Distributed Processing of SETI Data

Eric Korpela

As you have read in prior chapters, researchers have been performing progressively more sensitive SETI searches since the late 50s. Each search has been limited by the technologies available at the time. As as radio frequency technologies have become more efficient and computers have become faster, the searches have increased in capacity and become more sensitive. It is most often the limits of the hardware that performs the calculations required to process the telescope data in a way that exposes any embedded signals. Shortly before the start of the 21st century, projects began to appear that exploited the processing capabilities of computers connected to the Internet in order to solve problems that required a large amount of computing power. The SETI@home project, which I and a group of researchers at the Space Sciences Laboratory of the University of California, Berkeley manage, was the first attempt to use large scale distributed computing to solve the problems of performing a sensitive search for narrow band radio signals from extraterrestrial civilizations. (Korpela et al. 2001) A follow-on project Astropulse searches for extraterrestrial signals with wider bandwidths and shorter time durations. Both projects are on-going at the present time (mid-2010).

Computation in radio SETI

Why would an enormous supercomputer would be necessary to detect radio signals from an alien civilization? It might seem to be a fairly simple signal processing task. One reason is that the parameters of any alien signal are unknown. Some of these parameters are intrinsic to the signal: frequency, frequency changes, bandwidth, encoding, and duration Others are properties of how the signal is sent and received: Is the transmitter on a planet, in orbit, in interstellar space? Is it directional or omnidirectional? Still others are unavoidable properties of how the signal propagated through space to get here. To perform a thorough search, we need to investigate a wide variety of these parameters.

One typical assumption made in SETI is that an alien civilization wishing to make contact with other races would broadcast a signal that is easily detectable and easily distinguishable from natural sources of radio emission. One way of achieving these goals is to send a narrow band signal. By concentrating the signal power in a very narrow frequency band, the signal can be made to stands out among the natural sources of noise which are broad band. A second way is to send a signal of short time duration which, in principle, would be detectable above the background noise for the duration that the signal is on. For reasons to be described later, much more processing power must be employed in order to detect this second type of signal.

Because of this, radio SETI efforts have concentrated on detecting narrow band signals. When searching for narrow band signals it is best to use a narrow search window (or channel) around a given frequency. The wider the channel, the more broad band noise is included in addition to any signal. This broadband noise limits the sensitivity of the system. Early systems used analog technology to create narrow bandpass filters that could observe at a single frequency channel. More recent systems use massive filter-banks banks of dedicated Discrete Fourier Transform[1] (DFT) processors to separate incoming signals into up to a two billion spectral channels, each of width ~1 Hz.

1 Other chapters in this book may use the term Fast Fourier Transform (FFT) which is a specific type of DFT. For these purposes, the terms are interchangeable.

There are, however, limitations to this technique. One limitation is that extraterrestrial signals are unlikely to be stable in frequency due to accelerations of the transmitter and receiver. For example, a receiver listening for signals at 1.4 GHz located on the surface of the earth undergoes acceleration of up to 3.4 cm/s^2 due to the earth's rotation. That may not seem like much, but it corresponds to a Doppler drift rate of 0.16 Hz/s. If uncorrected for this drift an alien transmission would move outside of a 1 Hz channel in about 6 seconds, effectively limiting the maximum integration time to 6 seconds. Because of the inverse relationship between maximum frequency resolution and integration time ($\Delta\nu=1/\Delta t$) there is an effective limit to the frequency resolution that can be obtained without correcting the received signal for this effect. ($\Delta\nu\sim0.4$ Hz)

In principle a correction can be made for most of the drift due to motions of the earth, but how does one correct for motions of an unknown planet? An alien civilization beaming signals directly at the earth could correct the outgoing signal for the motions of the transmitter, but a civilization transmitting an omnidirectional beacon could not make such an adjustment[2]. Therefore, to search for this type of signal at very narrow bandwidth (<<1 Hz) and the highest possible sensitivity, the correction for Doppler drift must be made at the receiving end and a search for signals performed at multiple Doppler drift rates. Repeating an analysis at multiple Doppler drift rates becomes compute intensive.

Other parameters of the signal are also unknown, for example: At what frequency it will it be transmitted? What is the bandwidth of the signal? Will the signal be pulsed, if so at what period? Fully investigating a wide range of these parameters requires proportionally larger computing power.

In addition to detecting a signal, we must be able to determine whether a signal is truly of celestial origin. The vast bulk of the narrow band signals received by a radio telescope will be radio frequency interference (RFI) generated locally. Fortunately RFI often has properties that allow it to be distinguished from extraterrestrial emission. RFI elimination requires some level of computing resources.

Performing all of these these calculations for even a small portion of the radio spectrum requires as much computational power as is available in the largest existing supercomputer. However, such computers are not typically made available to SETI researchers.

SETI@home

Fortunately, searching for signals in a data stream from a radio telescope is a task that is easily distributed. Data from an observation can be broken up into frequency bands that are essentially independent of one another. In addition, an observation of one portion of the sky is essentially independent of an observation of another part of the sky. This allows a large data set to be divided into small chunks that can be analyzed by a personal computer in a comparatively short time, making possible the distribution of the work to people willing to donate their spare CPU cycles.

[FIGURE 1 THIS PAGE OR LATER]

2 Actually, with enough expense, they could. Rather than building a single omnidirectional beacon they could build enough directional transmitters to cover the sky. Sixteen million five hundred thousand Arecibo class telescopes would do the job nicely, but that's probably overkill. Because small telescopes have a larger field of view, but require more power to send the same effective isotropic radiated power, there is a tradeoff between number of telescopes, the amount of uncorrected Doppler drift and the total power that could be transmitted without melting the transmitters. Calculating the optimum number is left for the reader, a colleague or another time.

[FIGURE 1 CAPTION: Figure 1. SETI@home and Astropulse use the National Astronomy and Ionospheric Center's 305 meter telescope at Arecibo, Puerto Rico. Photo courtesy of the NAIC - Arecibo Observatory, a facility of the NSF]

SETI@home conducts its observations at the National Astronomy and Ionospheric Center's 305 meter radio telescope in Arecibo, Puerto Rico. (Figure 1) The project uses ALFA, an array of seven receivers arranged in a hexagonal pattern with one in the middle, which is mounted in the enclosed dome-like structure seen suspended above the Arecibo telescope. SETI@home makes its observations in conjunction with other uses of the ALFA array. Currently this array is used to survey for pulsars near the plane of the galaxy, to map the distribution of hydrogen in all parts of the Galaxy visible from Arecibo, and to search for extragalactic hydrogen gas in isolated clouds or in nearby galaxies. This results in three main modes of observation. The pulsar surveys tend to track positions in the sky while accumulating data for 30 seconds to tens of minutes. The other surveys either utilize a drift scan mode where the receivers are held in position while objects in the sky drift by during the earth's rotation or a "basket-weave" mode in which the receiver tracks north and south while the sky drifts by, resulting in a zigzag path.

If the primary feed is stationary, objects in the sky pass through the fields of view ALFA receivers (0.05°) at the rate of the rotation of the earth (also known as the sidereal rate). An object would require about 13 seconds to transit the field. When used in basket-weave mode, less time is required for transit. When tracking, objects can remain in the field of view for large durations.

During the course of the these projects, the SETI@home will view most portions of the sky visible with the Arecibo telescope three or more times. This includes stars with declinations (the celestial equivalent of latitude) between -2° and 38° thoroughly covering about 25% of the sky.

The SETI@home system records a 2.5 MHz wide band from each of the two polarizations of the seven receivers (14 data streams in all) centered at the 1420 MHz Hydrogen line. Because the Hydrogen line would be of interest to astronomers of any species who were studying the galaxy, this frequency is considered one of the most likely locations for deliberate extraterrestrial transmissions. These 2.5 MHz bands are recorded continuously onto hot-swappable serial ATA disk drives using 2 bit complex samples. A 2TB drive holds the data for about 57 hours of observing. We are accumulating data at a rate of about 50 TB per year. This data is archived at the National Energy Research Scientific Computing Center at the Lawrence Berkeley Laboratory.

The full drives are shipped to Berkeley where they are subdivided into small "work units" using software appropriately known as a "splitter." The 2.5 MHz bandwidth data is divided into 256 sub-bands by means of a 2048 point DFT followed by 256 eight point inverse transforms. The 9766 Hz wide sub-bands are divided into lengths of 2^{20} samples. Each work unit corresponds to about 10 kHz of bandwidth and 107 seconds of duration. When the project began in 1999, these sizes were chosen such that a common desktop computer could perform our analysis procedure in less than a week. Thanks to Moore's Law, a current (2010) 4-core processor can typically process four of these work units in two hours.

Subsequent work units overlap by 20 to 30 seconds to allow full analysis of signals that may be within a beam transit time of the end of a work unit. Each of the work units data file are transferred to temporary storage (which typically holds one to three million work unit files) for distribution to users. The work unit files are stored there until the results for that work unit are received.

BOINC

The structure of the SETI@home server hardware has evolved over time from a single under-powered workstation to what is now several six foot tall racks of computers and disk drives. The software has evolved even more.

The original SETI@home server was a relatively small program that spoke a limited subset of hypertext transfer protocol (HTTP). Although it communicated over the standard HTTP port, it was only capable of processing requests from the SETI@home application, storing results, and returning a single work unit file. Despite this simplicity, it was easily overwhelmed when request rates became high. It had no means of monitoring behavior of users or validating that the result returned belonged to the work unit file that had been sent. It was easy for malicious people to attempt to both boost their credit standings or attempt to damage the integrity of our science database by returning invalid data for a large number of results. This server was also very specific to SETI@home. If we wanted to develop other volunteer computing application, we would need to develop a new server for each.

To alleviate some of these issues we have developed the Berkeley Open Infrastructure for Network Computing or BOINC. (Anderson 2004) Rather than using a special purpose HTTP server, BOINC utilizes standard web servers that support the Common Gateway Interface (CGI) or FastCGI for calling external programs for web page generation. Handling and monitoring connections is done by the web server, which is typically well optimized for the task. The BOINC software is divided into 1) "work generators," which in our case are our splitters described above. 2) A CGI "scheduler" which handles request from volunteers' computers and decides what work to distribute to each, downloads of the work units are performed using standard HTTP from any web server. 3) A CGI "file upload handler" that collects the results that are returned. 4) A "validator" which determines whether the returned results are likely to be correct, in our case by comparing results returned from two or more machines, and 5) an "assimilator" which stores the valid results. In our case our assimilated results are stored in our science database.

BOINC allows easy distribution of these tasks among multiple machines. In addition, it maintains statistics on each computer including processing speed estimates which are used to determine how much work to send. It maintains an estimate of the error rate for a machine, so a trustworthy machine might be trusted to generate a correct result without sending the same work to another computer, while an untrustworthy machine would always have its work checked by another machine.

Once a result has been returned to our server and validated, the assimilator process stores the time, sky coordinates, frequencies, etc. for each of the potential signals that was returned. The largest portion of the science database capacity is used for storing these parameters of potential signals. This database is currently (May 2010) about 2TB in size, and holds about 4 billion potential signals. Later, we'll discuss how we sift through that many signals to try to find the extraterrestrial ones.

The SETI@home application program

SETI@home volunteers download the BOINC client software through a link provided on the SETI@home web site (*http://setiathome.berkeley.edu*). Standard version are available for Microsoft Windows, Apple Macintosh, and Linux systems. Versions ported to many other systems are available through the BOINC website (*http://boinc.berkeley.edu*). After installation, the BOINC client will provide a list of projects that the volunteer can join. After joining SETI@home little user interaction is required. The BOINC client software will automatically contact the SETI@home server to request work. The server will reply containing the URLs at which the BOINC client will download the SETI@home application and each of the work unit file to be processed.

[FIGURE 2 THIS PAGE OR LATER]

[FIGURE 2 CAPTION: Figure 2. A screenshot of the SETI@home application graphical display. The bottom half of the screen presents the power spectrum currently being analyzed. The upper left shows analysis state and the results of the current analysis. The upper right section shows information about the data being processed and the user's statistics.]

If the user wishes to have more control over how work is processed they can set preferences as to whether work will be processed while the computer is in use or to set hours of the day when processing can be performed. For Microsoft Windows and Apple Macintosh the user has the option of using a BOINC screen saver. (Figure 2) This screen saver allows the processing application to generate graphics that will be displayed when the screen saver is active. If the running application does not generate graphics, the screen saver displays statistics about the running application.

After receiving a work unit the application performs a baseline smoothing on the data to remove any wide band ($\Delta\nu$ > 2 kHz) features. This prevents the application from confusing fluctuations in broad band noise (due in part to variations in the Hydrogen line emission as the field of view transits the sky) with intelligent signals. The application then begins the main data analysis loop, which is shown schematically in Figure 3.

[FIGURE 3 HERE OR LATER]

[FIGURE 3 is the text pseudo-code given below. It should be typeset in a fixed width font such as Courier or similar.

```
for Doppler drift rates from -100 Hz/s to +100 Hz {
   for bandwidths from 0.075 to 1220 Hz in 2X steps {
      Generate time ordered power spectra.
      Search for short duration signals above a constant threshold (spikes)
      for each frequency {
         Search for faint signals matching beam parameters (Gaussians)
         Search for groups of three evenly spaced signals (triplets)
         Search for faint repeating pulses (pulses)
      }
   }
}
```

FIGURE 3 CAPTION: Figure 3. A pseudo-code representation of the SETI@home processing method.
]

At the start of each passage through the loop, the data is transformed in to an accelerated frame of a given Doppler drift rate. The drift rates at which the application searches the data for signals vary from -30 Hz/sec to +30 Hz/sec (accelerations expected on a rapidly rotating planet) in steps as small as 0.0009 Hz/sec. The application also examines the data at Doppler drift rates out to ±100 Hz/sec (accelerations of the magnitude that would arise from a satellite in low orbit about an earth-like planet), but at a more coarse step of 0.015 Hz/sec. A signal from an alien world would be most likely to have a negative drift rate (as the accelerations involved would be away from the observer). Despite this, we examine both positive and negative drift rates for purpose of statistical comparison and to leave open the possibility of detecting a deliberately chirped extraterrestrial signal.

At each drift rate the application searches for signals at one or more bandwidths between 0.075 and 1221 Hz. This is accomplished by using DFTs of length 2^n (n=3,4,..,17) to transform the data into a number of time ordered power spectra. In order to avoid repeating work, not all bandwidths are examined at every Doppler drift rate. Only when the change in drift rate becomes significant compared to $1/\Delta\nu^2$ is another DFT of that length computed. Therefore 32k-point transforms are performed one quarter as often at 64k-point transforms.

The transformed data is examined for signals that exceed 24 times the mean noise power. This threshold corresponds to 2.0×10^{-25} W/m^2 at our finest frequency resolutions, or the equivalent of detecting a cheap cell phone on one of the moons of Saturn. The SETI@home application reports any such "spike" signals in the result transmission.

If there is sufficient time resolution in the transformed data (n<15) and the SETI receiver is not tracking an object on the sky the application examines it for signals which match the parameters of the telescope beam. As a radio source drifts through the field of view, the measured power will vary depending upon the beam profile of the telescope. This profile is approximately Gaussian. The SETI@home application performs a χ^2 curve fit on any signals which exceed 3.2 times the mean noise power and reports those for which the goodness of fit exceeds a certain level. This power level typically corresponds to 2.1×10^{-25} W/m^2.

The application then divides transformed data at each frequency into chunks with duration equal to the time required for an object to transit the telescope field of view. These chunks are examined for pulsed signals using two algorithms. The first algorithm, the triplet finder searches each chunk for three evenly spaced signals that each exceed 9.1 times the mean noise power (as little as 2.5×10^{-25} W/m^2), and reports any detected signals.

The second algorithm is a modified fast folding algorithm (FFA). A folding algorithm divides the data into chunks of duration equal to the period being searched and co-adds them in order to improve signal to noise ratio. An FFA performs this function on a large number of periods without duplicating additions. The SETI@home folding algorithm searches roughly $N \log N$ pulse periods, where N is the length of the input array, between 2 samples and $N/3$ samples. During a typical run of the application this typically means half a million periods between 2 ms and 10 s. The threshold for detection of a pulsed signal is computed dynamically to match the number of co-added samples, and can be as low as 0.04 times the mean noise power for pulses with periods less than 10 ms. This corresponds to pulse energies of about 4.4×10^{-27} J/m^2.

This processing loop requires over 5 trillion floating point operations (teraFLOP). For an average work unit the SETI@home application would report one spike signals, one Gaussian, one pulsed signal, and one triplet signal.

Astropulse

[FIGURE 4 THIS PAGE OR LATER]

[FIGURE 4 CAPTION: Figure 4. An illustration of dispersion of a broadband pulse. The upper figure shows the waveform of a pulse that has undergone a small amount of dispersion. The lower figure shows a pulse after more dispersion has slowed the low frequency components. Note

the high frequencies are located at the left side of the figure which indicates they arrive first.]

One advantage of the BOINC infrastructure is that adding an additional application that processes a different data format is relatively straightforward. All that is necessary is to build the application, a work generator, a validator, and an assimilator. I mentioned earlier that one possibility is that an extraterrestrial wishing to attract attention might send a short (microsecond) duration broad band pulse rather than a narrow band signal of long duration. Detecting such a pulse presents some challenges because of how a broad band signal interacts with the tenuous gas that fills interstellar space. In most of the volume of space the gas through which a signal would traverse is at least partially ionized into a plasma of positive ions and free electrons. As the radio wave passes an electron, the electric and magnetic fields in the wave try to shake the electron at the frequency of the wave. The longer the wavelength (which also means the lower the frequency) the more the electron is able to interact. This interaction tends to slow the speed at which the radio wave propagates. This process spreads out a wide band signal by delaying the low frequencies more than it delays the high frequencies. (See Figure 4). This process is called dispersion, and it can be reversed with mathematical manipulations similar to those used by SETI@home to correct for Doppler drift. Fortunately, this process only depends upon how many free electrons lie on the line of site from the transmitter to the observer rather than on the details of that distribution. This quantity is, in detail, the integral of the electron density along the line of sight. Astronomers call it the Dispersion Measure (DM) and usually report it value in cm^{-3} parsecs.

Unfortunately, we don't know where the transmitter is, so we don't know how many electrons are between it and us. So we correct for reasonable values of galactic dispersion where a signal might be seen, from 49.5 to 830 cm^{-3} pc. Because an extraterrestrial might transmit a signal that is negatively dispersed either as an indication the signal is artificial, or as precompensation for dispersion toward the target of the signal, and because seeing negatively chirped interference help us to characterize the interference in our data we also look at the same range in negative dispersion as well.

In one way, Astropulse uses a simpler method than SETI@home; because we are looking for a broad band pulse we don't want to divide the recoded data by frequency. The work generating splitter for Astropulse merely needs to divided the data in chunks of about 13 seconds duration (a typical beam transit time). Both these and the Astropulse application are sent out to our volunteers. No additional action is required on the part of SETI@home volunteers to receive Astropulse work, although they may opt out if they wish. Because Astropulse work units take ten times longer to process than SETI@home work units, the BOINC server must check to be sure the volunteer's computer is capable of processing the data in a reasonable amount of time before assigning Astropulse work to it.

The algorithm is fairly simple. We dedisperse the data at a specific dispersion measure, which generates a time series representing the signal power in the dedispersed frame with a time resolution of 0.4 μs. If any events are above threshold they are reported. We then co-add adjacent bins, to improve sensitivity to longer timescale pulses, again looking for events above threshold. We repeat this co-add 8 more times, examining the data for signals at timescales from 0.4 to 204.8 μs. Then we move on to the next dispersion measure (usually stepping 0.05 cm^{-3} pc). At some dispersion measures we perform a folding algorithm similar to that used by SETI@home to detect repeating pulses. More details of the Astropulse algorithms have been presented by Von Korff (2010). We have set the thresholds such that ~1 pulse will be detected in a work unit filled with random noise. As we have discovered, there are many dispersed terrestrial signals that result in many signals be detected on average.

Post-processing

When the applications have done their work, the job isn't done. Typically the application programs return a few potential signals per work unit. Of course, not all of these signals are evidence of extraterrestrial intelligence.

Some of the signals are due to errors made in the processing computers. Typical numeric processors, memory and disk systems are fairly reliable. However SETI@home and Astropulse thousands of years of CPU time per day, magnifying even low error rates. Event if undetected errors occur only on average every 10^{18} machine instructions, SETI@home would see several per day. To combat these effects our validator examines each signal to see if the parameters match their permitted values. We also send each work unit to multiple volunteers, and cross check the returned values to verify accuracy.

A large number of the signals in the database are evidence of terrestrial intelligence. Sources of narrow band radio emission are ubiquitous where human technology is present. The sources of dispersed broad band emission (primarily radars) are even stronger. Even at the Arecibo observatory, where care is taken to minimize interference, this noise is present, due to airport and air defense radars, local equipment, aircraft, satellites, and other transmitters. Most of the time these terrestrial emissions are fairly easy to distinguish from an extraterrestrial signal.

It's possible to mitigate the effect of radars both at the telescope, and in our data processing pipeline. At the Arecibo telescope, the observatory maintains an antenna which monitors the most powerful radar and a device known as the "radar blanker" that predicts when the radar pulses will arrive. An observer can use the prediction to replace the telescope data with a noise-like signal during the time when the radar pulse might arrive. We have developed a second system that works in a similar fashion after the fact by examining our recoded data for the radar signals. We can then fit the known radar patterns to what is seen and remove one or more of the contaminating radar patterns. This has greatly reduced the number of RFI signals that are being stored in the SETI@home and Astropulse databases.

[FIGURE 5 FULL PAGE, HERE OR LATER]

[FIGURE 5 CAPTION: Figure 5. These plots show the frequency distribution of pulses detected by SETI@home. The upper panel shows all pulses. The middle panel shows pulses determined to be due to persistent interference sources. The lower panel shows the pulse frequency distribution after the interference has been removed.]

A large fraction of RFI consists of continuous narrow band signal generated at or near the observatory. We use this property to detect it signals in the zones containing it. The RFI frequency zones are typically quite narrow. We have identified 35,000 frequencies, covering less than 1% of our total bandwidth, which are subject to frequent interference. These zones contain between 5% and 20% of the detected signals depending upon signal type. For example, the top panel of Figure 5 shows the frequency distribution of 378,362,077 potential pulsed signals detected by SETI@home between July 5, 2006 and September 16, 2009. The vertical bands that are present indicate frequencies that are over-represented and are probable RFI frequencies. We use a statistical analysis to determine which frequencies appear too frequently on differing sky positions to be due to noise processes. Those frequencies define the exclusion zones. Pulses determined to be within these zones (6.6% of the total) are shown in the middle figure. The lower figure show the distribution of pulses that remain after those within zones have been removed.

Other RFI sources are of short duration and repeat on time scales of hours to days. So any signal that repeats after a short time when the telescope is viewing a different portion of the sky should also be rejected. After RFI is removed, the bulk of the remaining signals are due to random fluctuations in the noise background mimicking an extraterrestrial signal. One means of sorting out the true extraterrestrials is by looking for persistent signals. We expect that an extraterrestrial signal will be present at a similar frequency the next time the same celestial location is examined. We have developed a program called a Near Time Persistency Checker (NTPCkr) that sorts through the database looking for persistent signals. When it finds one, it sends it off to the RFI removal program, to make sure that it is not due to RFI. Nearly every time, the RFI removal program finds that the signal was due to an RFI event. For those aren't flagged as RFI are added to a candidate list which is then proposed for telescope time to reobserve them. Thus far no reobservation has confirmed the detection of a candidate.

Distributed Thinking

Each candidate on the candidate list must be verified by a human being before being confirmed at a target for reobservation, primarily because automated means of RFI detection are insufficient. Temporary RFI sources often appear and disappear or shift frequencies in ways that cannot be easily detected by automated software. We are hoping to develop a means for our volunteers to help identify RFI in time vs frequency "waterfall plots" by first training them on manufactured data. They will then be able to examine our actual candidates and give an opinion on whether each signal is clean or due to interference. We're hoping this will help to speed through identifying candidates for reobservation from among the thousands of possibilities.

Distributed Development

There are other ways to distribute the SETI workload. One is distributed software development. Shortly before transitioning to the BOINC infrastructure, we released the source code to SETI@home, eventually settling on the General Public License (GPL) for our code. This enabled several developments. First, many bugs within the source code were brought to our attention. Most were minor, but some would have limited our ability to correctly identify candidates if we hadn't corrected for their effects. Second, was optimization of the code. There has always been an element of competition to SETI@home. People compete to see who can do the most work in the least time. Many of these volunteers developed optimized versions of SETI@home. Some found new algorithms to perform the same functions; others added support for single instruction multiple data instruction sets. Many of these contributions have been returned to us and included in the application we distribute. The current SETI@home application runs in about one twelfth the time that the original version would take, despite doing many times as much work.

SETIQUEST, a project run by the SETI Institute (*http://www.setiquest.org*), has even more ambitious distributed development goals. In addition to providing the source code for the existing SETI Institute data processing routines, they invite participants to download data and develop their own algorithms for detecting signals within the data. They are hoping to develop a group of citizen scientists which will help to improve current and future SETI searches.

The future of SETI@home

SETI@home was originally slated to process two years worth of data from the Arecibo telescope. The strong public response and new improvements to the application software have kept us going for eleven years. Recently we've started deploying versions of SETI@home that run on graphics processing units (GPUs) that are capable of highly parallel operations. SETI@home can compute on the GPU up to 30 times faster than the CPU on systems that contain a compatible GPU.

Despite this, SETI research lives in a perpetual state of being starved for computation resources. In the past twelve months we have discussed three new algorithms with other SETI researchers, one that we will probably implement soon will make very little change to the time require to process a work unit. The second would increase our processing time by factors of ten to one hundred. We are considering it for the future. The third, if implemented fully, would easily require all of the compute cycles executed by all of the computers that have ever existed on Earth in order to examine a small fraction of our data. If Moore's law continues, perhaps this will be possible before we realize.

SETI@home currently samples only a small portion of the radio spectrum, and a small portion of the sky. The two most obvious means of expanding its capabilities are to expand the sky coverage and widen the frequency bandwidth. The primary impediment to larger bandwidth is the SETI@home data recorder (which can record at 80 Mbps for a total recorded bandwidth of 40 MHz) and the available storage for maintaining the data. Very large baseline interferometry (VLBI) data recorders recorders in use at many observatories can record at 4 Gbps (total recorded bandwidth up to 2 GHz). That's enough to fill a 2TB disk drive in one hour. At current prices it would cost $1.2 million to buy disks to hold one year's worth of data when recorded at that rate. Needless to say, SETI@home doesn't have the financial resources to do that. But incremental improvements can be achieved for less cost.

The best means of expanding the sky coverage would be to add a SETI@home recorder system to a southern hemisphere radio telescope. This would allow us to increase our sky coverage from about 25% to 75%. We have considered this quite often, but thus far the resources to do so have not presented themselves.

As in any voluntary organization, it's important that SETI@home be responsive to the desires of its volunteers. The success of SETI@home is entirely dependent on the volunteers who provide the computing resources. We will continue working to keep our volunteers informed of our progress and to share with them the science behind SETI.

Acknowledgments

I'd like to thank the rest of the SETI@home/Astropulse and BOINC staff at the Space Sciences Laboratory: Josh Von Korff, Andrew Siemion, Jeff Cobb, Matt Lebofsky, Robert Bankay, Dan Werthimer, and David Anderson. SETI@home and Astropulse are supported by grants from NASA and the NSF and donations from the friends of SETI@home. We would also like to thank SETI@home volunteers everywhere for their valuable contribution of the processing power that makes SETI@home work. The observations for these projects were made at the NAIC Arecibo Observatory a facility of the NSF operated by Cornell University. This research uses resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

Korpela, E. J., Werthimer, D., Anderson, D., Cobb, J., and Lebofsky, M. 2000. "SETI@home-Massively distributed computing for SETI," *Computing in Science and Engineering*, 3 (1): 78-83.

David P. Anderson. "BOINC: A System for Public-Resource Computing and Storage," 5th IEEE/ACM International Workshop on Grid Computing, November 8, 2004, Pittsburgh, USA, pp 1-7.

Von Korff, J. 2010