Processing TESS Targets using I-LOFAR SETI pipeline

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

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. In preparation, I undertook several data processing exercises to develop the skill set needed to accomplish this goal. This report details the fifth project. The Bl backend system at I-LOFAR has two nodes, the head node (blh0) and the processing node (blc00). After installing and testing the newly installed SETI pipeline using data taken for the BL open data base. Data products similar to the high spectral resolution files used by BL can be created for I-LOFAR and searched using TurboSETI.

The targets that were observed to carry out the first SETI search are TESS targets. I-LOFAR observed six TESS targets for between 20 and 40 minutes each. The raw data needs to be reordered and combined into one file, which can then be channelized to form a filterbank file suitable for a SETI search. Then TurboSETI can be used to search the data for events that could have an extraterrestrial origin. This search's parameters must be defined to maximize the search and exclude RFI hits, but include possible SETI searches.

This report is divided into five sections. The first section describes why the targets were chosen. Section two describes the process of channelization to form a filterbank file. The third describes the parameters that were defined and the results of the SETI search. The fourth section defines the difficulties encounted when adapting TurboSETI for use on the I-LOFAR system. The final section describes the next steps needed.

2 TESS Targets

To conducted the first SETI search at I-LOFAR, six TESS targets were chosen and observed. In 2018, TESS and BL, announced a collaboration, where approximately 1,000 TESS targets will be searched. Most of these new targets are Earth sized rocky planets inside the habitable zone, which makes them optimal SETI targets. Furthermore, Systems observed by TESS are viewed edge on from Earth, this is important as the radio signal leakage from the Earth mostly occurs at the plane of the Earth's orbit. The theory is that if an extraterrestrial civilization has radio technology, radio signal leakage from the planet's plane, is expected. Our specific six TESS targets were chosen because they were the closest targets on the day of observation.

3 Channelization

Before searching for drifting signals, the raw data needs to channelized. A suitable frequency resolution must be determined, to detect narrow band signals. SETI uses three filterbank files, the high spectral resolution file used for SETI searches, the high time resolution file used for pulsars, and the medium resolution file which is used for spectral lines. The BL high spectral resolution file has a channel width of 3Hz and a time resolution of 18s. These files are used as they have an optimal resolution for detecting narrow band signals. Secondly, these files are smaller than the original files, making them quicker to search. Our high spectral resolution file for I-LOFAR has a channel width of 3Hz and a time resolution of .3s. These are filterbank files. Interesting the data isn't corrected for the effects of dispersion by the ISM on the signal. Firstly the time resolution of the high spectral resolution file is greatly reduced making incoherent dedispersion pointless. Secondly our targets are so close that the dispersion experienced by the signal is minimal.

The data is gathered over four ports at I-LOFAR. Therefore the raw data for each observation is contained in four files. We can use an extractor to reorder and combine the four files into one dada file, a header in the appropriate format must also be created. Then the filterbank file can be created by performing a fast Fourier Transform for a channelization value. I-LOFAR channels normally have a width of .195MHz, as a channel width of 3Hz is required, a channelization of 2 to the power of 16 is used. I-LOFAR has 488 channels, which means that we need channelization value of 31981568x. Now that the data has be channelized, we can search for drifting signals, in the data.

4 Using TurboSETI

The general procedure for searching for extraterrestrial signals using TurboSETI has three main steps. The first is a Doppler drift search, which attempts to identify drifting signals in the data. The Doppler drift search's parameters are the max drift rate to be searched and the S/N cutoff. The noise floor for I-LOFAR for our purposes id estimated to be 5.8 sigma, but a higher threshold of 6 is used. A higher threshold is used, as the noise is not perfectly Gaussian in shape, due to RFI, and possible SETI hits.

A sensible maximum drift rate can be reached by examining physically allowable accelerations, which have been determined by exoplanetary discoveries. In Price et al (2019) they conducted a search with a maximum drift rate of 4Hz, for an observing frequency of 2.275GHz. Therefore, as the High band array's observing frequency is 13 times lower, the maximum drift rate was set to 2hz for this search. After the Doppler drift search, the results are returned as hits. The hits can be filtered into events where several parameters are used to filtered out unwanted hits. The parameters that are applicable to us are are the SN cutoff and filter threshold.

The SN cutoff can be the same value as used in the Doppler drift search, or another value if desired. The filter threshold can have three value or "levels" 1,2, and 3. The first level returns hits above the SN cut, taking into account the check zero drift parameter, but without an on-source/off-source check. The second level returns hits that passed level 1 and that are in at least one on-source but no off-sources. The third returns events that passed level 2 and that are present in all on-sources. The last parameter checks zero drift can have the value either True or False. Signals with a drift rate of zero indicate no relative acceleration between the receiver and transmitter. Therefore this signal is likely RFI. Hence we didn't include signals with zero drift rate.

After the hits are filtered into events, they are returned in the form of a table. These plots must then be examined manually. The targets observed by I-LOFAR weren't observed in cadences. The Doppler drift search pipeline can be used, but the find event pipeline can not be used exactly as described above..

5 Problems Encountered

TurboSETI is still under development, hence there are still bugs, some of which we fixed. Furthermore, TurboSETI was written for use on the Parkes and GBT backend, this meant that some system parameters exclusive to these two telescopes were hard-wired into the code. For instance, the number of course channels needs to be adjusted to 488 channels for I-LOFAR. Furthermore, TurboSETI does not accept filterbank files that are less than 8 bit.

The most important error that we discovered was a root 2 error, this means that currently TurboSETI only finds local maxima. Basically they only search for 1-channel peaks in the (drift-corrected) spectra. It does not search correctly for anything with wider spectral occupancy. Hence the sensitivity of the algorithm is sqrt(2) worse than it should be. Solving this problem would increased search volume by 70 present. This is demonstrated in the diagram below.

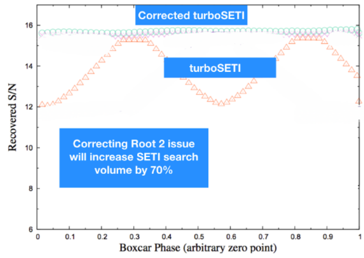

Figure 1: Root 2 Error

6 Results

As the data isn't in cadences, the data can be searched for drifting signals as described above. But the find event pipeline can't used. The hits can be plotted using a function in TurboSETI's parallel branch on github. As we don't have on and off targets the only way RFI can be excluded is to discard any plots were the drift rate is zero. A frequency allocation chart can also be used to identify sources of RFI. After the Doppler drift search we have hundreds of hits, this is filtered down by excluding signals with zero drift rate. We can then exclude hits that are below our signal cutoff, this leaves possible SETI hits. This list could be filtered more if filter thresholds could be used. In the diagram below a histogram of hits from the first 5 mintues of one of the TESS targets is shown. There are still many possible hits above the SN cutoff that could be SETI sources (identified in blue). Most of these hits are likely to have a terrestrial source.

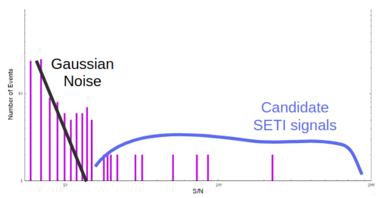

Figure 2: Histogram of Hits

Unfortunately the plot hits function in the parallel branch of TurboSETI on github does not currently appear to plot hits, not in the standard SETI format of six cadences, stacked on top of each other. Important I-LOFAR has a working SETI pipeline, that should be able to filter hits down to events fully. The root two error that was identified during the processing of the TESS data, when solved will significantly improve the current TurboSETI software.