Installation of TurboSETI on the BreakThrough Listen processing node

Aoife Brennan

Supervisor: Dr. Evan Keane

I-LOFAR

Acknowledgements

I would like to express my very great appreciation to Evan Keane for mentoring me during my internship. I would also like to offer my special thanks to David McKenna, for all the guidance during the internship. I would like to thank Steve Croft and Howard Isaacson for organizing and running the BL internship program. Finally, I would like to thank everyone at I-LOFAR and BL.

Abbreviations

Dispersion measurement	DM
Signal to Noise	SN
BreakThrough Listen	BL
Irish Low Frequency Array	I-LOFAR
Fast Radio Burst	FRB
Radio Frequency Interference	RFI
Interstellar Medium	ISM
Bandwidth	BW
Fast Fourier Transform	FFT
Search for Extraterrestrial Intelligence	SETI
Transiting exoplanet Survey Satellite	TESS
Polyphase Filterbank	PFB
Green Bank Telescope	GBT

1 Introduction

During the summer of 2020, I undertook an internship with I-LOFAR and the BL initiative. The internship's main aim was to set up the newly installed BL computer backend located on the I-LOFAR site. After accomplishing this, the first SETI search can be carried out on TESS data. The BL's main initiative is to search for intelligent extraterrestrial communications in the universe. Currently, BL uses the GBT and the Parkes telescope, for observations, after setting up the BL backend at I-LOFAR, the station can be used to carry out SETI search.

To carry out a SETI search, we need to install special software designed by the BL initiative. Firstly the software Blimpy, which is used to handle and manipulate the raw files. Secondly, the software TurboSETI, which is used to search data for narrowband drifting signals in filterbank data. The code's main purpose is to identify narrowband signals that are not produced in nature, which could have extraterrestrial origins. Both softwares and all dependencies need to be installed on the blc00 node.

First, I set up TurboSETI and Blimpy on my personal laptop; then, I set up TurboSETI on blc00. I used two tutorials to ensure that I had set up everything correctly using the data provided to test various components of TurboSETI. After setting up TurboSETI on blc00 I examined multiple .h5 and .fil files from different sources, one of my favorite targets is described in the report below.

This report contains four sections the first describes the process of how I set up TurboSETI on the BL processing node. The second gives background on how TurboSETI and Blimpy works, along with a brief demonstration and description of parameters used in the search. The third section contains a description of how I used TurbSETI to examine data on the BL open database. The last section contains my results along with a discussion.

2 Installation

As both softwares are underdevelopment, I downloaded the most update to date versions directly off of github using git clone. I first created a virtual environment in my personal directory on blc00, then I could install both Blimpy, TurboSETI and all the dependencies needed. After downloading, there were a few changes that I had to make to the source code. First, I had to remove the two error messages from the TurboSETI source code in `turbo_seti/turbo_seti/find_doppler/data_handler.py`

The error message

`"errmsg = "Error encountered when trying to open file: {}".format(self.filename)` in the function `def __split_h5(self):`,

the same error message in the function `def __init__` is also removed. Then I had to move the function def `_setup_freqs(self): from blimpy/blimpy/io /base_reader.py to blimpy/blimpy/` to `waterfall.py.` Furthermore, the plot event pipeline doesn't work on the main branch; the best versions of `plot_event.py and`

`plot_event_pipeline.py` are available on sofias repository (past BL intern). I created a directory in the same directory that I installed TurboSETI, for sofias repositories. Furthermore, there is a useful script to plot hits in the parallel branch on github, which can come in handy.

3 Operation

The software TurboSETI uses an tree search algorithm to search for narrowband signals with a Doppler drift, a feature expected to be observed from a non-zero acceleration source relative to the receivers. The amount of drift a signal experience is important as narrow channels/greater drift means, less power in each individual channel. This causes a smearing effect, that can be corrected for, in order to maximize the SN. As the drift data is unknown prior we must test a range of drift rates, and hence set a suitable maximum. This process can be computationally long, therefore it is very important to define a suitable range, I investigated this when testing TurboSETI on data from the open BL database, this is discussed in the next section.

Normally the target is observed of approximately 5 minutes, which is considered an "ON" observation. Then another target is observed for another 5 minutes, this is considered the "OFF" observation. These steps are repeated until 3 "ON" and 3 "OFF" observation have been recorded. The purpose of this methodology is to identify RFI, as RFI is terrestrial it should appear in both "ON" and "OFF" observations, hence if we recorded a narrowband signal in both the "ON" and "OFF" observation, the natural assumption is human technology. There are three main steps to using Turboseti; the first is a Doppler search; the results are returned as hits in (.dat) files. The Doppler search's main parameters are the max drift rate and SN cutoff, which define the max drift rate to be searched and the cutoff for the SN. Next, the hits are filtered into events where several parameters are used to filtered out unwanted hits.

These parameters are the SN cutoff, filter threshold, and a parameter for checking zero drift. The SN cutoff can be the same value as used in the Doppler search, or another value if desired. The filter threshold can have three value or "levels" 1,2, and 3. Filter threshold 1 filters out hits below the SN cutoff, taking into account the check zero drift parameter but doesn't take into account if the signal is observed in both the on and off observations. Filter threshold 2 accepts hits that passed level 1 criteria. Furthermore, the signal has to be present in at least one on observation, and not present in any of the off observations. Filter threshold 3 accepts events that passed level 2 criteria. The signal also must be present in all on observations.

The last parameter, "check zero drift" can have the value either true or false. This parameter lets the user include or excluded events with zero drift. The events are returned in the form of a table which is saved to a .csv file. Using the plot event pipeline, the events can be plotted, returning .png files. To test that everything was installed correctly, I used data provided in a tutorial created by Shane Smith, another BL intern, using the suggested parameters, below I've included one of my events plotted.

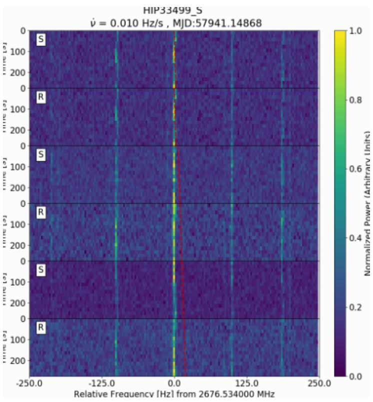

Figure 1: Plotted Event(HIP33499)

In the above image, the six cadences are stacked one on top of each other, the signal is the yellow line, and the red line is model, and has a different drift then the actual signal, due to the parameters that we've defined. RFI when present will appear as a vertical line.

4 ALPHACEN

After setting up TurboSETI on the blc00, I tested the software functionality using (.h5) and (.fil) files from several sources. One of the most exciting pointings was in the direction of Alpha Centauri, the data was taken by the Parkes telescope and takes the form of 6(.h5) files, three are "ON" and 3 are "OFF" files. Alpha Centauri is the closest star and planetary system, to the Earth's solar system, around 1.34 parsecs from our Sun. The system is a triple star system, with one confirmed planets, possibly in the habitable zone. The files are quite large around 14Gib per file, downloading and searching each file took several hours. All files are from the BL open database.

My procedure for searching for hits, and then filtering the hits into events follows the same procedure described in the previous section. For the Doppler search, I defined the S/N cutoff as 25, as described in Enriquez et al. (2017). I later tried to use a value of 10 as suggested in Price et all (2019), as an improvement on Enriquez et al. (2017). A sensible maximum drift rate can be reached examining physically allowable accelerations, which have been determined by exoplanetary discoveries. The four main components in determining drift rate are the rotation of Earth, the orbital motion of Earth, the rotation of the host body, and the host body's orbital motion. The paper Sheikh et al. (2019) suggests a maximum drift rate of 4Hz based on the methodology presented above, hence my maximum drift rate with be 4 Hz. It is also important to note that the drift rate can be approximated as a linear function, over short periods.

After I have generated my (.dat) files, I search for events using the same SN cutoff as the Doppler search. I first tried the filter threshold 0f 2 and got 837 events, I then tried a filter threshold of 3, AND got no events. As my search for filter threshold 2 revealed many events, I have chosen four events to plot, which are included below.

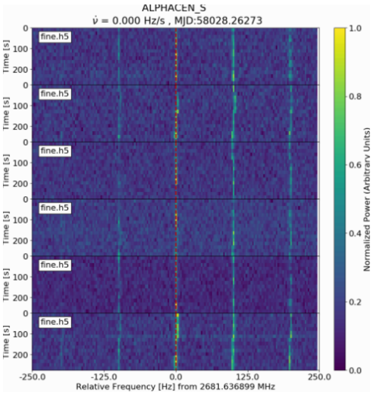

Figure 2: Plotted Event (ALPHACEN)

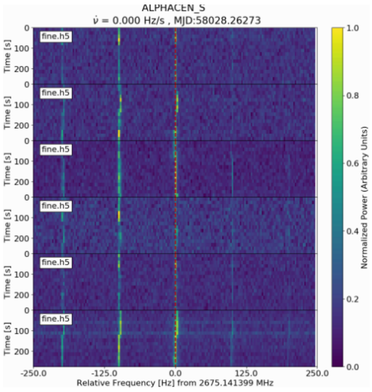

Figure 3: Plotted Event (ALPHACEN)

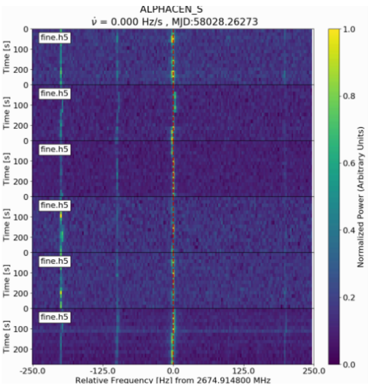

Figure 4: Plotted Event (ALPHACEN)

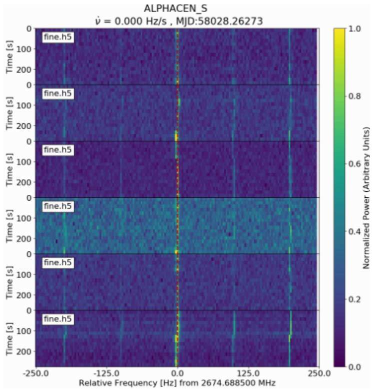

Figure 5: Plotted Event (ALPHACEN)

I then changed the SN cutoff to a value of 10, for both the Doppler search and event filtering, I used a filter threshold of 2, and got 306 events. I also tried a filter threshold of 3, but got no events. Below I have plotted 4 events, for filter threshold 2.

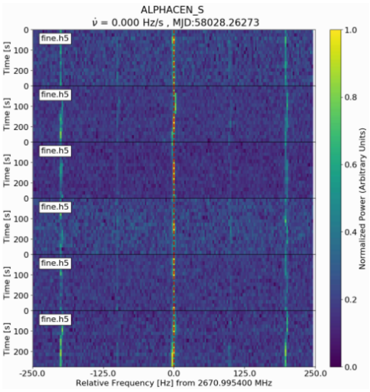

Figure 6: Plotted Event (ALPHACEN)

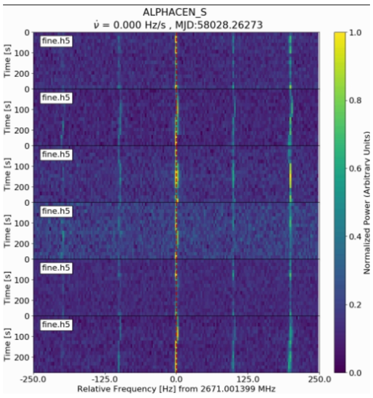

Figure 7: Plotted Event (ALPHACEN)

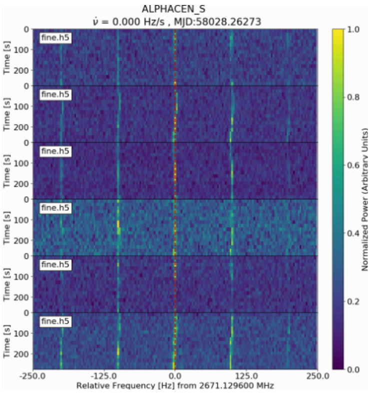

Figure 8: Plotted Event (ALPHACEN)

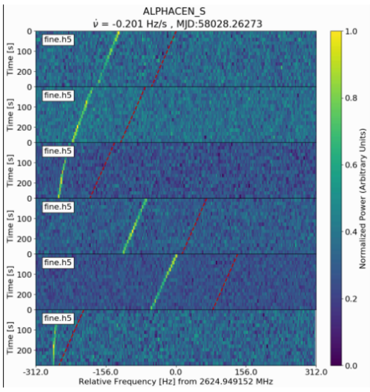

Figure 9: Plotted Event (ALPHACEN)

In all the event plots for Alpha Centauri, the signal is present in both the "OFF" an "ON" observations, which suggests ground based RFI. The signal appears to be around 2600 MHz, with zero drift, the only exceptional is figure 9, where the signal appears to drift, the model based on the parameters that I previously defined doesn't fit the model well. This apparent drifting signal, is likely caused by my parameters and not a signal not explained by RFI. The signal appears in both "ON" and"OFF" observations, this suggests ground based RFI again. The source of the RFI can be identified by looking at RFI chart, for Parkes. Looking at the RFI charts available on Parkes website, there is a well know signal around 2500MHz, produced by Globalstar. Globalstar is a satellite communications company based in America, that operates a low Earth orbit satellite constellation for satellite phone and low-speed data communications.

5 Conclusion

The Alpha Centauri system is a promising candidate for SETI searches, as the exoplanet Proxima b, is about 1.3 times more massive than Earth, which suggests that the exoplanet is a rocky world. Furthermore, Proxima b is also in the habitable zone, so there is a possibility that liquid water could exist. Alpha Centauri A is very similar to our own Sun, Alpha Centauri B is slightly smaller and cooler. Both stars are about 10 percent older than our solar system, which in theory means that a possible civilization would have time to develop. There is also the possibility of other Earth sized planets, but scientists debate this. The unique structure of the system has made the search for exoplanets a complicated one. Unfortunately I was not able to identify a signal not explained by RFI, hopefully future discoverers of exoplanets will prompt further SETI searches of Alpha Centauri.

One of the biggest challenges facing SETI scientists is RFI. To conclusively prove that a signal is not terrestrial, we need to understand all sources of RFI observed in the event waterfall plots. Developments have been made in this regard; a permanent RFI monitoring tower has been developed at Parkes. The tower is located approximately 1km away from the observatory to monitoring terrestrial RFI. The monitoring station utilized the existing structure, which was previously the Kennedy Radiotelescope. The tower operates by monitoring RFI using antennas, low noise amplifies, and a spectrum analyzer. The tower has been successful in monitoring medium and strong RFI. Currently, the tower can not be used for instantaneous viewing of the RF spectrum by the observers. In the near future, there are plans to make a network connection between the RFI monitoring station and the main observatory building, allowing real-time updates of the RFI environment.

In this project my implementation of TurboSETI, was successful, I was also able to search existing data from the BL open data base, and identify sources of RFI. There is still further development need for TurboSETI, as I encountered many errors, while installing and using TurboSETI. In the next task I hope to carry out the first SETI search for I-LOFAR using TESS data, hopefully identifying a signal that can not be explained by RFI.