Initial HI Survey Observations with ALFA in Drift Mode

A Proposal submitted by: Riccardo Giovanelli, Martha Haynes, Brian Kent & Amelie Saintonge (Cornell), Alessandro Boselli (Marseille), Noah Brosch (Wise Observatory), Christian Bruens (U. of Bonn), Barbara Catinella (NAIC), Jonathan Davies (Cardiff U.), Diego Garcia–Lambas (U. of Cordoba), Giuseppe Gavazzi (U. of Milan), Lyle Hoffman (Lafayette Coll.), Igor Karachentsev (Special Astrophys. Obs., Russia), Valentina Karachentseva (U. of Kiev), Becky Koopmann (Union College), Karen Masters (Cornell), Emmanuel Momjian (NAIC), Mary Putman (U. of Colorado), Jessica Rosenberg (U. of Colorado), Josep M. Solanes (U. of Barcelona), Kristine Spekkens (Cornell), Chris Springob (Cornell), Carlos Valotto (U. of Cordoba), Wim van Driel (Meudon), Liese van Zee (Indiana)

Background and Main Goals: The Arecibo telescope is the world's most sensitive instrument for studies of the HI emission in galaxies. The combination of its high sensitivity with the multi–pixel capability of the Arecibo L-band Feed Array (ALFA) will allow us to conduct sensitive surveys over large portions of the sky. A multiplicity of science programs will initiate the exploitation of the new capabilities afforded by ALFA. Several of those science programs will be best served by observations carried out in drift mode, i.e. one in which the telescope configuration remains fixed and the sky drifts by at the sidereal rate.

Drift mode observations will address scientific problems such as the nature and cosmic density dependence of the HI Mass Function, the possible existence of optically inert, baryon–rich, low mass halos, the structure of groups and nearby clusters of galaxies, the study of the peripheries of nearby galaxies and of the properties and nature of high velocity clouds, the abundance of low z HI absorbers, the detection of medium z OH Megamasers. For a more detailed descriptions of the science, we direct the attention of reviewers to the E–ALFA 'White Paper' delivered to NAIC in July 2003 (see alfa_extragal-whitepaper.pdf in alfa.naic.edu/extragal/alfa_extragal.html).

HI ALFA surveys will constitute an important astronomical legacy of the Arecibo Observatory for many years to come. They will require hundreds or even thousands of hours of telescope time. The standards on the process of delivery of data products to the community will correspondingly be very high.

The main goal of this "survey precursor" proposal is that of finalizing the optimization of the observing techniques that will use ALFA in drift mode, in the forthcoming extragalactic HI surveys, thus achieving 'closure' on the definition of the technical parameters of those surveys. The selection of the regions of the sky for the observations described below was made with the expectation that the observations will be able to deliver substantial science fall–out, with the understanding that the observations are to be carried out within a "shared risk" framework, as pointed out in NAIC's call for ALFA precursor proposals. Other observing modes will be investigated by other E–ALFA groups, in a coordinated effort with the present one.

Preparation for ALFA: ALFA will arrive on site at Arecibo in April 2004, opening new prospects for large–scale surveys at L–band. This proposal is in response to NAIC's announcement to welcome proposals for ALFA preliminary "shared risk" observations, aimed towards verifying the suitability of the multi–feed system for successful and efficient use in a protracted survey effort. Because the "quality bar" for all stages of data taking and processing will be very high for large surveys, special attention needs to be given, in advance, to a number of technical aspects, which naturally follow, and take advantage of, the commissioning tasks planned by NAIC for the first few months after delivery of the ALFA front–end. Among them:

- (a) Calibration scheme: how often? "stop-n-go" or "on-the-fly"? impact of continuum sources (esp. near galactic plane)?

- (b) Sidelobe contamination: const. declination drifts at constant (AZ,ZA) or not? Impact on contiguous strips which will have different sidelobe structure?
- (c) Bandpass calibration: 1-d or 2-d? Impact of standing waves.
- (d) Can the survey be bandpass–calibrated on the fly, i.e. is delivery of bandpass–calibrated data in near–real–time possible?
- (e) For some applications, "E–ALFA data" could have important "G–ALFA fallout". Recovery of galactic HI: bandpass calibration? (Daring) interpolation or fake an occasional frequency switch?
- (f) Sky tiling: optimal tile size?
- (g) Optimal sky weaving in multiple pass drift mode.
- (h) Accumulation of multiple drifts over same strip: noise behavior? Treatment of standing waves?
- (i) Test 3-d rfi excision (in spectral regime, by strip and by multiple passes).
- (j) Test front-end, "electronic gain" (i.e. conversion of backend counts to flux) stability timescales.
- (k) Test suitability of tiled data for continuum mapping.
- (l) Minimization of overhead.
- (m) Test match of NAIC–provided software with custom–provided processing pipeline.

While the precursor observations proposed here are principally aimed at resolving issues of a technical nature, they may also provide results of scientific–grade quality. We anticipate that these tests, jointly with the survey simulations and processing pipeline development currently under way, will be sufficient to satisfactorily characterize the final planning of future HI Drift ALFA surveys.

We wish to underscore that a substantial amount of work has been already carried out by NAIC staff and by members of the team, in preparation for the surveys:

- Data reduction procedures have been developed in IDL — assuming SDFITS but loaded into a *cormap*–compliant internal IDL structure —, for the application of baseband correction, rfi excision, map production and elaborate statistical analysis of the data.
- Automatic signal extraction algorithms, based on a correlation analysis of the data cubes, have been developed and tested.
- Preliminary tests of the performance of the WAPPs as spectral processors have been carried out, fed with the IF from the L–band wide, single pixel receiver.
- Data taking routines have been tested and recommendations have been made to the AO staff, on improvements, new implementations and header contents, that will facilitate the conduction of spectral line surveys.
- NSF support (AST-0307661) has been obtained, in the form of a 3–yr grant, towards the initial phase of HI ALFA surveys. Other proposals have been submitted to funding agencies.

Initial Observations: As we discuss in Appendix 1, an all–sky survey (dubbed "ALFALFA") may subdivide the Arecibo sky in "tiles", of $\sim 20^m \times 4^\circ =\sim 20$ square degrees. These tiles will form the basic units for data processing and public delivery. A fast survey such as ALFALFA may consist of a single or double pass in drift mode through each of the sky tiles. A medium–deep survey may consist of several passes (at different epochs) of each tile, each in drift mode, thereby maximizing efficiency, enhancing rfi removal and offering the opportunity for commensal continuum transient detection within the same dataset. In the case of multiple passes, denser than half beam sampling in Dec. may be desirable, in order to improve the ability for source centroiding and map gridding.

For this precursor proposal, we propose to make observations as specified as follows:

- (a) We propose to complete observations over a tile width of (4°) in Dec in single pass mode, covering 1^h in RA during a nighttime slot and for 1^h in RA during a mid-afternoon slot, in order to compare the suitability of daytime observations for survey work. Since it takes 19 strips to complete a tile, this will require 2 allocations of 19 sessions of 1^h each plus setup time.
- (b) We propose to survey more deeply a smaller region, 1° in Dec and 1^h in RA, in multi–drift (5 passes) mode, also in a nighttime slot. This strategy is aimed at verifying the stability of baselines, the impact of standing waves and the vulnerability of the ALFA system to thermal gradients, out-of-band rfi and different environmental variables. To survey 1° in Dec requires 5 passes of each of 5 strips, for a total of 25 sessions, also at nighttime.
- (c) We propose to obtain a few drift "strips" across the galactic plane, in order to test the suitability of bandpass correction schemes to conditions of rapidly varying, underlying continuum emission. These would best be done during nighttime.

The nightime observations listed above under (a) and (b) will be centered near Dec=+27° and thus will include several very nearby galaxies with measured primary distances, the region with the lowest cosmic density in the local Universe and the Perseus–Pisces supercluster main ridge in the background. Potentially useful scientific results may emerge from this precursor set of observations, in addition to providing technical closure on future survey parameter definition. The data units collected will also serve as tools to refine data processing procedures and our future ability to deliver on a timely fashion robust data products to the community.

To carry out the precursor observations for extragalactic HI surveys in drift mode, following the tiling concept discussed above and in more detail in Appendix 1, we request:

- 25 sessions between LST $00^h\ 30^m$ and $02^h\ 45^m$, during the nighttime, for a total of 56 hours
- 2 sessions between LST $02^h\ 45^m$ and $07^h\ 00^m$, and 2 days between 19^h and 21^h LST, during the nighttime, for a total of 13 hours
- 19 sessions of a $1^h.2$ period in the mid afternoon (1430 to 1700 AST; not in the galactic plane, but otherwise LST not specified), for a total of 23 hours

for a grand total of 92 hours of telescope time, with the ALFA multifeed system and the complete WAPP backend.

Software & other Requirements: We currently anticipate the need for NAIC software development in the areas noted below. Other issues may arise as our own efforts advance.

- *CIMA capabilities:* (1) firing of cal diode every $\sim$ 300 secs with minimum dead time. (2) start of new file (data block) at specified time without interruption of observations (i.e. drift for 1 hour, but write 20 minute "tile" strips separately). We will also need access to the results of the ALFA commissioning phase, ie. beam maps, gain and calibration curves, etc.
- *Monitoring and reduction software:* Our current software development effort uses IDL, taking advantage of the rich library of procedures developed at NAIC and elsewhere, including use of "WAS" routines. We will appreciate being informed about planned NAIC development of data quality monitoring and first look software and specification of Level I product.
- *Data formats:* SDFITS standards. We would appreciate access to finalized data structures in advance of observations and a chance to comment on such.

Result delivery: Data obtained under this program will be placed in the public domain immediately with suitable explanatory documentation to make its access and use feasible by other interested parties. The analysis of most of the tests will be made public within two to four weeks after the completion of data taking. Likewise, summary results of such tests will be provided in the form of public memos, as has already been done (See: http://www.astro.cornell.edu/~haynes/precursor.htm).

Appendix 1: Tiling the Sky with ALFALFA

Survey Tiles. In order to effectively manage a full (Arecibo) sky survey, that will cover $\sim 12,000^\circ$, it will be necessary to subdivide the sky in sectors, for each of which continuum maps and 3–d spectral data cubes can be processed and archived coherently. We shall refer to each of those sectors as a "tile".

Observations are most likely to be carried out in drift mode, so that gain, beam pattern and ALFA rotation angle remain constant through a given drift and, possibly for a whole declination strip. Calibration and data processing will be greatly simplified. In drift mode, ALFA would be rotated at an angle such that each beam sweeps a track separated from that of the next beam by 108", approximately half a beam width. In order to maintain beam patterns and gain constant for a given declination, it may be advisable to observe at fixed azimuth, e.g. sampling along the local meridian.

A single ALFA strip will cover $7 \times 108" = 12'.6$. Nineteen such strips will cover $12'.6 \times 19 = 3^\circ.99 \simeq 4^\circ$ in declination. This is a natural size for a tile, given its commensurability with standard coordinate units. In order for tiles to have a sensible aspect ratio, an RA extent of 20 minutes (tile size: $\sim 4^\circ$ to 5°) is right. For full declination coverage of 36°, the survey would consist of 648 tiles, each of approximately $20 \cos\delta$ square degrees, where δ is the declination of the tile. Gridding the map to half–beam grid point separation, each tile would consist of 200×133 grid points, albeit keeping data to a tighter RA grid separation may be desirable.

The spectral values will be written in 4 byte real format, so a single, one–polarization N–channel spectrum will be 4N bytes long. Assuming the spectral processor dump rate to be 1 sec (the beam will be oversampled in order to allow better rfi–excision capability), a single ALFA drift strip along the RA width of a tile will be $7 \times 2 \times 1200 \times 4N$ plus the space allocated to headers. The largest possible number of spectral channels envisaged for ALFALFA is N=8192, which for a bandwidth of 50 MHz (3 levels) would yield ~ 1.3 km s^{-1} channel separation and may make the data useful for a number of galactic studies, in addition to extragalactic ones. Other ALFALFA options are N=2048, 50 MHz bandwidth (9 levels) and N=4096, 100 MHz (3 levels). With N=8192, a single 7–beam strip across a tile will be 0.55 GBytes, plus headers, for the raw data. Other configurational options will be respectively 2 and 4 times smaller in size.

An important consideration in setting the size of the raw data block units is connected to the computational constraints. The data will be bandpass–corrected one ALFA declination, 7–beam strip at a time. If the length of the strip is restricted to the width of a single tile, and assuming that all seven tracks and both polarizations will be simultaneously processed, it is important for processing expediency that the raw data be loadable in memory all at once. With currently available, inexpensive workstations with 2 GBytes of memory, a 0.6 GByte data set has about the right size for efficient processing.

Bandpass correction and first-pass rfi excision will be applied to fully sampled ALFA strip segments of extent comparable with the tile width. After that, strips may be compressed by a factor of about 3 to 6 in the RA dimension, to approximate one–quarter to one half–beam sampling. Sampling somewhat more generously than the Nyquist rate helps with the quality of the gridding process. Polarizations may be added.