ASTRONOMICAL AND BIOCHEMICAL ORIGINS
AND THE SEARCH FOR LIFE IN THE UNIVERSE
Eds. C.B. Cosmovici, S. Bowyer and D. Werthimer

THE SERENDIP IV INTERFERENCE REJECTION AND SIGNAL DETECTION SYSTEM

JEFF COBB, CHARLES DONNELLY, STUART BOWYER, DAN WERTHIMER and MICHAEL LAMPTON
Space Sciences Laboratory, Center for EUV Astrophysics and Department of Astronomy, University of California, Berkeley, CA 94720-7450, USA

ABSTRACT

The SERENDIP Interference Rejection and ETI Notification system (SIREN) is a software package developed at UC Berkeley for use in the search for extraterrestrial intelligence (SETI). Originally designed for and used by the SERENDIP III project, it is being substantially rewritten for use with SERENDIP IV. This paper will present the SIREN system and discuss the algorithms employed. SIREN performs both interference rejection and signal detection. The end result of SIREN processing is a database of candidate ETI signals. The most interesting candidates will be the subject of a targeted search.

INTRODUCTION

The UC Berkeley SETI project, SERENDIP, has implemented its fourth generation in hardware and software design. Installation at the Arecibo Observatory in Puerto Rico is scheduled for early 1997. With the aid of a dedicated receiver, SERENDIP IV will carry out a multiple year sky survey in the 1.4 GHz band. As in previous generations of SERENDIP, this survey will be a commensal search. By piggybacking on other regularly scheduled observations, SERENDIP IV will be able to acquire data around the clock.

The SERENDIP IV observing cycle consists of real-time data analysis by the SERENDIP IV instrument, near-time data transfer across the Internet, off-line data reduction and candidate generation, and finally reobservation of the most interesting candidates. This is illustrated by Figure 1.

The SERENDIP IV version of SIREN builds on the extensive experience that we gained during the six years of SERENDIP III development and operation. See Donnelly *et al.* (1995) for a detailed discussion of SERENDIP III software and Bowyer *et al.* (this volume) for a discussion of SERENDIP III results.

REAL-TIME DATA ACQUISITION AND ANALYSIS

The 168 million channel SERENDIP IV instrument consists of forty spectrum analysis/postprocessing boards working in parallel. See Werthimer *et al.* (this volume) for details on hardware design. Each of the forty boards utilizes dedicated hardware to perform a 4.2 million point Fast Fourier Transform on a different sec-

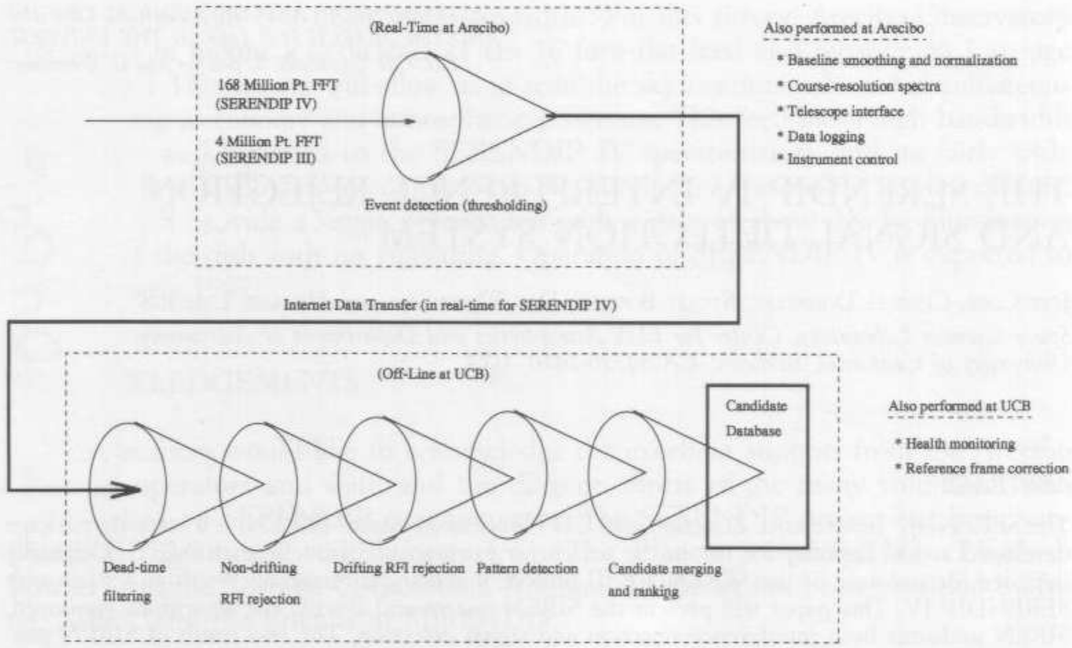

Figure 1 The SIREN data reduction sieve. Each cone represents a data reduction step.

tion of the band. The calculated power spectra enter the postprocessing portion of each board where baseline normalization, coarse resolution spectra computation, and event thresholding occur. Because a small error in knowledge of the average power per bin would lead to a large relative error in noise detection probability, we take care in establishing an accurate average baseline noise power spectrum. Baseline normalization is achieved with a sliding eight thousand channel local mean boxcar. To gain sensitivity to wide band or frequency modulated signals, a sequence of progressively coarser resolution spectra are computed by adding the powers of adjacent bins. Thresholding is performed based on mean spectral power, and signals that exceed the threshold are logged along with time, pointing coordinates, frequency, and power. The threshold is set such that the system is not overwhelmed with detections while at the same time rendering the probability of rejecting an ETI signal as low as possible. A threshold of 16 times the mean noise power has worked well for SERENDIP III, producing, on average, 25 detections per 4.2 million point power spectrum. Due to radio frequency interference (RFI) this detection rate is typically 50 times that expected from pure noise. Our system can accommodate this data rate and the RFI is removed during off-line data reduction. SERENDIP IV will be operating on a 100 MHz band centered at 1420 MHz. This area of the spectrum has less RFI than the 424-436 MHz band of SERENDIP III and should thus produce fewer detections per board if the threshold were kept the same. We will tune the threshold during SERENDIP IV installation to make maximum use of throughput capacity.

Signals exceeding the threshold are reported by the postprocessors to the control PC which then logs these signals to a workstation via a dedicated Ethernet. The workstation functions as a short term archive and data transfer machine. As data arrive from the PC, they are simultaneously written to a local file system and sent, via TCP/IP, across the Internet to UC Berkeley. A workstation at UC Berkeley receives the data and writes them to a file system mirroring the short term archive at Arecibo. Once successful data transfer is confirmed, the short term archive files at Arecibo are deleted.

OFF-LINE DATA REDUCTION

Off-line data reduction consists of first removing detections that were collected during times when the telescope slew rate was incompatible with our system and then applying a set of cluster analysis algorithms that identify and reject RFI. Instrument health is also monitored at this stage by the detection of an artificial signal that is periodically injected into the instrumentation.

Telescope slew rate is critical to both proper data acquisition and effective RFI rejection. SERENDIP receives the telescope's pointing coordinates every 10 seconds and interpolates the positions for spectra taken between each report. A too rapid slew rate leads to errors in this interpolation. Conversely, zero or very slow slewing makes it difficult to remove RFI. Data collected during both of these extremes are removed while those collected during normal slew rates are passed on to the RFI rejection algorithms. During SERENDIP III roughly 15% of the observations were rejected because of incompatible telescope slew rates.

During the RFI rejection stage, we remove interference from terrestrial and near-space sources (passband RFI), interference from observatory electronics (baseband RFI), spread spectrum RFI (defined below), and interference that drifts in frequency over time. A separate test for each of these RFI types is conducted on several hours of observational data at a time.

The passband and baseband RFI rejection algorithms calculate the expected distribution of detections over receiver and baseband frequency bins assuming a pure noise environment. Those bins containing an excessive number of detections are then examined for signals that persist over multiple telescope beams. Such signals are rejected as RFI.

Spread spectrum RFI is defined as the appearance of many detections in a single integration period, with these detections being spread out across the spectrum. The goal here is to reject spread spectrum RFI without rejecting a broad band ETI signal. The latter would appear as many detections grouped closely together in the spectrum. We have found empirically that this goal is achieved if we reject spectra for which the following inequality is valid:

$$\frac{\sum d^2}{x} > S$$

where d is the number of bins (0.6 Hz/bin) between adjacent events and x is the number of hits in the spectrum. The threshold S was tuned through our experience with SERENDIP III data and is set at 10^8.

SERENDIP III data shows that even though typically 98% of all events are identified and rejected as passband, baseband, and spread spectrum RFI, only 1% of the band is lost during the rejection process. The surviving events are passed on to the drifting RFI detection algorithm.

Signals that drift rapidly in frequency and are thus not detected by the passband RFI test are detected and rejected by the drifting RFI algorithm. Even though this is

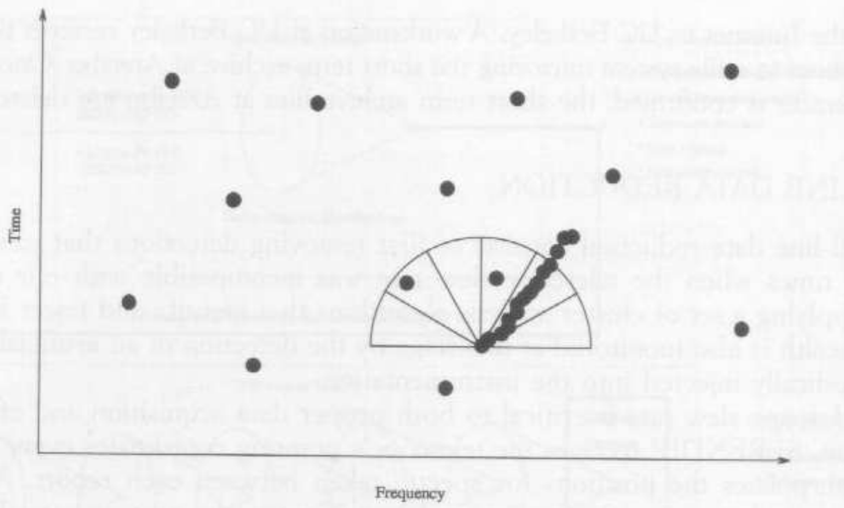

Figure 2 Illustration of SIREN's drifting RFI detection algorithm. A sectored semi-ellipse of frequency time space around each detection is analyzed for signals that exhibit time-coherency and persistence.

the most computationally intensive of our RFI tests, its throughput is good because it is conducted on very clean data. Event density in frequency time space is calculated for each dataset. A sectored semi-ellipse around each detection (see Figure 2) is analyzed for sectors containing an excess of detections. Such sectors are scrutinized for signals that are coherent over multiple pointings of the telescope. SERENDIP III has shown that 18% of remaining events are typically rejected at this stage.

PATTERN DETECTION

SIREN pattern detection algorithms are sensitive to signals that are persistent over single and multiple observations, signals that match the telescope's beam pattern and pulsed signals. We frequency correct all clean data to the solar system barycenter before subjecting it to the pattern detection algorithms. This removes the Doppler drift induced by our own acceleration.

Detected ETI signals might be of two types: directed or leakage. Accordingly, we apply two distinct searches for persistence in frequency over time. A directed signal is one for which the ETI have likely removed any Doppler drift induced by the acceleration of their transmitter. Thus, for this search, we apply the criteria of multiple detections within a narrow frequency range. The frequency range is set at 20 Hz, owing to variation due to our inherent pointing uncertainty; a strong signal could be detected a beamwidth away from its true source leading to frequency correction error. A set of signals containing one or more detections during each of two or more observations of a given sky position and which are within 20 Hz of one another is logged as a candidate of the directed type. The second type of persistent signal for which SIREN is sensitive is radio leakage. Radio leakage from another civilization will likely drift significantly in frequency over time spans of minutes to months, owing to the uncompensated transmitter acceleration. In this search, candidate signal sets can have a frequency range of up to 165 KHz (for SERENDIP IV at 1.4 GHz). We chose this window because a transmitter on a rapidly rotating planet could induce a frequency difference of this magnitude between detections made when the transmitter was at maximum velocity towards us and detections made when it was at maximum velocity away from us. However, if we simply widened the frequency window, we would be inundated with leakage candidates from noise alone. To prevent this, we apply additional criteria: leakage candidates are required to consist of two or more detections during each of two or more observations of a given sky position. In addition, the same-observation detections must either not vary more than 20 Hz in frequency or must display a coherent drift rate. These added constraints reduce the list of leakage candidates to the most interesting two percent. From the analysis of SERENDIP III data, we found that the number of candidates from both the directed and leakage algorithms is consistent with that expected from noise, indicating that these algorithms are working well.

Same-observation detections that are within 20 Hz of each other or which display a coherent drift rate are further scrutinized. Sets of three or more consecutive detections are analyzed for power distributions matching the telescope's Gaussian beam pattern. A good match indicates that the signal may have emanated from a point source. We also look for detections occurring at regular intervals within a single transit as this could indicate a pulsed signal.

CANDIDATE RANKING AND REOBSERVATION

A SERENDIP candidate is defined as a one beamwidth area on the sky that has been identified by one or more pattern detection algorithms. Candidates are ranked by three criteria. First, each candidate is given one or more scores, one score for each algorithm that identified it. Each score is the probability of the identified set of signals occurring from noise alone. Second, we ask how many algorithms identified the candidate. Finally, we determine if any of our candidates are at the coordinates of a nearby star or a known planetary system.

To determine the algorithm specific score, we calculate the *relative* probability of our detections occurring from noise, given the number of detections in the candidate area, the number of times we have observed the candidate area, the number of frequency bins assumed by the algorithm, and the actual frequency «tightness» of the detections. The relative probability, from Poisson statistics, is given by

$$P = \frac{n_e^{n_d}}{n_d!\left(\frac{f_{tot}}{f_{win}}\right)^{(n_d-1)}}\left[\frac{\Delta f + f_\sigma}{f_\sigma}\right]$$

where: n_e is the total number of events logged at any frequency in the candidate area; n_d is the number of detections within f_{win}; f_{win} is the frequency window assumed by the algorithm; f_{tot} is the total SERENDIP band observed; Δf is the maximum frequency separation for the detection set; f_σ is the expected frequency variance.

If the candidate was detected by the algorithm sensitive to directed signals, the dominant source of f_σ is position uncertainty. If the candidate was detected by the leakage algorithm, the dominant source of f_σ is Doppler drift induced by the motion of the transmitter.

The assessment of joint probability in the case of a candidate being identified by multiple algorithms is more problematic. The pattern detection algorithms are not orthogonal to one another, thus the probabilities do not simply multiply. The ranking of multiple-algorithm candidates is subjective but nonetheless useful in generating a list for reobservation.

The most interesting candidates are subjected to visual inspection for RFI contamination using time/frequency plots. Those surviving constitute the list of interesting candidates to be the subject of reobservation.

SERENDIP employs two types of candidate reobservation. The first type is simply chance reobservation during the normal sky survey. This typically happens every six months. As candidate locations are reobserved, additional detections or the lack thereof will alternately add or subtract from a candidate's score. In addition, SERENDIP has been awarded twenty-four hours of dedicated reobservation time. This time will be used to reobserve the top scoring candidates. During dedicated reobservation, sensitive narrow band on and off source observations will be used to characterize the candidate.

CONCLUSIONS

The 4 year SERENDIP III survey has provided us with the opportunity to confirm the efficacy of the SIREN algorithms. We have tuned the system and invented several new ways of looking at our data. All data reduction and analysis algorithms discussed have been implemented and are being used in the final SERENDIP III analysis. Coding changes needed to support SERENDIP IV are well under way. We expect to have SERENDIP IV up and running in early 1997.

ACKNOWLEDGEMENTS

The SERENDIP project is largely funded by the Planetary Society. The authors would also like to acknowledge generous donations from Mrs. Elizabeth Bowyer, Sun Microsystems, Toshiba, and Intel Corp. Xilinx, and the Friends of SERENDIP headed by Arthur C. Clarke have provided additional support. We also acknowledge science and engineering input from Michael Davis of Arecibo and the tireless volunteer efforts of Sabine Airieau and Dean Kakasu in assisting with the development of the SERENDIP IV software system. SERENDIP's successful efforts are also dependent on the excellent support from the entire Arecibo staff and in particular lead telescope operator Rey Valez and computer department head Arun Venkataraman.

REFERENCES

DONNELLY, C., WERTHIMER, D., BOWYER S. and COBB, J. (1995): The SERENDIP interference and signal detection system, in *Progress in the Search For Extraterrestrial Life*, (Shostak, S. ED.), ASP CONFERENCE SERIES, VOL. 74.

BOWYER, S., WERTHIMER, D., DONNELLY, C., COBB, J., NG, D. and LAMPTON, M. (1997): Twenty years of SERENDIP, the Berkeley SETI effort: past results and future plans (this volume).

WERTHIMER, D., BOWYER, S., NG, D., DONNELLY, C., COBB, J., LAMPTON, M. and AIRIEAU, S. (1997): The Berkeley SETI Program: SERENDIP IV instrumentation (this volume).