CANDIDATE IDENTIFICATION AND INTERFERENCE REMOVAL IN SETI@HOME. E. J. Korpela, J. Cobb, M. Lebofsky, A. Siemion, J. von Korff, R. Bankay, D. Werthimer and D. Anderson, Space Sciences Laboratory, Mail Code 7450, University of California, Berkeley, CA 94720-7450 (*korpela@ssl.berkeley.edu*).

SETI@home, a search for signals from extraterrestrial intelligence, has been recording data at the Arecibo radio telescope since 1999. [1] To date, SETI@home volunteers have detected over 4.2 billion potential signals. [2] While essentially all of these potential signals are due to random noise processes, radio frequency interference (RFI) or interference processes in the SETI@home instrumentation, it is possible that a true extraterrestrial transmission exists within this database. We describe the processes of interference removal used in the SETI@home post-processing pipeline, as well as those used to identify candidates worthy of further investigation.

Candidate Identification: Candidate identification ranks groups of signals by their persistence in time, their spacial proximity, dissimilarity to signals generated by random noise processes, dissimilarity to known interference sources, and their proximity to interesting celestial objects (nearby or solar type stars, known planetary systems, etc.)

Early in the project, candidate identification was an arduous process which was undertaken at intervals ranging from 6 months to more than a year. Because this process would access every signal in the SETI@home database several times, it was very I/O intensive and would require months to complete. To remove this shortcoming, we have designed a Near-Time Persistency Checker (NTPCkr).

The SETI@home pipeline keeps track of incoming potential signal spatial locations by pixelating the sky in an equal area pixelization scheme. When a signal comes in, the corresponding sky pixel is marked as "hot" and given a time-stamp. Since a given area of sky tends to be observed several times in a short period, this pixel is allowed to "cool" for several weeks. At this point, if no further signals for that pixel are received, it is marked as ready for analysis.

The NTPCkr then examines the signals within that pixel and adjacent pixels to determine a candidate score based upon the above criteria. It is our goal that the score represent the probability of the set of potential signals within the candidate arising in that position, with its associated position and frequency distribution, due to random noise processes. The candidates are ranked in order of this score.

Interference Removal: In the past, it has been our practice to perform interference removal on the entire set of potential signals detected by our instruments. Again, this method requires that the entire database be examined multiple times, which is inefficient.

Because narrow band interference is very unlikely to occur due to random noise processes, candidate groups containing interference are ranked very highly on our candidate lists. Therefore we run interference rejection on candidate groups in order of their ranking. After interference rejection, the candidate position is again marked as ready for analysis by the NTPCkr. We have implemented, or plan to implement several interference rejection routines.

Zone Interference Removal: Zone Interference Removal removes signals that are contained within a "zone," which is a region of parameter space known to contain a large number of invalid signals. The parameters that define a zone can include radio frequency, base-band frequency, period (for pulsed signals), detection time, or the identity of the receiver or software used for various stages of the analysis process. Signals determined to be within the zones are marked as interference and are excluded from candidate scoring computations. For example we have identified ~35,000 frequencies, covering ~1% of our band which are subject to frequent interference. These zones contain between 10% and 20% of the detected signals depending upon signal type.

Short-Term Fixed-Frequency Interference Removal: Some sources of interference are present at constant frequencies for periods of time ranging from hours to days. Because celestial objects stay in our field of view for seconds to minutes, we can use this properly to remove these sources of interference. By examining a time range around a potential signal we can calculate the probability of coincidence with another signal with similar frequency but seen at a different sky position. If this probability falls below a threshold (of order 10^{-4}) we conclude that the signals are due to an interference source.

Drifting Interference Source Removal: Some sources drift in frequency, even over short periods of time. For these methods we use the octant-excess drifting interference detection and removal method described in Cobb et al. [3] Adjacent signals in time and frequency, but at different sky positions are allocated into octants of frequency-time space surrounding the signal being examined. A significant statistical excess in an octant and the octant 180 degrees opposite indicates the presence of an RFI source drifting in frequency. Again, a probability computation is used to determine the likelihood that this excess is due to random noise, and if this computation falls below a threshold, the signal being examined is marked as being due to interference.

Acknowledgments: The SETI@home and Astropulse projects are funded by grants from NASA and the National Science Foundation, and by donations from the friends of SETI@home. Observations are made at the NAIC Arecibo Observatory, a facility of the NSF, administered by Cornell University.

References:

[1] Korpela, E. J. et al. (2000) *Computing in Science and Engineering, v3n1,* 78.

[2] http://setiathome.berkeley.edu/sci_status.html

[3] Cobb, J. et al. (2000) *Bioastronomy 99, 213,* 485.