Astronomical Data Analysis Software and Systems III
ASP Conference Series, Vol. NNN, 1994
D. R. Crabtree, R. J. Hanisch, and J. Barnes, eds.

The VLA's FIRST Survey

Robert H. Becker

Department of Physics, University of California, Davis, CA 95616, and Lawrence Livermore National Laboratory

Richard L. White

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218

David J. Helfand

Astronomy Department, Columbia University, 538 West 120th Street, New York, NY 10027

Abstract. Over the next decade the Very Large Array (VLA) will carry out a systematic survey of the northern sky at 20 cm wavelength in two configurations. We have selected as a moniker the acronym FIRST, which in uncompressed form reads Faint Images of the Radio Sky at Twenty-cm. The high-resolution survey will be done in B configuration and hence will achieve an angular resolution of 5 arcsec. It will cover a 10^4 square degree region centered on the north Galactic pole. Each field will be observed for ~ 3 minutes with a resulting rms noise level of ~ 0.2 mJy. The survey will result in a catalog of 10^6 discrete sources as well as $65,000$ images each composed of 4×10^6 pixels. To expedite the massive data analysis task this project entails, we have decided to utilize computers.

It is our intention to release to the community compressed, machine-readable copies of all the images as well as an annotated catalog of sources. The challenge will be to accomplish this with a minimum of resources. We hope to automate maximally the data analysis pipeline to achieve this end. This paper describes the history of the project, its current status, and our data analysis methods, followed by a brief resumé of FIRST's expected scientific impact.

1. Introduction

New astronomical surveys almost invariably herald advances in astrophysics. Surveys play a crucial role in identifying new classes of objects, in revealing structure in the Universe, and in expediting studies of individual objects. Each new generation of instrumentation usually warrants a new survey. Hence the ROSAT Sky Survey replaces the HEAO-I Sky Survey, the 2MASS survey (Kleinmann 1992) will replace the AFGRL survey, etc. Since radio telescopes tend to be unique, there are few matched surveys in the northern and southern hemispheres. The only reasonably uniform sky survey in the radio band is the 5 GHz survey carried out at Greenbank and Parkes; it is the best existing radio survey, with a limiting flux density of ~ 30 mJy and positional accuracy of ~ 1 arcmin. Both the relatively high flux density limit and the relatively large positional uncertainty limit the usefulness of this 5 GHz database. With this in mind, the National Radio Astronomy Observatory (NRAO) has decided to utilize the VLA to produce a next generation radio sky survey.

2. History of FIRST

There is anecdotal evidence that astronomers have been talking about a VLA Sky Survey since the VLA's inception. In the spring of 1990, having just completed the analysis of a VLA Galactic Plane survey, Rick White, David Helfand, and Bob Becker in a mood of elation were contemplating their next move and, as may have happened many times before to other astronomers, began fantasizing about a VLA Sky Survey. But this time, instead of just going out for a drink, we convinced ourselves that such a project was truly feasible. Three months later, we submitted what was, to the best of our knowledge, the first serious proposal to survey the sky at 20-cm wavelength with the VLA. The proposal called for observing $\sim$125,000 fields for 90 s each in a hybrid BC configuration with the objective of surveying all the sky north of $-30°$ declination. The proposal generated a great deal of discussion among the staff at NRAO, but ultimately was rejected without formal review. But the stage had been set.

Nine months later, NRAO received two all-sky proposals; a revised version of our original proposal, this time in C configuration, and a D-configuration proposal from Jim Condon. In fact, this was a healthy development. In writing a C configuration proposal, we were well aware of the scientific compromises involved. The primary difference among VLA configurations is angular resolution: the FWHM ranges from 54 arcsec in D-configuration to 2 arcsec in A-configuration. A secondary consideration is available bandwidth (high-resolution images must be made in spectral line mode). High resolution images are desirable because they facilitate the identification of optical counterparts. Low resolution images are desirable because they are more sensitive to extended, low surface brightness sources. The submission of both high and low resolution proposals insured a serious discussion of the best way to carry out a VLA survey.

NRAO sent both proposals out for review and there appeared to be a consensus that the time was right for a VLA Sky survey, but a great deal of uncertainty as to which survey to do.

To facilitate an objective, optimal decision, trial observations were scheduled in the spring and summer of 1992 to compare the merits of survey results from C and D-configuration observations. Two scientific panels were convened in October 1992 and March 1993 to review the situation and make recommendations to NRAO. The outcome of these two meetings was a surprise to everyone involved. Rather than sacrifice the advantages derived from either a high or a low resolution survey, it was decided that two surveys would be made, one in D configuration and one in B configuration. The D-configuration survey was chartered to cover 85% of the sky to a 2 mJy flux density limit. The B-configuration survey was chartered to cover the 25% of the sky centered on the North Galactic Pole (NGP; chosen to coincide with the Sloan Digital Sky Survey [Gunn & Knapp 1993]) to a 1 mJy flux density limit. Observations for both surveys would start in 1993. The rest of this paper will be restricted to discussion of the high-resolution, B-configuration survey.

3. Why a High Resolution Survey?

The VLA is a versatile instrument. It is capable of observing at six wavelengths in four different configurations. It is not sufficient to decide to do a survey, one must also decide which survey to do. The choice of a frequency was an easy one to make. The field of view of the VLA scales linearly with wavelength (inversely with frequency.) A survey at 6 cm would require $\sim$10 times as many observations as a survey at 20 cm covering the same area of sky. The same logic would suggest that the VLA's longest wavelength, 90 cm, would be even better than 20 cm. However, the surface density of bright sources is much higher at 90 cm so that a single snapshot of data is insufficient to achieve the high dynamic range required for reaching the theoretical sensitivity of the VLA. Furthermore, the 90 cm receivers are more than an order of magnitude less sensitive than the 20 cm receivers.

The relatively high resolution of the B configuration ($\sim$5 arcsec) can also be justified despite the extra difficulty that high-resolution imaging entails. There are two primary drawbacks to going to higher resolution. A B-configuration image has $\sim 10^2$ times as many resolution elements per unit area as a D-configuration image, hence 10^2 times as much mass storage is needed to archive images and the computational time to produce each image increases by a factor of ~ 20. In addition, VLA images will degrade if the bandwidth chosen is too wide (bandwidth smearing). The farther apart the antennas, the smaller is the acceptable bandwidth. In B configuration, this constraint forces the observations to be made in spectral line mode, resulting in the loss of 50% of the available bandpass compared to continuum mode.

However, the scientific advantages are worth the operational disadvantages. In particular, the high resolution results in better positional accuracy for all sources found in the survey and, for a significant fraction of sources, results in morphological classification based on a source's brightness distribution. Much of the power of the radio survey will come from the association of individual radio sources with optical counterparts. But unambiguous association demands a radio error box small enough to exclude random coincidences, where "small enough" depends on both the density of potential optical counterparts and the fraction of optical sources that have radio emission. Approximately 50% of the radio sources in the FIRST survey will have optical counterparts brighter than 23rd magnitude. At this level, sub-arcsecond positions are necessary for reliable identifications.

The morphology of the brightness distribution of a radio source is also important for making a correct optical identification. Many sources are complex; for example, radio triples are common. In such cases, the optical counterpart will be located at the position of a weak central component, far from the brighter radio lobes. Without high resolution, misidentification is likely. The morphology is also important for classifying the nature of the radio emission (e.g., distinguishing between FR I and FR II radio sources.)

4. The Spring, 1993 B Array Observations

During April and May 1993, ~ 120 hours of observations were scheduled. To achieve a flux density limit of 1 mJy (5σ), each field was observed for ~ 3 minutes. The data were taken in spectral-line mode using two 25 MHz bandpasses, each divided into seven 3 MHz channels in two polarizations. The field centers are distributed in the sky in a variable hexagonal grid with an average spacing of 26 arcmin between nearest neighbors. A total of ~ 2200 fields were observed in a narrow declination strip $2.75°$ wide passing through the NGP and spanning $\sim 9^{\rm h}$ in RA.

5. Data Analysis

The data analysis procedures required to generate high-resolution, widefield images from B-configuration VLA data are computer intensive, and at first glance, are beyond the capability of computers available to astronomers. Imaging the full primary beam requires a 2048×2048 image, but ignores the likelihood of bright sources in sidelobes beyond the first zero of the beam pattern.

Proper self-calibration of the data normally requires mapping all the sources contributing to the observed visibility function. To avoid repeatedly creating such large images we have taken advantage of existing radio catalogs, in particular the Greenbank 1400 MHz survey (White & Becker 1992). The first step in self-calibration is to generate a tapered map of the primary beam and satellite maps of all the outlying sources found in the Greenbank catalog out to $10°$ that will appear brighter than 2 mJy. The resulting maps are searched for sources and a source list is written to a disk file. These sources are then used as input to a self-calibration script which makes small maps around each and every discrete source, thus avoiding the need to make large images. With this methodology we can self-calibrate ~ 50 fields a day using a Sparc 10/41 processor. In general, we do a 3 iteration phase self-calibration on any field containing at least one source brighter than 30 mJy. All the observations from Spring 1993 have already been self-calibrated.

Once the data are self-calibrated, the field is imaged into a 2k by 2k map using the AIPS task WFCLN. We typically clean the images to a flux density level of 0.5 mJy ($\sim 3.5\sigma$). A Sparc 10/41 processor with ~ 25 megabytes of memory can create ~ 25 images/day. The images so produced are distorted because of the 3-D nature of the Celestial Sphere in contrast to the 2-D FFT done by WFCLN. But, since the observations were made as short duration snapshots, the distortion is correctable using the AIPS task OHGEO which transforms the image onto the correct coordinate grid. The overlapping areas of adjacent images will be summed to increase the sensitivity of the survey. All the images from Spring 1993 will be created by January 1994. As of October 1, we have generated 600 images. In Figure 1 we show the distribution of peak flux densities and rms noise levels for these images. The vast majority of the images ($\sim 95\%$) achieve the theoretical noise limit. In Figure 2 we display one of these images to illustrate what the community can expect from the survey.

Although the necessary software does not yet exist, we intend to search the final images to extract discrete sources and construct a survey catalog. The catalog will contain source positions accurate to better than an arcsec, flux densities, and morphological information. We also plan to distribute the images in compressed format (White, Postman, & Lattanzi 1992; White 1993) so that the entire survey will reside on several CD-ROMs. A sample compressed image is shown in Figure 3.

6. Scientific Potential of the FIRST Survey

The most important discoveries that have been made using the POSS — quasars, starburst galaxies, the large-scale structure of the Universe — were not anticipated by those who carried out the original survey. And many of these discoveries have come as a result of comparing work done at other wavelengths with this unparalleled optical database. Likewise, we cannot anticipate all of the new source classes and the astrophysical insights that FIRST survey will yield, but we are confident that cross correlation of the maps and catalogs we produce with data from other wavelength regimes will be enormously productive. To maximize the utility of the survey, it has been designed to cover the same 10^4 square degrees of sky as the Sloan Digital Sky Survey (SDSS) which will, by the end of the decade, have produced five-color CCD images of this entire area to $m_v = 24$ and will be in the process of collecting spectra from 10^6 of the objects therein. The subarcsecond positions our survey will produce will allow immediate optical identification of $> 50\%$ of our radio catalog from the SDSS database, further enhancing its utility in infrared and high energy astrophysics. While an exhaustive description of the scientific impact this project will have is beyond the scope of this paper, we outline below a few examples of areas in which the FIRST survey will make fundamental contributions.

6.1. Quasars and active galaxies

At 20-cm flux densities above ~ 3 mJy, active galactic nuclei dominate radio source counts. In the thirty years since the discovery of quasars, nearly 10,000 galaxy systems with extraordinarily powerful, compact nuclei have been catalogued, and progress in understanding the mechanisms which power them has been substantial. Nonetheless, we still argue about density vs. luminosity vs. luminosity-dependent density evolution for quasars, and estimates of the typical quasar lifetime are uncertain by roughly two orders of magnitude. The connection between IRAS starburst galaxies and quasars, the importance of mergers, the origin of the powerful γ-ray emission recently discovered by CGRO, the contribution of various classes of objects to the cosmic X-ray background, the importance of relativistic beaming for BL Lacs and their connection to radio galaxies, as well as generalized "unified models" of AGN are all topics of current active research. Work toward answers for each of these outstanding questions will be materially aided by a deep, uniform, large-area radio survey.

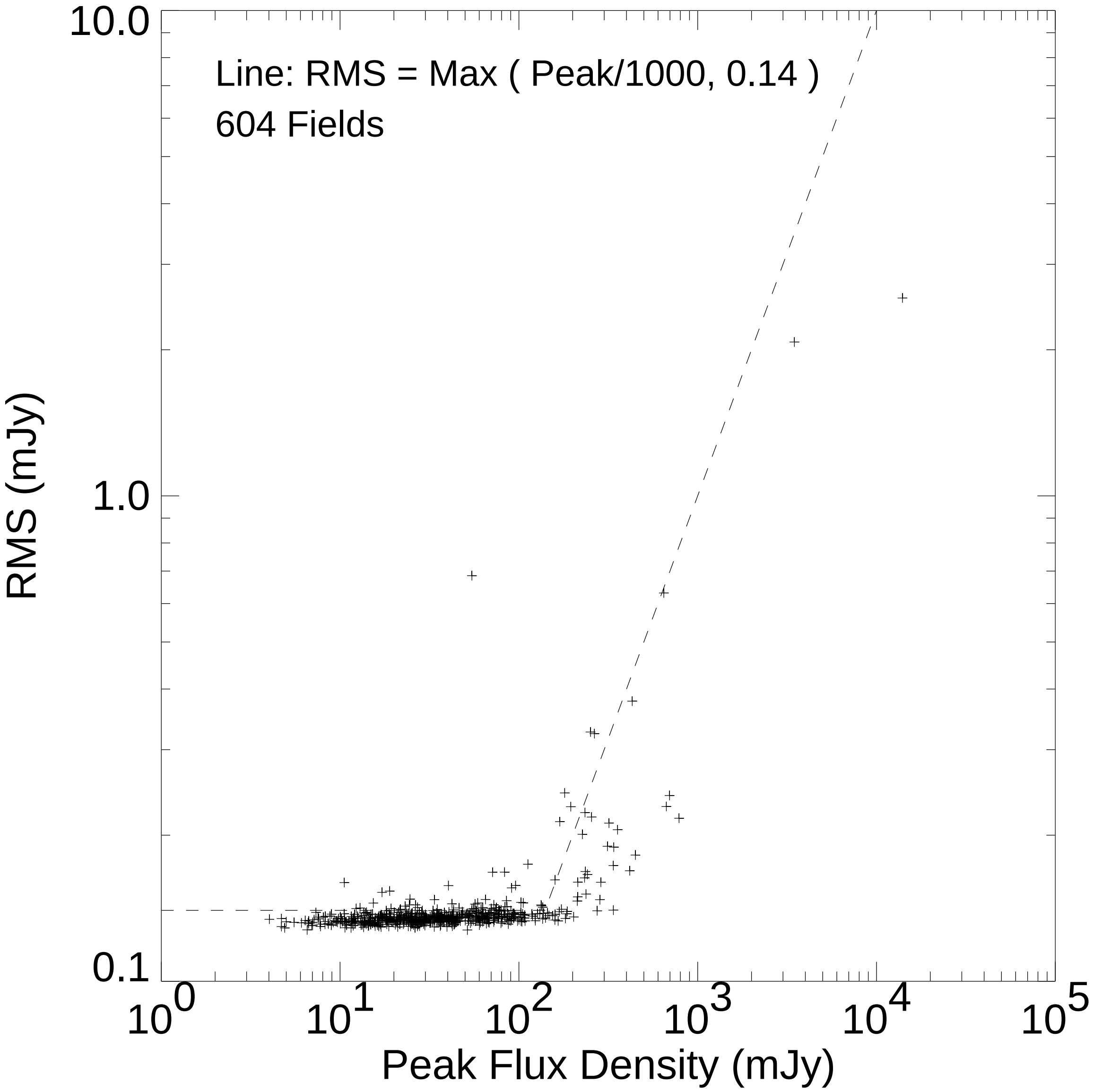

Figure 1. RMS noise in 604 cleaned maps versus peak flux density in the map for fields calibrated and mapped using the automatic procedure described in text. Fields with sources brighter than ~ 150 mJy are dynamic-range limited, while other fields reach the rms limit set by thermal noise in the 20 cm VLA receivers (~ 0.14 mJy). Only 1 field out of 600 has failed to reach an acceptable rms level and will require manual intervention; such a low failure rate is certainly acceptable.

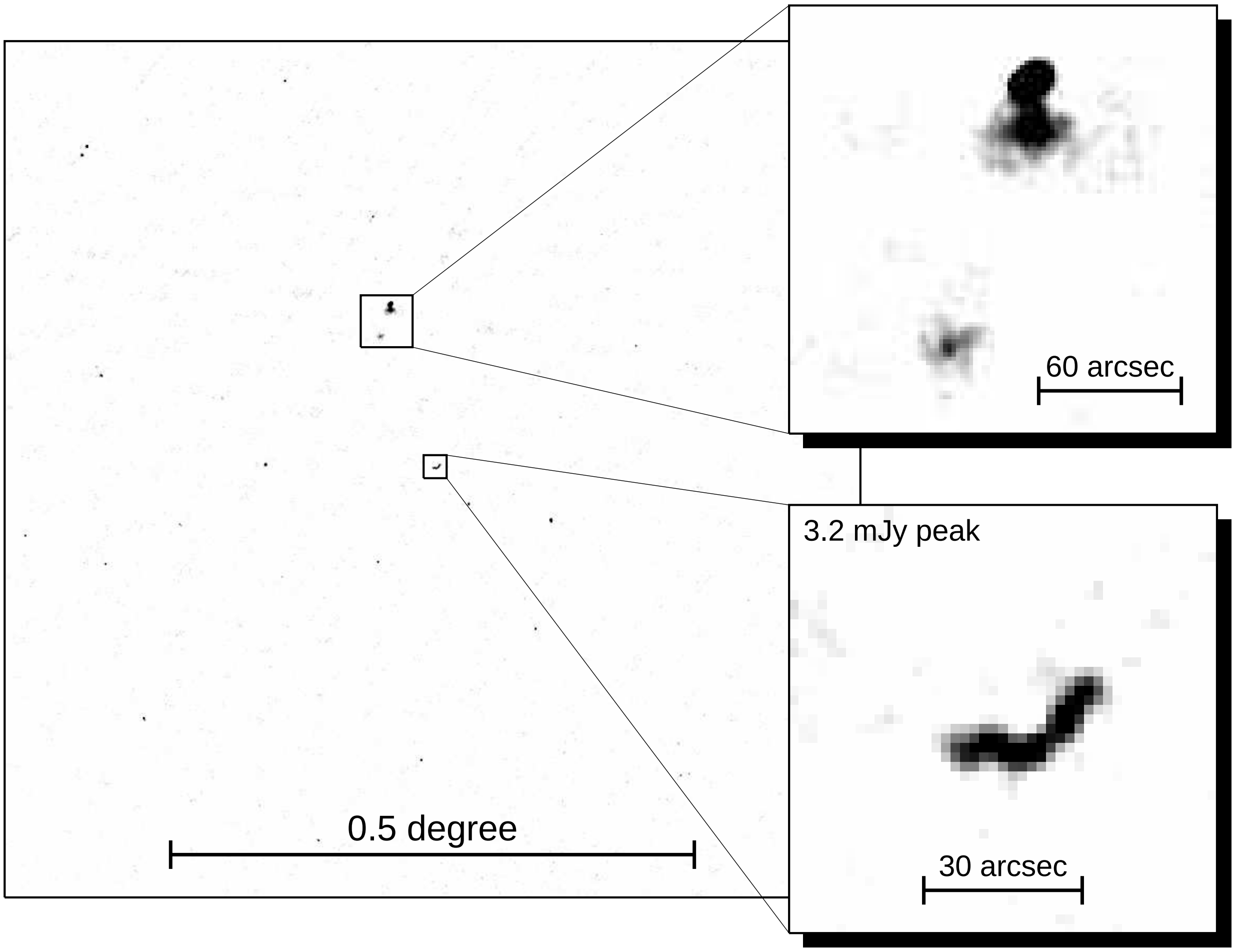

Figure 2. Sample map for the field 10330+30456 generated by our automatic procedures. The large image has 1634 × 1634 1.8 arcsec pixels; the smaller images show details of two interesting sources.

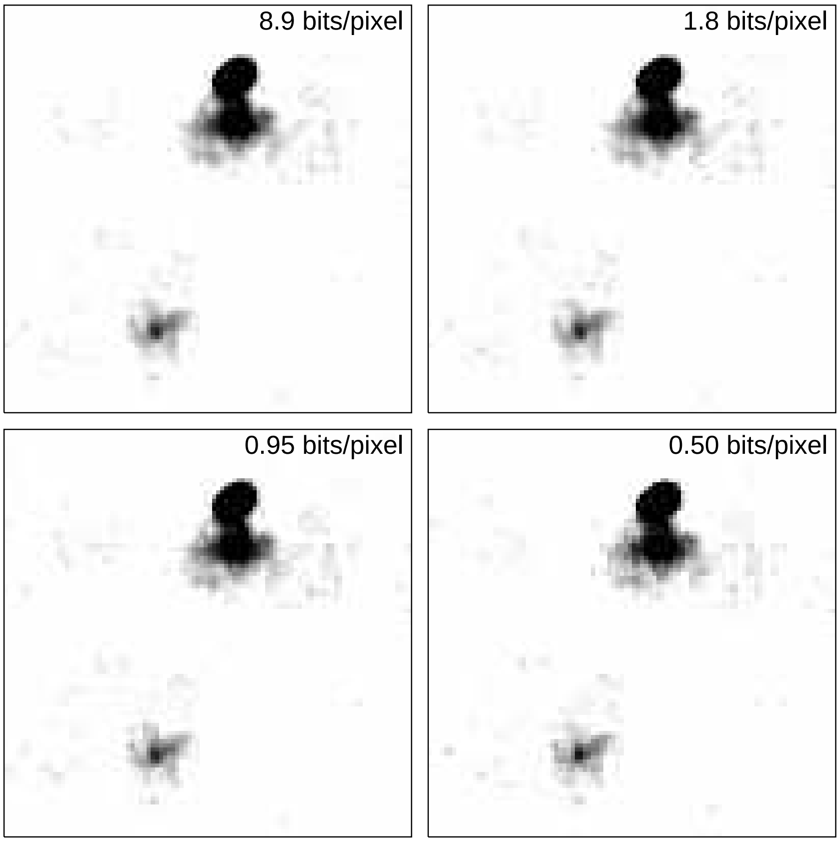

Figure 3. Compressed versions of the image in Fig. 2. The original image has 32 bits/pixel. The 8.9 bits/pixel image differs from the original by at most 0.5 μJy. At 1 bit/pixel the B-configuration images covering 10,000 deg^2 will fit on $\sim$8 CD-ROMs.

Using the Cambridge University Automated Plate Machine (APM) scans of the POSS (McMahon, private communication) and data from a 25 deg^2 pilot for our VLA survey, we have demonstrated that, using optical color selection plus radio identification, we will be able to generate immediately a complete, flux-limited catalog of $\sim 20,000$ radio-loud quasars brighter than $B = 19.5$ for use in evolution studies. Since this sample will be optically bright, it will also be ideal for detailed spectroscopic studies. Cross-correlation with the ROSAT all-sky survey will increase the currently small sample of known BL Lacs by more than an order of magnitude from a little over 100 to several thousand; furthermore, since the 1 mJy limit of the survey is equivalent to the lowest flux density ever found from X-ray selected samples of these objects (Stocke 1990) and is far below the threshold for radio-selected objects, this sample will also be complete and flux-limited. The enormous benefit this new sample will have for the study of all objects in which relativistic beaming is important can be gauged by the progress brought about by the Kuhr *et al.* (1981) complete radio sample, the flux density limit of which is 1 Jy, 10^3 times our threshold. In all, the FIRST survey will detect over half a million AGN, and most will have optical counterparts in the SDSS. None of the current problems in the field will be immune from attack with this powerful resource.

6.2. Galaxy evolution

Below about 3 mJy, the radio source $\log N - \log S$ slope changes markedly, indicating the presence of a new population. Work over the past decade (Windhorst *et al.* 1985; Oort 1987; Thuan & Condon 1987) has demonstrated convincingly that this population consists largely of actively star-forming galaxies and, as such, holds an important key to the subject of galaxy evolution. At 1 mJy, the majority of our sources will be starburst galaxies at redshifts from 0.1 to 0.6 and, perhaps, far beyond (*e.g.*, IRAS 10214+4724 at $z = 2.3$ [Rowan-Robinson *et al.* 1991]). Every galaxy in the IRAS Faint Source Catalog will be detectable and, with the resultant subarcsecond radio positions, immediately identifiable. Many fainter objects, detectable only by future far-IR missions such as ISO and SIRTF, will also be identified, distinguished from AGN by their optical and radio extents. Questions such as the importance of mergers to star formation history, the shape of the IMF, the extent of mass-loss in supernova-driven winds, and the importance of local environment in galaxy evolution will be addressable with this enormous sample. Moreover, the combination of the radio data from our catalog with data from IRAS, CGRO, SDSS, etc., will allow the discovery of many rare objects such as IRAS 10214+4724, which often hold vital clues to the behavior of less extreme systems.

6.3. Large-scale structure and dark matter

The evolution of large scale structure in the Universe, from the era probed by the COBE observations of the microwave background to the present, is one of the central issues of modern astrophysics and, owing to the important role of dark matter in this evolution, of particle physics as well. The enormous amount of work undertaken over the past decade to map the local large-scale structure has resulted in a relatively clear picture of the distribution of luminous matter at redshifts $z < 0.1$. The challenge now is to extend our knowledge of this structure to much higher redshifts so that its evolution becomes apparent. For high redshifts, our quasar sample will provide a useful start on the largest scales, but it is in the redshift range 0.1–0.5 that real progress will be forthcoming by tracing the matter distribution with our starburst sample. Other avenues of approach include the detection of a large cluster sample through selection of head-tail radio galaxies, and the selection of large-separation gravitational lens candidates via an optical filter on the sample of close doubles (demanding equal relative intensities for the two components in the two wavelength bands).

6.4. Galactic astronomy

A half dozen categories of radio-emitting stars provide us with information on topics ranging from main sequence and pre-supernova mass loss to surface magnetic activity and its evolution. Only about 200 radio stars are known, and virtually all of these emerged from targeted surveys of optically selected objects. FIRST provides a unique opportunity to produce unbiased, flux-limited, and complete samples of stellar radio sources. The accurate positions of even the faintest radio sources will lead to immediate and unambiguous optical identifications from existing POSS and ESO plate scan databases, as virtually all detectable objects are brighter than these plate limits. (Note that although bright stellar objects are not dense on the sky, radio stars are so rare that one must have subarcsecond radio positions to identify them in a large area survey.) From enriching our understanding of the evolution of dynamo-generated magnetic fields on late type stars (which will emerge from a comparison with the ROSAT database), to identifying nearby, coeval, moving groups of red dwarfs, our survey will open a new era in stellar radio astronomy.

Other Galactic objects may also turn up in the survey. Recently, a 0.5 Jy radio source in the southern hemisphere was identified as the nearest millisecond pulsar yet discovered (Johnston *et al.* 1993). Comparison of our survey with the Texas 80 cm interferometric survey and the ongoing, low-frequency Westerbork survey will immediately identify good candidates for these fascinating objects from their uniquely steep spectral indices, providing the first wide-area pulsar survey unbiased by the problems of traditional search techniques such as low duty cycles, high dispersion measures, short periods, and binary accelerations.

7. Summary

The FIRST project will produce a radio image of the sky with 20 times the sensitivity and 10 times the angular resolution of the best centimetric surveys now available. If history provides any guide, many new, unforeseen discoveries will emerge both from an analysis of the radio sources themselves and through comparison with catalogs and sky maps in other spectral regimes. Our proposal, accepted by the NRAO, that there be no proprietary period for the survey data, is designed to maximize the scientific return of the project by making the raw data, calibrated UV data, survey images, and source lists avaliable to the community as soon as they are produced. We encourage the widest possible use of the FIRST Survey data products and welcome suggestions from the community concerning their production and distribution.

References

Gunn, J. E., & Knapp, G. R. 1993, in *Surveys in Astronomy*, ed. B. T. Soifer, ASP Conference Proceedings, in press

Johnston, S., *et al.* 1993, Nature, 361, 613

Kleinmann, S. G., 1992, in *Robotic Telescopes in the 1990's*, ed. A. V. Filippenko (San Francisco: Astron. Soc. Pacific), p. 203

Kuhr, H., *et al.* 1981, A&AS, 45, 365

Oort, M. J. A., 1987, Ph.D. thesis, University of Leiden

Rowan-Robinson, M. *et al.* 1991, Nature, 351, 719

Stocke, J. T. *et al.* 1990, ApJ, 348, 141

Thuan, T. X., & Condon, J. J. 1987, ApJ, 322, L9

White, R. L, & Becker, R. H. 1992, ApJS, 79, 331

White, R. L., Postman, M., & Lattanzi, M. G. 1992, *Digitized Optical Sky Surveys*, eds. H. T. MacGillivray & E. B. Thomson (Amsterdam: Kluwer), p. 167

White, R. L. 1993, in *Space and Earth Science Data Compression Workshop*, NASA Conference Publication 3183, ed. James C. Tilton, p. 117

Windhorst, R. A., *et al.* 1985, ApJ, 289, 494