Removing radio interference from contaminated astronomical spectra using an independent reference signal and closure relations

F. H. Briggs

Kapteyn Astronomical Institute, University of Groningen Postbus 800 9700 AV Groningen, The Netherlands; fbriggs@astro.rug.nl

J. F. Bell & M. J. Kesteven

CSIRO Australia Telescope National Facility, PO Box 76 Epping NSW 1710, Australia; jbell@atnf.csiro.au, mkesteve@atnf.csiro.au

ABSTRACT

The growing level of radio frequency interference (RFI) is a recognized problem for research in radio astronomy. This paper describes an intuitive but powerful RFI cancellation technique that is suitable for radio spectroscopy where time-averages are recorded. An RFI "reference signal," is constructed from the cross power spectrum of the signals from the two polarizations of a reference horn pointed at the source of the RFI signal. The RFI signal paths obey simple phase and amplitude closure relations, which allows computation of the RFI contamination in the astronomical data and the corrections to be applied to the astronomical spectra.

Since the method is immune to the effects of multipath scattering in both the astronomy and reference signal channels, "clean copies" of the RFI signal are not required.

The method could be generalized (1) to interferometer arrays, (2) to correct for scattered solar radiation that causes spectral "standing waves" in single-dish spectroscopy, and (3) to pulsar survey and timing applications where a digital correlator plays an important role in broadband pulse dedispersion.

Future large radio telescopes, such as the proposed LOFAR and SKA arrays, will require a high degree of RFI suppression and could implement the technique proposed here with the benefit of faster electronics, greater digital precision and higher data rates.

Subject headings: instrumentation: detectors — Methods: analytical, interference suppression

1. Introduction

The growing level of radio frequency interference (RFI) is a recognized problem for research in radio astronomy. Fortunately, the technological advances that are giving rise to the increasing background of radiation – through increased telecommunications and wide-spread use of high speed electronics – are also providing some of the tools necessary for separating astronomical signals from undesirable RFI contamination. New radio telescopes will necessarily have RFI suppression, excision and cancellation algorithms intrinsic to their designs. No one technical solution will make radio observations immune to interference; successful mitigation is most likely to be a hierarchical or progressive approach throughout the telescope, combining new instrumentation and algorithms for signal conditioning and processing (Ekers & Bell 2000; Ellingson 1999).

Techniques from the communication industry that are finding application in radio astronomy experiments include (1) adaptive beam forming with array telescopes that steer nulls of the instrument reception pattern in the directions of sources of RFI (Hampson et al. 1998; Ellingson & Hampson 2000; Leshem et al. 2000; Smolders & Hampson 2000; Kewley et al. 2000), (2) parametric signal modeling techniques, where the RFI signal is received and decoded to obtain a high signal-to-noise ratio (SNR) reference signal for subtraction from the astronomical data (Ellingson et al. 2000; Leshem & van der Veen 1999a; Leshem & van der Veen 1999b), and (3) adaptive filtering using a reference horn to obtain a high SNR copy of the RFI for real-time cancellation from the signal path ahead of the standard radio astronomical backend processors (Barnbaum & Bradley 1998).

This paper describes an intuitive but powerful RFI cancellation technique that is suitable for radio spectroscopy where time-averages are recorded. The method requires computation of cross power spectra between the RFI contaminated astronomical signals and high signal-to-noise ratio RFI "reference signals" obtained from a receiving system that senses the RFI but not the astronomical signal. The correction term that removes the unwanted RFI is computed from closure relations obeyed by the RFI signal. The test applications reported here derived the reference signal either from a separate horn antenna aimed at the RFI source or from a second feed horn at the focus of the Parkes telescope, as illustrated in Fig 1. For these experiments, we recorded digitally sampled baseband signals from two polarizations for both the reference and astronomy feeds, and then we performed the cross correlations in software off-line. However, the method could use correlation spectrometers of the sort already in use at radio observatories. With minor design enhancements, future generation correlators could incorporate this technique with the additional benefit of the faster electronics, greater digital precision, and higher data rates that technological advance promises.

There are a number of advantages to performing the RFI in a "post-correlation" stage. Foremost is that the RFI subtraction remains an option in the data reduction path, rather than a commitment made online and permanently. Furthermore, the correlation method is effectively a coherent subtraction, since the correlation functions retain the information describing relative phase

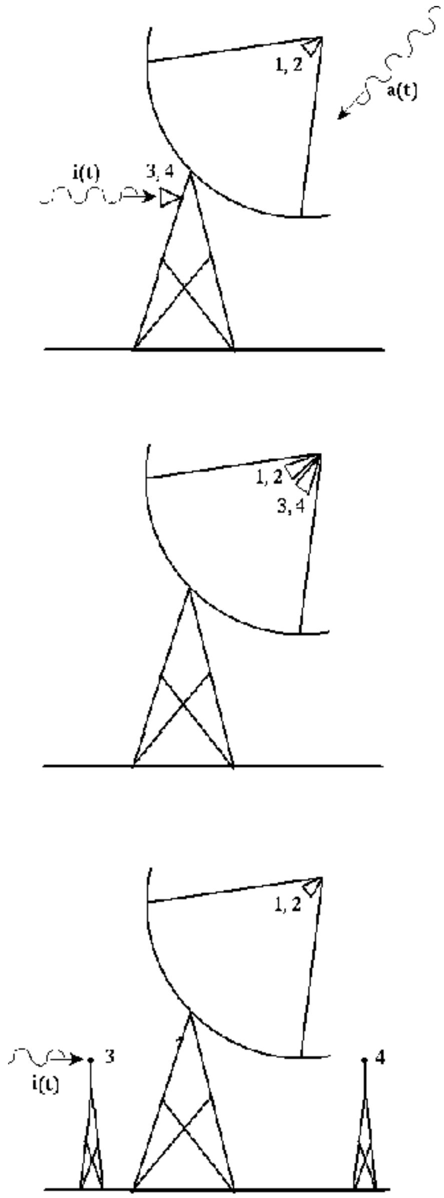

Fig. 1.— Three configurations applicable to the analysis in this paper. *Top* The four basebands are recorded from the 2 polarizations of the Parkes telescope feed and 2 polarizations of a reference horn directed at the RFI source. *Center* The four basebands correspond to 2 polarizations from each of two Parkes feeds. *Bottom* A proposed system for optimal application of the RFI subtraction technique described here.

between the RFI entering in the astronomy data stream and the RFI entering the reference antenna. As we show in this paper, this means that the RFI noise power is largely subtracted, leaving only the usual components of system noise.

This paper provides a description of the post-correlation RFI cancellation technique and illustrates its success with data from the Parkes telescope. A mathematical overview shows (1) why unknown multipaths do not cause the algorithm to break down, (2) how to simply construct a suitable RFI reference spectrum, and (3) how to build an inverse filter to obtain immunity to low signal levels at frequencies that suffer destructive interference by multipathing in the reference horn signal path.

2. Mathematical description of the method

In our mathematical model, we make the assumption that the RFI source emits a single signal $i(t)$. (At the RFI source, the signal from a single power amplifier feeds an antenna of unknown, but irrelevant, polarization.) The RFI that appears differently in the recorded data channels at Parkes has experienced scatterings with different path lengths and amplitudes, so that the received signals are linear sums of time delayed versions of the original broadcast $i(t)$.

In fact, the model is applicable to multiple interferers within the spectrometer band, provided they do not overlap in frequency.

Consider for the moment a single interferer, which propagates through the four signal paths s_1, s_2, s_3, and s_4 that will be processed: there are two astronomical channels, $s_1(t)$ and $s_2(t)$, which convey the voltages $a_A(t)$ and $a_B(t)$ from the celestial sources for the two independent polarizations from the Parkes Telescope receiver along with contamination from the RFI signal $i(t)$. Radiation from astronomical sources may be partially polarized, causing $a_A(t)$ and $a_B(t)$ to be correlated to some degree. The two reference channels carry $s_3(t)$ and $s_4(t)$, containing the representations of $i(t)$ but negligible signal from the celestial sources. For example, the measured signal in channel 1 comprises channel noise $n_1(t)$, plus the convolutions of the impulse responses for the multiple scattering paths for the interference signal and the true astronomy signal: $s_1(t) = H_A(t)\bigstar a_A(t) + H_1(t)\bigstar i(t) + n_1(t)$. Here $H_A(t)$ and $H_1(t)$ are impulse responses for the astronomy signal path and interference respectively, and $\bigstar$ is the convolution operator.

For many purposes, an intuitive picture of the multipathing results from considering the scattering sites to be achromatic mirror-like scatterers each with relative effective areas $G_{A,j}$ and $G_{1,k}$, and attaching the path delay to each separate version of the astronomy and RFI signals, $s_1(t) = \sum_j G_{A,j} a_A(t-\tau_{1,j}) + \sum_k G_{1,k} i(t-\tau_{1,k}) + n_1(t)$. The time delays τ are determined by the different path lengths $L_{1,j}$ to give $\tau_{1,j} = L_{1,j}/c$, where c is the speed of light.

All the signal paths are vulnerable to corruption by stochastic noise. The noise terms $n_1(t)$, $n_2(t)$, $n_3(t)$, and $n_4(t)$ should ideally be uncorrelated among the different data paths. Unfortunately, in real astronomical systems, there is likely to be low level coupling between the two orthogonal polarizations of a feed horn or common stray radiation pickup from spillover that will make a weakly correlated noise floor in some of the cross power spectra. This will form a systematic limitation to the accuracy of the subtraction.

In this experiment, the goal is to explore the usefulness of cross-correlation spectra to correct for the effects of RFI in time-averaged spectra. These spectra are products of scaled sums of the of the Fourier transforms of the astronomy signals, the RFI and noise. In the tests with real data in section 4, we compute estimates of the complex spectra

$$\begin{aligned} S_1(f) &= g_A A_A + g_1 I + N_1 \\ S_2(f) &= g_B A_B + g_2 I + N_2 \\ S_3(f) &= g_3 I + N_3 \\ S_4(f) &= g_4 I + N_4 \end{aligned} \tag{1}$$

from Fourier transforms of finite length time series of discrete samples of the four signals. The transforms contain contributions from the celestial sources $A_A(f)$ and $A_B(f)$, the RFI $I(f)$, and the noise in each channel $N_i(f)$, modulated by the associated complex voltage gains, which are the Fourier transforms of the impulse responses $H(t)$. The gains for each channel separate into dependencies on (1) the path delay L/c, which appears in a frequency dependent phase term, according to the Shift Theorem of Fourier transforms, and (2) a possible additional frequency dependence $g(f)$

of each delay path:

$$g_1(f) = \sum g_{1,k}(f) e^{i2\pi f L_{1,k}/c} \tag{2}$$

These complex gain and delay factors are sufficiently general to include complicated scatterers and propagation through dispersive and lossy media.

The real power spectra for the four data channels have the following form, once terms that average toward zero are omitted and the complex gains are assumed to be constant over the time span for which the spectra are computed:

$$\begin{aligned}
P_1(f) &= < S_1 S_1^* > \\
&= |g_A|^2 < |A_A|^2 > \\
&\quad + |g_1|^2 < |I|^2 > + < |N_1|^2 > \\
P_2(f) &= < S_2 S_2^* > \\
&= |g_B|^2 < |A_B|^2 > \\
&\quad + |g_2|^2 < |I|^2 > + < |N_2|^2 > \\
P_3(f) &= < S_3 S_3^* > \\
&= |g_3|^2 < |I|^2 > + < |N_3|^2 > \\
P_4(f) &= < S_4 S_4^* > \\
&= |g_4|^2 < |I|^2 > + < |N_4|^2 >
\end{aligned} \tag{3}$$

We use the superscript symbol * to represent complex conjugation and the $< ... >$ notation to signify averages over an integration time t_{int}; in the tests we describe in section 4, we find the method is effective for t_{int} as long as $\sim$1 sec. We adopt a normalization where the power levels $< |A_A|^2 > \approx < |I|^2 > \approx < |N_1|^2 >$ so that, for example, in data channel 1 the signal to noise ratio $SNR_1 \sim |g_A|^2$, and the interference to noise ratio $INR_1 \sim |g_1|^2$.

The goal of the RFI cancellation will be to form estimates of the $|g_1|^2 < |I|^2 >$ and $|g_2|^2 < |I|^2 >$ terms and then subtract them from $P_1(f)$ and $P_2(f)$, while leaving the astronomical signal (and noise) behind.

This discussion has assumed that the complex gain terms are constant over the integration time t_{int}, allowing us to separate the interference from the gain in expressions such as:

$$< g_1 I g_1^* I^* > = |g_1|^2 < |I|^2 > \tag{4}$$

In anticipation of the discussion of section 4, we note that this assumption will fail for extended integration times since the scattering paths that lead the RFI signal to the telescope feed will change, resulting in loss of precision in the cancellation scheme and leading to substantial residuals in the corrected spectra.

The complex cross power spectra for all combinations of the four data channels have the following form when the leading contributions are retained:

$$\begin{aligned}
C_{12}(f) &= < S_1 S_2^* > \\
&= g_A g_B^* < A_A A_B^* > \\
&\quad + g_1 g_2^* < |I|^2 > \\
&\quad + g_1 < I N_2^* > + g_2^* < N_1 I^* > \\
&\quad + < N_1 N_2^* > \\
C_{ij}(f) &= < S_i S_j^* > \\
&= g_i g_j^* < |I|^2 > \\
&\quad + g_i < I N_j^* > + g_j^* < N_i I^* > \\
&\quad + < N_i N_j^* > \\
&\quad \text{for i} \neq \text{j}, \quad \text{j} > 2
\end{aligned} \tag{5}$$

In order to cancel the dominant RFI terms in the power spectra Eqns 3, we need to compute quantities of the form $|g_1|^2 < |I|^2 >$, which can be subtracted from the measured $< |S_1|^2 >$. One possibility would be a combination of auto and cross power spectra of the form

$$\begin{aligned}
|g_1|^2 < |I|^2 > &= \frac{g_1 g_3^* g_1^* g_3}{g_3^* g_3} < |I|^2 > \\
&\approx \frac{|C_{13}|^2}{< |S_3|^2 >} \\
&\approx \frac{|C_{13}|^2}{|g_1|^2 < |I|^2 > + < |N_3|^2 >}
\end{aligned} \tag{6}$$

However, the problem encountered with this approach is that $< |S_3|^2 >$ is biased by the term $< |N_3|^2 >$, which averages over time to the total power spectrum of the noise in data channel 3. This might be compensated by calibrating the noise spectrum, in order to improve the estimate of $|g_3|^2 < |I|^2 > = < |S_3|^2 > - < |N_3|^2 >$ to use in the denominator of Eqn 7.

An alternate combination that avoids this bias forms the estimate of $|g_1|^2 < |I|^2 >$ from three cross power spectra, in which the noise terms have the form $< N_i N_j^* >$ and $< I N_j^* >$ which average toward zero as the integration time increases:

$$|g_1|^2 < |I|^2 > = \frac{g_1 g_3^* g_1^* g_4}{g_3^* g_4} < |I|^2 >$$

$$\approx \frac{C_{13}C_{14}^*}{C_{34}^*} \quad (7)$$

$$|g_2|^2 < |I|^2 > \;=\; \frac{g_2 g_3^* g_2^* g_4}{g_3^* g_4} < |I|^2 >$$

$$\approx \frac{C_{23}C_{24}^*}{C_{34}^*} \quad (8)$$

$$g_1 g_2^* < |I|^2 > \;=\; \frac{g_1 g_4^* g_2^* g_3}{g_3 g_4^*} < |I|^2 >$$

$$\approx \frac{C_{14}C_{23}^*}{C_{34}} \quad (9)$$

The expressions involving the measured cross power spectra are "approximate," since the cross power spectra result from finite integrations, and the noise terms will limit the precision of the cancellation.

The occurrence of the C_{34} term in the denominator for Eqns 7-9 indicates there will be a problem in implementing a correction scheme in frequency ranges where C_{34} becomes small or zero. In many situations when the signal to noise ratio for the spectra in the numerator is very high, the cross power spectra in the numerator will also be small or zero whenever C_{34} is small, so that divergence will be canceled. A simple means to avoid division by small numbers in the presence of noise in this kind of situation is to create a filter of the following form:

$$|g_1|^2 < |I|^2 > \;\approx\; \frac{C_{13}C_{14}^*C_{34}}{\psi(f) + C_{34}C_{34}^*} \quad (10)$$

$$|g_2|^2 < |I|^2 > \;\approx\; \frac{C_{23}C_{24}^*C_{34}}{\psi(f) + C_{34}C_{34}^*} \quad (11)$$

$$g_1 g_2^* < |I|^2 > \;\approx\; \frac{C_{14}C_{23}^*C_{34}^*}{\psi(f) + C_{34}C_{34}^*} \quad (12)$$

where $\psi(f)$ is the square of the power spectrum of the noise present in $C_{34}(f)$. Whenever $\psi(f)$ becomes small compared to $C_{34}(f)$, the expressions Eqns 10-12 revert to Eqns 7-9. When the noise exceeds the signal power in C_{34}, the computed correction tends to zero. In practice, during the test described here, a constant ψ_0 was used in place of $\psi(f)$. Alternatively, division by zero can be avoided by testing the amplitude of $C_{34}(f)$ for significance above a noise threshold and setting the correction to zero when the significance criterion is not met.

These corrections for the the auto correlation spectra (Eqns 7 and 8) are expected to be real valued. Therefore a logical test for the accuracy of the correction is that phases computed for the frequencies of strong RFI signal should be close to zero. In fact, this is a statement of phase closure. Note that the denominators of Eqns 10-12 are purely real, and the numerators, such as $C_{13}C_{14}^*C_{34}$, form logical triangles for computing closure phases.

An amplitude closure relation

$$\frac{C_{13}C_{24}}{C_{23}C_{14}} = \frac{g_1 g_3^* g_2 g_4^*}{g_2 g_3^* g_1 g_4^*} = 1 \quad (13)$$

can also be constructed and tested. It too will suffer from divergence of the quotient in frequency ranges where the cross power spectra are noisy and C_{23} and C_{14} have small amplitude. Here we avoid including C_{12}, which typically has significant cross correlated power in addition to the RFI signal, due either to polarized celestial flux or cross talk between the channels.

3. Noise and the Accuracy of the Cancellation

In this section we assess the importance of the Interference-to-Noise ratio. First we expand the autocorrelation spectra, keeping all cross terms, including those that average toward zero. Then $P_1(f)$ becomes

$$\begin{aligned} P_1(f) &= |g_A|^2 < |A_A|^2 > + \, |g_1|^2 < |I|^2 > + \, 2\mathrm{Re}\,[g_1 g_A^* < IA^* > + \, g_1 < IN_1^* >] \\ &+ \, 2\mathrm{Re}\,[g_A < AN_1^* >] \, + < |N_1|^2 > \end{aligned} \tag{14}$$

The complex correction spectra were described by equations 7, 8 and 9. Including the cross terms and noting that when

$$g_3 \text{ and } \mathrm{g}_4 \gg \mathrm{g}_1 \text{ and } \mathrm{g}_2 \tag{15}$$

and the interference power to noise power ratios

$$\begin{aligned} INR_3 &= \frac{|g_3|^2 < |I|^2 >}{< |N_3|^2 >} \approx |g_3|^2 \gg 1 \\ INR_4 &= \frac{|g_4|^2 < |I|^2 >}{< |N_4|^2 >} \approx |g_4|^2 \gg 1 \end{aligned} \tag{16}$$

then the correction CX_1 for $P_1(f)$ becomes:

$$\begin{aligned} CX_1 &\approx \frac{C_{13} C_{14}^*}{C_{34}^*} \\ &\approx |g_1|^2 < |I|^2 > + \, 2\mathrm{Re}\,[g_1 g_A^* < IA^* > + \, g_1 < IN_1^* >] \\ &+ \frac{g_A g_1^*}{g_3^*} < AN_3^* > + \frac{g_A^* g_1}{g_4^*} < A^* N_4 > \\ &+ \frac{g_1^*}{g_3^*} < N_1 N_3^* > + \frac{g_1}{g_4} < N_1^* N_4 > - \frac{|g_1|^2}{g_3^* g_4} < N_3^* N_4 > \end{aligned} \tag{17}$$

The terms in Eqn 17 involving I all appear in the power spectrum of Eqn 14, so that application of this correction CX_1 to $P_1(f)$ leads to a result with no residual contamination by the RFI:

$$\begin{aligned} P_1(f) - CX_1 &= |g_A|^2 < |A_A|^2 > + \, < |N_1|^2 > \\ &+ \, 2\mathrm{Re}\,[g_A < AN_1^* >] \, - \frac{g_A g_1^*}{g_3^*} < AN_3^* > - \frac{g_A^* g_1}{g_4^*} < A^* N_4 > \\ &- \frac{g_1^*}{g_3^*} < N_1 N_3^* > - \frac{g_1}{g_4} < N_1^* N_4 > + \frac{|g_1|^2}{g_3^* g_4} < N_3^* N_4 > \end{aligned} \tag{18}$$

The complete cancellation of the RFI terms is consistent with the concept that post correlation subtraction is equivalent to coherent subtraction of the RFI electric field $i(t)$ in the time domain, which should leave no trace of the RFI signal nor an increase in noise power. Provided the INR for the reference horn channels is substantially greater than the INR for the Parkes feed channels, the noise terms due to N_3 and N_4 will be smaller by a factor of order $\beta \sim \sqrt{INR_3/INR_1} \approx |g_3|/|g_1|$ than the normal noise contributions arising from N_1 plus the astronomical signal power. The terms in Eqn 18 other than $|g_A|^2 < |A_A|^2 >$ and $< |N_1|^2 >$, such as $(g_A^* g_1/g_4^*) < A^* N_4 > \propto |g_A| \beta^{-1} t^{-1/2}$ and $(g_1^*/g_3^*) < N_1 N_3^* > \propto \beta^{-1} t^{-1/2}$, average toward zero (in inverse proportion to the square root of the integration time), provided the noise in the signal channels is independent. The next higher order terms, which are not included in Eqn 18, have dependencies such as $< IN_1^* >< N_1^* N_4 > / g_4 < |I|^2 > \propto t^{-1}$ and $g_1^* < IN_3^* >< I * N_1 > / g_3^* < |I|^2 > \propto \beta^{-1} t^{-1}$, which converge toward zero faster than the dominant noise terms.

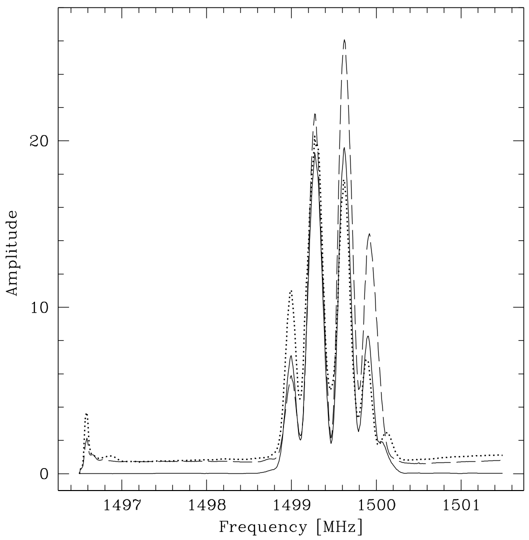

Fig. 2.— Scan averaged RFI spectra measured with the reference horn for scan SRT00502. Solid line is the cross power spectrum $C_{34}(f)$ defined in Eqn 5. Dashed and dotted lines indicate the auto-correlation spectra $P_3(f)$ and $P_4(f)$ in Eqn 3. The spectra have been passband calibrated using approximate gain curves for the 5 MHz filters. There are 512 frequency channels covering a 5 MHz band.

4. The test data

The astronomical dataset used in testing these algorithms is a dual linear polarization data stream from the CSIRO ATNF 64m telescope at Parkes in Australia. One configuration has two polarizations from the central beam of the the Parkes multibeam receiver (Staveley-Smith et al. 1996) and two polarizations from a reference horn aimed at an interfering source (Bell et al. 2000). A second configuration uses both polarizations from two beams of the Parkes Multibeam system, which are directed at slightly different areas of the sky. The center frequency of the datasets was 1499 MHz in each case.

The datasets we used are labeled SRT00501, SRT00502 and SRT00601 (Bell et al. 2000). The main interfering source is a NSW Government digital point-to-point microwave link. Examples of the time-averaged spectrum for the RFI signal are shown in Fig 2. Further details are available in the ACA databases (Australian_Communications_Authority 1998; Sarkissian 1999). The 4 signals were downconverted to base band and passed through 5 MHz low pass filters. Each signal was then digitized with 2-bit precision at a 20MHz sampling rate to achieve a factor 2 oversampling and recorded using the CPSR recorder (van Straten et al. 2000).

The data processing for these tests simulates a radio astronomy backend by computing power spectra and cross power spectra in software. The sampled voltages are treated in 8192 sample blocks (410μsec durations). Fourier transformation of each block yields a spectrum of 4096 independent complex coefficients.

Examples of 25 second averages of the power spectra are shown in Fig 3. The RFI spectrum is significantly different in the spectra for all four data channels. Since the data are $2\times$ oversampled, we kept the total power and cross power spectra of the lower 2048 complex Fourier coefficients. The RFI subtraction steps were performed on these spectra, as illustrated below, and subsequently the spectra were block averaged by 4 to keep power spectra of length 512 spectral channels for further calibration and display at 9.8 kHz resolution.

Examples of the cross power spectra are shown in Fig 4. Both C_{12} and C_{34} have broadband correlated power, which is clear in the integrated spectra both as a significant non-zero amplitude and as a well defined trend in phase across the band. The phase gradients that are clearly visible in the $C_{13}(f), C_{14}(f), C_{23}(f)$, and $C_{24}(f)$ spectra indicate differential path delays. The delay for $C_{13}(f), C_{14}(f)$ and for the channel numbers above 300 in $C_{23}(f)$, and $C_{24}(f)$ amounts to 15 time steps (0.75μsec). The steeper phase gradient for channels below 300 in $C_{23}(f)$, and $C_{24}(f)$ imply path delays of 60 time steps $= 3\mu$sec or a path length of $\sim$900m. Apparently, the frequency dependent reception pattern of the far sidelobes of the Parkes dish, coupled with complicated scatterers in the field, can lead to quite complex spectral dependence for the RFI. A strength of our method is the simplicity with which it handles this complex multipathing.

Examples of the complex correction spectra defined in Eqns 7-9 are shown in Fig 5. The spectra for Correction A and Correction B have zero phase over the spectral range where the signal is significantly non-zero, confirming that phase clo-

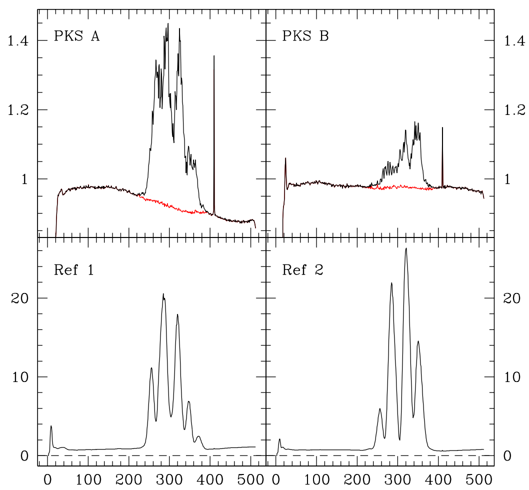

Fig. 3.— The Power Spectra PKS A= $P_1(f)$, PKS B= $P_2(f)$, Ref 1= $P_3(f)$, and Ref 2= $P_4(f)$ for scan SRT00502. These spectra are the averages of $\sim$25 seconds of data. A passband calibration has been applied to compensate for the gain dependence of the 5 MHz band limiting filters. The upper panels show the spectra both before and after cancellation.

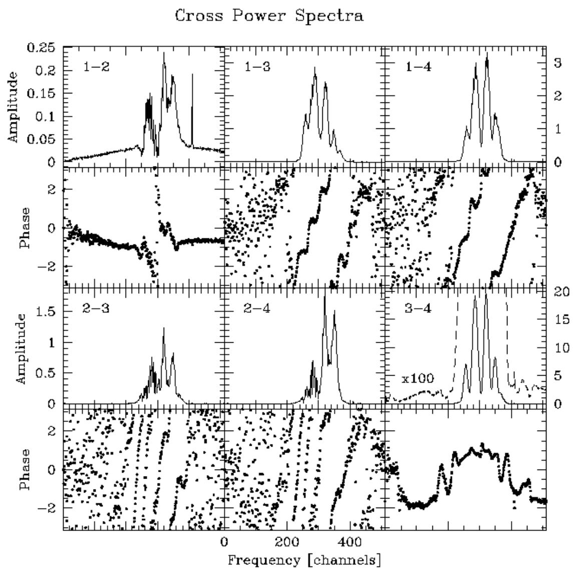

Fig. 4.— The Cross Power Spectra $C_{12}(f), C_{13}(f), C_{14}(f), C_{23}(f), C_{24}(f)$, and $C_{34}(f)$ for scan SRT00502. These spectra are the averages of $\sim$25 seconds of data. The spectrum $C_{34}(f)$ at lower right is also plotted with a rescaling of a factor of 100 in order to display the noise level away from the frequencies containing strong RFI.