#### Lessons for the Future HI Surveys from ALFALFA



<u>Arecibo Legacy</u> <u>Fast ALFA</u> (ALFALFA) Survey





Martha Haynes Cornell University For the ALFALFA team

> ASTRON/PHISCC14 Mar 18, 2014





#### It is a radio "camera"







#### Arecibo L-band Feed Array (ALFA)!





#### Arecibo Legacy Fast ALFA Survey

- One of several major surveys undertaken at Arecibo, exploiting its ALFA multibeam capability
- An extragalactic spectral line survey (mainly HI)
- Covers ~7000 sq deg of high galactic latitude sky
- 1345-1435 MHz (-2000 to +17500 km/s for HI line)
- 5 km/s resolution (100 MHz/4096 channels)
- 2-pass, drift mode (total int. time per beam ~ 40 sec)
- ~2 mJy rms (per spectral resolution element)
- 4400 hrs of telescope time
- Started Feb 4, 2005; completed Oct 26, 2012
- 58 (published) + 5 (submitted) refereed papers to date
- 11 PhDs completed; 9+ underway; + Undergrad ALFALFA Team
- An "open collaboration": let's do science!

#### http://egg.astro.cornell.edu/alfalfa

E A



#### ALFALFA: A 2<sup>nd</sup> generation HI survey

In comparison with opt/IR, the HI view is largely immature

#### ALFALFA:

- Designed to explore the HI mass function over a cosmologically significant volume
  - Higher sensitivity than previous surveys
  - Higher spectral resolution => low mass halos
  - Higher angular resolution => most probable optical (stellar) counterparts
  - Deeper: 3X HIPASS median redshift => volume
  - Wider area than surveys (other than HIPASS) => nearby volumes for lowest M<sub>HI</sub> => <u>cosmologically significant volume</u>

#### We're here to talk about the 3<sup>rd</sup> generation!!!!!!!!!!

LFA



#### ALFALFA: A 2<sup>nd</sup> generation HI survey

In comparison with opt/IR, the HI view is largely immature

#### ALFALFA:

- Designed to explore the HI mass function over a cosmologically significant volume
  - Higher sensitivity than previous surveys
  - Higher spectral resolution => low mass halos
  - Higher angular resolution => most probable optical (stellar) counterparts
  - Deeper: 3X HIPASS median redshift => volume
  - Wider area than surveys (other than HIPASS) => nearby volumes for lowest M<sub>HI</sub> => <u>cosmologically significant volume</u>

#### 3<sup>rd</sup> generation MUCH MORE DEMANDING THAN ALFALFA

LEA



## Arecibo Legacy Fast ALFA Survey

Important issues to consider

- 1<sup>st</sup> priority: Forefront science today for tomorrow
- Best observing efficiency with this telescope/facility
- Survey "Management": Who's in charge now and for the duration?
- Software: Robust, easy-to-use and well documented
  - Survey planning, execution, bookkeeping
  - Processing pipelines (raw => calibrated => images+catalogs)
  - Analysis tools (source extraction/parameters/crosscorrel'n)
  - Data access tools (internal/external)
- Data management plan: Broad use of good data
  - Quality: Bad data is bad; poor data is worse!
  - Archive: What, where, how, who
  - Products: Robust, easy-to-use, and well-documented.
- Counterpart identification: Value-added adds a lot of value.
- Data rights/authorship: The more the merrier with clear rules.
- People and funding: Who's really going to do the work?



- 1. Census of HI in the Local Universe over cosmologically significant volume
- 2. Determination of the faint end of the HI Mass Function and the abundance of low mass gas rich halos
- 3. Environmental variation in the HI Mass Function
- 4. Blind survey for HI tidal remnants
- 5. Determination of the HI Diameter Function
- 6. The low HI column density environment of galaxies
- 7. The nature of HVC's around the MW (and beyond?)
- 8. HI absorbers and the link to Ly  $\alpha$  absorbers
- 9. OH Megamasers at intermediate redshift 0.16 < z < 0.25



- 1. Census of HI in the Local Universe over cosmologically significant volume
- 2. Determination of the faint end of the HI Mass Function and the abundance of low mass gas rich halos
- 3. Environmental variation in the HI Mass Function
- 4. Blind survey for HI tidal remnants
- 5. Determination of the HI Diameter Function
- 6. The low HI column density environment of galaxies
- 7. The nature of HVC's around the MW (and beyond?)
- 8. HI absorbers and the link to Ly  $\alpha$  absorbers
- 9. OH Megamasers at intermediate redshift 0.16 < z < 0.25



- 1. Census of HI in the Local Universe over cosmologically significant volume
- 2. Determination of the faint end of the HI Mass Function and the abundance of low mass gas rich halos
- 3. Environmental variation in the HI Mass Function

Dictate requirements for:

- Sky coverage (a lot!)
- Bandwidth (enough redshift coverage; avoid strong RFI)

- Velocity resolution (low mass galaxies)
- Sensitivity (volume sensitivity)
- These in turn dictate observing strategy @ Arecibo
  - Drift scan, 2-pass, "minimum-intrusion" strategy



# Drift scanning

-5.9 dB



-3.0 dB

8.5 dB



#### ALFALFA Survey 2005-12





- 1. Census of HI in the Local Universe over cosmologically significant volume
- 2. Determination of the faint end of the HI Mass Function and the abundance of low mass gas rich halos
- 3. Environmental variation in the HI Mass Function
  - Get some science done EARLY!
    - Specific areas to target first (e.g. Virgo)
    - Partial analysis to verify science requirements being met

- Demonstrate progress to periodic reviews
- Unexpected delays (very frustrating):
  - Proposed for 4 year survey; time allocated over 6.5
  - Some LST blocks allocated to other program
  - Lost several months/semester due to maintenance (telescope painting; ALFA modification)



## Maximize Observing Efficiency

- Telescope time is precious and competition for it is stiff.
- ALFALFA requires a lot of telescope time and generates a lot of data (modest by 3<sup>rd</sup> generation standards)!
  - Arecibo and ALFA are complex instruments to use.
- ALFALFA science goals demand high quality data.
  - The legacy nature of ALFALFA raises the standards for data product generation and delivery. RFI is nasty and inevitable.
- The A2010 observing proposal was approved pending periodic reviews of our ability to perform the survey.

The ALFALFA technique delivers >99% "open shutter" time. A few early science goals identified; those areas observed first & papers/PhD theses completed. Optical counterpart identification done simultaneously.



## Arecibo Legacy Fast ALFA Survey

- Important issues to consider
- 1<sup>st</sup> priority: Forefront science for tomorrow
- Best observing efficiency with this telescope/facility
- Survey "Management": Who's in charge for the duration?
- Software: Robust, easy-to-use and well documented
  - Survey planning, execution, bookkeeping
  - Processing pipelines
  - Analysis tools
  - Data access tools
- Data management plan: Broad use of good data
  - Quality: Bad data is bad; poor data is worse!
  - Archive: What, where, how, who
  - Products: Robust, easy-to-use, and well-documented.
- Counterpart identification: Value-added adds a lot of value.
- Data rights/authorship: The more the merrier + clear rules.
- Manpower and funding: Who's really going to do the work?





#### Survey Management: Who?



- Point-of-contact w/ telescope operations
- Oversee execution of survey observations
- Monitoring data quality => short-timescale feedback
- Oversee software development, data management, data product development
- Oversee documentation, bookkeeping, long-term history
- Monitor progress in support of survey science goals
- Communicate with team and with scientific community
- "Manage" science programs (avoid conflicts, protect students, maximize output/impact)
- Undergraduate ALFALFA team "management" ("broader impact" important to U.S. funding schemes)



## Arecibo Legacy Fast ALFA Survey

- Important issues to consider
- 1<sup>st</sup> priority: Forefront science for tomorrow
- Best observing efficiency with this telescope/facility
- Survey "Management": Who's in charge for the duration?
- Software: Robust, easy-to-use and well documented
  - Survey planning, execution, bookkeeping
  - Processing pipelines
  - Analysis tools
  - Data access tools
- Data management plan: Broad use of good data
  - Quality: Bad data is bad; poor data is worse!
  - Archive: What, where, how, who
  - Products: Robust, easy-to-use, and well-documented.
- Counterpart identification: Value-added adds a lot of value.
- Data rights/authorship: The more the merrier + clear rules.
- Manpower and funding: Who's really going to do the work?





#### Software pipelines: From 2D to 3D



# A Drift scan, before bandpass correction (bpd)





A Drift scan, after
bandpass correction (bpd)

FA

🛞 🛞 🏈

#### Useful GUIs and VO tools



## Mining ALFALFA: signal detection



HI flux density sensitivity depends on HI line width => but <u>well-behaved</u>

> Amélie Saintonge (Cornell) PhD thesis

- Source finding done in Fourier domain using matched filter algorithm (Saintonge 2007, AJ, 133, 2087)
- Use f>1420 MHz range for tests
- Biggest problems: Baselines/RFI



## Mining ALFALFA: signal detection

g

{n

Baseline ALFALFA computation was designed to run on 2005 vintage desktop computers.

While the source extractor was part of the pipeline, the source parameter measurement was done by an astronomer.

This was feasible for ALFALFA. We wouldn't recommend it for Generation III.

But: pay attention Be prepared to have to make compromises for reasons of practicality. What does the science really demand?



## Mining ALFALFA: spectral stacking



Silvia Fabello (MPA) PhD thesis

Fabello+ 2010 MNRAS 411, 993 Fabello+ 2011 MNRAS 416, 1739 Fabello+ 2012 MNRAS 427, 2841 Hallenbeck+ 2012 AstroJ 144,87

> ALFALFA pipeline tracks RFI/continuum (weights)=> Spectral stacking to dig deeper



## Mining ALFALFA: spectral stacking



Silvia Fabello (MPA) PhD thesis

Stacking is really powerful: Plan on it!

Fabello+ 2010 MNRAS 411, 993 Fabello+ 2011 MNRAS 416, 1739 Fabello+ 2012 MNRAS 427, 2841 Hallenbeck+ 2012 AstroJ 144,87

> ALFALFA pipeline tracks RFI/continuum (weights)=> Spectral stacking to dig deeper



#### Connection to other databases



## Arecibo Legacy Fast ALFA Survey

- Important issues to consider
- 1<sup>st</sup> priority: Forefront science for tomorrow
- Best observing efficiency with this telescope/facility
- Survey "Management": Who's in charge for the duration?
- Software: Robust, easy-to-use and well documented
  - Survey planning, execution, bookkeeping
  - Processing pipelines
  - Analysis tools
  - Data access tools
- Data management plan: Broad use of good data
  - Quality: Bad data is bad; poor data is worse!
  - Archive: What, where, how, who
  - Products: Robust, easy-to-use, and well-documented.
- Counterpart identification: Value-added adds a lot of value.
- Data rights/authorship: The more the merrier + clear rules.
- Manpower and funding: Who's really going to do the work?





# **Identifying Optical Counterparts**

- ALFALFA source centroids good to ~18" (depends on S/N) ALFALFA catalogs include:
  - the HI centroid position
  - the position of the most probable
     OC
  - OC's SDSS PhotoObjID and SpecObjID (where applicable)

#### Of 15855 sources in a.40:

- 1013 have no OC
- 844 of those could be HVCs (or UCHVCs/LG minihalos)
- 199 (<2%) extragalactic</li>
- Of those, <50 are "isolated" "(Almost) Dark Galaxies"
- Talk by Gyula "Josh" Josza
- Posters by Karen Lee-Waddell & Luke Leisman







#### The data management plan



- Level 0: The raw data, recorded at the telescope
- Level 1: Bandpass subtracted, calibrated, RFI-flagged
- Level 2: 3D grids (line) + continuum maps, "flatfielded" Continuum used to update astrometry/flux scale
- Level 3: Catalogs, spectra, optical counterparts

#### For ALFALFA

- Minimal replication (disk I/O)
- Data products kept small => laptops
- Total amount of data < 15T (trivial...!)</li>
- Unclear issue: permanent archive



## Arecibo Legacy Fast ALFA Survey

- Important issues to consider
- 1<sup>st</sup> priority: Forefront science for tomorrow
- Best observing efficiency with this telescope/facility
- Survey "Management": Who's in charge for the duration?
- Software: Robust, easy-to-use and well documented
  - Survey planning, execution, bookkeeping
  - Processing pipelines
  - Analysis tools
  - Data access tools
- Data management plan: Broad use of good data
  - Quality: Bad data is bad; poor data is worse!
  - Archive: What, where, how, who
  - Products: Robust, easy-to-use, and well-documented.
- Counterpart identification: Value-added adds a lot of value.
- Data rights/authorship: The more the merrier + clear rules.
- Manpower and funding: Who's really going to do the work?





#### Harvesting ALFALFA

- ALFALFA is astronomer-power-limited.
- The solution has been to allow more people to join





#### The ALFALFA team: An open collaboration







Heavy student involvement: • 11 PhDs to date • 9+ PhDs underway •166 ugrads (so far!)

LFA





# Team guidelines



#### Joining ALFALFA

ALFALFA is an open collaboration. Anyone with an interest in the science that can be done with the ALFALFA dataset and the willingness to contribute to the collective effort is welcome to join. While the ALFALFA team plan to make the data available to public access on a short time scale through periodic data releases, direct participation in the collaboration offers the chance to interact directly with other team members. In the spirit of collaboration, members of the team exchange technical and scientific expertise, gain understanding of the data sets in earlier stages of processing, provide training in data processing tasks and contribute to setting priorities as the project moves forward.

All <u>science projects</u> by collaboration members are posted on the ALFALFA website, to inform others of who's doing what and on what timescale. Being a member of the team means sharing in the responsibilities of the effort. See the recommended guidelines for science projects and authorship.

If you have an idea that could benefit from ALFALFA data and think that joining the collaboration may be beneficial for its development, contact Riccardo Giovanelli (riccardo@astro.cornell.edu, 607-255-6505).





#### **Recommended Criteria for ALFALFA Science Projects Authorship**

The following guidelines were adopted following the ALFALFA team meetings in May and June 2005.

0.0 A member of the ALFALFA collaboration who wishes to work on a specific project that relies on ALFALFA data should inform the Oversight Committee (OC). A description of the project scientific goals, instruments to be used and extent of use of ALFALFA data should be indicated, as well as a timetable for completion of the project. The plan of work will be posted on the ALFALFA website and other ALFALFA group members informed. Any member of the group interested in the project should feel free to contact the provisional project leader and propose to join, provided that a useful contribution can be made within the given timetable. The ultimate constitution of the team is however up to the project leader. Progress reports will be required; continuation of project leadership will be linked to consistency with the proposed schedule. Posted projects, especially PhD thesis projects, will thus be "protected" within the collaboration, and external users of the data will be advised of the ongoing internal activities of the group.

0.1 In case of partial conflict of interests between different proposed projects, differences will be preferentially resolved amicably by team members. The OC can act as arbitrator if required.

0.2 In preparation of publication of ALFALFA and ALFALFA-related results by group members, an authorship list should be sent to the OC. The OC may thus advise or recommend modifications, for increased fairness.

0.3 One category of group members should be considered potential co-authors on all ALFALFA and ALFALFA-related results: that of individuals who : (a) acted for more than 100 hours of telescope time as "designated observers", within the last year or

(b) processed and inspected more than 100 hours of telescope time worth of data, within the last year

or

(c) made major contributions to the HI data processing and management software.

0.4 In the publication of strictly ALFALFA HI data, results or catalogs, the authorship list will include those in 0.3. Inclusion of any others is up to the consideration of the team leader.

0.5 For a paper that uses publicly posted but scientifically unprocessed data, such as source lists, level 1 or level 2 data sets, the same considerations described above apply, if the initiators are ALFALFA collaboration members.

0.6 For a paper that uses published ALFALFA data, the usual criteria of professional courtesy apply.

0.7 For all publications, an effort should be made to protect the projects led by PhD students, and to emphasize their contributions. For the papers which are a substantive part of students' theses, "dilution" of credit should be minimized, i.e. members of the collaboration who didn't make a substantive contribution to that specific project should be encouraged to exclude themselves from authorship.

0.8 "Surprise" or "unexpected" discovery reports: "Hot" sources will occasionally be found during various stages of data inspection. The discoverer will have the option of first authorship after clearance with the OC, whether the finding is the result of blind luck or planned search. Discovery papers of "hot" sources found through automated signal identification algorithms will be led by authors selected with a "merit" criterion that involves level of commitment to observations, data processing, software development and other overall contributions to the survey.



#### The Arecibo Legacy Fast ALFA Survey

Main People Science Schedule Data Documentation Links Publications Undergrads Non-experts News/Events Observing/Data Team

#### **Publications related to ALFALFA (Most recent first)**

#### **Refereed papers**

- 63. The HI Mass Function and Velocity Width Function of Void Galaxies in the Arecibo Legacy Fast ALFA Survey Moorman, C.M., Vogeley, M.S., Hoyle, F., Pan, D.C., Haynes, M.P. & Giovanelli, R. 2014, MNRAS (submitted) preprint available after acceptance
- 62. HIghMass High HI Mass, HI-rich Galaxies at z ~ 0: Sample Definition, Optical and Hα Imaging, and Star Formation Properties Huang, S., Haynes, M.P., Giovanelli, R., Hallenbeck, G., Jones, M.G., Adams, E.A.K., Brinchmann, J., Chengalur, J.N, Hunt, L.K., Masters, K.L., Matsushita, S., Saintonge, A. & Spekkens, K. 2014, Ap.J. (submitted) preprint available after acceptance
- HIghMass High HI Mass, HI-rich Galaxies at z ~ 0: High-Resolution VLA Imaging of UGC 9037 and UGC 12506 Hallenbeck, G., Huang, S., Haynes, M.P., Giovanelli, R., Adams, E.A.K., Brinchmann, J., Chengalur, J.N, Hunt, L.K., Masters, K.L., Saintonge, A. & Spekkens, K. 2014, A.J. (submitted)

preprint available after acceptance

60. ALFALFA Discovery of the Nearby Gas-rich Dwarf Galaxy Leo P. V. Neutral Gas Dynamics and Kinematics

Bernstein-Cooper, E.Z., Cannon, J.M., Elson, E.C., Warren, S.R., Chengalur, J.N., Skillman, E.D., Adams, E.A.K., Bolatto, A.D., Giovanelli, R., Haynes, M.P., McQuinn, K.B.W., Pardy, S.A., Rhode, K.L., Salzer, J.J. 2014, A.J. (submitted) preprint available after acceptance

59. Distance Determinations to SHIELD Galaxies from HST Imaging

McQuinn, K.B.W., Cannon, J.M., Dolphin, A., Skillman, E.D., Salzer, J.J., Haynes, M.P., Adams, E.A.K., Cave, I., Elson, E.C., Giovanelli, R., Ott, J. & Saintonge, A. 2014, Ap.J. (accepted) astro-ph/1402.3723

58. ALFALFA Discovery of the Nearby Gas-Rich Dwarf Galaxy Leo P. IV. Distance Measurement from LBT Optical Imaging

McQuinn, K.B.W., Skillman, E.D., Berg, D., Cannon, J.M., Adams, E.A.K., Dolphin, A., Giovanelli, R., Haynes, M.P., Rhode, K.L. & Salzer, J.J. 2013, AJ 146, 145



#### The Arecibo Legacy Fast ALFA Survey

Main People Science Schedule Data Documentation Links Publications Undergrads Non-experts News/Events Observing/Data Team

#### **Publications related to ALFALFA (Most recent first)**

#### **Refereed papers**

63. The HI Mass Function and Velocity Width Function of Void Galaxies in the Arecibo Legacy Fast ALFA Survey Moorman, C.M., Vogeley, M.S., Hoyle, F., Pan, D.C., Haynes, M.P. & Giovanelli, R. 2014, MNRAS (submitted) preprint available after acceptance

62. <u>HIghMass</u> - High HI Mass, HI-rich Galaxies at z ~ 0; Sample Definition, Optical and Hα Imaging, and Star Formation Properties

Huang, S., Hynes, M.P., Giovanelli, R., Hallenbeck, G., Dnec, M.G., Adams, D.A.K., Brinchmann, J., Chengalur, J.N, Hunt, L.K., Masters, K.L., Matsushita, S., Saintonge, A. & Spekkens, K. 2014, Ap.J. (submitted) preprint available after acceptance

61. HIghMass - High HI Mass, HI-rich Galaxies at z ~ 0: High-Beselution VLA Imaging of UGC 9037 and UGC 12506

Hallenbeck, S., Huang, S., Haynes, M.P., Giovanelli, R., Adams, E.A.K. Brinchmann, J., Chengalur, J.N, Hunt, L.K., Masters, K.L., Saintonge, A. & Spekkens, K. 2014, A.J. (submitted)

preprint available after acceptance

60. ALEALEA Discovery of the Nearby Gas-rich Dwarf Galaxy Leo P. V. Neutral Gas Dynamics and Kinematics

Sernstein-Cooper, E.Z., Cannon, J.M., Elson, E.C., Warren, S.R., Chengalur, J.N., Skillman, E.D., Adams, E.A.K., Polatto, A.D., Giovanelli, R., Haynes, M.P., McQuinn, K.B.W., Pardy, S.A., Rhode, K.L., Salzer, J.J. 2014, A.J. (submitted) preprint available after acceptance

59. Distance Determinations to SHIELD Galaxies from HST Imaging

McQuinn, K.B.W., Cannon, J.M., Dolphin, A., Skillman, E.D., Salzer, J.J., Haynes, M.C., Adams, E.A.K, Cave, I., Elson, E.C., Giovanelli, R., Ott, J. & Saintonge, A. 2014, Ap.J. (accepted) astro-ph/1402.3723

 ALFALFA Discovery of the Nearby Gas-Rich Dwarf Galaxy Leo P. IV. Distance Measurement from LBT Optical Imaging McQuinn, K.B.W., Skillman E.D., Berg, D., Dannon, J.M., Adams, E.A.K., Dolphin, A., Giovanelli, R., Haynes, M.P., Rhode, K.L. & Salzer, J.J. 2013, AJ 146, 145

#### Lessons learned from ALFALFA



The HI population is not the optical population

- Virtually all SF galaxies contain HI.
- Low mass, HI-dominated galaxies are important!
- ... So there is great science to be done; do some of it EARLY.

A successful survey needs to be "managed":

- The team needs to be efficient and effective, to show steady progress, and to recognize needs of all.
- The observatory/funding source needs to be committed to survey success.

All successful surveys will be astronomer-power -limited.

- New scientific doors will be opened, beyond that planned.
- The harvest will be most bountiful with more farmers.

A successful survey will produce great science but will require a lot of long-term commitment and hard work

• The pay off will be in the accomplishment: "more fun than human beings should be allowed to have"

