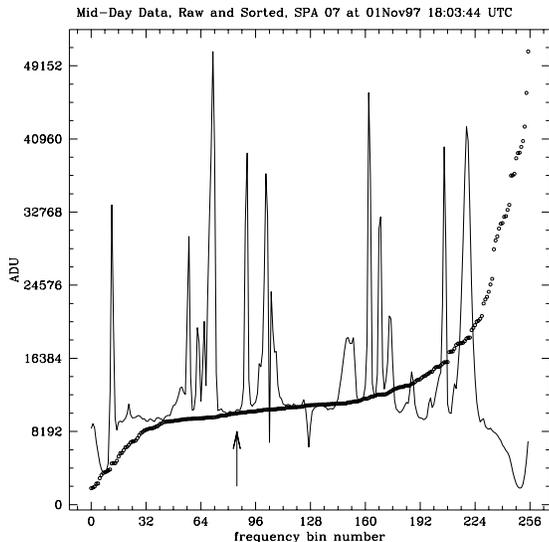
Figure 3.7: A sorted dirty spectrum, showing the rationale for using a 33% point as the estimate of the baseline level. The roll-off on the left of the sorted curve corresponds to the filter skirts, while the upward slope on the right hand side is due to the RFI. Plot courtesy of I. Avruch [50].

minute contains four samples lying exactly on LMST quarter minute intervals. The re-sampling involves placing a window around the 15 second tick, and stretching the window until it includes three non-cal points. The median of these three points is then used as the re-sampled point, and it is placed exactly at a 15 second tick.

3.5.3 Inertial frame

The radial frequency with respect to the CMB rest frame is computed from the best estimate of the pointing of the sky beam. Doppler shifts due to the earth's motion with respect to the CMB are corrected by shifting bins in the spectra appropriately. The Doppler shift includes a component which is constant in time, due to the motion of the local group of galaxies with respect to the CMB rest frame. More important is the differential Doppler shift, due to the rotation of the earth, and its orbit around the sun, since this would tend to smear the signal as data from different days is co-added. The biggest velocity is that of the earth around the sun, for which $v/c = 10^{-4}$. This is

just the fractional shift in frequency, so at 235 MHz the Doppler component is about 24 kHz, causing a total shift of around 5 bins. Because of the relatively small velocities involved and the fairly broad bin-width of our spectrometer, the shifts involved are small. This is fortunate, since significant shifts would tend to smear the RFI into neighboring clean bins.

3.5.4 Baseline ripple removal

We have characterized the ripples in the passband by connecting a load at 300 K to the front end in place of the antenna for periods of time. A long time average gives the plot in 3.8, which shows the averaged response of channel one over the full 32 MHz band. The ripples are systematic structure in frequency which are removed from the already co-added image by dividing all the spectra by the load template.
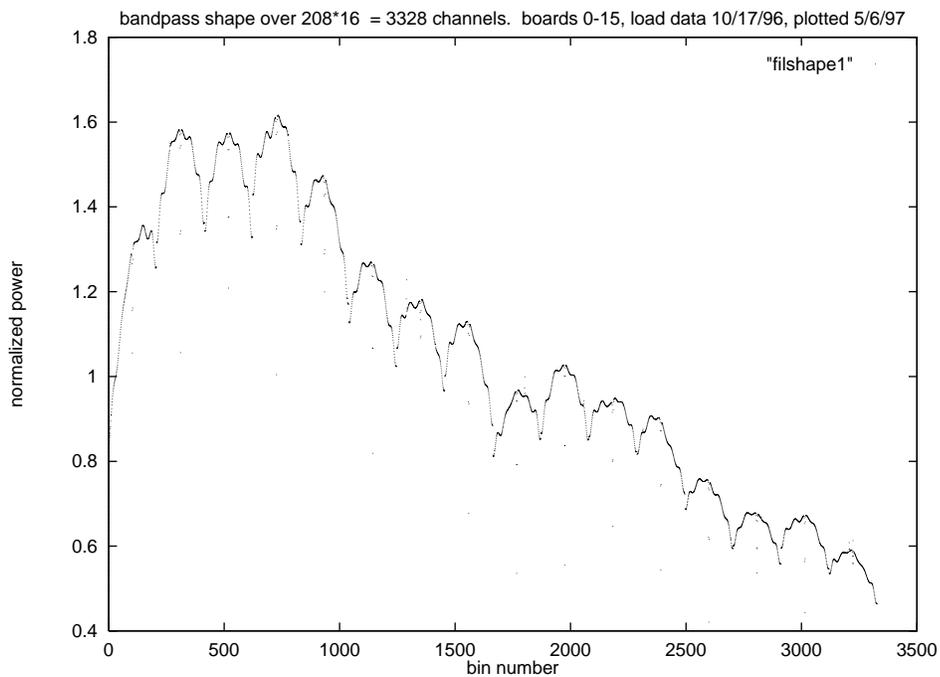

Figure 3.8: Load characterization of the channel one front end. The 16 individual board spectra are visible and show the anti-alias filter ripples, while the overall character is due to the various bandpass filters in the system.

3.5.5   Mapping bin number to frequency

The following simple formula gives the center frequency of a bin in MHz, as a function of board number, B (0–15) and bin number within a board, A (0–255). It assumes, as is the case in this experiment, that the first LO is set to 176.0 MHz, that the bank of sixteen second local oscillators is spaced at 2.0 MHz intervals starting at 44.0 MHz and that the FFT runs off a digital clock of frequency 39.384615 MHz.

$$\nu_\circ(A, B) = 0.0096153846 \times (A - 127.5) + 2.0 \times B + 220 \qquad (3.16)$$

3.5.6   Reduction procedure

Avruch has developed a pipelined data reduction procedure. One data tape, containing approximately 13 days worth of data, is loaded to disk and reduced at a time. Only data taken between 1 am and 5 am AST are reduced. By reference to the email status reports, our experiment log, and table 3.2, data files corresponding to days when known problems have been logged are deleted. The remaining data files are processed by the reduction program, which produces as output maps of signal intensity with coordinates of frequency and local mean sidereal time (LMST). The matched filtering to be discussed in §3.6 is applied to these maps to try and extract the expected signal from the noise field.

3.5.7   Preliminary results and pointing of the sky beam

Figure 3.9 shows the increase in sky background noise due to the galactic synchrotron background as our beam drifts through the Milky Way. Each plot is an estimate of the time evolution of total power from a single spectrometer board. An RFI-robust estimate of power is obtained by sorting each spectrum, excising outliers, and averaging the remaining data. Overplotted is the expected curve from a model of our galaxy derived from two all-sky surveys. We see that the fit is particularly good when the galactic background is strong, but gets noisier as the sky beam drifts off
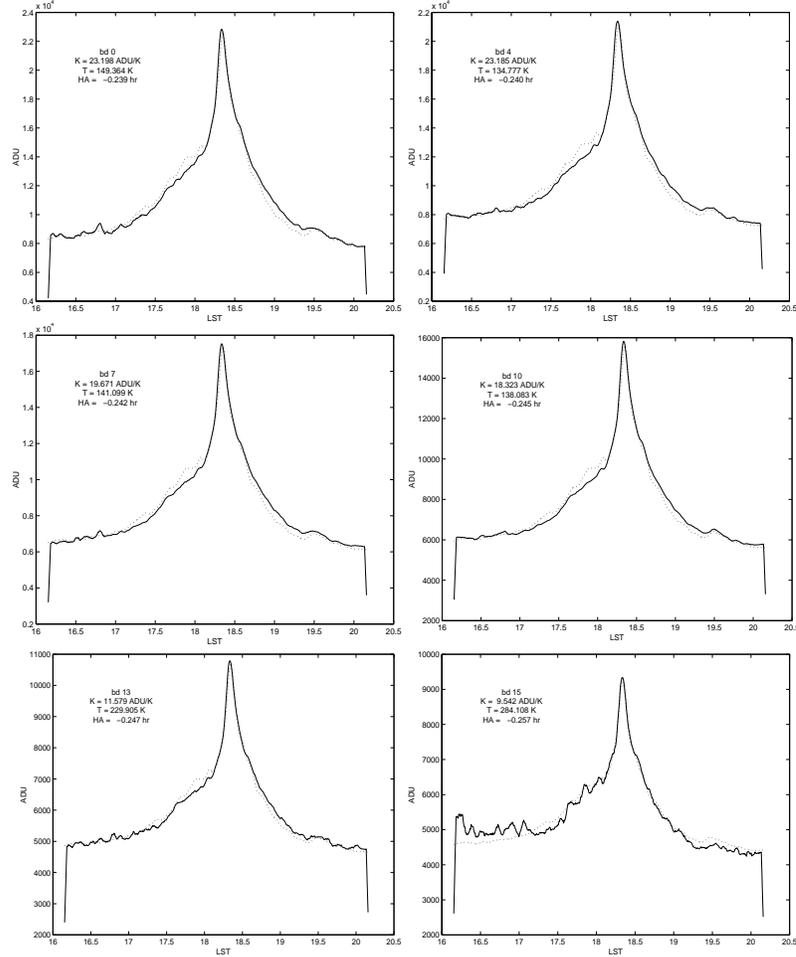
Figure 3.9: The background temperature of the Milky Way as detected in the continuum on various boards. Solid line is our measurement, dotted line is a model. Plot is courtesy of I. Avruch [50].

the galactic center. In these plots, a fit is made to all-sky maps in order to calibrate the second reduction.

Unfortunately, when pointed off the Galactic center, the total power plots are not as well behaved. Figure 3.10 shows continuum power over the same LMST range for ten consecutive days. Similar steps to those used in the "on Galaxy" plots are taken to excise outliers. However there are significant differences in the plots from day to day, showing that they are not an accurate representation of the true sky background. The fluctuations are broadband, since (although these results are not presented), the
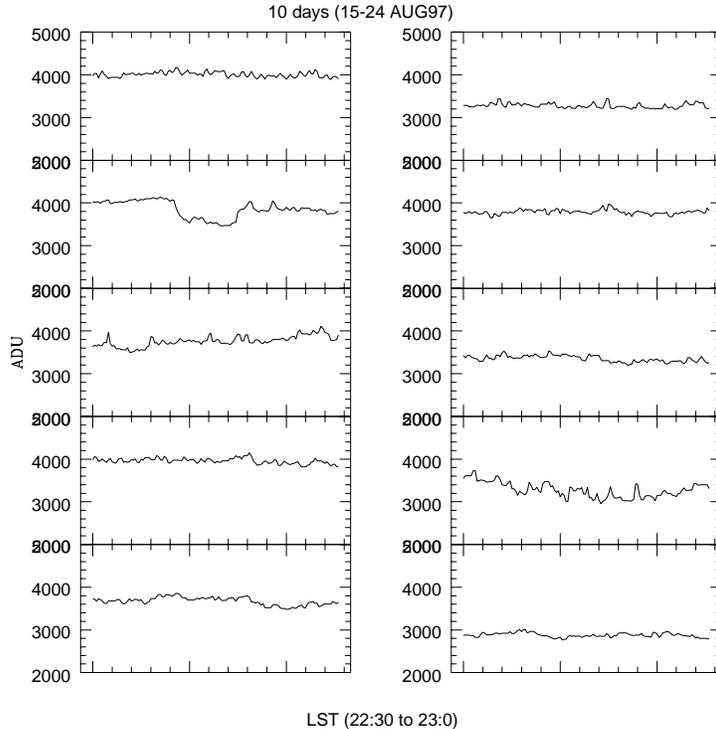
Figure 3.10: Continuum power over the same LMST range on ten consecutive days. Plot is courtesy of I. Avruch [50].

same shapes are duplicated from board to board. The most likely source of these fluctuations is some sort of broadband interference. Gain fluctuations are thought to be less likely explanation, since they are multiplicative rather than additive, and so would be expected to show up in the on-Galaxy drift scans as well.

Confirming our model of beam pointing, figure 3.11 represents 28 seconds of data centered on 10:49:35 UTC on MJD 50254. The observation is by Ingrid Stairs and Joseph Taylor using our experiment's feed and RF system as a test signal input to a new pulsar timing instrument, the Princeton Mark IV. The pulsar has drifting sub-pulses, which explains the "multiple-spikes" appearance of the peak [76].

The galaxy and pulsar observations, along with calculations based on the feed geometry are all consistent with a feed pointing of declination $\delta = -7°$ and hour angle $HA = -16$ minutes.
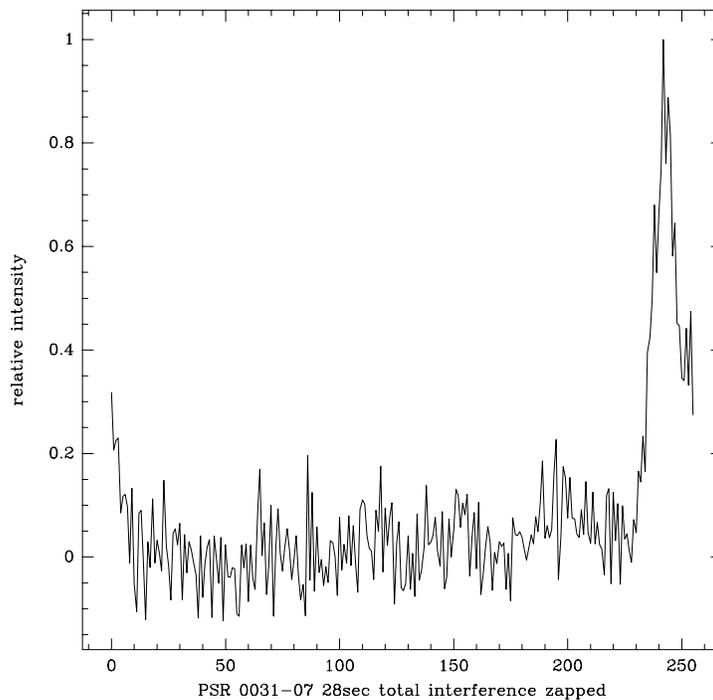
100

Figure 3.11: Pulsar $0031 - 07$, detected using our feed and RF by a group from Princeton. Plot is courtesy of I. Stairs [76].

3.6 Matched filter correlation processing for signal detection

The expected signal has structure in both time and frequency. We use a matched filter technique to extract the protocluster signal from the noise. This section describes the techniques we developed to pull out the expected signal, while rejecting as far as possible the remaining stationary RFI. The author developed the methods discussed in this section.

3.6.1 Simulated signal in simulated Gaussian noise

To start our explorations, we tried a simple matched filter simulation in Matlab. We computer generated a signal with a character similar to that of the expected

signal, and hid it in computer generated Gaussian noise. The output of the co-adding process are data maps with coordinates of time and frequency, and the signal has a characteristic width in each of these dimensions. In the initial simulation we use for a test signal a two dimensional Gaussian bump, which is plotted in a 256 by 256 point array. We add Gaussian pseudo-random numbers to each element of the array. The peak of the bump is 0.5, while the Gaussian distributed noise has unit standard deviation, and hence unit variance.

The matched filter processing involves cross-correlating the signal plus noise field with a template the same shape as the signal. We shift the template to the center of the noise field and normalize it, so that the position of the signal in the filtered output, and its amplitude, are unchanged. The implementation of the matched filter correlation uses our old friend the FFT again—we apply a 2-dimensional FFT to both the image and the template, multiply (this is not matrix multiplication, but is element by element) the one array by the complex conjugate of the other, and apply an inverse transform. Apart from slight differences at the boundaries this is equivalent, by the convolution theorem, to direct cross-correlation; however this process is considerably faster than direct cross-correlation. Figure 3.12 shows the signal and the template.
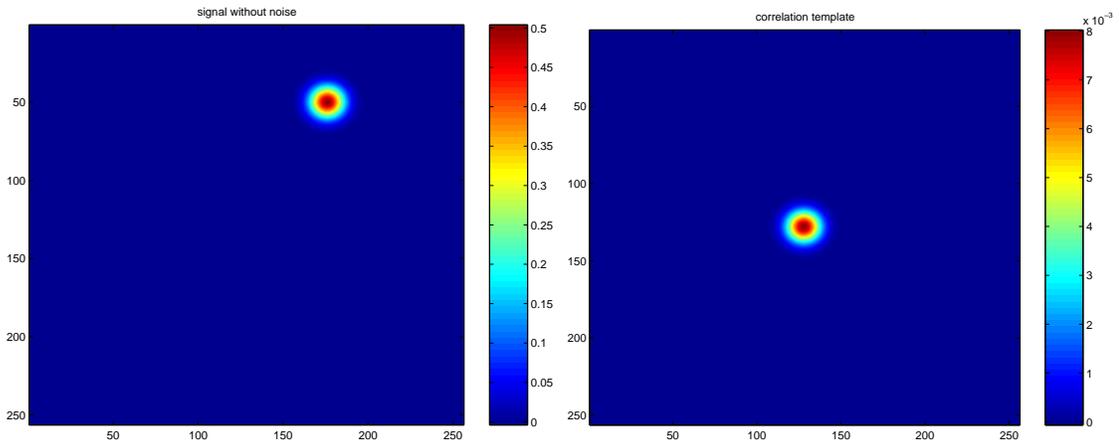
(a) simulated signal

(b) correlation template

Figure 3.12: Simulated signal, a two dimensional Gaussian bump, with a peak of 0.5, and the correlation template, which is the same shape as the original signal but shifted to the center of the field, and normalized.
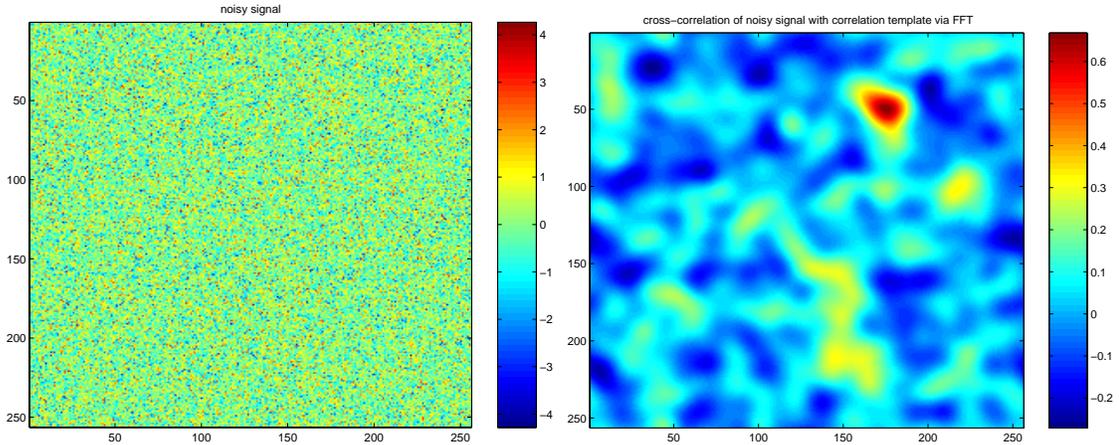
(a) simulated signal hidden in noise

(b) cross correlation with template

Figure 3.13: Signal of peak amplitude 0.5 with Gaussian noise of unit variance added. Signal to noise ratio (peak-to-rms) is thus -3 dB. Right hand plot is the cross-correlation of the noisy field with the template.

The results, shown in figure 3.13 are encouraging. The left hand plot shows that the signal is almost indistinguishable in the Gaussian noise field, while cross-correlating draws it out quite clearly, as shown in the right hand plot. Matched filtering in a Gaussian noise field is a very well behaved problem, and pulling a signal out of a more realistic noise field is expected to be much more challenging.

3.6.2   Simulated signal in a real noise field

We then extended our simulation to use real data for the noise field. We took one of our frequency-time plots from the co-adding process. and dropped in a simulated signal, with structure in time and frequency. Once again we use a 2-dimensional Gaussian bump, this time with parameters consistent with the expected signal—a 1 MHz linewidth and drift time of three minutes. Figures 3.14 and 3.15 shows the simulated signal alone (unit amplitude), and that signal, scaled to an amplitude of 0.005, and added to a real image. The signal is invisible in the unfiltered plot. There is also a vertical band in the middle of the plot. This is an artifact, a result of creating a one hour LMST plot from two half hour plots output from the co-adding procedure—since our calibration procedure is not particularly robust it is not too surprising that see differences in the intensity levels in the two plots.

The stationary RFI shows up as the horizontal lines in the plot. It is particularly strong at high and low frequencies, also scattered throughout the plot with some "quiet" bands in the low-center frequencies. Even the quiet bands, however, show a higher than expected noise level. We first tried to pull the signal out with a simple cross-correlation with exactly the shape of the simulated signal, the method used in the pure simulation above. As discussed there, this method worked well with simulated Gaussian noise but the result when operating on the real data map is unimpressive (see figure 3.16). The stationary RFI lines have spread into broad bands which dominate the plot and obscure the desired signal.

We need a filter that is well matched to the signal, but that is insensitive to the horizontal stationary RFI bands. We made such a filter by modifying the Fourier transform of the cross correlation template to remove the constant components in

the "transformed time" dimension. The cross-correlation is performed in the Fourier domain in any case, so this modification is easily implemented by zeroing out the appropriate spectral components in the template transform before multiplying. The result of the cross correlation is shown in figure 3.17. It can be seen that the effect of the horizontal lines have been reduced and the signal stands out quite clearly. There are some artifacts too; in particular, the vertical gain artifact has not been removed, since we have not removed constant components in the transformed frequency domain.

Part of the problem with the matched frequency approach is that it forces all features in the reduced maps to look rather like the signal we are looking for. As a result there are numerous spurious objects with the appropriate structure in time and frequency. However these are concentrated in the region of the strong RFI lines. We veto them when scrutinizing the filtered maps noting that these are co-linear with other artifacts, and, by referencing the original unfiltered plot, we note that they arise in severely RFI corrupted areas. They arise because the RFI is somewhat time-variable. We have only filtered out the DC part leaving fluctuations which are on average greater than the fluctuations in quiet regions of the spectrum. The simulated signal stands out because it is not co-linear with other like objects. In general, isolated features are the most interesting.

As a further refinement, we next take out the constant component in the frequency direction. The tradeoff is that any changes to the matched filter to improve its noise rejection characteristics also degrade the match to the signal. Figure 3.18 the constant component has been removed in both dimensions. The vertical artifact is gone. We consider the performance of this matched filter to be quite good, and we decided to use it to process all of our maps.

Note that all processing is done in the Fourier domain mostly because it's much faster than direct cross-correlation, and also because it's easier sometimes to think of the processing as a multiplication in the Fourier domain which excludes the constant components, rather than a cross-correlation on the direct image. However some insight can be gained as to how the correlation process works by looking at the inverse transform of the modified filter function. This correlation template is presented in figure 3.19. For the case of both constant components removed the correlation template

now has valleys, vertical and horizontal, about the width of the Gaussian bump in the middle and getting shallower towards the edge of the image. These depressions cause features that are constant in time or frequency to be subtracted out when the correlation is performed.

We have analyzed our maps with this stationary RFI rejecting cross-correlation procedure. The resultant maps, and the raw co-added maps from which they are derived, are presented in appendix A. We have scrutinized the maps, and discuss our conclusions in the next, and final, chapter.
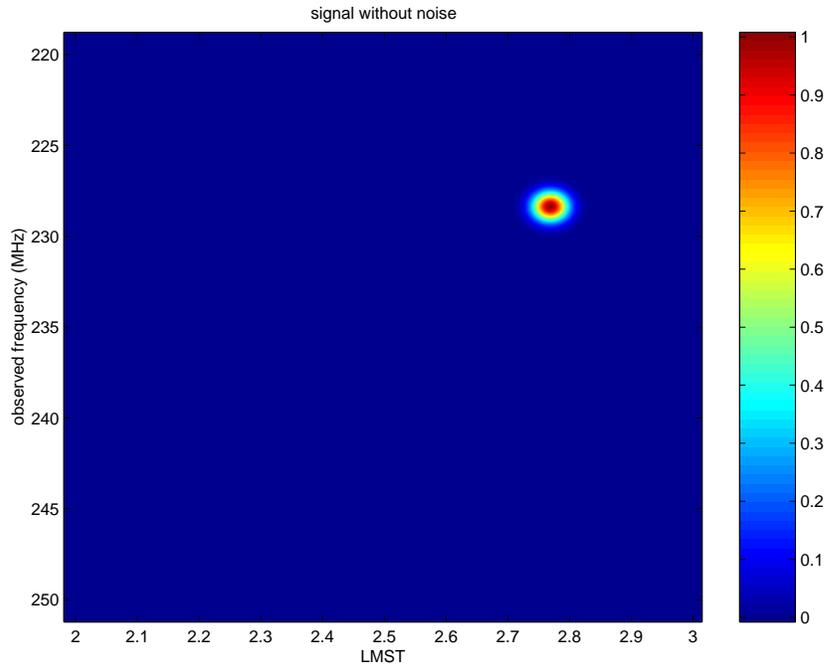
Figure 3.14: Simulated signal with 3 minute drift time, and 1 MHz bandwidth.
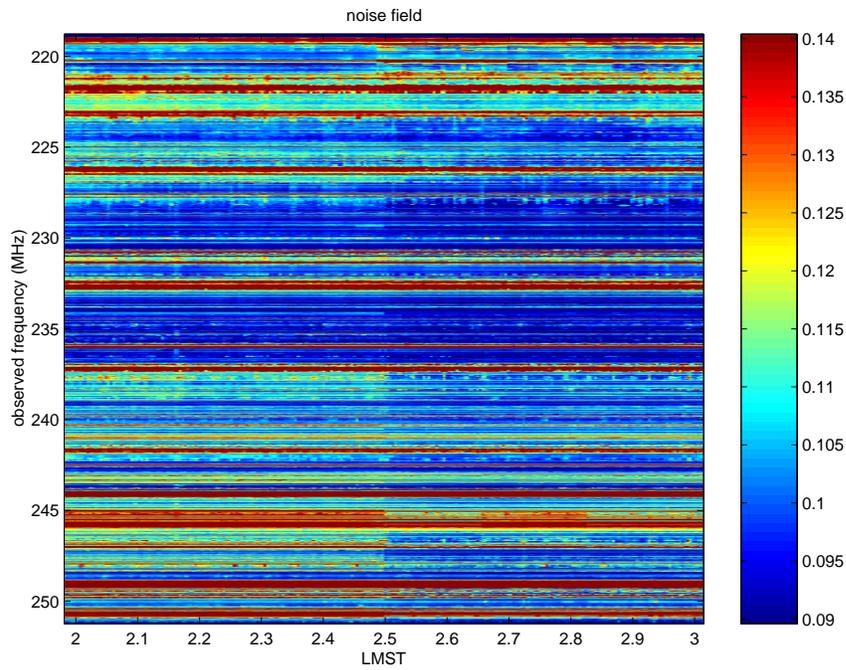
Figure 3.15: The simulated signal, (amplitude 0.005), hidden in a real data map.
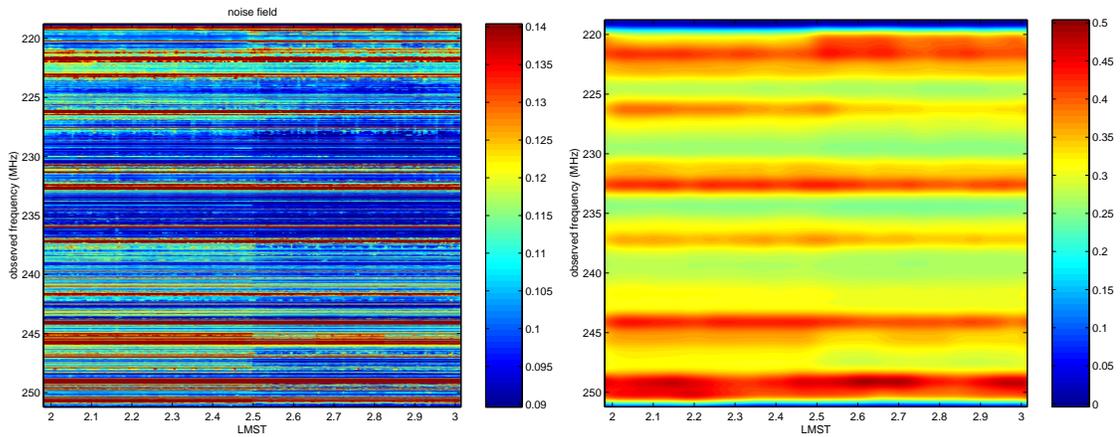
(a) signal hidden in a real noise field   (b) cross correlation with Gaussian

Figure 3.16: Simulated signal hidden in real image, and first attempt at matched filter processing.
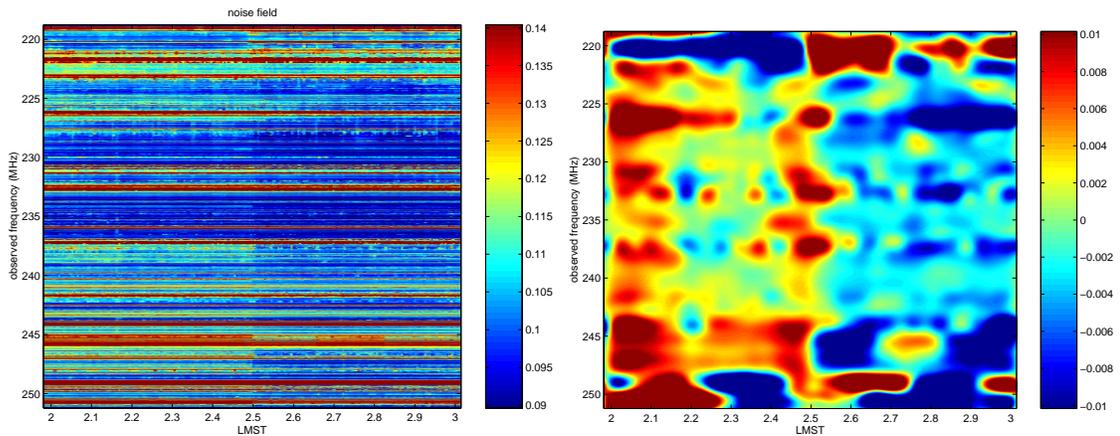
(a) signal hidden in a real noise field   (b) cross correlation with modified Gaussian

Figure 3.17: Simulated signal hidden in real image, and second attempt at matched filter processing, taking out dc in one dimension only.

Figure 3.18: The filtered image, constant components in both dimensions rejected.

Figure 3.19: The RFI rejecting correlation template.

109

Chapter 4

Results and conclusions

In this chapter we examine the images output from the co-adding program, and tabulate candidate detections. We confirm or veto these by looking at maps derived from subsets of the data. We relate the results to the underlying cosmology. We conclude with some suggestions for improving the experimental setup, and the data analysis.
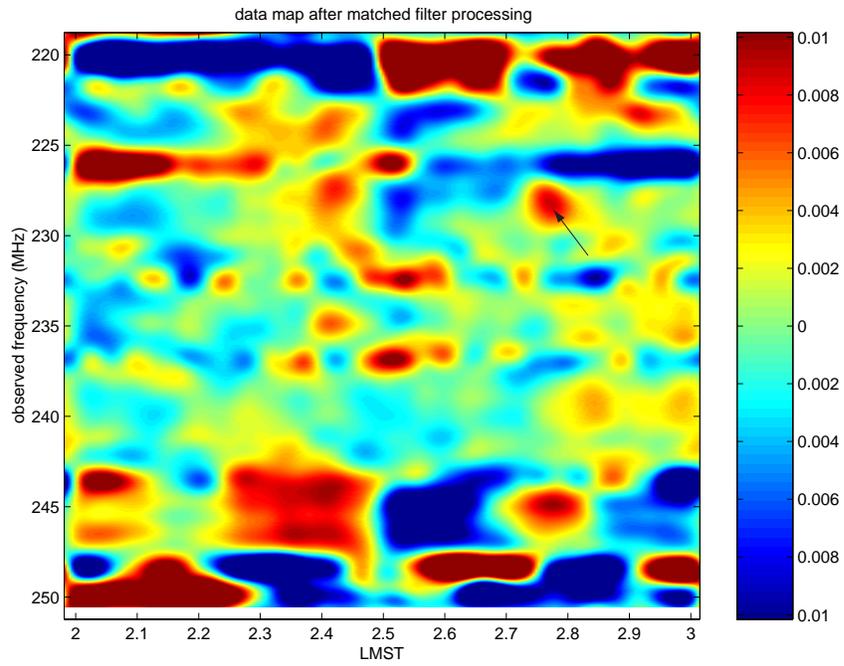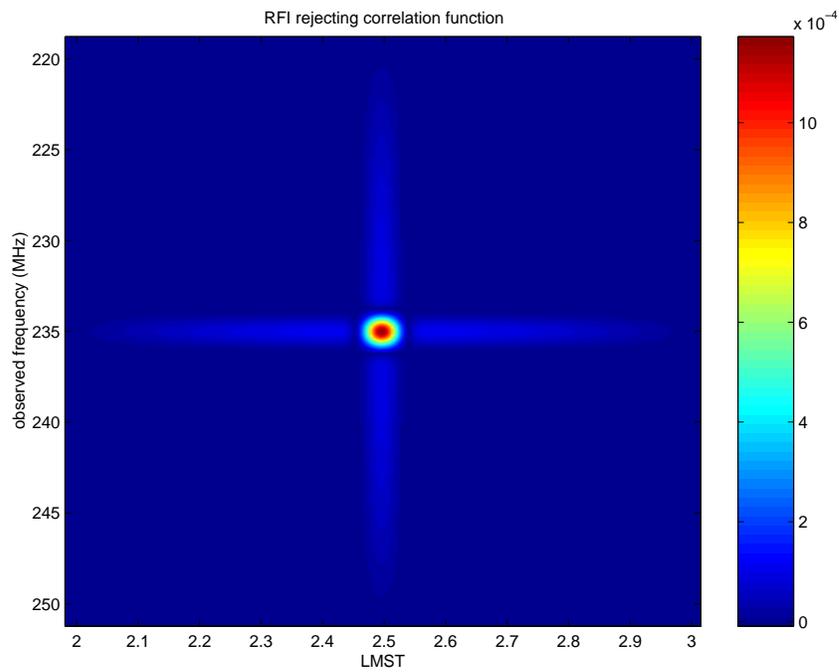
4.1   Analysis of filtered data maps

The unfiltered and filtered maps from the first reduction are presented, side by side, in appendix A. The maps are calibrated in units of the calibration diode power. The calibration diode temperature as a function of frequency has already been presented in figure 2.16. These curves are used to express the maps in absolute temperature.

The matched filter approach tends to force all features in the reduced maps to look like the expected signal. Because the maps are rich with RFI, there are numerous spurious objects with the appropriate structure in time and frequency. However these tend to be concentrated in the region of the strong RFI lines. We veto them when scrutinizing the filtered maps, noting that they are co-linear with other similar artifacts. Also, if we reference the original unfiltered plot, we note that they arise mostly in severely RFI corrupted areas. Because the RFI is somewhat time-variable,

an area with lots of RFI is more likely to generate spurious features. Our matched filter takes out the constant part of the RFI bands leaving fluctuations which have a greater variance than the fluctuations in quiet regions of the spectrum.

4.1.1 Sensitivity threshold

A direct way to compute the noise level in the maps is by selecting a clean section of the map and computing the rms in that region. Then, using the calibration data, this is converted to Kelvin. The RMS noise in the quietest regions of the map is computed in this way to be approximately 15 K. This is rather a high number, three orders of magnitude larger than the expected signal, which, as estimated in §1.6, is about 12 mK. Clearly the RFI is dominating the noise in the system. Not only is the RFI limited system temperature much higher than that expected from thermal noise, but the RFI, being non-Gaussian, does not decrease in amplitude as fast as $\sqrt{\tau}$ (as would be the case for Gaussian noise).

This approach to estimating the noise does not account for the processing gain of the matched filtering. To account for this we use a "test signal" method to estimate sensitivity. The simulated signal added and extracted in figure 3.18 had an amplitude of 0.005. Given where it was positioned in the map at about 228 MHz we note that the diode contributes about 3100 K to the system temperature at that frequency. This means that the contribution of the simulated signal to the system temperature is about 15 K. However, in figure 3.18, the SNR of the test signal in the filtered plot is much better than would be needed for a detection. We were able to lower the peak amplitude to 0.0008, corresponding to about 2.5 K, and still just barely detect the signal. This suggests that 2.5 K is a fair estimate of the threshold sensitivity for detection in RFI quiet regions.

4.1.2 Identification of candidate signals

Table 4.1 lists the coordinates of interesting objects extracted through scrutiny by the author of the filtered maps. This scrutiny is somewhat subjective, since there

is structure at low levels all over the maps. We use the unfiltered maps to rule out candidate objects which happen to coincide with strong RFI lines, since we expect the higher variance to cause spurious detections in these regions. The maps on the edges of the sidereal time range processed are noisier than those towards the center. This is because fewer data points went into producing them. This "edge effect' occurs because we only process data from 1 am to 5 am solar time, and sidereal time drifts with respect to solar time. The candidate signals are drawn from the quieter maps towards the middle of the time range.

The coordinates of an object are its frequency and sidereal time. Using the map's color scale and the calibration diode measurement, we estimate the peak antenna temperature due to each object and note it in the table.

number	frequency (MHz)	LMST	T (K)	comment
1	229	1.65	15	diffuse, paired with 2
2	229	1.78	15	diffuse, paired with 1
3	228	2.44	18	diffuse
4	235	2.50	13	compact
5	229	2.78	9	diffuse
6	229	3.48	25	somewhat diffuse
7	229.5	4.4	15	compact

Table 4.1: Coordinates of candidate celestial objects.

4.1.3   Confirmation or veto using partial data maps

An intermediate data file is written for each tape processed. To confirm or veto a given detection we go back and look at these intermediate results, which we call "partial data" maps. Each of these three maps contains approximately one third of the data. The "partial data" maps are also presented in the appendix.

If an object is truly celestial it should appear at the same sidereal time equally in all of the maps, and there are typically partial maps in which the candidate does not

appear at all. None of our candidate signals pass this simple test; all of the signals are seen to appear very strongly in one of the maps alone. Not surprisingly, considering the high level of RFI in our data, our experiment has resulted in a non-detection.

Note that I have restricted my attention in this thesis to the first reduction, which used the calibration diode. Recent work has led us to conclude that the data from the second reduction seems to be considerably quieter (about a factor of 5 in temperature) than the data from the first reduction. There are three possible reasons for this: first, we used the calibration diode in the first reduction, and its misfiring may have introduced noisy artifacts into the first reduction maps, and that these exceed the fluctuations due to gain variations alone; second, more data is co-added in the second reduction, since we have access to data taken after the calibration diode was shut down; and third, the combination of median filtering and averaging used in the second reduction may have benefited the SNR. Continuing work will focus on the second reduction, as we consider it to be the better method. Ian Avruch has analysed data from the second reduction in his thesis [50]. I present a small sample of data from the second reduction in appendix B.

4.2   Limits on protoclusters

The calculations of §1.6 show that a protocluster containing $10^{14} M_\odot$ of neutral atomic hydrogen (with a velocity dispersion scaled to $\sim 1000 \ \mathrm{km \, s^{-1}}$, corresponding to the bandwidth of our matched filter) would produce an excess antenna temperature of about 1.1 mK. This is a very large protocluster, and the sensitivity level in the data maps presented here is a much larger 2.5 K. The best we can do is set a rather weak limit. The antenna temperature due to a cloud is directly proportional to the mass of neutral hydrogen in the cloud. Thus the smallest cloud we can detect is ~ 2000 times larger than this. Our non-detection suggests that there are no HI clouds more massive than $2 \times 10^{17} M_\odot$ in the volume of space we searched.

Note that the maps were scrutinized with a view to finding positive bumps, indicating emission lines. Note, however, that the matched filter does not reject absorption lines.

It would produce a valley in response to an absorption feature drifting through the beam. Emission features correspond to clouds with a spin temperature exceeding the CMB temperature, while absorption features are expected from clouds which are cooler than the CMB at the observed redshift.

4.3 Suggestions for a future experiments

We present here some suggestions for improving the experimental equipment and software. These suggestions are for the benefit of intrepid souls considering repeating this experiment.

4.3.1 Optics

Our 9 GHz models suggest that there may be some small benefit in implementing a new feed design. However, given the marginal improvement it is still hard to justify building a significantly more complicated unit. The main point to be made here is that in RFI limited conditions it is probably better to optimize the raw sensitivity, rather than the SEFD. A reworking of the calculations in chapter §2.1.2 should yield the appropriate feed positioning parameters.

4.3.2 Receiver

The unreliable calibration diode has been a real headache. When installed it seemed to be working fine, and it was only in later scrutiny of the data that we noticed the intermittent firing. The problem could be due to something as simple as an unreliable power connector. Power is routed to the calibration diode via a BNC connector, which is not at all weatherproof, and may have failed in the humid environment at Arecibo.

In designing the calibration diode circuit again it would be nice to use a low loss RF switch to replace the antenna by the diode during firing. Part of the problem we have had is the limited infrastructure for communicating between the control room

and the upstairs electronics rack. We are limited to two cables for two channels of RF output, shared for power. Short of laying in additional cables we would have to design some kind of multiplexing scheme to control the switch in the upstairs box.

Active RFI suppression

A group under Richard Bradley at NRAO, Charlottesville, is currently working on adaptive RFI suppression [77]. The concept involves using FIR filters with an RFI monitor antenna as input to a digital signal processor (DSP) based adaptive controller. The controller modifies the FIR coefficients to cancel RFI over a fairly wide band. They achieve about 40 dB suppression over a band of about 1 MHz in an early working prototype. Early in this project we experimented informally with passive RFI suppression, using a probe antenna and a "trombone" phase shift transmission line to couple a 180° phase shifted version of an interfering carrier back into the receiver. We achieved significant cancellation of the VHF television channel 13. Unfortunately the cancellation was narrowband—we could choose to cancel audio or video, but not both simultaneously. Given the difficulty of suppressing RFI over a broad band with a simple method, it seems that a broadband cancellation scheme similar to the one being developed by NRAO would be extremely useful. However this is a hardware intensive developmental system, and does not represent a quick fix for RFI problems.

4.3.3 Spectrometer

One unfortunate property of the spectrometer board is that when signals exceed the available range, they wrap around, rather than saturate with a flat top. Effectively we lose the most significant bit in the data. If the spectra are well behaved (that is, smoothly varying) this is fairly straightforward to correct in software. Unfortunately when the spectra are corrupted with spiky RFI it is much more difficult to detect the overflow, since strong carriers can wrap around multiple times. The confusion of points all over the spectrum make it difficult to write an algorithm to track the wraps and thus re-constitute the spectrum. Wraparound could be handled in hardware by having the spectrometer board pick up activity on the overflow bits, and saturate the

output at its maximum possible value when they are active, thus relieving the data processing software (and the programmer) of a rather tedious task.

The spectrometer board was designed in 1993, using an FFT ASIC which was selected from the best devices available in 1991. We are thus using an instrument designed with eight year old technology. With the rapid advances in integrated circuit technology, which have continued unabated through this decade, it is likely that different approaches to the design would suggest themselves, were we to design the instrument again. A design using digital signal processor (DSP) technology may well prove to be a better approach if we were to start again with a clean slate. A DSP approach would most probably be more software intensive, and thus would result in a more flexible design. Such alternative processor based approaches should be considered for future spectrometer designs.

4.3.4   Experiment control and status

One of the major limitations of the current software setup is the lack of a multitasking operating system. Multitasking would significantly simplify the data backup task, and, most importantly, would enable us to log in remotely to monitor both system performance, and the data, in real-time. Our budget limited us to an IBM compatible personal computer, and, at the time we set up the system, free Linux multitasking software was not as widely available, or as well tested, as it is now. We were also worried about the fact that a multitasking operating system is not a real-time system, with no guarantee of maximum latency between the request for service, and that service actually being provided.

In practice our concerns may have been unnecessary. The data in the system is asynchronously buffered in hardware such that all realistic latencies would cause no problems, particularly if one is careful about the other tasks which are allowed to run on the system. Professor Cordes' instrument, the Arecibo Observatory Fourier Transform Machine, (AOFTM) uses our spectrometer boards attached to a Unix workstation, with no apparent problems.

It would also be nice to have more real-time processing in the data logging computer. Many features come to mind, for example real time calibration, and real time RFI detection and excision.

4.3.5 Data analysis

The data analysis task was a struggle. We spent an enormous amount of time and effort trying to figure out the nature and source of artifacts in the data. Overriding everything was the RFI, both narrowband and broadband. The broadband RFI is intermittent, and is difficult to separate from other artifacts such as gain fluctuations. The improvements to the front end of the system suggested above would definitely ease the data analysis burden. More real time feedback on data quality would also be extremely helpful, since we are mostly not on the site, leaving the equipment to run autonomously.

Improvements to matched filter processing

For the matched filter processing, we used a very basic template, with the most simple possible profile of a Gaussian in both dimensions. This is probably a good approximation for the time dimension, but, from the theory, we expect there to be more complicated frequency shapes, including multi-peaked structures. Also, it is not clear what the overall bandwidth of the signal might be. It would be productive to extend the analysis to a library of possible expected signals, based on the suggestions of theorists. Our MIT collaborator, Ian Avruch, is scrutinizing the raw maps, rather than filtering, since there is less likelihood of suppressing unusually shaped signals.

A future improvement to this processing has been suggested by Dr. Michael Davis. The correlation filtering alone broadens the signal. His suggestion is to compute not just signal (the output of the matched filter) but also signal to noise ratio. To compute the SNR we subtract the matched filter result from the original data, and then compute the filter-weighted RMS at each point. A true signal has a very characteristic signal to noise ratio plot in two dimensions, which is significantly sharper than

just the signal plot. This helps to locate the signal more accurately in the data map.

4.4   Future experiments at Arecibo

Is it possible to detect protogalactic structure in HI at Arecibo, given the RFI environment? That is a difficult question to answer with certainty, since little is known observationally, or even theoretically, about the nature of this structure. We would not want to discourage future attempts, since a successful detection would be a very significant result. Indeed, future experiments may make use of the equipment we built, which is well tuned to the experiment, and is staying at Arecibo. And the idea of running a background experiment is also fundamentally sound. However experimenters should survey the RFI in detail, and have a clear idea of how bad the situation is. We suggest that they have definite plans for dealing with it before starting to take data. It should also be noted that we have observed the RFI getting stronger during the course of our experiments, although the situation seems to have stabilized.

It is perhaps possible that improved algorithms applied to reanalysis of the existing data may help, but we believe that, given the sensitive nature of the measurement, substantive improvements can only be gained through improvements to the raw data quality. Extending the system to a two beam configuration, as originally planned for this experiment, may have some merit. Certainly this would double the amount of time "on source" but the real benefit would accrue through rejecting interference. Given the very weak signals, the algorithms which would be applied to reject RFI using two beams should also be thought through at the outset. We would recommend a carefully designed and tested front end calibration scheme, so that gain fluctuations are removed, or at least easily distinguished, from other artifacts. And the baseline ripples should be investigated, understood, and if possible, removed. If the resources are available this is just the sort of experiment which could benefit from an active RFI cancelling such as that discussed in §4.3.2.

4.5   Closing remarks

This experiment was based on an extremely elegant idea. By setting up a system to run in the background of the largest radio telescope in the world, we were able to achieve indefinite amounts of observing time. This allowed us to attempt a deeper search than would have been possible in a time allocated situation. The relatively low frequency associated with the hyperfine line at high redshift allowed us to piggy-back onto the Arecibo reflector with a very simple antenna. With a very limited budget we built and installed a receiver, spectrometer and control software, all of which have performed admirably.

Unfortunately the extremely hostile RFI environment has made the data analysis an unenviable task. We have spent the bulk of our data analysis effort trying to separate out and understand the various artifacts in the data. Along the way we have gained experience in, and developed methods for, dealing with RFI corrupted data. This is, unfortunately, a skill which is likely to have increasing value in the field of radio astronomy.

Ultimately the solution to the RFI problem may depend on siting experiments in more remote sites. The ultimate would be an observatory on the far lunar surface, but in the near term the Giant Metrewave Radio Telescope (GMRT) in India is probably the most interesting instrument for future attempts at this measurement.

Appendix A

Frequency-LMST maps
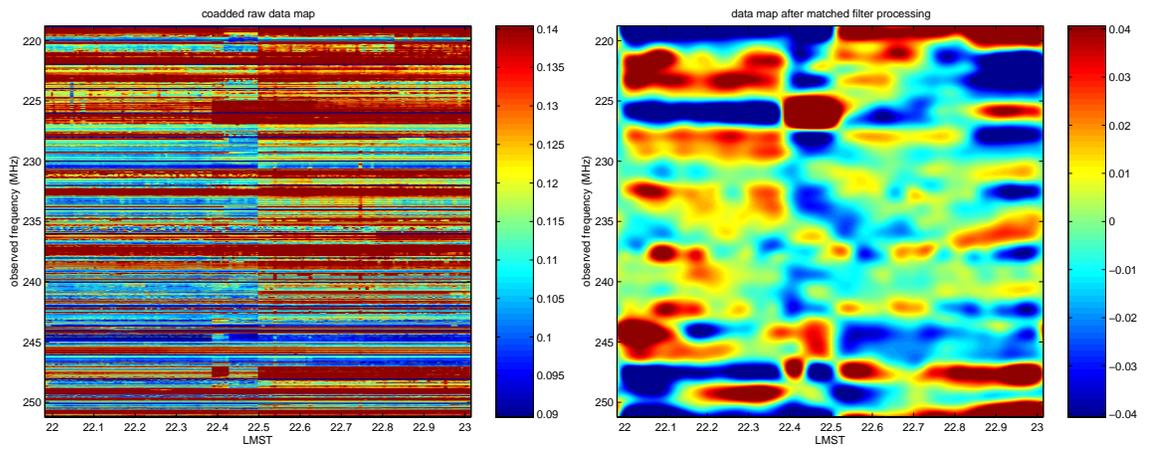
(a) raw

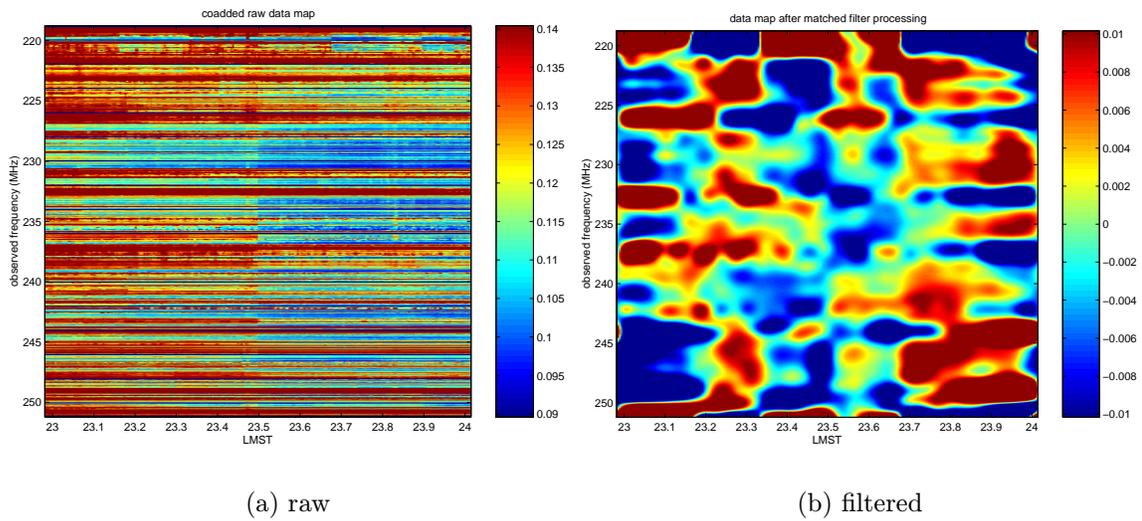(b) filtered

Figure A.1: Sidereal time 22–23 h, all data

(a) raw   (b) filtered

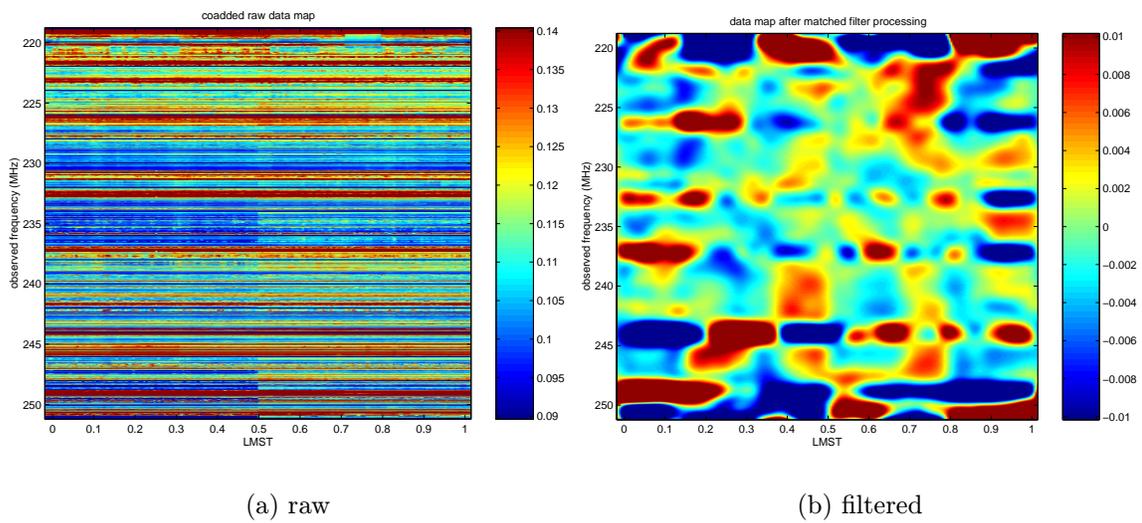Figure A.2: Sidereal time 23–00 h, all data
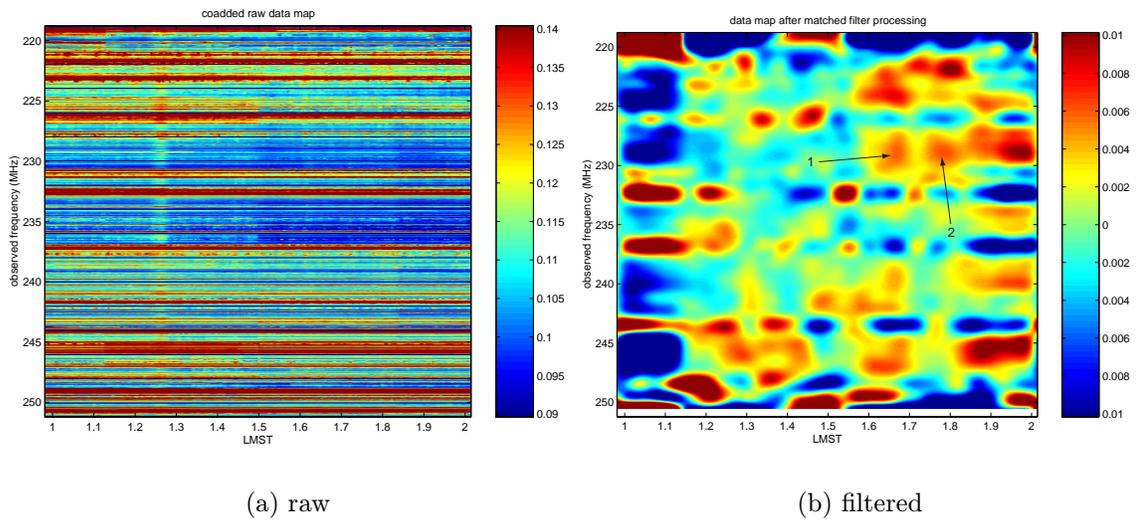

(a) raw   (b) filtered

Figure A.3: Sidereal time 00–01 h, all data

121

(a) raw   (b) filtered

Figure A.4: Sidereal time 01–02 h, all data
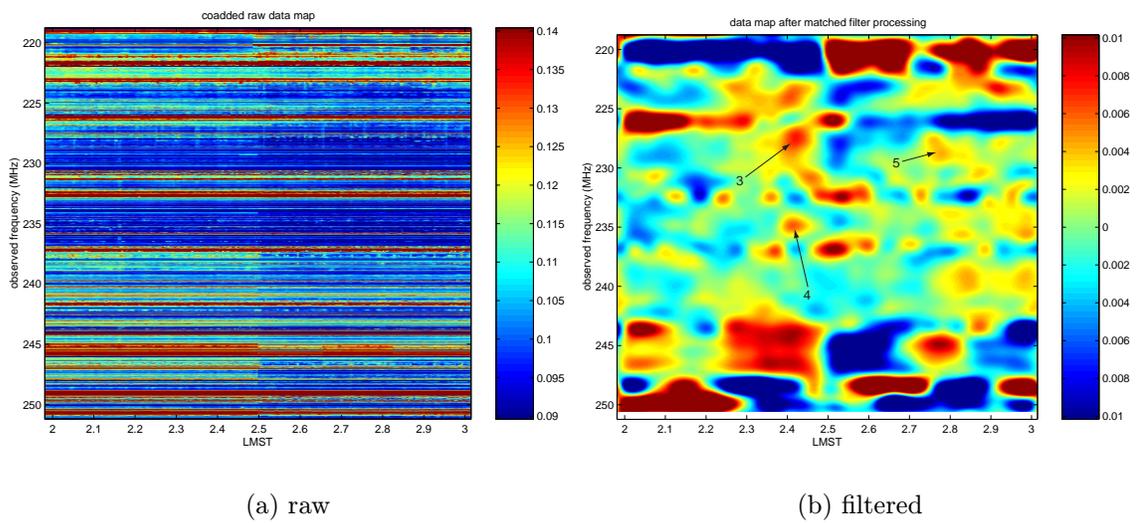
(a) raw   (b) filtered

Figure A.5: Sidereal time 02–03 h, all data
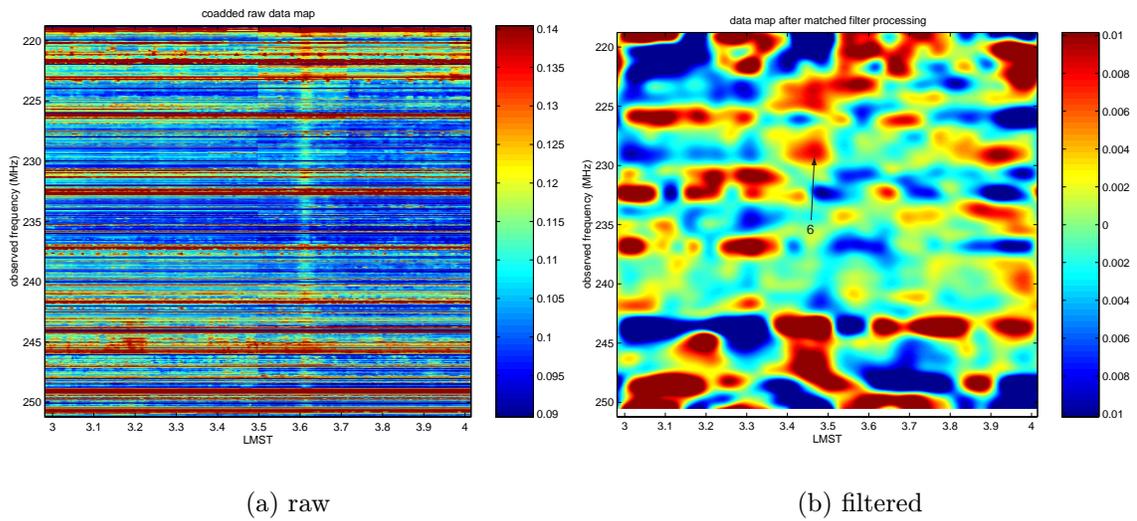
(a) raw   (b) filtered

Figure A.6: Sidereal time 03–04 h, all data
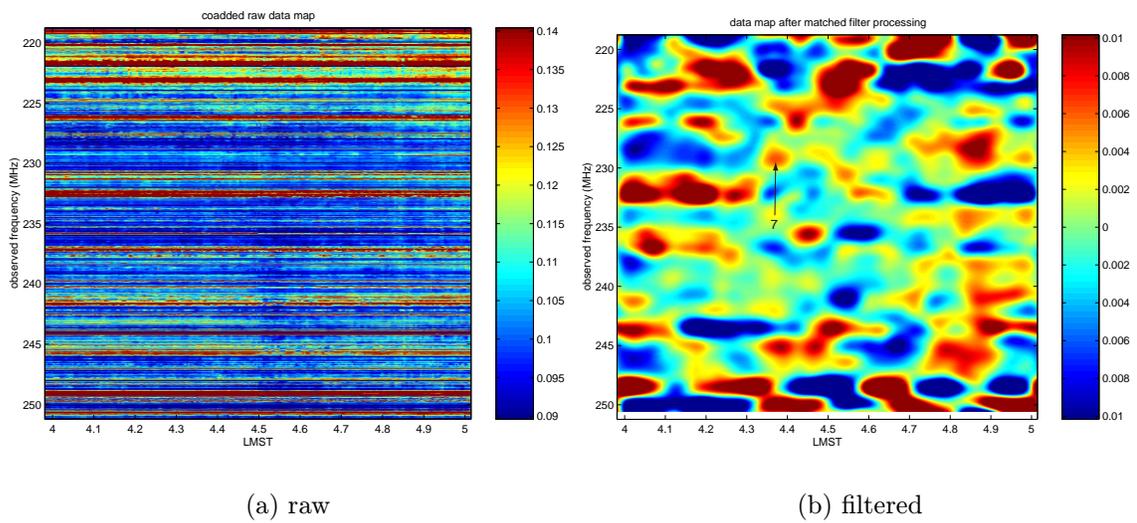
(a) raw   (b) filtered

Figure A.7: Sidereal time 04–05 h, all data
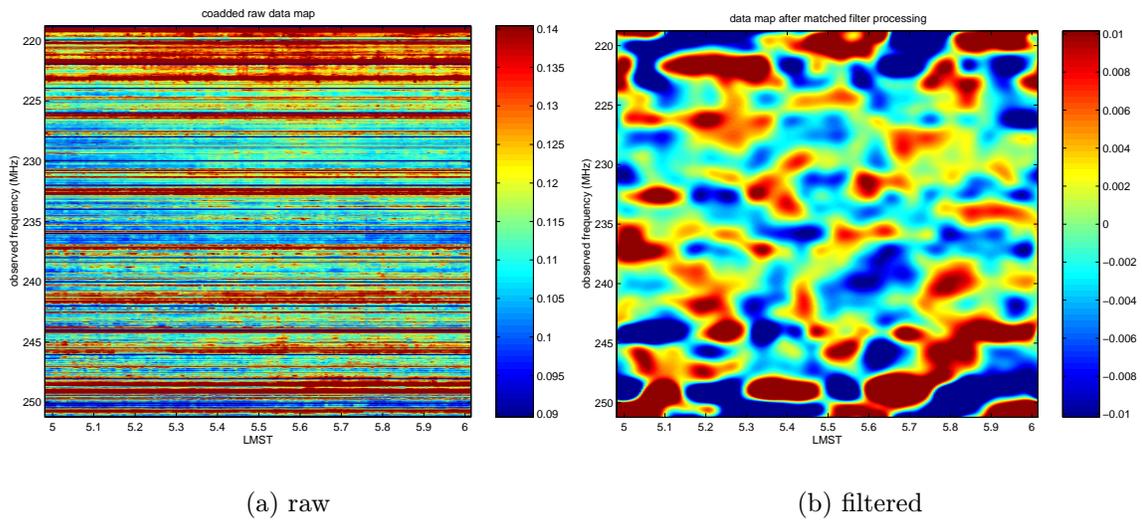
(a) raw

(b) filtered

Figure A.8: Sidereal time 05–06 h, all data
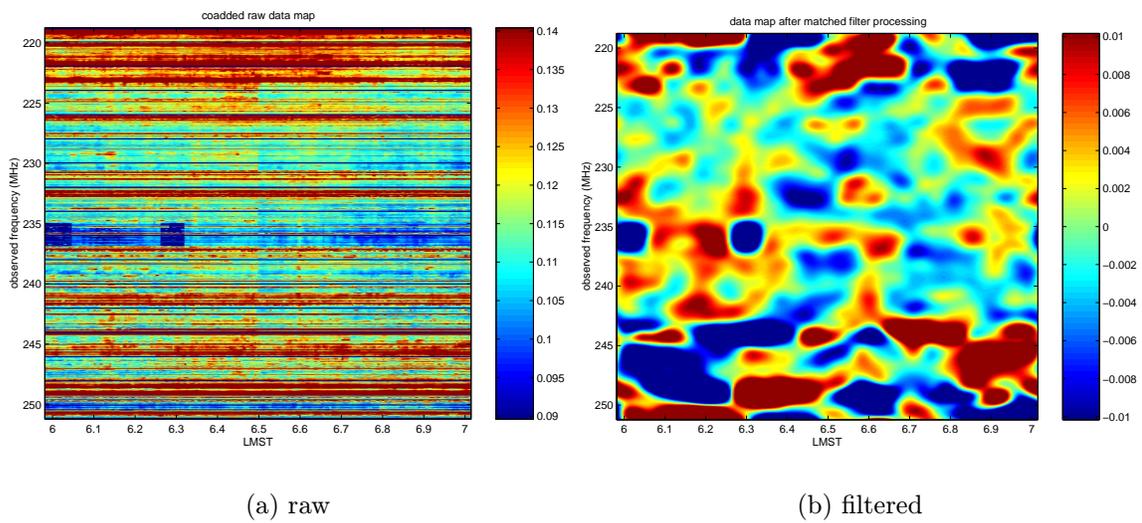
(a) raw

(b) filtered

Figure A.9: Sidereal time 06–07 h, all data

124
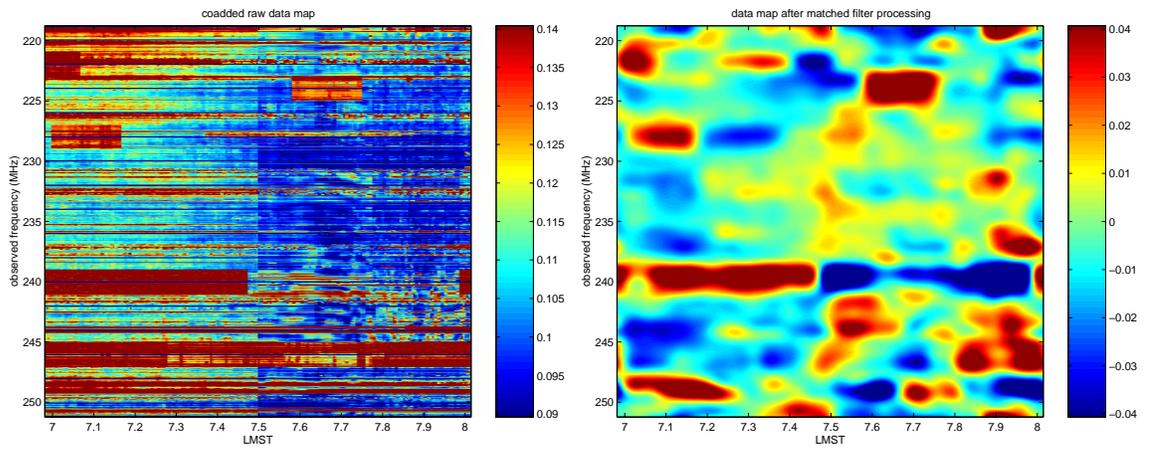
(a) raw   (b) filtered

Figure A.10: Sidereal time 07–08 h, all data
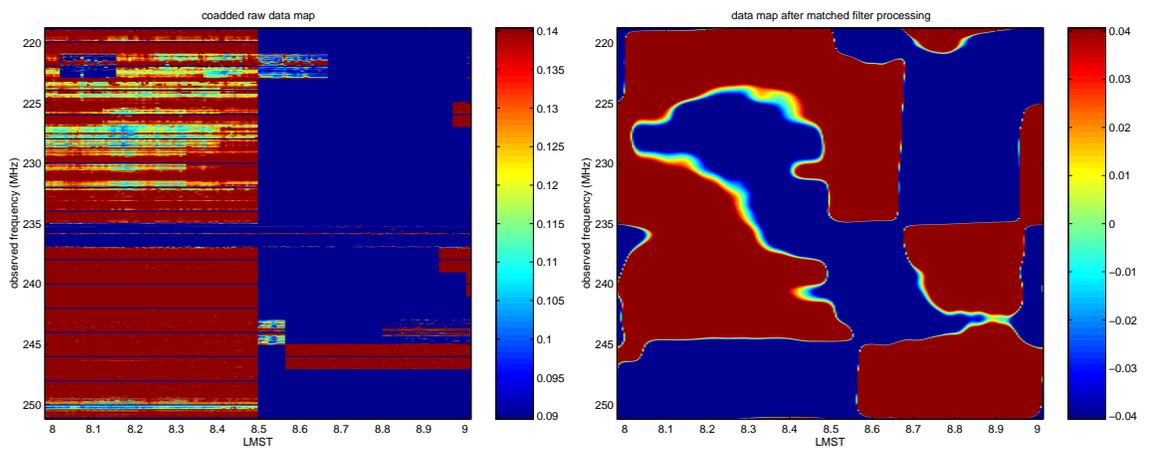
(a) raw   (b) filtered

Figure A.11: Sidereal time 08–09 h, all data
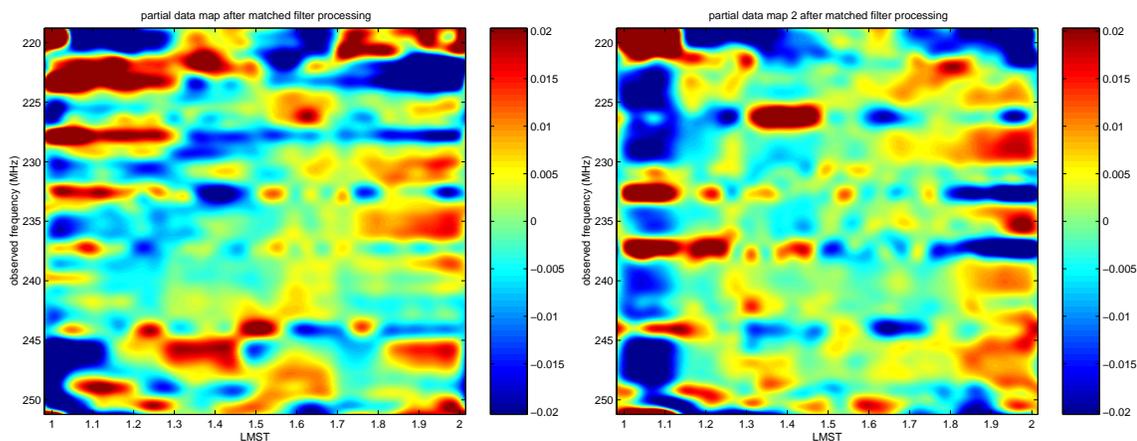
(a) part 1   (b) part 2

Figure A.12: Sidereal time 01–02 h, partial data
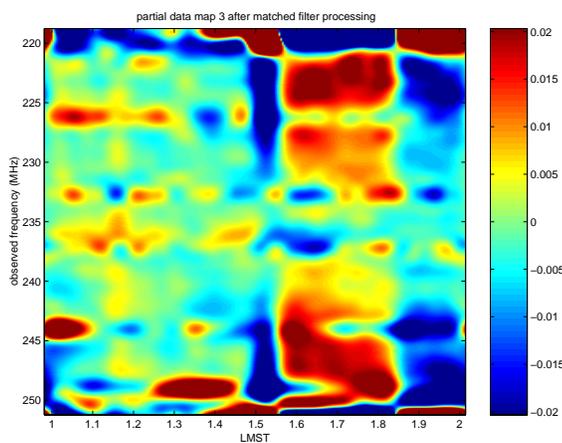
(a) part 3

Figure A.13: Sidereal time 01–02 h, partial data
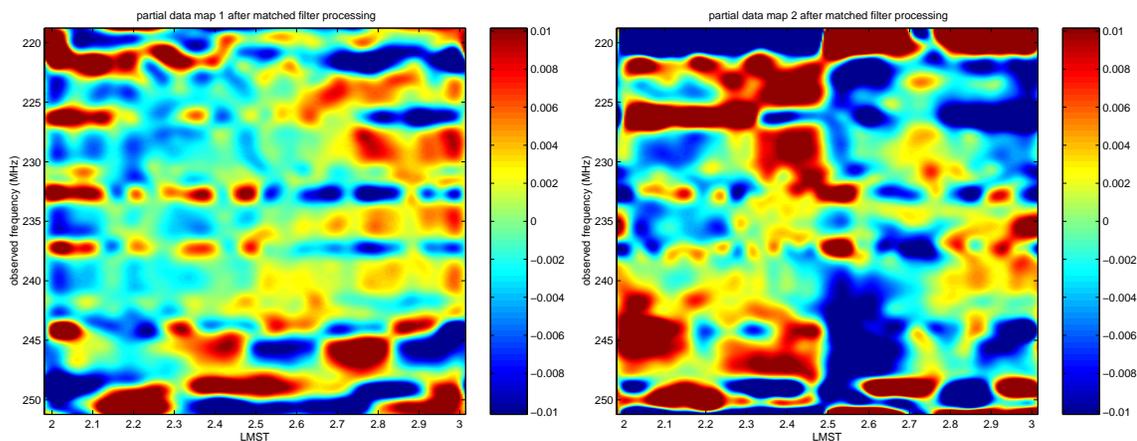
(a) part 1   (b) part 2

Figure A.14: Sidereal time 02–03 h, partial data
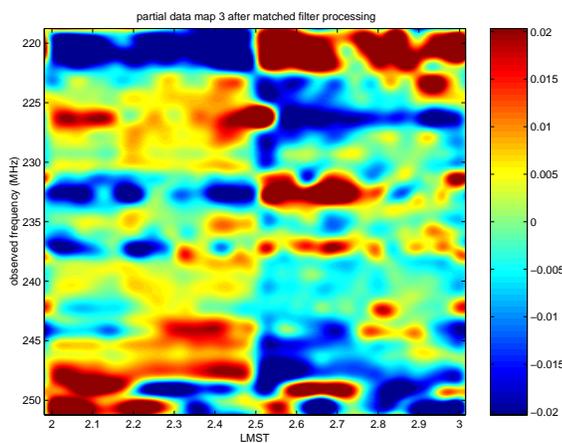
(a) part 3

Figure A.15: Sidereal time 02–03 h, partial data

127
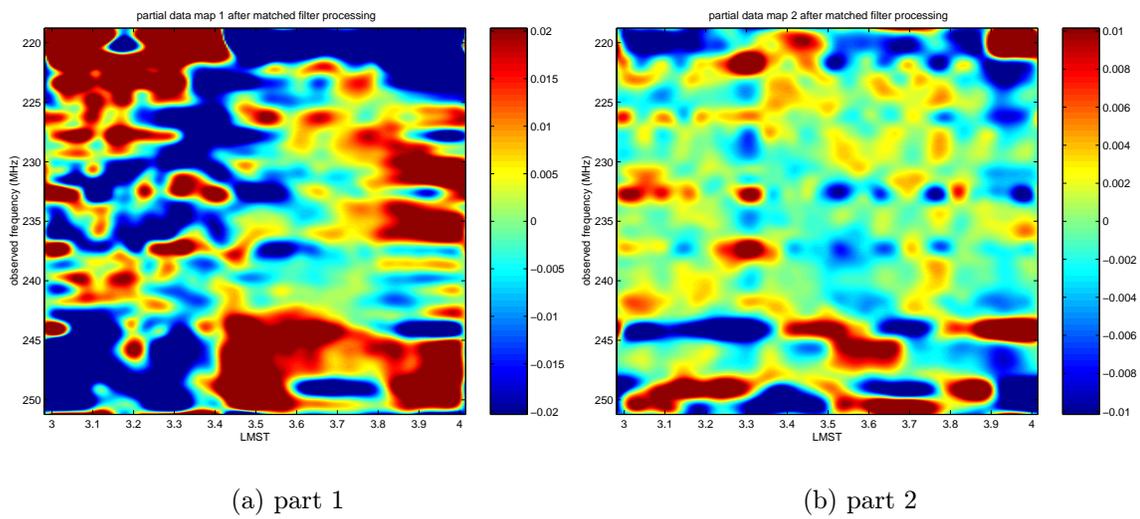
(a) part 1   (b) part 2

Figure A.16: Sidereal time 03–04 h, partial data
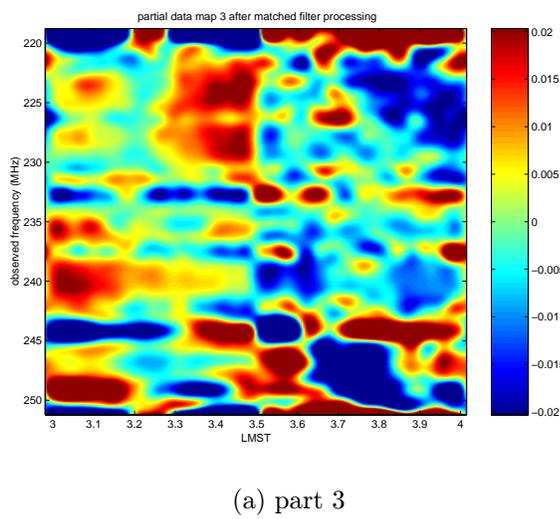
(a) part 3

Figure A.17: Sidereal time 03–04 h, partial data
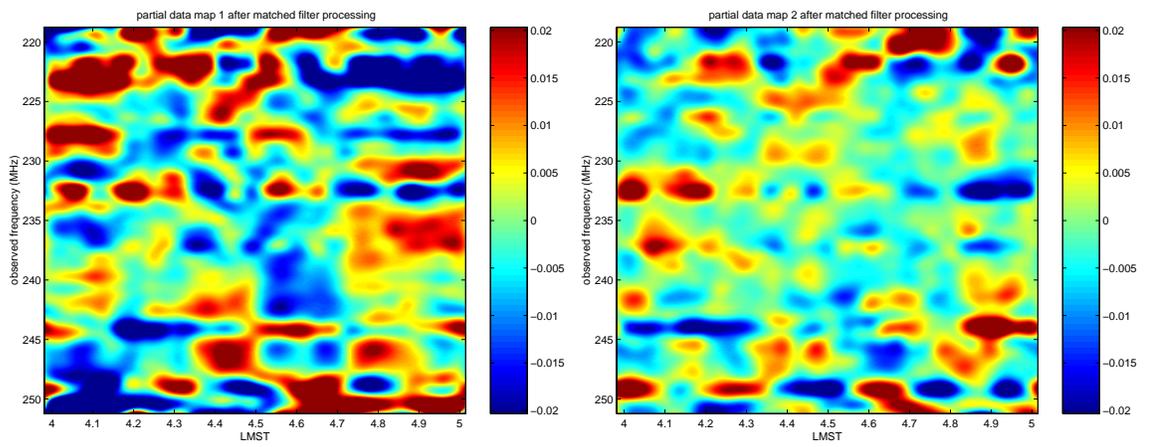
(a) part 1　　　　　　　　　　　　　(b) part 2

Figure A.18: Sidereal time 04–05 h, partial data

Appendix B

Sample second reduction maps

(a) raw　　　　　　　　　　　　　　　　(b) filtered
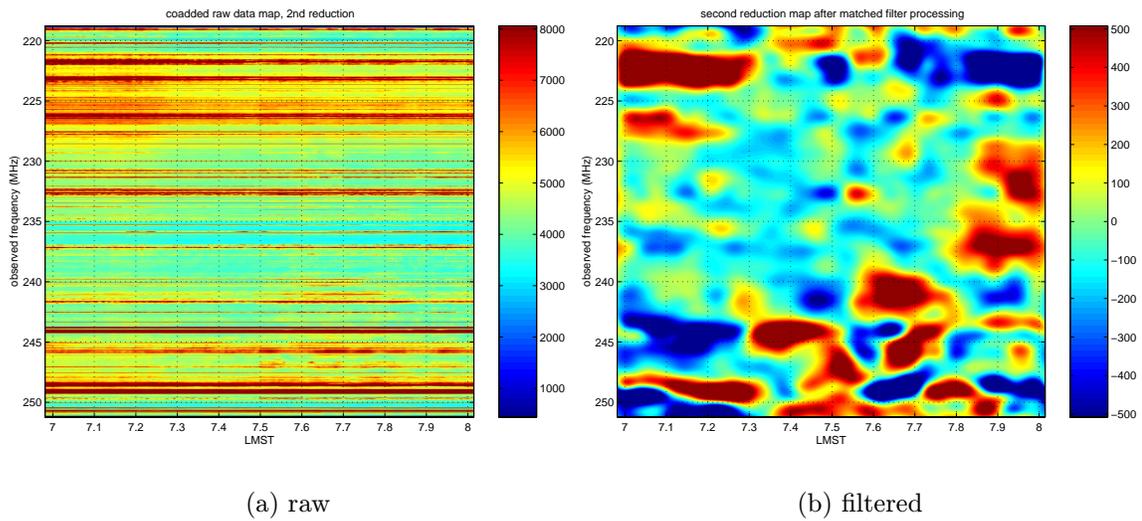
Figure B.1: Sidereal time 07–08 h, all data, 2nd reduction.

(a) part 1   (b) part 2
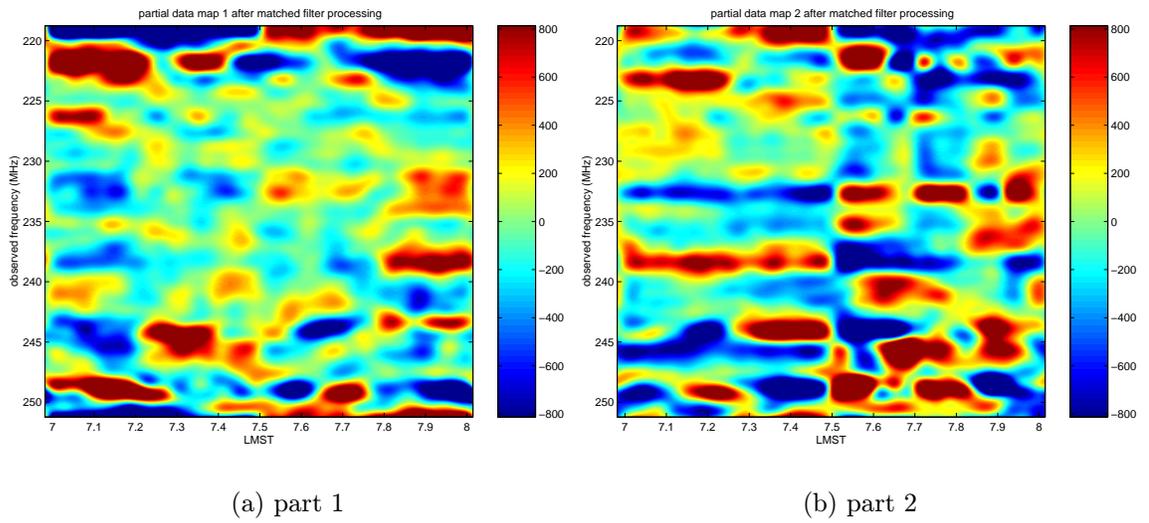
Figure B.2: Sidereal time 07–08 h, partial data, 2nd reduction.

(a) part 3   (b) part 4
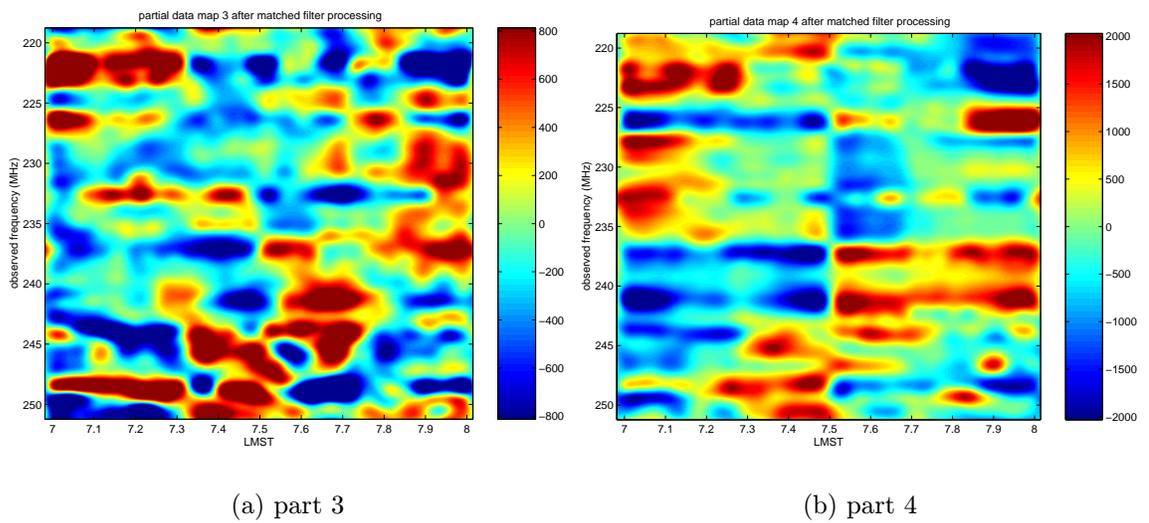
Figure B.3: Sidereal time 07–08 h, partial data, 2nd reduction.

131

Appendix C

LMST to galactic coordinate transformation
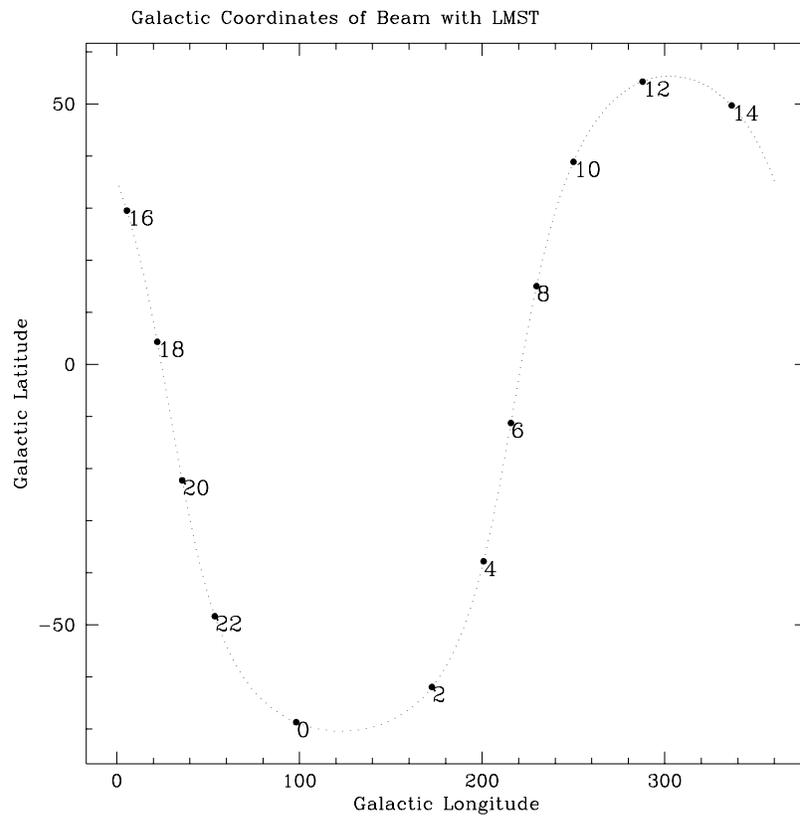
Figure C.1: Transformation between LMST and galactic coordinates for our beam pointing. Plot courtesy of I. Avruch.

Appendix D

Spectrometer board schematic

Appendix E

Spectrometer timing diagrams
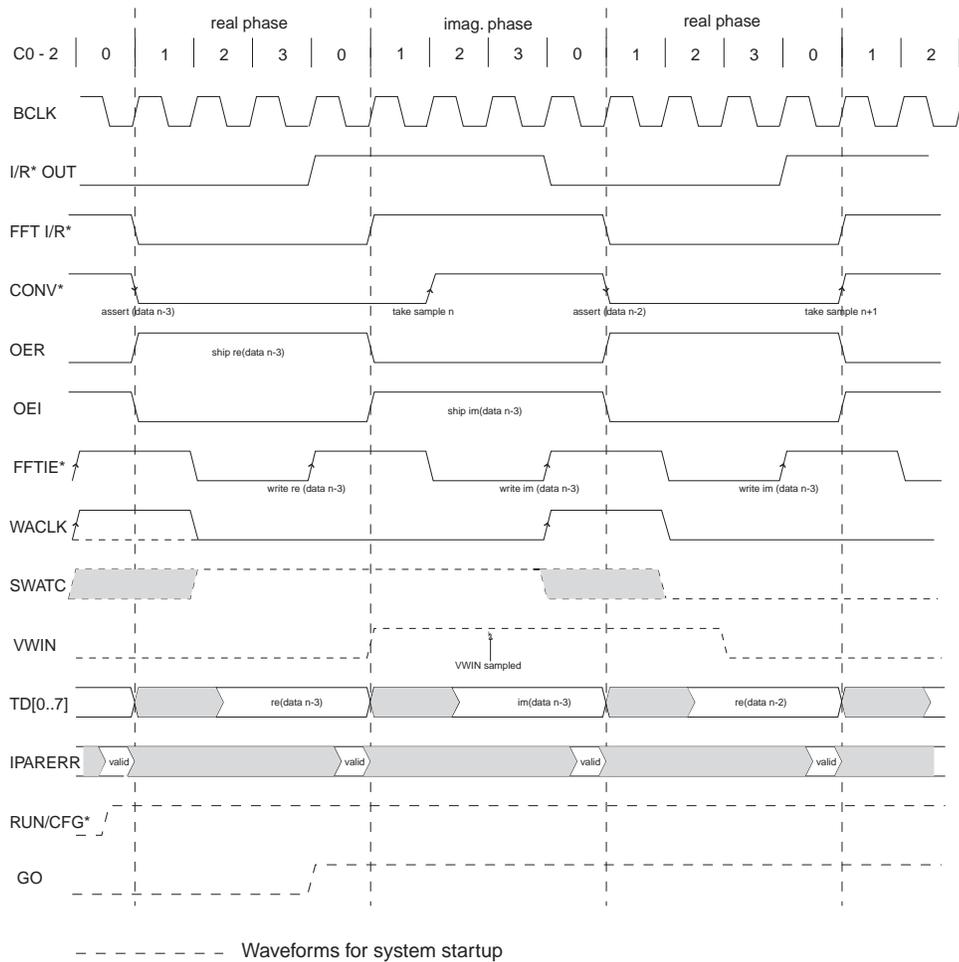
Figure E.1: FFT spectrometer—input timing diagram
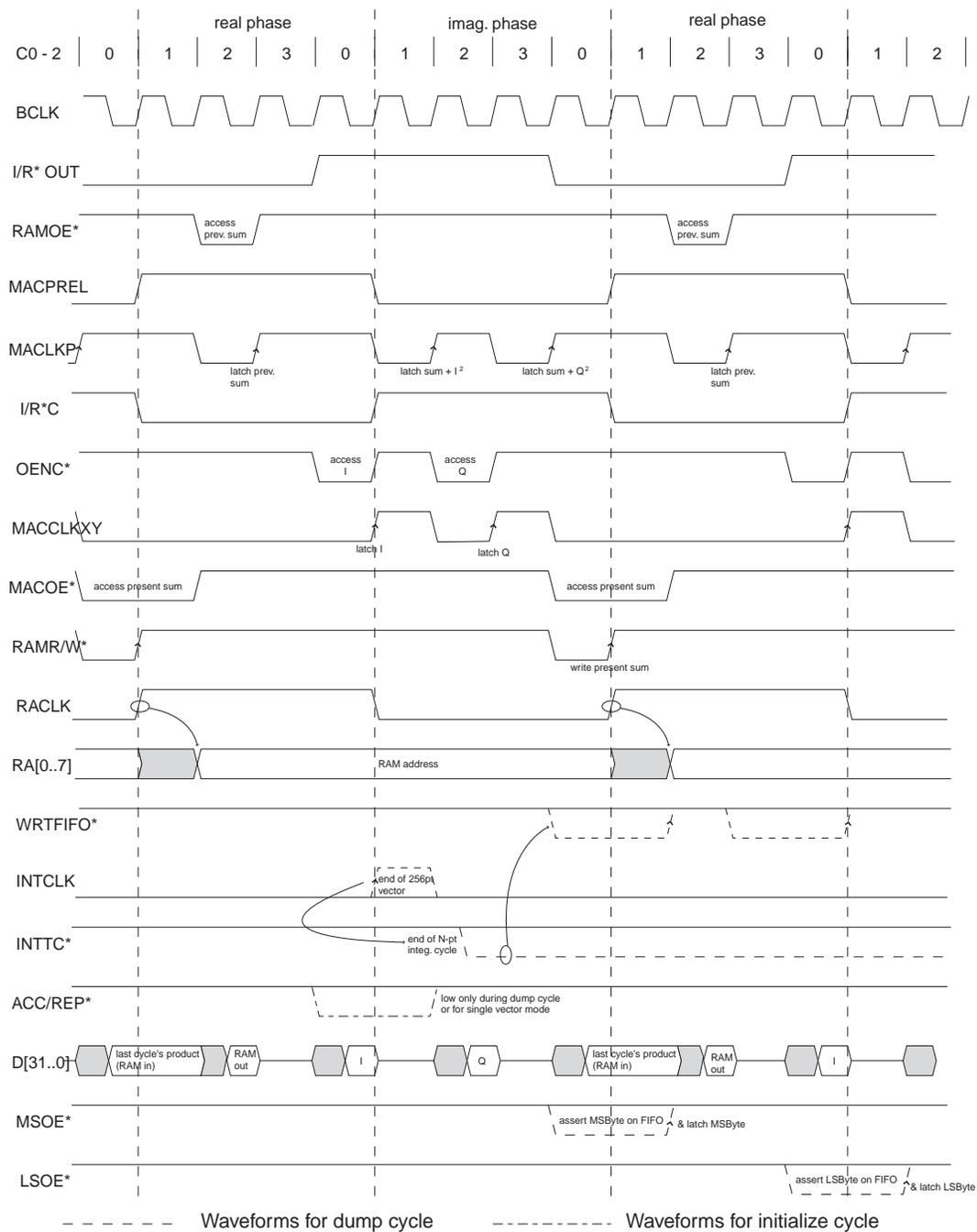
Figure E.2: FFT spectrometer—Austek output and power computation timing
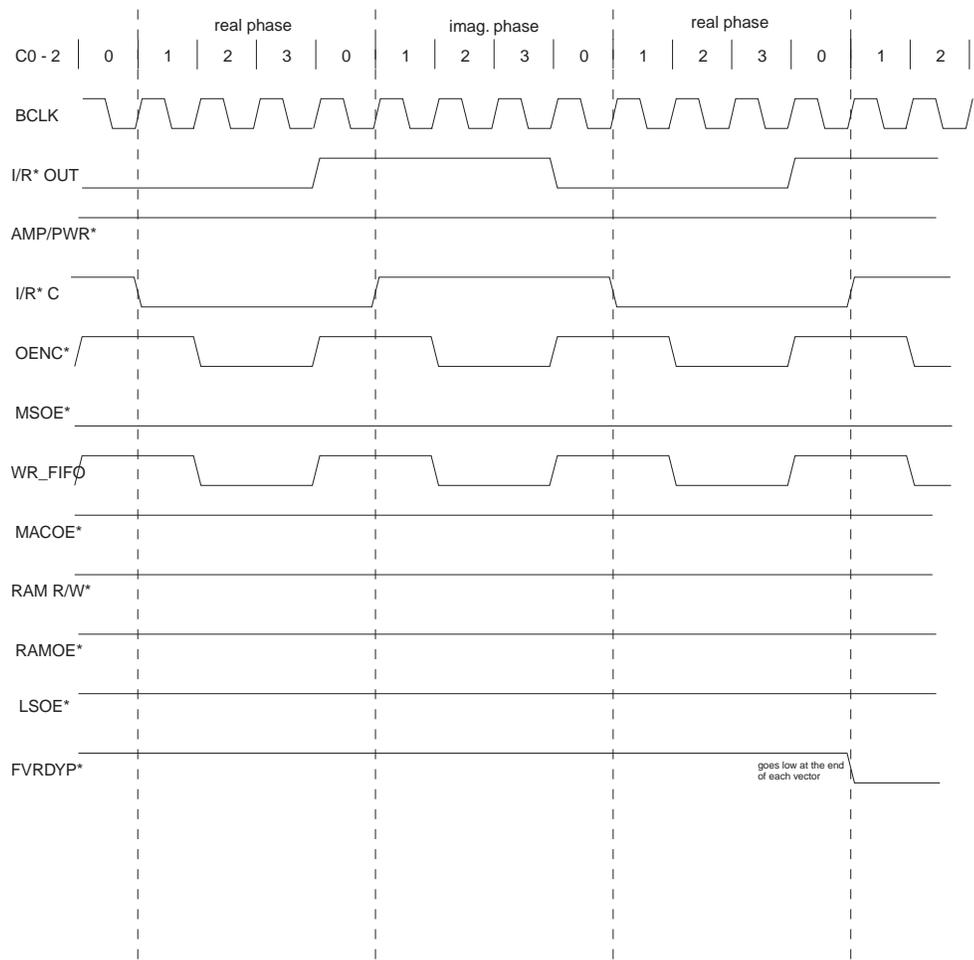
Figure E.3: FFT spectrometer—complex amplitude mode timing
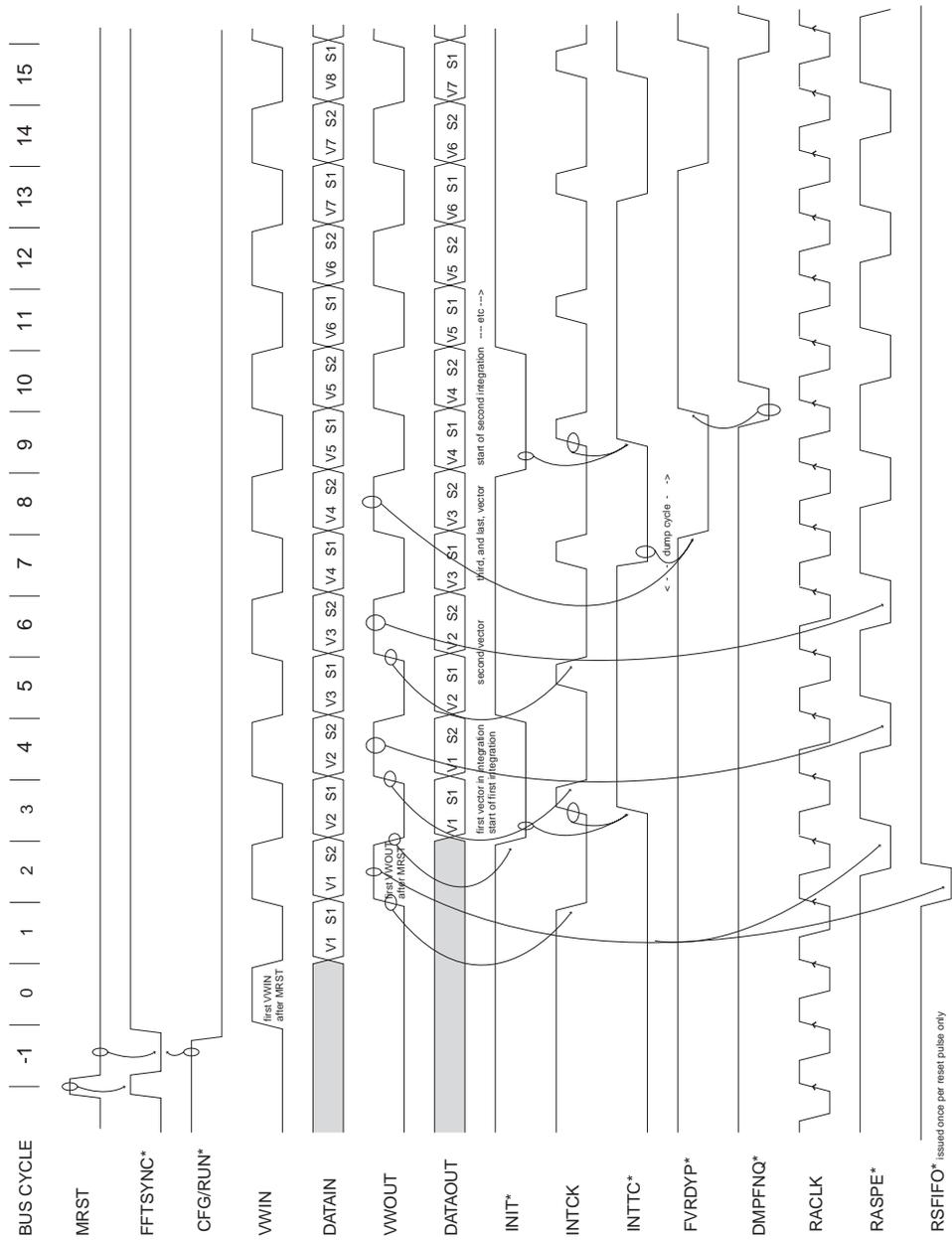
Figure E.4: FFT spectrometer—long time scale timing

137

Appendix F

Data file format

```
Arecibo Early Hydrogen Search   File Format  Ver 0004    03 Sept 1996
=============================   ====================    ============

Field name     # bytes   Description
----------     -------   -----------

I.  File Header
--------------

File start       4       *Always* the ASCII string 'ASEH'
                         (Arecibo Search for Early Hydrogen)
                         Identifies the file as from this experiment

Version          4       ASCII string starting at '0000' and incrementing
                         by 1 each time the file format is revised.
                         The software will then be able to automatically
                         detect & read older files after the rev date.

Date & Time     26       ASCII in the (sample) format 'Tue Jun 13 12:23:38 1995\n
\0'

# of Boards      2       ASCII, total number of spectrometer boards
                         currently in the system, example '32'

Cal Interval     4       ASCII, number of integrations between cal source firings
.

Scale bits       2       Setting of scales in Austek chip.
                         Format is ASCII conversion of HEX number, eg. '1f' for 1
f.

Spare            2       Currently unused.

Self Test       32       Reserved for results of the system self test.
                         Exact format still to be decided.  One byte for
                         each board up to 32.
                         ASCII characters converted from HEX.

Status message  10       bytes 1-2: 2-byte word (little endian) whose bits store
                                    hard error flags reported during the *previous*
                                    data file.
                                    PARITY_ERROR            0x0001
                                    OVERFLOW_ERROR          0x0002
                                    DISK_ERROR              0x0004
                                    DISK_FULL_ERROR         0x0008
                                    FAULT_ASSERT            0x0010
```

```
                               NO_RESPONSE_ERROR        0x0080
                  bytes 3-4: 2-byte word whose bits store tape errors from

                               previous data file.
                               MISSING_TAPE             0x0001
                               OLD_TAPE_NEED_NEW         0x0002
                               VOLUME_ID_MISMATCH       0x0004
                               SPACE_FORWARD_ERROR      0x0008
                               FILE_MISCOUNT            0x0010
                               NO_LOG_FILE              0x0020
                               NEED_NEW_TAPE_NEXT       0x0040
                               LAST_WRITE_FAILED_ERROR  0x0080
                               TAPE_FULL_ERROR          0x0100
                  bytes 5-6: 2-byte word holding IRIG errors from
                             previous file.
                               BAD_IRIG_ERROR           0x0001
                  bytes 7-10: 4 byte word holding the cumulative number
                               of bytes written to the current archive tape.

Comment          290      Operator comment field - 290 ASCII characters,
                          which come from the text file whose name is
                          specified from the menu.
                 -----
file hdr  tot    376      Bytes
                 =====


The spectra follow in binary records according to the format below
(which repeats as required).  All data is little endian.

II. Data Records
----------------
A. Record Header

binary GMT       4        binary seconds since midnight on January 1 1970 UTC

binary ms        2        binary ms to go with s above

Cal State        1        Binary number reporting whether the cal was on, off,
                          or being switcheD immediately prior to this data being
                           ready to dump.
                               00 => CAL OFF
                               FF => CAL ON
                               01 => CAL START (The cal wasn't on entire record.)
                               02 => CAL STOP (The cal wasn't off entire record.)

Board Number     1        Binary, the number (i.e. address) of the board
                          from which *this* spectrum was read.

Int Count        4        Integration count in binary (9.765625 counts/ms)

Window type      1        Binary number describing which of the windows
                          is currently being used.

Spare            1        Currently unused.

Error code       2        Two binary bytes reporting if parity or overflow
                          error were observed during this dump.
                 -----
rec hdr tot      16       Bytes
                 -----


B. Data          512      512 binary bytes of spectral data,  each pair of
                          bytes being one 16 bit point in the power spectrum.
                 -----
Record total     528      Bytes
                 =====
```

Appendix G

Cost of material

This is an estimate of the cost of materials used to build the spectrometer. No attempt is made to estimate labor costs (which were significant).

Antenna feeds: $1000

Receivers: $4000 (upstairs) + $2000 (downstairs)

First local oscillator: $200

IF channelizer: $2,300

Local oscillator array: $1000

Mixer/digitizer array: $360 per board, 18 boards, subtotal $6,480

FFT spectrometer: about $400 per board, 35 boards, subtotal $14,000

Backend computer (including various interface boards): $3,500

Cables and connectors: $1000

Power supplies: $500

Sub-racks, assorted hardware: $500

Grand total: $36,480

Appendix H

Work credit summary

This section summarizes the direct contributions of the author[1] to the design, building and installation of the instrument described in this thesis. A summary of my direct contributions to the data analysis is also given.

Overall

I was responsible for overall system design. I helped to build circuit boards (and supervised our able undergraduate helpers in soldering). I transported the system to Arecibo, where, with the assistance of the electronics department, was primarily responsible for installation, wiring, debugging, testing, and support. Also primarily responsible for scheduling and ordering parts.

Antenna feeds

The full size helices in use at Arecibo were built and installed by the Arecibo Electronics Department. Using a vector network analyzer, I measured the impedance of the unit we used in place on the catwalk. I built the 9 GHz scale models discussed in chapter 2, and did the anechoic chamber measurements. The analytic optimization of the optics in chapter 2 is also my own work.

---

[1]As seems most natural, I will use the first person singular in this section.

Receivers

The Arecibo Electronics Department built the original RF front end. I largely rebuilt this (at the level of connectorized components), specifying and procuring new low loss RF filters, and adding the calibration diode and associated directional couplers. I bench tested the receiver, and characterized its noise temperature with cal on and cal off.

First local oscillator

I designed, built and tested this PLL oscillator, and optimized the output spectrum by selecting the loop filter. The original idea for the design is due to Paul Horowitz.

IF channelizer

This design is by Paul Horowitz for the BETA SETI project. I respecified and bought the IF filters, and built and tested the unit.

Local oscillator array

Also a design by Paul Horowitz for the BETA SETI project. I. Avruch built the unit for this experiment. I provided support, helped to re-specify tuned components for our frequencies, and helped with testing and machining the enclosure.

Mixer/digitizer

This is yet another BETA SETI design by Paul Horowitz. I developed an alignment procedure, and calibrated and tested our boards. Also, I provided support to the Arecibo Electronics Department in building the mixer/digitizer boards for the AOFTM.

FFT spectrometer

I was responsible for the entire design, prototyping and debugging of these boards (with the exception of the printed circuit board layout, by Nick Shectman, and the microcontroller code by Paul Horowitz). The boards include a PAL state machine, for which I programmed all the code.

Backend computer

I designed and tested the spectrometer interface board, and wrote, with I. Avruch, the original executive loop software which programmed and interrogated the spectrometer boards and downloaded data. I defined the first version of the data file format.

Data analysis

I developed the simulations of co-adding comparing average and median in the presence of RFI. I also developed the modified matched filters, ran them on the co-added data maps, and scrutinized and extracted the candidate signals from the filtered maps. I. Avruch wrote the co-adding programs and piped through the data. I had regular discussions with him on procedures and algorithms for the co-adding procedures. Also, I did some preliminary analysis on the raw data.

Other contributions—BETA and AOFTM

I provided support for both these projects. Some of my design contributions to BETA include working out the timing for a PAL on the feature recognizer board and reviewing the timing on the 4 Megapoint FFT board. I also worked with Suhail Shah, on the telescope switching hardware. I helped extensively with the purchasing of parts, and the building and installation of the system. For the AOFTM I talked with Jim Cordes at the outset to design the spectrometer with appropriate features for pulsar work. I provided a great deal of telephone and email support to his student, Dirk Koechner, when he was building the fast interface to a Sun workstation. Also, I helped Jim Cordes debug the system, meeting him at Arecibo on two occasions.

Talks and publications

I gave a talk at the URSI meeting in Boulder, Colorado; also a poster at AAS, San Antonio, Texas both in January 1996. More recently I gave a talk at the Highly Redshifted Radio Lines Symposium, in Green Bank, West Virginia, October 1997.

Bibliography

[1] Penzias A. A. and Wilson R. W. 1965, *Ap. J.*, **142**, 419

[2] Dicke R. H., Peebles P. J. E., Roll P. G. and Wilkinson D.T. 1965, *Ap. J.*, **142**, 414

[3] Lubin P., Villela T., Epstein G. and Smoot G. F. 1985, *Ap.J.*, **298**, L1

[4] Smoot G. F. *et al* 1992, *Ap.J.*, **396**, L1

[5] Thommes E., Meisenheimer K., Fockenbrock R., Hippelein H., Röser H.-J. and Beckwith S. 1998, *MNRAS*, **293**, L6

[6] Dey A., Spinrad H., Stern D., Graham J. R. and Chaffee F.H. 1998, *Ap.J.* L (in print, astro-ph/9803137)

[7] Sunyaev R. A., and Zel'dovich Ya. B. 1972, *Astr. Astrophys.*, **20**, 189

[8] Fixsen D. J. *et al* 1994, *Ap.J.*, **420**, 445

[9] Lineweaver C.H. *et al* 1995, *Ap.J.*, **448**, 482

[10] Kolb E.W. and Turner M.S. 1990, *The Early Universe*, Addison-Wesley

[11] Sachs R. K. and Wolfe A.M. 1967, *Ap.J.*, **147**, 73

[12] Peebles P.J.E. 1993, *Principles of Physical Cosmology*, Princeton University Press

[13] Rees M. 1995, *Perspectives in Astrophysical Cosmology*, Cambridge University Press

[14] Shu F.H. 1982, *The Physical Universe*, University Science Books

[15] Kaufmann W.J., 1987 *Universe*, Second Edition, W.H. Freeman and Company

[16] Kellerman K.I. and Owen F.N. 1988, in *Galactic and Extragalactic Radio Astronomy*, G.L. Verschuur and K.I. Kellerman, Editors, Springer Verlag, 563

[17] Woltjer N. 1990, in *Active Galactic Nuclei*, T.J-L. Courvoisier and M. Meyer, Editors, Swiss Society for Astrophysics and Astronomy, Springer Verlag, 1

[18] Schneider D.P., Schmidt M. and Gunn J.E. 1991, *Astron.J.*, **102**(3), 837

[19] Gunn J.E. and Peterson B.A. 1965, *Ap.J.*, **142**, 1633

[20] Giallongo, E., D'Odorico, S., Fontana, A., McMahon, R.G., Savaglio, S., Cristiani, S., Molaro, P. and Trevese D. 1994, *Ap.J.*, **425**, L1

[21] Cohen, J.G., Cowie, L.L., Hogg, D.W., Songaila, A., Blandford, R., Hu, E.M., Shopbell, P. 1996, *Ap.J.*, **471**, L5

[22] Steidel, C.C. and Hamilton D. 1992, *Astron.J.*, **104**, 3

[23] Steidel, C.C., Adelberger, K.L., Dickinson, M., Giavalisco, M., Pettini, M. and Kellogg, M. 1998, *Ap.J.*, **492**, 428

[24] Brown R.L. and Vanden Bout P.A. 1992, *Ap.J.*, **397**, L19

[25] Brown R.L. and Vanden Bout P.A. 1993, *Ap.J.*, **412**, L21

[26] Guilloteau S., Omont A., McMahon R.G., Cox P. and Petitjean P. 1997, *A&A*, **328**, L1

[27] Ewen H.I. and Purcell E.M. 1951, *Nature*, **168** 356

[28] Giovanelli R. and Haynes M.P. 1988, in *Galactic and Extragalactic Radio Astronomy*, G.L. Verschuur and K.I. Kellerman, Editors, Springer Verlag, 522

[29] Swarup G. 1996, in *Cold Gas at High Redshift*, M.N. Bremner *et al* eds, Kluwer Academic Publishers, 457

[30] Subramanian K. and Padmanabhan T. 1993, *MNRAS*, **265**, 101

[31] Kumar A., Padmanabhan T. and Subramanian K. 1995, *MNRAS*, **272**, 544

[32] Bagla J. S., Nath B. and Padmanabhan T. 1997, *MNRAS*, **289**, 671

[33] Madau P., Meiksen A. and Rees M.J., *Ap.J.* **475**, 429

[34] Scott D., and Rees M. 1990, *Mon. Not. R Astr. Soc.* **247**, 510

[35] Maoli R., Ferrucci V., Melchiorri F., Signore M. and Tosti D. 1996, *Ap.J.*, **457**, 1

[36] Davies R. D., Pedlar A. and Mirabel I. F. 1978, *MNRAS*, **182**, 727

[37] Bebbington D. H. O. 1986, *MNRAS*, **218**, 577

[38] Hardy E. and Noreau L. 1987, *Astron. J.*, **94**(6), 1469

[39] Noreau L. and Hardy E. 1988, *Astron.J.*, **96**(6), 1845

[40] Wieringa M.H., de Bruyn, A.G., and Katgert, P. 1992, *A&A*, **256**, 331

[41] Uson J.M., Bagri D.S. and Cornwell T.J. 1991, *Ap.J.*, **377**, L65

[42] Uson J.M., Bagri D.S. and Cornwell T.J. 1991, *Phys. Rev. Lett.*, **67**, 24

[43] Briggs F.H., Sorar E. and Taramopoulos A. 1993, *Ap.J.*, **415**, L99

[44] Burke B.F., Horowitz P., Weintroub J. and Davis M.M. 1993, *High-z Hydrogen Search—Proposal for observing time*

[45] Leigh, D.L. and Horowitz, P. 1997, in *Astronomical and Biochemical Origins and the Search for Life in the Universe*, Proc. IAU Colloq. No. 161, C.B. Cosmovici, S. Bowyer and D. Werthimer, Editors Capri: Editrice Compositori, 601

[46] Pauliny-Toth I.I.K. and Shakeshaft J.R. 1962, *MNRAS*, **124**, 61

[47] Northeast Radio Observatory Corporation (NEROC) Internal Report 1968, *Comparison of Large Sperical and Parabolic Radio Telescopes*

[48] Burke B.F. and Graham-Smith F. 1997 *An Introduction to Radio Astronomy*, Cambridge University Press

[49] Kraus J.D. 1986, *Radio Astronomy*, 2nd Edition, Cygnus-Quasar Books

[50] Avruch I. M. 1998, *Ph.D. Thesis*, Massachusetts Institute of Technology

[51] Kraus J.D. 1988, *Antennas*, 2nd Edition, McGraw-Hill

[52] Johnson R.C., Editor, 1993, *Antenna Engineering Handbook*, 3rd Edition, McGraw-Hill

[53] Emerson D. 1994, *Private communication*

[54] Kraus J.D. 1977, *IEEE Trans. Ants. Prop.*, **AP-25**, 913

[55] Kildall P.S. 1990, *IEEE Trans. on Ant. and Prop.*, **38**(10), 1537

[56] Castro E.I. 1994 *Private communication*

[57] Horowitz P. and Hill W. 1989, *The Art of Electronics*, Cambridge University Press

[58] Press W. H., Teukolsky S. A., Vetterling W. T. and Flannery B. P. 1992, *Numerical Recipes in C—The Art of Scientific Computing*, 2nd Edition, Cambridge University Press

[59] Leigh D.L. 1998, *Ph.D. Thesis*, Harvard University

[60] Austek Microsystems 1990, *A41102 Frequency Domain Processor Product Specification*, rev 4.4, Australia

[61] Harris 1978, *Proc. IEEE*, **66**, 1

[62] Oppenheim A.V. and Schafer R.W. 1989, *Discrete-Time Signal Processing*, Prentice Hall

[63] Hagen J.B. 1996, *Radio Frequency Electronics—Circuits and Applications*, Cambridge University Press

[64] van Vleck J.H. and Middleton D. 1966, *Proc. IEEE*, **54(1)**, 2

[65] Knuth D.E. 1981, *The Art of Computer Programming—Seminumerical Algorithms*, Addison-Wesley

[66] Davis M.M. 1993, 1994 *Private communication*

[67] Werthimer D. and Ng D. 1994, *Private communication*

[68] Cordes J. 1993, *A DFT Filter Bank Based on the Austek Chip*, Arecibo Upgrade Notes

[69] Koechner D. 1995, *Masters Thesis*, Cornell University

[70] http://www.iridium.com

[71] http://www.teledesic.com

[72] Browne M. W. 20 March 1998, *New York Times*, A19

[73] 18 April 1998, *The Economist*, 77

[74] Davis M.M. 1997, *Private communication*

[75] Lupton R. 1993, *Statistics in Theory and Practice*, Princeton University Press

[76] 1996 Stairs I. and Taylor J.H. *Private communication*

[77] Bradley R., Wilson S., Barnbaum C. and Wang B. 1996, *An Adaptive Interference Cancelling Receiver for Radio Astronomy—Theory* NRAO Electronics Division Internal Report No. 305

The remaining references are not explicitly cited in the text.

Conference abstracts and proceedings:

[78] Avruch I.M., Burke B.F., Davis M.M., Horowitz P., Leigh D.L. and Weintroub J. 1995, *Bulletin of the American Astronomical Society*, **27-4** 53-04

[79] Avruch I.M. and Weintroub J. 1996, *Proc. URSI National Radio Science Meeting*, Boulder, Colorado, 339

[80] Weintroub J., Horowitz P., Avruch I.M. and Burke B.F. 1998, *Proc. Highly Redshifted Radio Line Symposium*, Green Bank, West Virginia, in press

A classic book on the Fourier transform:

[81] Bracewell R.N. 1986, *The Fourier Transform and Its Applications*, 2nd Edition McGraw-Hill

A recent short review of cosmology:

[82] Lineweaver C. H. 1998, *astro-ph*, 9803100

Computer tool references:

[83] Abrahams P.W. and Larson B.R. 1992 *Unix for the Impatient*, Addison-Wesley

[84] Lamport L. 1994, LaTeX *: A Document Preparation System User's Guide and Reference Manual*, 2nd Edition, Addison-Wesley

[85] Hahn J. 1991, LaTeX *For Everyone*, Personal TeXInc.