Realtime, Object-oriented Reduction of Parkes Multibeam Data using AIPS++

D. G. Barnes

School of Physics, The University of Melbourne, Parkville, VIC 3052, Australia; Email: dbarnes@physics.unimelb.edu.au

Abstract. An overview of the Australia Telescope National Facility (ATNF) Parkes Multibeam Software is presented. The new thirteen-beam Parkes 21 cm Multibeam Receiver is being used for the neutral hydrogen (HI) Parkes All Sky Survey (HIPASS). This survey will search the entire southern sky for HI in the redshift range -1200 km s^{-1} to $+12600$ km s^{-1}; with a limiting column density of $N_{\rm HI} \simeq 5 \times 10^{17}$ cm^{-2}. Observations for the survey began in late February, 1997, and will continue through to the year 2000.

A complete reduction package for the HIPASS survey has been developed, based on the AIPS++ library. The major software component is realtime, and uses advanced inter-process communication coupled to a graphical user interface, provided by AIPS++, to apply bandpass removal, flux calibration, velocity frame conversion and spectral smoothing to 26 spectra of 1024 channels each, every five seconds. AIPS++ connections have been added to ATNF-developed visualization software to provide on-line visual monitoring of the data quality. The non-realtime component of the software is responsible for gridding the spectra into position-velocity cubes; typically 200000 spectra are gridded into an $8^\circ \times 8^\circ$ cube.

1. Introduction

1.1. The Parkes 21 cm Multibeam Receiver

The Parkes 21 cm Multibeam Receiver (hereafter "Multibeam") is described in detail by Staveley-Smith et al. (1996). The Multibeam is a thirteen-beam, cooled 21 cm receiver designed by the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) Division of Telecommunications and Industrial Physics and built by the Australia Telescope National Facility (ATNF). The Multibeam was installed at the prime focus of the ATNF-operated Parkes 64 m radio telescope on January 21, 1997. The Multibeam was funded by the ATNF and a large Australian Research Council grant, to undertake the ambitious HI (neutral hydrogen) Parkes All Sky Survey (HIPASS) project. The collaborating institutions are the Universities of Melbourne, Western Sydney, Sydney and Wales Cardiff; Mount Stromlo Observatory, Jodrell Bank and the ATNF.

The thirteen 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 receivers each. The feed horns are all identical, have diameters at the focal plane of 240 mm, and are sensitive to orthogonal linear polarisations of radiation in the frequency range 1.27–1.47 GHz. The instantaneous bandwidth is 64 MHz for each of the 26 channels.

1.2. The HIPASS Project

The Multibeam was funded to undertake the HIPASS project whose primary aim is to make deep, large-area surveys for neutral hydrogen emission from external galaxies. HIPASS consists of two major surveys:

1. a survey for 21 cm emission from redshift -1200 km s^{-1} to $+12600$ km s^{-1}, over the entire southern sky south of declination $+2°$, with an effective integration time of 430 s per pointing; and

2. a deeper (by a factor of two) survey of a part of the "Zone of Avoidance" (ZOA) region—specifically $l = 213°$ to $33°$ with $|b| < 5°$, with velocity range the same as for the all sky survey.

The scientific potential of these two major surveys is great. The HI mass and luminosity functions for the nearby Universe will be determined better than ever before. The HIPASS project will provide completely new information on the distribution of galaxies, the density parameter, the space density of optically rare and invisible galaxies, and on group and supercluster dynamics. In particular, the ZOA survey will search a part of the sky in which typically 10 magnitudes of extinction at optical wavelengths conceal the southern crossing of the Local Supercluster and the likely connection between the Hydra-Centaurus and Pavo-Indus-Telescopium superclusters.

Survey parameters. Staveley-Smith (1997) investigated observational techniques for optimising the sensitivity, and uniformity of sensitivity, of the all sky and ZOA surveys. Subsequently, an active scanning approach was selected for both surveys, whereby the telescope is driven across the sky at a rate of $1°$ min^{-1} and the Multibeam correlator is programmed to cycle every five seconds. For the all sky survey, $8.6°$ scans are made along lines of equal right ascension, spaced by 7.0 arcmin; ZOA survey scans of length $8.6°$ are taken along lines of equal galactic latitude, spaced by 1.4 arcmin.[1]

The total observing time for the all sky survey will be close to 3000 hr, during which time 17000 scans will be acquired, each scan file containing approximately 2700 spectra. For the ZOA survey, 10000 scans will be collected over a total observing time of 1700 hr. Given the large predicted investment of telescope, observer and analyst time, it became clear during the planning phase for the project that realtime calibration of the survey data was desirable, if not necessary. This would enable realtime visualization of the data, thereby

[1] During the commissioning of the Multibeam it became clear that any movement within the focus cabin led to horrendous baseline ripple in bandpass corrected spectra, thus parallactification is not enabled for HIPASS scanning. As a result, the scans made by the twelve non-central beams are slightly curved on the sky. In practice this does not matter, since both surveys have tremendous redundancy built into the scan scheduling.

providing immediate feedback to observers on system integrity and survey data quality. It was also evident early on that radio frequency interference (RFI) had the potential to seriously contaminate a significant fraction of the survey data. Thus any automated data handling algorithms had to be robust to a wide variety of RFI conditions, yet operate in realtime.

When development of a realtime software solution for the HIPASS project commenced, there were no systems in existence, to our knowledge, which could realise the dual objectives of realtime and robust data processing of radio astronomical data. It was necessary to develop a dedicated software package for the HIPASS project.

1.3. AIPS++

The AIPS++ (Astronomical Information Processing System) project (Croes, 1993; Glendenning, 1996) is a new, modern, object-oriented C++ software package designed to process (principally) radio astronomy data. AIPS++ is being developed by an international consortium of seven partners who amongst them operate most of the world's synthesis arrays, and many of the best single dish telescopes. At the core of AIPS++ is an extensive set of classes for storing and manipulating (mathematically and logically) large multi-dimensional arrays of data—the bread and butter of radio astronomy.

The decision to base the Multibeam Software on AIPS++ was taken for several strong reasons. First and foremost, AIPS++ promised to provide (and indeed has provided) what is essentially a *realtime environment* in which to access, manipulate and store or pipe (to further processes) radio telescope data. Secondly, through its complex but rich MeasurementSet storage paradigm, AIPS++ gives the programmer and user *direct access to their data* in the AIPS++ shell ('glish'). This was seen as a particular advantage for the trialling of new algorithms prior to implementing binary clients, and thus improved the speed of development and debugging of the Multibeam Software. Furthermore, at the time development of the Multibeam Software commenced, AIPS++ already provided *advanced, object-oriented techniques for inter-process communication*, and simple methods for displaying and controlling a *graphical user interface* (GUI).

2. The Multibeam Software–an Overview

The Multibeam correlator is programmed to write data into RPFITS (Radiophysics FITS) format files. These files were originally designed to store complex, cross-correlation, visibility data from the Australia Telescope Compact Array, and have been modified to store real auto-correlation data from the Multibeam receiver. In HIPASS mode, the correlator cycle time is 5 s, and spectra are written for each beam and polarisation at the end of each cycle. There are two polarisations per beam, so a total of 26 spectra are written each cycle, each having 1024 channels. Each channel value is stored as a single precision floating point number, which takes four bytes of storage. Thus the HIPASS raw data rate is 104 kb/cycle, or 1.2 Mb/min. The HIPASS RPFITS file is referred to as an HPF (HIPASS FITS) file, and is closed and reopened each cycle by the correlator to ensure that the very latest data can be read from the end of the file.

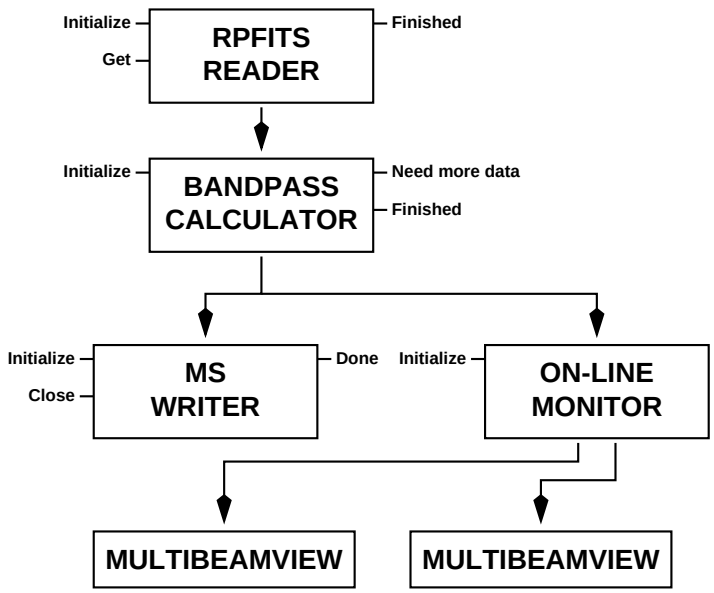

Figure 1. Flow diagram for the realtime Multibeam Software. The data is transported along the solid arrows as a glish record, which can contain several cycles of data, but normally contains only one. Left tags indicate the important commands that can be sent to each client; right tags indicate status events emitted by the clients, to which the LiveData script must respond.

The overall flow diagram for the realtime component of the Multibeam Software is shown in Figure 1. The Multibeam Software reads the HPF files in realtime, or as close to realtime as possible, and applies robust bandpass and baseline removal, velocity frame conversion, and spectral smoothing to all 26 spectra every cycle. The calibrated data is then written to an AIPS++ MeasurementSet. At this stage, realtime operations cease.[2] A complex, event-based glish script is used to control realtime processing of data by a set of binary glish clients; this script must recognise when new HPF files are available, queue these files, and set up control structures to process the next queued file whenever the processing system becomes idle. The glish control script, "LiveData", can be operated from the glish command line, or via an informative GUI (Figure 2). The GUI is written in glish. The script also has hooks so that the observer or data reducer can monitor the data on-screen using adapted visualisation tools from the ATNF "karma" library. In general, two on-line monitors are run concurrently to display the data for both polarizations of a selected beam. Bundles of spectra, usually 26 at a time, are transported in LiveData between the binary clients by means of a "Multibeam Glish Record", structured specifically for the Multibeam Software.

[2]However, we have plans for pipelining the data into a pseudo-realtime gridding client running on a remote machine.

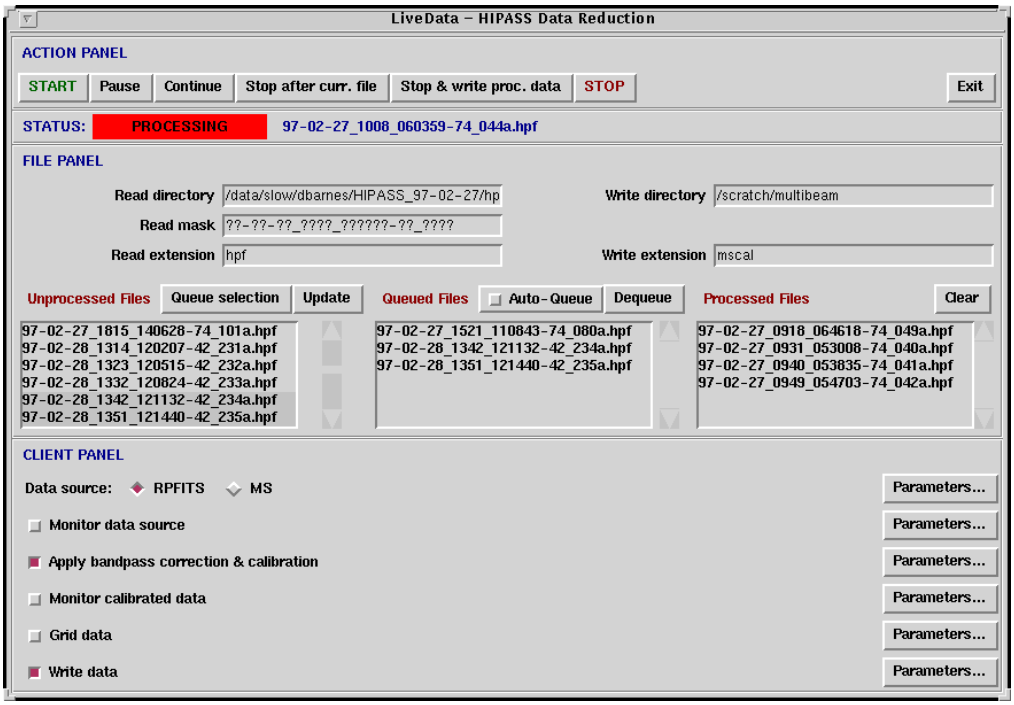

Figure 2. The LiveData graphical user interface—the user can dequeue files, including those which have been automatically queued, or change the order of queued files if the reduction system is falling behind the observing system.

Off-line, a number of MeasurementSets, typically of order 100 covering an area of sky of order 70 square degrees, are collected to be gridded into a position-velocity cube. The gridding software is relatively straightforward, although in lieu of a machine with a vast amount of memory (say, at least 1 Gb), the cubes need to be built in a number of velocity "slabs", and "glued" together afterwards.

3. Reading Realtime Telescope Data

A binary client was designed to read HPF files in realtime, and prepare a glish record structure (ie. a Multibeam Glish Record) containing an entire cycle's worth of data from the HPF file ("RPFITS READER" in Figure 1). The glish record is emitted from the client, and appropriate control structures in the LiveData glish script forward the record on to other clients. The RPFITS READER is able to package more than one cycle of data into a single Multibeam Glish Record. This can be useful when the reduction system is being operated in "catch-up" mode, and speed can be gained from the lower overhead of passing fewer, but larger, glish records to and from clients.

An alternative to the RPFITS READER was the combination of a client which converts HPF files to MeasurementSets, and a client which extracts Multibeam Glish Records from MeasurementSets. Whilst this approach proved too slow

because of the intermediate writing and reading of a MeasurementSet, this client is a useful utility which enables inspection of the contents of HPF files using the AIPS++ TableBrowser application.

4. Realtime Data Manipulation

A binary glish client (the "BANDPASS CALCULATOR") has been developed for the principal purpose of robustly removing the gross bandpass from the spectra in realtime, and calibrating the spectra using the system temperature values written to the HPF file by the correlator. This client also provides further data manipulation capabilities, specifically robust zeroth order baseline removal, spectral smoothing and velocity frame conversion. In cases where the system temperatures written by the correlator are known to be wrong, the client can prescale the raw spectra such that an approximate calibration is made. The BANDPASS CALCULATOR is the most sophisticated component of the Multibeam Software, and is also the most time consuming in the realtime pipeline. A brief description of the components of the BANDPASS CALCULATOR are given below; for further details refer to Barnes et al. (1998).

The BANDPASS CALCULATOR keeps spectral data in a four-dimensional buffer—the four axes are cycle, beam, polarization and channel number. For approximately two minutes from the commencement of a particular scan, the client fills this buffer without actually making any calculations, since at least two minutes of data are required for generation of a bandpass estimate. After this filling period, the behaviour of the client changes, and instead of simply emitting a request for more data, it begins to calculate and apply bandpass removal, baseline removal, spectral smoothing and velocity frame conversion; culminating in the emission of corrected spectra in a Multibeam Glish Record.

4.1. Bandpass and Baseline Removal

The spectra written by the correlator suffer a number of effects which must be removed during processing to obtain useful spectra of the sky. The most important of these effects is the presence of a "bandpass" in the correlator output spectra. The bandpass reflects the combined shape of all the filtering applied to the data by the receiver and correlator sub-systems, and may also include artifacts from strong contiuum sources in or even out of the beams of the telescope. Whilst the bandpass for spectra from different beams of the Multibeam may exhibit large scale similarities, they are expected to differ at smaller scales, since each of the 26 signals pass through independent receivers, amplifiers, down-converters, filters and correlators.

For the HIPASS project, where the telescope is actively driven across the sky, traditional bandpass removal techniques (eg. signal–reference subtraction) are not suitable. Instead, a statistical estimate of the bandpass at the position and time that a particular spectrum—the *target* spectrum—was acquired is made from a set of earlier and later spectra observed by the same beam—these are the *reference* spectra. The reference spectra are selected individually for each target spectrum, and must be independent (with respect to the telescope beam) measures of the HI sky to that of the target spectrum, but taken with timestamps near that of the target spectrum, so that time variations of the bandpass are

not significant. Furthermore, valid reference spectra must be from the same right ascension (or galactic latitude) scan series as the target spectrum, since movements of the receiver with respect to the telescope surface introduce gross phase changes in the bandpass.

Dividing the target spectrum by the statistical bandpass estimate yields the bandpass-corrected spectrum. This spectrum is then scaled using system temperature information, yielding a calibrated spectrum. The overall shape of the resultant spectrum is referred to as the *baseline* of the calibrated spectrum, and ideally should be flat. It is often the case, however, that an offset exists such that the mean (or median) of all the channel values for a calibrated spectrum is non-zero; and it is occasionally the case that low order curvature is present in bandpass-corrected spectra. Such curvature may be due to very long wavelength ripple of which only a portion of the cycle is visible in the spectra, or it may be some more complicated artifact of hardware or software processing. The final stage in calibration of the correlator spectra involves the removal of any flux density offset by subtracting from each channel value the median of all channel values for a particular spectrum.

4.2. Spectral Smoothing

Some form of smoothing is required to lessen the ringing that is associated with strong Galactic HI emission, and is caused by the sidelobes in the spectral response of the system. In some cases, ringing can be seen throughout the entire 21 cm spectrum from the correlator. Several smoothing techniques were applied to the data, in order to assess the best trade-off between loss of velocity resolution, and suppression of the ringing. A Tukey 25% filter was selected, and implemented in a C++ class where the filter is applied to the spectra in the Fourier domain. Although the suppression of the spectral sidelobes by this filter is not as strong as by a Hanning filter, this filter still gives good sidelobe suppression, and the loss in spectral resolution is only 15 per cent.

4.3. Velocity Tracking

Observations for the HIPASS project are made in topocentric mode, whereby the actual observing frequencies for each channel remain fixed throughout the survey. Consequently, the velocity range observed varies over the course of a year, and some correction must be made to the spectra. A C++ class was implemented to address this problem, and does so in the Fourier domain. The AIPS++ Measures classes were used to calculate the required velocity shift which was converted to a phase gradient; a subsequent inverse Fourier transform yields a spectrum in a fixed, barycentric reference frame. Note that the velocity frame conversion is applied *after* bandpass removal, since many components of the bandpass are internally generated in a fixed frame.

4.4. Realtime Monitoring Software

As one of the primary objectives of the development of the Multibeam Software was to enable realtime visualization of both raw and calibrated data, some time was spent investigating the best way to do this. The two options available at the time were AipsView, a young visualization tool which can be operated from glish; and the ATNF visualization software, specifically "kview". Whilst

AipsView could already accept glish events, it was necessary to introduce a client between AipsView and LiveData to translate the Multibeam Glish Records into simple three-dimensional glish arrays (having channel, cycle and beam number as axes) that AipsView could handle. This client (the "ON-LINE MONITOR"), can be used to mask out a particular channel range, or set of beams, or even apply averaging to the data to reduce the CPU utilization of the client which actually writes the spectra to the display.

Unfortunately though, we found AipsView non-intuitive to use, and it lacked some features that the ATNF in-house visualization software already had.[3] Furthermore, at the time, AipsView was not supported on the DEC alpha architecture. It turned out that we were able to add a simple C++ wrapper to one of the main ATNF applications, kview, so that it could handle glish arrays of the format being emitted by the ON-LINE MONITOR. This modification was very straightforward, and provided us with the visualization client ("MULTIBEAMVIEW") that is currently in use at Parkes for on-line monitoring of HIPASS data, and which retains all of the features and the same user interface as kview.

5. Data Storage

The calibrated, bandpass-corrected spectra are piped by LiveData to the final client in the realtime system, the MS WRITER. This client accepts incoming Multibeam Glish Records, and writes them to an AIPS++ MeasurementSet. We refer to the MeasurementSets written by the on-line reduction system as MSCAL files. The MS WRITER has been designed to write and read AIPS++ MeasurementSets which contain HIPASS data. The MSCAL files can be inspected directly using the AIPS++ TableBrowser application. Prior to archiving on CD-ROM media, the MSCAL files are converted to SDFITS (Single Dish FITS) files; this conversion does not take place in realtime.

6. Pseudo-realtime and Non-realtime Gridding

For the all sky survey, 75 scans are needed to produce full survey coverage cubes of the sky; for the ZOA survey, 375 scans are needed. Since a complete set of scans for a given cube are acquired over a number of months, gridding of spectra into final survey cubes is not done at the telescope. Indeed, as of September 1997, only two fully sampled all sky survey cubes had been observed and generated.

The gridding is done by a single glish client (the "LARGE SCALE GRIDDER") designed specifically to generate HIPASS cubes; although it can in principle grid spectral line data from any single dish radio telescope. Because of the large data bandwidth needed by the LARGE SCALE GRIDDER (the MSCAL files for a ZOA cube occupy 4.3 Gb), it was necessary to design this client to read MeasurementSets and write position-velocity FITS images (cubes) itself, ie. without transferring data through glish. The gridding algorithm at present is very simple,

[3]Indeed, AipsView is now frozen at its current release, and will be superceded by a number of AIPS++ applications based on the AIPS++ Display Library.

and will probably be altered prior to publication of the survey. For a particular pixel in the cube, the value for the ith channel is calculated by taking the median of all the ith channel fluxes from spectra having position stamps that are within six minutes of arc of the fixed pixel position. Whilst this gridding technique is robust to even quite strong interference, it unfortunately introduces a downward bias in the gridded fluxes, since this algorithm fails to take account of the beam shape.

7. Conclusion

AIPS++ was adopted by the HIPASS project in late 1995 as the platform for the planned realtime, object-oriented data reduction pipeline. The resultant Multibeam Software was successfully operating at the Parkes telescope from day one of the HIPASS project, in February 1997. As of November 1997, approximately 45 Gb of data have been robustly bandpass corrected, smoothed, and velocity corrected, *mostly in real time,* using a combination of AIPS++ infrastructure, and binary clients and glish scripts developed by us but based on the AIPS++ library. The HIPASS project members now look forward to extracting astronomy from calibrated HI cubes, rather than manually grinding their way through gigabytes of unprocessed spectra!

Acknowledgments. I am sincerely grateful to my collaborators on the development of the Multibeam Software for their ideas, time and direction: Lister Staveley-Smith, Taisheng Ye, Tom Oosterloo, Mike Kesteven, Warwick Wilson and Richard Gooch. We are grateful to the Australia Telescope National Facility, the Universities of Melbourne, Western Sydney, Sydney, and Wales Cardiff; Mount Stromlo Observatory, Jodrell Bank and the Australian Research Council for supporting the development of the Parkes Multibeam Receiver. We acknowledge the large investment of time and personnel towards the design and construction of the receiver by the CSIRO Division of Telecommunications and Industrial Physics. We were delighted with the frequent and high-quality help provided by the many AIPS++ programmers during the planning and programming of the Parkes Multibeam Software. DGB acknowledges the support of an Australian Postgraduate Award, and is grateful for the financial assistance to attend ADASS '97 provided by the ADASS POC and a Melbourne Abroad Scholarship.

References

Barnes, D. G., Staveley-Smith, L., Ye, T., Oosterloo, T. 1998, this volume

Croes, G. A. 1993, in ASP Conf. Ser., Vol. 52, Astronomical Data Analysis Software and Systems II, ed. R. J. Hanisch, R. J. V. Brissenden & Jeannette Barnes (San Francisco: ASP), 156

Glendenning, B. E. 1996, in ASP Conf. Ser., Vol. 101, Astronomical Data Analysis Software and Systems V, ed. George H. Jacoby & Jeannette Barnes (San Francisco: ASP), 271

Staveley-Smith, L. 1997, PASA, 14, 111

Staveley-Smith, L., Wilson, W. E., Bird, T. S., Disney, M. J., Ekers, R. D., Freeman, K. C., Haynes, R. F., Sinclair, M. W., Vaile, R. A., Webster, R. L., & Wright, A. E. 1996, PASA, 13, 243