Mon. Not. R. Astron. Soc. **322**, 486–498 (2001)

The H I Parkes All Sky Survey: southern observations, calibration and robust imaging

D. G. Barnes,[1,2★†] L. Staveley-Smith,[1] W. J. G. de Blok,[1] T. Oosterloo,[3], I. M. Stewart,[4] A. E. Wright,[1] G. D. Banks,[5] R. Bhathal,[6] P. J. Boyce,[7] M. R. Calabretta,[1] M. J. Disney,[5] M. J. Drinkwater,[8] R. D. Ekers,[1] K. C. Freeman,[9] B. K. Gibson,[2] A. J. Green,[10] R. F. Haynes,[1] P. te Lintel Hekkert,[1] P. A. Henning,[11] H. Jerjen,[9] S. Juraszek,[10] M. J. Kesteven,[1] V. A. Kilborn,[8] P. M. Knezek,[12] B. Koribalski,[1] R. C. Kraan-Korteweg,[13] D. F. Malin,[14] M. Marquarding,[1] R. F. Minchin,[5] J. R. Mould,[9] R. M. Price,[11] M. E. Putman,[9] S. D. Ryder,[14] E. M. Sadler,[10] A. Schröder,[15] F. Stootman,[6] R. L. Webster,[8] W. E. Wilson[1] and T. Ye[1]

[1]*Australia Telescope National Facility, CSIRO, PO Box 76, Epping, NSW 1710, Australia*
[2]*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia*
[3]*Netherlands Foundation for Research in Astronomy, PO Box 2, 7990 AA Dwingeloo, the Netherlands*
[4]*Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH*
[5]*Department of Physics and Astronomy, University of Wales, Cardiff CF24 3YB*
[6]*Department of Physics, University of Western Sydney Macarthur, Campbelltown, NSW 2560, Australia*
[7]*Department of Physics, University of Bristol, Tyndall Avenue, Bristol BS8 1TL*
[8]*School of Physics, University of Melbourne, VIC 3010, Australia*
[9]*Research School of Astronomy and Astrophysics, Australian National University, Weston Creek, ACT 2611, Australia*
[10]*School of Physics, University of Sydney, NSW 2006, Australia*
[11]*Institute for Astrophysics, University of New Mexico, Albuquerque, NM 87131, USA*
[12]*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[13]*Departamento de Astronomia, Universidad de Guanajuato, Guanajuato, GTO 36000, Mexico*
[14]*Anglo Australian Observatory, PO Box 296, Epping, NSW 1710, Australia*
[15]*Observatoire de la Côte d'Azur, Nice, 06304 Nice Cedex 4, France*

Accepted 2000 October 2. Received 2000 September 28; in original form 2000 June 1

ABSTRACT

The acquisition of H I Parkes All Sky Survey (HIPASS) southern sky data commenced at the Australia Telescope National Facility's Parkes 64-m telescope in 1997 February, and was completed in 2000 March. HIPASS is the deepest H I survey yet of the sky south of declination $+2°$, and is sensitive to emission out to $170\,h_{75}^{-1}$ Mpc. The characteristic root mean square noise in the survey images is 13.3 mJy. This paper describes the survey observations, which comprise 23 020 eight-degree scans of 9-min duration, and details the techniques used to calibrate and image the data. The processing algorithms are successfully designed to be statistically robust to the presence of interference signals, and are particular to imaging point (or nearly point) sources. Specifically, a major improvement in image quality is obtained by designing a *median-gridding* algorithm which uses the median estimator in place of the mean estimator.

Key words: instrumentation: detectors – methods: observational – methods: statistical – techniques: image processing – surveys – radio lines: galaxies.

1 INTRODUCTION

The H I Parkes All Sky Survey (HIPASS) is a survey for neutral hydrogen (H I) emission from extragalactic objects in the radial velocity range $-1280 < cz < 12\,700\,\mathrm{km\,s^{-1}}$, over the entire southern sky south of declination $+2°$, with an effective

★ E-mail: dbarnes@swin.edu.au
† Present address: Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia.

© 2001 RAS

Table 1. Parameters for the H I Parkes All Sky Survey. The H I mass limit is calculated for a galaxy profile width of $200\,\mathrm{km\,s^{-1}}$.

Sky coverage	$\delta < 2°$
Integration time per beam	450 s
Average system temperature	19 K (35 Jy)
Central beam efficiency	63 per cent
Central beam FWHM	14.0 arcmin
Average beam FWHM	14.3 arcmin
Velocity range	$-1280 < cz < 12\,700\,\mathrm{km\,s^{-1}}$
Channel separation	$13.2\,\mathrm{km\,s^{-1}}$ at $z = 0$
Velocity resolution	$18.0\,\mathrm{km\,s^{-1}}$
3σ positional accuracy	3 arcmin
3σ detection limit	$40\,\mathrm{mJy\,beam^{-1}}$
3σ H I mass limit	$10^6 \times d^2_{\mathrm{Mpc}}\,\mathrm{M}_\odot$

integration time of 450 s per beam. The survey parameters are shown in Table 1. HIPASS observations utilize the new 21-cm Multibeam receiver installed at the prime focus of the Parkes 64-m radio telescope,[1] began in 1997 February, and were completed in 2000 March.

The scientific potential of the survey is tremendous: the H I mass function for the nearby Universe will be determined better than ever before. HIPASS will provide completely new information on the distribution of galaxies, the cosmological density parameter, the space density of rare and optically invisible galaxies, and on group and supercluster dynamics. New results from the all-sky survey have included the discovery of a leading arm to the Magellanic system (Putman et al. 1998) and the identification of 10 new members of the Centaurus A group of galaxies, adding to the 28 already known within the area surveyed (Banks et al. 1999). Present work is directed towards the generation of a complete catalogue of southern sky H I sources.

This paper describes the observing, calibration and imaging techniques used for the survey. In Section 2 the Multibeam hardware and the survey observing technique are described. In Section 3 the algorithms designed for the on-line processing of the data are described; these algorithms are directed at robust bandpass removal and flux calibration, residual baseline subtraction, spectral smoothing and Doppler tracking. Technical details on the implementation of some of the algorithms are given in Barnes (1998). Images of the H I sky are generated using the median-gridding technique introduced in Section 4, and are post-processed into final survey images using the techniques given in Section 5. The primary survey product is a set of data cubes, measuring $8° \times 8°$ on the sky, and having 1024 individual maps of H I emission at mean increments of $13.4\,\mathrm{km\,s^{-1}}$ covering the velocity range given in Table 1.

2 OBSERVING

2.1 The Multibeam system

The Parkes 21-cm Multibeam system comprises a cooled, 13 beam receiver and a digital correlator. The receiver was installed at the prime focus of the Parkes 64-m radio telescope on 1997 January 21. The 13 circular feed horns of the Multibeam are positioned in a hexagonal arrangement on the focal plane, with a single central feed, and inner and outer rings of six horns each (see Fig. 1; see also Staveley-Smith et al. 1996, Wilson et al. 1997 and Sinclair et al. 1997). Physically, the 13 feed horns are identical, having diameters of 240 mm at the focal plane, narrowing in steps towards the receiver end. The 26 receivers (two per feed horn) are sensitive to orthogonal linear polarizations of radiation in the frequency range 1.27–1.47 GHz. The dewar is cooled to ~80 K surrounding the feed horns, and to less than 20 K at the low-noise amplifiers. At 1.42 GHz, the mean beamwidth (full width at half power) is 14.3 arcmin, and the peak responses of the 13 beams are projected on to the sky with typical separations of 30 arcmin, or just over two beamwidths.

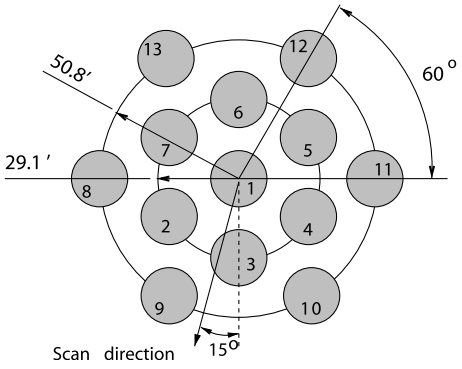

Figure 1. Multibeam receiver configuration on the sky. The radii of the inner and outer rings of beams are 29.1 and 50.8 arcmin respectively.

The Multibeam correlator has an instantaneous bandwidth of 64 MHz divided into 1024 channels for all 26 receivers. This wide bandwidth offers a velocity range of −1280 to $12\,700\,\mathrm{km\,s^{-1}}$ (when the receivers are tuned to 1394.5 MHz), and a mean channel spacing of $13.4\,\mathrm{km\,s^{-1}}$. The correlator chip is identical to that used in the new Arecibo system (Canaris 1993).

2.2 Scanning the sky

Observations for HIPASS commenced on 1997 February 27, and are taken by scanning the telescope in Dec. strips of length 8°. The Parkes telescope has an alt-azimuth mount, and while in principle the receiver could be rotated during each scan to compensate for the apparent rotation of the beam pattern on the sky, this is not done, as it has a detrimental effect on spectral baseline stability. Therefore, prior to each scan, the receiver assembly is rotated by the parallactic angle at the scan mid-point, plus another 15°, to obtain *approximately* uniform coverage of the sky (see Fig. 1 and also Staveley-Smith 1997). The footprint of the receiver on the sky is ~1.°7, so that each scan maps out an $8° \times 1.7°$ area of sky with reasonably uniform coverage. The exact coverage depends on the rate of change of parallactic angle (η) during the scan. This is highest near the zenith ($\mathrm{d}\eta/\mathrm{d}\delta = \sin\eta\cot z$, where δ is Dec. and z is zenith angle). Generally, observations within 20° of the zenith have been avoided, as significant changes in parallactic angle ($|\Delta\eta| > 10°$) become possible.

To obtain full coverage of the sky at full sensitivity, subsequent scans are displaced by 7 arcmin in RA, as shown in Fig. 2. Since the Nyquist rate is $\lambda/2D = 5.7$ arcmin (where λ is the observing wavelength and D is the telescope diameter, in the same units), this scan displacement ensures that the sky is mapped at close to the Nyquist rate by *each* of the 13 beams, and gives sufficient redundancy that data can be edited by purely automated, statistically robust procedures. The scan rate is $1°\ \mathrm{min^{-1}}$, and the correlator signal is recorded every 5 s. Measurements of the widths

[1] The Parkes telescope is part of the Australia Telescope which is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO.

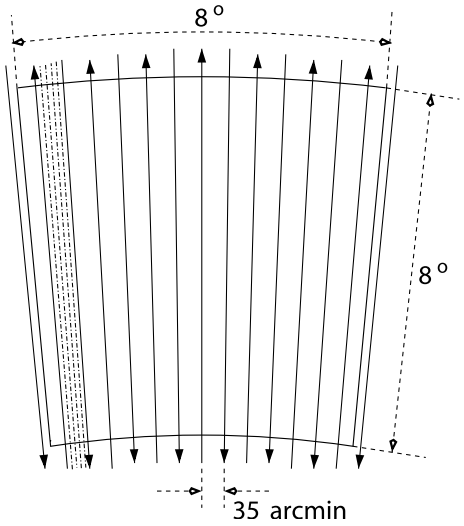

Figure 2. HIPASS scan pattern on the sky for an $8° \times 8°$ field. Five sets of 15 scans each are acquired, with adjacent scans within each set separated by 35 arcmin, and each set displaced from the previous set by 7 arcmin in RA. The solid lines with arrows depict one set of 15 scans; the arrows indicate the direction of each scan. The final scan density, for a single beam of the Multibeam receiver, is shown by the dot-dashed lines near the left-hand edge of the field.

of the beams give full widths at half power of 14.0 arcmin for the central beam, 14.1 arcmin for beams in the inner ring, and 14.5 arcmin for beams in the outer ring. The mean observing beamwidth is therefore 14.3 arcmin, but is extended to 14.7 arcmin in the Dec. direction because data are acquired while the telescope is scanning. The total integration time of the survey is 7×10^3 s deg^{-2}, or 450 s beam^{-1}.

Although the Dec. scans are separated by 7 arcmin, the observations are made in five separate sets, with the scans in each set spaced by 35 arcmin. Care was taken to make the observations of a given part of the sky at well-separated times so that interference, if present, did not corrupt all data for that direction. This seems to have worked well, in general, with little sign of any solar or narrowband interference in the final images, except in a few unfortunate cases where more than ~60 per cent of the data have been acquired during the daytime, over a period of just a few days. In terms of narrowband interference, only the 11th harmonic of the 128-MHz sampler clock at 1408 MHz ($cz \simeq 2640$ km s^{-1}) consistently appears at moderate levels in the final cubes. This harmonic was removed in early 1999 by enclosing the correlator in a Faraday cage.

2.3 Data acquisition

The Multibeam correlator is programmed to write data directly into a disk file. The correlator cycle time is 5 s, and spectra are written for each beam and polarization at the end of each cycle. There are two polarizations per beam, so a total of 26 spectra are written each cycle, each with 1024 channels. Each channel value is stored as a single-precision floating point number, occupying four bytes of storage. Thus the HIPASS raw data rate is 104 kb cycle^{-1}, or 1.2 Mb min^{-1}. The correlator file is closed and reopened each cycle to ensure that the very latest data can be read for near real-time processing. All HIPASS data are archived in their unprocessed state.

Coordinate information from the telescope is generally not synchronized with data from the correlator. As a result, coordinates in the data files refer to the position of the telescope a few seconds before the mean time-stamps applied to the data. A linear interpolation of telescope positions has been applied throughout the survey, typically resulting in corrections of a few arcmin in the Dec. direction. A comparison of the positions of more than 200 sources selected from the PKSCAT90 2700-MHz continuum source catalogue (Wright & Otrupcek 1990) with positions measured from the final survey images exhibits a standard deviation of 0.8 arcmin in RA, and slightly less than this in Dec. There appear to be no significant systematic position offsets.

3 SPECTRAL PROCESSING

3.1 Requirements

Spectra generated by the Multibeam correlator contain the usual structure seen in spectra from other H I instruments: a bandpass spectrum which is the sum of the sky, ground and receiver temperatures multiplied by the product of the filters in the receiver chain; superimposed on this is a noise whose amplitude depends inversely on the square root of integration time and channel bandwidth; occasional baseline ripple whose amplitude is a complicated function of the telescope geometry and the location and strength of 20-cm continuum sources in the sky; and other internal effects such as *ringing* due to strong, sharply peaked sources.

Any H I line emission – whose amplitude is typically a few thousandths of that of the system bandpass spectrum – lies on top of the structure described above. For HIPASS, algorithms were designed: to remove the system bandpass spectrum, to calibrate the residual spectrum, to remove any remaining baseline (DC) offset, to suppress strong ringing effects, and to shift the resultant spectrum from the topocentric observing frame to a fixed frame of reference. Since radio-frequency interference (RFI) had the potential to seriously contaminate a substantial fraction of the survey data, the algorithms were developed to be statistically robust to a moderate fraction of bad data. Furthermore, as the Parkes beam is significantly larger on the sky than the typical extragalactic H I source, the algorithms were designed specifically to identify compact emission regions.

3.2 Robust bandpass estimation

The dominant component in a raw correlator spectrum is the system bandpass spectrum, which varies with time because of, e.g., slow fluctuations in the physical temperature of the receivers, and external influences, such as atmospheric conditions which of course change with pointing direction. Traditionally, bandpass removal is accomplished by observing in a *signal/reference* mode. In this mode, an extended integration (e.g., 200 s) is acquired while the telescope tracks the target position, yielding a signal spectrum. A second integration, usually of the same duration, is then acquired while the telescope is pointed toward a nearby position (hopefully) free of line or continuum emission, yielding a reference spectrum. The reference spectrum is assumed to be a good estimate of the bandpass spectrum, and is subsequently removed from the signal spectrum by division. For longer on-source integrations times, this process is repeated many times, and the quotient spectra averaged. It is normal to spend as much time off-source as on-source for this observing mode, since division by the (noisy) reference spectrum increases the noise in the resultant

quotient spectrum accordingly. It is also normal to select a reference position such that the sidereal telescope track with respect to the ground is similar to the track for the signal observation, in order to re-create the ground spill-over contribution to the bandpass spectrum as well as possible.

For HIPASS, where the telescope is actively scanned across the sky at $1^\circ\,\mathrm{min}^{-1}$, the signal/reference method of bandpass removal needs to be modified. The basic method for removing the bandpass from a selected (*target*) HIPASS spectrum is first to *estimate* the shape of the bandpass spectrum at the time the spectrum was acquired, and then to divide the target spectrum by the bandpass estimate. The bandpass estimate is determined from a set of earlier and/or later spectra observed by the same feed of the Multibeam receiver. These *reference* spectra are selected individually for the target spectrum, and must satisfy a number of criteria to be suitable. Suitable reference spectra should be independent measures of the H I sky to that of the target spectrum, but should be acquired nearby in time so that temporal variations of the bandpass are kept to a minimum. Since the bandpass estimate is made from a number of spectra, the increase in spectral noise caused by bandpass removal can be made negligible.

Actual estimation of the bandpass is done by taking a channel-by-channel median of the reference spectra, i.e., the bandpass estimate is the *median* reference spectrum, not the *mean* reference spectrum. Despite the standard error on the median statistic being 25.3 per cent greater than that on the mean statistic (for normal distributions; Kendall & Stuart 1963; Freund 1971), the median statistic is robust to a high fraction of outlying data points, and is independent of the magnitude of deviation of the outlying points. The mean statistic, on the other hand, has the characteristic that *no* data point can be arbitrarily distant from the general trend without affecting the statistic.

Formally, the HIPASS bandpass estimate is determined independently for all 26 feeds (13 beams, two polarizations). Since the 13 beams track out nearly parallel paths on the sky, it is sufficient to generate a single list of *cycle numbers*, or equivalently *integration time-stamps*, which refer to integrations which contain spectra that are valid for estimating the bandpass spectrum at the instant of the target spectrum. This set of cycle numbers, $\mathcal{M}_\mathrm{T}$, is calculated once for the central beam, and assumed to be good for all other beams. Only in the case of unusual observing circumstances would this assumption be invalid, e.g., spiral scanning, or rapid rotation of the receiver while acquiring data. The suitability of a particular cycle for inclusion in the set $\mathcal{M}_\mathrm{T}$ is determined as follows.

(i) Only spectra from the same Dec. scan as the target spectrum are used as reference spectra. Between scans, the receiver is rotated slightly, thus changing the observing geometry, and hence the sidelobe radiation pattern.

(ii) The displacement on the sky from the target integration position must be at least 15 arcmin, in order to measure an independent part of the sky to that viewed by the target integration. Even though the beam half-width is only 7–8 arcmin, depending on the position of the beam on the focal plane, this minimum distance is set conservatively to prevent a particularly strong point source, or instead a nearby and extended source, from being partially removed during bandpass removal.

(iii) The displacement in time from the integration time-stamp of the target cycle must be less than 120 s. This is principally a restriction to enable near real-time processing of the data, and provide immediate feedback to the user on the data quality. However, this criterion also ensures that the spectra used to generate the bandpass estimate are not drawn from a long period of time over which the bandpass may have varied significantly.

Once the set of cycle numbers ($\mathcal{M}_\mathrm{T}$) is determined, the bandpass estimate is generated channel by channel, for each polarization of each beam of the receiver. For beam b (1…13) and polarization p (1 or 2) of the target cycle c_T, the bandpass estimate as a function of channel number n (1…1024) is:

$$B_{c_\mathrm{T},b,p}(n) = \mathrm{median}\{S_{c,b,p}(n) : c \in \mathcal{M}_\mathrm{T}\}, \quad (1)$$

where $S_{c,b,p}$ is the raw correlator spectrum for cycle c, beam b and polarization p. There are two points worth special mention regarding equation (1): first, in this formalism the bandpass estimate for the nth channel is independent of the bandpass estimate for any other channel and, secondly, an immediate consequence of the choice of the median statistic is that spatially unresolved spectral line sources will be absent from *all* bandpass estimates, provided that the set $\mathcal{M}_\mathrm{T}$ is sufficiently large (compared to the source density on the sky), independent of (ii) above. In particular cases, though, the median estimate may be slightly biased by the presence of a strong source in a subset of $\mathcal{M}_\mathrm{T}$; see Fig. 3.

3.3 Bandpass removal and calibration

Calibration of the spectra is done concurrently with bandpass removal, since system temperatures are recorded for every

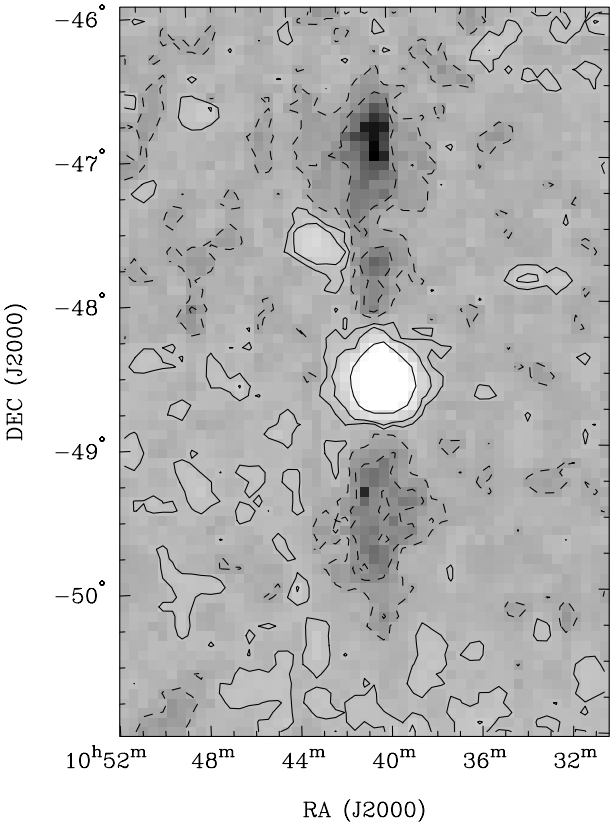

Figure 3. An example of negative *bandpass sidelobes* generated by bandpass correction north and south of a strong H I line source: the source at the centre is ESO 214 − G017, and the fainter source to the north east is ESO 264 − G035. The contours are at −50 and −20 mJy (dashed), and 20, 50 and 200 mJy (solid).

spectrum written to the correlator files. The system temperature, $T_{c_T,b,p}$, of the raw target spectrum, in Jy, indicates the total power measured by the correlator, calibrated against a calibration diode which is constantly switched in and out of the signal path. The diode itself is occasionally calibrated against an extragalactic radio source of known amplitude, e.g., 1934−638 (14.9 Jy at 1420 MHz) or Hydra A (40.6 Jy at 1395 MHz). $T_{c_T,b,p}$ is read directly from the correlator file. The system temperature of the bandpass estimate ($E_{c_T,b,p}$) must be estimated. This is accomplished by treating the system temperature in exactly the same way as a single channel, and taking the median of the system temperatures for the cycles in $\mathcal{M}_T$, i.e.

$$E_{c_T,b,p} = \text{median}\{T_{c,b,p} : c \in \mathcal{M}_T\}. \quad (2)$$

Then the bandpass-removed, calibrated target spectrum ($\mathbf{S}_{c_T,b,p}$) is obtained by comparing the raw correlator spectrum to the bandpass estimate, scaled by the ratio of the system temperatures:

$$\mathbf{S}_{c_T,b,p} = T_{c_T,b,p} \times \left(\frac{S_{c_T,b,p}}{B_{c_T,b,p}} \times \frac{E_{c_T,b,p}}{T_{c_T,b,p}} - 1 \right)$$
$$= E_{c_T,b,p} \times \left(\frac{S_{c_T,b,p}}{B_{c_T,b,p}} \right) - T_{c_T,b,p}. \quad (3)$$

In practice, some H I sources are sufficiently bright that bandpass estimates for target cycles just prior to and just after the cycle when the source makes its closest approach to a beam axis are elevated by the source itself. This results in *bandpass sidelobes* which are depressions in the spectra north and south of strong H I sources. The depressions are generated during the division of the target cycle by a bandpass estimate containing positive H I flux. Note, though, that depressions do *not* occur in the sources themselves, unless they are spatially extended (compared to the Parkes beam and the restrictions on $\mathcal{M}_T$) or brighter than ∼1 Jy: this is prevented for all but the strongest sources by the second criterion in generating $\mathcal{M}_T$. In Fig. 3 an example of the sidelobes associated with a strong source, as they appear in a gridded sky image (see Section 4), is given.

HIPASS bandpass sidelobes are generated during the bandpass removal stage of the HIPASS spectral processing. Thus the H I emission from extended H I sources, including the Galaxy, the high-velocity clouds (HVCs), the Magellanic Clouds and Stream, and other nearby galaxies such as Circinus, is corrupted in the standard HIPASS calibrated data, and therefore also in the HIPASS sky images (see Section 4). However, since all unprocessed HIPASS spectra are archived, the spectra can be reprocessed using different techniques, e.g., algorithms which preserve spatially extended emission. Such an approach has already been applied to the data in order to successfully image HVCs and the Magellanic Stream (Putman et al. 1998).

3.4 Residual baseline removal

Whilst bandpass removal and calibration ordinarily yields spectra that are flat, excluding receiver noise and spectral line sources, there are cases where the baseline is not flat. The two predominant causes of non-flat baselines for the HIPASS project are *ringing* associated with strong Galactic H I emission, and continuum emission extending over the wavelength range 20–22 cm which can produce standing wave patterns in the telescope structure, thereby causing residual ripple after bandpass removal. The first of these – spectral ringing – is relatively straightforward to correct, by smoothing every spectrum with a well-chosen kernel.

3.4.1 *Spectral smoothing*

A characteristic of digital correlators is that their spectra often suffer from what is commonly known as the Gibbs ripple. The spectral response function of a digital correlator is the Fourier transform of the time-lag weighting function used in the correlator. In the Multibeam correlator, the lag weighting is nearly constant as a function of lag so that the spectral response function is, to first order, a sinc function. When narrow lines are present in the data, this spectral response causes strong ringing in the spectra that decays roughly as n^{-1}, where n is the number of channels away from the narrow line. Given that in many positions on the sky the Galactic H I emission is strong and contains (in terms of the spectral resolution of the Multibeam receiver) narrow emission lines, this ringing can seriously affect the data over a large velocity range (see Fig. 4 for an example). A standard procedure to suppress this ringing is to Hanning-smooth the data. Such a smoothing very effectively suppresses the sidelobe level (the first spectral sidelobe decreases from −6.7 to −15.9 dB) and more importantly, the sidelobe level decays as n^{-3}. Because of these effects, the region of the spectrum affected by the ringing is much smaller, and a larger fraction of the spectrum can be used.

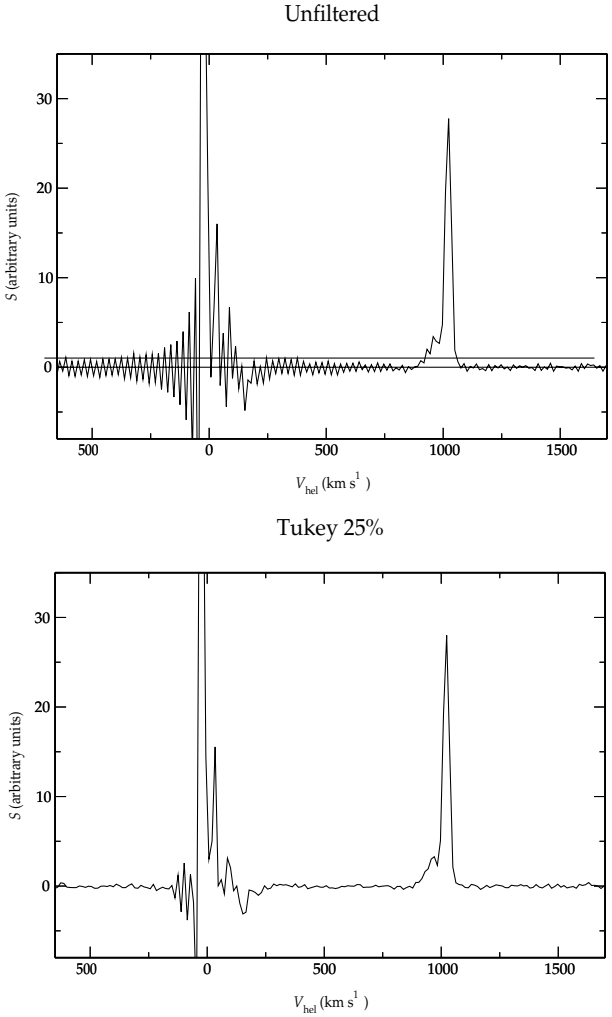

Figure 4. An unfiltered (top) and Tukey 25 per cent smoothed (bottom) HIPASS spectrum containing strong Galactic emission near 0 km s^{-1} and extragalactic emission at 1000 km s^{-1}. The ringing in the unfiltered spectrum is essentially removed in the smoothed spectrum.

However, the spectral resolution after Hanning smoothing is degraded by 67 per cent, and some detail in the signal is lost.

In order to suppress the Gibbs ringing in Multibeam spectra, while retaining most of the information in the signal, a Tukey filter was selected for smoothing purposes. This filter is applied in the lag domain, and can be written as

$$T_f(x) = \begin{cases} 1 & \text{for } |x| < f \times x_{\max} \\ \frac{1}{2} + \frac{1}{2}\cos\left(\pi \frac{|x| - fx_{\max}}{x_{\max} - fx_{\max}}\right) & \text{for } fx_{\max} \le |x| \le x_{\max}, \end{cases} \quad (4)$$

where $x_{\max}$ is the maximum lag, and $1 - f$ is the fraction of the lag spectrum that is tapered. The Tukey filter is very similar to a Hanning filter (in fact, it is *identical* for $f = 0$), and the sidelobe level of a filtered spectrum also decays as n^{-3}. Hence the region of the spectrum affected by the ringing is limited. The main difference with a Hanning filter is that the tapering is performed only on part of the lag spectrum as controlled by the parameter f. This means that the sidelobe level is higher, but also that the spectral resolution is degraded much less. For HIPASS spectra, $1 - f$ is set to 0.25, and the resolution is degraded by only 15 per cent compared to the full resolution (and the first sidelobe is at -6.9 dB). Fig. 4 also shows an example smoothed spectrum, illustrating that the ringing is much reduced and only a very small part of the spectral is not usable because of the ringing. Fig. 5 shows that the spectral resolution is much less affected by this Tukey filter than by the Hanning filter, and that more detail in the signal is retained.

3.5 Doppler tracking

Observations for HIPASS are made in topocentric mode, wherein the observing frequencies for each channel remain fixed throughout the survey. Consequently, the velocity range observed varies with time and position, and some correction must be made to the spectra to align them in a fixed reference frame. This is accomplished in the Fourier domain. First, for each spectrum, the velocity shift required to place the spectrum in the frame of the Solar System barycentre is calculated: this depends on observing epoch, sky position and antenna location. This shift is converted to a phase gradient across the spectrum, and is applied to the Fourier transform of the spectrum. The subsequent inverse transform yields a spectrum in the selected frame. The velocity frame conversion is applied *after* bandpass removal, since the bulk of the bandpass spectrum is internally generated in the observing frame.

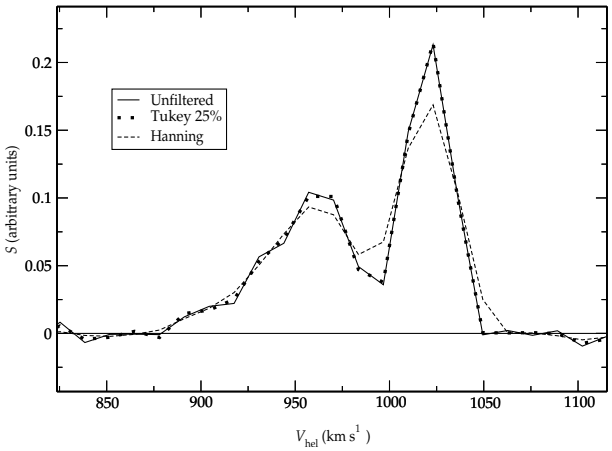

Figure 5. Comparison of the effects of Tukey 25 per cent and Hanning smooths for a strong, resolved signal.

4 IMAGING

4.1 Requirements

The result of the spectral processing of a *single* HIPASS scan is a set of calibrated H I spectra for 1300 unique positions on the sky. The scanning approach delivers typically 1500 spectra per square degree of sky, ~55 of which are acquired by each Multibeam feed. For the most part, these spectra are free of continuum ripple, of baseline curvature, and residual bandpass effects. Indeed, they are mostly free of H I line emission too, given the sparsity of galaxies on the sky. Furthermore, since the integration time is only 5 s, the root mean square (rms) noise level in the individual spectra is typically 72 mJy (see Section 4.3.2), and so only the very brightest H I sources are ever seen in single spectra. Some operation is therefore required to compile the individual spectra into high signal-to-noise images of the H I sky. As well as dramatically improving the visibility of low-flux sources, such an operation produces a more compact and natural representation of the H I sky. This operation is the primary task of the Multibeam imaging software.

Projection centres have been defined for 388 images covering the southern sky. The projection used is the orthographic projection (see Kellaway 1946) with reference (tangent) point at the centre of each image. Spatially the images measure $8° \times 8°$, where the width in RA is measured at the centre of the image. For the HIPASS pixel size of $4' \times 4'$ (see Section 4.2.2), the images have dimensions of 170×160. Spectrally, each image comprises 1024 planes (channels), extending in cz from -1280 to 12 700 km s^{-1}. Each image is generated from 75 HIPASS scans, and has typically a 10 per cent sky overlap (by area) with adjacent images.

4.2 Robust gridding

Conversion of the individual spectra into position-position-velocity cubes requires placing the spectra on a regular grid, i.e., *gridding*. For each pixel in the map, the gridding process must:

(1) determine which spectra will contribute to the pixel;
(2) reject those spectra which appear to contain corrupt data;
(3) ascribe a weight to each remaining spectrum, and
(4) calculate the value of the pixel based on the input data and weights.

The gridding algorithm used for the survey is optimized for point sources: more than ~90 per cent of HIPASS sources are expected to be small in angular size compared to the Multibeam resolution.

4.2.1 Algorithm

The objective of gridding is to reconstruct the flux at a certain position (*pixel*) on the sky, given an arbitrary number of single beam spectra measured at irregularly distributed positions near the target pixel. Fig. 6 shows a representative example of the problem: two observations A and B have been made near the target pixel P. Observation A measures a flux F'_A at a distance r_A from the pixel centre, and B measures a flux F'_B at a distance r_B from the pixel centre. Two possible sources C and D are shown in Fig. 6, having

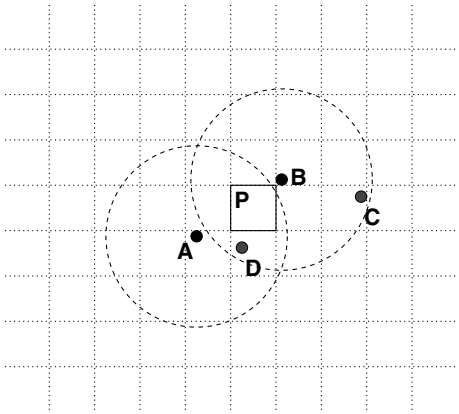

Figure 6. Diagram showing the positions of two observations *A* and *B* near a target pixel *P*, with two sources *C* and *D* in the field. The dashed circles indicate the half-power radii of the observations *A* and *B*.

intrinsic fluxes F_C and F_D respectively. It is, of course, generally true that $F'_A < (F_C + F_D)$ and $F'_B < (F_C + F_D)$.

Provided that the instrumental beam shape is known (or can be estimated), the flux in the pixel *P* can be *estimated* as follows: the beam shape (beam sensitivity) is introduced as a weighting function $w(r)$ which depends only on the distance from the beam centre r, with $w(0) \equiv 1$ and $w(\infty) \equiv 0$. Then

$$F_A = F'_A / w(r_A),$$

$$F_B = F'_B / w(r_B), \tag{5}$$

where F_A and F_B are the *reconstructed* summed fluxes of the sources *C* and *D* as measured by *A* and *B*. As $r_A > r_B$ and $w(r_A) < w(r_B)$, F_A is a less certain measurement of the flux in sky pixel *P* than is F_B. Then the weighted mean of F_A and F_B yields a (simple) statistical estimate of the flux F_e which may be assigned to the pixel *P*, with the weights (again) provided by the beam sensitivity $w(r)$:

$$F_e = \frac{w(r_A) \times F_A + w(r_B) \times F_B}{w(r_A) + w(r_B)}. \tag{6}$$

Substituting the reconstructed fluxes into this equation, and generalizing to *N* pointings, we find:

$$F_e = \frac{\sum_{i=1}^{N} F'_i}{\sum_{i=1}^{N} w_i} = \frac{\left(\sum_{i=1}^{N} F'_i\right)/N}{\left(\sum_{i=1}^{N} w_i\right)/N} = \frac{\text{mean}(F')}{\text{mean}(w)}. \tag{7}$$

Unfortunately, for real data, mean (F') can be severely corrupted by artefacts in the data like interference. In principle, this could be corrected by explicitly removing bad data, either manually or automatically, using some pre-defined quality criteria. In practice, however, manual data removal is tedious work subject to human error, and automatic data removal often depends on arbitrarily chosen criteria. A far better way is simply to replace the mean statistic used in equation (7) with a more statistically robust estimator – the median – in keeping with the philosophy applied elsewhere in the Multibeam processing. An estimator like the median is robust to a substantial presence of outliers, for example, measurements corrupted by interference. The statistic itself acts to identify the unusual elements in the sample, and requires no arbitrary criteria to do so. As long as the fraction of corrupt data in the sample is less than ~30–40 per cent (as is the case for nearly all samples of the survey data) the median statistic performs admirably. Thus

$$F_e = \frac{\text{median}(F')}{\text{median}(w)}, \tag{8}$$

which we define as *median gridding*. Note that the mean estimator, which we have replaced with the median, was only one of many that could be chosen. Furthermore, alternative weighting schemes are possible, e.g.,

$$F_e = \frac{\vartheta(w^n F')}{\vartheta(w^{n+1})}, \tag{9}$$

where ϑ is some statistical estimator (e.g., mean, median, etc.), and n is a measure of the importance of the weights. Generally, ϑ = mean and $n = 1$ will give the minimum variance result *for interference-free data*.

It is actually possible to estimate the image *error* per pixel by slightly modifying the gridding algorithm. Equation (9) can be even more generally written as

$$F_e = \frac{\varphi(w^n F')}{\vartheta(w^{n+1})}, \tag{10}$$

where different statistical estimators φ and ϑ are used in the flux and weight calculations respectively. In this form, with φ selected as a measure of dispersion, and ϑ selected as a mean-like estimator (e.g., mean, median, …), the resultant gridded image describes the error present in each pixel according to the selected dispersion measure. Though not used for the purposes of this paper, images constructed with robust dispersion measures are proving to be extremely important for assessing the significance of source detections in the HIPASS images.

Equation (7) (and therefore equations 8 and 9) neglects the displacements between the observation *A* and *B* and the sources *C* and *D* in Fig. 6. These observation–source displacements *cannot* be known without adopting an iterative sky reconstruction technique. Despite this, equation (8) works well for the HIPASS data, which is highly redundant and over-sampled. While the source *C* in Fig. 6 is within the half-power radius of observation *B*, it will only be measured at a fraction of its intrinsic strength, because the observation itself applies a beam-weighting to the sky source distribution. Source *C* will be invisible to observation *A*, thereby weighting its contribution to pixel *P* even lower. Generalizing to the case of highly over-sampled data, and applying a nearness-to-*P* requirement, the majority of measurements considered in equation (8) for pixel *P* will not detect source *C*. On the other hand, source *D* is likely to be detected by most observations clustered around *P*, and will therefore be reconstructed at a modest fraction of its intrinsic flux in pixel *P*. For a sky containing a single point source centred on pixel *P*, equation (8) will reconstruct exactly the flux of the point source.

4.2.2 *HIPASS-specific parameters*

The value of median (w) depends primarily on the weighting function $w(r)$ used and the radius r_{max} out to which spectra will be included for the calculation of a single-pixel value. For the HIPASS, the telescope beam is assumed to be a two-dimensional Gaussian function, so the weighting function takes the following form:

$$w(r) = \begin{cases} \exp\left[-\left(\frac{r}{\sigma}\right)^2/2\right] & \text{for } r \le r_{max} \\ 0 & \text{for } r > r_{max}. \end{cases} \tag{11}$$

The mean beamwidth of 14.3 arcmin is used, corresponding to $\sigma = 6.1$ arcmin. The generalization of median gridding to use a weighting function per beam, while simple, has not been made, since it is unlikely to affect the images in any significant way.

The choice of smoothing radius ($r_{\rm max}$) determines how many spectra contribute to each pixel value, and therefore partly determines the final image noise level. The smoothing radius also ultimately decides the final gridded beamsize because of quantization noise. Additionally, the choice of pixel size affects the correlation between pixels and, consequently, the apparent image noise. Tests have shown that smoothing radii less than 4 arcmin do not provide sufficient robustness; this is simply because the number of HIPASS spectra per pixel is too low (typically 30–40). Large smoothing radii (>10 arcmin) cause loss of resolution and positional accuracy. Following extensive testing, it was evident that a compromise was necessary between obtaining the lowest possible image noise and obtaining the highest possible resolution images. The HIPASS smoothing radius was selected to be 6 arcmin, and the pixel size to be 4 arcmin.

For the selected imaging parameters, the distribution of $w(r)$ was investigated. For a random distribution of observations, and a smoothing radius of $r_{\rm max}$, the median of the weights [median (w)] is simply the beam (weighting) function evaluated at the radius which divides the smoothing *area* equally in two, i.e., $w(r_{\rm max}/\sqrt{2})$. For $r_{\rm max} = 6$ arcmin, median $(w) = 1.28$. Tests during gridding HIPASS images show the distribution around this value to be very narrow, indicating that the input spectra are very nearly randomly distributed on the sky.

4.3 Properties of median-gridded images

4.3.1 Robustness

The motivation for the median-gridding algorithm was the potential for automated statistical robustness to corrupt data arising from externally generated interference or an observing system failure. This objective has been achieved, as Fig. 7 exhibits. The Global Positioning System (GPS) L3 beacon near 1380 MHz is powerful (typically more than 10 Jy is measured when the offending satellite is in the distant sidelobes of the telescope beams!) and is visible in up to ~5 per cent of HIPASS spectra. Without robust gridding, the GPS signals are preserved in the HIPASS sky images, manifested as severe declination striping for instances when the GPS signal persisted for several minutes, and as *beam imprints* on the sky when the GPS signal is brief (see Fig. 7, top panel). Median gridding, by comparison, is mostly impervious to the GPS signal, which corrupts only a small fraction of the spectra taken at that position on the sky (Fig. 7, lower panel). Only in cases where the GPS signal is present for an extended period of time will the median gridding fail to remove the offending signal completely, thus GPS signals are seen in a small but non-negligible number of HIPASS images.

4.3.2 Image noise

The expected rms noise in single-polarization HIPASS spectra is 72 mJy. This is calculated assuming a mean system temperature of 19 K (measured for 1350 spectra for each feed, sampled from 18 months of data), a correlator efficiency of 82 per cent, and a noise equivalent bandwidth of $1.10 \times 62.5 = 68.8$ kHz, where this adjustment is made for the Tukey 25 per cent smooth that is applied to the spectra. In order to determine the rms noise present in the *images*, the number of spectra contributing to each pixel is required. The total integration time for the survey is 450 s beam^{-1}, and for the $r_{\rm max} = 6$ arcmin smoothing radius, this corresponds to 220 s pixel^{-1} in the gridded images. The number of single-polarization spectra contributing to the calculation of any pixel therefore is ~88, since each spectrum takes 5 s to gather, and there are two independent linear polarizations.

Thus the rms noise expected in *mean* gridded HIPASS images is $72/\sqrt{88} = 7.7$ mJy, assuming that the rms noise in the individual spectra is close to a normal distribution. For *median* gridding though (see equation 8), the noise level is increased due to the variance on the median estimator being higher than that on the mean estimator by 25.3 per cent (Kendall & Stuart 1963; Freund 1971). Additionally, the (narrow) scatter in the weighting factor in equation (8) increases the noise in the images by a further 3.2 per cent (measured from the distribution of weights), giving an image

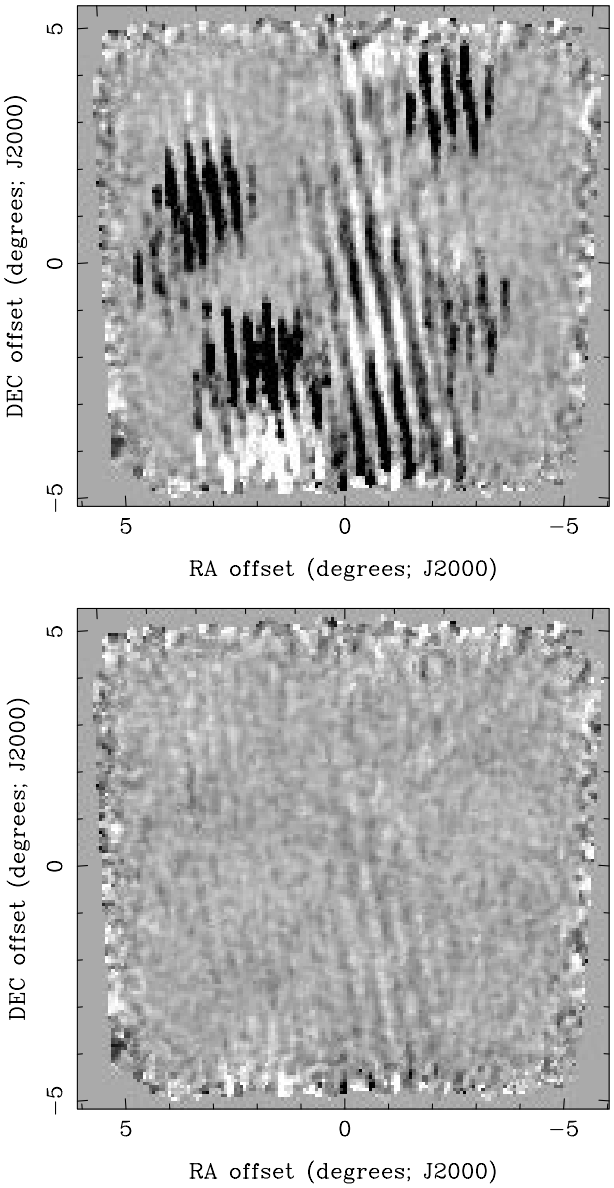

Figure 7. Planes from mean (top) and median-gridded HIPASS images near 1380 MHz. The powerful *L3 beacon* of the GPS has completely corrupted the mean gridded image (in this and nearby channels), yet is almost completely removed by the robust median-gridding algorithm.

Table 2. Image beam dimensions and noise as a function of peak flux, for point sources. See text for a description of the columns.

Source peak flux	Image noise	MEASURED IMAGE BEAM		INTRINSIC IMAGE BEAM	
		Shape	Error	Shape	Area
(mJy)	(mJy)	(arcmin × arcmin)	(arcmin)	(arcmin × arcmin)	(arcmin^2)
(1)	(2)	(3)	(4)	(5)	(6)
3.0×10^1	12.5	19×17	6	16.6×16.0	300
5.0×10^1	12.6	18×16	4	16.2×16.1	295
1.0×10^2	12.7	17×16	2	16.4×16.2	300
1.5×10^2	13.0	17.6×16.0	0.6	16.3×15.9	292
1.0×10^3	31.2	15.6×14.9	0.3	15.1×14.9	255
1.0×10^4	350	15.3×14.6	0.2	14.8×14.6	244

rms noise level of 9.9 mJy. Finally, the median weight term in equation (8) (i.e., the denominator) effectively globally scales the images by a factor of 1.28; thus the rms noise level measured in median-gridded HIPASS images should be 12.7 mJy at best.

Simulations of point sources placed at various offsets from the centre of a gridding top-hat, with input noise 72 mJy, and 88 unique observations per top-hat, confirm rms noise levels of 12.7 ± 0.2 mJy [see Table 2, columns (1) and (2)]. The rms noise is elevated for sources stronger than $\sim$1 Jy, because there is a much wider variation in flux in the samples made within the gridding top-hat; note that the signal-to-noise ratio (S/N) is typically around 30 or greater for such strong sources. For the HIPASS images themselves, measurements show the canonical rms noise level to be 13.3 mJy. Careful selection of a region devoid of H I sources, continuum sources, and residual solar ripple (see Section 5.1) enables the measurement of rms values approaching $\sim$12.8 mJy. Thus, in places, the theoretical noise limit of the survey is attained.

4.3.3 Beam properties

The HIPASS image beam is manufactured from 26 unique telescope beams, each of which has been elongated by the active scan (effectively convolved with a 5 arcmin top-hat in the Dec. direction), and then convolved with the gridding kernel, which is a 12 arcmin top-hat function. However, since the median gridding is a markedly non-linear process, the actual HIPASS image beam is statistical rather than definite. The beam depends not only on the smoothing radius and pixel size, but also on the sky sampling (and therefore position in the image), source strength (akin to S/N), and source shape. The only way to determine canonical values for the image beam, therefore, has been to construct and analyse simulations.

Dependence on signal-to-noise ratio. A simulation to assess the dependence of the beamwidth on S/N for point sources was designed as follows: a circle of radius $r_{\max} = 6$ arcmin was laid down having its centre at the origin of a Cartesian plane. A source of strength S mJy was placed at (x,0), with x measured in arcmin along the X-axis. 44 random positions within the circle were generated, and the measured flux calculated twice (once per polarization, each measuring the same flux, but having different noise contributions) for each position assuming a Gaussian beam of width 14.3 arcmin (the mean ungridded beam). The smearing of the beam in the dec. direction by the active scanning was introduced by actually taking the average flux over a set of positions displaced from each random position by displacements in the interval $[-2.5', 2.5']$. Normally distributed noise with rms 72 mJy was added to the 88 flux measurements, and the median flux measurement calculated and scaled by the median of the weight values for all positions, according to equation (8). Varying x and S, and repeating the simulation a large number of times, yields the gridded beam shape for various peak flux image signal-to-noise ratios S/13.3.

The results of these simulations are presented in Table 2. In column (3) the measured image beam dimensions are given for a series of input source peak flux values. The data provided are the mean widths for Gaussian fits to 50 independent simulations, and the standard deviations of the width distributions are given in column (4). Some component of the beam widths measured in this way, however, will be due to problems inherent in fitting Gaussians to relatively noisy data. Therefore a more meaningful estimate of the image beam has been obtained for each source by adding the 50 simulation images together and making one Gaussian fit to the resultant image. This yields the beam dimensions given in column (5), and the beam area in column (6). While columns (3) and (4) give the range of width measurements which in practice will be determined from Gaussian fits to HIPASS images,[2] columns (5) and (6) are more representative of the *intrinsic* image beam, and should be used for image analysis purposes.

Point sources. Simulations which inject false sources into genuine survey data prior to imaging show directly that median gridding produces correct *peak* fluxes (typically within 1 per cent of the expected value) for point sources. This is expected, since equation (8) has been constructed with this in mind. To calibrate the *integrated* flux scale, it is useful to consider the area of the image beam ($A_{\rm b}$), which in arcmin^2 is defined to be

$$A_{\rm b} \equiv 1.13 \times \frac{S_i p_x p_y}{S}. \qquad (12)$$

S is the integrated flux of the input (simulated) source, S_i is the summed flux in the output image *not* corrected for the beam, and p_x and p_y are the image pixel dimensions in arcmin. The pre-multiplier 1.13 is equal to $\pi/(4 \ln 2)$, the constant term in the area of a Gaussian beam. Inspection of the *integrated* flux in the simulated point sources gives $S_i = 1.59$ Jy pixel for $S = 1$ Jy, and thus $A_{\rm b} = 243$ arcmin^2. The beam area calculated for a 1.00 Jy source from Table 2 is 255 arcmin^2; thus the image beam for a 1.00-Jy source deviates from Gaussian by $\sim$5 per cent.

[2] Actually, the situation is slightly better than this, since the values in Table 2 are appropriate for fits made to a single plane of an image; in practice, source fits are normally made to intensity maps integrated over several channels.

Table 3. Flux corrections to be applied to measurements made on HIPASS images. See text for a description of the columns.

Source width (arcmin) (1)	Peak flux scale (2)	Flux-weighted beam area (arcmin^2) (3)
0	1.00	244
5	0.97	256
10	0.92	268
15	0.86	275
20	0.83	283
30	0.79	289
40	0.78	293

Extended sources. The median weight [median (w) in equation 8] can be interpreted as a correction for the case where a set of observations are made on and around a *point source* centred on a pixel in the output image. Without the median weight correction, the value assigned to the pixel will be consistently lower than the true peak flux, as the few observations made directly on-source are at the extremes of the distribution, and consequently are ignored by the median operation. In the case of an extended source, however, the median weight correction *over-corrects*, resulting in peak fluxes that are too high. For the extreme case of an infinitely extended source (i.e., any source larger than say 60 arcmin), the over-correction is by exactly the factor median (w) in equation (8), or typically $1.0/0.78 = 1.28$. Simulations have been used to provide a table of peak flux corrections for sources of various sizes, where the peak flux measured in the image is scaled by a factor between 1.0 (for point sources) and 0.78 (for *infinitely* extended sources). These corrections are given in Table 3, where column (1) gives the simulated source width, and column (2) gives the factor by which to multiply peak fluxes measured in HIPASS images to obtain correct source peak fluxes.

For *integrated* fluxes of extended sources, simulations have been used to measure the *flux-weighted beam area* as a function of source size, *for a fixed peak flux of 1.00 Jy.* The input *integrated* flux S is given by

$$S = \frac{14.3^2 + \theta^2}{14.3^2}, \tag{13}$$

where θ is the full width at half power of the simulated (Gaussian) source. The results are given in column (3) of Table 3, which gives the flux-weighted beam area to use when calculating integrated fluxes of point and extended sources in HIPASS images.

In summary, to measure the H I flux of a source found in a HIPASS image, it is necessary to first deduce (by fitting the source of interest and comparing the result to Table 2) whether the source is a point or extended source. For the first case (corresponding to probably more than 90 per cent of extragalactic H I sources visible in the HIPASS images), the source peak flux is exactly that measured in the HIPASS image, and the source integrated flux is the summed flux in the region of the source, divided by the appropriate flux-weighted beam area from Table 3 (in this case, 243 arcmin^2), multiplied by the pixel area (i.e., 16 arcmin^2). For a source which appears to be extended, *provided that the assumption that the source is intrinsically Gaussian is valid*, the source size must be estimated by deconvolving the fitted image size with a suitable beam from Table 2, and then the corresponding flux scale and flux-weighted beam area from Table 3 used in place of the values for the point source case. For extended sources which are not intrinsically Gaussian, it is not possible to provide generic beam shape or flux calibration tables. In such cases, simulations pertinent to the particular source geometry must be constructed. Note also that sources extended much beyond ~40 arcmin can be imaged poorly in HIPASS images because of the bandpass sidelobe effect described in Section 3.3.

5 POST-PROCESSING

5.1 Continuum ripple removal

As is the case for observations made with most H I observing instruments, continuum sources moving through any of the beams or sidelobes of the Multibeam generally manifest themselves as sinusoidal *ripple* in the spectral baseline. The main reflection path in the Parkes telescope, from the receiver directly to the dish centre and returning to the receiver, i.e., *twice* the focal length, is 52 m in length, which corresponds to a time displacement of 0.17 μs. This time delay (appearing in the lag spectrum as a spike) gives rise to a sine wave in the Fourier transform spectrum having this wave number or, equivalently, a period of 5.7 MHz in frequency space or 1.2×10^3 km s^{-1} in velocity space.

Solar emission is the most common cause of ripple, and is problematic for a high fraction (say ~40–60 per cent) of spectra acquired during daylight hours. However, the apparent movement of the Sun against the extragalactic sky means that solar ripple conditions vary over the course of a year, and consequently do not persistently affect spectra of a particular part of the sky. All other continuum sources which lie at fixed positions on the sky can inflict baseline ripple on all spectra within a few arcminutes of the sources. In particular, sightlines directed towards continuum sources, or near strong, extended sources like the Galactic plane or Cen A, always suffer from baseline ripple. When most or all of the spectra from a particular part of the sky are plagued by baseline ripple, the ripple will be embedded in the gridded images, and it therefore deteriorates image quality to a much greater degree than does solar ripple.

An additional baseline distortion seen in calibrated HIPASS spectra taken near strong continuum sources is a gradual rise in flux towards the high-velocity (low-frequency) end of the observing band. There is a clear dependence of the slope of this deviation on the strength of the underlying continuum source (see below). This distortion suggests that the response of the Multibeam receivers to strong continuum sources – and therefore mildly increased system temperature – is a noise spectrum whose amplitude is elevated at lower frequencies. The multiplication of this spectrum by the correlator bandpass, and subsequent division by a weak source bandpass spectrum generates a baseline which curves upward at low frequencies.

Inspection of the HIPASS cube data base shows that spectra taken from HIPASS cubes are well described by the expression

$$\mathbf{S}(n) = N(n) + R(n) + k \times B(n) + H(n), \tag{14}$$

where $\mathbf{S}(n)$ is the flux in channel n, $N(n)$ is the image noise, $R(n)$ is residual continuum ripple associated with solar and spill-over radiation, $B(n)$ is the bandpass curvature and continuum source ripple, and $H(n)$ is any H I signal present. Typically $N(n)$ approximates $1/f$-noise (Voss 1988) with an rms value of 13.3 mJy, and $R(n)$ is all but eliminated by robust statistics operating on spectra acquired under different observing conditions. For positions in HIPASS images coincident with strong continuum sources, the component $B(n)$ dominates, and to first order is constant in shape

for all sources. This important feature makes it possible to remove the bulk of spectral ripple and curvature from HIPASS images. The coefficient k is predominantly proportional to the increased system temperature (compared to the average system temperature) of the (ungridded) spectra which were combined to produce that part of the image. That is, the magnitude of k is correlated with the strength of any continuum sources at that position.

The principal steps in the method used to accomplish the removal of $B(n)$ from $\mathbf{S}(n)$ – hereafter the *scaled template method* – are the formation of a *canonical baseline* [i.e., a template for $B(n)$], and the statistically robust excision of $k \times B(n)$ from all spectra (pixel by pixel) across a HIPASS image. Both parts make use of the same basic algorithm, and at present are applied *directly* to the gridded images. Future reprocessing may be able to make use of the information gathered on $B(n)$ and apply it to spectra prior to gridding. The method by which the canonical baseline S_c is constructed is to form the weighted sum

$$S_c = \sum_{i=1}^{M} w_i S_i \tag{15}$$

of the continuum source spectra exhibiting the most severe spectral ripple and curvature in a HIPASS image. These spectra S_i are found by selecting those with the largest positive fluxes at the low-frequency end of the spectrum. The maximum signal-to-noise ratio of this weighted sum over terms which include uncorrelated noise of uniform variance is accomplished by setting the weights w_i proportional to the amplitude of the terms themselves. In the present case, that means setting the weights equal to the value of k for each spectrum in the sum. To determine k, note that, in the absence of significant contributions from the $R(n)$ or $H(n)$ terms in equation (14),

$$S_i = a \times S_j + N', \tag{16}$$

where S_i and S_j are two arbitrary spectra, $a = k_i/k_j$, and N' is a noise term given by

$$N' = N_i - a \times N_j. \tag{17}$$

Hence a scatter plot of S_i against S_j should give a straight line (plus noise) with a slope equal to k_i/k_j. Provided that the maximum excursion of that member of the pair with greater amplitude is significantly greater than the standard deviation of the noise, and that k_i is not very large compared to k_j, a linear regression can be performed to give the slope $a_{i,j}$.

The need to avoid contamination of the canonical baseline by H I signals is addressed by an iterative process that flags any points which depart by more than 5σ from the straight-line fit to the scatter plots, and then performs the regression without these points; this cycle continues until no further points are being flagged. This procedure works only if an H I signal appears (for a given channel) in one spectrum but not in the others. This is not true for the contribution from the Milky Way: these channels (less than 5 per cent of the total) are therefore not included in any of the scatter plots.

After formation via equation (15), the canonical baseline S_c is passed through a narrow median filter, typically nine channels wide, to further reduce the high spatial frequency noise component. S_c can then be used to remove continuum source distortion components $k \times B(n)$ from the whole cube. This is done in exactly the same way that S_c is formed, namely, by making scatter plots of each spectrum S_i against S_c and performing a linear regression to derive the slopes $a_{i,c} = k_i/k_c$. The cleaned spectra S_i' are then formed as follows:

$$S_i' = S_i - a_{i,c} \times S_c = N + R + H. \tag{18}$$

The noise variance is increased only slightly, since S_c is relatively noiseless compared to S_i, and the H I contribution H is undisturbed. A typical example of ripple and curvature seen in a HIPASS image, and the result of processing that image to remove the canonical baseline, is given in Fig. 8.

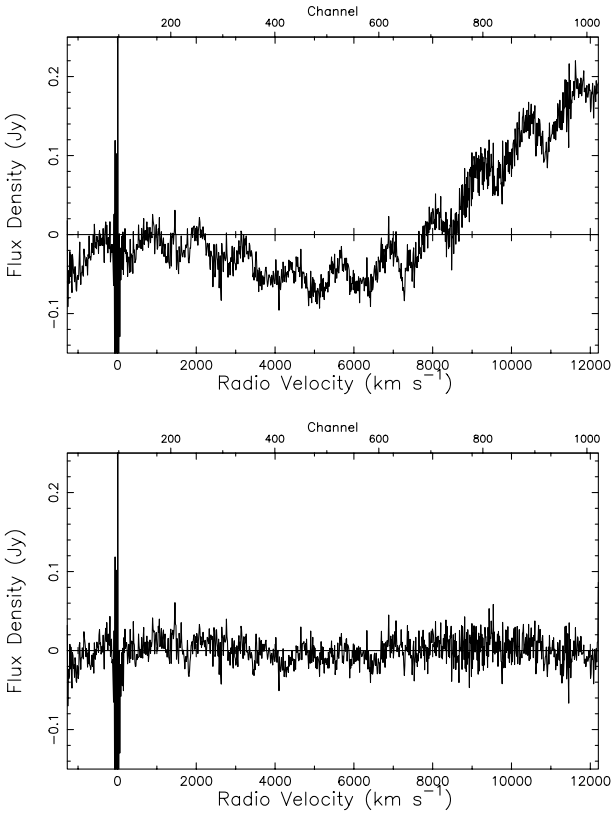

Figure 8. Moderate spectral baseline ripple and curvature as seen in a spectrum at the position of the 3.9-Jy continuum source NGC 4696 (top), and a spectrum at the same position in the image after *cleaning* with the scaled template method (bottom).

5.2 Mosaicking

There are 388 HIPASS images, each measuring $8° \times 8°$ on the sky. Because of the significant overlap in RA and Dec. of adjacent cubes, each discrete source can be examined with optimum sensitivity and with sufficient surrounding *blank field* in a single cube. However, for slightly extended sources, for removing cube edge-effects, and for making a large field for display purposes, it is desirable to mosaic several cubes together. The alternative to mosaicking, which is to regrid the original scan data, is often impractical due to the disk space and processing time needed.

For the ith data cube, a weight image $W_i(x, y)$ is defined to be the number of data points contributing to the flux density image $F_i(x, y)$ (which itself is the scaled median of these data points). The mosaicked image is then:

$$M(x, y) = \frac{\sum_i W_i(x, y) F_i(x, y)}{\sum_i W_i(x, y)}, \tag{19}$$

where it is assumed that all cubes and weight images have identical projections. Points with a weight $W_i < 0.15\langle W_i \rangle$ are excluded from the calculation to preserve robustness. In practice, HIPASS cubes have different tangent points, so are usually reprojected (e.g. with the MIRIAD task REGRID) to a common tangent point near the intended centre of the mosaic. The weight for each pixel is nominally inversely proportional to variance, so equation (19) is an optimum estimator for normal data. It does not give an identical cube to one produced by, say, regridding the calibrated scan data on to the same large output grid. This is because of the non-linear median estimator used in the gridder. Nevertheless, in practice, similar data quality is obtained, and it is much faster and more convenient to use equation (19) on already existing cubes than to start again from the scan data.

An example wide-field HIPASS image is shown in Fig. 9. The image is a $16^\circ \times 25^\circ$ map of integrated H I emission in the Sculptor region. The moment map has been generated over a velocity range of 200–550 km s^{-1}, which includes the main members of the Sculptor galaxy group, numerous other galaxies, and a large number of high-velocity clouds in the south, most of which comprise part of the Magellanic Stream. A detailed analysis of the neutral hydrogen distribution in the Sculptor group is in preparation (Koribalski et al., in preparation).

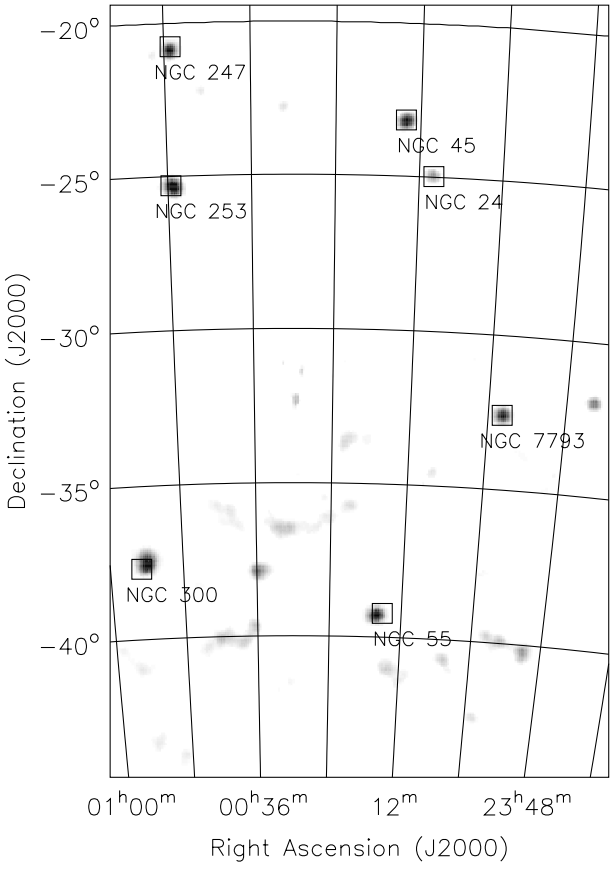

Figure 9. A sample of HIPASS data: a moment map of the Sculptor region, showing the H I line intensity integrated over a small velocity range. The seven (optically) brightest galaxies in the field are marked; H I emission is easily detected in these galaxies, and also in fainter galaxies throught the region, and in high-velocity cloud complexes towards the south of the image.

6 DATA AVAILABILITY

One of the important goals of the HIPASS project is to deliver to the astronomical community its images of the southern H I sky. In 1999 November, a preliminary release of data from 11 of the total 388 HIPASS images was made. This release, which provided web access to H I spectra in the Centaurus A region, was extended in 2000 May to the entire southern sky. The release pages can be accessed via the 'Parkes 21-cm Multibeam Project' page at the address:

`www.atnf.csiro.au/research/multibeam/multibeam.html`

The data in the Centaurus region is the subject of a paper by Banks et al. (1999). Downloaded spectra are extracted from HIPASS images which have been cleaned using the procedure described in Section 5.1, and smoothed spatially to a pixel size of 8 arcmin. Users of the data release are requested to acknowledge the Australia Telescope National Facility as follows:

`The Parkes telescope is part of the Australia Telescope which is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO.`

Full-resolution H I images will be released in late 2000.

7 SOFTWARE AVAILABILITY

The software which implements the spectral processing algorithms described in Section 3 is called LIVEDATA, and is available in the `atnf` package of AIPS++. The software which implements the median gridding algorithm is known as GRIDZILLA, and can also be found in AIPS++.

ACKNOWLEDGMENTS

The Multibeam system was funded by the Australia Telescope National Facility (ATNF) and an Australian Research Council grant. The collaborating institutions are the Universities of Melbourne, Western Sydney, Sydney and Cardiff, Mount Stromlo Observatory, Jodrell Bank Observatory and the ATNF. The Multibeam receiver and correlator was designed and built by the ATNF with assistance from the Australian Commonwealth Scientific and Industrial Research Organisation Division of Telecommunications and Industrial Physics. The low noise amplifiers were provided by Jodrell Bank Observatory. The Multibeam Survey Working Group is acknowledged for its role in planning and executing the HIPASS project. The HIPASS algorithms are implemented in code built upon AIPS++, and the authors are especially grateful for the extensive support provided by various AIPS++ consortium members, and especially for travel funding provided by the ATNF. We express particular thanks to the following AIPS++ personnel: Tim Cornwell, Ger van Diepen, Bob Garwood, Brian Glendenning, Neil Killeen and Darrell Schiebel. We are grateful to Richard Gooch for providing visualisation services in the on-line software implementation. Finally, we express our sincere gratitude to the staff of the Parkes Observatory who have provided magnificent observing support for the survey since the very first HIPASS observations in early 1997.

REFERENCES

Banks G. D. et al., 1999, ApJ, 524, 612
Barnes D. G., 1998, in Albrecht R., Hook R. N., Bushouse H. A., eds, ASP Conf. Ser. Vol. 145, Astronomical Data Analysis Software and Systems VII. Astron. Soc. Pac., San Francisco, p. 32
Canaris J., 1993, in Dowd A., ed., Proceedings of a Workshop on New Generation Digital Correlators. National Radio Astronomy Observatory, Tucson, p. 117
Freund J. E., 1971, Mathematical Statistics. Prentice-Hall, Englewood Cliffs
Kellaway G. P., 1946, Map Projections. Methuen & Co. Ltd., London
Kendall M. G., Stuart A., 1963, The Advanced Theory of Statistics. Griffin, London
Putman M. E. et al., 1998, Nat, 394, 752
Sinclair M. W., Wilson W. E., Mooney G. G., Sykes P. B., Bolton R. J., 1997, Proceedings of the 4th International Symposium on Low Temperature Electronics and High Temperature Superconductivity, Proc. Electrochem. Soc., 97, 336
Staveley-Smith L., 1997, Proc. Astron. Soc. Aust., 14, 111
Staveley-Smith L. et al., 1996, Proc. Astron. Soc. Aust., 13, 243
Voss R. F., 1988, in Peitgen H.-O., Saupe D., eds, ASP Conf. Ser. Vol. 145, The Science of Fractal Images. Springer-Verlag, New York
Wilson W. E. et al., 1997, Proceedings of the Workshop on Applications of Radio Science. Australian Academy of Science National Committee for Radio Science, Canberra, p. 254
Wright A. E., Otrupcek R., 1990, Parkes Catalogue, Australia Telescope National Facility

This paper has been typeset from a TEX/LATEX file prepared by the author.