Data Analysis of a Fast Radio Burst

Summer Internship with I-LOFAR and the BreakThrough Listen Initiative

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

Contents

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 first exercise that I completed. By completing this exercise, my understanding of radio astronomy is deeply enhanced. I have also acquired critical data processing skills, a greater knowledge of FRB's, and techniques for process data containing FRB's signals.

The data used for the first task was gathered using the Parkes telescope, located in New South Wales, Australia. The telescope pointed slightly off the Small Magellanic Cloud during this particular observation. The data contains a very famous FRB called the Lorimer burst. The Lorimer burst was first discovered in 2007 by Duncan Lorimer while processing data from 2001. In this report, I aim to extract this signal and confirm that the signal is indeed an FRB, with an origin outside of the Milky Way.

General techniques for processing pulsar data are applied; first, the process of incoherent dedispersion, followed by a pulse search. A signal not explained by RFI is observed when plotting time, DM, and SN on a colour plot. The signal, which is a FRB, is observed in the files 6,7 and D exclusively. Further analysis of the original data, shows the shape of the fast radio burst to be quadratic. Important parameters of the fast radio burst are then calculated, including SN, DM, and the signal duration. A summary of the results and discussion is at the end of this report.

2 Dedispersion

The data used for the first task was gathered using the Parkes telescope, located in New South Wales, Australia. The telescope pointed slightly off the Small Magellanic Cloud during this particular observation. The data takes the form of thirteen filterbank files, where each file is one beam. There are 96 frequency channels for each file that are 3MHz wide, with a sample time of 1000us. The file is a high time resolution file, and has been optimised for pulsar searches. Each file covers the same frequency band and time span but a different beam direction. The beam layout and the beam projection are both shown in the diagrams below.

Digitization is the process of converting an analog signal to a

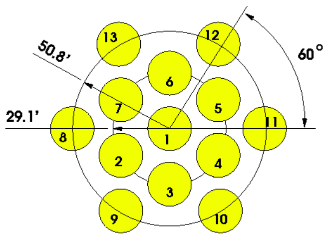

Figure 1: Beam Layout

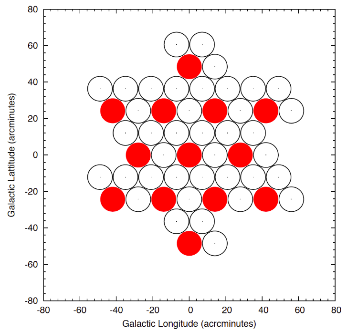

Figure 2: Beam Projection

digital signal. The two main characteristics are the sample rate and the number of bits per sample. The data used in this exercise has a rate of 1 bit per sample, comparatively the standard for I-LOFAR is 8 bit. There are some advantages to using 1 bit over 8 bit, for example, RFI is practically filtered out of the signal as the worst value RFI can give is a 1.

The data used is not raw; it has undergone previous processing to produce filterbank files. The filterbank files are created by a polyphase filter spectrometer, in which the signal is divided into several frequency channels and recorded at short time intervals after detection. As signals propagate through a medium, they experience dispersion. This dispersion is caused by several mediums, including the earth's ionosphere, interplanetary medium, galactic interstellar medium, intergalactic medium, host galaxy interstellar medium and the local FRB environment. The DM is defined as the integrated column density of free electrons (ne) along the line of sight,

$$DM = \int_a^0 nedl$$

This gives a frequency-dependent (in MHz) delay (in seconds) to be written as

$$\triangle t = \frac{DM}{2.41 * 10^{-4}}(\upsilon^2)$$

It is observed that higher frequency pulses arrive earlier than pulses of a lower frequency, where the delay in pulse arrival is inversely proportional to the frequency. The difference in arrival time for low frequencies (fl) and high frequencies (fh) is given by the equation,

$$\triangle t = 4.148803 * 10^{-3} \left[\left(\frac{fl}{GHz} \right)^{-2} - \left(\frac{fh}{GHz} \right)^{-2} \right] \left(\frac{DM}{cm^{-3}pc} \right)$$

Hence, there is a dependence on the group velocity as a radio wave propagates through an ionized medium. The effect of dispersion on a signal is shown in the graph below, where the higher frequencies arrive before the lower frequencies, giving a quadratic shape to the signal. If the signal experienced no dispersion, a straight line would be observed in the graph below instead.

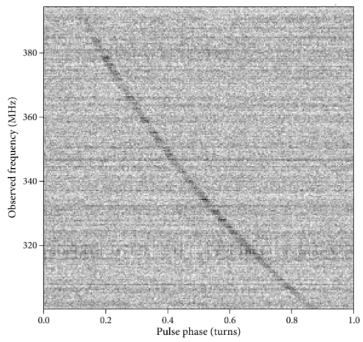

Figure 3: Dispersion of a signal

This effect can be corrected for, by applying the appropriate time delay to each frequency channel, this method is known as incoherent dedispersion. The process is an example of post-detection dedispersion, where signals have intensity-like quantities due to a squaring operation that has occured after the signal has been channelized in a spectrometer. After the frequency channels have been appropriately delayed, they are added together to form a dedispersed time series.

Incoherent dedispersion is limited by the width of the individual frequency channels that retain a certain amount of time delay. The strong dependence on observing frequency means that wider channels can be used at higher frequencies. Therefore the time step could theoretically get larger for the values for dispersion measurement increase. Since the source's dispersion measurement was unknown before processing, a large number of dispersion measurement trials must be searched. The trial values shouldn't be so fine that they slow down the program and not so large that they miss possible signals.

Another important consideration is the step size; a standard solution is to set the step size between the lowest and highest frequency channel equal to the data sampling interval. This step size becomes critical when observing a pulsar with a period of a few hundred ms. Because there are significant model uncertainties due to unmodeled electron density structures in the ISM, it is wise to search for DM values higher than predicted by the mode. Importantly, there can be residual dedispersion smearing if the exact value of dispersion measure DM is not used.

The signal, is also effected by other propagation effects, for example, scintillation, scattering and Faraday rotation. For this report only dispersion by the ISM is corrected for using the process of incoherent dedispersion. The software sigproc contains the command dedisperse, which preforms incoherent dedispersion on a filterbank file as described above. This command outputs a time series. For this data set, a maximum dispersion measurement of five hundred was used, a step size of one is suitable in this case. After dedispersion there are 500 files for each of the thirteen files. Interestingly without this dispersive delay it would be even more difficult to distinguish astrophysical signals from ground based RFI.

3 Seek

After correcting for the effects of propagation through the ISM, the data can be searched for pulses. The command seek can be used to conduct a periodicity search, the flag -pulse, can be used for a single pulse search. FRB's are usually individual pulses that do not display periodicity; therefore, a single pulse search will be more informative.

The sigproc single pulse search carries out a matched filtered search to detect single pulses. Matched filtering is an optimal method for detecting a signal of known shape in the presence of additive noise. The time series can be searched by first convoluting the time series with boxcar functions. For pulse duration greater than a single time sample, the SN must is normalized by the boxcar function. Any peaks in the time series are now considered possible pulses.

After the time series files are searched, the time series files, along with any other irrelevant files, are deleted in order to save space. Then the files are separated so that all the .pls files for each separate beam are together. The header is removed so that all the .pls files can be combined. This gives thirteen large files, one for each beam. The noise floor for I-LOFAR is 7 sigma, but a higher threshold of 8 was used as the noise is not perfectly Gaussian in shape, due to

RFI, and other interesting signals. The SN can be calculated by the following equation, where k is the Boltzmann's constant, Ae is the effective aperture area, Sv is the average pulsar power spectrum density, and Tsys is the temperature of the system,

$$SNR = \frac{1}{2k}\,(Ae)\,(Sv)\,(Tsys)^{-1}$$

Now that the SN has be determined any column with a SN value less then or equal to six is removed. The files have been properly processed, and can be plotted by extracting the relevant columns. The single pulse search file has five columns, the first is DM, the third is time, and the fourth SN.

4 Plotting

Plotting DM, SN, and time on a colour map, reveals constant lines that are observed in all plots. Since the lines are present in all the plots, its is assumed that they are RFI. In the graphs for files 6, 7, and D a signal is observed in the same location, that is absent in all the other graphs. Below all thirteen graphs are displayed.

Zooming in on the unidentified signals, it is apparent that the unexplained signals are single pulse. Examining the signal further, the DM and signal duration can be determined. First, using the awk command ignore the header line and ignores things after 8000 seconds as they are the brightest objects. The output is organized numerically, and the top ten results are printed. Examining this data,the DM values is between 360 - 380. To obtain a more precise value, plot DM vs SN, the DM with the highest SN, will be the closest to the correct value. This method is not very precise, but visually inspecting the graph and comparing with the data gives the following results in figure 4.

Beam	DM	S/N
6	375	30
7	364	26
D	372	28

Figure 4: Chart of results

A more accurate method is to fit the analytical form, which is described in Cordes and McLaughlin paper. The same process can be repeated for S/N versus width. Viewing the second column, gives a value of 4 with a SN = 30. This column means the width was 2(4-1) = 8 samples, so 8ms. This value is not exact as the trials are 1, 2, 4, 8, 16, 32, etc, the actual answer is in between. My final results for the fast radio bursts are SN = 30, DM = 375 and duration = 8ms.

To confirm that this signal is a fast radio burst, the original data needs to be examined. A signal with a good SN at a non-zero DM is not solid evidence of a signal, as non-astrophysical reasons can not be fully ruled out. For instance, if there are two narrow band bursts of interference at different frequencies at some DM trial value, they will line up, and a higher S/N is observed at this DM. In order to confirm that the signal is a fast radio burst the shape of the signal needs to be viewed, hopefully displaying the typical quadratic sweep.

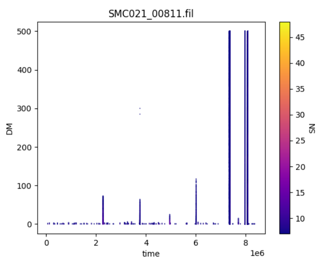

Figure 5: Beam One

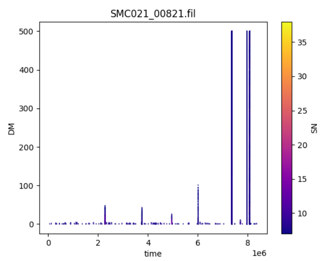

Figure 6: Beam Two

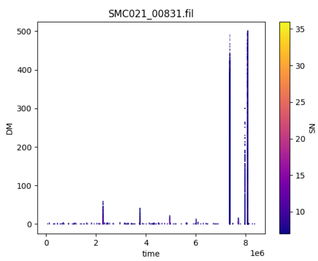

Figure 7: Beam Three

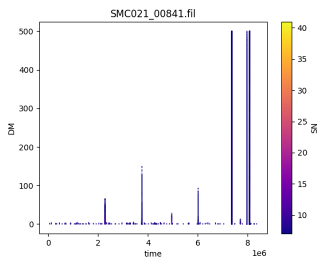

Figure 8: Beam Four

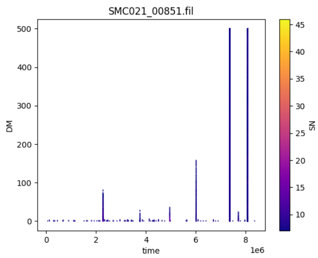

Figure 9: Beam Five

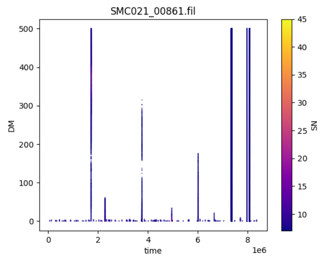

Figure 10: Beam Six

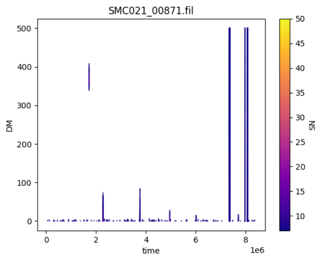

Figure 11: Beam Seven

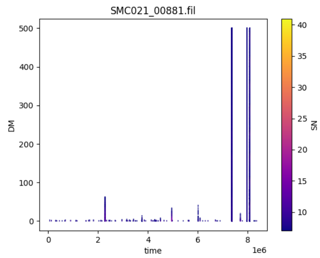

Figure 12: Beam Eight

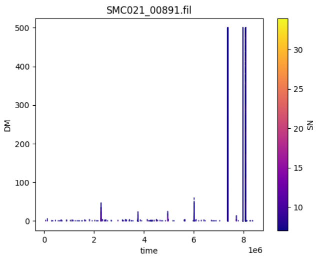

Figure 13: Beam Nine

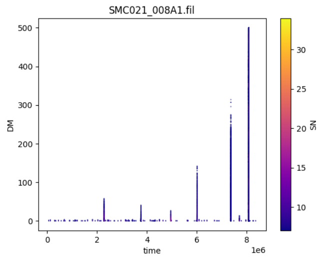

Figure 14: Beam A

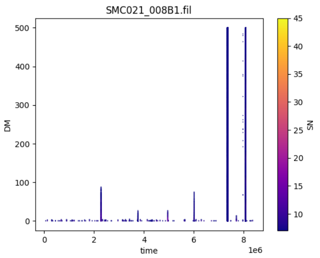

Figure 15: Beam B

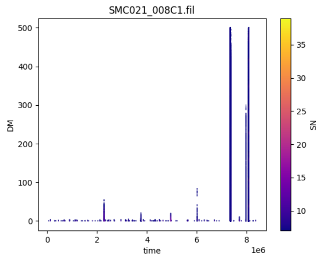

Figure 16: Beam C

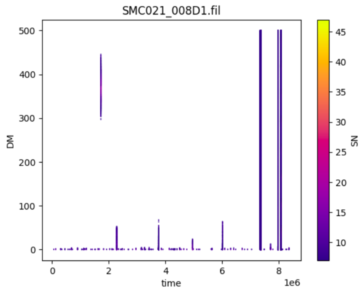

Figure 17: Beam D

5 Quadratic Sweep

Confirming that a signal is an FRB can be difficult in practice as the research community currently has no strict and standard formalism for defining an FRB. Although there is a loose set of criteria that most researchers agree on. These criteria include the pulse duration, brightness, and in particular, whether the DM is larger than expected for a Galactic source. In order to confirm that the signal is an FRB, the shape of the signal needs to be examined, as an FRB should have a characteristic quadratic sweep. For this step, I started with the original data.

Viewing the original data, it is obvious that the resolution is too low, as its just ones and zeros. We know that the signal is eight samples in duration; therefore, the data can reduce the resolution in time by 8x to gain a bit more dynamic range. This can be accomplished by using the sigproc command decimate to add every eight samples together, then chop out a 10-second chunk around the time of interest and save the output in ascii form in a file. Plotting in gnuplot, we can see the quadratic sweep in each of the three files 6,7 and D, as shown in the figure below.

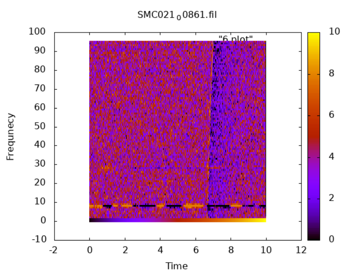

Figure 18: Beam 6

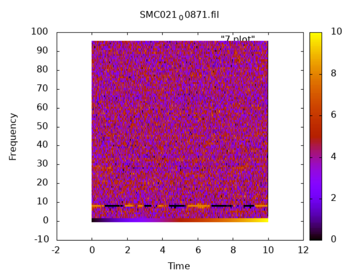

Figure 19: Beam 7

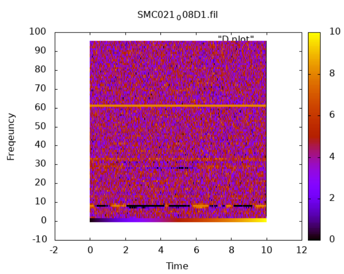

Figure 20: Beam D

Zooming in on the FRB, and plot the high frequencies on the top and the low frequencies on the bottom, a the classical FRB shape is clearly visible, as seen in the figure below. This signal was so bright that it saturated the primary detection beam of the receiver, causing a dip below the nominal baseline of the noise right after the pulse occurred.

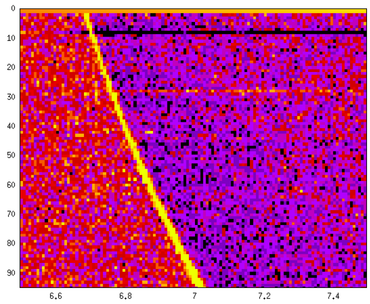

Figure 21: Quadratic Sweep

6 Results

FRBs have been a source of fascination and intense research for the scientific community, since their discovery in 2001 by Lorimer. Many aspects of FRBs are still largely unknown due partly to small all sky surveys. Firstly FRBs have large DM, which indicates an originate from outside the Milky Way. Secondly, most FRBs appear to be a single pulse, with just two known exceptions. Thirdly, due to their origin from outside the Milky Way, FRBs are very luminous. Furthermore, the cause of FRBs and their interworkings are still widely debated in the community today.

Currently, the FRB population is relatively small, consisting of more than 60 independent sources detected at 10 telescopes and arrays around the world. The observed population spans a large range in DM, pulse duration, and peak flux density, interestingly FRB have only be detected by radio telescopes thus far.

The burst that I have detected in the data is very famous and is known as the 'Lorimer burst', and is remarkable not only for its incredible brightness but also for its implied distance. In the original literature, the pulse's large dispersive delay was estimated to be roughly eight times greater than could be produced by the free electrons in the Milky Way. I have approximated the signal to have a DM of 375, SN of 30, and a duration of 8ms. The quadratic nature of the FRB has also been investigated by looking at the original data.

The brightness of the source can be calculated using the radiometer equation, where beta is a digitization loss factor, G is the gain of the telescope in Jy/K, Tsys is the system temperature, n pol is the number of polarisation's, BW is the bandwidth, and T obs is the observing time. For Parkes, beam 6 has a gain of 1.5 Jy/K. The bandwidth is 288MHz, and there are two polarization's. The duration of the pulse is 8ms. As the data is 1-bit digitized, only 80 percent of the information is obtained. Hence beta = 1/.8 = 1.25. The system temperature is 23K. Subbing these values into the equation below, the peak flux density in janskies is .6Jy for the FRB.

$$\sigma_\nu = \frac{\beta * G * T_{sys}}{\sqrt{n_{pol} * BW * T_{obs}}}$$

The chart below compares the flux density of the Lorimer burst to other bright sources in the universe.

Source	
Quasar(observed 5GHz)	.3
Lorimer Burst	.6
Milky Way (observed at 10 GHz)	10000
Sun (observed at 10 GHz)	4000000
Disturded Sun (observed at 20 GHz)	20000000

Figure 22: Bright Objects in the Universe (Jy)