# WHISP Reduction Document

Dolf Sijbring and Jurjen Kamphuis

September 12, 1995

# Contents

# Part 1

# The NEWSTAR phase

# 1 Prologue to the NEWSTAR phase

The objective of the WHISP project is to observe a large number of carefully selected galaxies in the 21-cm line with the WSRT, to reduce the data in a uniform way, and to store the results in a database for general use.
In this document we describe the procedures by which we propose to reduce and archive the WSRT data. We do this by following the steps in the reduction, meanwhile justifying the decisions we have taken along the way.
This is the first part of the document, where we describe the reduction of the *uv*-data and the Fourier transforming of the data to the map plane. For this we use the NEWSTAR reduction package.

# 2 Loading raw data from tape

In the first step of our reduction procedure the Westerbork data are read from DAT-tapes, using NSCAN (option load) from NEWSTAR. The current input unit is number 8. The name of the scan-file is the name of the observed galaxy (e.g. U11861(.SCN)). This file contains both the calibrators and the actual observation.

## 2.1 Items to record in a database

- name galaxy
- input tape name (e.g. W93003)
- sequence number of galaxy observation
- tape labels (e.g. 3677-8:10)
- order of calibrator(s) and observations (e.g. 3C286, U11861, 3C147)
- observing frequency and velocity (e.g. 1413.46 MHz and 1484 km/s)
- total bandwidth (e.g. 2.5 MHz)
- # of channels (e.g. 128)
- # of observing hours (e.g. 12:00)
- integration time (e.g. 120 seconds)
- shortest spacing (e.g. 36 m)

# 3 Calibrators: self-calibration and editing

## 3.1 Introduction

Our goal in this part of the reduction is to obtain accurate corrections for the gains and phases of the data. These corrections are derived from a comparison of the observed visibilities of known sources (calibrators) with theoretical visibilities from a model of that source. This comparison is done by making use of the *selfcal* method. For most observations there is one calibrator observation before and one after the intended observation. The *selfcal* method derives a correction for all (14) telescopes, which are copied, when approved, to the actual data. For each telescope an average correction for amplitude and phase is copied from calibrators to data.

## 3.2 Self-calibration

In the first step we run a blank *selfcal* (NCALIB) on the calibrator(s) and monitor the behaviour of the solutions. As a rule interferometer 9A will be excluded from the *selfcal* solution, because it almost always suffers from electromagnetic interference (EMI) of the sun. Typical gain and phase errors per channel are 80 W.U. for a 2.5 MHz, 128 channel, 120 seconds integration observation. When a (reasonable) solution is obtained, we inspect the *uv*-cube (NMAP for an indirect, or NGIDS for a direct image of the data) of the calibrator data, where we have subtracted the model from the data. If the calibrator data are of 'perfect' quality and the *selfcal* solutions are correct, we expect to see a cube with noise.

## 3.3 UV-inspection and editing

However, life's a bitch. A lot of data is hassled by imperfections, like correlator spikes, EMI or correlator problems. It is not our intention in this phase of the reduction to obtain perfect noisy residuals. Any editing of the data must be seen in the light of achieving good corrections for the actual data. This means that editing is only necessary if the bad data influence the corrections calculated for a telescope. In practice then, only errors occurring over large hour angle ranges, or with a high S/N, will have to be edited. If any editing has taken place the *selfcal* has to be redone.

There are many ways for marking bad data, some of which are:

- interactive blotting in NGIDS
- clipping on a value in NFLAG
- flagging by using the residual of data − model
- flagging by using the noise of the *selfcal* solution

The last two methods are only feasible when a selfcalibration of the data has been done. This is the case for the calibrators, but not for the actual data.

## 3.4 Copying telescope corrections to data

When both the *selfcal* solutions and the residual *uv*-cube are of good (enough) quality, the telescope corrections can be copied to the actual data (NCALIB option: SET, COPY). The corrections are averaged between the calibrators and are the same for all hour angles of the corrected observations.

## 3.5 Items to record in a database

- the interferometers excluded from the *selfcal* solution (e.g. 9A)
- short description of the editing applied (e.g. clipped all $>$ 400 W.U.)
- which calibrators are used to copy corrections from (e.g. 3C295 and 3C147)

# 4 UV-inspection and editing of galaxy data

## 4.1 Introduction

In this section we briefly describe the examination and editing of the observations of the galaxy. The observations are often polluted by imperfections caused by correlator problems, electromagnetic interference or malfunction of a telescope. Because these errors usually originate in the *uv*-plane, the observations can therefore best be studied in this domain.

## 4.2 UV-inspection

In order to inspect the *uv*-data we display the data-cube as maps with interferometer plotted against hour angle, as a function of frequency for both the XX and YY polarization (NMAP or NGIDS). We only produce the amplitude information because this parameter does not show large fluctuations per baseline and can therefore be reliably inspected. (In contrast, the phase can be a rapidly varying parameter and therefore it is difficult to judge its behaviour.) The *uv*-cube can be viewed from different angles (NGIDS). Bad *uv*-data will light up as single points or as stripes if the Look Up Tables (LUT's) are correctly adjusted.

Note: If strong *extended* continuum emission is present, it could be desirable to subtract this emission from the data before making interferometer - hour angle maps as a function of frequency. The continuum emission can be approximated by making a channel map of the UV data (NMAP). The continuum emission has to be modelled (NMODEL for point sources and NCLEAN for extended sources) in order to subtract it from the *uv*-data in the scan-file.

## 4.3 UV-editing

In judging the quality of the observations, knowledge about the response of the errors in the *uv*-data on the Fourier transformed images is necessary. A good starting point is that the quality of the maps in the image plane is all that counts: in the ideal case, the noise distribution in these maps is free from imperfections and approach the theoretical value. In practice, this means that the errors in the image plane should be suppressed well below the noise level in the maps. We only delete, therefore, bad *uv*-data that significantly disturb the (Fourier transformed) image.
As described in section 3, the editing of the *uv*-data can be done in several ways. In this stage of the reduction, we only use the FLAG-option in NGIDS, which manually flags the data points, or the CLIP-option in NFLAG, which flags the data above a certain value. Because the eye is sensitive to coherent structures, the FLAG-option in NGIDS seems the most suitable way to edit extended low S/N corrupted UV data, but it is more laborious.
A list of edited data points can be obtained from the scan-file (NFLAG option: FLAG; FLAG-OPTION=FLIST; OPS-FLIST=GET).

## 4.4 Criteria for quality assessment

We use for the evaluation of the *uv*-data a one-word description, which summarizes the quality of the data. We devide the data in the following 6 quality groups:

- excellent: no editing of the data is necessary.
- good: minor editing of the data; <10% of the data is flagged; e.g. a few hour-angle scans or 1 or 2 (parts of) baselines, up to 1 missing telescope, but not A, B, C or D.
- reasonable: editing of >10% of data necessary, but <25%; e.g. 1 telescope flagged in XX and YY up to one of telescopes A, B, C or D.
- usable: >25% of the data is flagged, but <50%; e.g. a few telesopes are not recorded (including A, B, C or D); reobserving is an option.
- bad: >50% of the data is lost; e.g. XX or YY lost; reobserving must be considered seriously.
- unusable: <25% of the data is usable; object must be reobserved.

## 4.5 Items to record in a data base

- short description of the editing applied (e.g. clipped all > 400 W.U.)
- quality assessment of the observations

# 5 Fourier transformation

The final reduction step in the *uv*-domain is the (Fast) Fourier transformation (FFT) of the data to the sky plane. The output is a set of maps, the so-called channel maps, consisting of the sky brightness distribution as a function of frequency. In the FFT we have to choose the 'format' of the most important part of the data-base, i.e. the 'format' of the actual data of the galaxy. In particular, we must choose the (angular) size of the channel maps, the sizes of the pixels, the different kinds of tapers and the (angular) resolutions of the data. The last item is extensively discussed in section 6. Here we also decide which antenna patterns (i.e. the point spread function) we want to create for removing most of the instrumental effects in the channel maps.
In the section 5.1 we summarize the standards for the channel maps and antenna patterns (only for the full resolution data). In section 5.2 we briefly discuss the justification of these standards (for the $30''$ and $60''$ resolutions see section 6) .

## 5.1 Standards for channel maps and antenna patterns

- pixel sizes are $5'' \times 5''$
- FFT size is $1024 \times 1024$
- field size of channel map is $512 \times 512$ pixels ($=42\rlap{.}'7 \times 42\rlap{.}'7$)
- field size of antenna pattern is $1024 \times 1024$ pixels
- one antenna pattern every $25 \times (\frac{\#channels}{128})$ channel maps
- *uv*-weighting is inversely proportional to track density of *uv*-data (standard)
- *uv*-tapering function is Gaussian
- convolving function is Exponential-Sinc
- type of coordinates are B1950
- uniform velocity taper

## 5.2 Justification

We only discuss here the standards of the maps for the full resolution data. The 'smoothed' data will be discussed in section 6.
After carefully calibrating and editing, the *uv*-data of the actual observations is Fourier transformed (NMAP) to the sky plane as a function of frequency (or heliocentric radial velocity). Since the antennas have a limited field of view, the sizes of the channel maps should be roughly as large as the size of the primary beam, which at $\lambda$=21-cm emission is given by $\cos^6(\mathrm{C} \times \mathrm{RF})$ ; (RF is radius in degrees times frequency in GHz; C is a constant

weakly dependent on frequency). Furthermore, the sampling of the pixels should satisfy the Nyquist rate, about 2 pixels per beam (this sampling also implies a complete coverage of the *uv*-plane). Because the angular resolution for a maximum baseline of 2700 meter is $11'' \times 11''/\sin\delta$ (no $w_t$, see below), the pixel size is chosen to be $5'' \times 5''$ . A map size of $512 \times 512$ pixels is therefore sufficient to cover the primary beam at the HPBW. Moreover, the field size is sufficiently large for an adequate sampling in the *uv*-plane ($\Delta u = \frac{1}{fieldsize}$). In order to control the form of the synthesized beam to some extent, a taper (w) is applied to the *uv*-data. This taper consists of two different tapers. The *first* taper ($w_u$) regulates the area density of the data (i.e. visibilities) in the *uv*-plane. Though the WSRT fills the *uv*-plane at regular intervals on elliptical tracks, the visibilities are more densely sampled at short baselines, because the measurements are taken at uniform increments in hour angle. To obtain a uniform area density of data points throughout the *uv*-plane, a weighting (inversely proportional to the track density) has been applied. A *second* taper ($w_t$) has been introduced to reduce the depth of the sidelobes of the synthesized beam. We apply a Gaussian that reduces the visibilities to 25% of their value at the maximum baseline. This results in a reduction of the depth of the first sidelobe from 13% (no $w_t$) to 5%. (Note: $w = w_u \times w_t$)

A third taper is necessarily introduced by the FFT. The FFT regrids and resamples the *uv*-data. This leads to two effects. First, the maps are multiplied by a convolving function. Second, the maps are replicated at intervals equal to the sizes of the maps, which introduces aliasing. Therefore, the choice of the convolving function is of main concern in the use of the FFT. It would be ideal to have a rectangle as a tapering function in the map plane. Since a rectangle is a sinc in the *uv*-domain, which requires too much computer time to convolve with, another solution would be to use a truncated sinc. Instead of a sharp cut off, this gives weird effects at the edges of the maps. We choose therefore the widely used compromise of an Exp-Sinc function (i.e. sinc $\times$ exponential). This is, by experience, a good alternative and reduces the aliasing to acceptable levels. Also, to suppress the aliasing, the FFT size of the maps is taken as $1024 \times 1024$, four times as large as the final maps.

In order to have the possibility of CLEANing the entire data-cube, we produce antenna patterns that are twice as large as the channel maps. Because the instrumental effects depend on frequency, we make antenna patterns at about every 25 channels, sufficient to suppress the unwanted effects below the noise level.

# 6 Smoothing

## 6.1 Introduction

In order to emphasize different scale sizes in the emission, we also smooth the original data to $30'' \times 30''$ and $60'' \times 60''$ angular resolution. The smoothing will be done in the *uv*-domain with NEWSTAR, by a circular weighting of the *uv*-data. This has three advantages over smoothing in the xy-plane: 1) it is 4 times faster (about 1.5 times faster than *uv*-smoothing in GIPSY); 2) no 'lost' edges; 3) we can select the mapped area of the smoothed maps independently of the full resolution maps. This last feature is very convenient if one wants to include a larger area of the sky in the smoothed maps (with a larger pixel spacing) than in the full resolution maps (e.g. for searching serendipitous H I clouds).

## 6.2 Different kinds of smoothing

In Table 1 we list the beam-sizes resulting from three different smoothing methods possible in NEWSTAR and their combinations. The normal taper method in general leads to non-circular beams (not clear from the table, since the declination of the galaxy is 73° ). The *uv*-area selecting method, where part of the *uv*-area is left out, leads to non-circular beams with large sidelobes (even with a taper). In order to obtain circular beams, the best method, we believe, is to use the circular weighting method. The difference between the beam resulting from the circular-weighting method and the 'ordinary' xy-plane smoothing is nowhere larger than 0.02%.

Table 1: Different types of smooth in NMAP (at a declination of 73° )

| # | name | taper | value | circ-weight | value | *uv*-area | beam-size |
|---|---|---|---|---|---|---|---|
| 1 | full | gauss | 2548 | natural | - | $*$ | $12.01\times11.55$ |
| 2 | ga1274 | gauss | 1274 | natural | - | $*$ | $19.07\times18.36$ |
| 3 | ga1274.cw1274 | gauss | 1274 | gauss | 1274 | $*$ | $26.27\times25.83$ |
| 4 | nat.cw1274 | natural | - | gauss | 1274 | $*$ | $18.36\times18.31$ |
| 5 | nat.nat.uv1274 | natural | - | natural | - | 1274 | $22.82\times21.78$ |
| 6 | ga1274.nat.uv1274 | gauss | 1274 | natural | - | 1274 | $25.02\times24.90$ |
| 7 | ga1274.cw1274.uv1274 | gauss | 1274 | gauss | 1274 | 1274 | $29.17\times28.47$ |
| 8 | nat.cw1274.uv1274 | natural | - | gauss | 1274 | 1274 | $24.86\times24.70$ |

nat=natural means no weighting
ga=gaussian means gaussian taper, not circular
cw=circular weighting means gaussian taper, circular
uv means *uv*-area selected

Table 2: Smoothing at different declinations

| | Declination | beam-size |
|---|---|---|
| c_weight = 1200 m | | |
| | 15 | 39.81×20.19 |
| | 20 | 32.30×20.06 |
| | 25 | 25.06×20.18 |
| | 30 | 21.36×21.26 |
| | 45 | 19.78×19.73 |
| | 60 | 19.48×19.44 |
| | 75 | 19.45×19.35 |
| | 90 | 19.38×19.38 |
| c_weight = 780 m | | |
| | 20 | 34.54×30.28 |

## 6.3 Declination effects on the circular weighting smoothing

In Tabel 2 we have investigated how the beam-size depends on the declination of the observation when a constant circular weight value of 1200 m is used. From the first part of the table it is clear that at low declinations (<30°) the circular weighting leads to a non-circular beam. This is because the projected baseline at these low declinations becomes less than 1200 m. At a lower circular weight value of 780 m (necessary to achieve a beam-size of 30″ , see below) the non-circularity of the beam is already much less severe. For even larger beam-sizes of 60″ the circular weighting provides no problems at all.

## 6.4 Beams resulting from different taper values

In Table 3 we have listed for a declination of 30°, how the beam-size depends on the taper value selected for the circular weighting. To achieve beam-sizes of 30″ and 60″ we find taper values of 780 and 390 m, respectively. The table shows that the resulting beams are indeed circular within a few tenth of an arcsecond. For reference we have also included the full resolution beam with a normal tapering of the data.

As indicated above, we can select our mapped area of the sky in the smoothed maps independently from the full resolution maps with the circular weighting method. We propose to make 512 × 512 maps of the smoothed 30″ maps with a pixel size of 10″ × 10″, so that we have 4 times the area of the full resolution maps. These maps are than ideal for searching for serendipitous H I. Beyond these dimensions ($85\farcm3 \times 85\farcm3$) the primary beam attenuation is already too severe to observe any H I gas. This seems to us a more economic way to use

Table 3: Smoothing with different taper values: declination is 30°

| taper value | beam-size |
|---|---|
| 1200 | 21.36×21.26 |
| 1000 | 24.78×23.73 |
| 800 | 29.40×29.67 |
| 790 | 29.74×29.67 |
| 780 | 30.08×30.03 |
| 600 | 38.83×38.57 |
| 400 | 57.20×57.83 |
| 390 | 59.69×59.32 |
| 380 | 61.26×60.88 |
| 300 | 77.63×77.17 |
| full | 23.35×11.68 |

our available disk, tape and cpu resources. For the maps with 60″ beam the map size is 256 × 256 with a pixel size of 20″ × 20″.
In summary, for smoothing the data to 30″ angular resolution we produce 512 × 512 maps with a pixel size of 10″ × 10″ using a value for the circular weight taper of 780 m. For smoothing the data to 60″ angular resolution we produce 256 × 256 maps with a pixel size of 20″ × 20″ using a value for the circular weight taper of 390 m. The taper values of the circular weights (780 and 390 m) are independent of the declination of the observations. In observations were large hour angle ranges are missing, additional steps have to be taken to make the beams circular (e.g. smoothing in the xy-plane with a 'correcting' beam).

## 6.5 Items to record in a database

- status of the 3 resolution maps (done = yes)
- description of map quality (single word valuation (good, reasonable, etc) + description of any map defects left (if any))
- noise in maps for all resolutions (e.g. 1.4 mJy/beam, 2 mJy/beam and 2.9 mJy/beam)

# 7 From NEWSTAR to GIPSY

## 7.1 Writing FITS files

Here we describe the first phase of the archiving of the data and the change from the NEWSTAR environment to the image reduction package GIPSY where the final reduction of the data cubes in the map plane will be done. At this stage of the reduction, the *uv*-data have been edited, calibrated and Fourier transformed to channel maps and matching antenna patterns. In order to archive and to continue the reduction in GIPSY (possibly at a later time), we store these files on DAT-tape. We use GIPSY because it provides a good collection of programs for handling three-dimensional data-cubes.

## 7.2 Archiving data on tape: phase 1

We will record the data resulting from the NEWSTAR reduction on two tapes. The edited SCN-file will be archived with 'tar' on a tape with label-format WS000n, where 000n is a number. On this tape UV-data from several galaxies can be stored. This tape is for internal use only and can also be considered as a backup for the maps, since the maps can easily be derived from these scan-files. The FITS-files will be stored on a tape with label-format WF000n with 32 bits real precision (NMAP with option=wrlfits, cubic=yes, fits-scale=wu). These tapes will (at the moment) contain maps of one galaxy: maps at 3 different resolutions (full, 30″, 60″) + matching antenna patterns. At a later stage also the cleaned versions of the 3 data cubes will be stored on this tape. Also a blotched cube at 60″ resolution, which has been used as a mask to clean all 3 resolution cubes will be archived. For the naming of the different files we refer to Appendix A. The names of the cubes can be found with MTLIST in GIPSY in the comment field.

## 7.3 Loading data into GIPSY

In order to load the channel maps (and antenna patterns) from disk into GIPSY (with RFITS), we have to rename their FITS files to 'file00000n.mt' with n is an integer. In RFITS, these files get the number n-1 (option filenumber=n-1). When the files are read from tape the filenumbers are simply the numbers on tape (starting with 0).
The data cubes have for 127 channel maps in GIPSY a third axis (i.e. the velocity/frequency axis) ranging from -63 to 63 instead of 1 to 127 used in the NEWSTAR format.

## 7.4 Items to record in a database

- names of archive tapes

# Part 2

# The GIPSY phase

# 8 Prologue to the GIPSY phase

The objective of the WHISP project is to observe a large number of carefully selected galaxies in the 21-cm line with the WSRT, to reduce the data in a uniform way, and to store the results in a database for general use.
In this document we describe the procedures by which we propose to reduce and archive the WSRT data. We do this by following the steps in the reduction, meanwhile justifying the decisions we have taken along the way.
This is the second part of the document, where we describe the reduction of the data in the map plane. For this we use the GIPSY reduction package.

# 9 Continuum subtraction

## 9.1 Standard procedure

In normal cases the continuum subtraction will be done in GIPSY using CONREM. At every pixel position CONREM fits a polynomial through the data cube. The standard procedure is to fit a first order polynomial (NPOLY=1). For the definition of the continuum emission we will, in a 128 (actually 127) channel observation, not use, besides the line channels and any channels containing isolated galactic foreground line emission, channels $-63$ to $-59$ at the low-velocity side of the band and channels 52 to 63 at the high velocity side. These channels have already a degraded noise level due to the bandpass of the Westerbork observations. CONREM still gives good results for continuum sources as strong as (several) hundred W.U. For a 64 channel observation the channels $-31$ and $-30$ at the low velocity side and channels 26 to 31 at the high velocity side of the band are routinely left out of the continuum subtraction.

As an extra check, we will form two continuum maps, one defined from the low velocity side and another from the high velocity side of the cube. The difference of these two maps is inspected to see whether any residuals are left (because e.g. the extent of the HI was underestimated). If any residuals are still visible, the range of continuum channels will be redefined.

After the continuum has been subtracted the continuum free cube will be checked for residual non-line-emission impurities.

## 9.2 Considerations for strong continuum sources

The grating rings of very strong (point) sources ($> 300$ W.U.) can leave residual rings when trying to remove them by fitting a first order polynomial. This is because the radius of the grating rings is proportional to the frequency of the observations. A 1000 W.U. point source e.g. placed at a distance of $1700''$ of the center of the map leaves residual rings of order 2 W.U. (the primary beam corrected flux density of such a source would be $\sim 6000$ W.U.!) after continuum subtraction, while the noise in our maps is of the order 0.5 W.U. (before the use of the MFFE). At the moment we use two possible remedies for such special cases:

- Use NPOLY=2 in CONREM. With this option in CONREM a second order polynomial is fitted to every pixel position. When used on the cube with a 1000 W.U. point source, this procedure leaves no visible residuals in the channel maps. The bandwidth of our 12h observations is, at the moment, not larger than 2.5 MHz. The behavior of the grating ring response can be well represented by a second order polynomial in this case. For larger bandwidths NPOLY=3 might be necessary.

- Subtract the continuum directly from the *uv*-data. This can be done in NEWSTAR, either by cleaning an extended source or be searching the position and flux of a point source with NMODEL (option FIND). This method, however, is very labour-intensive and should only be used in very special, difficult cases.

For data with continuum sources with peak flux densities <200 W.U. we will use the standard continuum subtraction, where CONREM is used with NPOLY=1. In all other cases we will use one of the special remedies.

## 9.3 Items to record in a database

- channels used for continuum subtraction in CONREM (e.g. $-58$:$-45$ and 40:51)
- noise in line channels for 3 angular resolutions (e.g. 1.5 mJy/beam, 2.1 mJy/beam and 3 mJy/beam)
- description of line map defects (if any)
- flux + position of continuum source(s) (if subtracted from *uv*-data)

# 10 Cleaning of maps

## 10.1 Defining search areas

In order to remove instrumental effects like grating rings and a negative baseline level the continuum-free maps containing line emission are cleaned. Although we could search clean-components over the entire maps, this is not done for both practical and theoretical reasons. The practical reason is, of course: time considerations. Searching clean components over the entire maps is much more time consuming than over a restricted area of the maps. The theoretical reason is that when the search area for clean-components is too large the number of independent positions in this area will exceed the number of measured *uv*-points (~30000) and the total flux in the clean-components will no longer converge (see for an example Figure 1).

Because of these reasons one usually defines search areas in the maps containing line emission. These search areas are made interactively by defining polygons around the area(s) of emission. The areas defined by the polygons are used as masks in the cleaning process, where only emission inside the masks will be searched for clean-components. Every clean-

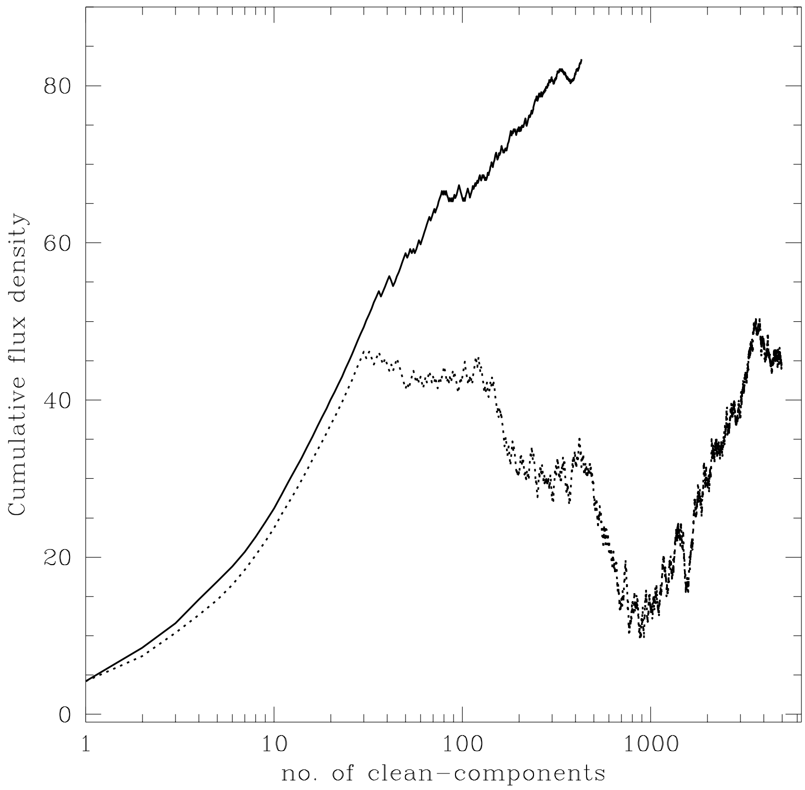


Figure 1: Comparison between the cumulative flux densities of two cleaning attempts using different search areas in the same map. Both areas have been cleaned to half the noise level. The dashed line represents the cleaning of an area of 300×300 pixels, i.e. an area with ≫30000 pixels. The solid line represents an area slightly larger than the area with emission. Note the logarithmic x-axis. The 'dirty'-flux density is about 80.

component, with its antenna-pattern, however, is subtracted from the entire map to ensure that no instrumental effects remain in the maps.
In a later stage the flux density of the clean-components will be used to determine the global profile of the galaxy. This is one of the reasons why it is important to include all the emission in the masks. For defining the search areas one would, ideally, like to have maps with the highest possible signal to noise. Unfortunately, the S/N ratio is a local quantity in the map, depending on the geometry and flux density of the source and on the beam size and noise level of the observation. A way to improve the S/N ratio for *extended* low-level emission regions is to smooth the maps. On the other hand, the S/N ratio of point sources will degrade in such a process. The best we can do is to use a compromise between these two extremes. As a first step, we will use our 60″ maps to define the search areas for the cleaning interactively. Before selecting the search areas we will apply a Hanning smoothing on our uniform tapered 60″ data. This will increase the S/N ratio of line emission 'extended' in velocity. In defining the polygons we will not follow the geometry of the emission very tightly, but only globally (see also §12.2 for additional remarks on defining masks for velocity fields).

## 10.2 Cleaning

After the masks have been defined, the 60″ uniform tapered maps corresponding to these masks are cleaned. The cleaning is stopped when the flux density of the clean-components is below a certain threshold: the clean limit. We will clean our maps to a clean limit of $\frac{1}{2}\sigma$, where $\sigma$ is the r.m.s. noise in the maps.
As a precaution, we will redefine the masks by using the cleaned Hanning tapered 60″ channel maps, where the instrumental effects (like a negative bowl and grating rings) have been reduced below the level of the noise. Without the instrumental effects it is much easier to define the exact areas of emission. This time, however, we first clip our Hanning smoothed 60″ maps to a level of $1.8\sigma$. The search areas are now no longer polygons but defined tightly around the 60″ emission (and some of the noise). After the clipping, we remove, by blotching, the areas where only noise peaks ($> 1.8\sigma$) are found.
This alternative way of defining search areas has several advantages: 1) speeding up the cleaning process (especially those of the full and 30″ resolution data); 2) the ring-like structures outside the search areas, which are a consequence of cleaning noise, are much reduced; 3) a larger percentage of the flux measured with single dish is retained in the global profile; 4) better S/N in the total HI and velocity field maps. A disadvantage of this method is that the uniform tapered 60″ data might suffer from 'creating emission' in the total HI map. This is due to the fact that we only pick out positive noise peaks at the edges of the emission (though the effect is reduced, since we are defining our masks from the Hanning smoothed data and not the uniform data). The full and 30″ resolution data are free from the effect.
The redefined masks are used to clean the 60″ maps again and will be used later to make the total HI map. After the 60″ cube has been cleaned again, the masks will be scaled to the cubes of the full and 30″ resolution and used in the cleaning of those cubes.

To every group of $25 \times (\frac{\#channels}{128})$ channel maps belongs one antenna-pattern. This antenna-pattern is used to clean the appropriate maps.
When the cleaning is finished, the (primary beam corrected) total flux density of the clean-components and residual 'dirty' flux (cf. §11.1) are stored in every map header. Also in a table belonging to the map the flux densities and positions of all clean components are stored. The cube with residual or restored maps will next be inspected to see if any emission is still missed in the search areas. This is particularly important when a large galaxy or two or more galaxies in one field are observed. In these cases the grating rings may intersect with the line emission, making it difficult to define the search areas well. If any emission is still in the residual maps we will have to define the masks again and redo the cleaning. If no residual, except noise, is visible in the maps we can restore the clean-components with the appropriate 'clean' beam.

# 11 The global profile

## 11.1 The flux density in a global profile

After the maps have been cleaned, both the flux density in the clean components (CFLUX) and the residual flux density (RFLUX), calculated from adding intensities and dividing by the dirty beam over a rectangular area around the search area in the residual map, are stored in the header of the map. The usual way to derive a global profile for a galaxy is to add these two, primary beam corrected, numbers for every channel. This way the flux density as a function of velocity can be derived. The most appropriate cube for deriving the values of CFLUX and RFLUX is the 60″ data cube.
There is a snag, however. The pitfalls concerning the value of CFLUX have already been described in §10.1. The value of RFLUX, however, is also not unambiguous. The value of RFLUX depends on the size of the rectangle that is used for the estimate of the residual flux. The effect is shown in Figure 2. In the upper panel we have plotted the estimated flux in a box with half-axis i, with its center shifted to position (−4,−4). As map we used

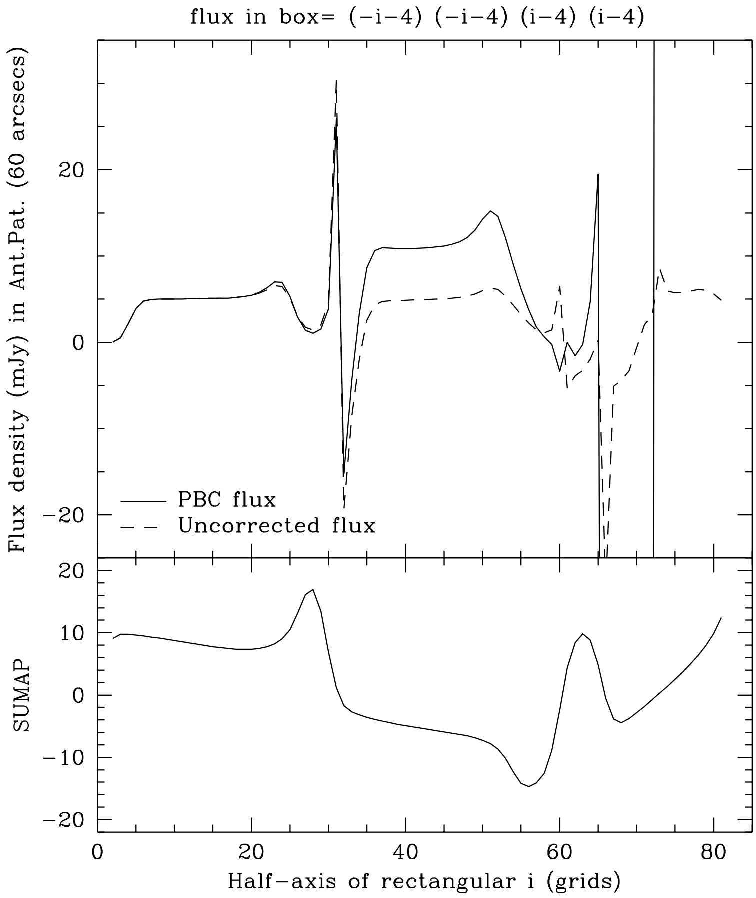


Figure 2: (Upper panel) Flux density measured in an antenna pattern within a box with axes 2 × i and its center shifted to position (-4,-4). Flux densities corrected and uncorrected for primary beam attenuation are shown. (Lower panel) Corresponding SUMAP of the box.

an antenna pattern to simulate a point source of 5 mJy. In the lower panel we plotted the SUMAP, i.e. the integral over the antenna pattern, of the box.

Figure 2 clearly shows that with many box sizes one finds the wrong flux density. Especially the primary beam corrected flux density only has a small range ($5 < i < 25$ in this particular case) of box sizes which give a reasonable estimate of the 'true' flux density of 5 mJy. The uncorrected flux density is somewhat better, but still 50% of the box sizes give erroneous estimates. The estimates are very bad were the value of SUMAP, as a function of area, is changing fast and close to zero. This happens near the grating rings.

The residual flux density of maps cleaned to a level of half or one third the r.m.s. can therefore still be considerable, when the box size is chosen wrongly. It is, however, difficult to determine the right box size *a priori*, because the combined behaviour of the sum, SUMAP and primary beam correction is different for every situation.

We believe the only way to solve this problem in general, is to use only the CFLUX in our global profile and account for the residual flux in the error estimate of flux density. The value of RFLUX will be recorded in the table of the global profile data. As an internal check we will compare the global profiles derived from the $60''$ and $30''$ cubes.

The global profile will be recorded both as a table (for all 3 resolutions + Hanning smoothed $60''$ data) and as a figure in the database (values from the Hanning smoothed $60''$ data). We will record the primary beam corrected value of CFLUX, but the uncorrected values of RFLUX and DFLUX (=dirty flux). The ratio of RFLUX over DFLUX can serve as an estimate of the error in the value of CFLUX.

## 11.2 Items to record in a database

- table of global profile values (velocity, CFLUX, RFLUX and DFLUX)
- figure of the global profile based on CFLUX only
- estimate of total H I flux + error
- H I global profile width at 50% and 20% + error
- systemic velocity derived from error-weighted global profile widths

# 12 2-dimensional reduction

Besides the reduction of the data to 1 dimension (the global profile), we also want to give the reduction to 2 dimensions: the total HI map and the radial velocity field. The products of the 2-dimensional reduction should be regarded with more skepsis, because, especially in the case of velocity fields, there are more roads than one that lead to Rome. We regard them as first estimates, that may highlight (peculiar) properties of the HI and can be used in the future as initial estimates in CUBEFIT.

## 12.1 The total HI map

The most straightforward reduction to 2 dimensions is the total HI map. The final masks used for the cleaning are used again (see §10.2) to select the areas of HI emission, setting the rest of the maps to zero. This way we add as little as possible noise of maps containing no emission at a certain position to that position in the total HI map. The masks are *not* defined very tightly around the HI emission of the full and 30″ maps to avoid the 'creation' of emission by selecting only positive noise peaks at the edges of the HI emission. After the areas have been selected the maps in the cube are added, for all three resolutions. Finally, the maps are corrected for primary beam attenuation.
The total HI map is also used to create a signal-to-noise map, where the noise depends on the position in the map and is calculated using:

$$\sigma_{Tot.HI} = n_l^{\frac{1}{2}}\, \sigma_{c.s.} \sqrt{(n_c + n_l)/(n_c - 1)}.$$

Here $\sigma_{Tot.HI}$ and $\sigma_{c.s.}$ are the noise in the total HI map and in each individual, continuum subtracted map, respectively. The number of line channels is denoted by $n_l$ and the number of continuum channels by $n_c$. The formula is valid for data with a uniform taper. This S/N map will be used later for the definition of the area were we can sensibly derive a velocity field.

## 12.2 The radial velocity field

The velocity field can be derived in many ways. In the past mostly three methods have been used for deriving the velocity field: (1) the intensity weighted mean velocity; (2) the velocity of the peak of the profile; and (3) by fitting a Gaussian to the profile and taking the velocity corresponding to the peak of the Gaussian. All methods have their advantages and disadvantages. The method of intensity weighted mean velocity (IWMV) has the advantage that it is simply a statistical descriptive method with no assumptions about models etc. The disadvantage is that if the profiles are asymmetric the IWMV no longer is the same as the velocity of the peak (the velocity of the peak is what we believe to be, in the absence of detailed models, the best approximation to the (radially projected) velocity of the rotation curve).
The second method, which takes the velocity of the peak of the profile, avoids the problem with asymmetric profiles. However, the velocity fields derived with this method look noisy,

because the possible velocities in the map are discrete (only channel-map velocities are possible) and, also, the velocity of the peak is sensitive to the r.m.s. noise at positions of low level emission. Also the contours in the inner part of the velocity field are rather box-shaped (see Figure 3 upper right panel). This method probably gives the best estimate of the steep rise of the rotation velocity in the inner parts of a galaxy.

The third method, of fitting Gaussians, is doing remarkably well in finding the velocity of the peak in asymmetric profiles (but not as good as the previous method). It has the advantage over the 2nd method that the velocities in the map are not discrete and less sensitive to the r.m.s. noise, therefore these velocity fields look much smoother. This method also has the possibility to fit more than one Gaussian to the velocity profile. The main disadvantage (although maybe a minor one in practice) is that we assume that the profiles are intrinsically Gaussians.

We have implemented a fourth method which may combine some of the advantages of the first and third method. We regard the velocity profile as a frequency distribution and (try) to determine the mode of this 'distribution'. We can determine the mode by using the skewness and standard deviation of the 'distribution' to correct the IWMV. The

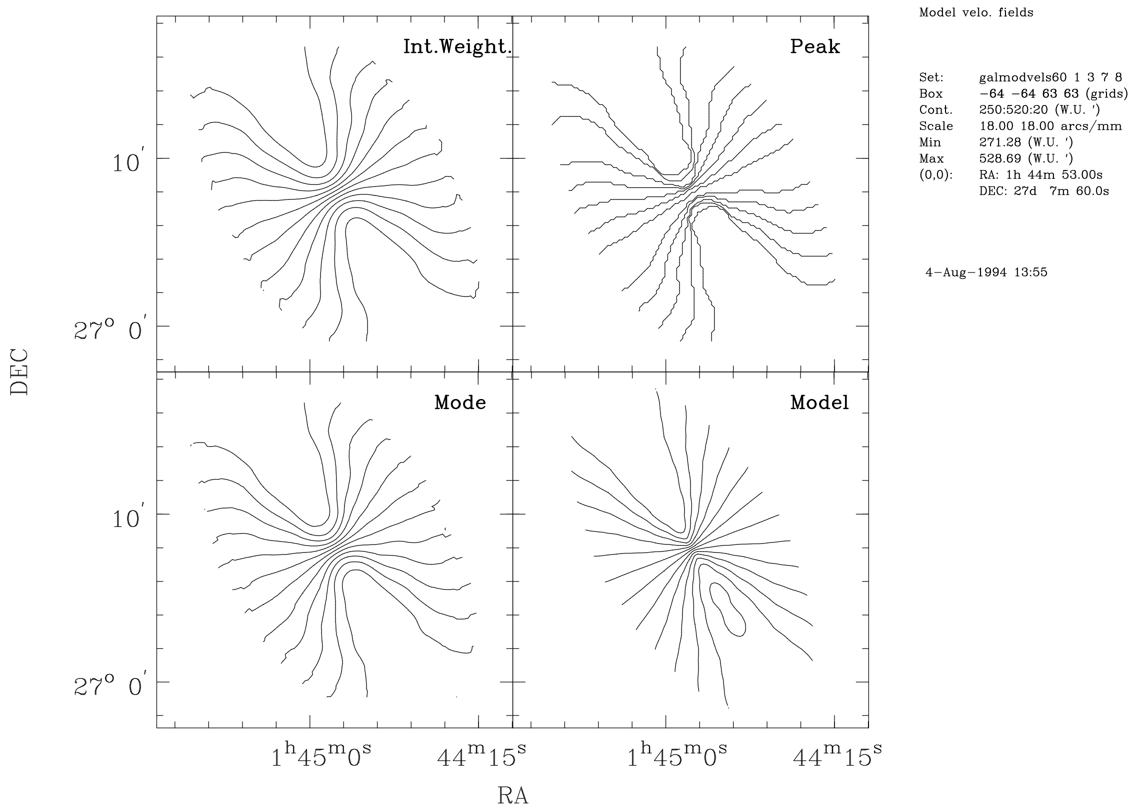


Figure 3: Comparison between different methods for deriving the velocity field. The map is derived from a model-cube smoothed to a resolution of 60″. Velocity field derived from intensity weighted mean (*upper left*); velocity of profile peak (*upper right*); the 'mode' of the profile (*lower left*). The lower right panel shows the unsmoothed model velocity field.

skewness can be derived from the third moment of the velocity profile. This mode-method is a statistical descriptive method with no assumptions about models etc. In practice the velocity fields look very much like the velocity fields derived by the Gaussian-fit-method.
In Figure 3 we show velocity fields derived from a GALMOD model, that has been smoothed to a resolution of 60″. We show the velocity field resulting from the IWMV, peak and mode-methods, together with the unsmoothed model velocity field made using VELFI. The mode-method gives a steeper rise of velocities in the inner part of the galaxy than the IWMV-method and less boxy and noisy contours than the peak-method. In Figure 4 we show the run of velocities along the major axis for the 4 velocity fields of Figure 3. Also added is the velocity curve derived with GAUFIT, i.e. by fitting Gaussians. The last one coincides exactly with the velocity curve derived with the mode-method.
At the moment we can not say which method is best. The IWMV, peak and mode methods are quite straightforward and can be obtained in an automated way. We therefore store the results of these three methods on tape (not much space is involved here) as three different maps in one small cube (which also contains maps of the velocity dispersion, peak value of the velocity profile, dispersion in intensity, number of subsets contributing to the velocity field, S/N of the total H I map and radio continuum).
In deriving the velocity fields, the cleaning masks (cf. §10.2) are used again to select the

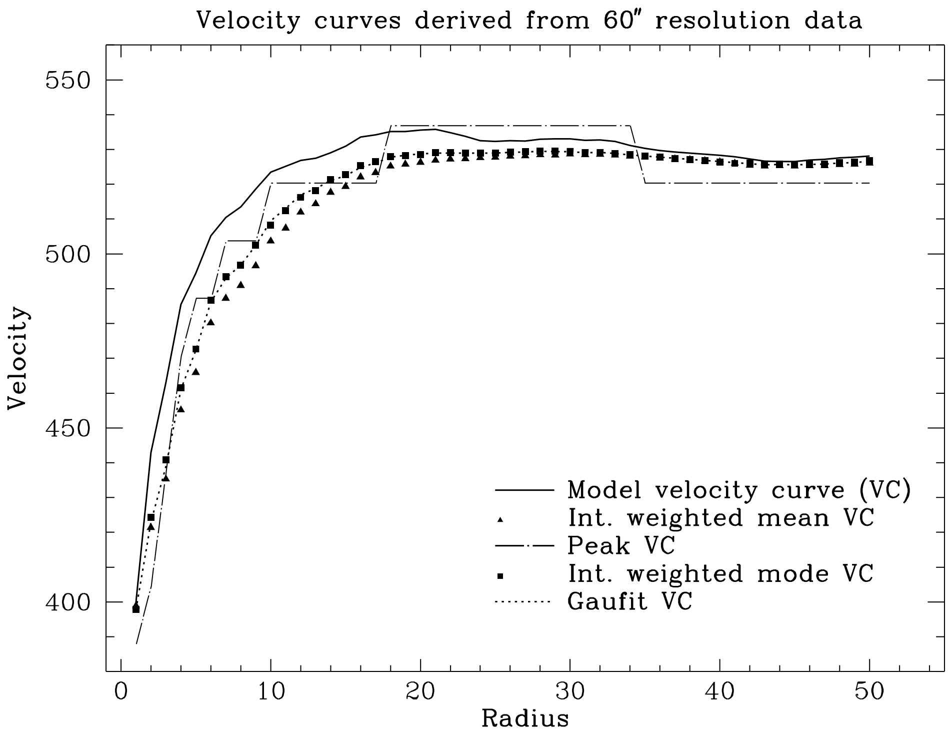


Figure 4: Comparison between velocity curves of different methods for deriving the velocity field. The velocity curves are taken along the major axis and derived from the velocity fields in Figure 3. Also added is the velocity curve derived with fitting Gaussians.

areas from which we calculate the velocity field. We have added no further restrictions to the areas used to derive the velocity field. However we like to point out that the information in the velocity field maps can only be appreciated in combination with the S/N maps derived from the total HI maps. Reliable estimates of the velocity at any positions requires a suitable HI flux at that position.

## 12.3 The XV-diagram along the major axis

We can easily derive, along the major and minor axis, while using the optically measured central position, an XV-diagram of the galaxies from the cleaned cubes for all three resolutions. For all galaxies in the field of view and falling within a reasonable velocity range, the optical central position and position angle (P.A.) are taken from the PGC (which are taken from the RC3). If no optically measured value of the P.A. is available or if this value is clearly in error, the P.A. is derived by fitting an ellips to the total HI distribution. Such P.A. estimates should be regarded with caution. The prime purpose of the XV-diagrams is to locate interesting features in a quick, graphical way.

## 12.4 Items to record in a database

- HI dimensions to characteristic surface density (e.g. 1 $M_\odot/pc^2$) or half total flux value
- surface density of HI at characteristic radius (e.g. $\frac{1}{2}D_{25}$) or average
- optical position and P.A. from PGC (or P.A. from fitting an ellips)
- $V_{max}$ (determined from the velocity field as a first estimate)

# 13 Archiving of GIPSY results

## 13.1 Archiving data on tape: phase 2

The results of the GIPSY reduction phase are: cleaned data cubes at 3 resolutions (full, $30''$, $60''$); a cube with first analysis results: total HI, velocity fields derived from different methods, a velocity dispersion map, a S/N map, a radio continuum map, etc.; and XV-diagrams along major and minor axis at 3 resolutions. All these results, together with a cube containing the masks used as search areas for cleaning the $60''$ cube, will be stored on the DAT-tape WF000n in GDS format. The tape WF000n is the same as the one used in phase 1 of the archiving for this target galaxy. We save the data in GDS format in order not to lose any of the wealth of information that is stored in the headers.

The file containing a table of the global profile flux densities is also stored on the tape WF000n, together with the file with data from the PGC galaxies falling within the observations of the target galaxy.

Contour plots of the total HI, the velocity field(s) and the global profile will be stored on this tape in a later stage.

## 13.2 Items to record in a database

- tape number of global profile file + pictures

# APPENDIX

## A Format of names for different file-types

For the naming of the different files we have chosen the following formats:

| | |
|---|---|
| - <name>.SCN | : calibrated scan-file of galaxy <name> + calibrators |
| - <name> | : full resolution raw cube (name is in comment) |
| - <name>S30 | : cube smoothed to $30''$ |
| - <name>S60 | : cube smoothed to $60''$ |
| - <name>AP | : full resolution antenna patterns |
| - <name>S30AP | : antenna patterns smoothed to $30''$ |
| - <name>S60AP | : antenna patterns smoothed to $60''$ |
| - <name>MASK60 | : search area masks defined with cleaned cube at $60''$ , |
| | : (e.g. used for derivation of total H I map and velocity field map) |
| - <name>CL | : full resolution cleaned cube (name is in comment) |
| - <name>S30CL | : cube smoothed to $30''$ and cleaned |
| - <name>S60CL | : cube smoothed to $60''$ and cleaned |
| - <name>–2DIM | : Cube containing data reduced to 2 dimensions, in the order: |
| | : (0) Total H I map (primary beam corrected) |
| | : (1) Intensity weighted mean velocity field map |
| | : (2) Dispersion in intensity weighted velocity along profile |
| | : (3) Velocity field using the velocity of peak value in profile |
| | : (4) Peak value in profile |
| | : (5) Dispersion in intensities |
| | : (6) Number of subset contributing to 2 dimensional reduction |
| | : (7) Intensity weighted mode velocity field |
| | : (8) Signal-to-noise ratio map of total H I signal |
| | : (9) Radio continuum map of field of view |
| - <name>S30–2DIM | : idem for $30''$ data cube |
| - <name>S60–2DIM | : idem for $60''$ data cube |
| - <name>XV | : XV-diagrams at full resolution along major and minor axes |
| - <name>S30XV | : XV-diagrams at $30''$ resolution along major and minor axes |
| - <name>S60XV | : XV-diagrams at $60''$ resolution along major and minor axes |

The raw data are stored in FITS-format. The file <name>.SCN will be stored on a different tape. The names of the cubes can be found with MTLIST in GIPSY, which shows the FITS-comment field. The processed data (from <name>MASK60 on) are stored in GDS-format. The names of the files with XV-diagrams is extended by there PGC number when more than one galaxy is present in the observations.

# APPENDIX

## B Data contained in REDUCED database

The REDUCED database contains all information on the reduction of the observations. This database is for internal use only.

1 : P(rincipal) G(alaxies) C(atalogue) number
2 : UGC number
3 : Pointing centre of observation
4 : Date of observation
5 : UV-tape number
6 : Tape labels
7 : Sequence number of observation
8 : Grouping of calibrator(s) and observation in the SCAN-file
9 : Central velocity of the band (km/s)
10: Bandwidth (MHz)
11: Number of channels
12: Observing time $T_{obs}$ in hours
13: Integration time of observation in seconds
14: Shortest baseline (36, 54, 72 or 90 meters)
15: Quality mark of maps + defects (e.g. full:good:none)
16: Noise estimate in maps (3 values)
17: Output tape number for calibrated SCAN-files
18: Output tape number for FITS-files (maps, etc.)
19: Output tape number for global profile file + pictures
20: Channel numbers used for continuum subtraction
21: $\Delta$V of HI

# APPENDIX

## C Data contained in WHISP database

The following is a proposal for data to be collected in the WHISP catalogue. This proposal is still flexible and will be decided on later. In the future this database will become a service to the astronomical community.

1 : P(rincipal) G(alaxies) C(atalogue) number
2 : Other designations (NGC, UGC, etc.)
3 : Coordinates (1950 or 2000) (from PGC)
4 : Type (from PGC)
5 : (optical) Inclination (from PGC)
6 : (optical) Dimensions (from PGC)
7 : (optical) Position Angle (from PGC)
8 : B-Magnitude (from PGC)
9 : Total HI mass or Total HI flux + error
10: HI global profile width at 50%, 20% + error
11: Systemic velocity (from HI global profile) + error
12: HI dimensions to characteristic surface density (e.g. 1 $M_\odot/pc^2$) or half total flux value
13: Surface density of HI at characteristic radius (e.g. $\frac{1}{2}D_{25}$) or average
14: $V_{max}$ (determined from the velocity field as a first estimate)
15: Distance used for calculating total HI mass
16: Total HI maps at 3 resolutions
17: Velocity field maps at 3 resolutions
18: Global profile (both table and graph)
19: XV-diagrams at 3 resolutions (along major and minor axis)
20: Optical picture
21: Kinematical P.A. and inclination (to be added after extensive analysis)
22: Peculiarities (subjects to be noted must be well described)

# APPENDIX

## D Short observations

### D.1 Introduction

The short observations are a side-product of WHISP: observations, typically a few hours, of galaxies for which no H I data appear to be available, and which may be suitable candidates for full-synthesis observations.[1] These observations just fill the holes in the WSRT observing schedules. The reduction of the data is somewhat different from that of the standard 12 hours observations.

### D.2 NEWSTAR phase

The reduction of the *uv*-data is essentially the same as for the 12 hours observations. We Fourier transform the *uv*-data, with a Gaussian taper of 390 meter, to channel maps with a $60''$ angular resolution in the direction of the axis of the highest resolution.

### D.3 GIPSY phase

Because the channel maps contain essentially one-dimensional information, the reduction in the map plane is different from the standard procedure described earlier in this document. After applying a Hanning taper over the channel maps, we made a sequence of position - velocity (PV) maps along the axis with the highest resolution and with different integration lengths perpendicular to this axis. We selected the PV-map with the highest signal-to-noise ratio as the most appropriate map to derive the global H I properties of a galaxy. The continuum emission is removed by fitting a line in radial velocity in the H I -free channel maps.
The selected PV-map has been CLEANed to one third of the r.m.s. noise level with the corresponding antenna pattern (determined in the same way as the PV-map).

### D.4 Global profile

The only direct result we derived from these observations are the global profiles, i.e. the flux density as function of radial velocity. The flux density at each radial velocity is determined as the sum of the flux in the CLEAN components and the flux in the residuals.

[1] We have written an article on the short observations ("Short H I synthesis observations of spiral galaxies with the WSRT", appeared in A&A 1995).