

Astropulse and Fly's Eye: SETI Searches for Transient Radio Signals Using Distributed Computing

J. Von Korff,^{1,2} A. Siemion,^{3,4} E. Korpela,¹ D. Werthimer,^{1,3}
P. McMahon,^{3,5} J. Cobb,¹ M. Lebofsky,¹ D. Anderson,¹ B. Bankay,¹
G. Bower,⁴ G. Foster,⁴ J. van Leeuwen,⁴ W. Mallard,³ and M. Wagner³

¹*University of California, Berkeley - Space Sciences Lab, Berkeley, CA 94720*

²*vonkorff@ssl.berkeley.edu*

³*University of California, Berkeley - Berkeley Wireless Research Center, Berkeley, CA 94720*

⁴*University of California, Berkeley - Department of Astronomy, Berkeley, CA 94720*

⁵*Stanford University - Department of Computer Science, Stanford, CA, 94305*

⁶*University of California, Berkeley - Department of Physics, Berkeley, CA 94720*

Abstract. Berkeley conducts 7 SETI programs at IR, visible and radio wavelengths. Here we review two of the newest efforts, Astropulse and Fly's Eye.

A variety of possible sources of microsecond to millisecond radio pulses have been suggested in the last several decades, among them such exotic events as evaporating primordial black holes, hyper-flares from neutron stars, emissions from cosmic strings or perhaps extraterrestrial civilizations, but to-date few searches have been conducted capable of detecting them.

We are carrying out two searches in hopes of finding and characterizing these μ s to ms time scale dispersed radio pulses. These two observing programs are orthogonal in search space; the Allen Telescope Array's (ATA) "Fly's Eye" experiment observes a 100 square degree field by pointing each 6m ATA antenna in a different direction; by contrast, the Astropulse sky survey at Arecibo is extremely sensitive but has 1/3,000 of the instantaneous sky coverage. Astropulse's multibeam data is transferred via the internet to the computers of millions of volunteers.

The Fly's Eye was successfully installed at the ATA in December of 2007, and to-date approximately 450 hours of observation has been performed. We have detected three pulsars (B0329+54, B0355+54, B0950+08) and six giant pulses from the Crab pulsar in our diagnostic pointing data. We have not yet detected any other convincing bursts of astronomical origin in our survey data.

1. Astropulse Telescope and Hardware

Arecibo Observatory scans approximately $\frac{1}{3}$ of the sky, between declinations of -1.33 and 38.03 degrees. This means that Astropulse cannot see the galactic center (around -29 dec) but can see many known pulsars, including the Crab.

The ALFA receiver has 7 dual-polarization beams on the sky, each with a $3.5'$ beamwidth. The central beam has a gain of 11 K Jy^{-1} , and the other beams have 8.5 K Jy^{-1} . The system temperature is 30 K .

We collect and store the 14 signals, saving a copy of each file at NERSC, the National Energy Research Scientific Computing Center. In all, we have taken 100 TB of data so far from ALFA multibeam. Volunteers' PCs will then process the data and send the results back to Berkeley, see Section 2.3.

2. Astropulse Algorithm and its implementation

2.1. Dedispersion

Between a radio pulse's source (i.e. ET, pulsar, or neutron star) and our detector, a radio pulse must travel through the Interstellar Medium (ISM), causing dispersion (Wilson et al. 2009) so that the high frequency components travel slightly faster.

We have to reconstruct the original pulse, bringing together the component frequencies and reassembling them so that the signal has a short duration again.

We have two choices for our methodology: coherent dedispersion and incoherent dedispersion. Astropulse uses coherent dedispersion, whereas other radio surveys use incoherent dedispersion. Incoherent dedispersion is much more computationally efficient, and for longer timescales it's almost as good as coherent dedispersion. However, Astropulse would be unable to examine the $0.4 \mu\text{s}$ timescale without coherent dedispersion.

Incoherent dedispersion means that the signal is divided up into sub-bands, and the power vs. time of each sub-band is recorded. The method is called "incoherent" for this reason – the phase information about individual frequencies is lost; only the total power of each subband is retained. Then, the sub-bands are realigned at all possible dispersion measures, in an effort to find one DM at which the components align to produce a large power.

However, this project is limited in two ways. First, the goal of recording power vs. time makes sense only on a timescale greater than $\frac{1}{d\nu}$, where $d\nu$ is the width of each sub-band. This is because of time-frequency uncertainty. Second, if $\Delta\tau$ is the time over which the pulse is dispersed, then in each sub-band the pulse is dispersed by $\Delta\tau \cdot \frac{d\nu}{\Delta\nu}$. So the method cannot localize the pulse better than this. Combining these two limits, we find that the minimum timescale for incoherent dedispersion happens when

$$dt = 12.8 \mu\text{s} \left(\frac{DM}{56.791} \right)^{0.5} \left(\frac{\nu_0}{1.42\text{GHz}} \right)^{-1.5} \quad (1)$$

Where we have calibrated to the Crab's DM , and ν_0 is the center frequency. For the Crab pulsar, this is a limit of $12.8 \mu\text{s}$. For a more distant source, the limit might be as much as $50 \mu\text{s}$. But coherent dedispersion deals with amplitude rather than power, and attains arbitrary resolution.

We perform coherent dedispersion using a deconvolution, which can be accomplished using FFTs in time $O(N \log N)$.

2.2. Algorithm Logic

Astropulse loops through the data at several nested levels, starting at a dm of 896 samples (49.5 pc cm^{-3}):

1. Large and small dm chunk: blocks of 128 or 16 dms at a time.
2. Data chunks of size $32768 = 2^{15}$ samples. Compute FFT of the data for use in convolution.
3. All dms within a small dm chunk, and sign of the dm (positive and negative.)

We consider negative dms for a few reasons: First, extraterrestrial intelligences might communicate using signals dispersed with negative dms. And second, signals detected at both positive and negative dm might be a sign of RFI.

4. Combine samples at scales (or “co-adds”) of 2^0 to 2^9 samples.
5. Samples within the data chunk

If the power in one sample (or rather 2^ℓ samples) is above a certain threshold, then Astropulse reports a pulse.

The Fast Folding Algorithm The Fast Folding Algorithm (FFA) is described in Staelin (1969). In our case, it operates on a time series of bins, not samples. Each bin is a sum of either 16 or 128 samples, depending on whether it was created during the large or small dm chunk loop. The FFA runs a nested loop, similar to the single pulse algorithm, looping over frequencies (the highest level), then subfrequencies, coadds, and bins (the lowest level).

2.3. BOINC

Astropulse runs on the BOINC platform, which stands for “Berkeley Open Infrastructure for Network Computing.” BOINC is a set of programs that organizes volunteers’ home computers to perform scientific calculations. In any BOINC project, a researcher has a computing problem that can run in parallel, that is, on several machines at once. Perhaps the problem involves searching a physical space (for Astropulse, the sky), and performing the same computation on each point in that space (for Astropulse, dedispersion.) The first BOINC project, SETI@home, searched exclusively for narrowband transmissions across the sky, which could be a signal from extraterrestrial intelligence. Or, the space could instead be a parameter space, for instance a space of potential climate models (climateprediction.net) or protein configurations (Rosetta@home).

The researcher for a BOINC project need not be affiliated with UC Berkeley, or with the BOINC development team at Berkeley (although we happen to be so affiliated.) BOINC is open source, and can be downloaded, compiled, and operated by anyone with sufficient technical skills; many projects currently exist outside Berkeley.

Likewise, volunteers need no particular technical knowledge. They just visit the BOINC web page and download the “client” programs. Astropulse has access to around 500,000 volunteers, each of whose machines might have 2 GFLOPs of processing power, and be on 1/3 of the time, for a total of 300 TFLOPs – approaching the speed of the world’s fastest supercomputer.

3. Fly’s Eye

The Allen Telescope Array has several advantages over other telescopes worldwide for performing transient searches, particularly when the search is for bright pulses. The ATA has 42 independently-steerable dishes, each 6m in diameter. The beam size for individual ATA dishes is considerably larger than that for most other telescopes, such as VLA, NRAO Green Bank, Parkes, Arecibo, Westerbork and Effelsberg. This means that the ATA can instantaneously observe a far larger portion of the sky than is possible with other telescopes. Conversely, when using the ATA dishes independently, the sensitivity of the ATA is far lower than that of other telescopes.

The Fly’s Eye instrument was purpose built to search for bright radio pulses of millisecond duration at the ATA. The instrument consists of 44 independent spectrometers using 11 CASPER IBOBs. Each spectrometer processes a bandwidth of 210MHz, and produces a 128-channel power spectrum at a rate of 1600Hz (i.e. 1600 spectra are outputted by each spectrometer per second). Therefore each spectrum represents time domain data of length $1/1600\text{Hz}=0.000625\text{s}=0.625\text{ms}$, and hence pulses as short as 0.625ms can be resolved¹.

We have to-date performed roughly 400 hours of observing with the Fly’s Eye.

3.1. Fly’s Eye Offline Processing

The analysis required for the Fly’s Eye experiment is, in principle, fairly simple – we wish to search over a wide range of dispersion measures to find large individual pulses. Specifically our processing requires that all the data be dedispersed with dispersion measures ranging from $50\text{ cm}^{-3}\text{ pc}$ to $2000\text{ cm}^{-3}\text{ pc}$. At each dispersion measure the data needs to be searched for ‘bright’ pulses.

The processing chain is in practice significantly more complicated than this description suggests. Processing is performed on compute clusters, with input data formatted, divided and assigned to worker nodes for processing. In the worker node flow, the data is equalized, RFI rejection is performed, and finally a pulse search is performed through the range of dispersion measures. The results are written to a database where they can be subsequently queried. The key feature of the results is a table that lists, in order of decreasing significance, the pulses that were found and the dispersion measures they were located at.

¹Pulses of duration $<0.625\text{ms}$ can also be detected provided that they are sufficiently bright, but their length cannot be determined with a precision greater than the single spectrum length.

Average power equalization is performed on the frequency spectrum equalized values $P'_i(t)$. We compute the average power over all frequency channels for a single integration (time sample t). The power average is defined as $\overline{P'(t)} = \frac{1}{N} \sum_{i=0}^{N-1} P'_i(t)$. N is the number of channels (for Fly's Eye this is always 128). The motivation for why it is possible to normalize the power is that we expect pulses to be dispersed over many time samples, so this procedure should not remove extraterrestrial pulses.

Our strategy for mitigating constant narrowband RFI is simply to identify the channels that are affected, and to exclude them from further processing. This channel rejection is typically performed manually by looking at a set of spectra and identifying obviously infected channels, which are then automatically excluded in subsequent processing runs.

Intermittent RFI is often quite difficult to automatically distinguish from genuine astronomical pulses, and we followed a conservative approach to try to ensure that we do not accidentally excise dispersed pulses. Our statistic for intermittent RFI is the variance of a single channel over a 10 minute data chunk, $\sigma_i^2 = \left(\frac{1}{T_0} \sum_{t=0}^{T_0-1} (P''_i(t))^2 \right) - \left(\frac{1}{T_0} \sum_{t=0}^{T_0-1} (P''_i(t)) \right)^2$. Curve-fitting determines a σ_i^2 outside which it is likely that channel i contains time-varying RFI. Future reprocessing will likely use a more robust method, such as that based on a kurtosis estimator Nita et al. (2007).

Our final RFI mitigation technique is manual – in our results it is easy to see high- σ hits that are a result of RFI: these hits appear as simultaneous detections at many dispersion measures.

3.2. Detection of Giant Pulses from the Crab Nebula

A suitable test of transient detection capability is to observe the Crab pulsar and attempt to detect giant pulses from it. We were able to detect several such pulses, which appeared only at the expected dispersion measure. Wideband RFI appeared across a wide range of dispersion measures.

Acknowledgments. Astropulse has been supported in part by NSF Grants AST-0808175 and AST-0307956, and NASA Grant NNX09AN69G. We are grateful for the support of the NAIC Arecibo Observatory, a facility of the NSF, managed by Cornell University. We would also like to thank the Allen Telescope Array a facility managed by the University of California and the SETI Institute.

References

- Nita G. M., Gary D. E., Liu Z., 2007, PASP 119, 805
 Staelin D. H., 1969, Proc. IEEE 57, 724
 Wilson T. L., Rohlfs K., Hüttemeister S., 2009, Tools of Radio Astronomy, Springer