

THE ARECIBO LEGACY FAST ALFA SURVEY: THE ALFALFA EXTRAGALACTIC HI SOURCE CATALOG

MARTHA P. HAYNES¹, RICCARDO GIOVANELLI¹, BRIAN R. KENT², ELIZABETH A.K. ADAMS^{3,4}, THOMAS J. BALONEK⁵, DAVID W. CRAIG⁶, DEREK FERTIG⁷, ROSE FINN⁸, CARLO GIOVANARDI⁹, GREGORY HALLENBECK¹⁰, KELLEY M. HESS^{4,3}, G. LYLE HOFFMAN¹¹, SHAN HUANG¹, MICHAEL G. JONES¹², REBECCA A. KOOPMANN¹⁴, DAVID A. KORNREICH¹, LUKAS LEISMAN¹³, JEFFREY MILLER¹⁵, CRYSTAL MOORMAN¹⁶, JESSICA O'CONNOR¹⁷, AILEEN O'DONOGHUE¹⁵, EMMANOUIL PAPASTERGIS^{4,18}, PARKER TROISCHT¹⁹, DAVID STARK²⁰ LI XIAO²¹

¹Cornell Center for Astrophysics and Planetary Science, Space Sciences Building, Cornell University, Ithaca, NY 14853, USA; haynes@astro.cornell.edu

²National Radio Astronomy Observatory, 520 Edgemont Rd., Charlottesville, VA 22901, USA

³ASTRON, Netherlands Institute for Radio Astronomy, Postbus 2, 7900 AA, Dwingeloo, The Netherlands

⁴Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD, Groningen, The Netherlands

⁵Department of Physics and Astronomy, Colgate University, Hamilton, NY 13346, USA

⁶West Texas A&M University Department of Chemistry and Physics, 2403 Russell Long Blvd. Canyon, TX 79015, USA

⁷Langley High School, 6520 Georgetown Pike, McLean, VA 22101, USA

⁸Physics Department, Siena College, Loudonville, NY 12211, USA

⁹Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy

¹⁰Washington and Jefferson College, Department of Computing and Information Studies, 60 S Lincoln Street, Washington PA, 15301, USA

¹¹Department of Physics, Hugel Science Center, Lafayette College, Easton, PA 18042, USA

¹²Instituto de Astrofísica de Andalucía, CSIC, Glorieta de la Astronomía s/n E-18008, Granada, Spain

¹³Department of Physics and Astronomy, Valparaiso University, Neils Science Center, 1610 Campus Drive East, Valparaiso, IN 46383, USA

¹⁴Department of Physics and Astronomy, Union College, 807 Union Street, Schenectady, NY 12308, USA

¹⁵St. Lawrence University, 23 Romoda Drive, Canton, NY 13617, USA

¹⁶Department of Physics, Lynchburg College, 1501 Lakeside Drive, Lynchburg, VA 24501, USA

¹⁷Department of Physics and Astronomy, George Mason University, 4400 University Drive, MSN: 3F3, Fairfax, VA 22030, USA

¹⁸Credit Risk Modeling Department, Coperative Rabobank U.A., Croeselaan 18, Utrecht NL-3521CB, The Netherlands

¹⁹Department of Physics, Hartwick College, Oneonta, NY 13820, USA

²⁰Kavli Institute for the Physics and Mathematics of the Universe (WPI), The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan

²¹National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100012, China

ABSTRACT

We present the catalog of ~ 31500 extragalactic HI line sources detected by the completed ALFALFA survey out to $z < 0.06$ including both high signal-to-noise ratio (> 6.5) detections and ones of lower quality which coincide in both position and recessional velocity with galaxies of known redshift. We review the observing technique, data reduction pipeline, and catalog construction process, focusing on details of particular relevance to understanding the catalog's compiled parameters. We further describe and make available the digital HI line spectra associated with the catalogued sources. In addition to the extragalactic HI line detections, we report nine confirmed OH megamasers and ten OH megamaser candidates at $0.16 < z < 0.22$ whose OH line signals are redshifted into the ALFALFA frequency band. Because of complexities in data collection and processing associated with the use of a feed-horn array on a complex single-dish antenna in the terrestrial radio frequency interference environment, we also present a list of suggestions and caveats for consideration by users of the ALFALFA extragalactic catalog for future scientific investigations.

1. INTRODUCTION

HI 21 cm line surveys provide a census of the extragalactic population of atomic gas-bearing galaxies. Because of the relatively simple physics involved in most HI line emission, conversion of the observed line flux into atomic hydrogen gas mass is straightforward, and the spectral nature of the emission provides observable measures of the redshift and

projected disk rotational velocity. While the molecular H_2 gas tends to concentrate in a small number of giant gas clouds principally in the inner regions, the HI disk traces the full extent of the gas layer. Star formation is linked more closely to the molecular H_2 gas (e.g. Kennicutt & Evans 2012; Saintonge et al. 2016; Catinella et al. 2018). However, in most galaxies, the HI fills a much larger fraction of interstellar space and contributes most of the cool gas mass, thus representing the fuel reservoir and potential for future star formation.

The Arecibo Legacy Fast ALFA (ALFALFA) Survey used the seven-horn Arecibo L-band Feed Array (ALFA) to map nearly 7000 deg^2 of high Galactic latitude sky accessible to the Arecibo telescope over ~ 4400 nighttime hours between 2005 and 2011. ALFALFA was conducted as a “blind” survey: at each position, the entire frequency range from 1335-1435 MHz, corresponding to heliocentric velocities $-2000 < cz < 18000 \text{ km s}^{-1}$, was searched for line emission. As described in detail by Giovanelli et al. (2005), the ALFALFA survey design was largely dictated by the principal science goal of determining the faint end of the HI mass function (HIMF), and the overall abundance of low mass gas-rich halos (e.g. Martin et al. 2010; Papastergis et al. 2011; Martin et al. 2012; Papastergis et al. 2013). Additional objectives include: how the HIMF might vary with environment (e.g. Moorman et al. 2014; Jones et al. 2016b, 2018), how the HI-bearing population differs from optically-selected ones (e.g. Huang et al. 2012a,b; Gavazzi et al. 2013), using the HI distribution to look for tidal debris on large angular scales (e.g. Lee-Waddell et al. 2014, 2016; Leisman et al. 2016), and establishing metrics for the normal HI content of galaxies (e.g. Toribio et al. 2011; Odekon et al. 2016). As the least clustered local ($z \sim 0$) galaxy population (Martin et al. 2012), the HI-bearing population traces how galaxies evolve when left on their own, in relative isolation.

ALFALFA has also discovered an number of enigmatic objects such as the nearby faint dwarf Leo P (Giovanelli et al. 2013), the very metal poor Leoncino (Hirschauer et al. 2016) and the highly HI-dominated Coma P (Janowiecki et al. 2015; Ball et al. 2018). Additionally, ALFALFA has provided the opportunity to survey classes of galaxies such as extremely low HI mass dwarfs (e.g. Cannon et al. 2011; Teich et al. 2016; McNichols et al. 2016), galaxies with extremely high HI-to-stellar mass ratios (e.g. Adams et al. 2015b; Janowiecki et al. 2015; Janesh et al. 2015, 2017) and HI-bearing ultra diffuse galaxies (Leisman et al. 2017). The vast majority ($> 98\%$) of extragalactic ALFALFA sources can be associated with at least one likely stellar counterpart, and the majority of the “dark” objects are likely associated with tidal debris in interacting systems (e.g. Haynes et al. 2007; Koopmann et al. 2008; Lee-Waddell et al. 2014; Leisman et al. 2016). A few dark galaxy candidates remain intriguing, and continuing work seeks to identify associated starlight and constrain their dynamics and star formation history (e.g. Kent 2010; Giovanelli et al. 2010; Cannon et al. 2015).

As a complement to Jones et al. (2018) which presents the derived HIMF and its dependence on local environment, this paper presents the extragalactic HI catalog extracted from the completed ALFALFA survey. Because of the overlapping nature of the drift scan survey and improved availability of the public optical imaging used to identify optical counterparts (OCs) of ALFALFA HI sources, this catalog, presented in Table 2, supersedes and replaces previous releases (Giovanelli et al. 2007; Saintonge et al. 2008; Kent et al. 2008; Stierwalt et al. 2009; Martin et al. 2009; Haynes et al. 2011). In addition to the catalog of ALFALFA HI line detections, nine sources are identified with OH megamasers (OHMs) and ten are flagged as being OHM candidates.

Section 2 reviews the important aspects of the ALFALFA survey observational program and data reduction process which has led to the production of the extragalactic dataset presented in Section 3. Section 4 summarizes a number of important points, realities and caveats about the survey and its resultant data products which readers are encouraged to keep in mind. Appendix A presents details of the data acquisition and processing pipeline used to produce the ALFALFA catalog.

To allow direct comparison with the vast majority of extant works on HI line emission at low redshift, we use the observed rest frame and do not apply cosmological corrections dependent on redshift; those amount to at most a few percent for the most distant sources. Details of this choice are given in the text.

2. THE ALFALFA SURVEY

The ALFALFA survey was intended to cover two sky areas at high Galactic latitude, one in the northern Galactic hemisphere $07^h 30^m < \text{R.A.} < 16^h 30^m$, $0^\circ < \text{Dec.} < +36^\circ$ and one in the southern hemisphere, $22^h < \text{R.A.} < 03^h$, $0^\circ < \text{Dec.} < +36^\circ$. For various practical reasons, the final sky area, depicted here in Figure 1, also shown in Figure 1 of Jones et al. (2018), is somewhat reduced near the edges.

As discussed in Giovanelli et al. (2005) and Giovanelli & Haynes (2015), the ALFALFA survey was designed particularly to sample the HIMF over a fair cosmological distance of $\simeq 100$ Mpc, therefore setting minimum requirements on the survey volume sensitivity and areal coverage. For a survey with a telescope characterized by a given system temperature T_{sys} and gain G , the science-driven need to detect a given HI mass M_{HI} of HI line width $W_{50} \text{ km s}^{-1}$ at

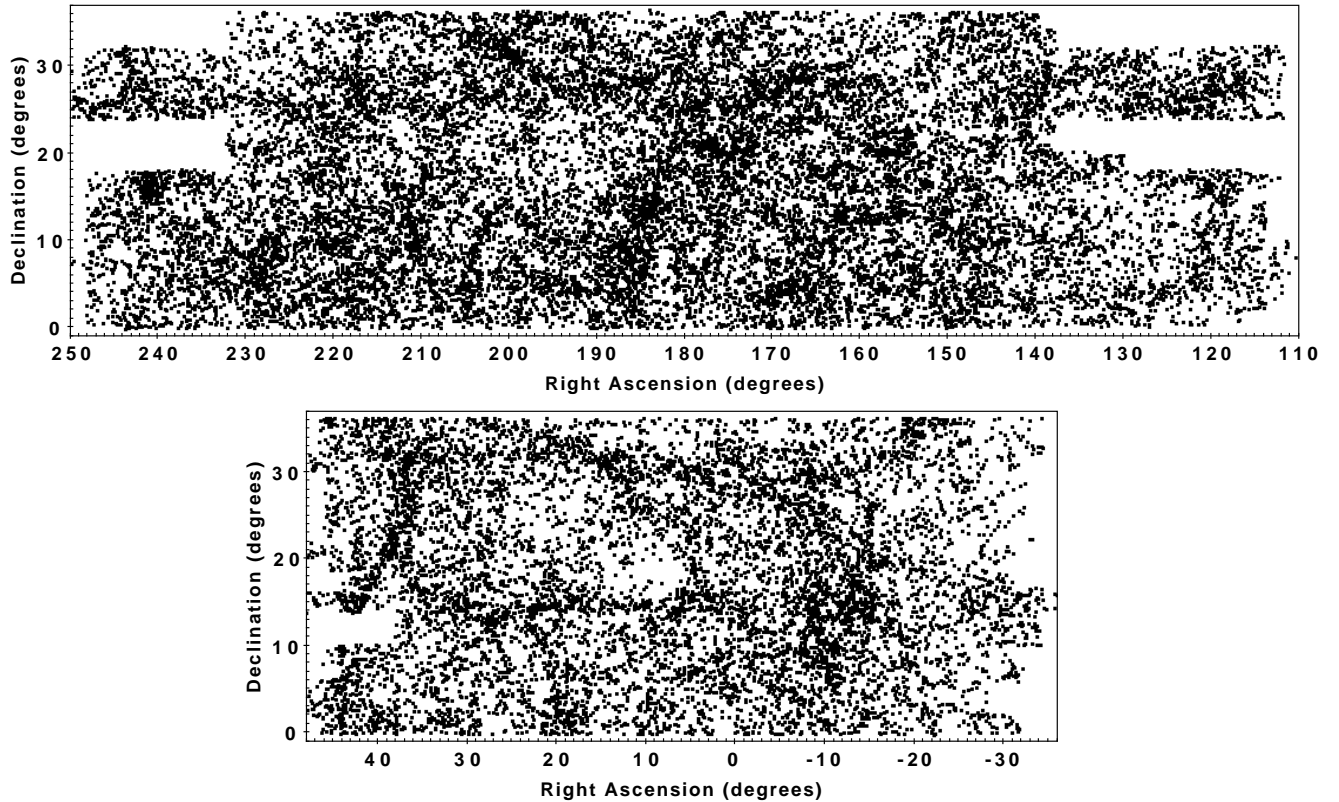


Figure 1. Sky distribution of ALFALFA sources included in Table 2 in the northern (top) and southern Galactic hemispheres (bottom), showing the roughness of boundaries imposed by practical and scheduling constraints.

a distance D_{Mpc} translates to a required integration time t_{int} in seconds

$$t_{int} \propto (T_{sys}/G)^2 M_{HI}^{-2} D_{Mpc}^4 W_{50}^{-2\gamma} \quad (1)$$

where $\gamma \simeq -1/2$ for $W_{50} < 200 \text{ km s}^{-1}$, $\gamma \simeq -1$ for $W_{50} > 200 \text{ km s}^{-1}$ (Giovanelli et al. 2005; Giovanelli & Haynes 2015). The ALFALFA HIMF science goal dictates that the survey cover a very wide solid angle $\Omega_{survey} \sim 7000 \text{ deg}^2$ with an average integration time of ~ 48 seconds per beam solid angle after combination of all drifts from all beams and polarizations across each spatial point. The sheer amount of telescope time (thousands of hours) needed to accomplish such wide sky coverage in turn demanded an observing strategy that exploited Arecibo’s large collecting area, the mapping capability of the ALFA instrument and the spectral power of its backend spectrometer to maximum observing efficiency.

2.1. Drift scan technique

As discussed in detail in Giovanelli et al. (2005), ALFALFA was conducted as a drift scan survey using the 7-feed horn array ALFA. The ALFA feed horn configuration delivers a central, higher gain beam surrounded by a ring of 6 equally-spaced somewhat lower gain beams. For most of the survey, the azimuth arm of the telescope was positioned on the meridian at a pre-assigned J2000.0 declination, with a spacing of $14.6'$ between primary drift centers. The feed array was rotated by 19 degrees so that the Earth-rotation drift-scan tracks of individual beams were equally spaced by $2.1'$ in declination in J2000.0 coordinates. A second, parallel drift pass of the same region of the sky was acquired later, with the center beam offset from the first by $7.3'$ (half the spacing to the next primary center beam positioning) so that the final sampling in declination was $1.05'$. Because hardware limits do not allow pointing straight overhead, coverage of declinations located close to the zenith (Dec. = $+18^\circ 21'$) with similar parallel tracks required the telescope to be positioned off-meridian and the array rotated by a different amount, depending on the declination of the array center. The spectra were acquired covering a 100 MHz bandwidth centered at 1385 MHz using the WAPP (Wide-band Arecibo Pulsar Processor) spectrometer, yielding 4096 “channels” per spectrum, equally spaced in frequency, for each of two linear polarizations of each of the seven feed horns (a total of 14 spectra). Parameters of the ALFALFA observing setup and specifications are summarized in Table 1.

The drift scan observations were conducted in observing runs that typically lasted four to nine hours at a time,

Table 1. ALFALFA Technical Details

Number of beams	7
Polarizations per beam	2 linear
Beam size (FWHM)	$3'.8 \times 3'.3$
Gain	11 K/Jy (central beam) and 8.5 K/Jy (peripheral)
T_{sys}	26-30 Kelvin
Frequency range	1335-1435 MHz
cz_{\odot} range	-2000 to 17912 km s ⁻¹
Bandwidth (total)	100 MHz
Correlator lags (spectral channels)	4096
Channel spacing	24.4 kHz (5.1 km s ⁻¹ at 1420.4058 MHz)
Spectral resolution	10 km s ⁻¹ , after Hanning smoothing
Autocorrelation sampling	3 level
Avg. channel rms	2.0 mJy/channel
Map rms	1.86 mJy/beam
Effective map t_{int}	48 sec (beam solid angle) ⁻¹
5 σ survey sensitivity	0.72 Jy km s ⁻¹ for $W_{50}=200$ km s ⁻¹ at t_{int}
Single drift sky coverage	600 sec \times 14.6' (all beams)
Drift scan size on disk	213 MB
Grid sky coverage	2.4 \times 2.4 degrees
Grid center spacing	8 min in R.A. and 2° in Dec.
Grid cz_{\odot} coverage	a) \sim -2000 to 3300 km s ⁻¹ b) \sim 2500 to 7950 km s ⁻¹ c) \sim 7200 to 12800 km s ⁻¹ d) \sim 12100 to 17912 km s ⁻¹
Grid cz overlap	140 channels
Grid size on disk	381 MB

normally without interruption, yielding an exceptionally high efficiency of “open-shutter time”. Once data acquisition for an observing run began, 14 individual spectra (polarizations/beams) were recorded each second at 99% time efficiency except for two adjustments made every 600 seconds. First, minor pointing corrections were made to maintain the pointing of the ALFA central beam in constant declination J2000.0 coordinates; it may be noted that this approach insured that adjacent or contiguous drift scans taken several years apart would thus remain parallel in that coordinate frame. The corrections from current epoch to J2000 coordinates depend on source position and over the seven-year period of data-taking amounted in some positions to several arcminutes.

In addition to the minor position update, the data acquisition sequence was interrupted every 600 seconds to allow the injection of a calibration noise diode for one second; because of hardware notifications (“hand-shaking”), this procedure, described in more detail in Appendix A, took in practice between four and seven seconds, still less than the time for a source to cross a single ALFA beam (14 seconds). No other adjustments were made. This “minimum intrusion” approach allowed tracking of separate polarization/beam/spectral behavior over the timescale of hours to compensate for systematic variations (e.g., drifts in “electronic gain”). Occasionally, hardware failures led the observing sequence to be aborted. In such cases, power levels were readjusted before data acquisition was restarted. Because of the desire to calibrate using a significant number (at least nine) of calibration samples, drift sequences of less than 90 minutes were discarded. In general, the lack of power readjustment and minimal telescope motion delivered very high overall data quality and robust system calibration.

2.2. Radio frequency interference

A major complication of observing the HI 21 cm line in the 1335-1435 MHz range is introduced by the presence of human-generated radio frequency interference (RFI), typically over relatively narrow ranges in frequency occupying a few MHz or substantially less. Some RFI is predictable, some is (nearly) omnipresent, and some is transient. Most RFI is polarized and some is very strong, causing a rise in the system temperature (T_{sys}) and sometimes introducing spectral standing waves (due to multiple reflections/scattering within the Arecibo telescope optical path). RFI mitigation was addressed in several different ways. To make possible the identification of RFI by statistical differences in power

levels, a second drift across each part of the sky was undertaken, typically with the second pass centered halfway between adjacent tracks of the first pass and acquired 3-9 months later than the first. Since doppler tracking was not implemented, the offset in the time of data acquisition allowed the discrimination of fixed-frequency RFI from cosmic sources. Each spot on the sky was included in multiple drift scans, and beams/polarizations, such that a statistical comparison of subsets of data could be checked for inconsistencies caused by bursts of RFI.

The strongest and most persistent (except for a period of a few months for its replacement) RFI feature arises from the FAA radar at the San Juan airport centered near 1350 MHz. The airport radar transmission is pulsed, polarized, azimuth-dependent and not picked up equally by all beams. When it is strong, harmonics generated within the Arecibo spectral chain may show up at 1380 and (sometimes) 1405 and 1410 MHz. Another common RFI source, evident in shorter bursts of 60-180 seconds at a time, is associated with the Nuclear DETonation (NUDET) detection system aboard the Global Positioning System (GPS) satellites. Many other transient RFI sources were present, arising from spurious transmissions, faulty equipment, etc. In order to address RFI contamination, each individual polarization/beam spectral drift scan was run through an RFI flagging routine and then examined by an expert who could accept or reject the pipelined flags and/or set additional ones. While laborious, this procedure of data flagging produced a spectral mask which maintains a record of flagged spectral pixels, important for identifying RFI “holes” in the 21 cm line sky, as the spectrum at each grid point is associated with a spectral weight at each frequency/velocity point.

Similar to the depictions of the typical spectral weights in previous ALFALFA data release papers, e.g., Figure 1 of [Giovaneli et al. \(2007\)](#) and Figure 6 of [Martin et al. \(2010\)](#), Figure 2 shows the normalized weight per spectral channel derived from the entire set of ALFALFA grids (top panel) and for two different declination strips of grids covering the northern Galactic ALFALFA regions (bottom panel). The most prominent reduced-weight features reflect contamination by the San Juan airport FAA radar near 1350 MHz ($\sim 15600 \text{ km s}^{-1}$) and modulations of it at 1380 MHz ($\sim 8800 \text{ km s}^{-1}$), 1405 MHz ($\sim 3300 \text{ km s}^{-1}$), and 1410 MHz ($\sim 2200 \text{ km s}^{-1}$). As found earlier by [Giovaneli et al. \(2007\)](#), on average, about 85% of the total bandpass was RFI-free with normalized weight > 0.9 . 94% of the bandpass carries spectral weight of > 0.5 ; that value can serve as an acceptable limit on data quality. Because of the large percentage of channels corrupted by the FAA radar systems at frequencies below 1350 MHz, statistical studies requiring volume completeness should be restricted to galaxies within the corresponding velocity limit of $cz < 15000 \text{ km s}^{-1}$ ([Martin et al. 2010](#)).

A more detailed discussion of the drift scan data acquisition, calibration, processing and RFI flagging process is presented in Appendix A.

2.3. Grid production

Upon acquisition of all the drift scans covering a region of sky, all of the relevant spectra were combined to produce a 3-dimensional spectral grid; further details of this process are given in Appendix A.2. In the spatial domain, standard ALFALFA grids are $2.4^\circ \times 2.4^\circ$, evenly sampled at $1'$, so that the spatial dimensions of a grid are 144×144 . Grid centers are pre-determined, separated by 8 minutes in R.A. (e.g., 23^h00^m , 23^h08^m , 23^h16^m , etc) and 2° in Dec. from $+01^\circ$ to $+35^\circ$. In order to keep grid files small enough to be processed and analyzed on typical 2005-era desktops, four separate grids were produced at each grid center covering four separate but partially overlapping frequency ranges corresponding to four velocity ranges: $-2000 < cz < 3300 \text{ km s}^{-1}$, $2500 < cz < 7950 \text{ km s}^{-1}$, $7200 < cz < 12800 \text{ km s}^{-1}$, and $12100 < cz < 17900 \text{ km s}^{-1}$ (see Table 1). The gridding procedure also produces, for every grid point, a record of all of the drift scans, beams and polarizations which contribute to the intensity for each spectral value. While the time for a source to drift across a single ALFA beam is about 14 seconds, the effective integration time after grid construction is typically $t_{int} \simeq 48$ seconds per beam solid angle. It can be less where significant data are missing. In order to track data quality, a normalized weight is recorded for each spectral value.

In addition, the gridding procedure changes the spectral intensities from Kelvins in antenna temperature to mJy in flux density, correcting for zenith angle variations in the gain of the telescope. The flux density scale, set initially by measuring the power injected by the noise diode (see Section 2.1 and Appendix A), is corroborated by comparing the ALFALFA flux densities of unresolved continuum sources in contiguous grids along the same declination strip with the catalogued flux densities at 1400 MHz of the same sources as reported by the NRAO VLA Sky Survey (NVSS; [Condon et al. 1998](#)); such comparison typically involves hundreds of sources over a spring strip of grids centered at the same declination. If the discrepancy in measured fluxes was significant, all fluxes in the involved grids were corrected by a small multiplicative factor to bring them in line with the NVSS values. In no case was the average continuum offset found to be greater than 4% and usually was within 2%.

The combination of drifts taken at different epochs with small variations in calibration, the “blind” baselining done

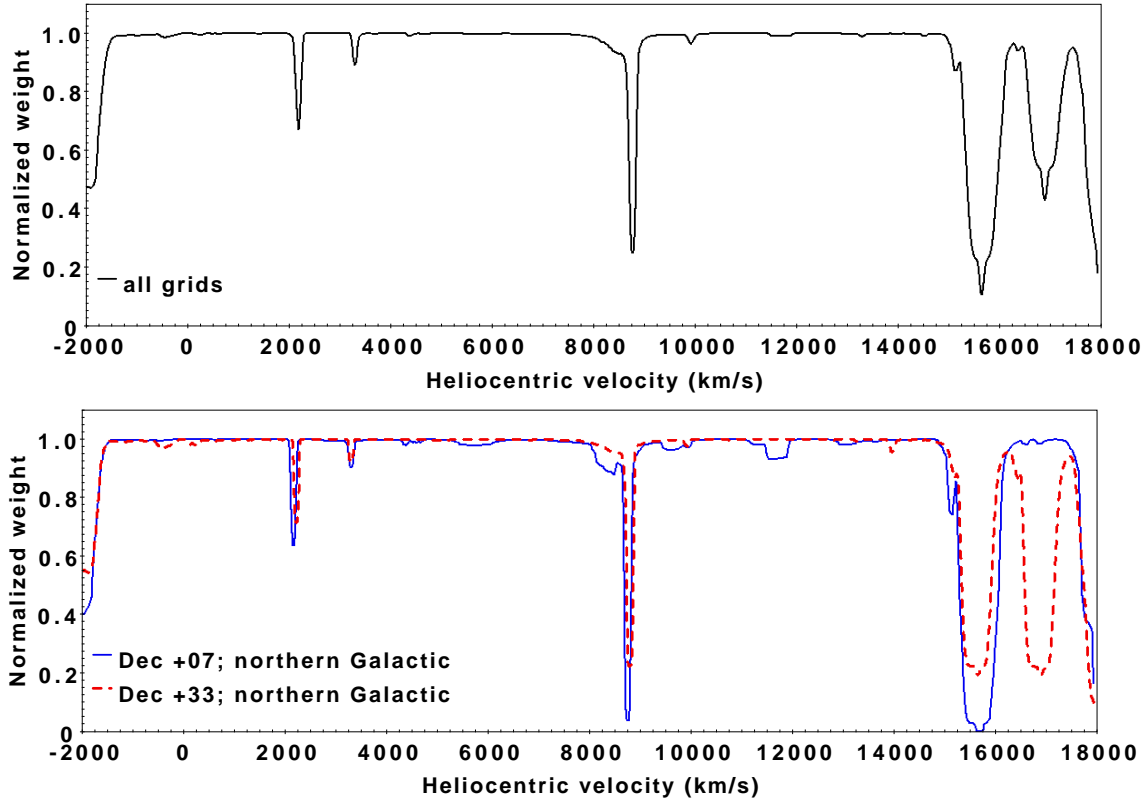


Figure 2. The normalized spectral weight at each of the final (non-overlapping) 3672 frequency channels displayed in the corresponding velocity units for the entire set of ALFALFA grids (upper) and for two strips of grids across the northern Galactic hemisphere (lower): at Dec. = $+07^\circ$ (blue line) and $+33^\circ$ (dashed red line). The main cause of missing data (reduced spectral weight) is RFI, most notably the 1350 MHz ($cz \sim 15600$ km/s) FAA radar at the San Juan airport. Narrower features at 8800, 3300 and 2200 km/s are modulations of the FAA radar occurring within the WAPP spectrometer. Comparison of the panels illustrates the varying nature of the RFI during different observing periods and the serious contamination caused by the 1345-1350 MHz radar system at velocities above 15000 km s^{-1} .

during baseline subtraction and the drift nature of the data acquisition produce various systematic blemishes in the spectral grids. Partial correction of those blemishes is achieved by (a) re-baselining the gridded data along the spectral dimension and (b) performing a similar task in the spatial dimensions, something akin to the flatfielding of optical images. For most grids, these procedures were performed using standard automated routines, discussed further in Appendix A.

2.4. Source identification

Signal extraction is initiated on the fully processed 3-D spectral grids in the Fourier domain with an automated matched filter algorithm described by Saintonge (2007) to produce a list of candidate HI detections. Each grid and its associated candidate catalog are then inspected together in a visualization environment called GridView which allows users to manipulate the spectral grids as well as to overlay datasets from various redshift catalogs and databases such as the Sloan Digital Sky Survey (SDSS; York et al. 2000), SkyView (McGlynn & Scollick 1994), the Second Palomar Observatory Digital Sky Survey (POSS-II; Djorgovski et al. 1998), and the NASA Extragalactic Database (NED). In the vicinity of candidate detections, both polarizations, existing catalog entries and redshifts, and imaging databases are examined closely to identify, where applicable, the most probable optical counterpart (OC), and possibly to corroborate (or reject) the candidate HI line signal. The process of identification of OCs is described in Section 4.1 of Haynes et al. (2011). The close examination of the ALFALFA grids as well as other relevant databases allows the identification with a higher than normal degree of confidence of low-signal-to-noise (SNR) sources which coincide in both space and redshift with likely optical counterparts. These latter sources, designated as the “priors”, are useful because they probe fainter flux levels, as illustrated in Figure 12 of Haynes et al. (2011).

2.5. Parameter measurement

The ALFALFA source catalogs have been produced in a uniform manner to yield a set of consistent parameters for each object within and across grids. Once an HI source is identified, a thumbnail grid is then extracted from a subgrid covering at least $7' \times 7'$ and imported into a measurement environment called GalFlux.

The algorithm used in the GalFlux program to measure the HI line flux, systemic velocity and velocity width is depicted in [Springob et al. \(2005b\)](#) and described by [Haynes et al. \(1999\)](#). For a given integrated spectral profile, peak flux density levels f_p of each spectral horn are selected by eye. A zeroth moment map is created by integrating the flux over these channels with pixel values given by

$$M_0 = \int S_\nu dv \quad [\text{mJy beam}^{-1} \text{ km s}^{-1}] \quad (2)$$

Ellipses are fit to the moment map image at multiple isophotal levels; for the catalogued measurements presented here, the isophotal half-peak intensity level for the moment map is used; see [Section A.2](#) for further details.

The total HI line flux density is derived from the source image, taking into account the telescope beam pattern ([Shostak & Allen 1980](#); [Kent 2011](#)). The source image is spatially integrated over the solid angle covered by the pixels within the chosen isophotal fit. This summed flux density s_ν is divided by the sum of the beam values B sampled at the position of the image pixels, given by

$$S_\nu = \frac{\sum_{x_0} \sum_{y_0} s_\nu(\Delta x, \Delta y)}{\sum_{x_0} \sum_{y_0} B(\Delta x, \Delta y)} \quad [\text{mJy}] \quad (3)$$

The HI line flux density S_{21} is then summed over all velocity channels containing signal and a statistical error is estimated. It is important to note that this HI line flux density measurement technique is optimized for source sizes of approximately the beam area or smaller. [Haynes et al. \(2011\)](#) performed a comparison of the HI flux densities measured by ALFALFA versus those reported by [Springob et al. \(2005a\)](#) and by HIPASS ([Meyer et al. 2004](#)). We have repeated the comparisons for the larger sample reported here with statistically indistinguishable results; analysis of differences is complicated by the various corrections required to account for pointing and source extent in the [Springob et al. \(2005a\)](#) sample and by SNR effects. While ALFALFA is a blind mapping survey and should recover all of the flux for extended sources, the pipeline processing employed in producing the catalog presented here may miss flux from the most extended or highly asymmetric sources. The catalogued HI line flux densities given in [Table 2](#) are derived over the area encompassed by the isophote at 50% of the peak power. This isophote is typically comparable to or larger than the beam (at the full width at half maximum power: FWHM) while the vast majority of sources are unresolved. The integrated flux densities for very extended sources or with significant angular asymmetries can be misestimated by the pipeline algorithm. A special catalog with parameters of extended sources is in the process of construction ([Hoffman et al.](#) in prep).

The global HI line velocity and velocity width provide measures of the galaxy's systemic recessional velocity and projected disk rotational velocity. The optimal definition of particularly the width measurement depends in part on the science objective (redshift, measures of rotational velocity) and data quality factors such as the impact of turbulence and SNR (e.g. [Schneider et al. 1986](#); [Bicay & Giovanelli 1986](#); [Chengalur et al. 1993](#); [Catinella et al. 2007](#)). The algorithm used here is nearly identical to that presented by [Chengalur et al. \(1993\)](#) and [Springob et al. \(2005b\)](#). To derive a measure of the profile width, polynomials (usually lines) are fit to the channels at both edges of the emission between 15% and 85% of the peak flux. The velocity width W_{50} is then defined as the difference between the velocities corresponding to the fitted polynomial at a level of 50% of the maximum value of flux on each horn. In practice, the maximum flux value is adopted as the observed peak flux minus the rms, f_p -rms, in order to correct for the contribution of noise. The average of the two velocities at 50% of f_p -rms is taken as the systemic velocity V_{50} . Similar measurements are also made at the 20% level for the velocity and associated width.

In addition to statistical uncertainties in the integrated HI flux density, velocity and velocity width, a subjective estimate of an additional systematic uncertainty is obtained by examination of the minimum and maximum extents of the HI emission signal (see [A.2](#)). In most cases, the statistical uncertainty is larger than the systematic one and the latter can be ignored. However, in cases of low SNR, very narrow velocity width and/or shallow outer profile slope, the statistical error is clearly too small and thus the adopted uncertainty is the sum in quadrature of the two. Velocity width measurements play an important part in determining the signal-to-noise ratio (SNR) of a spectrum, especially for sources with $W_{50} > 400 \text{ km s}^{-1}$ ([Haynes et al. 2011](#)).

Velocities and velocity widths, centroid sky positions from the isophotal ellipse fits, SNR, etc., are all measured and stored in individual source files with the spectral profiles. Optical identification is also made based on previous HI observations ([Springob et al. 2005b](#)) and visual inspection of the imaging databases from the SDSS, POSS-II, and Two

Micron All Sky Survey (2MASS; [Skrutskie et al. 2006](#)). Each source file can be reviewed in a catalog viewing program with a complete history of how the source was measured and the kinematic parameters derived. This tool, called GalCat (for Galaxy Catalog), connects to sites that use uniform International Virtual Observatory Alliance (IVOA) web service standard protocols ([Graham et al. 2007](#)).

Previous ALFALFA data releases have included a category of objects without OCs which lie within the velocity regime associated with the Galactic HI emission including the well-known HI high velocity clouds (HVCs). HVCs are often very extended, exceeding in size that of a single ALFALFA grid. For such objects, ALFALFA may identify multiple sources associated with a single cloud or complex, and only the flux in the knots of emission will be measured. A very few of the most compact HVCs dubbed the ‘‘Ultra-Compact’’ HVCs (UCHVCs) prove to be very nearby galaxies in or near the Local Group ([Giovanelli et al. 2010, 2013](#); [Adams et al. 2015a](#); [Janesh et al. 2015, 2017](#)), but the majority are likely associated with Galactic phenomena. Because ALFALFA was not optimized to probe the velocity range associated with Galactic HI, the HVCs identified by the standard ALFALFA pipeline must be interpreted with understanding of its limitations ([Adams et al. 2013](#); [Bianchi et al. 2017](#)). For this reason, they are not included here.

3. THE ALFALFA EXTRAGALACTIC SOURCE CATALOG

3.1. *Extragalactic HI Sources*

Table 2 presents the principal results of the ALFALFA extragalactic HI survey. The contents of Table 2 are largely similar to those presented in previous ALFALFA data release catalogs and are as follows:

- Col. 1: Entry number in the Arecibo General Catalog (AGC), a private database of extragalactic objects maintained by M.P.H. and R.G. The AGC entry corresponds to both the HI line source and the OC where one is assigned. In the absence of a feasible OC, the AGC number corresponds only to the HI detection. An AGC number is assigned to all ALFALFA sources; it is intended to be used as the basic cross reference for identifying and tracking ALFALFA sources as new data acquired in overlapping regions has superseded older results. Note that in previous ALFALFA catalogs, an index number was used, a practice no longer employed. The designation of an ALFALFA source referring only to its HI emission (without regard to its OC) should be given using the prefix ‘‘HI’’ followed by the position of the HI centroid as given in Col. 3 of Table 2.
- Col. 2: Common name of the associated OC, where applicable. Further discussion of the process of assigning optical counterparts has been discussed in Section 4.1 of [Haynes et al. \(2011\)](#).
- Col. 3: Centroid (J2000) of the HI line source, in hhhmss.sSddmmss, after correction for systematic telescope pointing errors, which are on the order of $20''$ and depend on declination. The systematic pointing corrections are derived from an astrometric solution for the NVSS radio continuum sources ([Condon et al. 1998](#)) found in the grids. As discussed in [Giovanelli et al. \(2007\)](#) and [Kent et al. \(2008\)](#), the assessment of centroiding errors is complicated by the nature of 3-D grid construction from the 2-D drift scans that were often acquired in widely separated observing runs, and, for resolved/confused sources of unknown source structure.
- Col. 4: Centroid (J2000) of the most probable OC, in hhhmss.sSddmmss, associated with the HI line source, where applicable. The OC has been identified and its likelihood assessed following the process discussed in Section 4.1 of [Haynes et al. \(2011\)](#). The median positional offset of the OC from the HI centroid is about $18''$ and depends on SNR following Equation 1 of [Haynes et al. \(2011\)](#). In rare low SNR instances, it can exceed $1'$. It should be noted that only one OC is assigned per HI source although, in some cases, confusion caused by multiple sources (either HI or OCs) within the telescope beam is a possibility.
- Col. 5: Heliocentric velocity of the HI line source, cz_{\odot} in km s^{-1} in the observed frame, measured as the midpoint between the channels at which the line flux density drops to 50% of each of the two peaks on the low and high velocity horns of the profile (or of one, if only one is present); see also [Springob et al. \(2005b\)](#). Values adopt the optical convention $\delta\lambda/\lambda$, not the ‘‘radio’’ one ($\delta\nu/\nu$). The statistical uncertainty on cz_{\odot} to be adopted is half the error on the width W_{50} tabulated in Col. 6.
- Col. 6: Velocity width of the HI line profile, W_{50} in km s^{-1} , measured at the 50% level of each of the two peaks, as described in Col. 5 and corrected for instrumental broadening following Equation 1 of [Springob et al. \(2005a\)](#). No corrections due to turbulent motions, disk inclination or cosmological effects are applied. The estimated uncertainty on W_{50} , σ_W , in km s^{-1} , follows in parentheses. This error is the sum in quadrature of two components: a statistical error dependent on the SNR of the HI signal, and a systematic error associated with

the user’s confidence in the definition of spectral boundaries of the signal and the applied baseline fit; see 2.5 and A.2. In the majority of cases, the systematic error is smaller than the statistical error and can be ignored.

- Col. 7: Velocity width of the HI line profile, W_{20} in km s^{-1} , similar to W_{50} , but measured at the 20% level of each of the two peaks. Note that the algorithm used to determine widths by fitting a polynomial to each horn between 15% and 85% of the peak flux is optimized to measure W_{50} and σ_W at 50% of the peak, not at the lower value of 20%. While some authors prefer to use the lower value, we find that it is less robust and its error harder to quantify, particularly at lower SNRs.
- Col. 8: Integrated HI line flux density of the source, S_{21} , in Jy km s^{-1} . The estimated uncertainty of the integrated flux density $\sigma_{S_{21}}$, in Jy km s^{-1} , is given in parentheses and has been derived following the same procedure as used to measure the uncertainty in the velocity width W_{50} (Col. 6). As discussed in Section 2.5 and the Appendix A.2, line flux density values included in Table 2 have been extracted from the spatial integration of the 3-D grid over a window of at least $7' \times 7'$ and corrected for the survey beam over the same area. The algorithm used may underestimate the flux of very extended and/or asymmetric sources.
- Col. 9: Signal-to-noise ratio SNR of the detection, estimated as

$$SNR = \left(\frac{1000S_{21}}{W_{50}} \right) \frac{w_{smo}^{1/2}}{\sigma_{rms}} \quad (4)$$

where S_{21} is the integrated flux density in Jy km s^{-1} , as listed in Col. 8. The ratio $1000S_{21}/W_{50}$ is the mean flux density across the feature in mJy . In this definition of SNR, w_{smo} is a smoothing width expressed as the number of spectral resolution bins of 10 km s^{-1} bridging half of the signal width and σ_{rms} is the rms noise figure across the spectrum measured in mJy at 10 km s^{-1} resolution, as tabulated in Col. 10. The ALFALFA raw spectra are sampled at $24.4 \text{ kHz} \sim 5.5 \text{ km s}^{-1}$ at $z \sim 0$, and, as in previous ALFALFA data releases (e.g. Giovanelli et al. 2007), w_{smo} is adopted as either $W_{50}/(2 \times 10)$ for $W_{50} < 400 \text{ km s}^{-1}$ or $400/(2 \times 10) = 20$ for $W_{50} \geq 400 \text{ km s}^{-1}$.

- Col. 10: Noise figure of the spatially integrated spectral profile, σ_{rms} , in mJy . The noise figure as tabulated is the rms as measured over the signal- and RFI-free portions of the spectrum, after Hanning smoothing to a spectral resolution of 10 km s^{-1} .
- Col. 11: Adopted distance D_H and its uncertainty σ_D , both in Mpc . For objects with $cz_{\odot} > 6000 \text{ km s}^{-1}$, the distance is simply estimated as cz_{cmb}/H_0 where cz_{cmb} is the recessional velocity measured in the Cosmic Microwave Background reference frame (Lineweaver et al. 1996) and H_0 is the Hubble constant, adopted to be $70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. For objects with $cz_{cmb} < 6000 \text{ km s}^{-1}$, we use the local peculiar velocity model of Masters (2005), which is based in large part on the SFI++ catalog of galaxies (Springob et al. 2007) and results from analysis of the peculiar motions of galaxies, groups, and clusters, using a combination of primary distances from the literature and secondary distances from the Tully-Fisher relation. The resulting model includes two attractors, with infall onto the Virgo Cluster and the Hydra-Centaurus Supercluster, as well as a quadrupole and a dipole component. The transition from one distance estimation method to the other is selected to be at $cz_{\odot} = 6000 \text{ km s}^{-1}$ because the uncertainties in each method become comparable at that distance. Where available, primary distances in the published literature are adopted; we also use secondary distances, mainly from Tully et al. (2013), for galaxies with $cz_{cmb} < 6000 \text{ km s}^{-1}$. When the galaxy is a known member of a group (Springob et al. 2007), the group systemic recessional velocity cz_{cmb} is used to determine the distance estimate according to the general prescription just described. Where primary distances are not available, objects in the Virgo region are assigned to likely Virgo substructures (Mei et al. 2007) and then the distances to those subclusters are adopted following Hallenbeck et al. (2012). Errors on the distance are generated by running 1000 Monte Carlo iterations where peculiar velocities are drawn from a normal distribution each time as described in Sec. 4.1 of Jones et al. (2018). Such errors are likely underestimates in the vicinity of major attractors such as Virgo. It should be noted that the values quoted here are Hubble distances, not co-moving or luminosity distances.
- Col. 12: Logarithm of the HI mass M_{HI} , in solar units, computed via the standard formula $M_{HI} = 2.356 \times 10^5 D_H^2 S_{21}$ and assuming the distance given in Col. 11 (not the luminosity distance). The uncertainty $\sigma_{\log M_{HI}}$ is derived, following Jones et al. (2018), by combining the uncertainty in the integrated HI line flux and the

distance with a minimum of 10% uncertainty. The latter minimum is set to prevent the error from ever getting unrealistically small and to include the systematic uncertainty in the flux calibration. The uncertainty in $\log M_{HI}$ then is

$$\sigma_{\log M_{HI}} = \frac{\sqrt{\left(\frac{\sigma_{S_{21}}}{S_{21}}\right)^2 + \left(\frac{2\sigma_D}{D}\right)^2 + 0.1^2}}{\ln 10} \quad (5)$$

It should be noted that the HI mass values given here do not include a correction for HI self-absorption.

- Col. 13: The HI source detection category code, used to distinguish the high SNR sources from the lower ones associated with OCs of comparable redshift. Code 1 refers to the 25434 sources of highest quality. Quality is assessed based on several indicators: there is a good match in signal characteristics match between the two independent polarizations observed by ALFALFA, a spatial extent consistent with the telescope beam (or larger), an RFI-free spectral profile, and an approximate minimum SNR threshold of 6.5 (Saintonge 2007). These exclusion criteria lead to the rejection of some candidate detections with $\text{SNR} > 6.5$; likewise, some features with SNR slightly below this soft threshold are included because of optimal overall characteristics of the feature, such as well-defined spatial extent, broad velocity width, and obvious association with an OC. We estimate that the detections with code 1 in Table 2 and associated with an OC are nearly 100% reliable; the completeness and reliability of the $\alpha.40$ catalog are discussed in Section 6 of Haynes et al. (2011).

Code 2 refers to the 6068 sources categorized as “priors”. They are sources of low SNR ($\lesssim 6.5$), which would ordinarily not be considered reliable detections by the criteria set for code 1, but which have been matched with OCs with known optical redshifts coincident (to within their uncertainties) with those measured in the HI line. We include them in our catalog because they are very likely to be real. As defined, the classification of a source as a prior depends not just on the redshift match with a likely counterpart but also on the profile shape, polarization match and location relative to RFI so that the SNR as defined is not a good statistical indicator of their reliability. In fact, 15 of the priors have $\text{SNR} < 3$. In general, the priors should not be used in statistical studies which require well-defined completeness limits; this point is further discussed in Section 6 of Haynes et al. (2011). Because of the substantially more complete SDSS spectroscopic coverage of the ALFALFA region in the northern Galactic hemisphere than the southern, the number of priors is substantially higher in the former than the latter.

It should be noted that objects without optical counterparts and cz in the range of Galactic HI emission, including the high velocity clouds, are not included in Table 2; in previous ALFALFA data releases, such sources have been included and identified as HI source category “9” sources. They are discussed in Giovanelli et al. (2010); Adams et al. (2013) and a complete catalog will be presented elsewhere.

3.2. HI Spectra of ALFALFA Extragalactic HI Sources

In addition to the measured and derived properties presented in Table 2, a final representative 2-D HI line spectrum for each source has been extracted over a window in each grid outlined by the isophote at half of peak intensity. Spectral files are provided both in ASCII and FITS formats and include values for each spectral channel of the frequency, heliocentric velocity (cz), flux density, the value of the subtracted polynomial baseline, and the normalized weight. We emphasize that the algorithm used to derive the total HI line flux presented in Table 2 integrates the source image over the solid angle covered by pixels contained within half-peak isophote and then applies a correction for the beam pattern as indicated in Equation 3. Caution should be exercised for those channels for which the normalized weight is low, e.g. < 0.5 . These spectra will be available at <http://egg.astro.cornell.edu/alfalfa/data/> and through the NASA Extragalactic Database (NED: <https://ned.ipac.caltech.edu/>).

3.3. OHM Megamasers and Candidate OHMs

As pointed out by Briggs (1998), OH megamasers (OHMs) redshifted from 18 cm rest wavelength into the targeted HI bandpass can be a source of contamination in blind HI line surveys of the local Universe. For ALFALFA, the relevant OHM redshift range is $0.167 < z < 0.244$. Several OHM candidates were presented in Haynes et al. (2011). In a few cases, the ALFALFA signals have been recognized as matching the OH emission of previously-known OHMs discovered by Darling & Giovanelli (2006). Suess et al. (2016) conducted a concerted program of optical spectroscopy to assess the likelihood of OHM interlopers. Here we present in Table 3, the small number of HI sources which have been flagged as possible interloping OH megamasers. 19 sources have been thus identified, nine of which coincide

Table 2. ALFALFA Extragalactic HI Source Catalog

AGC ID	Name	HI position J2000	OC position J2000	cz_{\odot} km s $^{-1}$	W_{50} km s $^{-1}$	W_{20} km s $^{-1}$	$f SdV$ Jy-km s $^{-1}$	SNR	rms mJy	D_H Mpc	$\log M_{HI}$ $\log M_{\odot}$	Code
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
105367		000000.4+052636	000000.8+052633	11983	274(39)	281	1.14(0.08)	8.1	1.91	166.0(2.3)	9.87(0.05)	1
333313		000000.9+245432	235959.4+245427	11181	313(20)	333	1.80(0.09)	11.3	2.02	154.8(2.3)	10.01(0.05)	1
331060	478-009b	000002.5+230505	000003.4+230515	4463	160(4)	184	1.96(0.07)	14.7	2.35	50.6(10.4)	9.07(0.18)	1
331061	456-013	000002.5+155220	000002.1+155254	6007	260(45)	268	1.13(0.09)	6.5	2.40	85.2(2.4)	9.29(0.06)	1
104570		000001.6+324230	000001.2+324237	10614	245(6)	250	0.86(0.07)	6.6	1.86	147.0(2.3)	9.64(0.06)	1
331405		000003.3+260059	000003.5+260050	10409	315(8)	345	2.62(0.09)	16.1	2.05	143.8(2.2)	10.11(0.05)	1
102896		000006.8+281207	000006.0+281207	16254	406(17)	433	2.37(0.12)	11.2	2.31	227.4(2.2)	10.46(0.05)	1
630358	382-015	000007.5-000249	000007.8-000226	7089	70(9)	103	2.47(0.06)	29.7	2.20	96.2(2.3)	9.73(0.05)	1
105368		000010.9+041654	000011.7+041637	3845	83(6)	94	0.72(0.06)	7.5	2.33	54.2(2.2)	8.70(0.07)	1
331066	382-016	000011.5+010723	000012.7+010712	7370	214(22)	299	2.30(0.11)	13.3	2.64	100.2(2.2)	9.74(0.05)	1
102571		000017.2+272359	000017.3+272403	4654	104(3)	124	2.00(0.06)	19.0	2.29	65.9(2.1)	9.31(0.05)	1
102728		000021.2+310038	000021.4+310119	566	21(6)	36	0.31(0.03)	7.5	1.92	9.1(2.2)	6.78(0.22)	1
331067	517-010	000020.0+343641	000022.2+343658	12687	104(13)	149	0.99(0.08)	7.8	2.75	176.7(2.3)	9.86(0.06)	1
104678		000022.4+204808	000022.3+204748	6852	190(9)	220	2.49(0.11)	13.6	2.95	92.9(2.2)	9.70(0.05)	1
105370		000027.7+053256	000029.6+053323	13133	245(6)	252	1.04(0.09)	6.6	2.26	182.5(2.3)	9.91(0.06)	1
12893	456-014	000028.0+171315	000028.1+171309	1105	71(2)	85	2.30(0.05)	29.2	2.07	12.8(4.4)	7.95(0.30)	1
12896	478-010	000030.1+261928	000031.4+261931	7653	170(10)	217	3.14(0.08)	22.0	2.44	104.5(2.3)	9.91(0.05)	1
102729		000032.1+305152	000032.0+305209	4618	53(6)	71	0.70(0.04)	10.5	2.02	65.4(2.1)	8.85(0.06)	1
331070		000033.8+224645	000033.6+224642	11715	80(3)	100	1.49(0.06)	16.4	2.26	162.4(2.3)	9.97(0.05)	1
12895		000039.6+200333	000038.3+200332	6746	162(2)	181	3.82(0.07)	29.9	2.23	91.3(2.3)	9.88(0.05)	1

Table 2 is published in its entirety in machine-readable format. A portion is shown here for guidance regarding its form and content.

with likely OC’s which have known optical redshifts in the appropriate redshift; the remaining ten sources have not been confirmed but should be considered as candidate OHMs. All 19 sources are listed in Table 3, but separated into confirmed (top) and candidates (bottom). Because of the uncertainties, we give only basic parameters for these sources and reiterate that the final ten OHM candidates need further confirmation. The frequency of the center of the line signal f_{sys} , in MHz, is included along with its approximate velocity width, line flux density, SNR and rms. Furthermore, there may be additional OHMs lurking but not yet identified in the ALFALFA catalog, consistent with the estimate of 9_{-6}^{+73} found by [Suess et al. \(2016\)](#).

4. SUMMARY AND CAVEATS

This paper presents the catalog of extragalactic HI line sources from the completed ALFALFA HI line survey. Previous papers, notably [Giovanelli et al. \(2005\)](#); [Saintonge \(2007\)](#); [Giovanelli et al. \(2007\)](#); [Kent et al. \(2008\)](#); [Haynes et al. \(2011\)](#), have presented further details on the survey design, observing strategy, signal extraction technique, survey sensitivity and completeness. Although the minimum-intrusion drift-scan technique attempts to minimize the impact of the complex optics of the Arecibo telescope and the realities of the terrestrial environment at L-band, the source catalog presented in Table 2 should be used with appreciation of numerous caveats. Here we list a few:

- **HI-selection:** The population of galaxies detected by emission of the HI 21cm line is dominated by relatively low luminosity, star-forming galaxies. In fact, virtually all star-forming galaxies contain a cool neutral component of their interstellar medium. Therefore the galaxies detected by ALFALFA are preferentially bluer and have lower surface brightness, lower luminosity and lower metallicity than comparable populations detected by their optical broadband flux.
- **Completeness:** Although the ALFALFA source population is statistically well behaved as illustrated by Figure 11 of [Haynes et al. \(2011\)](#), the survey is flux limited in a manner that depends on the velocity width, e.g. Figure 12 of [Haynes et al. \(2011\)](#). Hence completeness corrections need to be carefully considered for any statistical analysis for which they are important (e.g., gas fraction scaling relations).
- **Cosmological corrections:** In order to allow immediate comparison with the vast majority of extant literature and HI line data compilations, we have elected not to apply cosmological corrections to the values reported in Table 2. As a result, velocities are presented in the “observed” rest frame, simply as cz , the HI line flux densities are given in the commonly adopted hybrid units of Jy-km s^{-1} (as opposed to units of Jy-Hz), and distances are Hubble distances D_H . Careful discussions of the nature and impact of cosmological corrections are presented by [Hogg \(1999\)](#) and, of particular relevance to HI studies, [Meyer et al. \(2017\)](#). At large distances, the “true” HI mass would be derived use the standard equation (e.g. equation 1 of [Giovanelli et al. 2005](#)) adopting the luminosity distance D_L and dividing by a factor of $(1+z)^2$ to account for the fact that the HI flux in observed rest frame units is an overestimate (e.g. equation 46 of [Meyer et al. 2017](#)). Because $D_L = (1+z) D_C$ where D_C is the co-moving distance, substitution of D_C into the standard equation leads to the factors of $(1+z)$ cancelling out, so that the HI mass can be derived by the standard equation without any z terms but with the co-moving distance (and the integrated flux in the observed rest frame, as is typical for extant HI surveys) instead of the Hubble distance. At the outer edge of ALFALFA, $z = 0.06$, the difference between D_C and D_H is less than 2% so that the maximum systematic error in the HI mass due to using the Hubble distance is $\sim 3\%$. Future surveys that explore the Universe beyond that probed by ALFALFA should be careful to follow the detailed discussion presented in [Meyer et al. \(2017\)](#).
- **Integrated HI line fluxes:** As discussed in Section 2.5 and noted in the description of column 8 of Table 2, the HI line flux densities reported in the catalog here have been extracted from the spatial integration over a window of at least $7' \times 7'$, with an applied correction factor which models the beam response pattern over the same area. For this reason, the fluxes derived from the 3-D grids may underestimate the fluxes of very extended or highly asymmetric sources. Fluxes should match best when the HI extent is smaller than or comparable to a single ALFA beam. A special catalog with parameters of extended sources is in the process of construction ([Hoffman et al.](#) in prep).
- **Matching with other databases by position:** When performing automated matches to other catalogs, we strongly advise the use of the OC positions where given. The HI centroid positions are on average good to only $\sim 20''$ and their accuracy depends on SNR. For low SNR sources, offsets can exceed $1'$. If the HI position is used in an automated matching, many valid matches may be missed or false ones found. At the same time, we admit

Table 3. ALFALFA OH Megamaser Candidate Catalog

AGC ID	Name	HI position J2000	OC position J2000	f_{cent} MHz	W_{50} km s ⁻¹	$\int SdV$ Jy-km/s	SNR	rms mJy	z
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
114529	SDSS J015001.57+240235.8	015001.9+240223	015001.6+240236	1384.12	613	2.88	9.7	2.17	0.204368 ^a
121379	IRAS 02524+2046	025517.8+205918	025517.1+205857	1412.20	85	1.95	17.4	2.70	0.181402 ^b
181310	SDSS J082312.61+275139.8	082311.7+275157	082312.7+275138	1427.62	46	2.17	15.9	2.18	0.167830 ^{a,b}
219215	SDSS J111125.06+052045.9	111126.0+052044	111125.1+052046	1360.73	45	0.96	16.6	1.90	0.225213 ^a
219828	2MASX J11551476+3130026	115518.6+312933	115514.7+313003	1376.31	216	0.79	4.3	2.77	0.215989 ^c
229493	IRAS F12072+3054	120949.3+303812	120948.3+303750	1428.35	123	1.10	9.4	2.37	0.170000 ^d
229487	SDSS J120948.28+303749.5	121548.9+351149	121548.8+351100	1415.83	35	0.49	6.5	2.80	0.166056 ^c
257959	SDSS J155537.94+143905.6	155537.7+143906	155537.9+143906	1386.22	206	2.65	17.0	2.42	0.203568 ^a
333320	IRAS F23129+2548	231520.8+260508	231521.4+260432	1414.25	509	1.78	7.8	2.00	0.178913 ^b
102708	GALEXASC J000335.98+253204.4	000337.0+253215	000336.1+253204	1426.76	234	0.91	5.7	2.33	^e
102850	2MASX J00295817+3058322	002958.8+305739	002958.2+305832	1423.24	53	0.46	6.7	2.09	^e
114732	GALEXASC J010107.09+094624.0	010110.7+094626	010107.1+094621	1425.72	23	0.42	7.4	2.53	^e
115713	GALEXASC J014135.31+165731.5	014134.1+165718	014135.2+165731	1424.82	133	1.36	11.6	2.27	^e
115018	GALEXASC J015847.27+073204.2	015847.1+073159	015847.2+073202	1387.23	157	0.84	6.5	2.31	^e
124351	GALEXASC J021750.80+072429.3	021751.0+072447	021750.9+072428	1422.98	87	0.63	5.6	2.65	^e
749309	GALEXASC J101101.08+274012.9	101102.9+274020	101101.1+274012	1400.18	80	0.72	8.1	2.20	^e
219835		113034.2+322208	113034.2+322208	1390.52	155	0.85	6.4	2.36	^e
249507	IRAS F14014+3009	140341.6+295500	140340.3+295456	1414.88	216	1.65	12.3	2.04	^e
322231		223605.9+095743	223605.4+095726	1429.27	40	0.55	7.9	2.41	^e

^aOHM confirmed by (Stuess et al. 2016).^bOHM included in catalog of Darling & Giovanelli (2006).^cOC has coincident redshift in SDSS.^dOHM discovered by Morganti et al. (2006)^eOC has no reported redshift; OH detection and/or redshift requires confirmation.

that some of the OC assignments are somewhat subjective, for example, situations where two likely OCs fall within the beam; some of the assignments made here are certainly incorrect. As further information, particularly redshifts, becomes available, we invite comment on the current database and would plan to provide updated (and improved) versions of the catalog presented here.

- **HI column densities:** The majority of sources detected by ALFALFA have angular sizes much smaller than the telescope beam. While the line flux integral yields an accurate measure of the HI mass, information on the spatial extent and morphology of the HI distribution and velocity field is not present. Only in cases of very extended sources can HI column densities derived from ALFALFA be meaningful. In the absence of source resolution, measures of the HI column density reflect the *lower limit* on HI column density *averaged* over the areal extent subtended by the beam. Much higher column density knots could easily be present.
- **Assessing the HI extent and distribution:** Likewise, except for sources of angular extent larger than $\sim 5'$, follow-up HI synthesis imaging is necessary to obtain direct information on the HI column density distribution to measure the morphology and inclination of the HI layer, and to derive dynamical parameters such as the mass contained within the HI radius or the shape of the rotation curve. However, it should be noted that synthesis observations can resolve out the diffuse, low surface density gas, thus missing flux from extended distributions that exceed the scale of the shortest interferometer baseline used for such studies.
- **HI self-absorption:** Catalogued ALFALFA fluxes include no correction for HI self-absorption. Users may wish to implement their own corrections. [Giovanelli et al. \(1994\)](#) and [Jones et al. \(2018\)](#) both find that the correction for a typical L* galaxy is likely quite small, $\sim 10\%$, but might be as high as $\sim 30\%$ for edge-on galaxies.
- **Global velocity width measures:** In Table 2, we present HI line widths measured by fitting a polynomial on both horns of the profile between 15-85% of the peak flux on either side. The catalogued values W_{50} and W_{20} give the full widths at 50% and 20% of the peak flux as measured between the polynomial fits on either side. As has been shown by [Bicay & Giovanelli \(1986\)](#), the value of W_{50} is shown to be more robust particularly at lower values of SNR but other approaches may also be valid. The relationship of the global HI width to rotational velocity measures obtained from stellar absorption lines or nebular emission lines is complicated and may depend on galaxy properties such as surface brightness (e.g. [Catinella et al. 2007](#)).
- **Corrections to observed widths:** The values of W_{50} presented here have been corrected for instrumental broadening but not for other factors such as turbulence ([Fouque et al. 1990](#)) and cosmological stretch. Furthermore, they reflect only the projected component of the HI layer's rotational velocity.
- **Distances:** While the distance estimation routine represents the best available information regarding the distance to each source (that we are aware of), the reality is that it employs a highly inhomogeneous collection of primary and secondary distances, and group assignments. Furthermore, the flow model itself can be double or triple valued in the vicinity of very dense structures. Thus, the distance estimates should be considered with caution. [Jones et al. \(2018\)](#) find that for calculations with the complete catalog the distance estimates are unlikely to result in significant uncertainty or bias, but this may not be true for any individual object.
- **HI masses:** Under the assumption that the HI is optically thin, we derive the HI masses simply from the integrated HI line fluxes and the distances. As discussed in Section 3, the error on $\log M_{HI}$ is estimated as the combination of the uncertainties in the S_{21} and the distance with an additional allowance for systematic uncertainty in the flux calibration. As mentioned above, we have applied no correction for HI self-absorption, and the distances used are Hubble distances. For the most distant ALFALFA sources, as noted above, the latter effect will introduce a bias of similar scale to either the self-absorption or the flux calibration. The impact of the uncertainties in HI masses is discussed in [Jones et al. \(2018\)](#).
- **Impact of terrestrial interference:** RFI impacts some portions of the ALFALFA spectrum severely. The spectral data products for each HI source contain a column with the normalized weight at each individual frequency/velocity channel; low weight channels should be treated with caution. Largely because of contamination by the San Juan airport FAA radar at 1345-1350 MHz (see Figure 2), statistical studies requiring a high degree of volume completeness should be restricted to galaxies with $cz_{cmb} < 15000 \text{ km s}^{-1}$.

- Spectral stacking: The 3-D HI grids used by ALFALFA have demonstrated the power of spectral stacking (Fabello et al. 2011a,b, 2012; Hallenbeck et al. 2012; Brown et al. 2015; Odekon et al. 2016; Brown et al. 2017) to sample selected galaxy populations to low HI mass levels. An important part of the spectral stacking using ALFALFA has been the retention of the “weights” record of missing data and RFI flagging.
- Confusion: As discussed by Giovanelli & Haynes (2015) and Jones et al. (2016a), single dish surveys are most efficient in delivering large statistical samples for which resolution is not required over volumes where the average separation between galaxies is larger than the beam area. At even modest distances, the impact of confusion within the beam needs to be carefully considered. The impact of confusion on the ALFALFA catalog presented in Table 2 is estimated to be relatively minor (Jones et al. 2016a), although individual cases of confused sources are not hard to find.

Statistical studies of the HI-bearing galaxy population still sample relatively small numbers of galaxies in comparison with spectral surveys at optical wavelengths. When it was initiated in 2005, ALFALFA followed on the heels of the HI Parkes Sky Survey (HIPASS; Barnes et al. 2001), the first blind HI survey to cover a large volume. The much larger collecting area of the Arecibo dish offered improvements in sensitivity and angular resolution, advances in spectrometer capability allowed an increase in spectral bandwidth and resolution, and the adopted minimum-intrusion drift scan technique conducted during nighttime only delivered very high data quality. As a result, the source density of ALFALFA (~ 5 sources per square degree) is more than $25\times$ higher than that of HIPASS. The principal aim of ALFALFA has been to survey a wide area of the extragalactic sky over a cosmologically significant but local volume. Future surveys with single dish telescopes should focus on deeper surveys of the local Universe or intensity mapping applications which can actually benefit from lack of resolution. On-going and planned surveys with interferometric arrays will continue to sample the extragalactic HI sky, offering increased resolution to map the HI distribution and velocity field and extending beyond the local Universe (Giovanelli & Haynes 2015). While this paper presents the ALFALFA harvest, we are confident that the scientific seeds from ALFALFA promise a future yield that is even more bountiful.

Acknowledgments. The authors acknowledge the work of the entire ALFALFA collaboration who have contributed to the many aspects of the survey over the years. The ALFALFA team at Cornell has been supported by NSF grants AST-0607007, AST-1107390 and AST-1714828 and by grants from the Brinson Foundation. Participation of the Undergraduate ALFALFA Team has been made possible by NSF grants AST-0724918, AST-0725267, AST-0725380, AST-0902211, AST-0903394, AST-1211005, AST-1637339, AST-1637271, AST-1637299, AST-1637262 and AST-1637276. EAKA is supported by the WISE research programme, which is financed by the Netherlands Organisation for Scientific Research (NWO). BRK acknowledges the National Radio Astronomy Observatory (NRAO). The NRAO is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. MGJ acknowledges support from the grant AYA2015-65973-C3-1-R (MINECO/FEDER, UE). We thank Dmitry Makarov for comments and suggestions on cross-identifications.

This work is based on observations made with the Arecibo Observatory. The Arecibo Observatory has been operated by SRI International under a cooperative agreement with the National Science Foundation (AST-1100968), and in alliance with Ana G. Mendez-Universidad Metropolitana, and the Universities Space Research Association. We thank the staff of the Arecibo Observatory especially Phil Perillat, Ganesh Rajagopalan, Arun Venkataraman, Hector Hernandez, and the telescope operations group for their outstanding support of the ALFALFA survey program.

We acknowledge the use of NASA’s *SkyView* facility (<http://skyview.gsfc.nasa.gov>) located at NASA Goddard Space Flight Center and the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. The Digitized Sky Surveys were produced at the Space Telescope Science Institute under U.S. Government grant NAG W-2166. The images of these surveys are based on photographic data obtained using the Oschin Schmidt Telescope on Palomar Mountain and the UK Schmidt Telescope. The plates were processed into the present compressed digital form with the permission of these institutions. The Second Palomar Observatory Sky Survey (POSS-II) was made by the California Institute of Technology with funds from the National Science Foundation, the National Geographic Society, the Sloan Foundation, the Samuel Oschin Foundation, and the Eastman Kodak Corporation.

This research used data from the Sloan Digital Sky Survey Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck

Society, and the Higher Education Funding Council for England.

Facilities: Arecibo

Software: IDL

REFERENCES

- Adams, E. A. K., Giovanelli, R., & Haynes, M. P. 2013, *ApJ*, 768, 77
- Adams, E. A. K., Cannon, J. M., Rhode, K. L., et al. 2015a, *A&A*, 580, A134
- Adams, E. A. K., Faerman, Y., Janesh, W. F., et al. 2015b, *A&A*, 573, L3
- Ball, C., Cannon, J. M., Leisman, L., et al. 2018, *AJ*, 155, 65
- Barnes, D. G., Staveley-Smith, L., de Blok, W. J. G., et al. 2001, *MNRAS*, 322, 486
- Bianchi, S., Giovanardi, C., Smith, M. W. L., et al. 2017, *A&A*, 597, A130
- Bicay, M. D., & Giovanelli, R. 1986, *AJ*, 91, 705
- Briggs, F. H. 1998, *A&A*, 336, 815
- Brown, T., Catinella, B., Cortese, L., et al. 2015, *MNRAS*, 452, 2479
- . 2017, *MNRAS*, 466, 1275
- Cannon, J. M., Giovanelli, R., Haynes, M. P., et al. 2011, *ApJL*, 739, L22
- Cannon, J. M., Martinkus, C. P., Leisman, L., et al. 2015, *AJ*, 149, 72
- Catinella, B., Haynes, M. P., & Giovanelli, R. 2007, *AJ*, 134, 334
- Catinella, B., Saintonge, A., Janowiecki, S., et al. 2018, *MNRAS*, 476, 875
- Chengalur, J. N., Salpeter, E. E., & Terzian, Y. 1993, *ApJ*, 419, 30
- Condon, J., Cotton, W., E.W., G., et al. 1998, *AJ*, 115, 1693
- Darling, J., & Giovanelli, R. 2006, *AJ*, 132, 2596
- Djorgovski, S. G., Gal, R. R., Odewahn, S. C., et al. 1998, in *Wide Field Surveys in Cosmology*, ed. S. Colombi, Y. Mellier, & B. Raban, 89
- Fabello, S., Catinella, B., Giovanelli, R., et al. 2011a, *MNRAS*, 411, 993
- Fabello, S., Kauffmann, G., Catinella, B., et al. 2011b, *MNRAS*, 416, 1739
- . 2012, *MNRAS*, 427, 2841
- Fouque, P., Bottinelli, L., Gougouenheim, L., & Paturel, G. 1990, *ApJ*, 349, 1
- Gavazzi, G., Fumagalli, M., Fossati, M., et al. 2013, *A&A*, 553, A89
- Giovanelli, R., & Haynes, M. P. 2015, *A&A Rv*, 24, 1
- Giovanelli, R., Haynes, M. P., Kent, B. R., & Adams, E. A. K. 2010, *ApJL*, 708, L22
- Giovanelli, R., Haynes, M. P., Salzer, J. J., et al. 1994, *AJ*, 107, 2036
- Giovanelli, R., Haynes, M. P., Kent, B. R., et al. 2005, *AJ*, 130, 2598
- . 2007, *AJ*, 133, 2569
- Giovanelli, R., Haynes, M. P., Adams, E. A. K., et al. 2013, *AJ*, 146, 15
- Graham, M. J., Fitzpatrick, M. J., & McGlynn, T. A., eds. 2007, *Astronomical Society of the Pacific Conference Series*, Vol. 382, *The National Virtual Observatory: Tools and Techniques for Astronomical Research*
- Hallenbeck, G., Papastergis, E., Huang, S., et al. 2012, *AJ*, 144, 87
- Haynes, M. P., Giovanelli, R., Chamaraux, P., et al. 1999, *AJ*, 117, 2039
- Haynes, M. P., Giovanelli, R., & Kent, B. R. 2007, *ApJL*, 665, L19
- Haynes, M. P., Giovanelli, R., Martin, A. M., et al. 2011, *AJ*, 142, 170
- Hirschauer, A. S., Salzer, J. J., Skillman, E. D., et al. 2016, *ApJ*, 822, 108
- Hogg, D. W. 1999, *ArXiv Astrophysics e-prints*, astro-ph/9905116
- Huang, S., Haynes, M. P., Giovanelli, R., & Brinchmann, J. 2012a, *ApJ*, 756, 113
- Huang, S., Haynes, M. P., Giovanelli, R., et al. 2012b, *AJ*, 143, 133
- Janesh, W., Rhode, K. L., Salzer, J. J., et al. 2017, *ApJL*, 837, L16
- . 2015, *ApJ*, 811, 35
- Janowiecki, S., Leisman, L., Józsa, G., et al. 2015, *ApJ*, 801, 96
- Jones, M. G., Haynes, M. P., Giovanelli, R., & Moorman, C. 2018, *MNRAS*, 477, 2
- Jones, M. G., Haynes, M. P., Giovanelli, R., & Papastergis, E. 2016a, *MNRAS*, 455, 1574
- Jones, M. G., Papastergis, E., Haynes, M. P., & Giovanelli, R. 2016b, *MNRAS*, arXiv:1510.07050
- Kennicutt, R. C., & Evans, N. J. 2012, *ARA&A*, 50, 531
- Kent, B. R. 2010, *ApJ*, 725, 2333
- . 2011, *Astronomical Data Analysis Software and Systems XX*. ASP Conference Proceeding, 443, 625
- Kent, B. R., Giovanelli, R., Haynes, M. P., et al. 2008, *AJ*, 136, 713
- Koopmann, R. A., Giovanelli, R., Haynes, M. P., et al. 2008, *ApJL*, 682, L85
- Landsman, W. B. 1993, in *Astronomical Society of the Pacific Conference Series*, Vol. 52, *Astronomical Data Analysis Software and Systems II*, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, 246
- Lee-Waddell, K., Spekkens, K., Cuillandre, J.-C., et al. 2014, *MNRAS*, 443, 3601
- Lee-Waddell, K., Spekkens, K., Chandra, P., et al. 2016, *MNRAS*, 460, 2945
- Leisman, L., Haynes, M. P., Giovanelli, R., et al. 2016, *MNRAS*, 463, 1692
- Leisman, L., Haynes, M. P., Janowiecki, S., et al. 2017, *ApJ*, 842, 133
- Lineweaver, C. H., Tenorio, L., Smoot, G. F., et al. 1996, *ApJ*, 470, 38
- Martin, A. M., Giovanelli, R., Haynes, M. P., & Guzzo, L. 2012, *ApJ*, 750, 38
- Martin, A. M., Giovanelli, R., Haynes, M. P., et al. 2009, *ApJS*, 183, 214
- Martin, A. M., Papastergis, E., Giovanelli, R., et al. 2010, *ApJ*, 723, 1359
- Masters, K. L. 2005, PhD thesis, Cornell University, New York, USA
- McGlynn, T., & Scollick, K. 1994, in *Astronomical Society of the Pacific Conference Series*, Vol. 61, *Astronomical Data Analysis Software and Systems III*, ed. D. R. Crabtree, R. J. Hanisch, & J. Barnes, 34
- McNichols, A. T., Teich, Y. G., Nims, E., et al. 2016, *ApJ*, 832, 89

- Mei, S., Blakeslee, J. P., Côté, P., et al. 2007, *ApJ*, 655, 144
- Meyer, M., Robotham, A., Obreschkow, D., et al. 2017, *PASA*, 34, arXiv:1705.04210
- Meyer, M. J., Zwaan, M. A., Webster, R. L., et al. 2004, *MNRAS*, 350, 1195
- Moorman, C. M., Vogeley, M. S., Hoyle, F., et al. 2014, *MNRAS*, 444, 3559
- Morganti, R., de Zeeuw, P. T., Oosterloo, T. A., et al. 2006, *MNRAS*, 371, 157
- Odekon, M. C., Koopmann, R. A., Haynes, M. P., et al. 2016, *ApJ*, 824, 110
- Papastergis, E., Giovanelli, R., Haynes, M. P., Rodríguez-Puebla, A., & Jones, M. G. 2013, *ApJ*, 776, 43
- Papastergis, E., Martin, A. M., Giovanelli, R., & Haynes, M. P. 2011, *ApJ*, 739, 38
- Saintonge, A. 2007, *AJ*, 133, 2087
- Saintonge, A., Giovanelli, R., Haynes, M. P., et al. 2008, *AJ*, 135, 588
- Saintonge, A., Catinella, B., Cortese, L., et al. 2016, *MNRAS*, 462, 1749
- Schneider, S. E., Helou, G., Salpeter, E. E., & Terzian, Y. 1986, *AJ*, 92, 742
- Shostak, G. S., & Allen, R. J. 1980, *A&A*, 81, 167
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
- Springob, C. M., Haynes, M. P., & Giovanelli, R. 2005a, *ApJ*, 621, 215
- Springob, C. M., Haynes, M. P., Giovanelli, R., & Kent, B. R. 2005b, *ApJS*, 160, 149
- Springob, C. M., Masters, K. L., Haynes, M. P., Giovanelli, R., & Marinoni, C. 2007, *ApJS*, 172, 599
- Stierwalt, S., Haynes, M. P., Giovanelli, R., et al. 2009, *AJ*, 138, 338
- Suess, K. A., Darling, J., Haynes, M. P., & Giovanelli, R. 2016, *MNRAS*, 459, 220
- Teich, Y. G., McNichols, A. T., Nims, E., et al. 2016, *ApJ*, 832, 85
- Toribio, M. C., Solanes, J. M., Giovanelli, R., Haynes, M. P., & Masters, K. L. 2011, *ApJ*, 732, 92
- Tully, R. B., Courtois, H. M., Dolphin, A. E., et al. 2013, *AJ*, 146, 86
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579

APPENDIX

A. DATA PIPELINE

In this Appendix, we present details of the data reduction pipeline, developed largely by RG and BK, which has been used to process the ALFALFA survey observational data and produce its final data products.

A.1. Processing of the Drift Scan Data

The basic data reduction for the ALFALFA survey has been undertaken in specially-designed software developed in the *Interactive Data Language* (IDL) software environment. IDL is published by Harris Geospatial Solutions and is a dynamically typed, single namespace language with advanced graphics capabilities commonly employed in space science research. The procedure library developed by Arecibo staff member Phil Perillat is used in the ALFALFA software reduction pipeline for processing data from the WAPP backend. The pipeline makes use of the IDL User’s Library for data file input/output (Landsman 1993).

The raw spectral ALFALFA data consist of 1-second records of the individual spectra sampled by the two separate (linear) polarizations of each of the seven ALFA feed horns via the WAPP spectrometer. The initial FITS format files written to disk contain 600 such records, each containing the spectra for the 2 polarizations of the 7 beams. For processing, the original FITS files are first converted to array structures and saved in the native IDL format, so that each “drift” file is a 2 by 600 by 8 element array. In each array element, a large IDL structure is built with the original FITS header information, high-precision position and time stamp information, as well as a 4096 channel array containing the raw spectrum. The 8th array “beam” element carries redundant data and is kept merely for convenience.

As discussed in Giovanelli et al. (2005), a first stage of calibration is performed by injecting a noise diode (“cal”) into the system after every 600 seconds of a drift scan. Each one-second calibration record from the injected noise diode is written to an identical structure type, and saved in its own CALON file. An “off” calibration record is created from the final and first records of two adjacent drift files, and is also saved in its own CALOFF file, yielding a “triplet” of associated calibration files: the preceding CALOFF, the CALON, and the following CALOFF. The series of 600 second drift and calibration “triplets” file lists are created with scan number prefixes for import into the calibration pipeline. A temperature calibration ratio is computed by taking the cal values at 1400 MHz as provided and maintained by observatory staff and dividing by the difference of the average total power for both the CALON and CALOFF files. Corrections for the frequency dependence of the system are applied by the standard Observatory processing software. This is performed for each calibration file, beam, and polarization obtained at 600 second intervals over the whole observing period. Since the telescope is not moved and power levels are not normally adjusted during an observing run, systematic changes in the calibration ratio are an indication of gain drift, usually due to electronics. A sample calibration ratio for beam 0 and polarization A as well as the system temperature T_{sys} during an observing run is shown in Figure A1. Individual 10 minute calibrations may be affected by weather (lightning), RFI, or the presence

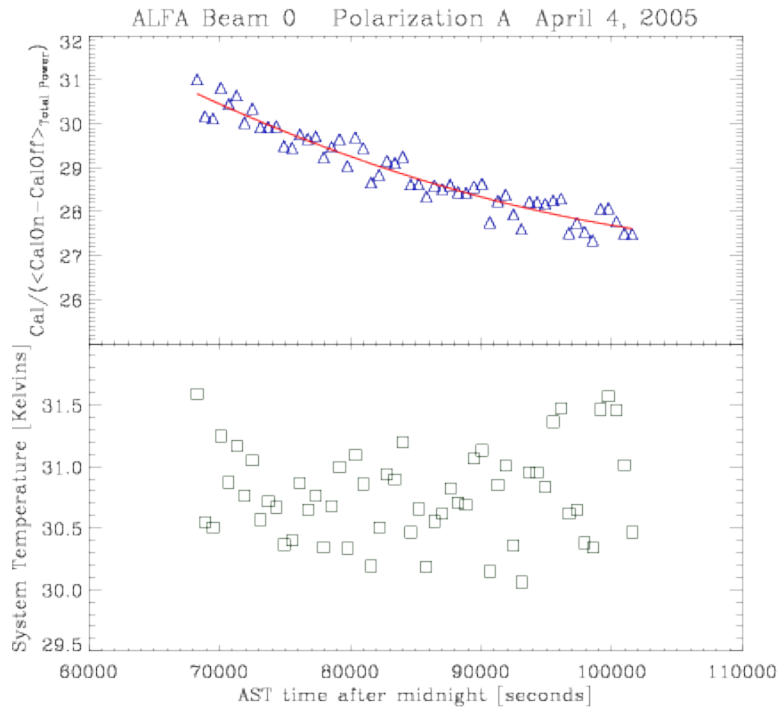


Figure A1. Sample calibration solution for one polarization of the center feed horn for a representative observing session (April 4, 2005). The upper plot shows the calibration ratio $\text{Cal}/(\text{TP}_{\text{calon}} - \text{TP}_{\text{caloff}})$ and a third order polynomial fit to the data. The lower plot shows the system temperature during the observing run. All data shown are for the same polarization (“A”) of the central beam (“Beam 0”).

of continuum emission in the CALON/CALOFF scans. As is evident in that Figure, there is a systematic drift in the calibration ratio, often observed over the course of the night, particularly in the summer; a common cause is believed associated with the ability of the dewar to cool the amplifiers as the ambient temperature declines during the night. Calibration is applied using the best fit to the observed variation rather than individual values, thereby reducing the scatter introduced by continuum sources and RFI in this simplistic approach to calibration.

Bandpass subtraction is performed on 2-dimensional time/frequency plots, one map at a time on each polarization and beam for a given drift file. The calibration process begins by performing a robust linear fit along the time dimension for each frequency channel. The rms is also computed for each channel, as is the fraction of time series records less than $2 \times \text{rms}$. After exclusion of outliers that deviate more than $2 \times \text{rms}$ from the fit, a bandpass value of either (1) the zeroth order coefficient to the linear fit c_0 , or (2) the median value of the strip is selected. This option is chosen by the user at the time of reduction. Based on experience, option 1 is usually preferable. The rms as a function of channel is iteratively fit with a 3rd order polynomial. Channels (including those at the bandpass edges) are flagged that deviate several standard deviations from the fit. In addition, channels are manually flagged, such as those around Galactic HI. A cubic spline is used to interpolate the bandpass across the flagged channels. An “off” bandpass is created as the normalized bandpass times the system temperature. A background total power continuum value is also computed for all time series records for all records and channels that have not been flagged, excluding point sources; the continuum contribution from these point sources is also stored away. Figure A2 shows the multiple diagnostics of the described calibration process.

The final calibrated and corrected bandpass is computed as

$$BP_{\text{corr}} = \frac{BP_{\text{on}} - BP_{\text{off}} T_{\text{sys}}}{BP_{\text{off}}} \quad (\text{A1})$$

All calibrated values are stored in units of Kelvins. Calibrated and reduced drift files are saved to disk for interference flagging. An example of the process of the bandpass calibration and subtraction process for a typical time-frequency drift dataset is shown in Figure A3.

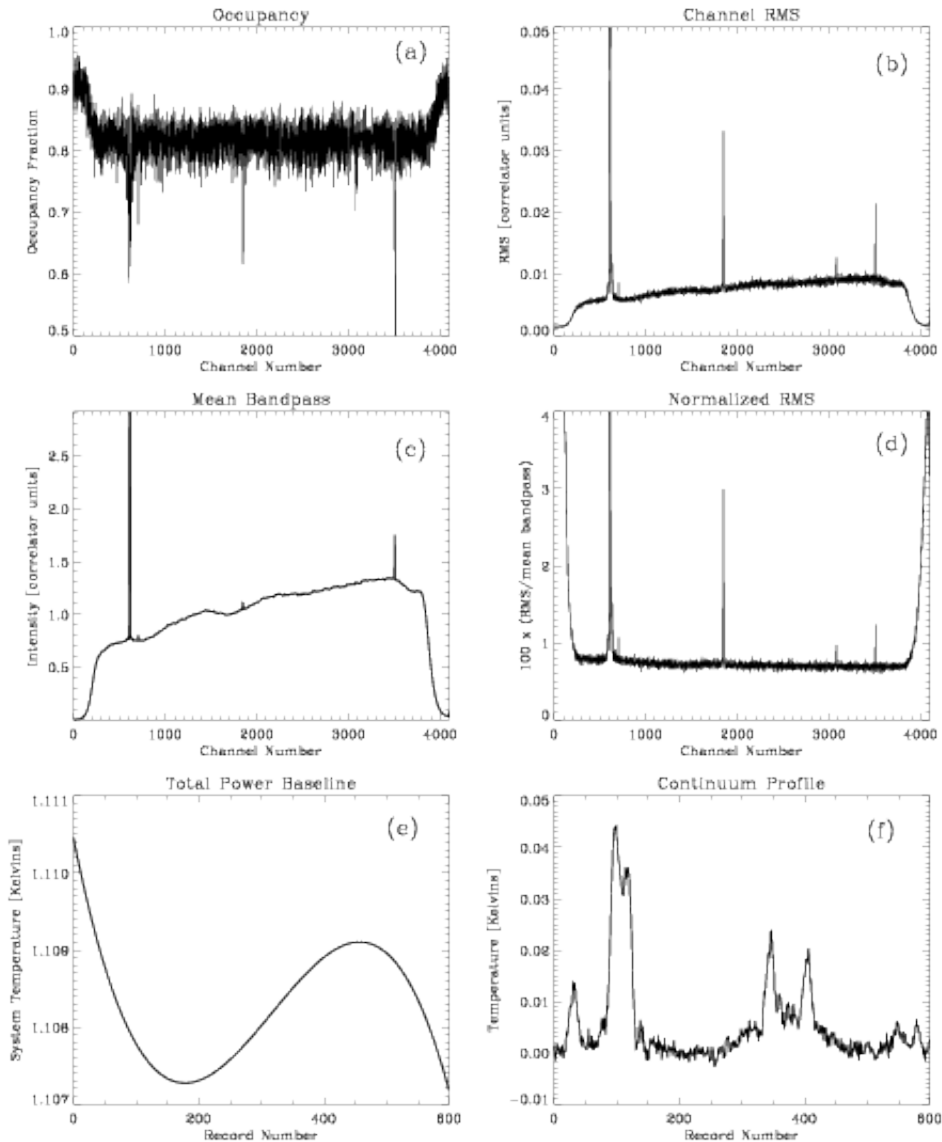


Figure A2. Plots showing the products, statistics, and diagnostics of the bandpass calibration for a single 600 sec drift. (a) shows the occupancy fraction of records that are within $2 \times \text{rms}$ of the time series fit for each channel; (b) shows the rms for each channel; (c) the mean bandpass of the strip; (d) the rms divided by the mean bandpass multiplied by a factor of 100; (e) the continuum baseline contribution after removal of point sources; and (f) the continuum “strip chart” integrated along *good* channels. All data shown are for the same polarization (“A”) of one of the outer beams (“Beam 4”).

Careful attention is given to flagging radio frequency interference for both improvement of data quality products and to decrease the likelihood of including spurious detections in the automated signal extraction process later on. The RFI flagging stage is where data quality is first assessed. Each bandpass calibrated 2D time/frequency plot is examined closely by eye. Areas of interference due to nearby airport radar and associated harmonics are individually flagged; these records and/or channels are excluded during the signal extraction and gridding process. An example flagging session is shown in Figure A4. The data products at this stage in the pipeline are known as Level I data products.

A.2. Construction and Analysis of 3-D Grids

ALFALFA “grids” are 3-D position-position-velocity “cubes” from which final source measurements are obtained. As discussed in Section 2.3, the grid centers are located at pre-determined locations, separated by 8^{min} of R.A. and centered at Declinations spaced by 2° between $+01^\circ$ and $+35^\circ$. The process of creating grids begins by scanning coordinate metadata saved in small size ($\sim 30\text{MB}$) files containing the sky positions of every spectrum recorded during

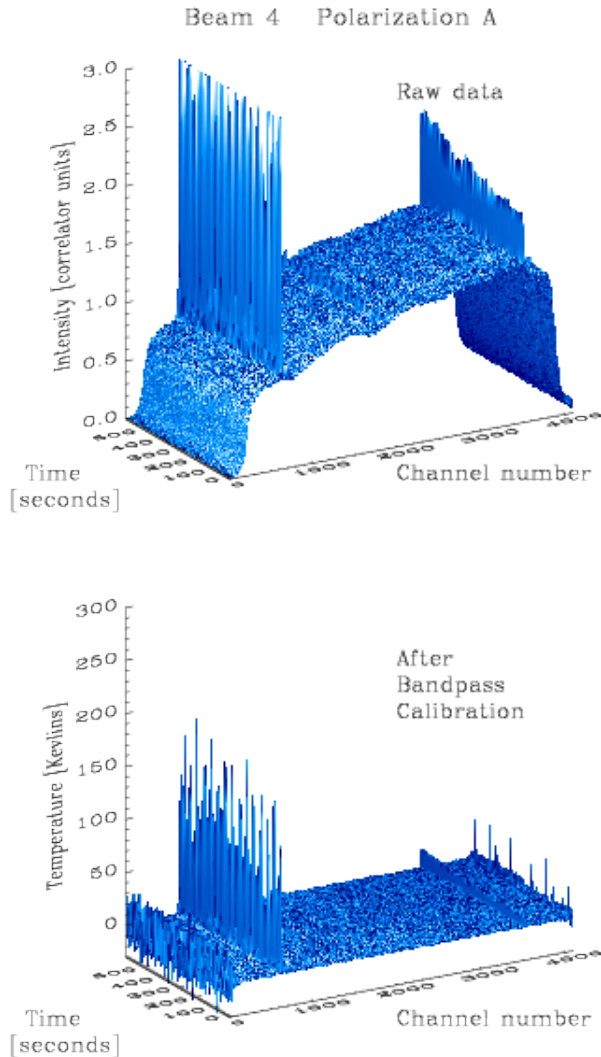


Figure A3. 2D surface illustrating the bandpass calibration and subtraction process for a representative time/frequency (spectrometer “channel number”) plot for one polarization (“A”) of a single beam (“Beam 4”) over a 600 sec drift. The upper plot shows the raw ALFALFA drift scan spectral data, clipped to an intensity of 3.0 for dynamic range convenience. The bottom plot shows the spectral data, scaled to Kelvins, after the bandpass calibration process. Features in both plots include the FAA radar signature near channel 615 (1350 MHz) and the Galactic HI line emission near channel 3500 (1420 MHz).

each observing session. A listing is compiled of any 600-second drift files that will contribute to any position within the specified grid boundaries. Each drift file is opened and data contributing to the grid are summed in the appropriate R.A./Dec./frequency bins weighted by a Gaussian kernel of size $2'$. Grid pixels are approximately $1'$ square depending on the declination range. Data are also included or ignored based on RFI flagging from the level I production process. A spectral weight map for each channel is also created based on contributing drifts at each grid point. The entire process is repeated for total power continuum maps and weights. Final grids are scaled to units of Jansky beam $^{-1}$ and divided by the appropriate weights for both spectral and continuum maps. In addition, because the synthesized beam area used to generate the grids is larger than that of the telescope at L-band (from $3.5'$ to $4'$), the grids are also multiplied by a gain dilution factor given by

$$G_{dilute} = 1 + \frac{W_{FWHM}^2}{3'.3 \times 3'.8} \quad (\text{A2})$$

where W_{FWHM} is the Gaussian kernel size of $2'$.

Velocity channels are first shifted to the heliocentric velocity frame such that channel 2047 (counting from zero) of all raw spectra are at a frequency of 1385 MHz and $cz_{\odot} = 7663 \text{ km s}^{-1}$. The process creates four spectral grids. Each contains a spectral HI grid of size 144×144 pixels in R.A. and declination, and 1024 channels in frequency space

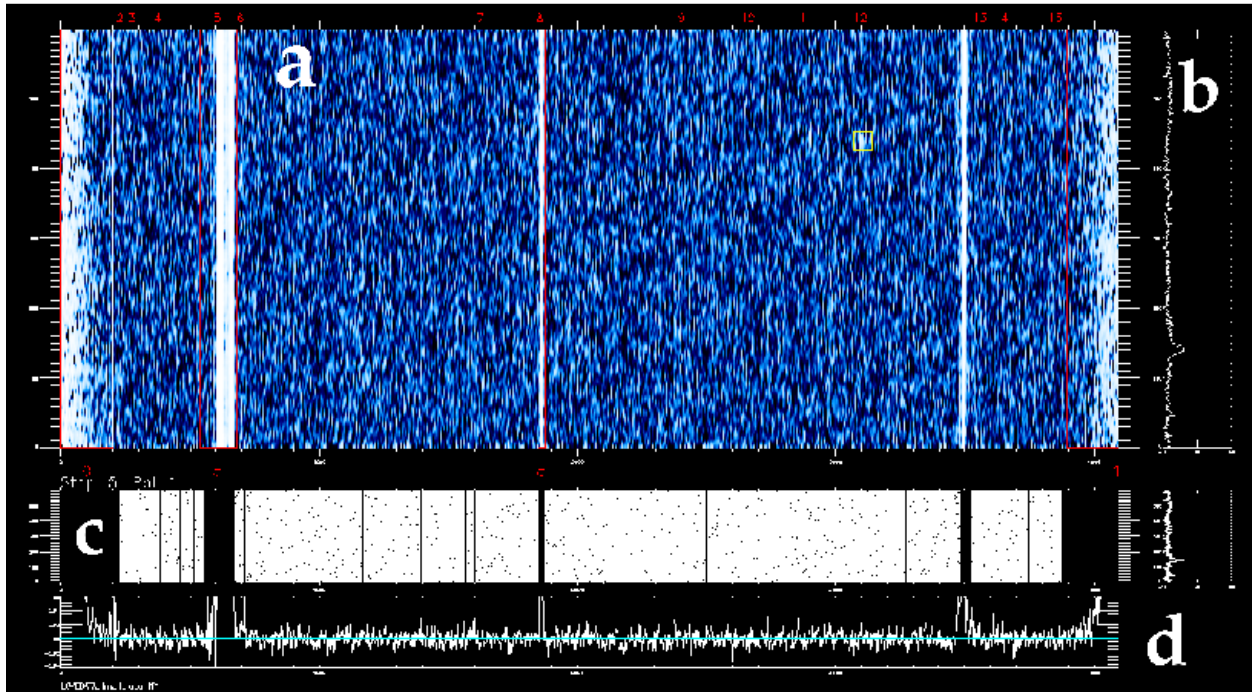


Figure A4. Example session of radio-interference flagging from April 4, 2005. The user selects regions of interference via predetermined box numbers or user-defined regions. Cross-referencing with optical and HI databases allows the user to check known information during the data reduction session. The figure sections are as follows: a) the main time vs. frequency plot of the data showing the FAA radar and Galactic HI, b) the continuum flux integrated along the strip, c) the map mask showing the pixels (in white) used by the bandpass calibration procedure to produce the continuum flux strip, and d) the average spectrum for all 600 records. Prominent features (indicated by white vertical stripes) include, from left to right, the strong FAA radar at 1350 MHz, an internally-generated modulation of the radar at 1380 MHz, and the Galactic HI emission at 1420 MHz. Red boxes indicated marked regions of RFI, and the yellow box indicates a galaxy previously identified in the AGC database at the corresponding position and redshift.

($\sim 5100 \text{ km s}^{-1}$ in cz space). The four spectral grids are identified by a letter designation for the cz_{\odot} range that they cover: (a) -2000 to 3300 km s^{-1} , (b) 2500 to 7950 km s^{-1} , (c) 7200 to 12800 km s^{-1} , and (d) 12000 to 17900 km s^{-1} . Each grid overlaps the next by $\sim 1000 \text{ km s}^{-1}$ in cz space. Each of the four grids occupies approximately 330 MB of disk space, with ancillary files attached containing a complete history of how the grid was constructed.

Baselining involves fitting polynomials to grid slices, i.e., R.A. vs. spectral channel maps separately for both polarizations. For most maps, a linear fit is subtracted in the spectral direction. Special cases may require excluding Galactic HI, high-velocity clouds, and high SNR extended detections from the fit. In areas where residual stray RFI may be present, baselining is also performed in the R.A. direction. A secondary process involves subtracting low order fits to R.A./Dec. maps for each spectral channel. The subtraction along each R.A. strip effectively “flatfields” the image in channels devoid of any signals. Extended signals are excluded from the fit, especially for channels containing Galactic HI or associated with bright galaxy HI line emission.

Retained along with the 3-D spectral grid is a continuum map over the same spatial area and a structure containing the normalized weight of each spectral grid point. The latter provides a record of data quality including the “gain scalloping” due to differences in the gains of the central versus peripheral beams, the possibility of missing beams/polarizations due to hardware failure, RFI excision, and the varying number of drifts covering each spatial point. Slices in the spectral dimension (“channel maps”) can be examined along with the continuum and comparable “weights maps” simultaneous (see Figure A5). The comparison of grids constructed from separate polarizations provides a further check on RFI contamination. The retention of such weight information also provides input for stacking software allowing the identification (and rejection) of sources for which the data quality do not meet a threshold for inclusion (e.g, high RFI excision). The extracted spectra also maintain a value of the normalize weight per spectral channel for similar reasons.

As mentioned in Section 2.4, HI line candidates are identified in the final grids by applying a matched filter algorithm in the Fourier domain (Saintonge 2007). The candidates are then examined and parameters measured individually using the tool GridView. An example of a GridView application is given in Figure A5. Importantly, GridView allows

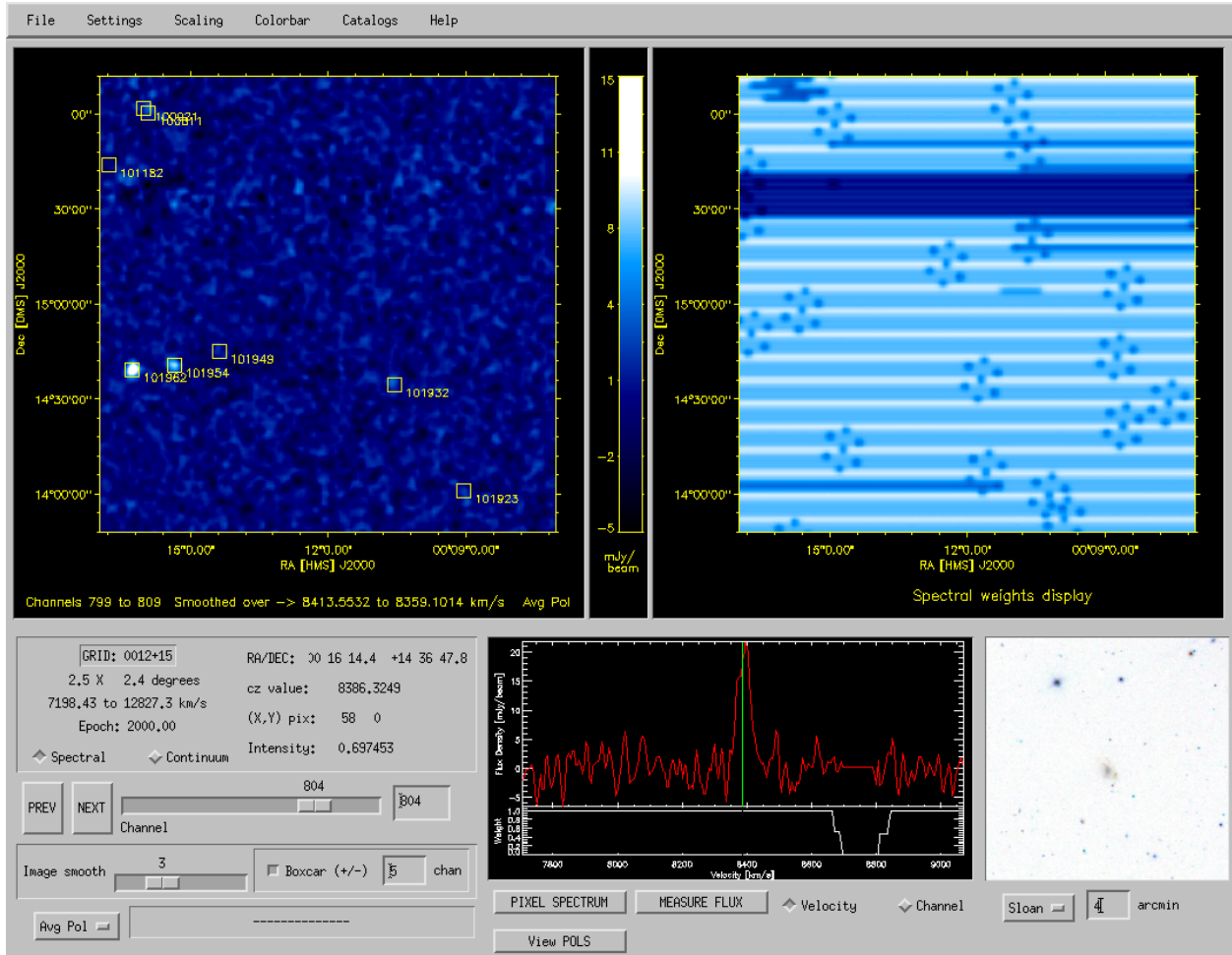


Figure A5. GridView application GUI created in IDL. The data cube visualization procedure incorporates various IVOA web service tools for use during the data analysis process. The figure shows both optical image viewing and catalog overlay on the HI line 0th moment map over an adjustable range of channels (left). The display to the right of the moment map shows the corresponding weight map of the same region; horizontal stripes in the weights map reflect the interleaved drift scan strategy and scalloped nature of coverage by beams of different gain. In that display, dark means low weight; there is a drift missing from this dataset. The repeated pattern of seven darker spots represent the interruption of data taking for firing of the noise diode (every 600 seconds); the seven spots are the seven ALFA beams. The target under analysis here is AGC 101962, matching the brightest white spot on the left just below the center of the moment map. Its spectrum is shown near the center bottom; the spectral weight display, plotted underneath the flux density indicates that which the spectral range near the galaxy’s HI emission is fine, but that the spectral region around 8700 km/s is of low weight, due to excision of RFI generated by the NUDET instrument on the GPS constellation of satellites and an internal modulation of the FAA radar. The right bottom window displays the SDSS r-band image of the field centered on the HI centroid. Users are able to manipulate the 3-D spectral grid with the various GUI controls, export maps, view polarizations separately, and examine the spectrum corresponding to any grid pixel.

the manipulation (smoothing, comparing polarizations etc) of channel maps as well as examination of the weight and continuum maps and the overlay and cross reference of complementary datasets, redshift catalogs and imaging databases. This functionality is important for the final catalog data quality and critical for the identification of the probable OCs.

Identified sources are then run through the routine GalFlux which measures fluxes and spectral parameters, allows cross-referencing with other databases and produces the final data products for each detection. The user selects the region over which the spectrum will be measured. A “postage stamp” is extracted, isophotes are fit in the spatial domain and the spectrum displayed. Elliptical isophotes are automatically computed at levels of half and one-quarter of the peak power and at levels of 100, 200, 300, 500 and 1000 mJy beam⁻¹; custom isophotes could also be fit. The interactive examination of the isophotes allows the user to refine the boundaries in both spatial and spectral dimensions and to refine the baseline fit. The user then can make adjustments to the spatial and spectral boundaries and can perform additional baseline fitting, smoothing etc. to determine the best spectral and spatial definition of the HI line

signal. Most often, the metric used to judge the best fit is the SNR; if the boundaries are too large, noise will drive the SNR down. In the majority of sources, such choices were not important, but this flexibility allows the user to make decisions about the best fit in the presence of low SNR, confusion, RFI contamination, polarization mismatch or proximity to Galactic HI line emission and to note evidence of extended sources or other peculiarity. Such examination can even result in source rejection; for example, if the feature is very narrow in velocity and its spatial definition is inconsistent with the beam pattern, the source is highly likely to be spurious. The half peak power isophote is typically larger than the ALFA beam FWHM and should contain all of the flux for sources of extent comparable to or smaller than the beam. It should be noted, however, that in the vast majority of sources, the true HI extent is not known but is expected to be smaller than the beam. As noted previously, flux may be missing from very extended or asymmetric sources when this pipeline approach is used. A separate work will present fluxes for the known extended sources (Hoffman *et al.*, in preparation).

Once the user is confident of the source definition, then the flux and velocity measurement algorithm, as described in Section 2.5, is applied to produce a display of the postage stamp, isophotal fits, the baselined spectrum and the associated weights over the isophote at 50% of the peak flux, and measures of the integrated flux, velocities, velocity widths and their error. In addition to statistical errors on the HI line flux density, velocity and velocity width, systematic uncertainties on each value are estimated by flagging minimum and maximum estimates of the extent of the spectral feature and then using the ratio of those values to estimate uncertainties. In most cases, the systematic uncertainty is less than the statistical one and can be ignored. In cases of e.g., low SNR, shallow profile outer slope or RFI contamination, the statistical error can be clearly too small and the uncertainty is set as the sum in quadrature of the two. Optical databases (SDSS, SkyView and NED) can be accessed and displayed so that the user can assign the most probable OC. The user assigns the detection category and can enter comments about the source, the extracted source, the optical environment, etc. The final step in the analysis of a source is the production of a final output file containing all of the above information.

An additional GUI, depicted in Figure A6 enables the examination of a catalog of HI line sources and their corresponding “source” file. Individual sources can be selected from the catalog listing (upper left); the isophotal fits (lower left), spectrum and associated weights (lower center) and corresponding optical field (lower right) are displayed along with a summary of the derived and assigned parameters. A user can update parameters if additional information becomes available; the latter capability has been an important component over the years.

Additional routines allow production of the catalogs, application of the adopted flow model (Masters 2005) and derivation of the HI masses. The full ALFALFA data reduction package, dubbed “Lovedata” (because we love our data), has been exported to over 50 sites.

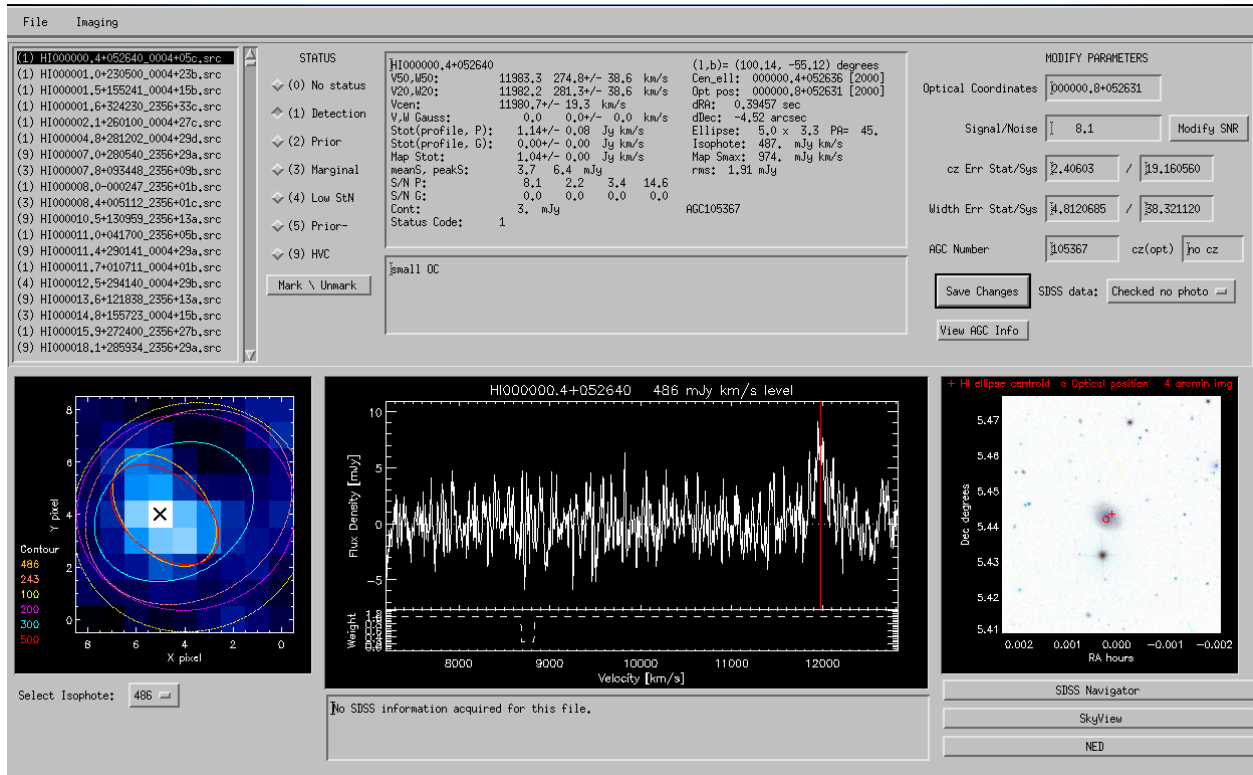


Figure A6. GalCat application GUI created in IDL. The catalog of ALFALFA sources is displayed as a list on the upper left. For each source selected from that list, a summary is presented in the interface, including the measured parameters (top center), the isophotal fits (bottom left), the extracted “postage stamp” spectrum and its normalized weight at each spectral point (bottom center), and the optical image (bottom right). The upper right panel offers the possibility to modify parameters as necessary.