THE ASTRONOMICAL JOURNAL, 120:3351–3361, 2000 December

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, Netherlands; fbriggs@astro.rug.nl

AND

J. F. BELL AND M. J. KESTEVEN
CSIRO Australia Telescope National Facility, P.O. Box 76, Epping, NSW 1710, Australia; jbell@atnf.csiro.au, mkesteve@atnf.csiro.au
Received 2000 June 16; accepted 2000 August 24

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.

Key words: instrumentation: detectors — methods: analytical

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.[1]

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 (Ellingson & Hampson 2000; Leshem, van der Veen, & Boonstra 2000; Smolders & Hampson 2000; Kewley et al. 2000),[2] (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, Bunton, & Bell 2000; Leshem & van der Veen 1999a, 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 Figure 1. For these experiments, we recorded digitally sampled base band 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,

[1] See also the presentation by S. Ellingson on Interference mitigation techniques, available at http://www.atnf.csiro.au/SKA/intmit/atnf/conf/.

[2] See also G. Hampson et al., The Adaptive Antenna Demonstrator, at http://www.nfra.nl/skai/archive/technical/index.shtml.

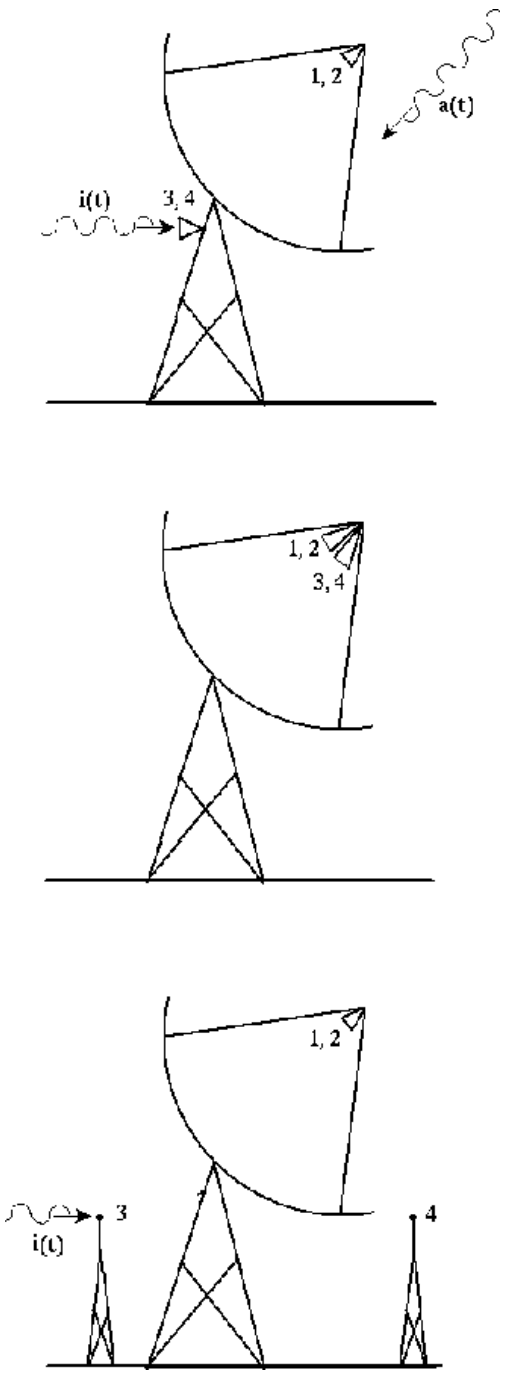

FIG. 1.—Three configurations applicable to the analysis in this paper. *Top*: The four base bands are recorded from the two polarizations of the Parkes telescope feed and two polarizations of a reference horn directed at the RFI source. *Center*: The four base bands correspond to two polarizations from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

rather than a commitment made on-line and permanently. Furthermore, the correlation method is effectively a coherent subtraction, since the correlation functions retain the information describing relative phase 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: $s_1(t) = H_A(t) \circledast a_A(t) + H_1(t) \circledast i(t) + n_1(t)$. Here $H_A(t)$ and $H_1(t)$ are impulse responses for the astronomy signal path and interference, respectively, and $\circledast$ 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 spill over 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 § 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:

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

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:

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

We use the superscript asterisk (*) to represent complex conjugation and the $\langle \ldots \rangle$ notation to signify averages over an integration time $t_{\rm int}$; in the tests we describe in § 4, we find the method is effective for $t_{\rm int}$ as long as ~ 1 s. We adopt a normalization where the power levels $\langle |A_A|^2 \rangle \approx \langle |I|^2 \rangle \approx \langle |N_1|^2 \rangle$ so that, for example, in data channel 1 the signal to noise ratio $\mathrm{SNR}_1 \sim |g_A|^2$, and the interference to noise ratio $\mathrm{INR}_1 \sim |g_1|^2$.

The goal of the RFI cancellation will be to form estimates of the $|g_1|^2 \langle |I|^2 \rangle$ and $|g_2|^2 \langle |I|^2 \rangle$ 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 $t_{\rm int}$, allowing us to separate the interference from the gain in expressions such as

$$\langle g_1 I g_1^* I^* \rangle = |g_1|^2 \langle |I|^2 \rangle . \tag{4}$$

In anticipation of the discussion of § 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:

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

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

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

However, the problem encountered with this approach is that $\langle |S_3|^2 \rangle$ is biased by the term $\langle |N_3|^2 \rangle$, 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 $|g_3|^2 \langle |I|^2 \rangle = \langle |S_3|^2 \rangle - \langle |N_3|^2 \rangle$ to use in the denominator of equation (6).

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

$$\begin{aligned} |g_1|^2 \langle |I|^2 \rangle &= \frac{g_1 g_3^* g_1^* g_4}{g_3^* g_4} \langle |I|^2 \rangle \\ &\approx \frac{C_{13} C_{14}^*}{C_{34}^*} , \end{aligned} \tag{7}$$

$$\begin{aligned} |g_2|^2 \langle |I|^2 \rangle &= \frac{g_2 g_3^* g_2^* g_4}{g_3^* g_4} \langle |I|^2 \rangle \\ &\approx \frac{C_{23} C_{24}^*}{C_{34}^*} , \end{aligned} \tag{8}$$

$$\begin{aligned} g_1 g_2^* \langle |I|^2 \rangle &= \frac{g_1 g_4^* g_2^* g_3}{g_3 g_4^*} \langle |I|^2 \rangle \\ &\approx \frac{C_{14} C_{23}^*}{C_{34}} . \end{aligned} \tag{9}$$

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 equations (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:

$$|g_1|^2\langle|I|^2\rangle \approx \frac{C_{13}\,C_{14}^*\,C_{34}}{\psi(f)+C_{34}\,C_{34}^*}\,, \tag{10}$$

$$|g_2|^2\langle|I|^2\rangle \approx \frac{C_{23}\,C_{24}^*\,C_{34}}{\psi(f)+C_{34}\,C_{34}^*}\,, \tag{11}$$

$$g_1\,g_2^*\langle|I|^2\rangle \approx \frac{C_{14}\,C_{23}^*\,C_{34}^*}{\psi(f)+C_{34}\,C_{34}^*}\,, \tag{12}$$

where $\psi(f)$ is the square of the power spectrum of the noise present in $C_{34}(f)$. Whenever $\psi(f)$ becomes small compared with $C_{34}(f)$, the expressions equations (10)–(12) revert to equations (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 autocorrelation spectra (eqs. [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 equations (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

$$\frac{C_{13}\,C_{24}}{C_{23}\,C_{14}} = \frac{g_1\,g_3^*\,g_2\,g_4^*}{g_2\,g_3^*\,g_1\,g_4^*} = 1 \tag{13}$$

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

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

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

$$g_3 \text{ and } g_4 \gg g_1 \text{ and } g_2 \tag{15}$$

and the interference power to noise power ratios

$$\mathrm{INR}_3 = \frac{|g_3|^2\langle|I|^2\rangle}{\langle|N_3|^2\rangle} \approx |g_3|^2 \gg 1\,,$$

$$\mathrm{INR}_4 = \frac{|g_4|^2\langle|I|^2\rangle}{\langle|N_4|^2\rangle} \approx |g_4|^2 \gg 1\,. \tag{16}$$

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

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

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

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

The complete cancellation of the RFI terms is consistent with the concept that postcorrelation 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 $\beta \sim (\mathrm{INR}_3/\mathrm{INR}_1)^{1/2} \approx |g_3|/|g_1|$ than the normal noise contributions arising from N_1 plus the astronomical signal power. The terms in equation (18) other than $|g_A|^2\langle|A_A|^2\rangle$ and $\langle|N_1|^2\rangle$, such as $(g_A^*\,g_1/g_4^*)\langle A^*N_4\rangle \propto |g_A|\,\beta^{-1}t^{-1/2}$ and $(g_1^*/g_3^*)\langle N_1\,N_3^*\rangle \propto \beta^{-1}t^{-1/2}$, 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 equation (18), have dependencies such as $\langle IN_1^*\rangle\langle N_1^*\,N_4\rangle/g_4\langle|I|^2\rangle \propto t^{-1}$ and $g_1^*\langle IN_3^*\rangle\langle I*N_1\rangle/g_3^*\langle|I|^2\rangle \propto \beta^{-1}t^{-1}$, which converge toward zero faster than the dominant noise terms.

4. THE TEST DATA

The astronomical data set used in testing these algorithms is a dual linear polarization data stream from the CSIRO ATNF 64 m telescope at Parkes in Australia. One configuration has two polarizations from the central beam of 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 data sets was 1499 MHz in each case.

The data sets 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 Figure 2. Further details are available in the Australian Communications Authority databases.[3] The four signals were down-converted to base band and passed through 5 MHz low-pass filters. Each signal was then digitized with 2-bit precision at a 20 MHz 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 (410 μs durations). Fourier transformation of each block yields a spectrum of 4096 independent complex coefficients.

Examples of 25 s averages of the power spectra are shown in Figure 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.

[3] Available at http://www.aca.gov.au/legal/legislation.htm. See also the presentation by J. Sarkissian on transmitter database visualization, available at http://www.atnf.csiro.au/SKA/intmit/atnf/conf/.

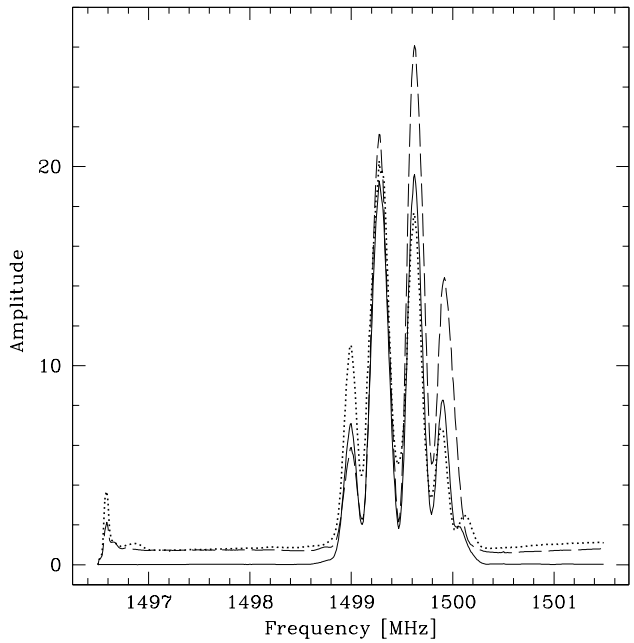

FIG. 2.—Scan-averaged RFI spectra measured with the reference horn for scan SRT00502. The solid line is the cross power spectrum $C_{34}(f)$ defined in eq. (5). Dashed and dotted lines indicate the autocorrelation spectra $P_3(f)$ and $P_4(f)$ in eq. (3). The spectra have been passband-calibrated using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

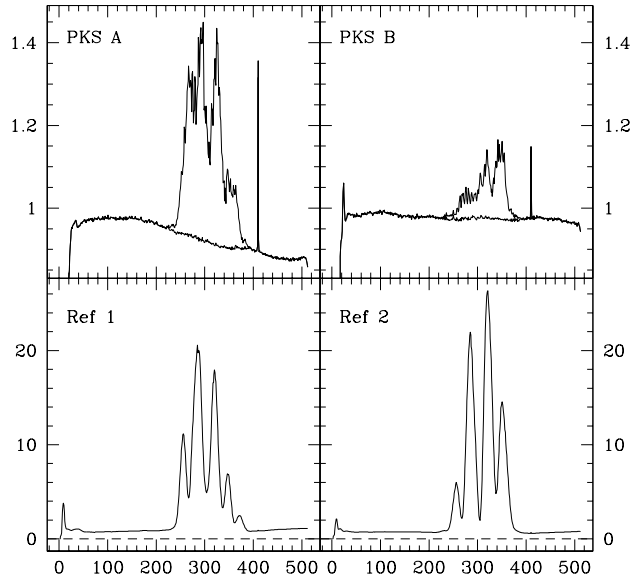

FIG. 3.—Power spectra PKS A = $P_1(f)$, PKS B = $P_2(f)$, Ref 1 = $P_3(f)$, and Ref 2 = $P_4(f)$ for scan SRT00502. These spectra are the averages of $\sim$25 s of data. A passband calibration has been applied to compensate for the gain dependence of the 5 MHz band limiting filters. The upper panels show the spectra both before and after cancellation. *See the electronic edition of the Journal for a color version of this figure.*

Examples of the cross power spectra are shown in Figure 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 (0.75 μs). The steeper phase gradient for channels below 300 in $C_{23}(f)$, and $C_{24}(f)$ imply path delays of 60 time steps = 3 μs or a path length of $\sim$900 m. Apparently, the frequency dependent reception pattern of the far side lobes 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 equations (7)-(9) are shown in Figure 5. The spectra for correction A and correction B have zero phase over the spectral range where the signal is significantly nonzero, confirming that phase closure applies.

Figure 6 shows averages of the amplitude closure quantity from equation (13), plotted as amplitude and phase across the spectrum. There is deterioration in the closure relation when the RFI signal is low, as expected.

Figure 2 compares three time-averaged power spectra representing the RFI signal sensed by the reference horn: power spectra for each of the two polarizations and one cross power spectrum. For use in correction schemes, the cross power spectrum $C_{34}(f)$ has the advantage that it is not contaminated by the positive bias of the receiver noise total power that is seen in the autocorrelation power spectra. In practice, the low-level correlated signal in $C_{34}(f)$, due to actual cross correlated broadband power or due to correlated quantization noise, may limit the accuracy of the cross power spectrum as an estimate of the RFI.

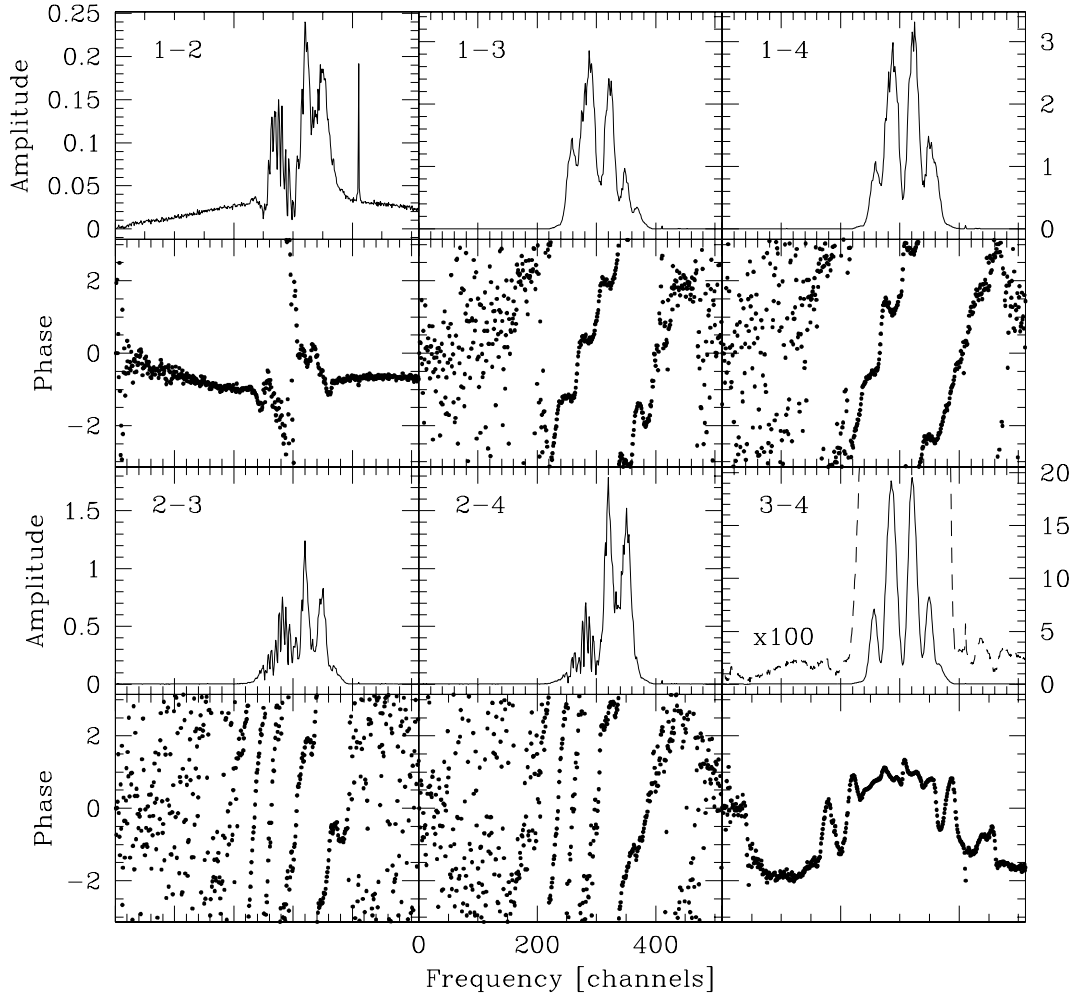

FIG. 4.—Cross power spectra $C_{12}(f)$, $C_{13}(f)$, $C_{14}(f)$, $C_{23}(f)$, $C_{24}(f)$, and $C_{34}(f)$ for scan SRT00502. These spectra are the averages of ~25 s of data. The spectrum $C_{34}(f)$ at the lower right is also plotted with a rescaling of a factor of 100 in order to display the noise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

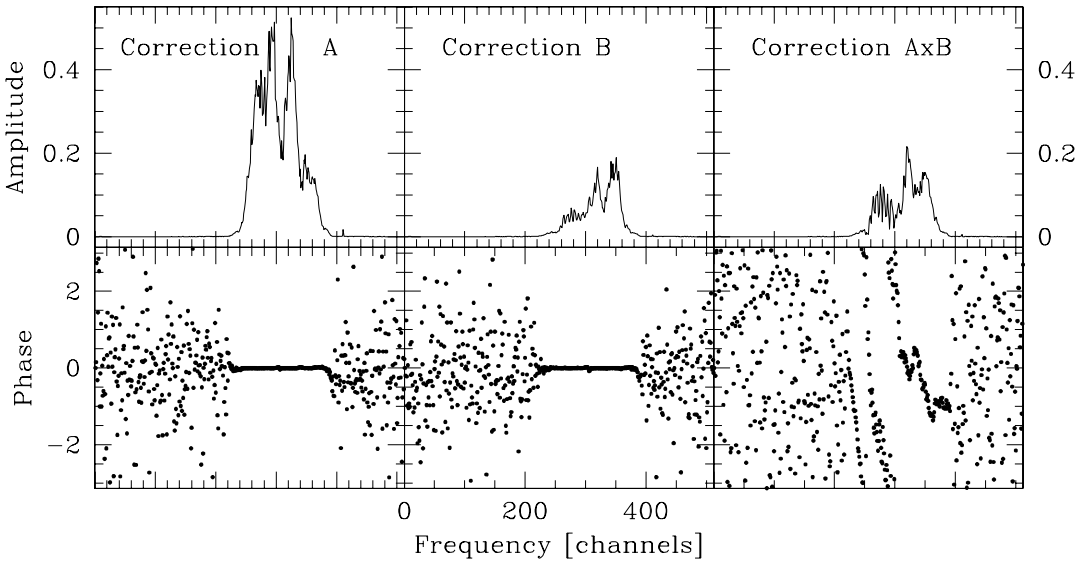

FIG. 5.—Complex correction spectra derived from eqs. (10)–(12) for scan SRT00502 for application to the Parkes spectra A, B, and A × B. These plots show the averages of ~25 s of data. The RFI subtraction was actually performed on 82 ms averages. A & B indicate the two polarizations.

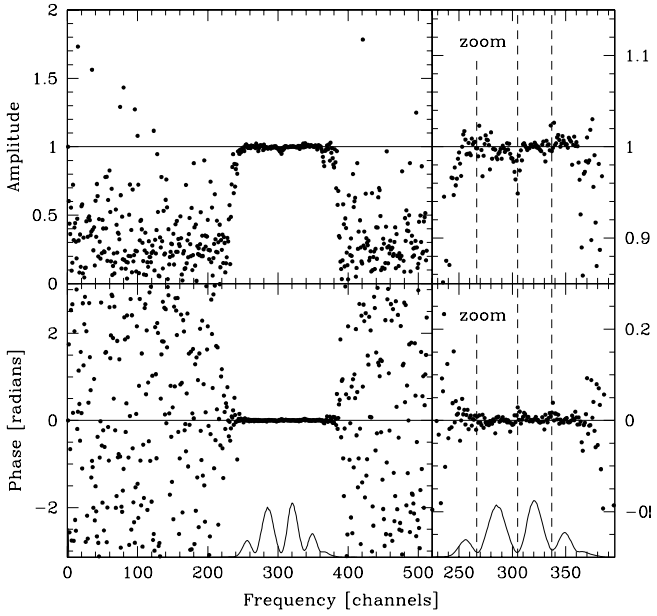

FIG. 6.—Complex closure quantity $C_{13}\,C_{24}/C_{23}\,C_{14}$, averaged for scan SRT00502. *Left*: full spectrum. *Right*: expanded scales. Reduced scale plots of the average RFI cross power spectrum from Fig. 2 are drawn in the bottom panels for comparison.

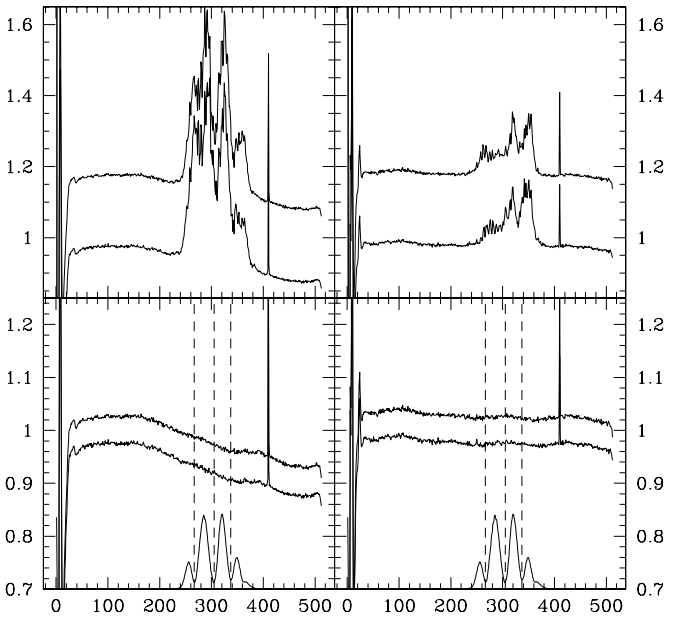

FIG. 7.—Comparison of two scans: SRT00501 and SRT00502. *Top Panels*: Uncorrected power spectra for the two Parkes polarizations. Spectra for SRT00501 are displaced vertically by 0.2 in amplitude. *Bottom Panels*: Corrected power spectra for the two Parkes polarizations. Spectra for SRT00501 are displaced vertically by 0.05 in amplitude. Reduced scale copies of the RFI cross power spectrum are included for reference with vertical dashed lines to indicate the minima in the RFI spectrum.

Figure 7 compares uncorrected and corrected spectra for two scans. For the purposes of display, the spectra have been crudely passband calibrated by dividing the spectra by gain templates formed from the scan average of the total power spectra of scan SRT00601, which was recorded while the sky frequency for the Parkes data channels was tuned off the RFI frequency. Since this gain template is in common to the processing for both scans SRT00501 and SRT00502, some of the common structure in the spectra in Figure 7 results from this common gain template.

The time dependence of the RFI sensed by the Parkes telescope is displayed in image format in Figure 8, side by side with the spectra after the RFI subtraction. These spectra received the same processing as described for the averages in Figure 7, with the additional step of subtracting a third order polynomial spectral baseline from each time step. This additional step was done to remove some faint

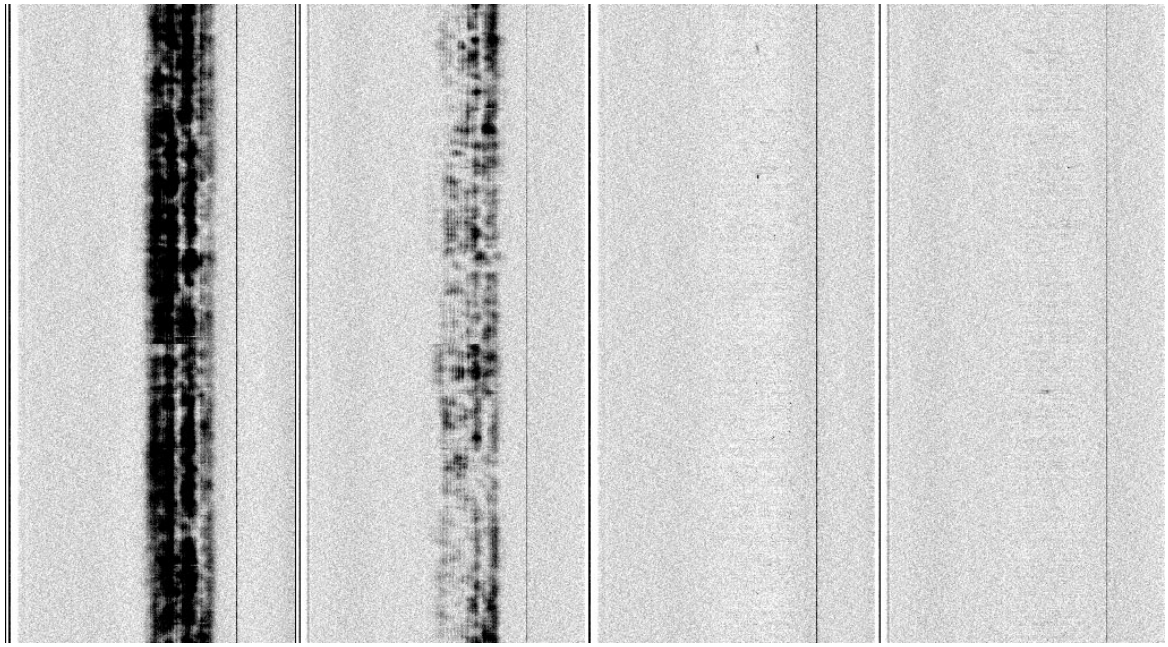

FIG. 8.—Dynamic power spectra over 564 time steps of 82 ms each (Scans SRT00501 and SRT00502). There are four spectra plotted in parallel with time increasing vertically. *Left*: The two Parkes polarizations prior to RFI subtraction. *Right*: The two Parkes polarizations after RFI subtraction. All spectra are passband calibrated to compensate for the frequency dependence of the 5 MHz filters. A third order polynomial spectral baseline was fitted to channels 30-235, 370-405, and 420-500 and subtracted for each time step.

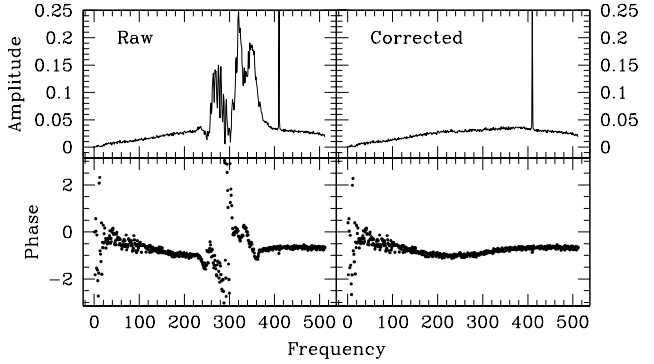

FIG. 9.—Raw and corrected Parkes A × B cross power spectra. *Left*: Raw, complex cross power spectrum. *Right*: Corrected complex spectrum.

variations in the total power level that occurred from one integration to the next.

The cross polarized spectrum from the Parkes A × B is shown in Figure 9. The figure includes the raw spectrum and the corrected spectrum after subtraction of the correction shown in Figure 5. Remaining in the corrected spectrum, there is a slow modulation of the power across the 5 MHz band, probably indicating that this power represents broad band noise signal that is scattered within the Parkes telescope structure with the same delay path lengths associated with traditional standing waves and path lengths of a few hundred feet.

The integration time over which the RFI corrections are applied is a critical parameter when the signal-to-noise ratio of the reference signal data path is low. If the integration time is too short, then some of the derived correction will be noise, and this will be folded into the resulting spectrum. On the other hand, if the integration time is long, the impulse response functions that couple the astronomy and RFI signals to the receiver will vary, and the mathematics in equation (4) and equations (7)-(9) will break down. Figure 10 shows a series of tests with a range of integration times ($t_{\rm int} \sim 8$ ms to 8 s) for application of the RFI subtraction. In all cases, the plots give the grand averages of the entire 25 s of the scan after application of the algorithm on the shorter data segments. The parameter ψ_o has an appropriate value in each case to produce a stable noise level and representation of spectral features in the frequency range away from the RFI contamination. For these data, the algorithm was most effective for integration times of only a $\sim$1 s or less. When treated on 8 s averages, substantial RFI remains unsubtracted.

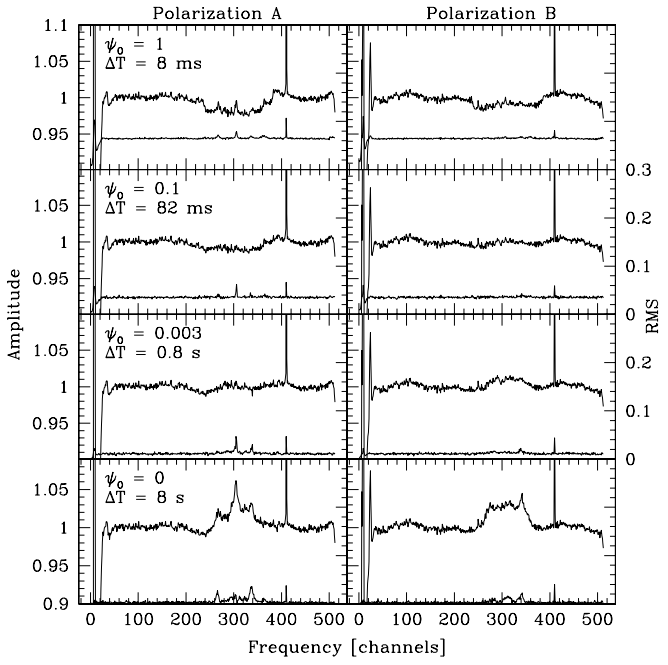

FIG. 10.—The effect of varying the time interval on which algorithm described in eqs. (10)–(12) is applied. The interval varies from 8 ms to 8 s in factors of 10. In each panel, the upper spectrum shows the corrected, calibrated, baselined spectrum; the lower spectrum is the rms scatter about the mean spectrum for each channel as a function of time. The typical value for rms should decrease by $10^{1/2}$ for each increase in factor 10 in integration time.

5. DISCUSSION OF LIMITATIONS

The wide range of delays for the scattered RFI led to problems in our initial experiments, which simulated an FX correlator with 2048 sample transforms. The RFI illustrated in Figure 4 has two strong components that are delayed by 15 and 60 time steps. A 60 time step delay is $\sim$3% of the time block being processed, so that the cross-correlation is not being performed on fully overlapped data streams. Increasing the window to 8192 was adequate to reduce the residuals to a level that was barely visible above the noise in the corrected spectra. An additional but less significant improvement resulted from applying an constant delay offset of 35 time steps to the reference signal at the input to the "FX correlator" to make it closer to the average of the principal delays in the Parkes data channels. A traditional time-domain lag correlation spectrometer would not encounter this problem, provided a sufficient number of lags are allocated to fully cover the range of delays experienced by the RFI.

In these tests at Parkes, there is a possibility that some uncancelled signal may be present because of a second transmitter operating at these same frequencies. The reference horn was pointed at the stronger, nearby transmitter, while the second transmitter, located at approximately three times the distance, can be scattered into the Parkes Telescope signal paths without being sensed by the reference horn.

The coarse digitization of the RFI reference signals will eventually form a limitation to the precision of the subtraction. Crude quantization generates an artificial noise floor throughout the spectrum, effectively scattering power out of the narrow band RFI. Since both polarizations from the reference horn are recorded at relatively high signal to noise ratio, the quantization noise is also correlated between the two data channels, so that there is corruption of the cross power spectrum as well as for the autocorrelation spectra. (The implication is that the term $\langle N_3 N_4 \rangle$ in equation (5) will not average to zero with increased integration.)

6. THE TOXICITY TEST

A crucial requirement of an RFI subtraction algorithm is that it must leave the astronomical signal of interest unaltered. To test the current method, we added simulated galaxy signals to the two Parkes input data streams. Independent Gaussian noise was filtered with a double-horned galaxy profile and the instrumental IF filter passbands and then injected into the data pipeline just before the correlation stage. Figure 11 shows a comparison between the

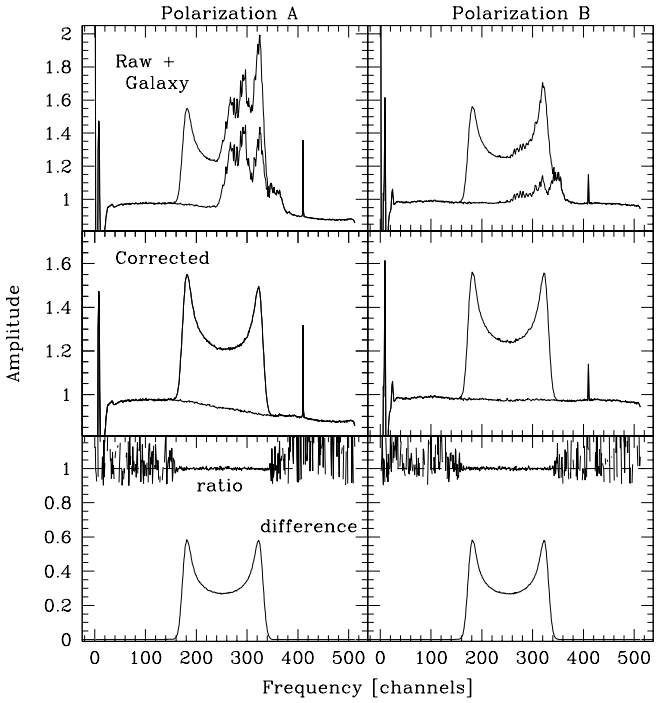

FIG. 11.—Survival of an injected synthetic galaxy signal in the astronomy channels through the RFI subtraction process. *Top*: Raw spectra—with and without the synthetic galaxy signal. *Center*: Corrected spectra—with and without the synthetic galaxy. *Bottom*: Difference between the corrected spectra and the ratio between the input synthetic spectrum and the difference spectrum. *See the electronic edition of the Journal for a color version of this figure.*

RFI corrected spectra both with and without the galaxy. To highlight the difference between the injected signal and the output after RFI subtraction, the bottom panel shows (1) the difference between the output with and without the added signal and (2) the ratio of this difference to the synthetic galaxy profile added to the input. The rms deviation of the ratio about unity is 0.005 for polarization A and 0.006 for polarization B. No systematic deviations are seen across the band, other than rises in noise level at the edges where the galaxy profile is approaching zero at the edge of the profile. The conclusion is that this method does no systematic harm to the astronomical signals.

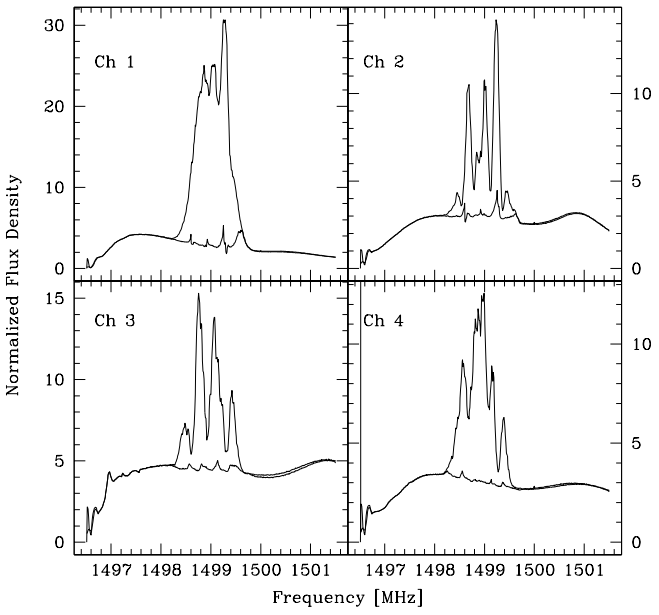

FIG. 12.—Autocorrelation spectra for both polarizations of two feeds of the Parkes multibeam system before and after RFI subtraction. An attempt was made to apply a passband calibration using the same passbands determined from the scan SRT006_01 for the reference horn experiment. The data were treated in ~82 ms averages, which in turn were integrated over the 20 s duration of the scan. *See the electronic edition of the Journal for a color version of this figure.*

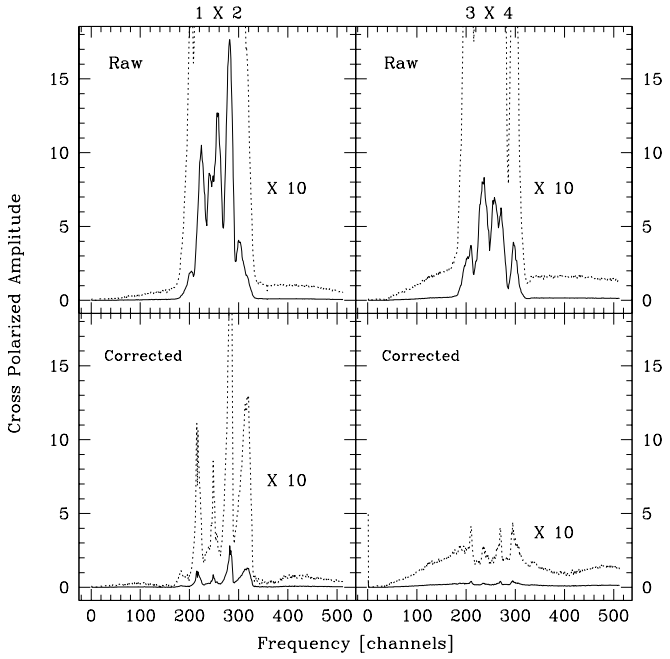

FIG. 13.—Cross polarization spectra for two feeds of the Parkes multibeam system before and after RFI subtraction for scan SRT00108. An attempt was made to apply a passband calibration using the same passbands determined from the scan SRT006_01 for the reference horn experiment. The data were treated in ~82 ms averages, which in turn were integrated over the 20 s duration of the scan. Dotted lines show the spectra amplified by a factor 10.

7. THE PARKES TWO FEED EXPERIMENT

The mathematical description (eqs. [3] to [12]) can be equally well applied to a case that uses a second feed from the Parkes Telescope as the "reference horn." Both feeds may receive astronomical signals, but since the feeds point different directions in the sky, these are independent astronomical signals, which will not correlate and therefore will not be subtracted from each other by this algorithm. The two feeds do sense the same RFI signal $i(t)$, although through different scattering paths. This is sufficient commonality that the cross power spectral approach should permit each feed to serve as the reference antenna for the other. Similar experiments have been reported.[4]

Figure 12 shows a comparison between the autocorrelation spectra measured for the Parkes two-feed experiment, before and after RFI subtraction. There is noticeable difference among the four channels in the effectiveness of the RFI subtraction. This probably results from the bias created by the broadband polarized flux or correlated noise in the

[4] See also the presentation by B. Sault on Cross-correlation approaches to interference elimination, available at http://www.atnf.csiro.au/SKA/WS and that by L. Kewley, R. Sault, & R. Ekers on Interference excision using the Parkes multibeam receiver, available at http://www.atnf.csiro.au/SKA/intmit/atnf/conf/.

cross polarized spectra 3×4 that is being used as the "reference" for data channels 1 and 2. As shown in Figure 13, this noise floor is higher in feed 2 (INR $\sim$ 351) than in feed 1 (INR $\sim$ 100:1), causing the RFI template spectra derived from the cross power spectra to be less faithful, which in turn leads to larger error in the correction spectra to be subtracted from feed 1. For comparison, the reference horn spectrum $C_{34}(f)$ in Figure 4 has INR $\sim$ 1000:1

8. CONCLUSIONS

RFI subtraction can be performed using cross power spectra between the astronomy data channels and RFI "reference" channels. In principle, the reference channel can also be an astronomy channel provided it carries an astronomy signal that is uncorrelated with the astronomy in the channel that is being corrected.

The tests made at Parkes demonstrate that a specifically designed reference sensor provided a higher signal-to-noise ratio reference signal—and consequently cleaner cancellation—than that obtained from a second horn feed at the Parkes Telescope focus, whose principal function is to illuminate the Parkes dish.

A refinement will be to implement this scheme using two reference antennas that are spatially separated (as in the lower diagram of Fig. 1) in order to avoid correlated noise contributions, while still obtaining as clear and stable path to the RFI source as possible. The cross power spectra from the two spatially separated antennas would form an optimal RFI reference spectrum C_{34} for use in the equations (10)–(12). To avoid problems with differential delay causing loss of coherence in the reference signal, the spectrometer would need to operate with spectral resolution $\Delta f = (\Delta\phi/2\pi)c/L \ll c/L$, where $\Delta\phi$ is the allowable phase rotation across a spectrometer channel and L is the spatial separation of the sensors. The 2.4 kHz spectrometer resolution emulated in software for the study reported here would allow spatial separations of up to 1200 m, if $\Delta\phi$ is required to be less than 0.02π radians.

There are a number of advantages to performing this type of "post-correlation" RFI subtraction:

1. Provided the required correlation products are recorded (i.e., the on-line system is capable of recording correlation functions with a sufficiently large number of delay lags), the RFI subtraction can be performed off-line, where it remains an option in the data reduction path, rather than a commitment made on-line and permanently.
2. The method is not vulnerable to the effects of sporadic RFI, which hurt many algorithms that have an initialization period while they acquire the RFI signal and optimize their cancellation parameters.
3. Nor is the result influenced by changes in beam shape during adaptive nulling.
4. The correlation method is effectively a coherent subtraction, since the correlation functions retain the information describing relative phase between the RFI entering in the astronomy data stream and the RFI entering the reference antenna. We showed in § 3, this means that the RFI noise power is largely subtracted, leaving only system noise.
5. Generalization of the method to an array of telescopes is straightforward but demands additional correlator capacity. If there are two reference signal sensors, labeled "x" and "y," that sense negligible astronomical signal, then their cross power spectrum $C_{xy}(f)$ containing a high INR signal can be used to correct any other power spectrum $C_{ij}(f)$ through the closure relation $C_{ix}C^*_{jy}/C^*_{xy}$. The i, j indices can denote orthogonal or parallel polarizations drawn from any combinations of antennas in the array or auto correlation, when $i = j$.
6. A modification of the method can be applied to pulsar data streams in which a digital correlator replaces the narrowband filter bank used in compensating for pulse dispersion. One possible implementation would construct cross power "coupling spectra" $X(f)$ that are valid for the time interval $t_{\rm int}$, during which the "g factors" of equation (1) are stable. The coupling spectrum is

$$X_{13}(f) = \frac{g_1^* g_4}{g_3^* g_4} \approx \frac{C_{14}^*}{C_{34}^*},$$

where C_{14} and C_{34} are averaged for up to 1 s as appropriate. Then the correction CX_1 can be computed and applied to cancel RFI in measurements of $P_1(f)$ on shorter timescales:

$$CX_1 = |g_1|^2\langle |I|^2\rangle = X_{13} g_1 g_3^* \langle |I|^2\rangle \approx X_{13}\langle S_1 S_3^*\rangle ,$$

where $\langle S_1 S_3^*\rangle = C_{13}$. Alternatively, one could construct the corrected time sequence $s_1(t)$ by subtracting the correction

$$SX_1 = g_1 I = X_{13}^* g_3 I \approx X_{13}^* S_3$$

from $S_1(f)$ and inverse Fourier transforming to obtain an RFI cancelled version of $s_1(t)$. Computationally efficient schemes could be implemented that include coherent dedispersion Hankins 1974 in the same transform operations as the RFI cancellation.
7. The method can be generalized to removal of solar radiation whose multi-path scattering effects give rise to the spectral "standing wave" problem. The important difference that the Sun generates a dual polarized signal with a variable polarized component; this will necessitate a processing path more akin to the subspace decomposition (Leshem et al. 2000),[5] in order to identify orthogonal components of the solar "RFI" signals.
8. The present generation of digital correlators, which typically operate with single bit to 9 level precision, could implement this method for use in moderate levels of RFI for testing and astronomical observation in the near future.

The disadvantages of the method are: (1) The data rates will be high, since the method requires a full-scale cross correlator, preferably with multi-bit precision to accept large SNR RFI reference signals, that must dump spectra after relatively short integrations of less than $\sim$1 s. Of course, these data rates are lower than recording the full base band, but they are substantially higher than a real-time adaptive filter approach, which would allow long spectral integrations once the RFI has been canceled. (2) When the interference is much stronger than the astronomical signal, a 1 or 2 bit sampler is "captured" so that the data stream consists primarily of ± 2, and the zero crossings are determined by the phase of the interferer. Under these condi-

[5] See also S. Ellingson on interference mitigation techniques, available at http://www.atnf.csiro.au/SKA/intmit/atnf/conf/.

tions, the astronomical data will be largely lost. These applications will require correlators with greater digital precision.

The case where multiple interferers occupy the same frequency band will require a greater number of sensors and more correlator capacity devoted to processing the RFI signals, as laid out in the analyses of Sault (1997) and Ellingson (1999).

The authors are grateful to W. van Stratten, M. Bailes, S. Anderson, S. Ellingson, R. Sault, P. Perillat, R. Ekers, J. Bunton, L. Kewley, M. Smith, and P. Sackett for helpful comments and discussion. F. B. is grateful to the ATNF in Epping, NSW, the Department of Astronomy at OSU, Columbus, OH, and to the IAS, Princeton, NJ, for their hospitality while this work was done.

REFERENCES

Barnbaum, C., & Bradley, R. 1998, AJ, 116, 2598
Bell, J. F., et al. 2000, Proc. Astron. Soc. Australia, submitted
Ekers, R. D., & Bell, J. F. 2000, in IAU Symp. 199, The Universe at Low Radio Frequencies (San Francisco: ASP), in press
Ellingson, S. W., Bunton, J. D., & Bell, J. F. 2000, in Proc. SPIE 4015, 400
Ellingson, S. W., & Hampson, G. A. 2000, IEEE Trans. Ant. Propag., submitted
Hankins, T. H. 1974, A&AS, 15, 363
Kewley, L., Sault, R. J., Bell, J. F., Gray, D., Kesteven, M. J., & Ekers, R. D. 2000, in preparation
Leshem, A., & van der Veen, A. J. 1999a, in IEEE Workshop on Signal Processing Advances in Wireless Communications 1999 (Piscataway, NJ: IEEE), 374
Leshem, A., & van der Veen, A. J. 1999b, in IEEE Workshop on Higher Order Statistics (Los Alamitos, CA: IEEE Comp. Soc.), 25
Leshem, A., van der Veen, A. J., & Boonstra, A. J. 2000, ApJS, in press
Smolders, B., & Hampson, G. A. 2000, IEEE Trans. Ant. Propag., submitted
Staveley-Smith, L., et al. 1996, Proc. Astron. Soc. Australia, 13, 243
van Straten, W., Britton, M. C., Bailes, M., Anderson, S. B., & Kulkarni, S. 2000, in ASP Conf. Ser. 202, Pulsar Astronomy—2000 and Beyond, ed. M. Kramer, N. Wex, & R. Wielebinski (San Franciso: ASP), 238