Data analysis for first pulsar data from I-LOFAR

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 second 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 pulsars, and have managed to process data containing pulsar signals.

The data used was gathered using the I-LOFAR array and contains the first ever pulsar observed by I-LOFAR. Since pulsars discovery in 1967 by Jocelyn Bell, scientists have discovered over 2,000 pulsars. Pulsars show an amazing range of characteristics, many with scientific importance, for example, PSR B1913+16, which has lead to a new understanding of gravitational waves. Pulsar surveys have been increasing in recent years due to newly built large radio telescope, with low-noise receivers, which is critical for observing pulsars, as they are weak radio sources.

My procedure follows the same methodology as described in my first report concerning FRB's. First, incoherent dedispersion, followed by a single pulse search and a periodicity search. Plotting both search results, pulse signals are observed with a periodic nature. After plotting the single pulse search and periodicity search, important characteristics of the signal are recorded, including the SN, DM, and pulse period. Hence, we can identify the pulsar by name using a catalog called psrcat. Following this, the power spectrum of the signal is examined after the true period is confirmed. The data can then be folded upon the epoch. Final results and discussion then follows at the end of this report.

2 Dedispersion

The data used has a frequency channel one of 166 MHz with a channel bandwidth of .195312 MHz. The observation has a sample time of 81.92 us and an observation length of 30.9 seconds. This information was obtained using the sigproc command "header" which returns a filterbank file's header. The data is digitization with a rate of 8 bits per sample, which is standard for I-LOFAR. In order to digitize a signal, the sampling rate must be at least twice the highest frequency, therefore we can't have any sample rate that we want, this is known as the Nyquist Theorem. In other words, the criterion is two samples per cycle of the maximum frequency difference, in order to reproduce the signal accurately.

The data is corrected for the effect of dispersion by the ISM, by incoherent dedispersion, with the same methodology as described in my first report. The command dedisperse from sigproc applies incoherent dedispersion by adding appropriate time delays to the frequency channels. After the frequency channels have been appropriately delayed, they are added together to form a dedispersed time series. As outlined in my first data analysis report, both the DM trial range and step size are important considerations for each source. For this data set, a maximum dispersion measurement of thirty is appropriate, and a step size of .002 is suitable in this case. After dedispersion is completed, a time series file is outputted.

3 Seek

After correcting for the effects of propagation through the intergalactic medium, the data is searched for pulses. Pulsars are weak radio sources; therefore, single pulses are observable from only strong radio sources. Pulsars exhibit strong periodicity; hence, a periodicity search is normally more effective. The command seek can be used to conduct a periodicity search, the flag -pulse, can be used for a single pulse search. The seek command reads in the time series file and performs a Fourier transform and harmonic summing, outputting a list of periodicity's and S/N, where each pair of columns is a different harmonic fold.

As pulsars have small duty cycles, their spectral power gets distributed into the harmonics. Fortunately, this power can be recovered by summing the harmonics. This is done by expands the amplitude spectrum by a factor of 2 and adding it to the original amplitude. This is repeated to add more harmonics. However, since the noise also increased as more harmonics are added, the S/N will reach a maximum at a harmonic fold related to the pulse width. If the pulse duration is W and pulse period is P, we can expect (P/W) harmonics. From the single pulse search, we have a width of .038, and a period of .25. Hence we have approximately 7 harmonics. The flag -s can be added to the seek command, to output a spectrum for each harmonic, which can be viewed using spec. Using the spec command, I viewed the outputted spectrum after the folds. After each summing we can see the fit improving, until a peak has formed at the fundamental frequency, we can also see harmonic present.

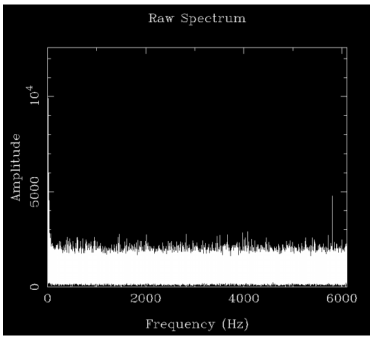

Figure 1: Raw Spectrum

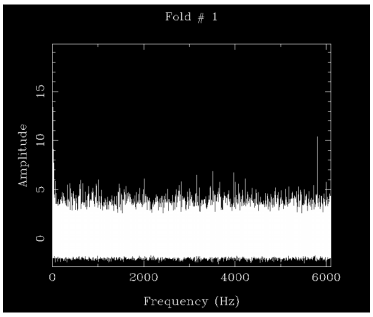

Figure 2: First Fold

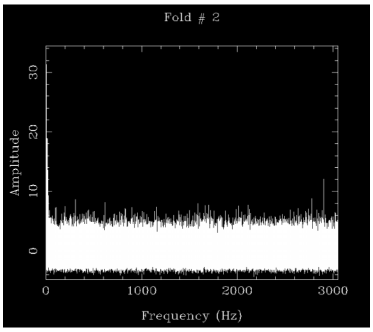

Figure 3: Second Fold

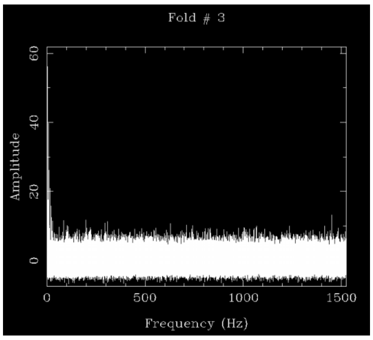

Figure 4: Fourth Fold

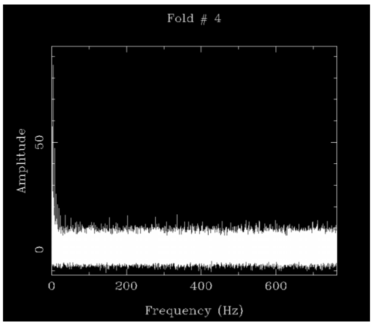

Figure 5: Fifth Fold

Any pulsar signals will appear as the strongest signal in the Fourier spectrum, where the pulsar's fundamental frequency is the strongest signal. After the command seek, two different files are produced, the single pulse search file .pls and the periodicity search files .prd. All irrelevant files can now be deleted, leaving the .pls and .prd files for each DM trial. The header is removed from each file, and all the .pls files are combined, then all the .prd files are combined, this leaves two files, one .pls and another .prd. The best command can be used to reduce the .prd file down to a more manageable list of candidates. The list of candidates will contain repeats of signals detected in several DM. The best command takes a (.prd) and optimizes the results by combining signals with the same period's same harmonics. The signal to noise level for this data is approximately 5, any columns that contain an SN value less than 5 are removed.

After file processing, the data can be plotted by extracting the relevant columns. The single pulse search file has five columns, the first is dispersion measurements, the third is time, and the fourth signal to noise. The .pls colour plot in figure one has time on the x axis and DM on the y axis, with SN as the z value. Final the .prd file is plotted, with period on the x axis and DM on the y axis, with SN as the z value, as seen in figure two below.

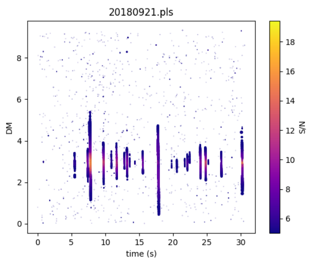

Figure 6: Single Pulse Search

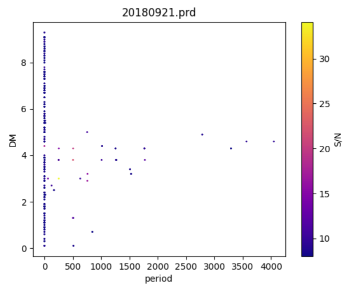

Figure 7: Periodicity Search

4 Identifying Pulsar

Examining the single pulse graph (figure one), a signal is clearly observable, this signal repeats at a set interval. From the first graph, the period of the pulsar can be estimated to be around .25 seconds. To confirm this, we need to examine the second graph, the periodicity graph (.prd). This graph gives a much more accurate period as pulsars are weak radio sources, but have a highly periodic nature. Looking at the second graph, there are lots of points around 20 - 30 ms, which corresponds to 50 Hz. A very strong periodicity for .25 seconds is also present, which corresponds to the signal in the first plot. To find the DM value, view the top results for SN, which gives a DM of 3. As the period of the pulsar has been determined, the pulsar can be identified by name using a pulsar catalog, like psrcat. Searching psrcat for a pulsar with a period of .25 seconds, there is only one result, the pulsar B0950+08. Furthermore, the period can be confirmed using a pulsar timing code.

5 Power Spectrum

The power spectrum of a time series describes the distribution of power into frequency components composing that signal. According to Fourier analysis, any physical signal can be decomposed into several discrete frequencies or a spectrum of frequencies over a continuous range. The power spectrum tells us about the expected or average power of a signal at each frequency in the spectrum. The most common approach to pulsar signal analysis is to divide the digitally sampled base band voltage into several frequency channels using a fast Fourier transform. The signal power is then detected in each channel, and integrated over a short period to produce a power spectrum estimate. This results in well-known trade offs and limitations on time and frequency resolution due to the competing effects of dispersive smearing and inverse channel bandwidth (Lorimer and Kramer 2004). The use of coherent dedispersion can significantly improve this trade off. This allows wider channels and hence higher time resolution to be achieved and is commonly used for high-precision pulsar timing. Traditional spectral estimation procedures are based on an implicit assumption that the input signal is stationary, or at least approximately so over the spectrum integration time-scale. In the figure below we see the power spectrum, created by sigproc as described above.

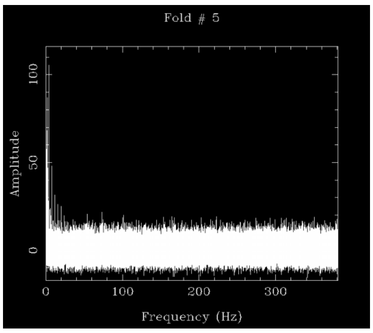

Figure 8: Power Spectrum

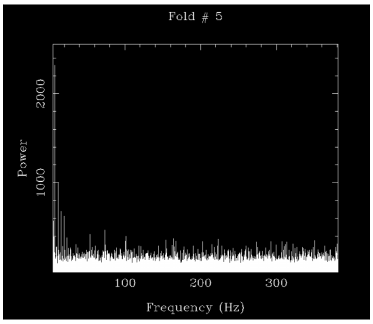

Figure 9: Zoomed in Power Spectrum

6 Folding

An integrated pulse profile is obtained by adding many individual pulses together; this gives a unique profile of the pulsar. The sigproc command fold can be used to fold the data for a known period upon the epoch. The folding algorithm used is a simple one: for each time sample, compute the phase based on a, possibly time-dependent, value of the pulse period and add that sample to the nearest phase bin of the appropriate profile. As we fold the date we can note that after folding the data out S/N value is worse then before folding. Plotting the folded profile, we get the graph below,

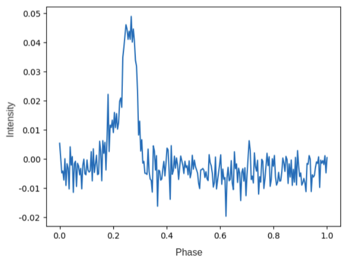

Figure 10: Pulse Profile

Now that we have the pulse profile we can determine the duty cycle, which is defined as the ”on” part of the profile which is the pulse. The pulse is approximately .15 in duration, which gives a duty cycle of 15 percent.

7 Summary

In summary we were able to process data gathered by I-LOFAR, containing the first pulsar detected by I-LOFAR using both single pulse search and and periodicity search. Plotting the results, and using the graph and data a pulsar with a period of .25 was identified. The spectrum after each harmonic summing was also viewed, and the power spectrum was also plotted. The pulsar was identified to be the pulsar B0950+08 using the psrcat catalog. After the period is confirmed the data can be folded upon the epoch, to give the pulse profile, a duty cycle of 15 percent is confirmed by the pulse profile.