

EXTRAGALACTIC SURVEYS WITH THE ARECIBO L-BAND FEED ARRAY

July 10, 2003

Contents

1	Introduction	2
2	Science justification, survey strategies, and requirements	2
3	Follow-up Observations	4
4	Synergies with Other Surveys	5
5	Hardware Specifications	6
6	Software Specifications	9
7	Organization	11
8	Education and Outreach Activities	12
9	Funding Support	12

Appendices

A	Scientific Justification: Wide Angle and Directed Surveys	15
B	Scientific Justification: Ultra- Deep Pencil-Beam Survey	25
C	Information on RFI at Arecibo in the 1230-1430MHz Bandwidth	28
D	Detailed Lists of Data Products and Relevant Software Tasks	31
E	Assignment of Responsibilities	33
F	Organizational Matters, Working Questions/Proposals	35
	F.1 Proposed Organizational Structure	36

1. Introduction

ALFA will enable a variety of survey projects which were not possible at NAIC before ALFA's advent. A number of those survey projects will be quite ambitious, requiring thousands of hours of telescope time and several calendar years to complete. These large ALFA surveys will need to be carefully evaluated and regulated, so that the enormous commitment by the NAIC has the greatest scientific impact.

Large E-ALFA surveys will likely present a sharp departure from the standard proposal process exercised at NAIC over the past decades. They will produce massive amounts of data in specific areas of research and will provide an important astronomical reference source for years to come, constituting a major legacy of the Arecibo Observatory. In a manner similar to efforts of comparable scope at other major observing facilities, the E-ALFA surveys should be expected to conform to a few basic criteria:

- They will be large, coherent efforts with fundamental scientific goals that cannot be satisfied by a combination of regular proposal efforts.
- The data products of the surveys must be of general interest to a broad scientific community. Therefore, a special effort must be made in the delivery of high quality data products and in the production of robust software tools that facilitate their access.
- All data products should be made accessible by the general community within the shortest possible time after data acquisition at the telescope, thus enabling maximally fertile and competitive scientific follow-up.

This document presents a brief summary of the science issues that would be tackled by E-ALFA surveys in the area of extragalactic Astronomy, the hardware and software requirements for their completion, the synergy with other ALFA and non-ALFA surveys, the scope of possible follow-up observing programs and general organizational details. The main goal of this document is to provide NAIC with useful input in its decisions leading to the delivery of an ALFA system, ancillary hardware and software, and its planning for optimally interfacing with the E-ALFA user community.

Note that some of these same principles are similar to those affecting other ALFA users, while others may be complementary, reflecting the special requirements of exploring the extragalactic universe.

2. Science justification, survey strategies, and requirements

The main categories of current scientific inquiry for extragalactic L-band science are:

- **The low-mass end of the HI mass function:** The relatively small number of low mass ($< 10^8 M_\odot$) sources detected in blind surveys has left large statistical uncertainties in the number density of the lowest HI mass sources. There are up to factors of 10 disagreement in the estimated number of sources at $10^7 M_\odot$. Because of their intrinsically weak signals, presently these sources have been detected only to small distances.
- **High-mass HI sources:** These are rare objects and may have surprising properties. One of the highest mass HI sources, Malin 1, has an extremely low-surface-brightness disk at optical wavelengths. The frequency of the rare sources that fall in the exponential tail of the HI mass function is uncertain, and given the peculiar nature of Malin 1, it is unlikely that optical surveys will provide the best means for identifying the highest-mass sources.
- **Environmental variations of the HI properties of galaxies:** Current observations hint at differences in the HI characteristics of galaxies from location to location. The cluster HI mass function appears to be shallower than the field mass function, which may be caused by a variety of environmental influences that might be explored in detail in nearby groups and clusters.
- **The low HI column density environment of galaxies:** How extended are galaxy disks? Are any subpopulations of the Milky Way's High Velocity Clouds (HVCs) extragalactic? Is there a system of HVCs around other galaxies? How far out and what is the filling factor?

- **Evolutionary changes in the HI mass function:** There is evidence that star formation has been changing rapidly over the past few billion years, perhaps changing the number density of HI systems as strongly as $(1+z)^{3.2}$. Thus, the overall density of HI in the universe and the shape of the HI mass function may be quite different even at a redshift of $z = 0.15$.
- **The link to Lyman- α absorbers:** The connection between the HI systems that generate QSO absorption lines and the HI systems seen in emission at 21 cm remains unclear. Observations at other wavelength regimes that permit the detection of low-redshift absorption against faint QSOs offer the possibility of determining where these absorption lines mostly arise.
- **Local large scale structure in the Zone of Avoidance (ZOA):** Detecting galaxies at 21 cm allows us to map out the galaxy distribution in regions of the sky where interstellar extinction makes it difficult to detect galaxies with optical techniques and where confusion with the galactic foreground hinders the identification of infrared tracers. Nearby structures fill large fractions of the sky, and the ZOA often interrupts them, making it unclear how they are interconnected.
- **OH Megamasers:** Observations of OH Megamasers hold the potential to trace the merger history of the Universe through cosmic time. ALFA surveys would have an important impact in the detection of such systems at low to intermediate redshifts.

Planned Major Surveys: Addressing the above topics involves balancing areal coverage on the sky against the required integration time. Long integrations allow the detection of a given mass HI source to a greater distance, but at a significant cost: the total volume coverage obtained in a fixed observing time drops inversely as the limiting distance achieved. In other words, integrating 16 times longer achieves 4 times lower noise, which permits detecting the same source twice as far away. However, the volume covered in, say 100 hours, will be only half as big as for a shallow survey that reaches half as far.

If the goal were simply to detect the greatest number of galaxies, the best strategy would be simply to divide the total available integration time uniformly over the area of the Arecibo sky. On the other hand, this strategy will sample only relatively nearby volumes for low-mass sources, potentially telling us nothing about low-mass sources in different environments. A strategy of observing deeply at a smaller number of widely-separated positions will maximize the variety of environments that are sampled. This could be measured by, for example, counting the number of 10 Mpc volumes intercepted by the survey.

Without going into great detail about integration times and survey strategies, we can identify five types of projects that address different aspects of the major science goals. They are summarized as follows; their basic requirements are given in Table 1.

A. All-Arecibo-Sky shallow survey: This survey is designed to maximize the number of galaxies detected. It will help to better establish the HI mass function in the local universe and give the most complete view of the Arecibo sky at 21 cm. ALFA will operate approximately 3 times faster than HIPASS, and will have the advantage of a significantly smaller beam, so fewer follow-up observations will be needed to disentangle confused sources. The all-sky coverage will provide the best opportunity for mapping out HVCs, identifying the HI environment around low- z Lyman- α absorption systems, and will provide low-sensitivity coverage of nearby groups and clusters. This is also the best strategy for finding the rare high-mass sources, and because these are detectable to > 100 Mpc in only ~ 10 s integrations, these will be detectable in a wide variety of independent volumes. In addition, a wide area shallow survey would potentially double the number of known OH Megamasers. *Coverage of 10,000 sq. deg. in single-beam spacing* (for a Gaussian beam solid angle of 15 sq. arcmin — 3.9'x3.4' at HPFW), *with 10 s integrations, will require about 1000 hours of telescope time.*

B. Ultra-deep survey: By integrating 50-100 hours per beam, it is possible to detect relatively high-mass HI sources ($> 10^9 M_\odot$) out to the bandpass limit of ALFA at $z \sim 0.15$. To distinguish between alternative evolutionary models, it will be necessary to detect approximately 40 galaxies. Based on the local density of $M_{HI} > 10^9 M_\odot$ sources, this will require coverage of ~ 0.4 sq. deg. The beam size will be ~ 0.5 Mpc at the limiting distance, so it is likely that galaxies in groups will be confused, but not isolated galaxies. This will be a challenging survey as it will be a severe test of systematic noise sources and it will be fighting strong interference. *Coverage of 0.4 sq. deg. in single-beam spacing will require between 700 and 1400 hours of telescope time.*

C. Medium-depth environmental survey: While the shallow survey will detect the largest number of galaxies of all masses, it will only detect low-mass ($\sim 10^7 M_\odot$) galaxies out to distances < 10 Mpc. This means that the low-mass tail of the HI mass function will only be sampled well within the local supercluster. Existing Arecibo HI surveys have already sampled low mass HI sources out to much greater distances, although with such small numbers of sources detected that the inferred environmental differences remain uncertain. A survey covering 10-20 times the volume of these surveys at comparable sensitivity (~ 1 mJy at 5 km/s resolution) would be able to explore variations in the low-mass tail in a statistically robust way. *Coverage of 2000 sq. deg. in single-beam spacing, at 60 sec integration per beam, will require about 1200 hours of telescope time.*

D. ZOA survey: Exploring galaxies and large scale structure within ~ 5 - 10° of the Galactic plane is distinct from the shallow survey for two reasons. First, Parkes is already surveying the ZOA at Arecibo declinations to a greater depth than in the HIPASS project, so that the gain of a shallow survey would be in resolution, not sensitivity. Second, there are special opportunities to “piggyback” with Galactic and pulsar surveys to achieve high sensitivity levels. This would permit a much more detailed mapping of galaxies in the ZOA than would be driven by extragalactic science goals alone. The parameters for a ZOA survey remain open to compromise with other ALFA surveys’ concerns.

E. Group and Cluster surveys: Several nearby regions and galaxies provide important targets for deeper surveys to explore the low column density HI environment on the outskirts and between galaxies. For these surveys it is useful to target regions that are well studied at other wavelengths to contrast properties of the neutral gas with other populations. In particular, the Virgo Cluster is the nearest cluster and is well-situated in the Arecibo declination range for a detailed study. Other high-density regions within the local supercluster have also been identified that appear to have unusual properties in, for example, the number density of dwarf systems, and several nearby galaxy groups provide an opportunity for exploring the intergalactic medium and draw comparisons with the Local Group. Exploring the environment around nearby isolated galaxies at high sensitivity would provide an important contrast for interpreting environmental effects. *For example, a survey of 400 sq. deg. around a cluster or group, in single beam spacing at 300 sec integration per beam, will require about 1200 hours of telescope time. One important restriction on the scheduling of such a project is that it may require a few to several hundred days of scheduling over a restricted LST range.*

Table 1. Basic Survey Requirements

All-Arecibo Sky shallow survey	10^4 sq.deg.	10 s/beam	1000 hours
Ultra-deep survey	0.4 sq.deg.	~ 300 ks/beam	~ 1100 hours
Medium-deep environmental survey	2000 sq.deg.	60 s/beam	1200 hours
ZOA survey	tbd	tbd	tbd
Group and cluster survey	(400 sq. deg.)	(300 s/beam)	(1200 hours)

The total durations of the surveys listed here are purely indicative. The degree of desirability of Nyquist sampling in solid angle is different for different scientific purposes, thus proposed time requests may eventually vary by factors of a few from the examples presented here. The advantages of drift mapping (telescope immobile during data taking) may in some cases outweigh the higher speed of OTF mode, and multi-purpose, commensal surveys may lead to compromises in the integration times. It should be noted that for most of the surveys proposed, a velocity resolution of ~ 5 km/s is needed to allow for the detection of the narrowest-line HI sources expected and to achieve good resolution of profile shapes. Some of the surveys may be combined or overlapped. For example, the medium-depth survey can be achieved at the lower-frequency end of the bandpass of group and cluster surveys, although this strategy would cover fewer independent volumes than a fixed-declination survey, and there may be problems with optical follow-up for background sources in the vicinity of nearby, bright galaxies. .

3. Follow-up Observations

It is expected that the brunt of follow-up activity after E-ALFA surveys will impact the Arecibo telescope. Single-feed observations at 21 cm of detected sources will be required for a variety of reasons: (i) confirmation of

detection and signal-to-noise improvement; (ii) higher spectral resolution desired; (iii) higher spatial sampling of resolved objects. Especially for the wide angle surveys, which will detect thousands of line sources, follow-up work can potentially generate requests for telescope time commensurate with that assigned to the surveys themselves — depending on how deep the signal extraction algorithms are trained to delve. Continuum and line follow-up Arecibo observations at other wavelengths will likely request relatively small amounts of time.

Another major fall-out of E-ALFA surveys will be given by requests of HI aperture synthesis observations, oriented mainly towards telescopes such as the VLA, WSRT, the GMRT and, in lesser measure, ATCA. At other wavelengths, observations of E-ALFA detections will likely follow at mm telescopes and arrays, while optical and UV photometry and spectroscopy will follow, especially to establish the characteristics of stellar populations associated with the HI objects. The possibility of simultaneous operation of ALFA and SIRTF offers very attractive options for mid- and far-IR follow-up work.

A basic data set that would be highly desirable for cross-referencing with E-ALFA sources would be narrow-band H α information about all sources and at all conceivable redshifts covered by E-ALFA. Such data would allow us to draw an immediate relation between the HI mass and the star formation rate. Narrow band filters are typically about 50 Å wide, so at least 15 filters would be required to cover much of the ALFA z range. Progress in the area of tunable filters presents interesting possibilities in this area.

4. Synergies with Other Surveys

Synergies with non-ALFA Surveys: The HI line observations from ALFA surveys will be combined with information at other wavelengths in order to best characterize the nature of the detected objects. The surveys that will have most relevant information available by the time the E-ALFA surveys will be executed are:

- 1. Optical: the PSS and Sloan surveys. The latter is a five-color imaging survey with the goal of mapping about 10,000 square degrees of the sky to 22 mag at a few percent accuracy and provide spectroscopy of all galaxies down to 19 mag. Sloan partly overlaps with the Arecibo sky. Future surveys with the LSST (scheduled for operation after 2008) will also provide rich cross-comparison with the E-ALFA survey archives.
- 2. Ultraviolet: the GALEX (launched in 2003) all-sky survey will provide imaging and spectroscopy in two bands to nearly the same depth as Sloan, depending on the spectral energy distribution of an object.
- 3. Near IR: the 2MASS survey of the entire sky in three near-IR bands (J, H, and K-short), recently released.
- 4. Far IR: still the old and trusty IRAS, with supplements from ISO. The ASTRO-F mission is scheduled for early 2004 and SIRTF is to be launched in August 2003. These telescopes will provide survey data much deeper than IRAS.
- 5. Radio Continuum: the VLA FIRST survey.
- 6. X-ray: RASS (ROSAT All Sky Survey).

The requirement for efficient use of these surveys in the process of source identification and follow-up is the availability of HI source positions of the highest possible accuracy. In this context, efforts towards maintaining ALFA to a high degree of pointing fidelity are urged. Wide field ALFA surveys will benefit from the natural calibration provided by continuum sources in the mapped fields.

Given the large number of expected detections, the process of cross-correlating source positions with those of different surveys should be automated. Since this is a problem generic to a number of surveys, existing tools will be tapped (see for example the experience at JHU or NVO). Format compatibility of source lists is clearly desired.

Synergies with other ALFA surveys: Multiple backends make commensal observing an exciting possibility that will allow the simultaneous achievement of various science goals while surveying the sky. Other consortia

are aimed at discovering pulsars (P–ALFA) and surveying the Galaxy, primarily in HI (G–ALFA). All consortia have the common goal of surveying the Galactic Plane ($b = \pm 5^\circ$), as well as the entire Arecibo sky to a shallower level. The Galactic Plane is the highest priority for both P–ALFA and G–ALFA; in terms of observing strategy, their flexibility is thus likely to be higher for surveys out of the plane (i.e. commensal with an All-Arecibo-Sky E–ALFA survey). The E–ALFA main goal in the Galactic Plane is a Zone of Avoidance (ZOA) survey, which would favor drift mode observing to most accurately locate optically-hidden galaxies. P–ALFA favors a long dwell time per position. The preferred G–ALFA observing method is drift mode, which is compatible with E–ALFA. The E–ALFA and G–ALFA surveys would also like to plan the survey for multiple passes over each region of sky. This approach would have the advantage of facilitating the discrimination between real sources and RFI, and could be useful for the detection of continuum transients. On the other hand, piggy-backing on a Galactic Plane P–ALFA survey would allow E–ALFA to sample a significantly deeper volume than possible in drift mode. Both options are thus attractive. Observing tests should be completed to determine the plausibility of using multiple pass, drift scanning for part of the P–ALFA survey. G–ALFA and E–ALFA should also work together in doing test observations to determine the best time of day to observe and the characteristics of the ALFA beam.

Both G–ALFA and E–ALFA are interested in online reduction of the data. Since these data reduction methods will have many similarities, these two groups should communicate and hopefully ease the software development burden. The G–ALFA group has the difficult task of dealing with highly extended structures and is planning preliminary observations to explore options for data reduction (e.g. frequency switching vs. position switching). There will most likely be an overlap of velocity coverage for E–ALFA and G–ALFA at the low-velocity end (-500 to 500 km/s) which includes the Galaxy, the Local Group, and high-velocity clouds (HVCs). An understanding will be formed between the two groups on the findings in this overlap region. The E–ALFA and G–ALFA groups desire a minimum velocity resolution of 5 km/s and 1 km/s, respectively. The different requirements of the surveys require separate spectral backends: while E–ALFA wishes to maximize bandwidth coverage, G–ALFA requires high spectral resolution. A parallel E–ALFA plus G–ALFA survey is very unlikely to satisfy both groups, by using the WAPPs as a common set of backends.

5. Hardware Specifications

While the development of hardware for data acquisition that will be employed by the E–ALFA science surveys is the purview of NAIC, it is expected that NAIC and E–ALFA will interact closely on decisions impacting hardware specification and use. Our requirements are summarized as follows:

- High sensitivity will be the key for E–ALFA surveys. Optimization of A_{eff}/T_{sys} across the entire ALFA band to the lowest possible level, with implementation of plans for a tertiary skirt, is strongly desired.
- Several different E–ALFA surveys will be proposed, covering a wide range of desired backend characteristics. While some of the surveys to be proposed may require only 50 MHz of dual polarization coverage for all seven feed horns, other surveys will be impossible without access to 200 MHz of coverage. Table 2 below presents a preliminary set of specifications for an ideal backend system that will satisfy all the currently foreseen E–ALFA survey demands.
- The success of E–ALFA science will strongly depend on our ability to combat RFI. RFI suppression, mitigation and excision both through vigilance and the development of hardware tools fall in the purview of NAIC. Software excision and mitigation tools will be developed by the community and NAIC staff, perhaps jointly. Modes to reduce or compensate for RFI must extend throughout the observing chain, from an IF system with sufficient dynamical range to avoid gain compression in the final signal, to insuring adequate sampling in the backend to avoid spectral degradation, removal of RFI in the software, filtering of the signal, placement of reference horns and providing for cross products between all horns in the ALFA system.
- As the development of backend instrumentation that will satisfy the requests of different sets of ALFA users will be subject to the necessity for compromise, we emphasize the need for transparency in order to maximize the spread of information among different groups. We urge NAIC to maintain an open forum on this issue, facilitating cross-fertilization of ideas.

- The maintenance of state-of-the-art network connectivity at the Observatory and to the outside world is urged for the ALFA era.
- NAIC should archive the raw data, possibly in collaboration with partners, and provide access to these data by members of the E-ALFA science team via network transfer from a **mainland** location.
- Plans must also be put in place that provide adequate hardware for data processing, storage and archiving both within NAIC, observing consortia and, in the case of the final data product archive, the scientific community at large.
- Compatibility with other national observatories is strongly encouraged.

Below we expand on the rationale for the above recommendations.

System Optimization: Continued efforts are encouraged to optimize system performance and to lower the system temperature and spillover, in particular, through the deployment of a tertiary mirror skirt. Likewise, optimization of the A_{eff}/T_{sys} across the entire ALFA frequency band and achievement of uniformity among the outer beams is important to the E-ALFA surveys. A detailed understanding of the variations in beam structure as a function of telescope configuration will be necessary in order to accurately correct for the large impact that the “dirty” ALFA beams will have on the characterization of HI sources, especially those with larger angular extent.

Backend: At the E-ALFA workshop, NAIC committed to delivering a minimum backend system (the expanded WAPPs) for ALFA with bandwidth coverage of 100 MHz at 3-level sampling or 50 MHz at 9-level sampling, in dual polarization mode for each of the 7 feeds. While this system may have acceptable specifications for some of the E-ALFA science projects, other surveys (e.g. the Ultra Deep Survey) will be severely hampered by such a backend. Even an upgrade of such a system to a 2 times 100 MHz bandwidth coverage would fall short for a very deep survey, which would require higher level sampling for RFI excision in an RFI-infested spectral range.

Surveys other than the Ultra Deep one may be successfully carried out by the “minimum” backend, although the opportunity of detection of the very largest HI masses would be reduced by a bandwidth restriction to 100 MHz. Even in these cases, a backend system with high level sampling would be desirable.

Table 2 below outlines the requirements for the “ideal” E-ALFA backend, one that would allow all of the surveys to achieve their science objectives with maximum efficiency. This list meets all currently foreseeable requirements of the surveys identified in Section 1.

Table 2. The “Ideal” E-ALFA Backend

Spec.	Loss for not Achieving	Comments
200 MHz BW	Ultra deep survey cannot be performed	The deep surveys will push ALFA’s limits more than other proposed surveys
8+ bit sampling	Reduction in RFI-excising capability Loss of useful BW	RFI could also be alleviated with (1) Narrow band filters &/or blankers (2) Flexible scheduling
≥ 8192 chan/pol	Inability to achieve science objectives	Max. channel sep. of 5 km/s over 200 MHz
2 Pols per Rx	Loss of sensitivity	
<1 ms data blanking	Time/sensitivity/obs. efficiency loss	<1 ms blanking allows efficient RFI excision
Cross-products & reference horns	Loss in useful BW and efficiency	Require a min. of 2 ref. horns with all X-products b/w ref. horns and ALFA feeds

For optimal limitation of the spectral spread of RFI features, multi-bit sampling is desirable. Preliminary studies indicate that spectrally sharp, strong RFI signals can be constrained within < 3 spectral channels of a correlator backend if ~ 8 bit sampling of the ACF is allowed. It should be noted that backends with 12–16 bit sampling are

already commercially available. Thus obtaining a backend with sufficient sampling to severely limit the signal degradation, in frequency/channel space, of extremely strong RFI sources is technically affordable.

RFI is also known to have an impulsive component with duration on the order of μs – ms. It is still unclear what the recurrence rate of such features is: were it to be high, high dump rates would be required in order to optimally excise RFI in software; if, however, the statistical properties of impulsive RFI were such that the data loss by deleting affected dumps at, e.g., 1 Hz dump rate were negligible, higher rates would not be necessary. Ongoing RFI monitoring at high time resolution by the pulsar community will produce useful data that will help resolve the dump rate question.

RFI Suppression and Mitigation: The presence of RFI is a strong concern because it might severely limit or even prohibit some E-ALFA science. E-ALFA surveys will observe to 1230 MHz where many strong RFI sources exist. Consequently, care needs to be taken to insure the maximum usable bandwidth is available over the maximum amount of time. Appendix C lists the main, known sources of RFI in the ALFA spectral domain at Arecibo.

NAIC must continue to practice due vigilance in minimizing the generation of RFI by on-site devices. We commend the efforts of NAIC staff towards protection of the radio spectrum in national and international spectrum management forums.

Below, we briefly discuss possible avenues in RFI mitigation.

- **Notched Filters.** These would be placed at the system front-end to prohibit certain frequencies from entering the IF path. The main advantage of this approach is the prevention of gain compression. Additionally, notched filters would ease the application of other RFI mitigation techniques discussed below. Among the disadvantages of a notched filter system are (1) the expense of installing 7 notched filter systems on ALFA, and (2) if additional burden of making them switchable, else they would prevent all observations with ALFA at the notched-out frequencies.
- **Radar Blankers.** In the case where the characteristics of an RFI signal are well known, a tuned blanker could be employed, much in the manner of the 1330/1350 MHz blanker now in place at Arecibo. The advantage of a system like this is that it can be turned on and off with ease, and could also be employed only on the E-ALFA backend path, simultaneously not restricting other ALFA surveys. Additionally, radar blankers do not have to compete for space in the ALFA receiver structure, but would be placed downstairs in the signal processing area. The main shortcoming is that the radar population in the ALFA band is large and diverse; thus, total suspension of data taking for each of a majority of radar systems may significantly reduce the on-source duty cycle.
- **Software.** If an RFI signal is confined to a few channels (best if 8+ bit sampling is available), and the data is dumped at a fairly rapid rate, the RFI signal could be removed from the data during the data reduction process, allowing for considerable flexibility in the RFI removal algorithms. Successful application of this technique is optimized with avoidance of gain compression, too slow data readout and sampling at too few levels.
- **Reference horn subtraction.** This hardware addition would considerably aid in the elimination of RFI at the software level. It could result in a fairly clean subtraction, but suffers from the same constraints as the software mitigation listed above.

Network Connectivity to Arecibo and to the data archive site: As in most aspects of astronomical survey undertakings, E-ALFA will place strong demands on network connectivity. We understand that there is currently some risk to Arecibo of the loss of its high-speed Internet connection when the current grant runs out at the end of August 2003. As generally recognized, this would be a disaster, as nearly all aspects of E-ALFA activities, from communication within E-ALFA and NAIC to observing to data transport to archive access, will rely heavily on the Internet.

Because limitations in bandwidth and speed to the island of Puerto Rico are likely to be more stringent than those placed on mainland sites, it is likely that at least the archive offering principal public access, but possibly

all archival storage, will be located off-site. It is critical that the highest performance access is available to the archive wherever it might be located.

Data Storage, Processing, & Transport: An E-ALFA survey will produce raw data at a minimum rate of > 1 GB/hr; this rate is likely to rise for next generation spectrometers. By many standards (e.g., LSST will produce 18 TB/night), this is not a large number. Processed data will be reduced in volume considerably by time averaging, but will have to be stored at a multitude of stages as it proceeds through the pipeline. We understand that raw and Level I data (see next Section) will be archived and maintained by NAIC. While NAIC has been acquiring hardware at sufficient pace to meet current needs, future planning for on-site monitoring, processing and disk storage must accommodate E-ALFA needs both on and off site.

The size of E-ALFA data sets raises issues concerning the mode of data delivery. While NAIC may assume responsibility for archival of level I data products, observing teams will require access to the raw data. For data transfer rates, we assume a minimum data production rate of 0.5 MB/sec. Three principal options exist for the transport of raw data: (a) network transfer, (b) magnetic tape and (c) hard disk. We recommend (a) that network transfer must be enabled to allow real-time access and (b) a collaborative effort between the E-ALFA participants and the NAIC needs to specify and develop a Linux-based hard disk storage/retrieval system that meets survey requirements and is both cost effective and easy to maintain.

Processing Hardware: While data storage presents a manageable problem, data processing for E-ALFA raises considerable challenges beyond the norm for current Arecibo extragalactic observers. Immediate and detailed planning is needed to avoid a data reduction bottleneck due to lack of hardware (plus manpower and software!) resources. NAIC will have to expand its current Linux cluster to accommodate Level I processing tasks. Observing teams need to plan and acquire the necessary high performance systems with sufficient disk storage and backup capability. Compatibility among systems must be insured if the processing task is distributed. Current indications suggest that Linux clusters are preferred, as they can provide the needed capability and architectural flexibility at the lowest cost.

Public and long-term archive hardware: The location and requirements for long-term public access to the E-ALFA data archive need to be specified.

6. Software Specifications

The development of the software necessary to monitor, process, analyze, archive and access survey data and data products presents a critical challenge to the success of E-ALFA. Of necessity, but also of desirability, software development will be a shared responsibility of NAIC and the E-ALFA science teams. Many of their tasks/requirements also apply to non E-ALFA surveys. NAIC, E-ALFA and the other ALFA concerns should work together to develop a systems approach to these tasks, as appropriate and possible, that will minimize cost, effort and the time to complete.

Software development tasks and their responsibilities can be broken into three distinct categories or “levels”, as shown in Table 3.

Table 3. Data & Software Processing Levels

Level	Description	Prime responsibility	Access
I	Observing and first stage processing	NAIC	public
II	Production of public data products	E-ALFA	public
III	Science toolkit	E-ALFA	proprietary?

While the responsibility for the production of observing and calibrating software and lowest level data products is expected to be assumed by NAIC, more elaborate stages of the data processing and associated software will be deliverable by E-ALFA groups. While level I tasks and data products will be mainly the responsibility of NAIC, substantial collaborative input from E-ALFA teams will be available. The basic understanding of who will be responsible for what is summarized in Table 4.

Table 4. Software Responsibilities

NAIC	<ul style="list-style-type: none"> • will provide the tools needed for telescope control, on-line data monitoring, trouble shooting, raw data storage and record keeping needed by the science team: <ul style="list-style-type: none"> ◦ to carry out the observations in an effective and efficient manner and ◦ to conduct real-time assessment of data quality. • will archive the raw data and provide access to it. • will conduct, evaluate and verify a first level of data processing. • will archive the Level I data products in standardized data formats and provide the tools needed to access and use them. • will provide access to the software needed for Level I processing. • will provide the technical information fundamental to the Level II processing (e.g. calibration, rfi identification and characterization, pointing, beam characterization etc.). • will independently verify the Level II data products. • will archive the Level II data products and provide access to them. • will provide access to the software needed for Level II processing.
E-ALFA team	<ul style="list-style-type: none"> • will conduct, evaluate and verify a second level of data processing. • will provide to NAIC the Level II data products in standardized data formats and the tools needed to access and use them. • will provide to NAIC the software needed for Level II processing.
NAIC + E-ALFA	<ul style="list-style-type: none"> • will insure capability for portal to NVO.

Software tasks: The tables in Appendix D give a preliminary itemized listing of software required for E-ALFA. Some portions may be of general use to other ALFA users. The most advanced elaboration, leading to level III data products, might/will be proprietary; diffusion will take place by publication through regular channels. As for public products, access should be provided through NAIC. HI source catalogs might be produced not only by groups associated with the E-ALFA consortium but also by others once public data access is made possible. Because different detection algorithms might yield different catalogs, separate science analysis groups might produce separate catalogs. Once published, they should all be accessible through the E-ALFA archive, with appropriate documentation.

Data formats: Because of the ease of transport and availability of software tools/utilities, raw data should be written and archived in SDFITS format. Further downstream processed data and data products of levels II and III should also be exported in SDFITS whenever appropriate. Advance planning for the physical location and maintenance of the entire archive should begin promptly. A coordinated effort with other ALFA concerns is highly desirable. The development of access tools and the necessary links to the NVO should also be pursued.

Reduction platforms: There are several options for processing data, most notably Classic AIPS, AIPS++, CLASS, GIPSY, IDL and MIRIAD. We anticipate that Level I processing at Arecibo will be conducted using IDL. AIPS++ has been used for Parkes Multibeam data and has some really good tools like the Parkes Multibeam Gridder, as well as developed algorithms for RFI excision. The G-ALFA consortium is discussing similar issues; collaboration in the processing of HI data, wherever the HI might be found, should be encouraged to maximize synergy and avoid duplication of effort. At this stage, there is no unanimous E-ALFA choice of a processing platform. It is, however, stressed that whatever the platform, the processing pipeline should be well documented and exportable, and that output data products conform to SDFITS standards and provide convenient access by a variety of software packages/tools.

7. Organization

The formation of a survey consortium should provide a framework within which extragalactic HI surveys with ALFA will facilitate large science programs significantly greater in scope and ambition than typical Arecibo observing programs. Such surveys will require a large investment of telescope time and hardware development, manpower resources, planning and management. At the same time, such large projects can derive significant benefit through coordination and standardization along the lines of the HST Treasury and SIRTf Legacy programs. As such, E-ALFA surveys carried out under the consortium paradigm should be governed in a manner that balances both the investment of telescope time with the investment of individuals who carry out the surveys. The nature of NAIC as a US NSF-sponsored national center plays a role in defining the paradigm. Both NAIC and the science team will necessarily have to work together closely to insure the success of the science program.

Here, we present a summary of the principles, objectives and circumstances that might form the basis for the organizational structure of an E-ALFA consortium. A detailed list of responsibility assignments between NAIC and the E-ALFA consortium is also presented in Appendix E. In the end, the E-ALFA consortium structure must accommodate the objectives and constraints while enabling the effective accomplishment of science goals. In this draft, the proposed model is not firm, but rather an outline that the interested community and NAIC can use for a proper definition.

The **fundamental principles** that should govern E-ALFA surveys are listed in the Introduction to this document.

The **objectives** of the consortium or survey team should be:

- To achieve the maximum science goals
- To have the broadest possible scientific impact
- To make the best use of the telescope and available resources
- To engage the broadest community of researchers
- To provide the best education and training for students and their teachers
- To raise the broadest level of public interest

The **operational principles** which define the **function** of a “consortium” are:

- To plan, execute and manage the survey observations on behalf of the larger community
- To coordinate the survey for maximum efficiency and maximum impact
- To contribute expertise, software and documentation to NAIC in concert with the Observatory’s efforts on behalf of the consortium.
- To deliver standardized, robust, high quality data products that enable further research

Competition with the other science “consortia” (P-ALFA and G-ALFA) will encourage optimization and compromise. In some cases, the choice to undertake parallel surveys would expedite the allocation of telescope time but it may also introduce inefficiencies: it is the function of the consortia, in concert with NAIC, to argue in favor of the best circumstance for the science output.

Tasks and Assignment of Responsibilities: The E-ALFA consortium and NAIC will work in a collaborative partnership to achieve the success of E-ALFA science. To accomplish the stated science goals, NAIC and the E-ALFA consortium/science teams will work together to plan, design and carry out the surveys, acquire and develop the needed resources, archive the data and data products, and provide community access, educational opportunities and public information and outreach. The assignment of responsibilities, which will be more completely specified in a memorandum of understanding between NAIC and the consortium/science team, is included as Appendix E.

Consortium Organization: As demonstrated by the attendance at the March 2003 workshop, the scientists interested in E-ALFA are a broad and diverse group, driven by a variety of scientific motivations and well-conditioned to multiwavelength studies. Several different E-ALFA surveys are likely to be conducted, which will

assemble different, perhaps partly overlapping, observing teams. These teams may develop their own internal set of operational rules. The E-ALFA “consortium” will evolve into an organization which will include members of the survey teams, as well as representatives of the community at large. A Coordinating Committee should eventually be formed, which will provide the main forum within which will take place a well-tended interplay between different observing teams, between different consortia and between consortia and NAIC.

At this stage, E-ALFA has constituted a preliminary structure which, by election, has chosen three coordinators — two US based (Riccardo Giovanelli and Steven Schneider) plus one based abroad (Lister Staveley-Smith) —, mainly charged with the responsibility of editing this draft and temporarily serving as an interface between the E-ALFA community and NAIC. The evolution of this organization will take place at future gatherings of E-ALFA interested scientists, and influenced by NAIC and other ALFA consortia self-imposed guidelines. In the meantime, ventilation of important issues in the community will be necessary. Among them: the definition of consortium membership, the interplay between consortium members and observing team members, the role of the consortium pre- and during survey work, the definition of timescales for membership, levels of commitment, the structure of a Coordinating Committee, its constitution, nomination/election process and duration of charges.

Preliminary ideas on the above matters have been assembled by members of the “Organizational Matters” Working Group. Their questions and suggestions are enclosed in Appendix F.

8. Education and Outreach Activities

In accordance with the agency requirements of both NAIC and the U.S. NSF-funded participants, education and outreach activities should be actively incorporated into E-ALFA activities. Other countries may encourage or require similar development under grant awards. These efforts should be the joint responsibility of both NAIC and the consortium members. As part of its mission, NAIC should continue to be a driving agent behind the education and outreach initiatives associated with ALFA in general and should assist the E-ALFA consortium in developing materials and programs specifically for E-ALFA. Educational opportunities should be provided for students at all stages of their college and graduate careers, including hands-on training at Arecibo as well as in software development, data processing and analysis at their home institutions. Outreach activities should not only focus on the outstanding Angel Ramos Foundation Visitor Center but also involve the institutions associated with team members and in the various states and countries in which they are located. Examples of activities associated with education and outreach appropriate to E-ALFA and their primary responsibility assignments are included in Table 5.

9. Funding Support

In order to insure success and broad participation in the E-ALFA surveys, participants need access to resources covering all aspects of survey planning, software development, carrying out the observing program, data processing and analysis, education and outreach. Current participants find themselves in varying employment circumstances with varying opportunities for funding.

Factors which complicate financial support for participants of the E-ALFA surveys include:

- the multi-year duration of the surveys
- their legacy nature, that is, that they produce data products that enable further research by a broader community
- the major manpower effort involved in their planning, execution software development and data processing and verification
- the computer hardware resources necessary for data storage, data processing and scientific analysis
- the diverse, international and collaborative nature of the consortium/teams
- the national center character of NAIC and the Arecibo Observatory

While some of these complications are fairly new to extragalactic HI science at Arecibo, they are not new to other large facilities, namely the NRAO, NOAO, 2MASS, SDSS, SIRTf, HST, Chandra, XMM, etc. How participants

Table 5. Example Education and Outreach Activities

NAIC	Scientific workshops	1st one held Mar2003
NAIC	E-ALFA Web site	Technical information for astronomers about ALFA & EALFA
NAIC	E-ALFA Web site	Outreach materials, including E-ALFA science
NAIC	Documentation website	Repository for E-ALFA-specific documentation
NAIC	Software website	Repository for E-ALFA-specific software
NAIC	Observing status website	Updated during survey(s) (Need to specify access?)
NAIC+EALFA	PPT presentation	Set of slides that summarize E-ALFA for general use
NAIC+EALFA	Scientific meeting(s) poster	e.g. AAS June 2004/Jan 2005?
NAIC+EALFA	E-ALFA "Ask an Astronomer"	Possibly use http://curious.astro.cornell.edu ?
NAIC	E-ALFA AOVEF display	E-ALFA-specific display (copies available to others)
NAIC	NVO contact/exchange	Coordinated with G-ALFA?
NAIC+EALFA	Student involvement program	At AO and home institutions
EALFA	Undergrad problems/exercises	To encourage incorporation of E-ALFA into courses
EALFA	Graduate problems/exercises	To encourage incorporation of E-ALFA into courses
EALFA	Graduate seminar	Cornell Astro 620 Spring 2004 (Cordes/Giovanelli/Haynes) "Large Surveys in Radio Astronomy"

in those large survey projects are supported and coordinated varies significantly, and not all paradigms are equally applicable to the circumstances that affect E-ALFA.

The E-ALFA surveys offer enormous potential to explore the nearby universe and to gain insight into how gas disks assemble and evolve. They will produce unique data products that will enable and stimulate follow-up research by a much greater community. Therefore it is in the interest of all parties – the survey teams, the NAIC and the funding agencies - to secure sufficient resources to insure that the full potential of the E-ALFA surveys is achieved.

US NSF Funding The 2000 Astronomy and Astrophysics Survey Committee (AASC) report "Astronomy and Astrophysics in the New Millennium" urged that the funding agencies provide adequate funding for data analysis along with facility support. In a recent experiment, NRAO has begun to offer support for instrumentation development and graduate student thesis research associated with use of the Green Bank Telescope. The latter program is being expanded to the VLA and VLBA within the limits of available funding.

While the NRAO/GBT programs are valuable first steps, they (a) do not apply to software development and (b) do not allow for student tuition payments. Therefore, they alone do not address the needs of faculty at most US institutions. University faculty participants need support for summer salary, travel, capital equipment, publications, student stipends and tuition, as well as applicable indirect costs; in the case of E-ALFA, they may also need an augmentation for communications costs. For a truly level playing field, sources of funding for all of these costs must be identified.

Tenure-track faculty at many U.S. institutions must demonstrate their ability to raise adequate support for all of their research initiatives, including student support. Thus, to enable junior faculty to participate, funding for E-ALFA must be commensurate with that available from other possible sources (e.g., NASA).

The decoupling of E-ALFA support from the grants program funding process places the NSF-supported individual investigator at risk of being granted telescope time but not having the funding resources to carry out the observations, software development and subsequent analysis, or vice versa. In the case of the major survey efforts involved in E-ALFA, covering all phases of the project from concept development through to scientific analysis, the timescale for science results may exceed the grant duration. Because of the diverse nature of the likely E-ALFA survey teams and the importance to the NAIC and the broader astronomical community of the success of these surveys, possible mechanisms for long term continuity of funding, tied to project review, should be explored. A collaborative proposal between NAIC and the consortium and its participants might be the best

fit with the current NSF peer review process.

For full impact, NSF funding should also be provided, as suggested by the 2000 AASC report, for analysis of the E-ALFA Level II data products by the broader community.

The broader impact, educational and outreach potential of the E-ALFA surveys are enormous. Likewise, funding support for those activities should be approached collectively as a partnership by the E-ALFA teams, the other X-ALFA consortia and the NAIC to both maximize and optimize their effectiveness.

EU Funding The international interest in E-ALFA is evident. Specifically, concerning the interest shown from within the European Union: among the 65 E-ALFA consortia members presently registered on its website, 22 are from astronomical institutes in the European Union, representing a total of 11 Institutes in 5 countries (France, Germany, Italy, Spain and the United Kingdom).

The European Commission, within its Sixth Framework Programme (FP6), covering the period 2004-2008, has funds for various kinds of projects. The most appropriate for the E-ALFA project appears to be the inclusion of research themes related to it in a Research Training Network of European Institutes.

The main purpose of such networks is the employment of postdocs, who do not necessarily need to be citizens of a European Union country, thereby also providing potential job opportunities for young astronomers from the US and elsewhere. The proposal for a Network can include other research themes as well. In practice, at most a dozen Institutes can form a network and employ about one postdoc per institute in total.

There are in principle different options for the employment of postdocs under the FP6 Marie Curie Actions: host-driven-actions (Marie Curie Research Training Networks, etc.); individual-driven actions (Marie Curie Individual Fellowships); and excellence promotion and recognition (Marie Curie Excellence Grants, etc.).

The different options of benefit to the E-ALFA Consortium will be investigated further. Information on the FP6 and the Marie Curie Actions can be found on the website (<http://www.cordis.lu/fp6>) of *CORDIS*, the Community Research and Development Information Service.

APPENDICES

A. Scientific Justification: Wide Angle and Directed Surveys

In this section we elaborate on the science justification for the large area surveys.

Introduction: A shallow All-Arecibo Sky Survey will provide a more sensitive, higher resolution (spatial and velocity resolution) look at the 21 cm sky than has been possible with HIPASS or HIJASS and will provide a tremendous improvement in the volume of space surveyed over the Arecibo Dual-Beam Survey (Rosenberg & Schneider 2000), the Arecibo HI Slice Survey (Zwaan et al. 1997), the Arecibo Survey (Spitzak & Schneider 1998), and all of the earlier surveys. The survey will provide much better statistics on the HI mass function than has been obtainable with previous surveys as well as giving a look at the change in the mass function with galaxy environment. The survey will also nicely combine with the data from the Galactic survey for studying high velocity clouds up to high Galactic latitudes and will provide the basis for many other scientific investigations. We discuss the basic parameters of the survey and some of the interesting science that we will be able to do with it here.

Observing Strategy: Our present thought about the observing strategy for this survey would be to get 12 seconds of integration per point using 2 driven 6 second scans. The duplication of scans will greatly improve the interference mitigation and data checking. The main concern with using 6 second scans is the effect of driving the system on baseline stability and noise characteristics of the data. These issues will need to be tested to determine whether this survey strategy will allow us to meet the scientific goals. Presently the Galactic group is considering a double driftscan survey of the entire Arecibo sky which would give twice the observing time discussed here. This increase in observing time would provide greater sensitivity to this survey if we were to piggy-back the observations and would be desirable if it does not detract from the other science that is of interest to the extragalactic consortium.

For some important pieces of the shallow survey science discussed here, including studies of large scale structure beyond $z = 0.1$, searching for OH megamasers at $z \sim 0.25$, and studying absorption line systems at high redshift, the 200 MHz backend is a necessity or, for the case of the megamasers, would greatly improve the quality of the science. Therefore, a 200 Mhz backend is very desirable for the shallow survey science goals.

Simulating Galaxy Detections: John Spitzak has put together a simulation of an ALFA survey with adjustable parameters for the purpose of studying the expected detection statistics for a given survey strategy. In particular we would like to investigate the statistics of galaxy detections as a function of mass and distance as a way to understand how well we will be able to compute the HI mass function since that is a primary goal of the survey. The simulation is designed to include a density map (not yet implemented) to study the statistics of galaxy detections as a function of environment.

The basic outline of the simulation is as follows:

- The Arecibo sky will be divided into eqi-volume, equal velocity-depth cubes which will be populated with galaxies.
- The cubes will be populated with galaxies based on the HI mass function that has been defined (the parameters can be defined for any run of the simulation and can be tied to the density of the cube).

$$\Phi(M) = \ln(10) \cdot \Phi_* M^{(\alpha+1)} e^{-M} \quad (\text{A1})$$

- The scaling of the mass function (and possibly the input mass function slope) will be determined from an estimate of the density in each cube. The major overdensities and underdensities that pass through the Arecibo strip include the overdensities: Virgo Cluster, Local supercluster, Pisces-Perseus supercluster - ZOA, Coma cluster, A1367 cluster, and the Coma Wall as well as the underdensities: Local Void - ZOA, Monoceros Void - ZOA, Orion Void - ZOA, Taurus Void - ZOA, Microscopium Void, Virgo Void, and the Coma Void. Those regions designated with "ZOA" appear to extend into the Zone of Avoidance. Right now the simulation has a uniform density universe, but we will later include scaling based on these structures.

- The integrated flux can be determined for each galaxy:

$$M_{HI} = 2.356 \times 10^5 D^2 \cdot S \cdot W_{20} \quad (\text{A2})$$

- The measured line widths from the Arecibo Dual-Beam Survey are used to parameterize the distribution of line widths for the galaxies. First, there is an upper limit to the line width for a given galaxy HI mass. The upper limit was defined by a fit to the data as:

$$W_{20} = 211.765 \cdot \log M - 1524.706 \quad (\text{A3})$$

the lower limit to the line width is assumed to be 40 km s^{-1} but this is an adjustable parameter. The line width distribution is defined by a Gaussian up to 100 km s^{-1} and by an exponential above 100 km s^{-1} as follows:

$$P(x) = (A/\sigma\sqrt{2\pi})e^{(x-x_0)^2/2\sigma^2} \quad (\text{A4})$$

$$P(x) = Ae^{-a(x-x_0)} \quad (\text{A5})$$

$A=1, \sigma = 100, x_0 = 100, a=0.007$

- The S/N of the galaxy is determined from:

$$\sigma = T_{sys}[\text{Jy}]/\sqrt{t\Delta\nu} \quad (\text{A6})$$

$T_{sys} = 3.5 \text{ Jy}$ (outer beams) $\Delta\nu = 50 \times 10^3 \text{ Hz}$ (after Hanning smoothing)

The effective noise is:

$$N_{eff} = v_{res}\sigma(W_{20}/v_{res})^{0.75}/(300/v_{res})^{0.25} \quad (\text{A7})$$

so the S/N is given by:

$$S/N = S \cdot W_{20}/N_{eff} = (4.24 \times 10^{-6}/\sigma) \cdot M_{HI}D^{-2}(300/v_{res})^{0.25}W_{20}^{-0.75} \quad (\text{A8})$$

$$S/N = 2.7 \times 10^{-4}t^{1/2}M_{HI}D^{-2}(300/v_{res})^{0.25}W_{20}^{-0.75} \quad (\text{A9})$$

We use 8σ as our galaxy detection limit for the simulations in the following sections. The mass function is defined by: $\alpha = 1.5$, $\Phi_* = 0.005$, and $M_* = 7.6 \times 10^9$.

Survey Sensitivity: We have looked at the sensitivity to galaxies in 12 seconds of observation considering only the sensitivity of the outer beams (therefore there will be part of the survey that is deeper). We also have not included the effect of averaging the polarizations in these calculations. Using this estimate of the sensitivities we have found that we should get a S/N of 8 for:

- an M_{HI*} galaxy, width of 200 km s^{-1} at 180 Mpc ($12,600 \text{ km s}^{-1}$)
- a $10^6 M_\odot$ galaxy, width of 50 km s^{-1} at 3.8 Mpc (267 km s^{-1})
- a $10^7 M_\odot$ galaxy, width of 50 km s^{-1} at 12.1 Mpc (844 km s^{-1})
- a $7 \times 10^{10} M_\odot$ galaxy, width 200 km s^{-1} at 600 Mpc ($42,000 \text{ km s}^{-1}$ or the edge of a 200 MHz band)
- a $3.6 \times 10^{10} M_\odot$ galaxy, width 200 km/s at $z = 0.1$ (429 Mpc, $30,000 \text{ km s}^{-1}$)

For a single driftscan (or two driven scans with half the time which would be preferable), interleaving scans to obtain a Nyquist sampled map covering 34° of declination, these observations should take 350 days or 8400 hours. This depth is probably useful for both the shallow survey and the ZOA work so at least for now we will combine the two.

Galaxy Counts from Simulations: The simulation is presently for a uniform density universe, but it gives us a look at the numbers of galaxies we might expect to detect. We detect $\sim 23,300$ galaxies over the whole Arecibo sky out to $40,000 \text{ km s}^{-1}$ (assuming the mass function parameters from Rosenberg & Schneider 2002) and use these results in the figured. Using the HIPASS values (Zwaan et al. 2003), the expected numbers are $\sim 35,300$ due to a significantly higher average density.

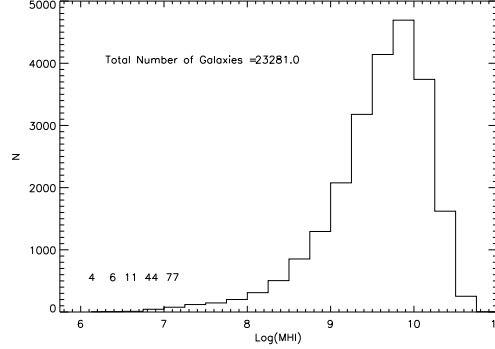


Fig. 1.— The distribution of expected galaxy detections as a function of HI mass. The numbers above the curve on the low mass end of the histogram give the numbers of galaxies in the corresponding bins.

Figure 1 shows the distribution of the galaxies by their HI mass. The numbers above the curve on the low mass end of the histogram give the numbers of galaxies in the corresponding bins. We expect to detect a handful of source at around $10^6 M_{\odot}$, but by $10^7 M_{\odot}$, the statistics should be pretty good and above $10^7 M_{\odot}$ we will have enough sources to study the shape of the HI mass function in different environments.

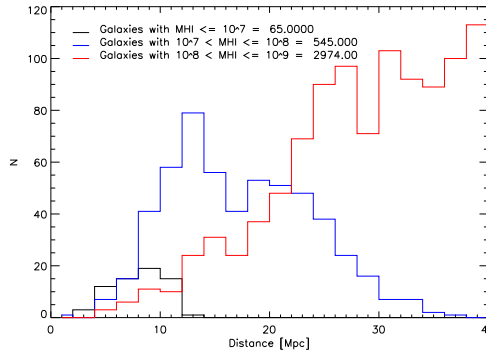


Fig. 2.— The distribution of expected galaxy velocities (distances where H_0 is assumed to be $70 \text{ km s}^{-1} \text{ Mpc}^{-1}$) for low mass sources in the sample.

Figure 2 shows the distribution of distances (expressed in velocity for $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$) for the low mass sources. While the lowest mass sources will only be detected very near by, there will still be a significant number of $10^7 M_{\odot}$ sources detected beyond 20 Mpc (1400 km s^{-1}) where velocity/distance confusion becomes somewhat less significant.

Figure 3 shows the full distribution of galaxy distances in the sample in order to demonstrate the population of galaxies that we would expect to detect beyond $z = 0.1$. There are a significant number of galaxies that should be detectable at these large distances which will be important for mapping out large scale structure.

Comparison with HIPASS and HIJASS Surveys: An All-Arecibo sky survey provides 2 direct benefits over HIPASS and HIJASS with its improved spatial and velocity resolution. The higher spatial resolution will be of benefit because HIPASS and HIJASS science is complicated by the confusion of sources in the large beam. The HIPASS follow-up needed at ATCA are enormous and therefore are limited to the higher flux sources. It will be years before the sources are followed-up (if ever), so an Arecibo survey will not encounter as many issues

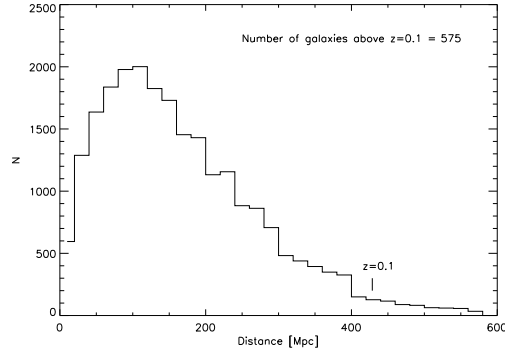


Fig. 3.— The distribution of expected galaxy velocities (distances where H_0 is assumed to be $70 \text{ km s}^{-1} \text{ Mpc}^{-1}$) for all sources.

with source confusion as HIPASS and will therefore be able to do science with the data without time consuming interferometry follow-up. The higher velocity resolution of ALFA will be especially useful in detecting edge on galaxies that have peak fluxes near the noise limit - the edge of a double peak spectrum is much sharper with the higher velocity resolution which should make it easier to automatically detect these sources (and detecting wide things near the noise limit is not easy!).

The HIJASS survey faces similar problems to HIPASS in terms of source confusion and wider velocity channels. HIJASS has an additional problem that there is very bad interference in the $4500\text{--}6000 \text{ km s}^{-1}$ (within which range a lot of the interesting large scale structure e.g., Pisces-Perseus). In addition, HIJASS is not scheduled to observe any more in the Arecibo range (a 4×4 degree region in Virgo and a few other areas have been covered at this point) for the next few years because they are concentrating on the region of the SDSS first release.

Discovering Isolated HI Clouds: With any new survey there is always the chance of discovering new objects. A blind survey like the one we are proposing has the potential to reveal new kinds of objects. One puzzle for numerical modelers of galaxy formation is that the process has been so efficient. Most of the gas seems to reside in the big bright galaxies. Yet, QSO absorption line studies indicate many absorption line systems that do not appear to be associated with bright galaxies. The Virgo survey will enable us to not only explore the faint end of the mass function, but to also look for HI clouds not associated with optical sources.

HI Mass Function Studies: The shape of the HI mass function in the local universe and the cosmological mass density of HI are important parameters for modeling the formation and evolution of galaxies. Blind 21 cm surveys are the only way to determine these parameters, but in the past these surveys have suffered from extremely poor statistics (50 – 300 sources) particularly relative to optical luminosity function surveys which typically sample tens of thousands of galaxies. HIPASS was the first 21 cm survey to have significant galaxy detection statistics (~ 7000 galaxies), but there are limitations with the HIPASS data as discussed in §4.1. The survey proposed here will detect $\sim 11,000$ galaxies between 20 Mpc and 145 Mpc where distance errors are minimized and where galaxy detections are not limited to the most massive sources. With the mass function well determined down to several times $10^6 M_\odot$, we will be able to fix the faint end much better than it has been done up to this point. This survey will allow us to resolve the controversy over the faint-end slope of the HI mass function (Zwaan et al. 2003, Rosenberg & Schneider 2002, Zwaan et al. 1997).

Previous studies of the HI mass function have concentrated on determining an “average” shape of the field mass function. However, we know that galaxy environment effects the galaxy content of galaxies through star formation, tidal interactions, and merging. Because of these processes we would not expect the mass function to be the same in all environments, but the relationship between the HI mass function and environment can not be studied with the small samples that have been available. Only with the much larger number of galaxies that we expect to detect in this ALFA shallow survey will we be able to look at the mass function in different environments and begin to study how the processes which transform galaxies affects the HI distribution function. The combination of the statistical study of galaxy populations with environment and the ability to map out the HI content of galaxies with respect to high density regions like Virgo (see §4.6) will allow us to understand how gas in galaxies

is affected by environment in much greater detail.

Zone of Avoidance Studies: The obscuration due to dust and the high stellar density in our Galaxy varies from place to place within the Milky Way. Overall, it blocks our optical view of the extragalactic universe over about 20 percent of the sky, less in the infrared. This sky coverage limitation does not pose a problem for the study of galaxies themselves, as there is no reason to believe that the population of obscured galaxies should differ from those in optically unobscured regions. However, to understand the Local Group’s motion requires mapping the surrounding mass inhomogeneity, measured in practice by galaxy over- and under-densities, across the entire sky, including the optically-obscured zone of avoidance (ZOA). Further, an accurate knowledge of the mass distribution within our neighborhood is essential if we are to understand the dynamical evolution of the Local Group from kinematic studies (e.g., Peebles et al. 2001), although previous surveys near the Galactic plane indicate we are probably not missing any important local giant galaxies (Kraan-Korteweg, Henning, & Andernach 2000).

As first demonstrated by Kerr & Henning (1987), galaxies that contain HI can be systematically found in the regions of thickest obscuration and IR confusion, through detection of their 21-cm emission. Taking advantage of the velocity information immediately available from the redshift of the 21-cm line, this technique also fills in the ZOA in redshift space, valuable toward mapping the velocity field in the ZOA.

The ZOA in the Arecibo sky cuts through some important known large-scale structures. The major known overdensity probed will be the Pisces-Perseus supercluster. The main voids include the Local, Orion, and Taurus voids, and the edge of the Monoceros void. At this sensitivity (see below) we anticipate uncovering many unknown structures, as well.

Currently, we envision two possible basic survey strategies. First, we consider the style of the “shallow survey”, with 12 sec integration. Scientifically, this is interesting, because it extends to the north the region of the ZOA mapped by the Parkes multibeam ZOA project, and even with this “shallow”, relatively short integration time, goes deeper than Parkes ZOA (e.g. Parkes ZOA sees an M^* galaxy to 70 Mpc, compared to the 180 Mpc estimated for “shallow” ALFA) A shallow ALFA/ZOA survey would be enormously more sensitive than the Dwingeloo Obscured Galaxies Survey, which covered this area but was sensitive only to very nearby objects, and massive spirals to its survey limit of 4000 km s^{-1} (Henning et al. 1998, Rivers 2000). An integration time of 12 seconds with ALFA would uncover a galaxy with HI mass of only $6 \times 10^8 M_\odot$ (linewidth of 100 km s^{-1} , with S/N of 8) at the distance of the Pisces-Perseus supercluster (5000 km s^{-1} , or 71 Mpc, with $H_o=70 \text{ km s}^{-1} \text{ Mpc}^{-1}$). This would allow for excellent mapping of the hidden structures in this rich area, including any backside infall, very important for mapping the extent of over- and under-densities.

If ALFA/ZOA were conducted in the same style as the “shallow” all-Arecibo sky ALFA, then scheduling the two surveys as one optimizes telescope scheduling. Also, we emphasize the desirability of the Nyquist-sampled mapping envisioned for shallow ALFA, which is particularly important for accurately locating optically-hidden sources in the ZOA. If ALFA ZOA is taken as the $|b| \leq 5^\circ$ portion of the Arecibo sky, then the resulting 1000 square degrees would take 700 hours of the 8400 estimated for all-sky shallow ALFA.

Second, ALFA/ZOA could well take advantage of the very similar needs of the GALFA low-latitude survey. As currently envisioned by the GALFA consortium, the Galactic HI Low Latitude Survey would cover the Galactic plane region ($|b| \leq 5^\circ$) in drift mode with Nyquist sampling, and 200 sec integration. Were ALFA ZOA to be conducted simultaneously, we could reach sensitivities 4 times greater than the shallow survey, far surpassing any other ZOA or all-sky HI survey. A 200 sec integration would find galaxies at S/N = 8 with the low HI mass of about $10^8 M_\odot$ at Pisces-Perseus, and $9 \times 10^9 M_\odot$ at $z=0.1$.

Deep Surveying of Nearby Groups and Clusters: There are a handful of nearby groups and clusters that we will be able to study in great detail with this survey, thereby providing the best information on gas-rich galaxies in dense environments that we have to date. We discuss the data we will get in the Virgo Cluster below.

One of the most significant groups that we will be able to study with this survey is the Canes Venatici I Group at about 5 Mpc. Only the southern half of the group is in the Arecibo declination range, but it is extremely rich in low mass HI systems. Given the relative orientations of the Arecibo strip and the Local Supercluster, there is no other region where discovery rates would be expected to come close to what we’ll find in this region. The group is near enough that we will be able to detect galaxies down to $10^6 M_\odot$, thereby allowing us to study low

HI mass sources in high density regions.

In the same region of the sky as the Canes Venatici I Group are Canes Venatici II and Coma I at 10 – 20 Mpc. At these distances we will not be able to detect the $10^6 M_{\odot}$ sources, but we will detect a large number of sources at a few $\times 10^7 M_{\odot}$ in these groups. We will be able to detect sources down to a similar level in the Virgo Cluster as well (see next section).

Additionally, we will be surveying the regions of some other major overdensities including the Pisces-Perseus supercluster, the Coma cluster, the A1367 cluster, and the Coma wall. There are also ~ 20 additional groups (smaller overdensities) at velocities less than 1000 km s^{-1} which we should be able to study in great detail.

Understanding the distribution of gaseous dwarfs in the Local Group is also an important question in understanding the formation and evolution of our own local overdensity. We will be able to survey the Local Group for HI-rich dwarf galaxies. Several known Local Group galaxies, mainly dwarfs, will be observable with this survey including: M33 (Sc; 850 kpc), Leo I (dSph; 205 kpc), Leo II (dSph; 270 kpc), And II (dSph; 680 kpc), And VI (dSph; 775 kpc), LGS 3 (dIrr/dSph; 620 kpc), Peg DIG (dIrr/dSph; 760 kpc), IC 1613 (dIrrV; 715 kpc), and Sextans B (dIrrIV-V; 1.3 Mpc; technically not part of the Local Group). The first 4 of these dwarf galaxies remain undetected in HI. Depending on their star formation history, we may expect to detect gas in the vicinity of these galaxies, i.e. if the galaxy was recently forming stars we may find its star formation fuel is still nearby and only recently stripped from the dwarf. M33 and the latter dwarfs have already been found to contain HI, but this survey will assess the full extent of the gas and determine if there are nearby high-velocity clouds accreting onto these galaxies.

A 21cm Survey of the Virgo Cluster and its Surroundings: The Virgo Cluster is the largest nearby grouping of galaxies and is the obvious place to study the individual and statistical properties of galaxies in the cluster environment. The cluster is probably still assembling itself from a number of smaller sub-groupings though in the central regions galaxy interactions will have been common because of the relatively short crossing time $\approx 0.1 H_0^{-1}$. Virgo is the nearest x-ray cluster with emission being detected as much as 6 deg from the cluster center. This indicates that there is a substantial inter-galactic gas that the galaxies move through and presumably this is what modifies the gas content of cluster galaxies compared to those in the field. Almost all previously published work on the HI properties of the cluster galaxies has been obtained from pointed observations of optically identified sources, here we propose to carry out a fully sampled HI survey of the cluster down to a mass limit of $\sim 10^7 M_{\odot}$. The entire cluster resides in the region observable by Arecibo.

- **Cluster structure:** Previous observations of Virgo indicate a complex velocity structure and a number of possible infalling clouds (Binggeli et al. 1993). Given that the infalling population is likely to be gas-rich compared to the established cluster population the best way of studying and distinguishing between the different sub-groupings is to consider the velocity structure as a function of HI mass.
- **Environmental dependence of the HI mass function:** Optical luminosity functions show large variations as a function of environment with steeper faint end slopes being in clusters (e.g., Lobo et al. 1997). Recently large numbers of faint cluster galaxies have been identified in the Virgo cluster (Sabatini et al. 2003) leading to a cluster luminosity function which is consistent with standard CDM models, but very different to the global luminosity function derived by 2df and Sloan. Galaxies evolve by converting gas into stars and so an interesting way of determining the details of this evolution is to compare the luminosity function and the HI mass function in different environments. Tantalizing evidence of an environmental dependence comes from the Arecibo Dual-Beam Survey (Rosenberg & Schneider 2002) which indicates that the HI mass function is much flatter in the Virgo Cluster than in the field. By considering the relationship between the optical and the HI, we will be able to assess the relative importance of star formation and gas stripping as means of consuming the gas.
- **Beyond the cluster edge:** An HI survey of the cluster would cover also its far edge, which is not well defined, in particular the void behind Virgo. Recent high-resolution simulations in a flat, Λ dominated Universe (Gottlober et al. 2003 MNRAS, submitted) indicate that the voids in the L_* galaxy distributions contain low-mass halos. Their prediction is that a typical void with a diameter of $\sim 20 h^{-1}$ Mpc would contain $\sim 10^5$ objects with mass greater than $10^7 M_{\odot}$. Given that the Virgo survey would be limited in sensitivity to $\sim 10^8 M_{\odot}$ for objects behind the Virgo cluster, we may expect to detect a few hundred such

objects. Their presence in the void, the HI mass function, and dynamical properties would pose significant constraints on modern cosmological models.

Mapping the the Distribution of Luminous and Dark Matter to $z = 0.1$: We will be able to detect $4 \times 10^{10} M_{\odot}$ galaxies out to $z = 0.1$ with this survey (if the hardware allows us to observe to $z = 0.1$). While this is the high mass end of the HI mass function, these galaxies will allow us to trace out the large scale structure at these distances inside and outside of the Zone of Avoidance providing an excellent picture of the large scale features in luminous and dark matter out to $z = 0.1$.

High Velocity Cloud Searches: An All-Sky ALFA HI Survey will greatly improve our understanding of how gas is accreted onto galaxies. Previous surveys of the high-velocity clouds (HVCs) of neutral hydrogen around our Galaxy have been of substantially lower resolution (15.5' at best) and/or were unable to trace the connection between HVCs and Galactic HI emission (Putman et al. 2002; Wakker & van Woerden 1991). An ALFA HVC survey will trace important high-velocity structures, such as the Magellanic Stream and Complex C, at 5 times the resolution of previous surveys. It will also be 8 times more sensitive to unresolved small clouds, or ultra-compact HVCs (if any exist with central neutral column density above $\sim 10^{20} \text{ cm}^{-2}$). Several key questions with regard to high-velocity clouds (HVCs) will be addressed with an ALFA survey, including:

- What is the origin of HVCs? Is there a link between HVCs and Galactic emission? How does our Galaxy obtain fresh star formation fuel? Are HVCs infalling onto the Galaxy (e.g. Tripp et al. 2003), possibly as the remnants of satellites, or does the gas appear to be outflowing in some cases (de Avillez 2000)? Previous surveys have either not been suited to studying both Galactic emission and HVCs, or the spatial and velocity resolution of the survey was insufficient.
- Do HVCs show evidence of interaction with a halo medium? Is there an interface between the cold clouds and a hot halo medium? This question has become especially important of late with the finding of OVI, OVII, and OVIII absorption that appears to be associated with a hot halo medium (Sembach et al. 2003; Rasmussen, Kahn, & Paerels 2003). The OVI absorption often directly traces HVCs detected in HI and appears to be produced via collisional processes (Sembach et al. 2003). The resolution and sensitivity of this survey will allow for a close examination of the head-tail HVCs (e.g., Brüns et al. 2000) and the environment of a large percentage of the OVI absorbers. Recent numerical simulations (Quilis & Moore 2001; Konz, Brüns & Birk 2002) demonstrate that interacting HVCs show a low-intensity tail of diffuse HI gas and a cold compression front at the leading edge. This compression front shows structures on arcminute scales, well observable with Arecibo.
- How many ultra-compact HVCs are there? Previously detected CHVCs and mini-HVCs exhibit a strong correlation between peak column density and angular size such that clouds as small as the Arecibo beam would have peak column densities $< 10^{18} \text{ cm}^{-2}$ (Hoffman & Salpeter 2002) and therefore would not be detectable in a shallow survey. However, the large number of "missing satellites" might be extremely small angular-size, low mass, gaseous objects with detectable peak column density (Moore et al. 1999; Klypin et al. 1999). An 8σ detection of an ultra-compact HVC at 10 kpc, 100 kpc, and 1 Mpc would be 3.5, 350, and $3.5 \times 10^4 M_{\odot}$, respectively. If there is a substantial "missing satellite" population, we would expect to detect a large number of these objects in the general direction of M31/M33. A population of M33 satellites would have a typical angular diameter of about 5 arcmin, if the known (Galactic) CHVCs have distances of the order 100 to 200 kpc. Arecibo is the perfect instrument to detect such M33-CHVCs.

An All-sky ALFA survey will observe several specific high-velocity complexes which are of interest for answering the above questions. The tail of the Magellanic Stream, a high-velocity cloud which is unique in that its origin is clearly known as originating from the Magellanic Clouds, is included in the survey. The tail of the 100° long Magellanic Stream fans out and appears to break up into many small clumps, possibly where it is being overwhelmed by the Galaxy's halo (e.g., Stanimirovic et al. 2002; Putman et al. in preparation). The structure of the Magellanic tail may provide important clues to how satellites lose their gas as they spiral into the Galaxy. The Southern edge of the giant high-velocity cloud, Complex C, is also included in the survey. Since we will be observing Galactic velocities as well, we can trace if and how this complex is feeding star formation fuel into our Galaxy (e.g., Tripp et al. 2003). In addition, the anti-center high-velocity complex, which may represent the gaseous trail of the Sgr dwarf galaxy (Putman et al. 2003), will be observed.

High-velocity clouds have a typical linewidth of 25 km s^{-1} (e.g., de Heij, Braun & Burton 2002). Many HVCs show a two-component profile, with cores with very small linewidths ($2\text{-}10 \text{ km s}^{-1}$) and extended envelopes ($20\text{-}30 \text{ km s}^{-1}$) (Wakker & van Woerden 1997). The velocity resolution of the E-ALFA survey will be 10 km s^{-1} after Hanning smoothing. This resolution will be sufficient to investigate the kinematic structure of the clouds' extended envelopes, especially of the larger complexes. The data from the Galactic ALFA survey will have 1 km s^{-1} resolution. This will allow us to look at the detailed kinematics of some of the brighter clouds and cores.

Searches for Galaxies Near Low Redshift Lyman- α Absorbers: The spectra of QSOs and AGNs are strewn with absorption features from intervening low column density HI along the line of sight, yet their relationship to galaxies is still uncertain. Are these systems related to small, low surface brightness (LSB) galaxies (Impey & Bothun 1997) or to the extended halos of high surface brightness galaxies (Lin et al. 2000; Lanzetta et al. 1995)? Do the absorbers trace filaments of primordial material which are correlated with galaxies because they are overdensities in the same large scale structures (Davé et al. 1999; Stocke et al. 2000)?

At present there are only a handful of low redshift Lyman- α absorbers known and the regions around the higher column density systems ($N_{HI} \gtrsim 10^{14}$) have now been surveyed using Arecibo, the VLA, Parkes, and the Australia Telescope Compact Array. However, HST (the only instrument capable of identifying low redshift Lyman- α absorbers), has been **extremely** restricted in the QSOs it can study because virtually all of them are too faint to allow detailed HST spectroscopy in moderate observing times. This situation is about to change dramatically. In mid-2004 the Cosmic Origins Spectrograph (COS) will be placed aboard HST, allowing high SNR ($R=22,000$) spectra of 17th mag QSOs in only a few orbits; i.e., 10-30 times the throughput of the present spectrograph at comparable resolution. This new spectrograph will dramatically expand the available sample of QSOs that can be observed, thereby producing a new list of nearby absorption line systems. This survey will be an excellent starting point for determining whether there are gas-rich galaxies near the lines of sight to these absorbers.

Absorption Line Studies at High Redshift: The background source counts for this work at 1.4 GHz gives:

- 190 sources/steradian (665 sources in the Arecibo range) brighter than 1 Jy
- 840 sources/steradian (2940 sources in the Arecibo range) brighter than 0.4 Jy
- 5600 sources/steradian (19600 sources in the Arecibo range) brighter than 0.1 Jy

We still have to determine what column densities this will allow us to get to. However, it does seem that by looking in absorption, this survey has the potential to detect galaxies that are not detectable by other means.

Find Rare OH Megamasers near $z = 0.25$ OH Megamasers (OHM) are powerful line sources observed in the L band, arising from the nuclear molecular regions in merging galaxy systems. Approximately 100 such sources are known to date. Several of them are observed to have spectral features variable in strength, an effect which allows keen insights in the source structure and physics. Observations of OHMs hold the potential to being able to trace the merger history of the Universe. ALFA surveys would have an important impact in the detection of such systems at intermediate redshifts. We estimate below the detection rates of OHMs for three possible ALFA surveys.

For a 12-second integration, the expected rms noise in a 150 km s^{-1} channel (the average OHM line width) is roughly 0.67 mJy, and a 5σ detection would be 3.4 mJy. This corresponds to an OH luminosity range of roughly $\log L_{OH} = 2.85\text{-}3.26 L_{\odot}$ for observations spanning 1320–1420 MHz ($z = 0.263\text{-}0.174$). A survey of the entire Arecibo sky with an upper $\log L_{OH}$ cutoff of 4.4 should detect about 130 OH megamasers (we use the OH luminosity function determined by Darling & Giovanelli 2002). Note that this estimate is optimistic because it does not include the tendency for luminous OH megamasers to be broad.

OH detection rates would benefit from extended spectral coverage. An additional 100–200 MHz would enhance detection rates significantly, despite the higher redshifts and dimming of sources.

A Survey for the Community: While the above science will guarantee exciting results from this survey, much of the exciting science that will be done can not be summed up here. There will be many more uses for this survey by the community as a complement to ongoing work. This will be the deepest, highest resolution survey over this area of the sky. By virtue of its unique depth and resolution it will be a major resource. The survey will

be a starting point for anyone wanting to know the HI properties of a galaxy in this velocity range in this part of the sky, it will be a place to start a search for galaxies around low redshift Ly- α absorbers, and much more. As the first resource for this kind of work, it may also increase the interest in the larger community in the use of Arecibo for follow-up work on regions where the survey is not deep enough to answer the questions at hand.

REFERENCES

- Binggeli, B., Popescu, C. C., Tammann, G. A. 1993, *A&AS*, 98, 275.
- Brüms, C., Kerp, J., Kalberla, P. M. W., & Mebold, U. 2000, *A&A*, 357, 120.
- Darling, J. & Giovanelli, R. 2002, *ApJ*, 572, 810.
- Davé, R., Hernquist, L., Katz, N., & Weinberg, D. H. 1999, *ApJ*, 511, 521.
- de Avillez, M. A. 2000, *Ap&SS*, 272, 23.
- de Heij, V., Braun, R., & Burton, W. B. 2002, *A&A*, 392, 417.
- Grebel, E. K., Gallagher III, J. S., & Harbeck, D. 2003, *AJ*, 125, 1926.
- Henning, P.A., Kraan-Korteweg, R.C., Rivers, A.J., Loan, A.J., Lahav, O., & Burton, W.B. 1998, *AJ*, 115, 584
- Hoffman, G.L., & Salpeter, E.E. 2002, in *The Outer Edges of Dwarf Irregular Galaxies, 2002 Lowell Workshop On-line Proceedings*, eds. D. Hunter and S. Oey (www.lowell.edu/Workshop/Lowell02/Proceedings/poster/hoffman.htm).
- Impey, C. D. & Bothun, G. D. 1997, *AR&A*, 35, 267.
- Kerr, F.J., & Henning, P.A. 1987, *ApJ*, 320, L99.
- Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, *ApJ*, 522, 82.
- Konz, C., Brüms, C. & Birk, G.T. 2002, *A&A*, 391, 713.
- Kraan-Korteweg, R.C., Henning, P.A., & Andernach, H. 2000, eds. *ASP Conf. Ser. 218, Mapping the Hidden Universe: The Universe Behind the Milky Way - The Universe in HI* (San Francisco: ASP)
- Lanzetta, K. M., Bowen, D. B., Tytler, D., & Webb, J. K. 1995, *ApJ*, 442, 538.
- Lin, W. P., Börner, G., & Mo, H. J. 2000, *MNRAS*, 319, 517.
- Lobo et al. 1997, *A&A* 317, 385.
- Moore, B. et al. 1999, *ApJ*, 524, L19.
- Peebles, P.J.E., Phelps, S.D., Shaya, E.J., & Tully, R.B. 2001 *ApJ*, 554, 104
- Putman, M.E., Staveley-Smith, L., Freeman, K.C., Gibson, B.K., & Barnes, D.G. 2003, *ApJ*, 586, 170.
- Putman, M.E., et al. 2002, *AJ*, 123, 873.
- Quilis, V. & Moore, B. 2001, *ApJ*, 555, 95.
- Rasmussen, A., Kahn, S.M., & Paerels, F. 2003, in *The IGM/Galaxy Connection*, eds. J. Rosenberg & M. Putman, Kluwer Academic Publishers, 109.
- Rivers, A.J. 2000, Ph.D. thesis, Univ. of New Mexico
- Rosenberg, J. L. & Schneider, S. E. 2000, *ApJS*, 130, 177.
- Rosenberg, J. L. & Schneider, S. E. 2002, *ApJ*, 568, 1.

- Sabatini, S., Davies, J., Scaramella, R., Smith, R., Baes, M., Linder, S. M., Roberts, S., Testa, V. 2003, MNRAS, 341, 981.
- Solanes, J. M., Manrique, A., García-Gómez, C., González-Casado, G., Giovanelli, R., & Haynes, M. P. 2001, 548, 97.
- Stanimirovic, S., Dickey, J. M., Krco, M., & Brooks, A. M. 2002, ApJ, 576, 773.
- Sembach, K., et al. 2003, ApJS, 146, 165.
- Spitzak, J. G. & Schneider, S. E. 1998, ApJS, 119, 159.
- Stocke, J. T., Shull, J. M., Penton, S. V., Gibson, B. K., Giroux, M., L., McLin, K. M. 2000, astro-ph/0009190.
- Tripp, T. et al. 2003, AJ, 125, 3122.
- Wakker, B.P. & van Woerden, H. 1997, ARA&A, 35, 217.
- Wakker, B.P. & van Woerden, H. 1991. A&A, 250, 509.
- Zwaan, M., Briggs, F. H., Sprayberry, D. & Sorar, E. 1997, ApJ, 490, 173.
- Zwaan, M. et al. 2003, AJ, 125, 2842.

B. Scientific Justification: Ultra- Deep Pencil-Beam Survey

Introduction: An ultra-deep survey of an very small patch of sky is proposed. The purpose of this is to probe the gas content and evolution of the most distant galaxies detectable in HI at Arecibo and to investigate the properties of nearby galaxies, the IGM and the Cosmic Web at extremely low column densities.

Galaxy Evolution across Cosmic Time: The rate of conversion of gas into stars is one of the fundamental quantities which describe the evolution of galaxies and remains a difficult measurement to make for all but the most nearby galaxy populations (Madau et al. 1996, MNRAS 283, 1388; Haarsma et al. 2000, ApJ, 544, 641). At redshifts between zero and unity, the global comoving star-formation rate appears to increase by an order of magnitude. Models which attempt to explain the global evolution of galaxies (e.g. Pei & Fall 1999, ApJ, 522, 604) over this redshift range, predict a corresponding sharp increase in the cosmic comoving gas density, with $\Omega_{HI} \propto (1+z)^{3.2}$. However, the limited data we have on the gas density at these redshifts from ultra-violet observations of damped Lyman- α (DLA) systems (e.g. Storrie-Lombardi & Wolfe 2000, ApJ, 543, 552) suggests that the redshift density of absorbers with $N_H > 2 \times 10^{20} \text{ cm}^{-2}$ only increases as $dN_{DLA}/dz = 0.05(1+z)^{1.1}$, which is consistent with constant comoving gas density, if the absorbers do not evolve. This discrepancy may be a limitation of the observations or with the models, or with both. Lyman- α observations in particular are difficult to make at $z < 1.6$ when the line is no longer shifted into the optical regime, making suitably bright background probes harder to find. In addition, the expansion of the Universe makes DLA systems rarer at low redshift. Furthermore, extinction of the background probe by a foreground galaxy associated with a DLA system is a source of possible bias in the measurement of gas density. A much better method, where possible, is to observe HI in emission and to thus directly measure the change in cosmic gas density, without appealing to model-dependent interpretations of the data.

ALFA makes it feasible, in principle, to observe down to frequencies of 1225 MHz, corresponding to a redshift of 0.16, corresponding to a lookback time of $\sim 15\%$ of the age of the Universe. Whilst this is a fairly limited redshift in a cosmological context, it is sufficient to be able to distinguish between the evolutionary scenarios above. Indeed other relatively shallow surveys, such as IRAS and 2dF, have found detected significant evolution of luminosity and comoving number density over similarly modest redshift ranges (for IRAS galaxies, $n(z) \propto (1+z)^3$, Takeuchi et al. 2003, ApJ, 587, L89). However, significant integration times, careful control of observational parameters, a low system temperature, and mitigation of interference are required.

The Cosmic Web: Evidence from QSO absorption line studies shows that the intergalactic medium is populated with low-density ionized gas, forming an important, possibly dominant reservoir of the Universe's baryons (Shull et al. 1996, AJ, 111, 72; Penton et al. 2000, ApJ, 544, 150). The structure of the clouds is still a topic of debate: theoretical modeling suggests that the absorption arises in filamentary structure ('the cosmic web'), whereas some observers believe that the clouds are associated with the extreme outer halos of individual galaxies (Chen et al. 2001, ApJ, 556, 118). The largest cross sections would be provided by highly ionized halos characterized by CIV absorption, followed by mainly neutral Lyman limit/MgII absorbers of HI column densities greater than a few $\times 10^{17} \text{ cm}^{-2}$.

The distribution of the number of absorbers along a redshift path as a function of column density (the $f(N)$ distribution) is nearly a power-law with slope -1.5 over 10 orders of magnitude. Deviations from the power-law occur at column densities where clouds become optically thin to ionizing radiation (around a few times 10^{17} cm^{-2}). This region of column densities is poorly understood in the optical because the Lyman- α line lies on the flat part of the curve-of-growth, and in the radio because typical 21-cm observations do not reach this column density sensitivity. The deviation from the power-law might be related to the ionization edge in the gaseous disks of galaxies (Maloney 1993, ApJ, 414, 41). Our proposed observations reach a 6σ column density sensitivity of $2 \times 10^{16} \text{ cm}^{-2}$ per 5 km s^{-1} for gas filling the beam. The Doppler parameters of the clouds are typically 30 km s^{-1} (Shull et al. 2000, ApJ, 538, 13), which means that they would be resolved with the proposed 5 km s^{-1} resolution. An ultra-deep 'blind' ALFA survey would therefore be ideal for exploring this column density regime.

A deep survey (HIDEEP) has previously been undertaken with the Parkes multibeam system (Minchin et al. 2003, MNRAS, submitted). The ALFA survey has four significant advantages:

1. Depth. The proposed ALFA ultradeep survey has an integration time per beam which is a factor of 30 greater than HIDEEP, giving a column density sensitivity, for a source filling the beam, more than five

times better.

2. Higher resolution. The ultra-deep survey will have much better ($> 4\times$) spatial resolution than HIDEEP, allowing the same column density limit to be applied to sources at $> 4\times$ the distance. It also allows the positions of sources to be found much more accurately.
3. Higher sensitivity to unresolved sources. HIDEEP runs out of sources at around $10^8 M_\odot$, and has nothing to say about low column-density sources below this limit. For the same volume, the ALFA ultra-deep survey will detect objects with $250\times$ lower column density, in the case of unresolved sources with diameters estimated by other (optical, IR) methods..
4. Higher velocity resolution. The ultra-deep survey will be able to detect sources with velocity widths similar to QSO absorption-line systems, which would only contribute to a single channel in HIDEEP.

Frozen Disks: The cosmic ultra-violet background is believed to give rise to ionization edges in the low column-density outer regions of galaxies (as mentioned above). However, an alternative explanation due to Disney & Minchin (2003, in *The IGM/Galaxy Connect*, eds. Rosenberg & Putman, publ. Kluwer Academica, NL, p. 305) is due to the low column-density HI cooling to the temperature of the CMB. Such gas in ‘frozen disks’ would still be visible in absorption, thus contributing to the Lyman limit systems and the Lyman- α forest, but would not be seen in emission. If Lyman-limit or Lyman- α systems are detected in the ultra-deep survey, it would demonstrate that the cut-off seen in galaxies is due to ionization, not ‘frozen disks’. If no such systems are seen, then the survey would lend weight to the ‘frozen disk’ hypothesis.

Observational Goals: A fundamental requirement of the ultra-deep survey is the ability to survey a large volume to the highest practical redshift ($z \approx 0.16$). For galaxy evolution studies, and in order to sample galaxies which contribute most to the mass function at $z = 0$, we need to be sensitive to M_* and sub- M_* galaxies in the range $10^{9-10} M_\odot$. Simple arguments suggest that a sample of ~ 40 galaxies would be appropriate to distinguish, at the 99% confidence level, between models of the form $\Omega_{HI} \propto (1+z)^p$, where $p = 0, 3$. With a linewidth of 200 km s^{-1} , a galaxy with $M_{HI} = 10^{9.5} M_\odot$ has an integrated flux of $0.04 \text{ Jy km s}^{-1}$ and a mean flux density of only 0.2 mJy .

Observing Parameters: The detection of nearby low column density objects and distant galaxies requires long integration times. A possible survey that could achieve the goal of 40 galaxies at the highest emission redshifts possible with ALFA will take 70 hr/beam, would have an rms of 0.05 mJy/beam (260 times deeper than HIPASS) and would survey 0.36 deg^2 , corresponding to 8000 Mpc^3 , in 1000 hrs. The observations would be taken at night in driftscan mode.

Hardware Issues:

- Wide bandwidths (200 MHz) are required for this survey to be effective. A wide band will give us a simultaneous view of galaxies from $z = 0$ to 0.16. Coverage at high- z is necessary to study galaxy evolution. Coverage at low- z is necessary to obtain the highest column-density sensitivities for objects of a fixed size. It will also provide a zero-point for the low- z HI mass function, though it is likely that other ALFA surveys could do this more effectively. The 200 MHz band is not required to be in a single contiguous band. In fact, in the case where limited sampler resolution (i.e. the number of sampler levels) is available, a separate low-frequency band (where much of the interference lies) might be desirable.
- A resolution of 5 km s^{-1} is desired. More might be useful for interference mitigation. This corresponds to 8192 channels, or greater, over 200 MHz.
- Observations at high redshifts will lie near strong interference lines from radar installations. To mitigate against such interference, this project needs a spectrometer with good dynamic range. Currently, we are informed by NAIC staff that between 8 and 12 bits are required of the sampler and spectrometer in order that emissions from strong interferers do not disrupt data at other frequencies.
- The primary interference lines may be able to be removed from the data before or after correlation using a reference horn. We’d like the use of such techniques to be investigated, and we may be able to carry out proof-of-concept observations at Parkes. However, the spectrometer will need two extra inputs and the ability to correlate between 3 and 6 times as much data as for normal autocorrelation mode.

- Radar emissions are impulsive, and an extra degree of immunity against them is possible using blanking techniques. So we would like blanking to be possible at the 1 ms level. Blanking could be automated in the spectrometer hardware, or enforced by applying the phase and period of known radar transmitters.
- Knowledge of accurate calibration factors and beam shapes over a range of zenith angles will be required. Although we propose a drift scan survey, the observing mode will involve chasing the field over a range of hour angles.

C. Information on RFI at Arecibo in the 1230-1430MHz Bandwidth

As emphasized previously, E-ALFA science will be dependent on and possibly limited by our ability to suppress, mitigate and excise RFI. The need for considerable effort in the realm of RFI mitigation techniques and hardware is strongly emphasized.

In this section, we include some relevant information on the RFI environment at Arecibo in the range of frequencies covered by ALFA. This information will be updated as modifications are required and as information becomes available.

Table C.1 US Frequency Allocations

Frequency	Primary	Secondary
1215–1240 MHz	Radiolocation – Government Radionavigation Satellite – Government	
1240–1300	Radiolocation – Government only	Amateur – Nongovernment
1300–1350	Aeronautical Radionavigation – Mixed	Radiolocation – Government
1350–1390	Fixed, Mobile, Radiolocation – Government	
1390–1400	Fixed, Mobile, Radiolocation – Government	
1400–1427	Radio Astronomy – Mixed Earth Exploration Satellite – Mixed Space Research (Passive) – Mixed	
1427–1429	Space Operation – Mixed Fixed, Mobile – Government	Fixed – Nongovernment Land Mobile – Nongovernment only
1429–1432	Fixed, Mobile – Government	Fixed, land mobile – Nongovernment
1432–1435	Fixed, Mobile – Government	Fixed, land Mobile – Nongovernment
1435–1530	Mobile (Aeronautical Telemetry) – Nongov.	

Future RFI at Arecibo

- **The Galileo Project:** (ESA's next generation global navigation system). This will be operating in the ranges of 1164-1300 MHz & 1559-1610 MHz. The current plan is to begin deployment of the satellites in 2006 and have a 30 satellites (full global coverage) by 2008. Further information is available online at http://europa.eu.int/comm/dgs/energy_transport/galileo/programme/phases_en.htm
- **GPS-L2:** GPS L2 is at 1227.60 MHz, just outside the ALFA band. It is switching its usage from military to military + civilian over the next eight years. The GPS L2 signal has a 1.023MHz clock rate which will repeat every 20ms or 1.5s (depending upon the signal being sent). GPS satellites that will transmit the new L2-signal will be launched from 2003 onward. By 2008, there will be 18 modernized so called Block IIR satellites in orbit. Full operational capability with about 28 satellites will be reached three years later. Further information is available online at www.navcen.uscg.gov/gps/modernization/ModernizedL2CivilSignal.pdf or www.geoinformatics.com/issueonline/issues/2002/september_2002/pdf_september/06_07_dejong.pdf
- **Ultra-Wide Band Devices:** UWB devices are now permitted by the FCC. None of the devices currently allowed will fall within the ALFA bandwidth. More information is available in the FCC ruling document available online at http://www.naic.edu/~astro/RXstatus/Lnarrow/fcc_UWB.pdf

Common, known RFI seen at Arecibo in the frequency range 1220–1520 MHz

Frequency [MHz]	Occupancy	Strength ¹	Timing	BW	Identification
1217, 1227.5, 1265.0, 1313.0	–	–	–	–	Pico del Este
1222.32, 1231.28, 1240.24, 1249.2, 1258.16, 1267.12, 1276.08, 1285.04, 1294.00, 1302.96, 1311.92, 1320.88, 1329.84, 1338.80, 1347.76, 1356.72, 1365.68, 1374.64, 1383.60, 1392.56	Always in some mode ²	–	400 μ s+51 μ s	–	Puntas Salinas Frequency Hopping Radar
1241.7, 1246.2, 1256.7	–	–	400 μ s+51 μ s	–	Aerostat
1270.9/1289.9	Occasional (and only one of the two freq. at a time)	–	5 μ s, 12s rotation	–	FPS20-93a radar located in Ramey
1324, 1340	Occasional (only during war games)	–	5 μ s, 12s rotation	–	Naval "landing system"
1330 & 1350	Always	–	1 μ s pulse, 12s rotation	–	FAA Radars (Pico del Este)
1366.3, 1382.7	Occasional (only during war games)	–	–	–	Naval "landing" system
1370.1, 1387.3	Occasional (only during war games)	–	–	–	Naval "landing" system
1381.1	–	–	–	–	GPS L3 satellites

NOTES

- (1) Many of the military radars may no longer be active due to the shutdown of military facilities. A list of the RFI sources which may be removed from the spectrum will be added as it becomes clear.
- (2) The PRANG (Puntas Salinas) Frequency Hopping Radar: This radar runs in the range from 1220 - 1400 MHz and has the potential for completely killing most HI observations at Arecibo. Fortunately, through considerable work by the RFI folks at Arecibo, an agreement has been made between us and the folks at Puntas Salinas. If a request is put in in advance, Puntas Salinas can restrict the frequencies used to just two of a given pair (to be listed later).
- (3) Columns designated by a dash (-) will be filled in.

D. Detailed Lists of Data Products and Relevant Software Tasks

Table D.1 Level I Data & Software Tasks

Item	Requirements/Specs needed
Data Acquisition:	
User Interface (GUI)	Observing modes, displays, remote capabilities
Telescope Control	Motions, rates
Backend Control	Spectrometer setup, dump times, blanking
Radar Blanking	Multiple?
Data Recording	Dump times, formats, headers
Bandpass Calibration	Schemes per observing mode
Position Registration	Encoders, catalog Xreference, pointing info
Beam Pattern	Sidelobe/comalobe per beam as fn (AZ,ZA,freq)
Record Keeping	What info?, access?
Raw data Archive	Capacity, format, location
Basic documentation	E-ALFA specific + general
On-Line Monitoring:	
Monitor display	
Statistics tools	
RFI id & removal	Cross id with hilltop, known rfi
NVSS cross reference	Click to get NVSS id
1beam bandpass	Click to get a bandpass
1beam spectra	Click to get a spectrum
1beam coordinates	Click to get coordinates
Quick-look T-F maps	Time-frequency domain
Quick-look X-Y maps	Sky maps
Level I Processing:	
Bandpass subtraction	
Continuum subtraction	
Flux calibration	K to Jy scale
Higher level RFI excision	Pipeline processing (autoflag)
Astrometry	Real positions on the sky
Gridding	Scheme development
Level I CLEANing	First pass continuum removal
Level I spectra	Individual spectra in Jy-km/s per beam
Level I maps & cubes	"Dirty" maps
Continuum catalog	Flux in Jy/beam at position
Data validation process	
Level I access tools	
Level I documentation	
Level I software export	available for small project ALFA users
Level I archive	Contents? location? access?

Table D.2 Level II Data & Software Tasks

Item	Requirements/Specs needed
Level II Processing:	
Level II CLEANing	sidelobe/comalobe deconvolution?
Level II spectra	calibrated, at any position/nearest grid point
Level II maps	
Level II cubes	
Data validation	statistics/reality checks
Level II archive	archived at/through NAIC
Level II access tools	access through NAIC
Level II documentation	provided by E-ALFA to NAIC
Level II software export	available for small project ALFA users
Portal to NVO	in collaboration with NVO

Table D.3 Level III Data & Software Tasks

Item	Requirements/Specs needed
Detection toolkit:	
Baseline fluctuations	
Noise characteristics	
Sidelobe "cleaning" schemes	
Object finding algorithms	SExtractor, Xcorrelation, etc.
Template fitting	
Injection of "fake sources"	
Extended sources vs. point sources	
Detection documentation	
Data validation process	
HI source catalog	Made available to the public upon publication
General science toolkit:	
Survey simulator	(RG already has some)
ALFA simulator	Simulates what ALFA will see, including multibeaming aspects
Overlay tool	e.g. NVSS, DSS, SDSS, 2MASS
Catalog Xreference tool	ditto
Toolkit documentation	ditto

E. Assignment of Responsibilities

In this section, we detail a proposed assignment of responsibilities as summarized in Section 7. This list, when finalized, could serve as a basis for a memorandum of understanding between NAIC and the survey team(s).

NAIC responsibilities:

1. establishing the processes that will govern other activities, including but not limited to:
 - Proposal guidelines and schedule
 - Proposal review process and criteria
 - Project readiness review
 - Observing time allocation and schedule
 - Survey progress review and criteria
2. Facilitating the formation and activities of survey groups, including but not limited to:
 - Supporting ALFA web-based information dissemination
 - Organizing consortia workshops
 - Facilitating cross-consortia (X-ALFA) discussions
 - Providing on-site educational opportunities
 - Organizing X-ALFA outreach activities
3. Providing the hardware and software needed to acquire data, including but not limited to:
 - Telescope control and data acquisition software, record keeping tools etc.
 - On-line monitoring displays, statistical tools, quick-look tools
 - Beam characterization, calibration tools and information
 - RFI suppression/mitigation, radar blanking schemes, real-time RFI removal
 - Standardized data formats, software for export to standard SDFITS
4. Providing the archive for the raw data
5. Undertaking the first phase of data processing, including but not limited to:
 - Bandpass calibration and subtraction
 - Astrometric registration and gridding tools
 - Level I (first pass) spectra, flat but “dirty” maps and data cubes
 - Continuum catalog
 - Level I data archive
6. Verifying Level I data quality validation
7. Establishing the process for Level II data quality verification

E-ALFA responsibilities:

1. Specifying the hardware and software needs and how they will be contributed by the consortium
2. Specifying the optimal observing strategies
3. Planning and overseeing the observations (with assistance of NAIC staff, by agreement)
4. Assisting NAIC in Level I data processing
5. Developing, and providing to NAIC, the necessary software for Level II data processing
6. Undertaking the second phase of data processing, including but not limited to:
 - Fully continuum cleaned maps
 - Deconvolved spectral line maps (when appropriate)
 - Robust, gridded data cubes
 - Calibrated spectra at nearest grid point
 - HI source catalog
 - Data product access tools

7. Verifying Level II data quality
8. Providing Level II data products to NAIC (or designate institution)
9. Prompt publication of scientific results, HI catalogs

Shared NAIC and E-ALFA responsibilities:

1. Outlining a mutually agreeable framework for the collaboration with a clearly delineated assignment of tasks, delivery schedule and evaluation criteria
2. Insuring that the observations are carried out effectively and efficiently
3. Obtaining NSF support for the surveys for U.S. participants
4. Identifying mechanisms for individuals at non U.S. institutions to participate
5. Specifying the format of standardized data products
6. Archiving the Level II data products
7. Developing a portal to the NVO
8. Establishing mechanisms for graduate thesis work
9. Establishing educational opportunities for undergrads, grads, postdocs
10. Developing educational materials for K-12, undergrads, grads
11. Providing public information and outreach materials

F. Organizational Matters, Working Questions/Proposals

In establishing a framework for the E-ALFA consortium, the following items need to be considered carefully in light of the principles and objectives discussion in Sections 1 and 6.

A. Consortium **definition**:

1. What constitutes a “consortium”?

- One large, open group interested in all E-ALFA surveys
 - What about piggyback surveys with other X-ALFA groups?
 - How is effort allocated/rewarded?
- Hierarchical structure
 - Each E-ALFA survey is undertaken by a separate team
 - Membership in a team constitutes membership in “consortium”, but rules specified by teams.

B. Consortium **membership**:

1. What constitutes “participation in a consortium”?

- Personal effort
- Commitment of additional resources
- Timescales for commitment
- Return on commitment

2. What are rules for “joining”?

- Pre-survey
- Post-survey
- Data/software access groundrules
- “Buy-out” options

3. What are rules for “continuing membership”?

- Can people be kicked out?
- On what grounds?
- Who decides?
- What mechanism for conflict resolution

C. Consortium/team **leadership**

1. Function of leadership

- Survey design and planning
- Resource planning and identification
- Survey management
- Oversight and accountability
- Return on commitment

2. Rules for electing leadership

- Nomination process
- Election process

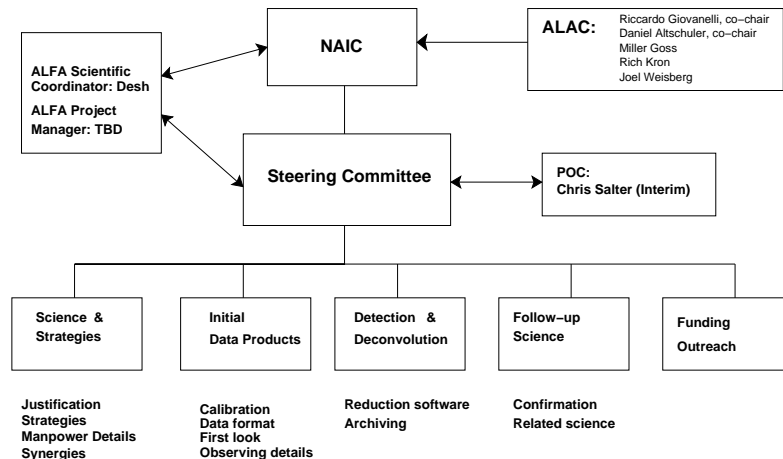
Some (hypothetical) questions whose answers must be clear from the proposed framework include:

1. Suppose a consortium forms that expresses its intent to propose to undertake a specific ALFA survey, but another group, not associated with the consortium for whatever reason, proposes to undertake the same survey. Is the first group guaranteed the telescope time and/or access to NAIC resources? Is the second group precluded from submitting a proposal? Who decides who gets the time, and based on what criteria?

2. Suppose an individual joins the consortium at the outset and plays a major role in the survey and proposal planning and execution (for example, by running simulations, developing deconvolution software, providing computer hardware, executing the observing program etc) but then, for whatever reason, does not participate in final data product delivery or science analysis and is not involved in the consortium later. What commitments are made to this individual by the consortium for early involvement in the project? Who decides? Who insures that commitments are carried forward?
3. The early phases of the project (from now until late 2005) are unlikely to produce scientific results. It thus seems reasonable that it might be unwise for some parties such as untenured faculty, current senior graduate students and current postdocs, to spend significant fractions of their time in the tasks of software and algorithm development and other “investment” activities. How will the balance be struck between adequate return for those who do undertake the groundwork and the later involvement of those who do not?
4. Suppose a survey project starts but after six months, several participants have not fulfilled their commitments, thereby putting the entire consortium at risk of not delivering data/software products on schedule. How is this handled? By whom?
5. Faculty at undergraduate colleges and other primarily-teaching institutions may not be able to commit large amounts of time, particularly during the academic year. How can they and their students be involved given such limitations in time commitment, and possibly, hardware resources and financial support?
6. There can be significant imbalance, particularly in the United States, between scientists at institutions supported on umbrella grants and those who receive support through individual investigator grants with regard to the availability of funds for travel, hardware resources and system support, student support, overhead and fringe benefit costs, etc. Are there mechanisms available through NAIC and/or the consortium concept that can alleviate or minimize the potential imbalance, particularly for the full duration of a survey (planning to analysis)?

F.1. Proposed Organizational Structure

A proposal for a possible organizational structure has been put forward, summarized in the accompanying diagram and governed by the rules below. At this stage, it does not answer many of the questions posed above but serves as a strawman for the purpose of discussion.



1. Steering committee comprised of three people who are voted in by all active consortium members. The chair of the steering committee rotates every six months. After serving as chair, that member steps down and a new committee member is elected. (This results in the first two chairs serving abbreviated terms, but allows for continuity within the committee itself.)
2. All other committees are open to volunteers. The leader of each committee is elected by vote from the active members of that committee. Committees are free to form subcommittees according to their needs.

3. An “active” member is a person whom is, or has been, actively working on some aspect of E–ALFA within the last six months.
4. The steering committee, with NAIC, is responsible for insuring the planned consortium meetings take place, and for organizing SOC and LOC for those meetings.
5. Nominations for replacement members of the steering committee & committee leaders will be given at each meeting and will continue for two weeks after the meeting (via a web-based submission form). Elections will then occur via a secure web form at the end of those two weeks. All active members will be allowed to vote.