THE EXTRAGALACTIC DISTANCE DATABASE: ALL DIGITAL H I PROFILE CATALOG

HÉLÈNE M. COURTOIS1,2, R. BRENT TULLY2, J. RICHARD FISHER3, NICOLAS BONHOMME1, MAXIMILIAN ZAVODNY2, AND AUSTIN BARNES2

1 Université Lyon 1, CNRS/IN2P3/INSU, Institut de Physique Nucléaire, Lyon, France
2 Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu HI 96822, USA
3 National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA

Received 2009 February 17; accepted 2009 September 25; published 2009 November 9

ABSTRACT

An important component of the Extragalactic Distance Database is a group of catalogs related to the measurement of H I line profile parameters. One of these is the All Digital H I catalog which contains an amalgam of information from new data and old. The new data result from observations with Arecibo and Parkes Telescopes and with the Green Bank Telescope, including continuing input since the award of the NRAO Cosmic Flows Large Program. The old data have been collected from archives, wherever available, particularly the Cornell University Digital H I Archive, the Nançay Telescope extragalactic H I archive, and the Australia Telescope H I archive. The catalog currently contains information on ∼15,000 profiles relating to ∼13,000 galaxies. The channel–flux per channel files, from whatever source, is carried through a common pipeline. The derived parameter of greatest interest is Wm50, the profile width at 50% of the mean flux. After appropriate adjustment, the parameter Wm50 is derived, the line width that statistically approximates the peak-to-peak maximum rotation velocity before correction for inclination, 2V sin i.

Key words: astronomical data bases: miscellaneous – catalogs – galaxies: distances and redshifts – radio lines: galaxies

Online-only material: color figures, machine-readable and VO tables

1. INTRODUCTION

The goal of the overall program facilitated by the Extragalactic Distance Database (EDD) is to obtain the densest and deepest possible coverage of galaxy distances and, hence, of line-of-sight peculiar velocities. We want to improve the local determination of the Hubble constant and measure departures toward the distribution of matter. We are giving consideration to the present discussion is to integrate the considerable amount of potential and the baryonic matter that shines. Considerable effort has been made to try to understand this link

Motions within galaxies are a response to the gravitational potential. If the H I gas is in equilibrium in a disk, rotating in circular orbits, then there is a simple relationship between the observed motions and the distribution of mass. The small dispersion in the relation between rotation rate and luminosity implies a strong correlation between the dark matter that dominates the potential and the baryonic matter that shines. Considerable effort has been made to try to understand this link

Motions within galaxies are a response to the gravitational potential. If the H I gas is in equilibrium in a disk, rotating in circular orbits, then there is a simple relationship between the observed motions and the distribution of mass. The small dispersion in the relation between rotation rate and luminosity implies a strong correlation between the dark matter that dominates the potential and the baryonic matter that shines. Considerable effort has been made to try to understand this link

Bell et al. 2003; Dutton et al. 2007. A focus of recent attention has been on the evaluation of how the correlation changes with look-back time (Conselice et al. 2005; Weiner et al. 2006). Our

RECEIVED 2009 FEBRUARY 17; ACCEPTED 2009 SEPTEMBER 25; PUBLISHED 2009 NOVEMBER 9

1.1. Historical Background

Motions within galaxies are a response to the gravitational potential. If the H I gas is in equilibrium in a disk, rotating in circular orbits, then there is a simple relationship between the observed motions and the distribution of mass. The small dispersion in the relation between rotation rate and luminosity implies a strong correlation between the dark matter that dominates the potential and the baryonic matter that shines. Considerable effort has been made to try to understand this link

Bell et al. 2003; Dutton et al. 2007. A focus of recent attention has been on the evaluation of how the correlation changes with look-back time (Conselice et al. 2005; Weiner et al. 2006). Our

Available at http://edd.ifa.hawaii.edu.
Neutral hydrogen is easily detected in nearby spiral and irregular galaxies with modern radio telescopes. The product of an observation with a single dish facility is a line profile that can be grossly characterized by three parameters: an integrated flux, a systemic Doppler shift from the rest wavelength, and a line width due to internal motions. The distances over which galaxies can be detected depend on the sensitivity of telescopes and the intrinsic gas content of galaxies.

Regarding telescopes, the bigger the better. Consider the situation of unresolved sources that is generally close to being met. Receivers and efficiencies being equal, the advantage of a big telescope in exposure time required to reach a given signal to noise (S/N) goes as the fourth power of the aperture. The S/N achieved in a unit time, $t_0$, depends on the square of the aperture, $D$, of the telescope. To reach a specific S/N requires a time $t$:

$$S/N \propto (t/t_0)^2 \propto D^4.$$ 

The Arecibo Telescope is presently by far the most sensitive single dish instrument for H I line studies. Unfortunately, it accesses only 30% of the sky. An additional advantage of a large telescope is a relatively small primary beam, hence reduced source confusion. A small beam only becomes a disadvantage when it is smaller than the dimensions of the source, a situation that can result in lost flux and a biased line width.

Regarding the properties of galaxies, it is instructive to consider the H I mass function (Zwaan et al. 2005). The cutoff at high mass is abrupt. Systems with higher H I mass than $3 \times 10^{10} M_\odot$ are rare. It can be supposed that larger gas reservoirs than this limit quickly get converted into stars. By contrast, many so-called dwarf galaxies have abundant H I, reflecting low time-averaged star formation rates. The result is a pile-up of 90% of H I masses in the two decade range $1.5 \times 10^8$–$1.5 \times 10^9 M_\odot$. Figure 1 demonstrates that most of the neutral gas in the $z = 0$ universe is locked up in galaxies (which are mostly spirals) with $\log M_{\text{HI}}/M_\odot = 9.2 \pm 1.0$. In the second panel, a histogram is presented of H I mass divided by profile line width (measured at 20% of peak intensity in units of 100 km s$^{-1}$). The sample is drawn from the Ursa Major Cluster at 17 Mpc and is essentially complete to an H I mass of $10^7 M_\odot$ (Tully et al. 1996). Eighty-five percent of the sources are contained within a single decade $2 \times 10^8$–$2 \times 10^9 M_\odot$ per 100 km s$^{-1}$. Total intrinsic fluxes are constrained to a two decade window (the left panel of the figure) and, since sources with greater fluxes tend to have larger line widths, intrinsic fluxes per spectral channel are constrained to a single decade (right panel). The curious consequence is that galaxies typed Sb–Sc–Sd can be detected in H I with comparable likelihood. A program to observe these kinds of galaxies can be expected to have a high level of completion within a volume dictated by the telescope, time available, and motivation. Current capabilities can be evaluated by giving consideration to Figure 2. Note the sharp upper cutoff in $M_{\text{HI}}$ which is rather flat with distance (increased volume).

It is one thing to detect H I in a galaxy and quite another to obtain a signal useful for the determination of distances. Non-pathological line profiles have characteristic features that are helpful. In the case of massive galaxies, most of the flux is at the high- and low-frequency extremes, originating from gas on the flat part of rotation curves. The consequence is profiles

![Figure 1](image1.png)

**Figure 1.** Left: fraction of H I mass in logarithmic mass intervals derived from the H I mass function of Zwaan et al. (2005). Right: H I mass per 100 km s$^{-1}$ of line width for a sample drawn from the Ursa Major Cluster complete to an H I mass limit of $10^7 M_\odot$.

![Figure 2](image2.png)

**Figure 2.** H I mass assuming distance = $cz$ km s$^{-1}$/75 km s$^{-1}$ Mpc$^{-1}$ vs. systemic velocity for 15,000 galaxies in the Lyon Extragalactic Database (LEDA). The solid curve traces the locus of an integrated flux of 1 Jy km s$^{-1}$, a practical limit with the Arecibo Telescope. (A color version of this figure is available in the online journal.)
with abrupt edges and flux where it is most useful to define these edges. For galaxies of sufficiently low mass, those with more slowly rising rotation curves, the line profiles lose the two-horn shape and instead can be approximately Gaussian. These systems tend to have less total H\textsc{i} but it is piled up over a small range of wavelengths, feeding the conspiracy of a common detectability level.

2. EXISTING H\textsc{i} INFORMATION

There has been a transformation in the last decade in the way data are taken at radio telescopes. In the early days, the observer made measurements from analog displays and the published output were profiles on a journal page. The catalog in EDD called Pre-Digital H\textsc{i} gives a compilation of information laboriously extracted from literature profiles from many dozens of sources (cf. Huchtmeier & Richter 1989). The great concern and attention in the compilation of this historical catalog was with profile line widths (the fluxes in the catalog are of mixed quality and may often suffer from beam dilution effects). Attention should be given to the parameter $W_20$, the line width at 20\% of peak intensity. Often, profiles for a given galaxy are available from multiple sources. There is a velocity cutoff of 3000 km s\(^{-1}\) for the entries in this catalog. A significant contribution came from observations by two of the authors (Fisher & Tully 1981). It is to be emphasized that the line widths given in the Pre-Digital H\textsc{i} catalog do not come from the literature source; they were all derived from published graph profiles by Fisher, Tully, or our colleague Cyrus Hall.

Some use will be made of this pre-digital information in what follows. This carefully accumulated information will be used as a basis of comparison with digital data processed by a machine. A value from the Pre-Digital H\textsc{i} catalog can even be preferred as a primary source in some cases of very large, nearby galaxies that could only be satisfactorily observed with the large beams of small telescopes that are no longer in service. Of course, care must be taken to avoid systematics in the translation of profile information between different systems.

The turn of events that marginalizes the Pre-Digital H\textsc{i} material is the ubiquitous availability of digital spectral information from radio telescope archives. There is no need to go to the telescope to access a rich load of data. Of course, the authors of each source will have analyzed their material in their own way. At issue for us, though, is the need to analyze all profile information from whatever source in a common way. One less than satisfactory approach would be to ingest parameters such as line widths, fluxes, and systemic velocities from literature sources and attempt to reconcile related measurements. However, more coherent results can be expected if a standard analysis procedure is applied to the directly observed spectra.

There are very large and very good compilations of H\textsc{i} material already available. The most important are assembled in separate catalogs within EDD. Foremost, there is the extensive database compiled by the Cornell group labeled Springob/Cornell H\textsc{i} (Springob et al. 2005) which presents material obtained with the Arecibo Telescope, the old 140 foot and 300 foot Green Bank Telescopes, and limited material from the Nan\c{c}ay and Effelsberg Telescopes. Additionally, the EDD contains tables and links relating to observations by other observers at individual telescopes. The results of a major observing program with the Nan\c{c}ay Telescope (Theureau et al. 2006) are accessed through the catalog H\textsc{i} Nan\c{c}ay. Key results from the H\textsc{i} Parkes Telescope All-Southern-Sky multibeam project (Koribalski et al. 2004) are reported in the catalog HIPASS 1000. Although the material available from archives is extensive, it is not sufficient. There are galaxies that are important to our program that have not been satisfactorily observed. Details regarding our samples will be described elsewhere but here is a flavor. One sample contains all suitably inclined and unobscured spirals within 3000 km s\(^{-1}\) brighter than $\sim 0.3 \ L^*$ (see the catalog $V3k \ M_K < -21$ in EDD). Another is of spirals out to 6000 km s\(^{-1}\) selected from the Infrared Astronomical Satellite Point Source–Redshift Catalog (IRAS PSC-z; see the catalog Saunders PSC\text{-}z in EDD). Yet another sample consists of galaxies extending out to 10,000 km s\(^{-1}\) with observed supernovae of Type Ia, included for the purpose of improving the zero-point calibration of the supernova scale.

For the present discussion, the point to be made is that we needed to supplement the archival material with our own observations, which required that we develop procedures for the analysis of our observations. We might have simply adopted a pre-existing procedure, say, incorporating one of the line width measures described by Springob et al. (2005). However, we decided that we would learn more if we developed our own algorithm, albeit one inspired by an approach considered by Springob et al. The exercise would provide independently derived results that could, in many cases, be compared with other sources to evaluate uncertainties.

The discussion continues in the following section with descriptions of our observations with the Arecibo, Green Bank, and Parkes Telescopes. In the ensuing four sections, there are descriptions of how we measure profile parameters, evaluations of product quality, and a brief discussion of adjustments to give a parameter of dynamical interest. In the section before the summary, there is a description of the catalog in EDD called All Digital H\textsc{i} which contains both the results of our own observations and the reprocessed results of data from the archives, all treated with the same procedures.

3. OBSERVATIONS WITH THE ARECIBO, GREEN BANK, AND PARKES TELESCOPES

Our program began once the Arecibo Telescope was brought back in service after installation of the Gregorian feed and ground screen. Since the start of the multibeam sky survey Arecibo Legacy Fast ALFA (ALFALFA), we have discontinued our own observations with the Arecibo Telescope, with the expectation that many of our sources in the survey range $-36^\circ < \delta < 0^\circ$ will be observed serendipitously with sufficient accuracy. Subsequently, we have been observing with the 100 m Green Bank Telescope (GBT) at declinations above $\delta = -45^\circ$ but excluding the Arecibo range. As of the third semester of 2008, this program has been awarded the status of a Large Program, now christened with the name Cosmic Flows. Results from this program are appearing in the All Digital H\textsc{i} catalog as they become available. Access to the remaining sky, at $\delta < -45^\circ$, requires use of Parkes Telescope in Australia. Observations with this facility began early in 2009.

3.1. Arecibo

The single-beam Arecibo observations of 330 galaxies were undertaken in two sessions: 1999 October and 2001 April, mostly between sunset and sunrise to avoid spectral baseline distortion from solar continuum emission. The 1999 session used the “L-Narrow” receiver for objects with known redshifts

---

4 http://www.vla.nrao.edu/astro/prop/largeprop/
and the “L-Wide” receiver for unknown redshifts that required a search over a wide velocity range. The 2001 session used only the “L-Wide” receiver with improved system noise temperature.

Eight correlator sections, each with 2048 spectral channels, were available. The unknown redshift search used four overlapping, 25 MHz bandwidth spectra for a velocity span from $-400$ to $+1800$ km s$^{-1}$, heliocentric, in each of two linear polarizations. The known redshift spectra were taken with four correlator sections, two polarizations with 12.5 MHz bandwidth, and two with 6.25 MHz bandwidth, all centered on the galaxy line profile. All spectral data were Hanning smoothed to produce correlator sections, two polarizations with 12.5 MHz bandwidth, respectively. Narrowband interference was edited manually and data values replaced with a linear interpolation. For most spectra a second-order baseline was least-squares fit to the data on either side of the line profile and subtracted from all spectral data values. In no case was the baseline curve higher than third order.

The basic Arecibo observation was 7 minutes on the object and 7 minutes on a blank-sky position on the same hour angle track. When time was available, weak line profiles were observed on more than one day, and the spectra were averaged. The line profile flux density scale was established with correlator observations of known continuum sources over the full scan range of the Arecibo Telescope to determine gain as a function of zenith distance.

### 3.2. Green Bank

The single-beam Robert C. Byrd GBT observations were carried out in the course of two programs, one program from 2001 to 2002 and one large project beginning from 2006 February and continuing through 2009 May. The observations were made day and night. Roughly, 1000 galaxies have been observed between these programs.

The earlier observations were conducted during commissioning of the GBT in the fall and winter of 2001/2002 as a background 21 cm observing program during times when the telescope was not occupied with tests or calibration. Simple on-off spectral line measurements were made to acquire global H$\text{I}$ profiles of galaxies at redshifts out to about 10,000 km s$^{-1}$. Integration times were between 10 and 60 minutes, and typical bandwidths were 5 or 10 MHz depending on the expected signal strength and line profile width. All observations used the fast Fourier transform (FFT) spectrometer which has 1024 channels for each of the two linear receiver polarizations. The system temperature was slightly under 20 K at high elevations and the 100 m aperture efficiency was roughly 70%. Intensity calibration of the H$\text{I}$ survey of galaxies during commissioning of the GBT is tied to the NRAO VLA Sky Survey (NVSS) flux density scale (Condon et al. 1998). About five dozen continuum sources with flux densities between 2.2 and 6.0 Jy were selected to avoid significant multi-source confusion with the 9 arcmin GBT beam. The continuum calibrators were observed with the same spectrometer and receiver configuration as was used to measure the H$\text{I}$ line profiles in the survey, with the exception that the spectrometer bandwidth was always 40 MHz centered on 1403 MHz to span most of the range of frequencies observed with smaller bandwidths.

The calibrator observing sequence was 2 minutes off, 2 minutes on, and 2 minutes off source in spectral line mode. The first off position was 38 arcmin toward lower right ascension than the source position, and the second off was the same distance toward greater right ascension. The hour angle track was, therefore, not exactly the same for the three observed positions for high declination objects, but this did not appear to degrade the spectral baselines significantly. The first task in the calibrator data reduction was to visually inspect the difference spectrum between the two off positions. The difference spectrum offset was typically less than about 60 mJy, as is expected from confusion noise with the GBT beam size, but a few offsets were as high as 300 mJy. These large offsets were possibly due to a moderately strong source in one of the off positions or, more likely, a bit of radiation from the Sun during the day. Since the observed source flux density was about 3 Jy, even the largest off-source baseline offset, after the two off spectra were averaged together, caused about 5% error in the measure source intensity. More typically, this source of error amounted to less than 1%. The statistics of the calibrator source measurements were not significantly improved by throwing out observations with larger off-position differences so all data were retained.

For the observations since 2006 use is made of the single beam L-band (1–2 GHz) receiver and the spectral line spectrometer as the backend detector. Data are taken with a 12.5 MHz bandwidth and 9-level sampling. Total power observations are made with a full calibration noise source switching cycle of 1 s. The spectrometer records data every 30 s. The spectral line is Doppler tracked in the barycentric velocity frame. Data are taken using linear polarizations. Redshifts from the Lyon Extragalactic Database (LEDA) or NASA Extragalactic Database (NED) were used to center the window.

The basic GBT observation procedure was to take a pair of on–off observations with 300 s on and 300 s off the target. Galaxies within 3000 km s$^{-1}$ usually required 1–3 scan pairs, while galaxies from the PSCz sample reaching up to 8000 km s$^{-1}$ required 10–15 scan pairs. A preliminary guess on how much observing time a target would need was derived from the 21 cm magnitude given in LEDA. In order to optimize the observing time, a target was observed with time split over several days. Data were reduced daily and evaluated in order to add observing time as needed until the signal reached the desired very high quality for a luminosity–line width distance measurement.

Individual observations are calibrated in Jy using the standard calibration procedure available at GBT. GBTIDL provides basic routines that can be used to calibrate and average spectra when the data are taken in standard, predefined observing modes. The calibration routines typically give a flux scale accurate to 10%. Well-known galaxies that can be used as H$\text{I}$ calibrators were also observed several times per month in order to be able to retrieve a more accurate flux calibration if needed in the future.

The calibrated data are then averaged, baseline subtracted using a polynomial fit usually of order 3, and smoothed with a simple Hanning filter. The final spectrum is stored with 1.6 km s$^{-1}$ resolution. It was usually binned at least once to 3.2 km s$^{-1}$ resolution for the H$\text{I}$ line width measurement.

### 3.3. Parkes

Observations with the Parkes Telescope make use of the seven-beam system in MX mode, with the target in the central beam and the six outer beams used to monitor the sky. Integration times are estimated based on fluxes obtained with the H$\text{I}$ Parkes All-Sky Survey (HIPASS) integrations of roughly 8 minutes per pointing, velocity resolution of 18 km s$^{-1}$, and rms sensitivity of 13 mJy per channel. For line widths adequate for our purposes
we attempt to obtain spectra with a peak S/N of at least 10 with a spectral resolution of 2 km s\(^{-1}\). Galaxies at 3000–4000 km s\(^{-1}\), which are the most common of our targets, typically require 60 minutes on the source. Profiles are evaluated at 30 minutes exposure and, if inadequate, the source is reobserved and profiles are summed. Daytime observations are avoided to minimize degradation by the Sun. The Parkes multibeam data are reduced using the graphical user interfaces Livedata, Gridzilla, and MIRIAD. The data are calibrated in Janskys using the procedure described by Barnes et al. (2001).

At press, a first observing run during 2009 January–February has been completed culminating in observations of 58 galaxies. Results from this run and the much larger archival sample of the 1000 brightest sources from the HIPASS program (Koribalski et al. 2004) are analyzed and included in EDD.

### 4. PROFILE LINE WIDTHS

The line width measure given in the Pre-Digital H\(_\text{I}\) catalog of EDD and used by us since the early paper by Tully & Fisher (1977) is \(W_{20}\), the line width at 20% of peak intensity. This is an appropriate moment to evaluate whether that parameter choice is optimal since we now do the analysis on digital data with a rigorous algorithm and apply the same procedure to all available material. The thinking behind the original choice was that a line width at a very low level of intensity with respect to the maximum is desirable to minimize dependencies on vagaries in the distribution of flux within the profile. The opposing constraint is the need to be above the noise level. With profiles deemed adequate, it was empirically determined that the 20% of peak intensity level is sufficiently out of the noise. Adequate profiles are characterized by peak signals at least seven times greater than the noise.

As we consider alternatives, we look to the study by Springob et al. (2005, SHGK). Their data are made available at the Cornell University Digital H\(_\text{I}\) Archive Web site.\(^6\) Those authors have given attention to a large body of high-quality data from their own observations and from the archives. They derived five separate line width parameters with automated algorithms. These distinct line width measures can be compared with each other and, for most of the galaxies within 3000 km s\(^{-1}\), with the \(W_{20}\) values from the Pre-Digital H\(_\text{I}\) catalog in EDD. Comparisons are shown in Figure 3 for 1110 galaxies considered to have good \(W_{20}\) measures. In 3% of cases, the line widths are discordant by more than 50 km s\(^{-1}\). These large differences are due to cataloging errors or gross errors due to noise. Automated procedures are vulnerable to occasional gross errors—as manifested by big differences between the five SHGK parameters in a small fraction of cases. In the following discussion, we clip instances with deviations greater than 50 km s\(^{-1}\) from the mean of \(W_{20} - W_X\), where \(X\) is one of the five line width parameters given by SHGK.

It turns out that we find a clear preference for one of the SHGK line width parameters. It is not the parameter advocated as optimal by SHGK. Those authors prefer the parameter \(W_{F50}\), the width at 50% of the peak minus rms flux with left and right edges evaluated independently with polynomial fits to the rising portions of the profile. In Table 1, we collect comparisons between our old \(W_{20}\) values and the five SHGK values (plus their parameter \(W_c\) which is \(W_{F50}\) with redshift, instrumental, and smoothing corrections). Of course, zero-point offsets are expected. The figure of merit is the rms dispersion. The SHGK parameter \(W_{M50}\) gives a significantly better correlation with our \(W_{20}\). Conveniently, it also gives rather close agreement in the zero point. Figure 4 gives a detailed comparison between \(W_{20}\) and the two SHGK line width parameters of greatest interest. In the case of the parameter \(W_{F50}\), the mean difference with respect to \(W_{20}\) is displaced from the peak of the distribution (the distribution is skewed) and (after clipping values more deviant than ±50 km s\(^{-1}\) from the mean) the dispersion is a rather substantial 17 km s\(^{-1}\). In the case of the parameter \(W_{M50}\), the histogram is symmetric and (after clipping values more deviant than ±50 km s\(^{-1}\) from the new mean) the dispersion is a reasonable 10 km s\(^{-1}\).

---

\(^6\) http://arecibo.tc.cornell.edu/hiarchive

---

**4. PROFILE LINE WIDTHS**

The line width measure given in the Pre-Digital H\(_\text{I}\) catalog of EDD and used by us since the early paper by Tully & Fisher (1977) is \(W_{20}\), the line width at 20% of peak intensity. This is an appropriate moment to evaluate whether that parameter choice is optimal since we now do the analysis on digital data with a rigorous algorithm and apply the same procedure to all available material. The thinking behind the original choice was that a line width at a very low level of intensity with respect to the maximum is desirable to minimize dependencies on vagaries in the distribution of flux within the profile. The opposing constraint is the need to be above the noise level. With profiles deemed adequate, it was empirically determined that the 20% of peak intensity level is sufficiently out of the noise. Adequate profiles are characterized by peak signals at least seven times greater than the noise.

As we consider alternatives, we look to the study by Springob et al. (2005, SHGK). Their data are made available at the Cornell University Digital H\(_\text{I}\) Archive Web site.\(^6\) Those authors have given attention to a large body of high-quality data from their own observations and from the archives. They derived five separate line width parameters with automated algorithms. These distinct line width measures can be compared with each other and, for most of the galaxies within 3000 km s\(^{-1}\), with the \(W_{20}\) values from the Pre-Digital H\(_\text{I}\) catalog in EDD. Comparisons are shown in Figure 3 for 1110 galaxies considered to have good \(W_{20}\) measures. In 3% of cases, the line widths are discordant by more than 50 km s\(^{-1}\). These large differences are due to cataloging errors or gross errors due to noise. Automated procedures are vulnerable to occasional gross errors—as manifested by big differences between the five SHGK parameters in a small fraction of cases. In the following discussion, we clip instances with deviations greater than 50 km s\(^{-1}\) from the mean of \(W_{20} - W_X\), where \(X\) is one of the five line width parameters given by SHGK.

It turns out that we find a clear preference for one of the SHGK line width parameters. It is not the parameter advocated as optimal by SHGK. Those authors prefer the parameter \(W_{F50}\), the width at 50% of the peak minus rms flux with left and right edges evaluated independently with polynomial fits to the rising portions of the profile. In Table 1, we collect comparisons between our old \(W_{20}\) values and the five SHGK values (plus their parameter \(W_c\) which is \(W_{F50}\) with redshift, instrumental, and smoothing corrections). Of course, zero-point offsets are expected. The figure of merit is the rms dispersion. The SHGK parameter \(W_{M50}\) gives a significantly better correlation with our \(W_{20}\). Conveniently, it also gives rather close agreement in the zero point. Figure 4 gives a detailed comparison between \(W_{20}\) and the two SHGK line width parameters of greatest interest. In the case of the parameter \(W_{F50}\), the mean difference with respect to \(W_{20}\) is displaced from the peak of the distribution (the distribution is skewed) and (after clipping values more deviant than ±50 km s\(^{-1}\) from the mean) the dispersion is a rather substantial 17 km s\(^{-1}\). In the case of the parameter \(W_{M50}\), the histogram is symmetric and (after clipping values more deviant than ±50 km s\(^{-1}\) from the new mean) the dispersion is a reasonable 10 km s\(^{-1}\).
The parameter $W_{M50}$ is the line width at 50% of the mean flux level within the $\text{H}$\textsubscript{i} signal. This construct has nice features. Using the mean flux level rather than the peaks serves to disengage the line width measure from details of the gas distribution. It gives a more natural transform over to single peak cases. The measurement is at a low level compared with the peak, statistically only slightly above the level that gives $W_{20}$ (hence the similar zero point). The main operational challenge is to define the window containing the signal. SHGK reserve most of their discussion for their line width parameter $W_{F50}$ and do not give details of the derivation of $W_{M50}$. We have developed an algorithm which might differ in minor details. An empirical test of that possibility will come from inter comparisons of our separate results with the same data, reported in the following section. To distinguish our parameter from that developed by SHGK, we refer to our measurement as $W_{m50}$; that is, with a lower case “m.”

The most sensitive detail with the derivation of $W_{m50}$ is the specification of the wavelength window for the summation of flux. The total flux detected within the single beam pointing is relatively well defined. However, the mean flux per channel can be significantly less well defined if channels in the wings of the profile are included or not because of noise.

Our specific recipe is to first determine the integrated flux within a window that is tight, yet sufficiently wide that the profile has reached the baseline level (call this the 100% window).

The histogram of $W_{20} - W_{X50}$ values is skewed with a tail to positive differences and a relatively large dispersion. The histogram of $W_{20} - W_{M50}$ values is symmetric and has lower dispersion.

(A color version of this figure is available in the online journal.)

Figure 4. Histograms of the differences in the line width value $W_{20}$ given in the “Pre-Digital H\textsubscript{i}” catalog of EDD and the line width values $W_{X50}$ given by SHGK where $X = F$ (dotted histogram in red) is the parameter preferred by those authors, a measure at 50% of peak flux, and $X = M$ (solid histogram in black) is a measure at 50% of the mean flux. The histogram of $W_{20} - W_{F50}$ values is skewed with a tail to positive differences and a relatively large dispersion. The histogram of $W_{20} - W_{M50}$ values is symmetric and has lower dispersion.

(A color version of this figure is available in the online journal.)

Figure 5. PGC 3671 (NGC 337A), PGC 6061 (UGC 1167), PGC 29096 (ESO 316-018), and PGC 26512 (NGC 2841), profiles with increasing line widths, observed with the NRAO 300', Arecibo, Byrd Green Bank, and NRAO 140 foot Telescopes, respectively.

(A color version of this figure is available in the online journal.)
Then the wavelengths are determined that exclude 5% of the integrated flux on each of the two wings (these enclose the 90% window). The mean flux per channel is then taken to be the sum of the flux within this 90% window divided by the number of spectral channels. Both the numerator and denominator in this calculation require interpolations since the wavelengths defining the window are not restricted by the discreteness of the spectral channels.

The line width is defined at the level of 50% of the mean flux per channel determined in the manner that has just been described. The intersection points at the two edges are defined by interpolations between observed flux–velocity points on the rising parts of the profile. SHGK use a more elaborate fitting scheme to define edges but it is not clear that the effort results in greater precision. In clean cases, the spectra rise abruptly and in ambiguous cases the uncertainty is not necessarily ameliorated by a particular fitting recipe.

The assignment of errors is a particularly challenging problem. Profiles can be messy in so many ways that we despaired of finding an algorithm that gives sensible results in all cases. Our overwhelming interest is the use of profile widths as a parameter in the measurement of distances. We consider that there is a threshold of acceptability; a profile may be of sufficient quality to be used in the determination of a distance, or it may not be. We link our error estimate to this threshold. Specifically, an adequate profile is assigned an error of less than or equal to 20 km s\(^{-1}\). Inadequate profiles are identified by errors greater than 20 km s\(^{-1}\). It has been found (Tully & Pierce 2000) that with line width errors constrained to this limit in quality the measure of line widths is not a dominant source of errors in the determination of distances.

The detailed error assignment is arrived at in two steps. The first step is automatically generated based on the signal (the mean flux per channel within the 90% window), \(S\), to rms noise, \(N\). Errors were evaluated from a training set. Errors of 8 km s\(^{-1}\) are assigned in the best cases, whenever the mean flux per channel is greater than 17 times noise. Errors degrade to 20 km s\(^{-1}\) by a mean flux per channel to noise of 2, and continue to increase as the signal degrades further. Specifically, the error \(e_W\) is assigned based on \(S/N\):

\[
\begin{align*}
e_W &= 8 \text{ km s}^{-1} & \text{if } S/N > 17 \\
e_W &= 21.6 - 0.8 S/N \text{ km s}^{-1} & \text{if } 2 < S/N < 17 \\
e_W &= 70 - 25 S/N \text{ km s}^{-1} & \text{if } S/N < 2.
\end{align*}
\]
Errors are not allowed to be less than the spectral resolution after smoothing. Our error estimates are conservative; roughly a factor 2 larger than values found in the literature associated with the same observations. Cross-correlations between data sets to be discussed later will demonstrate that our error estimates must on average be \( \sim 50\% \) greater than a 1\( \sigma \) value.

The second step involves a manual inspection, given to every profile. The fit and error estimate proposed by the computer algorithm is displayed on a monitor. The operator can accept or modify. It may be necessary to excise interference, or reposition the edges of the 100\% window (the most common occurrence), or smooth. The computer redisplays. The final decision by the operator is whether the error is appropriate, which is a binary decision to the question: is the profile adequate or inadequate for the purpose of measuring a distance? The error only needs to be changed one way or the other across the 20 km s\(^{-1}\) acceptance threshold if the carbon-based decision contradicts the silicon-based assignment.

4.1. Good, Bad, and Ugly

Examples of good profiles are seen in Figure 5. They are characterized by high flux “horns” at the extrema of the profiles, which arise from emission from the flat portion of rotation curves near maximum velocity. The profiles rise rapidly, leaving little uncertainty in the measurement of the widths. There is apparent splitting into two peaks even with the narrowest profile chosen as an example in this figure although typically the distinctness of two peaks is lost in narrow profiles. In this figure and others to follow, information regarding the source of the profile is encoded in the header above the profile. One is given the Principal Galaxies Catalog (PGC) number, a common name, the archival or new observation source in a code given in Table 3 and the telescope as identified in Table 2.

The catalog All Digital H\textsubscript{i} now contains information on approximately 13,000 galaxies. Of reliably measured profiles for apparently single targets, that for PGC 71392 (UGC 12591)
Table 2
Telescopes Contributing to the Database

| Telescope | Aperture | Beam   | Acronym              |
|-----------|----------|--------|----------------------|
| Arecibo   | 305 m    | 3′     | AOG-AOlf-ALFA        |
| Nançay    | 200 × 40 m | 4′ × 22′ | nan                  |
| GBT       | 100 m    | 9′     | GBT                  |
| Effelsberg| 100 m    | 9′     | Effs-Eff             |
| GB300     | 91 m     | 10′    | GB300                |
| Parkes    | 64 m     | 14′    | PAKS                 |
| GB140     | 43 m     | 21′    | GB140                |

Table 3
All Digital H\(_i\) Catalog Sources

| Code    | Literature Source          |
|---------|----------------------------|
| ksk2004 | Koribalski et al. 2004     |
| shg2005 | Springob et al. 2005       |
| hkk2005 | Huchtmeier et al. 2005     |
| tmc2006 | Theureau et al. 2006       |
| ghk2007 | Giovanelli et al. 2007     |
| sgk2008 | Saintonge et al. 2008      |
| kgh2008 | Kent et al. 2008           |
| ctf2009 | This paper                 |

is by far the widest, with \(W_{m50} = 989\) km s\(^{-1}\). The line profile and an image of the galaxy (typed SO/Sa) are shown in Figure 6. Giovanelli et al. (1986) have drawn attention to this unusual galaxy. That reference notes that there is probably absorption from a central continuum source affecting the spectrum.

At the other extreme, PGC 10314 (NGC 1058) is the most anorexic of galaxies in the current database. The profile is shown in Figure 7. This galaxy has been identified by Lewis (1975) and van der Kruit & Shostak (1984) as an example of a galaxy seen almost face on.

There are lots of bad spectra among the 15,000 profiles in All Digital H\(_i\). The two main reasons for bad spectra are poor S/N and confusion from multiple sources in the radio beam. It is easy to find examples of poor spectra; select cases with error assignments \(e_W > 20\) km s\(^{-1}\). We adopt the convention of assigning \(e_W = 100\) km s\(^{-1}\) in cases of confusion, and \(e_W = 500\) km s\(^{-1}\) in cases of uncertain or null detections, although we have not been consistent. Our fundamental convention is, if the profile is inadequate to the task of measuring a distance, then an error greater than 20 km s\(^{-1}\) is assigned. The exact value of an error assignment greater than 20 km s\(^{-1}\) has little rigor.

While it is not worthwhile to dwell on the bad, it is instructive to consider a few examples of the ugly. A cautionary example is illustrated in Figure 8. The two profiles were obtained with the Arecibo Telescope and the Green Bank 140 foot Telescope with, respectively, half-power beams of 3′ and 21′. The galaxy NGC 7814 has a diameter at the B band isophot of 25 mag as\(^{-2}\) of 5.5, larger than the Arecibo beam but much smaller than the 140 foot beam. Flux is lost in a single beam pointing with the Arecibo Telescope but not with a pointing involving the smaller telescope. The lost flux from the extremities of the galaxy with the Arecibo observation causes a pronounced reduction in the “horns” and affects the measurement of the profile width.

In Figure 9, one sees what seems to be a normal edge-on spiral galaxy but one horn is very pronounced and the other is almost unseen. A profile like this creates a problem if the measurement of the width is referenced to the peak flux. It creates a somewhat less problem with our derivation based on the mean flux.

In the case shown in Figure 11, the profile has gone beyond ugly to bad. There is a substantial wing on the long-wavelength side. The line width is acutely sensitive to the choice of level of measurement. The galaxy looks distorted. A galaxy interaction is suspected.

With IC 2511 seen in Figure 12 the profile is merely ugly. There is a wing on the short-wavelength side. The line width is sensitive to the choice of measurement flux level, but not sufficient to cause us to reject the profile by assigning an error greater than 20 km s\(^{-1}\). There is no evidence of an abnormality in the image of the galaxy. It can be appreciated that there is a continuum of situations between those shown in Figures 11 and 12, commonly aggravated by much worse confusion from noise. In the final analysis, profiles have been accepted or rejected
Figure 10. PGC 4063 (UGC 711). The profile at left obtained with the Arecibo Telescope is extremely asymmetric and is suspected to be biased because of beam attenuation. The profile at right was obtained with the five times larger beam of the Parkes Telescope. The asymmetry of the profile is not just a resolution effect. (A color version of this figure is available in the online journal.)

(assigned errors less or greater than 20 km s\(^{-1}\)) on the basis of visual inspection.

An example of contamination from multiple sources in the radio beam is provided by Figure 13. Both the galaxy at the center of the image and the fainter object 4.7 northeast have been observed with separate pointings with the Arecibo Telescope. The half-power beam diameter with that telescope is 3 arcmin. One of the sources is cleanly detected with a peak flux four times greater than the other. Lo and behold, it is the smaller fainter galaxy to the northeast. The visibly dominant galaxy is detected in H\(_{\text{i}}\) but the profile is messy and probably contaminated by the flux from the companion. We assign an error of 100 km s\(^{-1}\) (confused) to the line width of the brighter galaxy.

Another all-too-common situation is illustrated in Figure 14. The profile is anomalous, with a pronounced peak and shoulders. In the image, a second galaxy of unknown velocity is seen that would lie near the half-power level of the Arecibo Telescope beam. A third galaxy, fainter and more distant, could conceivably contribute to the confusion.

These examples give a reminder of the advantages of observations with different facilities. The relevant telescopes and the single beam half-power field diameters are listed in Table 2. The Nançay Telescope has an unusual beam shape that provides good resolution east–west but poor resolution north–south. The Arecibo Telescope provides the best resolution but attention must be given to possible loss of flux with large targets.

The examples given attention in this section provide the warning that there are lots of unacceptable line profiles for our purposes. Even among those identified as acceptable by the error estimate, comparisons when alternative observations are available reveal that 2%–3% are bad. Still, among 15,000 profiles there are an abundance of good data. In the following section, there is an evaluation of how good is good.

5. EVALUATION OF THE NEW \( W_{m50} \) PARAMETER

Comparisons between alternative line width parameters are illustrated in Figures 15 and 16. In each case, measures with
uncertainties greater than 20 km s$^{-1}$ are rejected so the comparisons are between data that are supposed to be good. In Figure 15, the comparison is between our $W_{m50}$ parameter reported in the All Digital H\textsc{i} catalog and the $W_{20}$ parameter in Pre-Digital H\textsc{i}. In the top left panel, the $W_{m50}$ values are derived exclusively from data extracted from the Cornell H\textsc{i} archive. Of 11 cases with line width measures that deviate by more than 50 km s$^{-1}$ from the mean in a sample of 1107 galaxies, five can be traced to confused profiles caused by near neighbors. The mean difference of $\langle W_{20} - W_{m50} \rangle = 15$ km s$^{-1}$ with the 1096 remaining galaxies is expected since $W_{20}$ is measured at a fainter level. The rms scatter after elimination of the 11 most deviant cases is a reasonable 11 km s$^{-1}$. In the top right panel, there is a slight augmentation of the sample through the inclusion of new data acquired by the authors. By plotting the difference in line width measures on the ordinate there is sufficient scale resolution to detect a weak dependence in the difference in line widths with the amplitude of rotation: $W_{20} - W_{m50} = 17.7 - 0.012 W_{m50}$. In the lower panel, it is seen that the histogram of the differences $W_{20} - W_{m50}$ is symmetric about the mean. The details of this comparison are reported at the bottom of Table 1.

The comparison in Figure 16 involves two separate algorithms to determine the line width at 50% of mean flux. There is the parameter given by SHGK referred to as $W_{M50}$ and the variant determined by our procedure reported in All Digital H\textsc{i} called $W_{m50}$. As in the previous figure, the data displayed in the top left panel draw exclusively from the Cornell H\textsc{i} archive while the data used to generate the right panel are slightly augmented by new observations. The results are substantially the same. There is an offset of 6.6 km s$^{-1}$ between the SHGK parameter $W_{M50}$ and our $W_{m50}$. The SHGK parameter is evaluated at a slightly lower flux, presumably because they evaluate the mean flux over a wider wavelength window than our 90% window. There is a small but significant dependence of the offset on rotation rate: $W_{M50} - W_{m50} = 10.8 - 0.015 W_{m50}$. Again, 1% of cases are deviant by greater than 50 km s$^{-1}$, usually because of confusion caused by a companion. Those aside, the rms difference from the mean is a satisfactory 9 km s$^{-1}$. However, as seen in the lower panel, the distribution is slightly skewed, with a tail to positive differences.

Besides the Cornell H\textsc{i} archive, another important source of H\textsc{i} profile information is LEDA (Paturel et al. 2003) and the
related \text{H\,i} archive\footnote{http://klun.obs-nancay.fr} associated with the “Kinematics of the Local Universe” (KLUN) project (Theureau et al. 2006). Derivative parameters and a discussion of results are presented by Theureau et al. (2007). Access to the KLUN tabular material and profiles is provided in EDD through the \text{H\,i} Nancay catalog.

A comparison with the KLUN results is not straightforward because of mixed use of the optical and radio conventions for transforming Doppler shifts to velocities. In the optical convention, which we use, one considers the shift in wavelength with respect to the rest value, \( V_{\text{opt}} = c(\lambda - \lambda_0)/\lambda_0 \), while in the radio convention one considers the shift in frequency, \( V_{\text{rad}} = c(\nu_0 - \nu)/\nu_0 \). Profiles displayed in the Nan\c{c}ay database are presented in the radio convention although it is to be noted that the same profiles made available through the NED, the NASA/IPAC Extragalactic Database,\footnote{nedwww.ipac.caltech.edu/forms/SearchSpectra.html} have been converted to the optical convention. The tabular information presented by Theureau et al. (2007) is mixed. Systemic velocities have been transformed to the optical convention but line widths appear to have remained in the radio convention. For a galaxy at 7000 km s\(^{-1}\) with a line width of \( \sim 400 \) km s\(^{-1}\) the line width in the optical convention is \( \sim 20 \) km s\(^{-1}\) wider than in the radio so the issue is significant.

We have made comparisons between KLUN line widths at 20\% of peak intensity, adjusted to the optical convention, and \( W_{m50} \) line width values drawn from the \textit{All Digital \text{H\,i} catalog. The comparison accepts only cases from the \textit{All Digital \text{H\,i} catalog with line width errors \( \leqslant 20 \) km s\(^{-1}\) and cases from the \text{H\,i} Nancay catalog with a line width quality index assigned by us of 1–3, see Figure 17. After rejection of four extreme outliers, the scatter is 13 km s\(^{-1}\). There is a hint of a correlation with \( W_{m50} \) as found in Figures 15 and 16. The comparison sample is restricted because for the moment only a small fraction of galaxies in the \text{H\,i} Nancay catalog have been assigned a quality
index by us. If galaxies without a quality index are accepted, then the comparison can be based on an order of magnitude larger sample of 992 galaxies. However, the scatter is then much worse. Even after clipping 48 cases with deviations greater than 80 km s\(^{-1}\) from the mean the scatter is a poor 22 km s\(^{-1}\). There will be a discussion in the following section of the results of our own analysis of profiles extracted from the Nançay database. It will be seen that when the data are treated in a uniform way there is good agreement between all sources, whether drawn from the Nançay or Cornell databases or derived from our new observations with GBT and Arecibo.

The weak but significant tilt in the difference plots as a function of rotation rate seen in Figures 15–17 is a consequence of a different line width definition. Our line widths are measured at relatively higher flux levels for small galaxies and approach the levels of the other measures for large galaxies. A result could be a slightly flatter slope for the luminosity–line width correlation. These small differences serve to emphasize that the line width measures are empirical constructs. We are reminded that the details of the construct may not be important but consistency is required if biases are to be avoided.

6. COMPARISON OF NEW AND OLD DATA

The comparisons in the previous section are between different measures of line widths with largely the same data. In this section, there is a comparison of new and old observations of identical targets, all analyzed with the procedure discussed in this paper. A point of detail: the line widths discussed in this section have received the small correction for spectral resolution discussed in the following section to facilitate the comparison of observations made with different telescopes and receivers.

Figure 18 shows the current status of comparisons between new and Cornell archival profiles, giving separate consideration to new GBT and Arecibo observations. The archival profiles come from a variety of telescopes, never GBT. In all cases, the line width measure is our \(W_{m50}\) parameter.

The rms scatter between line widths drawn from material out of the Cornell \(H\) \(I\) archive and line widths determined from new observations is a satisfactory 10 km s\(^{-1}\). If errors are partitioned equally, the implied uncertainty is 7 km s\(^{-1}\) in each of the new and old measures. The scatter is similar in the separate GBT and Arecibo comparisons.

There is a minor mystery in the zero-point offsets seen in both samples displayed in Figure 18. The line widths measured from the archival profiles are 3 km s\(^{-1}\) wider in the mean, a difference with 3\(\sigma\) significance. It may be a factor that the new profiles have higher \(S/N\) in the great majority of cases. This is not a negative reflection on the archival material but simply a consequence of our strategy of primarily re-observing objects with poor profiles (though recalling that profiles ascribed errors larger than 20 km s\(^{-1}\) are rejected in all comparison samples). With the 258 galaxies represented in Figure 18, the mean difference in the error assigned to the line width, archive minus new, is 6 km s\(^{-1}\).

An explanation for the zero-point offset might be that line widths are slightly overestimated with noisier profiles.

Another large component of the All Digital \(H\) \(I\) catalog is built from data extracted from the Nançay database. To be clear, the profile fits and derivative parameters for observations from the Nançay Telescope given through the All Digital \(H\) \(I\) catalog are based on the analysis procedures described in this paper; results from the original source of the data are found in the catalog \(H\) \(I\) Nançay. This is analogous to the distinction in the case of the material from Cornell, with original source material in catalog Springob/Cornell \(H\) \(I\) and re-analyzed material in catalog All Digital \(H\) \(I\). There are 720 galaxies in All Digital \(H\) \(I\) with both a satisfactory (\(eW \leq 20\) km s\(^{-1}\)) Nançay line width and a satisfactory line width either from the Cornell archive or new as reported here. The difference between new/Cornell and Nançay widths is \(<W_{new/cornell} - W_{nançay}>\) = 2.6 ± 0.4 with rms scatter 10.8 km s\(^{-1}\) after rejection of seven cases with excursions in excess of 4\(\sigma\). The comparison is shown graphically in Figure 19. The rms scatter is at a level that, in comparison with the discussion surrounding Figure 18, implies a characteristic uncertainty with the Nançay widths of 8 km s\(^{-1}\).

The offset of 2.6 km s\(^{-1}\) seen in Figure 19 is small but statistically significant. The comparison shown here is with widths corrected for line broadening. If the comparison is made on the directly observed \(W_{m50}\) line widths, then the difference is −0.1 km s\(^{-1}\). The offset arises through the broadening corrections discussed in the following section. The Nançay data tend to need larger corrections and our recipe may be slightly excessive for this sample. However, the problem is sufficiently small that it will not affect the measurement of distances.

At the most southerly latitudes, neutral hydrogen observations require use of the Parkes Telescope and results are becoming
available on their archive. An important contribution has come from the HIPASS (Koribalski et al. 2004). A comparison with material drawn from this database is summarized in the histogram of Figure 20. The difference in line widths measured off Parkes spectra with those measured off spectra from the Cornell or Nançay archives or from own new observations for 205 galaxies is $\langle W_{\text{other}} - W_{\text{parkes}} \rangle = 0.7 \pm 0.5$ with rms scatter 7.3 km s$^{-1}$. The excellent very low scatter can be attributed in part to the fact that the galaxies in the comparison tend to be nearby and easily detected in H$\alpha$.

### 7. LINE WIDTH ADJUSTMENTS

Several systematics affect line widths. Two that are well understood are a slight relativistic broadening and broadening because of finite spectral resolution. SHGK discussed these matters at length and we adopt a simplified version of their solution. We adjust line widths with the equation

$$W_{c_{m50}} = \frac{W_{m50}}{1 + z} - 2\Delta \nu \lambda,$$

(1)

where $cz$ is the heliocentric velocity of the galaxy, $\Delta \nu$ is the spectral resolution after smoothing, and $\lambda$ is determined empirically. SHGK, with a closely equivalent formula, gave a convoluted recipe for $\lambda$. Their complex description may be appropriate in their case because of a coupling between S/N and their peak measurement, hence their line width measure. With our parameter based on mean flux there is less systematic dependence on S/N. From tests on profiles with successively increased smoothing, we find that broadening is statistically described by Equation (1) if $\lambda = 0.25$. This correction is close to those advocated by Bottinelli et al. (1990) and Verheijen & Sancisi (2001).

---

**Figure 15.** Top left: comparison between the line width at 50% of mean flux in the catalog *All Digital H$\alpha$* and the line width at 20% of peak intensity in the catalog *Pre Digital H$\alpha$*. Top right: line width difference as a function of $W_{m50}$ with a least-squares fit superposed. Bottom: histogram of differences $W_{20} - W_{m50}$.

(A color version of this figure is available in the online journal.)
Figure 16. Top left: comparison between two alternative estimators of the line width at 50% of mean flux, \(W_{m50}\) from the All Digital \(\text{H}i\) catalog and \(W_{M50}\) from the Springob/Cornell \(\text{H}i\) catalog. Top right: line width differential plot with a least-squares fit as a function of \(W_{m50}\). Bottom: histogram of differences \(W_{M50} - W_{m50}\). The distribution is skewed, with a median difference of 5 km s\(^{-1}\).

(A color version of this figure is available in the online journal.)

The measured profile line width, whether it is our \(W_{m50}\) or another, is only a parameter of observational convenience and it is desirable to translate it into something more physically meaningful. Rectification to edge-on orientation is standard (division by \(\sin i\) where \(i\) is the inclination from face-on). It is also common, but less secure, to adjust the observed line width to correspond statistically with twice the maximum rotation velocity, \(V_{\text{max}}\). The adjustment is based upon samples with both global profiles and detailed rotation curves. Tully & Fouquè (1985) investigated the matter and provided a description that accounts for the effects of broadening by turbulent motions that transitions from a linear to quadratic correction as the unbroadened profile transitions from roughly boxcar in giant galaxies to Gaussian in dwarfs. Their formula is

\[
W_{R,\ell}^2 = W_{\ell}^2 + W_{t,\ell}^2[1 - 2e^{-(W_{\ell}/W_{c,\ell})^2}] + 2W_{\ell}W_{t,\ell}[1 - e^{-(W_{\ell}/W_{c,\ell})^2}],
\]

(2)

where the subscript \(\ell\) stands for the observed line width measure, \(W_{\ell}\) is the turbulent broadening for that observed measure, and \(W_{c,\ell}\) characterizes the transition from boxcar to Gaussian intrinsic profiles. In the case \(W_{\ell} = W_{20}\), Tully & Fouquè (1985) recommended \(W_{t} = 38\,\text{km s}^{-1}\) and \(W_{c} = 120\,\text{km s}^{-1}\).

More recently, the problem has been studied in detail by Verheijen & Sancisi (2001). They compared global profiles with detailed rotation curve information for galaxies observed with the Westerbork Synthesis Radio Telescope. They determined that the Tully–Fouquè value taken for \(W_{t}\) was too large that to get \(\langle W_{R} - 2V_{\text{max}}\sin i\rangle \simeq 0\) with the transformation of \(W_{20}\) to \(W_{R}\) of Equation (2) requires \(W_{t} = 22\,\text{km s}^{-1}\).

The transformations can be expected to be slightly different in detail with the new line width parameter \(W_{m50}\). Comparisons have been made with 35 galaxies in the Ursa Major Cluster with rotation curves determined from observations with the Westerbork Synthesis Radio Telescope and \(V_{\text{max}}\) values reported by Verheijen (2001). As anticipated by Verheijen & Sancisi
Figure 17. Difference between the Nançay/KLUN line width at 20% of peak intensity and the $W_{m50}$ line width parameter in the All Digital $\text{H}_\text{i}$ catalog plotted against the $W_{m50}$ parameter. A least-squares fit is superimposed. The slope has only a $2.2\sigma$ significance.

(A color version of this figure is available in the online journal.)

The effect of measuring the line width at a higher flux level above the baseline requires reduction of $W_{c,\ell}$ and, especially, $W_{t,\ell}$. The optimal fit results in the correlation seen in Figure 21. To differentiate from parameter variations discussed in earlier publications, we define $W_{m_{\text{mx}}} \equiv W_{R,m_{50}}$ and find a best fit for the parameters in Equation (2) with $W_{c,m_{50}} = 100 \text{ km s}^{-1}$ and $W_{t,m_{50}} = 9 \text{ km s}^{-1}$. With these parameters, observed $W_{m50}$ line widths are transformed into $W_{m_{\text{mx}}}$ line widths that agree with $2V_{\text{max}}$ with an rms scatter of 12 km s$^{-1}$ after deprojection (rms scatter 10 km s$^{-1}$ in the line of sight). This scatter is comparable to the $W_{m50}$ measurement accuracy.

A detailed study of the relationship between observed line widths and the intrinsic kinematics of galaxies was carried out by Singhal (2008). Such a study is particularly important if the interest is to understand the physical basis for the relationship between galaxy rotation and the light distribution. The slope of the correlation can be affected by the details of measurements and adjustments. For the practical matter of measuring distances the greater importance is to be consistent.

8. THE ALL DIGITAL $\text{H}_\text{i}$ CATALOG IN THE EDD

The All Digital $\text{H}_\text{i}$ catalog is accessed by selecting the “next” button on the EDD home page. It can be selected alone or in tandem with any of the other catalogs and either all or any fraction of the elements within the catalog can be selected. The tabular portion of the catalog is displayed with the “select” button and can be exported with the “download” button. Upon entering the tabular display, one can navigate to graphical displays of $\text{H}_\text{i}$ profiles by selecting on the common name of a galaxy (selecting on the PGC name in this, and any of the other catalogs, brings up a digital sky survey image of the galaxy). An example of what will be found is shown in Figure 22. The galaxy seen in this case is PGC 19996 = ESO 491-015. The image of the galaxy is drawn from the LEDA Web site and displayed with a field of 10 arcmin, roughly the beam size of the GBT, hence a scale reasonable for an inspection for contamination from any near neighbors. The left profile was acquired by the authors with observations using the GBT. The right profile is based on observations with the Green Bank 140 foot Telescope and was obtained from the Cornell archive. The two profiles are shown after treatment by the same analysis pipeline.

The results of the analysis are carried to Table 4, the main catalog, which includes the following information. The parameters given in Columns 3–6 are averaged over multiple observations. Those in Columns 7–18 are for an individual observation and the columns repeat if there are multiple observations. The first
Table 4
First 10 Lines of All Digital H I Catalog

| PGC Name/Profile | Vhel1 (km s⁻¹) | Wm501 (km s⁻¹) | \( \omega_{w} \) (km s⁻¹) | Nv1 | Source1 | Tel1 | Vhav (km s⁻¹) | Wmxav (km s⁻¹) | eWmxav (km s⁻¹) | We50 (km s⁻¹) | Wm501 (km s⁻¹) | eWm501 (km s⁻¹) | SN1 | Flux1 (Jy km s⁻¹) | Res1 | Ns1 | Fm501 (mJy) |
|------------------|----------------|----------------|-----------------|-----|--------|------|----------------|----------------|----------------|-------------|----------------|----------------|-----|----------------|------|-----|------------|
| 4 AGC331060      | 4458           | 154            | 16               | 1   | shg2005 | AOlf | 4458           | 173            | 162            | 154          | 16             | 8.5            | 1.85 | 8.5            | 2    | 5.3 |
| 6 AGC331061      | 6002           | 217            | 20               | 1   | shg2005 | AOlf | 6002           | 248            | 226            | 217          | 20             | 2.0            | 0.82 | 8.6            | 4    | 1.6 |
| 12 PG0000012     | 6548           | 400            | 19               | 1   | tmc2006 | Nanc | 6548           | 424            | 409            | 400          | 19             | 2.4            | 3.40 | 11.0           | 1    | 3.9 |
| 16 PG0000016     | 5591           | 151            | 20               | 1   | tmc2006 | Nanc | 5591           | 173            | 159            | 151          | 20             | 3.2            | 0.96 | 11.0           | 2    | 1.5 |
| 20 AGC331066     |                |                |                  |     | shg2005 | AOlf | 7380           | 269            | 245            | 50           | 3.8            | 2.40           | 8.52 | 8.7            | 4    | 2.8 |
| 29 AGC331067     |                |                |                  |     | shg2005 | AOlf | 12701          | 163            | 147            | 50           | 2.4            | 0.53           | 8.9  | 8.9            | 2    | 2.0 |
| 38 UGC12893      | 1108           | 78             | 19               | 1   | shg2005 | AOlf | 1108           | 87             | 82             | 78           | 19             | 3.8            | 2.41 | 8.5            | 1    | 13.8|
| 40 PG0000040     | 7282           | 289            | 20               | 1   | tmc2006 | Nanc | 7282           | 316            | 298            | 289          | 20             | 5.0            | 0.01 | 10.8           | 2    | 0.0 |
| 47 UGC12896      |                |                |                  |     | shg2005 | AOlf | 7676           | 181            | 172            | 25           | 3.6            | 2.61           | 8.8  | 4.1            | 5    | 5.9 |
| 53 UGC12895      | 6769           | 158            | 17               | 1   | shg2005 | AOlf | 6769           | 175            | 167            | 158          | 17             | 6.1            | 3.76 | 8.8            | 1    | 10.5|

(This table is available in its entirety in machine-readable and Virtual Observatory (VO) forms in the online journal. A portion is shown here for guidance regarding its form and content.)

Figure 19. Average of our new line widths and line widths derived from the Cornell archive data compared with line widths derived from data from the Nançay archive. All data are analyzed with the same pipeline designed to measure the line width at 50% of mean flux.

Figure 20. Average of our new line widths and line widths derived from the Cornell and Nançay archive data compared with line widths derived from data from the Parkes archive. All data are analyzed with the same pipeline designed to measure the line width at 50% of mean flux.

Figure 21. Comparison between the global profile parameter \( W_{mx} \) and twice the maximum rotation velocity determined from spatially resolved rotation curves. Few lines of the table are given in the print version of this paper.

1. Principal Galaxies Catalog name from LEDA.
2. Common name (click on name to view profiles).
3. Weighted average heliocentric velocity from all acceptable profiles (km s⁻¹). Weights are based on the inverse square of assigned errors.
4. Weighted average line width approximating twice the maximum rotation velocity before projection from all acceptable profiles, \( W_{mx} \) (km s⁻¹).
5. Uncertainty attributed to the line width, the inverse square root of the sum of individual weights (km s⁻¹).
6. Number of acceptable profiles (errors less than or equal to 20 km s⁻¹).
7. Source of observation. Source codes are identified in Table 3 for sources incorporated at the time of publication.
8. Telescope and receiver, see Table 2 for more information.
9. Heliocentric velocity, the midpoint between the low and high velocities that define \( W_{m50} \) (km s⁻¹).
10. \( W_{m50} \): line width at 50% of the mean flux per channel where the mean flux is calculated within the 90% window,
the range of velocities excluding $5\%$ of the integrated flux at each end of the profile ($\text{km s}^{-1}$).

11. Line width corrected for relativistic and instrumental broadening, $W_{m50} = \frac{W_{m50}}{1+z} + 2\Delta v\lambda$ where $\lambda = 0.25$ and $\Delta v$ is the product of the values in Columns 16 and 17.

12. Line width adjusted to statistically equal twice the maximum rotation velocity, before deprojection, derived from spatially resolved rotation curves ($\text{km s}^{-1}$). Statistically, $W_{mx} \sim 2V_{\text{max}}\sin i$. This parameter is only considered meaningful if $eW \leq 20$ km s$^{-1}$.

13. Uncertainty in the line width ($\text{km s}^{-1}$). Uncertainties less than or equal to $20$ km s$^{-1}$ are considered adequate for the purpose of determining a distance through the luminosity–linewidth correlation. An initial assignment of uncertainty, $eW$, is based on S/N: $eW = 8$ km s$^{-1}$ if S/N $\geq 17$; $eW = 21.6-0.8S/N$ if $17 > S/N > 2$; $eW = 70-25S/N$ if S/N $< 2$. If the spectral resolution after smoothing is greater than this assignment, then the error is increased to match the smoothed resolution. The uncertainty may have been modified by manual intervention to either increase it to above $20$ km s$^{-1}$ if the profile is too poor to be used for a distance measure or to decrease it to equal or below $20$ km s$^{-1}$ if the profile is considered adequate for this purpose. The error value $eW = 100$ km s$^{-1}$ is used to signal cases of confusion and the error value $eW = 500$ km s$^{-1}$ identifies a dubious/null detection.

14. Signal to noise. The signal is the mean flux per channel within the velocity range of the 90% window. The noise is calculated over 100 channels on each side of the signal outside the velocity range of the 100% window.

15. The flux is the signal integrated over all channels within the 100% window (Jy km s$^{-1}$). No attempt has been made to account for flux lost due to the finite beam size.

16. Channel resolution ($\text{km s}^{-1}$).

17. An integer $N$ indicates averaging over $N$ spectral channels in the profile that is displayed.

18. $F_{m50}$ is the flux level at 50% of the mean, the level at which the measurement of $W_{m50}$ is made.

Figure 22. Example of a graphical display accessed from the All Digital H\textsc{i} catalog. (A color version of this figure is available in the online journal.)
The most recent status of the full Table 4 catalog is made available with the electronic version of this paper.

9. SUMMARY

The purposes of the All Digital H\textsc{i} catalog in EDD are threefold: first, to make available the results of new observations made of galaxies in the 21 cm H\textsc{i} line; second, to make it easy to compare results with other observations and link to other information about the targets; third, to present a reanalysis of all archival data available in digital form to ensure that consistent line width information is available for essentially all galaxies that have been observed in H\textsc{i}.

Our preferred line width parameter is $W_{50}$, the profile width at 50% of the mean flux within the velocity window containing 90% of the total flux. This parameter is a variant of one of those introduced by the Cornell group (Springob et al. 2005). The availability within EDD of several very large catalogs of H\textsc{i} information facilitates comparisons and provides a way of culling bad data. It is satisfying to see the tight correlations between alternative profile descriptors. It is to be appreciated that various alternative profile descriptors have merit but it is important for the measurement of distances to maintain consistency. After accounting for zero-point offsets, rms scatter between alternatives is at the level of 10 km s$^{-1}$.

Based on comparisons with detailed rotation curve information, a statistical transformation is proposed that takes the observed global line widths to an approximation of the maximum rotation velocity $V_{\text{max}}$.

Presently the catalog All Digital H\textsc{i} contains 15,411 profiles providing information on 13,423 galaxies. Some 57% of the profiles originate from the Cornell database (8740), 21% originate from the Nan\c{c}ay database (3225), 7% come from the Arecibo Legacy Fast Alfa survey (1047), another 7% were extracted from the Parkes archive (997), 1% are an Effelsberg contribution independent of the Cornell database (176), and 8% result from new observations by our collaboration (1225). Currently, profiles for 10,580 galaxies are deemed acceptable. In 1330 cases, there are at least two acceptable profiles and in 81 cases there are three acceptable profiles. Inter comparisons between sources suggest that the characteristic accuracy of an individual acceptable profile width is 7 km s$^{-1}$.

New observations across the entire sky have been made possible by access to three fine radio telescopes. We made early observations with the refurbished Arecibo Telescope and expect to add fresh material coming from the wide field multibeam survey. At the Green Bank Telescope our ongoing project, Cosmic Flows, has been awarded the status of a Large Program. Observations of the deep southern sky began in 2009 with the Parkes Telescope in Australia. Equally important to us has been access to archival material from the Cornell Digital H\textsc{i} Archive, the Nan\c{c}ay Radio Telescope H\textsc{i} profiles of galaxies database, and the Australia Telescope online archive. Although electronic archives are a great innovation, the low-tech information gathered in the Pre-Digital H\textsc{i} catalog retains a great value and we thank Cyrus Hall for his role in assembling that material. We have made extensive use of NED, the NASA/IPAC Extragalactic Database operated by the Jet Propulsion Laboratory, California Institute of Technology, and the HyperLeda database hosted at the Université Lyon 1. R.B.T. is supported through the National Science Foundation award NSF0908846. Web access to the All Digital H\textsc{i} catalog is found at http://edd.ifa.hawaii.edu.

REFERENCES

Barnes, D. G., et al. 2001, MNRAS, 322, 486
Bell, E. F., Baugh, C. M., Cole, S., Frenk, C. S., & Lacey, C. G. 2003, MNRAS, 343, 367
Bottinelli, L., Gouguenheim, L., Fouque, P., & Paturel, G. 1990, A&AS, 82, 391
Catinella, B., Haynes, M. P., & Giovanelli, R. 2007, AJ, 134, 334
Catinella, B., Haynes, M. P., Giovanelli, R., Gardner, J. P., & Connolly, A. J. 2008, ApJ, 685, L13
Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., & Broderick, J. J. 1998, AJ, 115, 1693
Conselice, C. J., Bundy, K., Ellis, R. S., Brinchmann, J., Vogt, N. P., & Phillips, A. C. 2005, ApJ, 628, 160
Courteau, S. 1997, AJ, 114, 2402
Dale, D. A., Giovanelli, R., Haynes, M. P., Campusano, L. E., & Hardy, E. 1999, AJ, 118, 1489
Dutton, A. A., van den Bosch, F. C., Dekel, A., & Courteau, S. 2007, ApJ, 654, 27
Fisher, J. R., & Tully, R. B. 1981, ApJS, 47, 139
Giovanelli, R., Haynes, M. P., Rubin, V. C., & Ford, W. K., Jr. 1986, ApJ, 301, L7
Huchtmeier, W. K., & Richter, O.-G. (ed.) 1989, A General Catalog of H\textsc{i} Observations of Galaxies. The Reference Catalog (New York: Springer-Verlag)
Koribalski, B. S., et al. 2004, AJ, 128, 16
Lewis, B. M. 1975, A&A, 44, 147
Paturel, G., Petit, C., Prugniel, P., Theureau, G., Rousseau, J., Brouty, M., Dubois, P., & Cambresy, L. 2003, A&A, 412, 45
Singhal, A. 2008, PhD thesis, Univ. Virginia
Springob, C. M., Haynes, M. P., Giovanelli, R., & Kent, B. R. 2005, ApJS, 160, 149
Theureau, G., Hanski, M. O., Coudreau, N., Hallet, N., & Martin, J.-M. 2007, A&A, 465, 71
Theureau, G., Martin, J.-M., Cognard, I., & Borsenberger, J. 2006, in ASP Conf. Ser. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique (San Francisco, CA: ASP), 429
Tully, R. B., & Fisher, J. R. 1977, A&A, 54, 661
Tully, R. B., & Fouqu\é, P. 1985, ApJS, 58, 67
Tully, R. B., & Pierce, M. J. 2000, ApJ, 533, 744
Tully, R. B., Rizzi, L., Shaya, E. J., Courtois, H. M., Makarov, D. I., & Jacobs, B. A. 2009, AJ, 138, 323
Tully, R. B., Shaya, E. J., Karachentsev, I. D., Courtois, H. M., Kocovski, D. D., Rizzi, L., & Pelet, A. 2008, ApJ, 676, 184
Tully, R. B., Verheijen, M. A. W., Pierce, M. J., Huang, J.-S., & Wainscoat, R. J. 1996, AJ, 112, 2471
van der Kruit, P. C., & Shostak, G. S. 1984, A&A, 134, 258
Verheijen, M. A. W. 2001, ApJ, 563, 694
Verheijen, M. A. W., & Sancisi, R. 2001, A&A, 370, 765
Weiner, B. J., et al. 2006, ApJ, 653, 1049
Zwaan, M. A., Meyer, M. J., Staveley-Smith, L., & Webster, R. L. 2005, MNRAS, 359, L30

1956 COUTROIS ET AL. Vol. 138