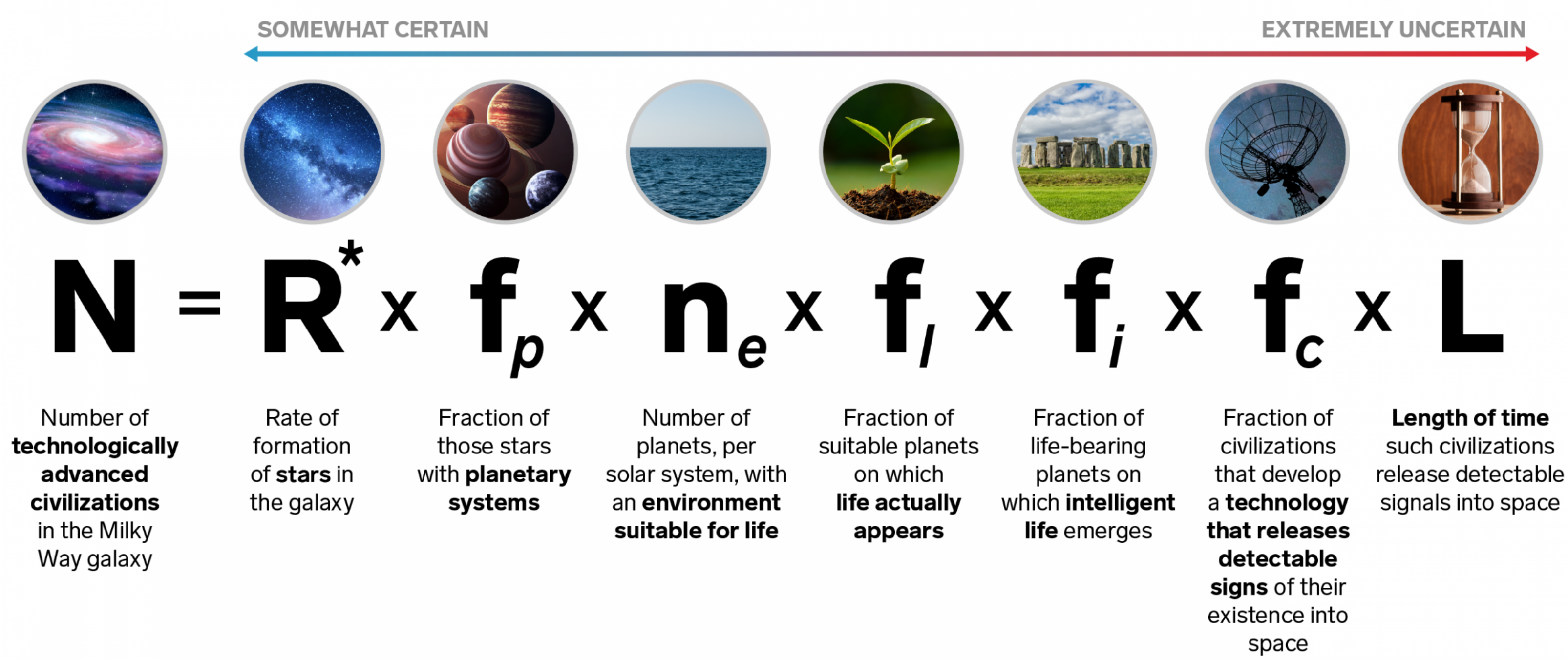
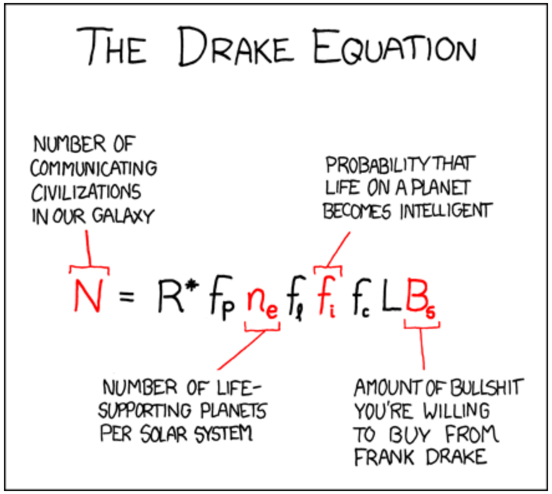
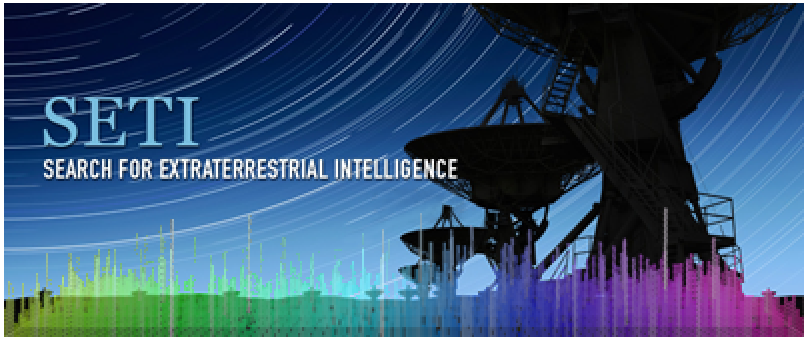
An interesting Example



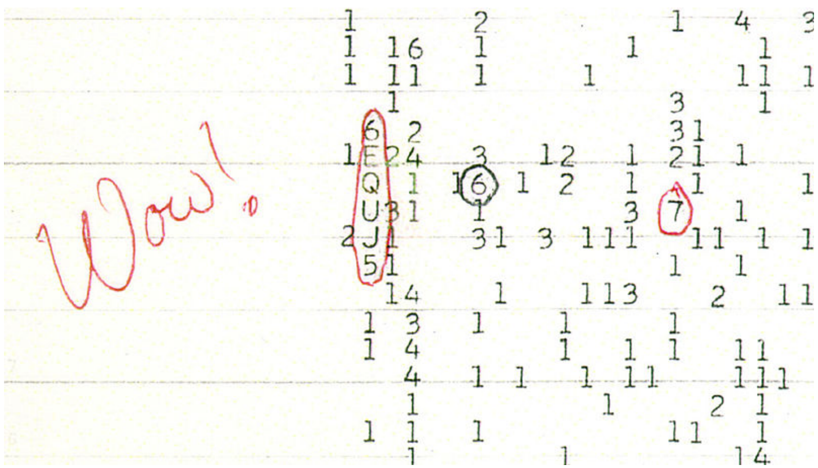


SETI: the search for Extraterrestrial Intelligence

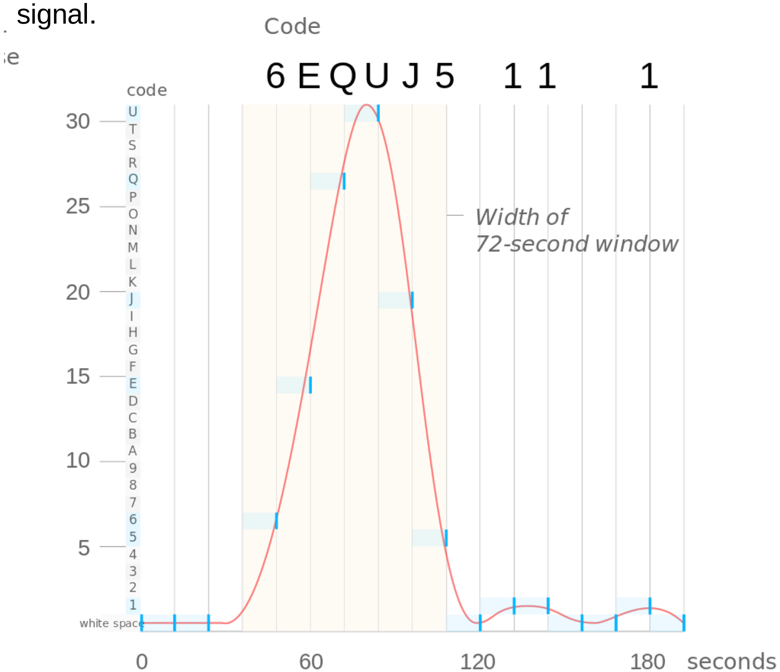


- uses radio telescopes to listen for narrow-bandwidth radio signals from space
- two goals:
  - to detect intelligent life outside Earth,
  - to prove the viability and practicality of the "volunteer computing" concept.

The search for extraterrestrial intelligence (SETI) project



A scan of a color copy of the original computer printout, taken several years after the 1977 arrival of the Wow! signal.



- Scientific investigation began shortly after the advent of radio in the early 1900s
- In 1960 F. Drake pointed the radio telescope 26 metres diameter at Green Bank at the stars Tau Ceti and Epsilon Eridani
- 400 kHz near the 1.420 GHz frequency – Nothing!
- August 15, 1977: Jerry Ehman witnessed a strong signal received by the telescope

SETI@home

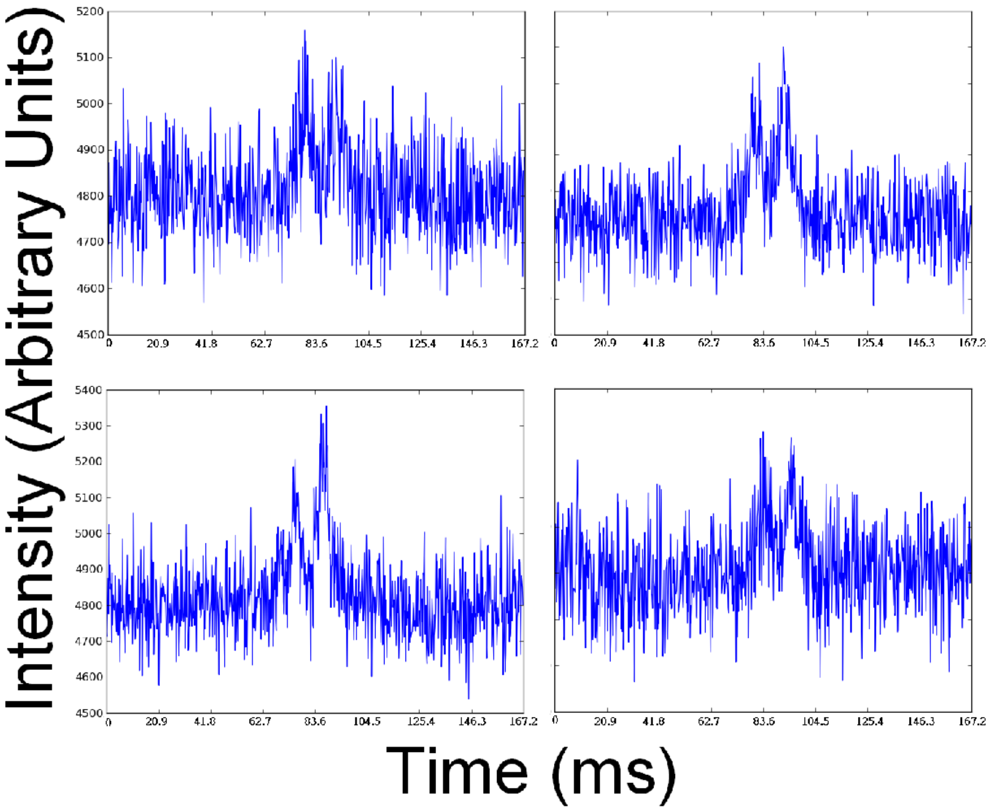
The Search for Extraterrestrial Intelligence

Arecibo Observatory in Puerto Rico (300 m dish), one of the world's largest filled-aperture (i.e. full dish) radio telescope, conducts SETI searches.

SETI@home

The Search for Extraterrestrial Intelligence

Another source for SETI: the Green Bank Radio Telescope (GBT) 100 m diameter: world's largest fully steerable radio telescope. In the center of the National Radio Quiet Zone.



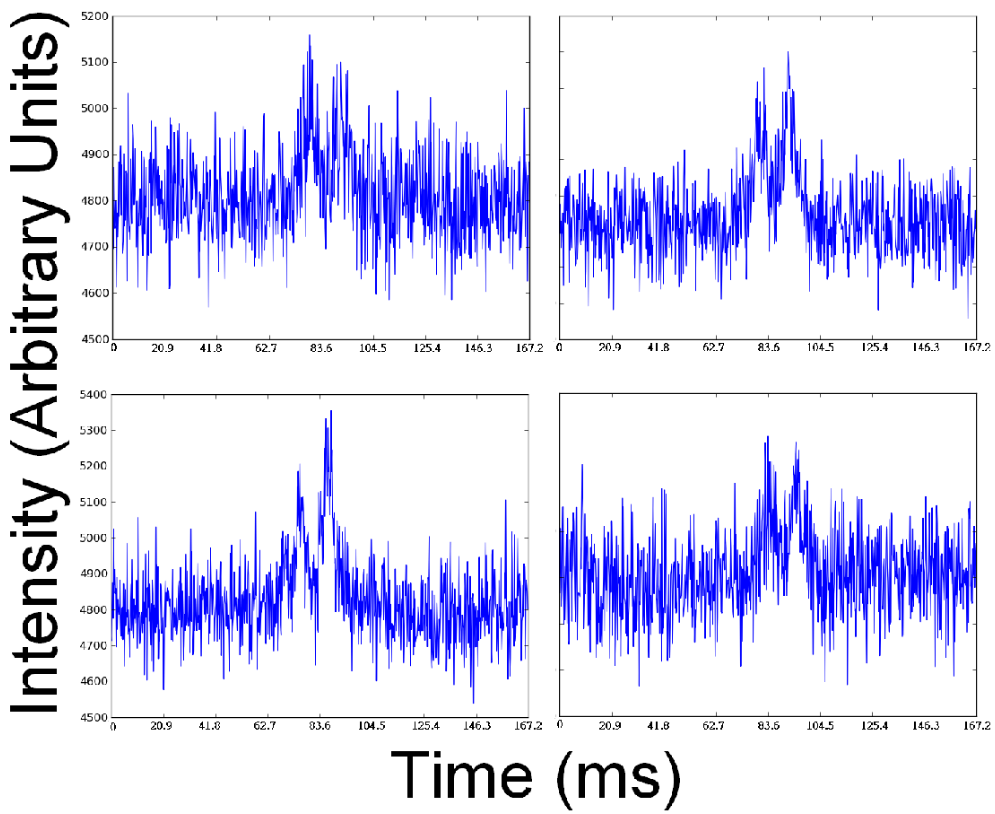
Screen shot of time domain signal samples

1) The data are taken while the telescope is used for other scientific programs.

2) The data are **digitized**, **stored**, and **sent** to the SETI@home facility.

3) The data are then **parsed** into small chunks in **frequency** and **time**.

4) Analyzed to search for **signals**: variations which cannot be ascribed to **noise**, and hence contain **information**

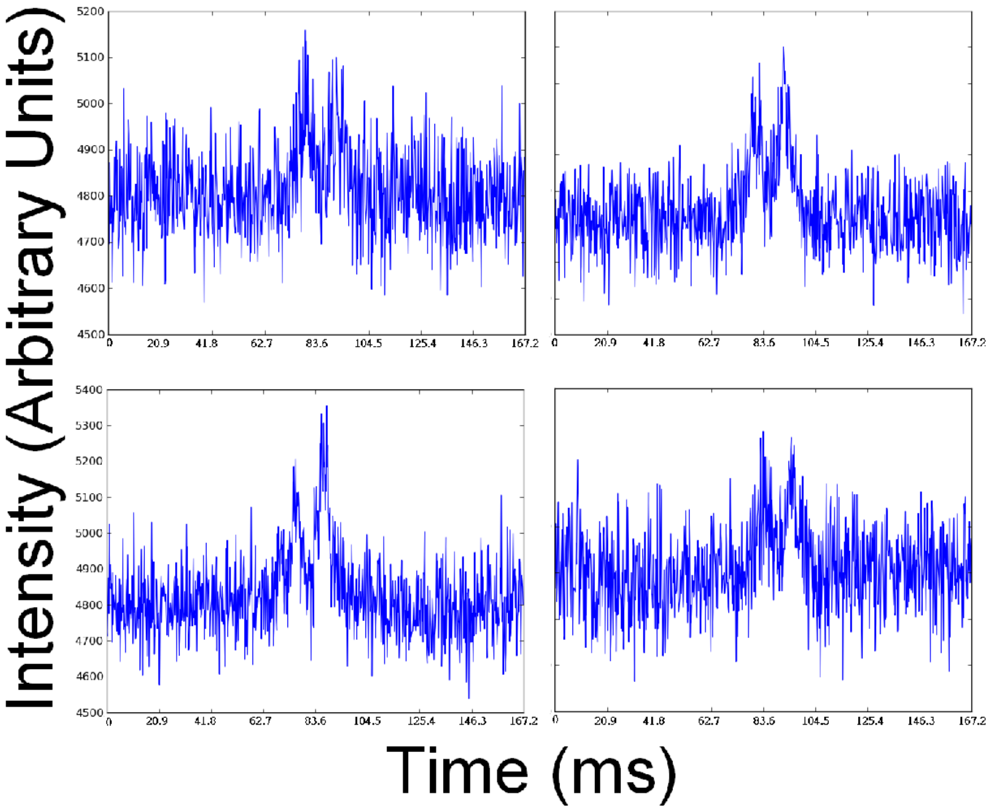


Screen shot of time domain signal samples

1) The data are taken while the telescope is used for other scientific programs.

POINTING:
Arecibo dish is fixed in place looking straight up at the zenith, but it can still be pointed through a a limited range: the telescope "sees" the sky between 2 degrees south declination and 38 degrees north declination and can track objects for 3 hours.
Actually though, SETI@home is not interested in tracking objects, but would rather have them drift through the telescope "**beam**" while the antennas sit **still**.

FREQUENCY:
Data are acquired in a **band** of the radio spectrum 2.5 MHz wide.
The SETI@home project breaks this wide band up into chunks of about 10 kHz each (actually 9765 Hz). This means that every 107 seconds of data recorded 256 blocks of data are produced.



Screen shot of time domain signal samples

2) The data are **digitized**, **stored**, and **sent** to the SETI@home facility.

The radio signal captured by the receiver is **amplified**, then **sampled** and **converted** to a ***binary digital form*** by an Analog-Digital Converter (ADC).

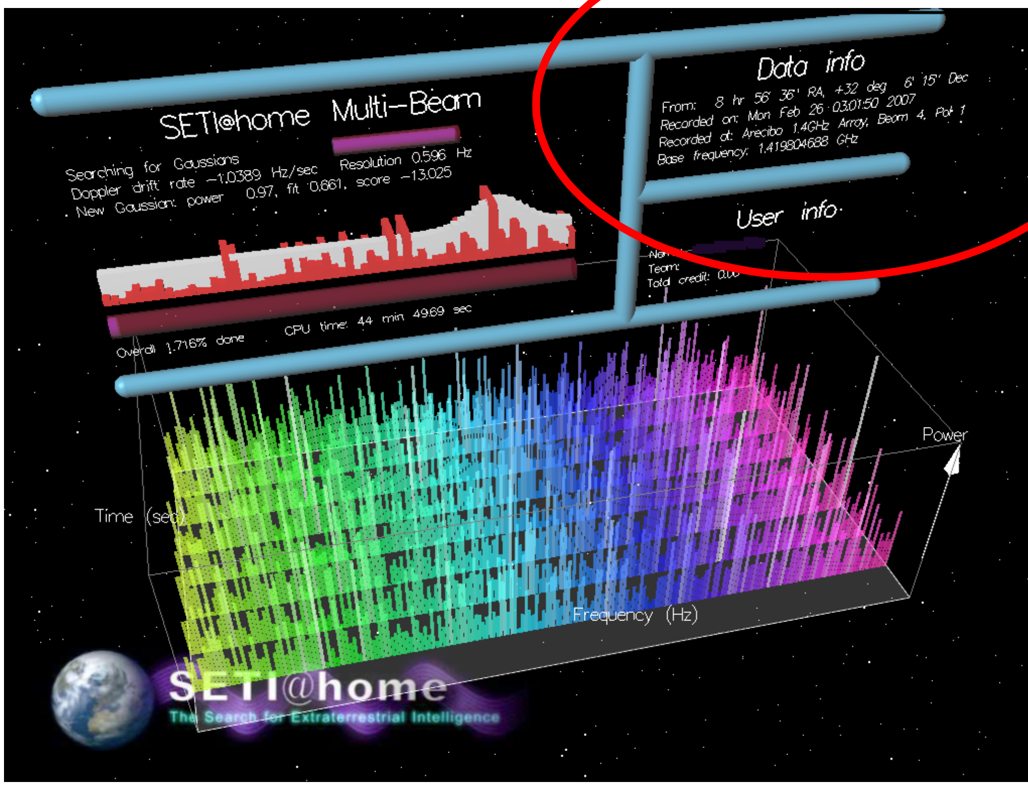
Observational data are **recorded** on 2TB SATA hard disk drives at Arecibo (2.5 days of observations each).
Then **sent** to Berkeley.

3) The data are then **parsed** into small chunks in **frequency** and **time**.

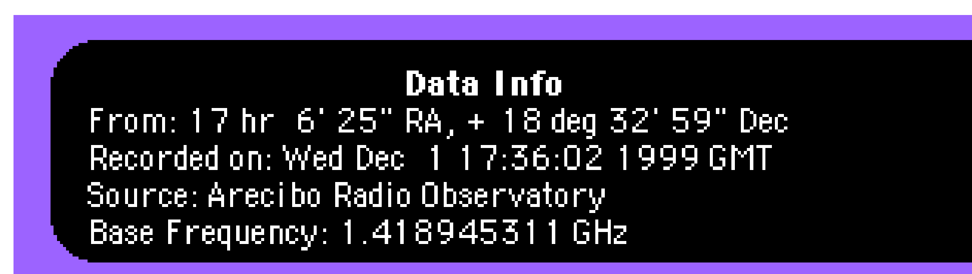
Divided in both time and frequency domains work units of 107 seconds of data or approximately 0.35 MB, which overlap in time but not in frequency.

These work units are then sent from the SETI@home server over the Internet to personal computers around the world to analyze.

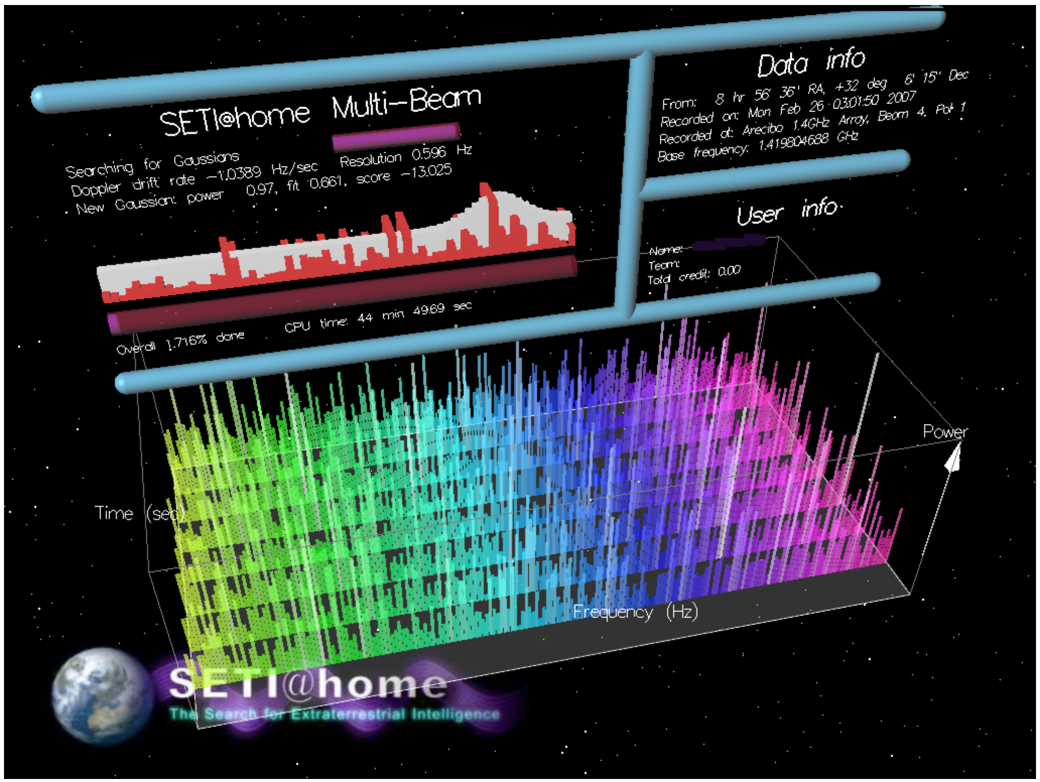
About 350 KB of actuall radio telescope data and about another 1k that describes the data (time data was taken, where in the sky, base frequency of this work-unit, etc…): the **header**



Screen shot of the screensaver for SETI@home, a distributed computing project in which volunteers donate idle computer power to analyze radio signals.



Contains information about the block of data currently being worked on.



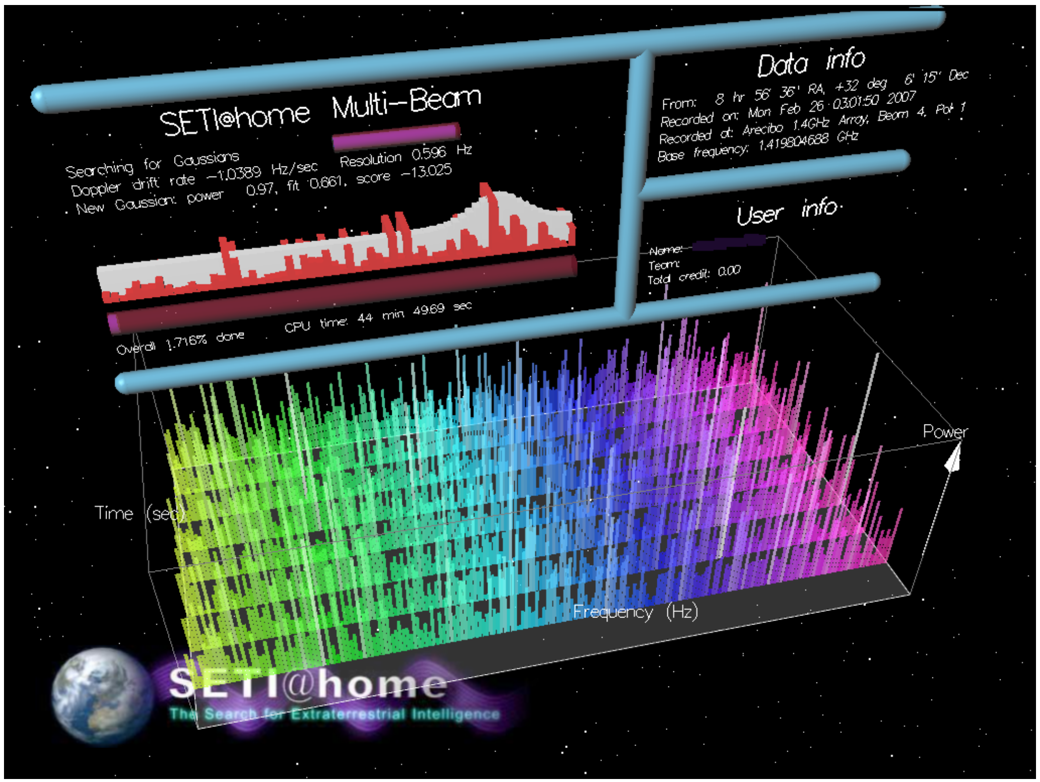
Screen shot of the screensaver for SETI@home, a distributed computing project in which volunteers donate idle computer power to analyze radio signals.

4) Analyzed to search for **signals**: variations which cannot be ascribed to **noise**, and hence contain **information**

The software searches for five types of signals that distinguish them from noise:

- Spikes in power spectra
- Gaussian rises and falls in transmission power, possibly representing the telescope beam's main lobe passing over a radio source
- Triplets — three power spikes in a row
- Pulsing signals that possibly represent a narrowband digital-style transmission
- Signal waveforms checked via autocorrelation.

- Spikes in **power spectra**
  - PRE-Processing:
  - Doing **Baseline Smoothing**
  - **Computing Fast Fourier Transform**

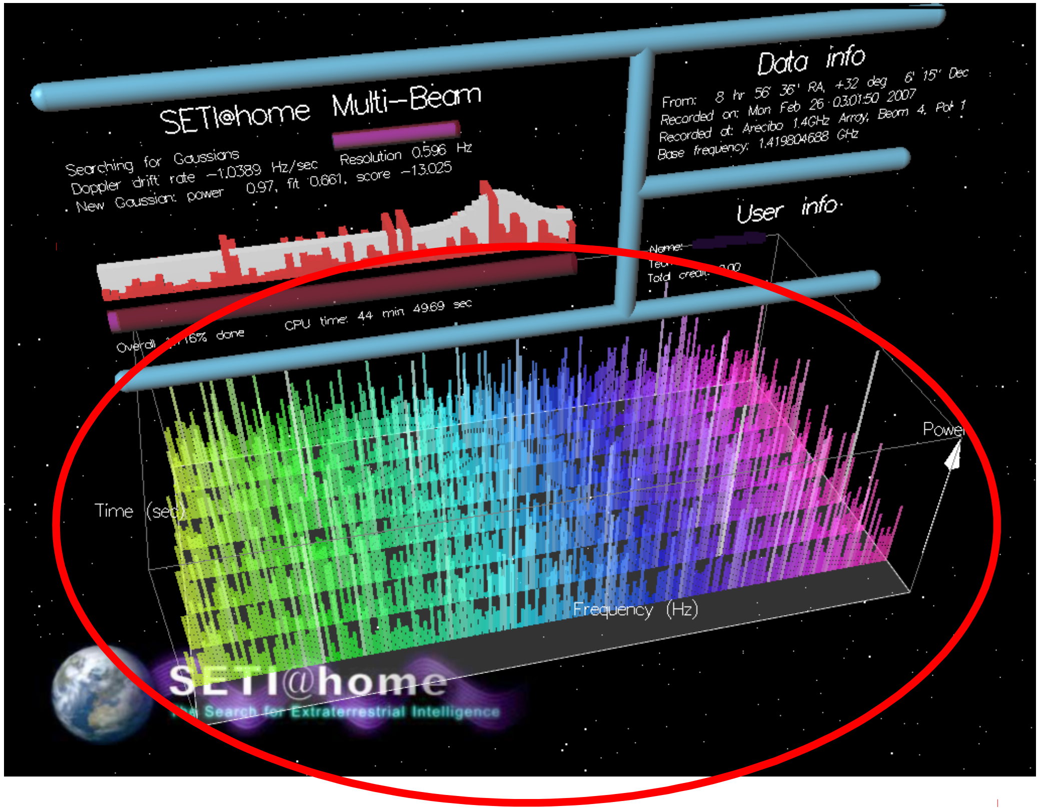


Screen shot of the screensaver for SETI@home, a distributed computing project in which volunteers donate idle computer power to analyze radio signals.

  - We are interested at narrow bandwidth signals, since broadband signals are most likely due to natural astronomical processes.
  - To reject broadband noise, the screensaver does an "average" through the data that eliminates this broadband noise and brings all the other narrow bandwidth signals down to a common "baseline" level.
  - Also, over the 107 seconds the signal sometimes gets slowly louder and/or softer. Baseline smoothing brings it all to the same level.

  - To turn a set of time-based data into a set of frequency-based data, we apply a mathematical operation called a "fast fourier transform" or FFT. For more information on the FFT, please wait for the rest of the course!

- **Computing Fast Fourier Transform**

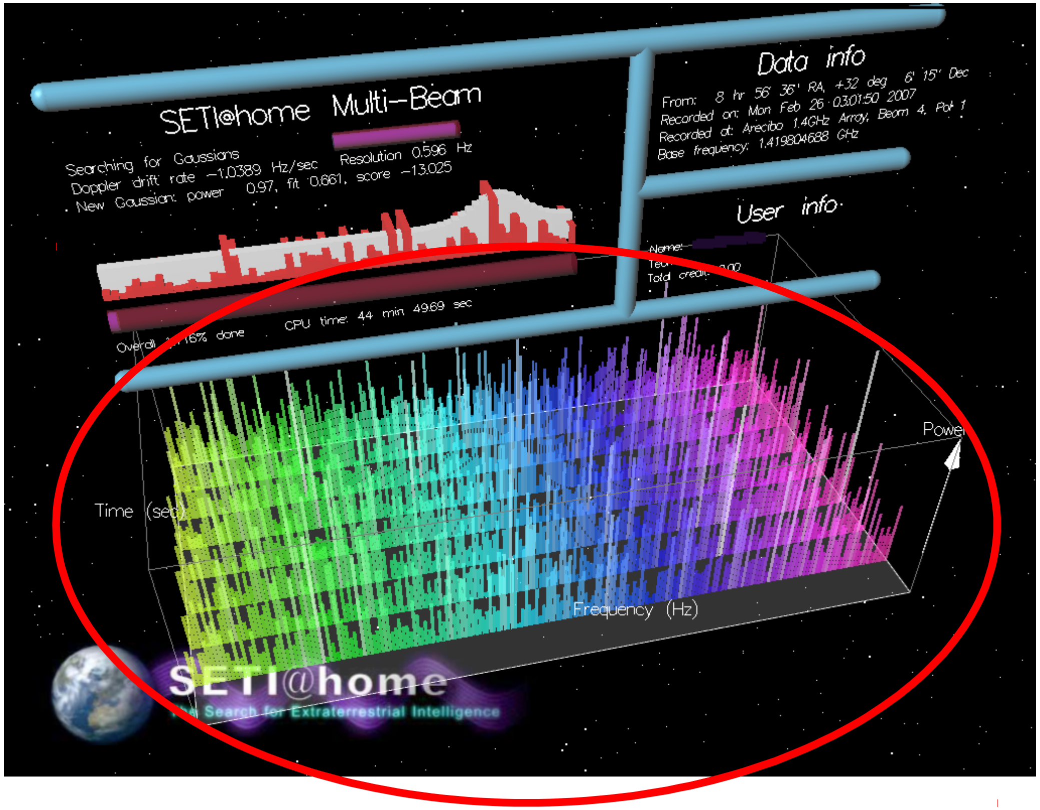


Screen shot of the screensaver for SETI@home, a distributed computing project in which volunteers donate idle computer power to analyze radio signals.

- The result of this processing is the graph produced in the lower frame of the screen saver.
- At the beginning of a work-unit, we do 15 different FFT's, each looking at the data with varying accuracy. We start looking for details as small as .075 Hz wide. There are tradeoffs when you are doing this kind of analysis. If you want to be very accurate in frequency, you have to observe the data for a longer time.
- At 0.075 Hz frequency resolution, we must look at chunks of data 13.42 seconds in length. To completely analyze our 107 second sample, we need to do 8 of these FFT's.

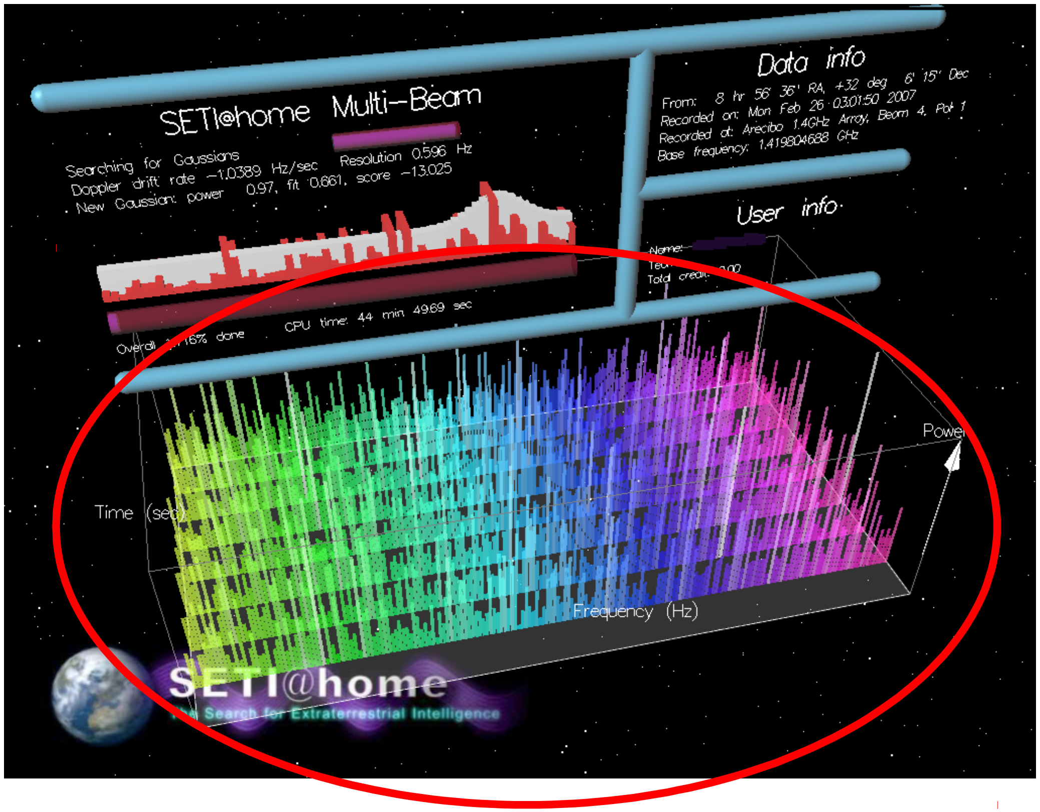
- When the frequency resolution is reduced to 0.14 Hz, we have 6.7 second samples of data. Less frequency resolution, but more time resolution.
- We have twice the number of these sub-sets to cover the 107 seconds of data!
- We employ 15 different frequency resolutions (0.075, 0.15, 0.3, 0.6, 1.2, 2.5, 5, 10, 20, 40, 75, 150, 300, 600, and 1200 Hz) in the analysis. With each halving of the frequency resolution we do twice the number of FFT's to cover the 107 seconds of data.

- **De-Chirping**
  - The signal may be emitted from a moving source and is received on a moving planet. → Doppler
  - The algorithm first takes the raw data and mathematically "undoes" a specific Doppler *acceleration* or "**chirp**".
  - It then feeds the resulting "de-accelerated" data to the FFT routine.
  - SETI@home tries to do this at many points between -50 Hz/sec to +50 Hz/sec.
  - At the finest frequency resolution of 0.075 Hz we check for 5409 different chirp rates between -10 Hz/sec and +10 Hz/sec!



Screen shot of the screensaver for SETI@home, a distributed computing project in which volunteers donate idle computer power to analyze radio signals.

- Therefore, **MANY** FFTs have to be computed for each 107.4 s data units



Screen shot of the screensaver for SETI@home, a distributed computing project in which volunteers donate idle computer power to analyze radio signals.

- **De-Chirping**
  - The signal may be emitted from a moving source and is received on a moving planet. → Doppler
  - The algorithm first takes the raw data and mathematically "undoes" a specific Doppler *acceleration* or "**chirp**" (drifting in frequency).
  - It then feeds the resulting "de-accelerated" data to the FFT routine.
  - SETI@home tries to do this at many points between -50 Hz/sec to +50 Hz/sec.
  - At the finest frequency resolution of 0.075 Hz we check for 5409 different chirp rates between -10 Hz/sec and +10 Hz/sec!

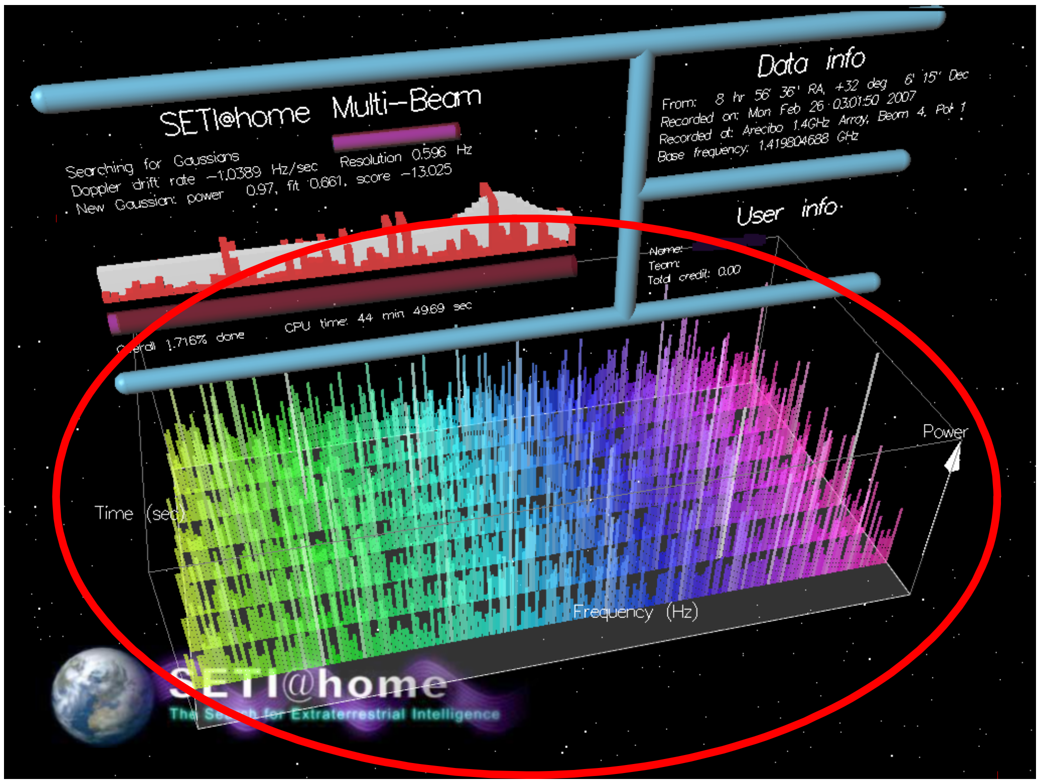
- Therefore, **MANY** FFTs have to be computed for each 107.4 s data units

- Spikes in **power spectra**

  - For a given signal, the power spectrum (or Power Spectral Density) gives a plot of the portion of a signal's power (energy per unit time) falling within given frequency bins.

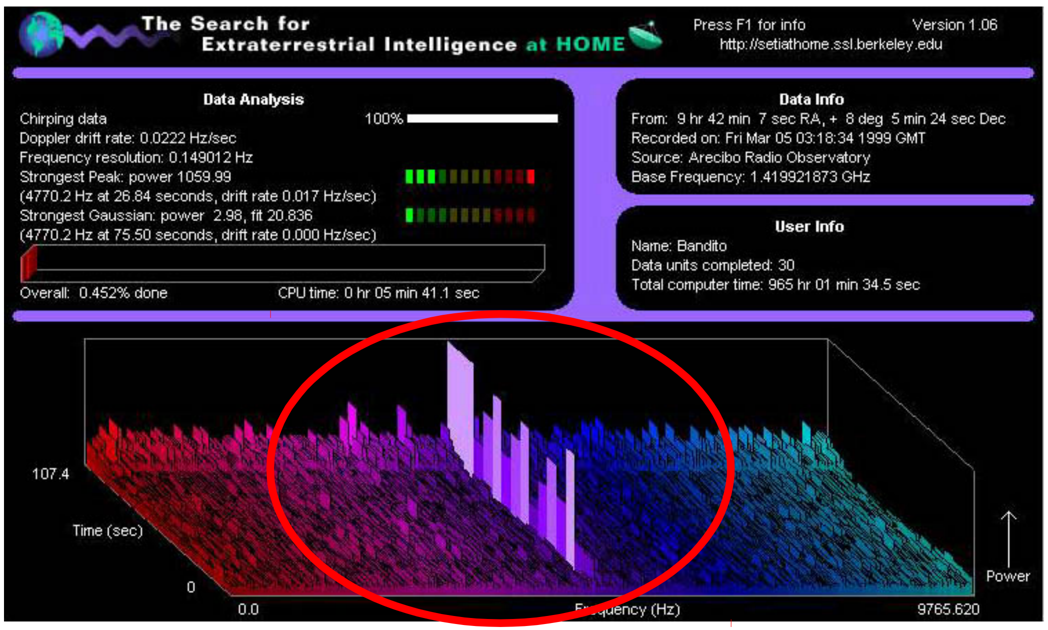
  - If we do not have to worry about limitations in the data—i.e. we have a continuous time series y(t) infinite in length—the power spectrum of the signal would be given simply by the Fourier transform: $P(\omega) \propto |\tilde{y}(\omega)|^2$

  - Since we are dealing with a **finite time** series and perform analysis of **sampled** and **digitized data**, we take a number of power spectra (from different sets of measurements over intervals of length T) and average the power spectra.



Screen shot of the screensaver for SETI@home, a distributed computing project in which volunteers donate idle computer power to analyze radio signals.

- Spikes in **power spectra**



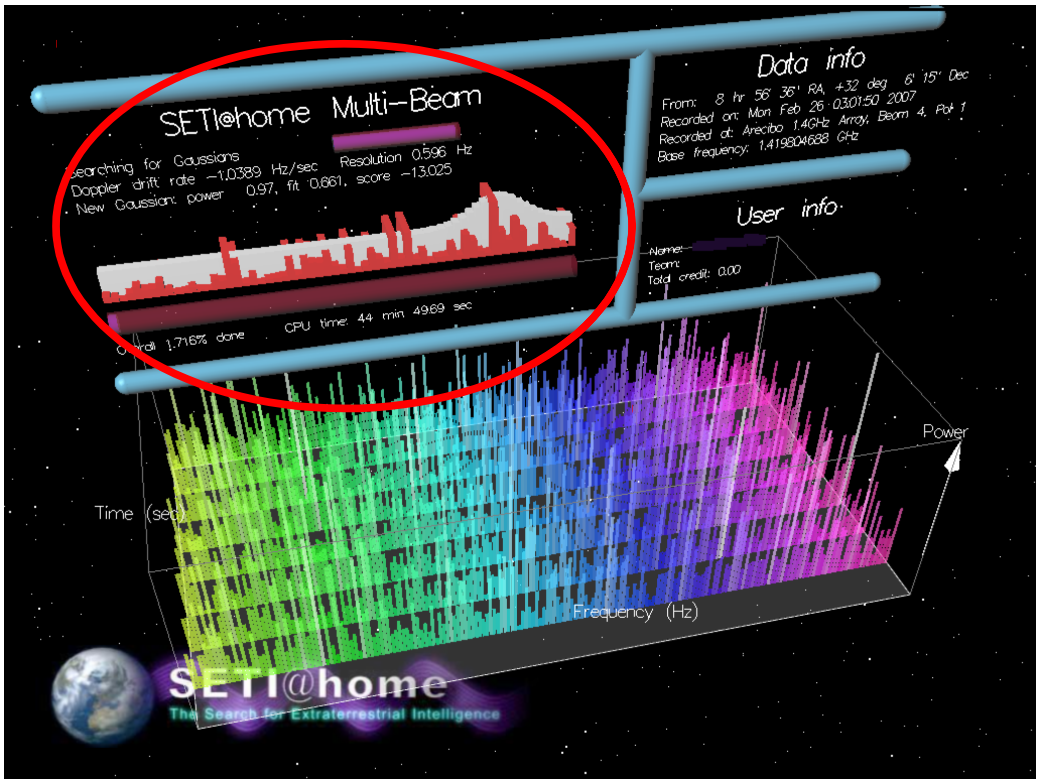
Graph that show narrowband interference. This form is usually generated by television, radio, and satellite transmitters. Note that the signal is more or less constant, lasting the entire 107 seconds of the data.

- For a given signal, the power spectrum (or Power Spectral Density) gives a plot of the portion of a signal's power (energy per unit time) falling within given frequency bins.

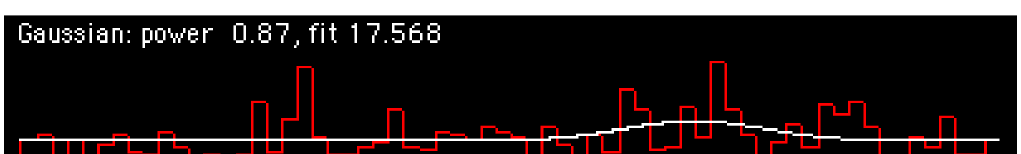
- If we do not have to worry about limitations in the data—i.e. we have a continuous time series y(t) infinite in length—the power spectrum of the signal would be given simply by the Fourier transform: $P(\omega) \propto |\tilde{y}(\omega)|^2$

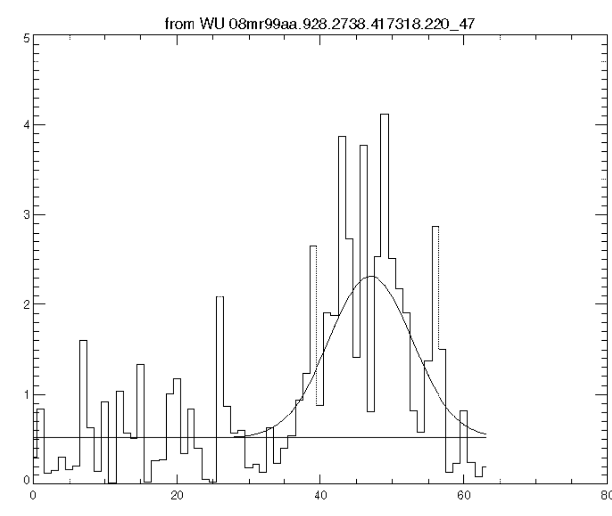
- Since we are dealing with a **finite time** series and perform analysis of **sampled** and **digitized data**, we take a number of power spectra (from different sets of measurements over intervals of length T) and average the power spectra.

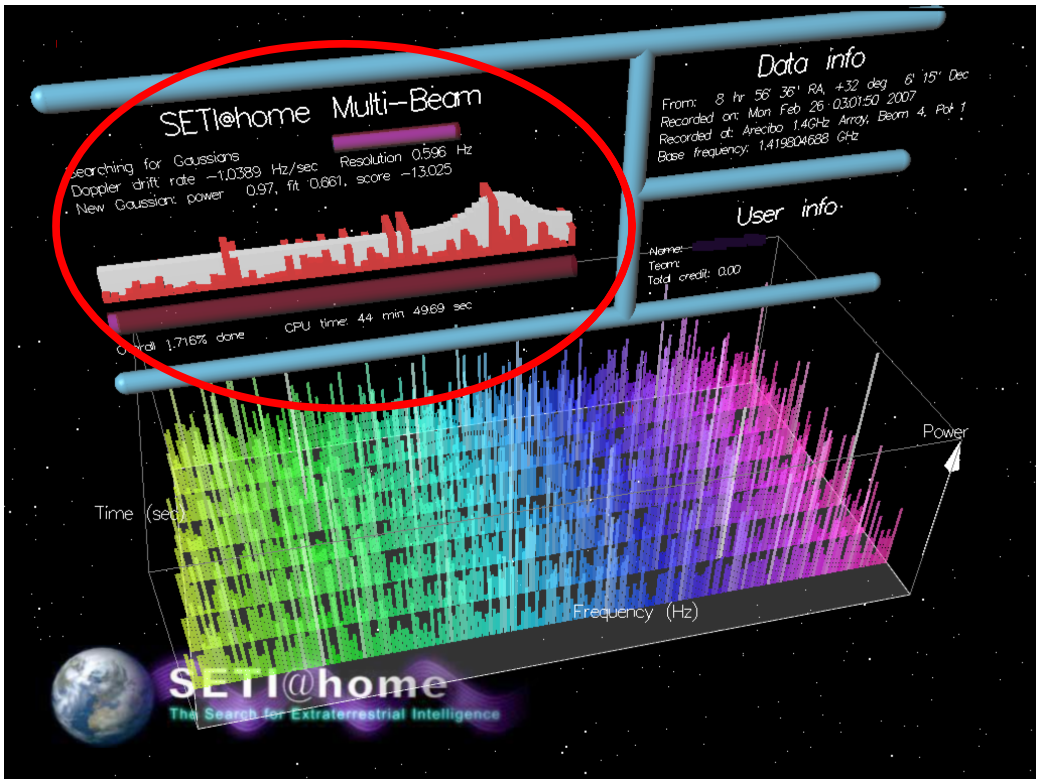
- Gaussian rises and falls in transmission power
  - A celestial object takes 12 seconds to pass through the telescope beam main lobe.
  - A signal from an earthbound source would not get louder and then softer over the 12 second period
  - When the time resolution is high enough, we can search the data for Gaussian rises
  - Since we are looking for these 12 second "Gaussians", the 107 seconds data units overlap by 15 seconds



Screen shot of the screensaver for SETI@home, a distributed computing project in which volunteers donate idle computer power to analyze radio signals.





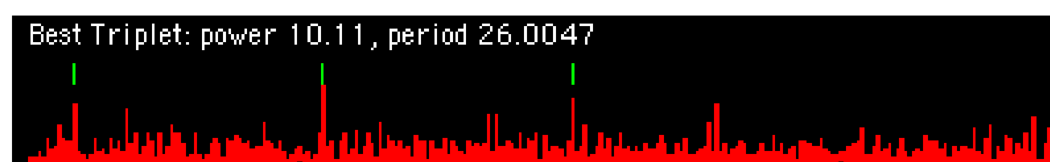


Screen shot of the screensaver for SETI@home, a distributed computing project in which volunteers donate idle computer power to analyze radio signals.

- Searching For Pulses / Triplets

  - Signal may be a series of closely or widely spaced set of pulses.
  - This is more economical power-wise at the transmitter than broadband signals.

  - It is done by the Fast folding algorithm: it computes superpositions of the signal modulo various window sizes simultaneously. (Check this out: https://arxiv.org/abs/1703.05581)



SETI: an interesting example

- Data are:
  - **Digitized** → sampled and converted to *digital form*
  - **Stored** → Copies, Redundancy, Compression
  - **Sent** → data description needed: the **header**

- Data contain:
  - **True Signal?**
  - **Noise(s)** → photon noise, digitization noise, thermal noise...
  - **Systematics** → telescope aperture, interferences...

- Data is pre-processed:
  - **Smoothing**
  - **Baseline subtraction**
  - **De-chirping**

- Data is analysed by:
  - **Fourier Transform**
  - **Peak search + Curve Fitting**
  - **Signal folding**
  - **Autocorrelation**