Searching for Crab Giant Pulses with EPIC

21 October 2019

Cassandra Whitton (cassandra.whitton@gmail.com), with the guidance of Adam Beardsley (adam.p.beardsley@gmail.com)

Summary

One of the primary upsides of using EPIC as the data-writing pipeline for a radio array is that it allows for high time resolution observations, which in turn allows us to use EPIC to search for transient radio sources in space. My project was to create a data-processing pipeline to perform a proof-of-concept transient detection using EPIC and the LWA telescope. Particularly, we chose to try to detect Crab Giant Pulses (CGPs) from the Crab Pulsar, as these have been detected by LWA previously. Ultimately, I did not get far enough in this project to detect a CGP; however, I am putting together this memo to detail filtering methods that I have used in the pipeline, as well as to catalog glitches and other interesting things that I have found in the data I sifted through. I have also uploaded the code I used (which is messy but probably workable) to a shared EPIC google drive, under the folder Cassie. This will probably be reorganized later by Adam.

Crab Giant Pulses: What to Look For

Where is the Crab Pulsar? I used the coordinates 05h34m31.93830s, +22d00m52.1758s, in the ICRS frame. These were obtained from (http://simbad.u-strasbg.fr/simbad/sim-id?Ident=Crab+Pulsar).

The following information regarding CGPs is pulled from Ellingson et. al. *Observations of Crab Giant Pulses in 20-84MHz Using LWA1* (https://iopscience.iop.org/article/10.1088/0004-637X/768/2/136/pdf).

This paper observes that the characteristic broadening time (which is, albeit with a correction factor, roughly the exponential decay time of the pulse) is 400 ms (at 76 MHz) to 1000 ms (at 44 MHz). This decay time notably has a fairly strong frequency-dependence. The paper notes also that the Crab Pulsar has time-varying characteristic broadening time which are “attributable to ionized clouds or filaments within the nebula crossing the line of sight”.

Due to the slowing effects of free electrons between us and the Crab pulsar, we expect to find that higher frequencies of EM waves reach us slightly earlier than those at lower frequencies.

This effect is characterized by the dispersion measure (DM), which for crab is 56.791 pc/cm^3. The arrival time difference is given by

$$\Delta t = 4.15 \text{ ms} * (DM) * [(f1 / GHz)^{-2} – (f2 / GHz)^{-2}]$$

where f1 and f2 are the two frequencies being observed, and DM is given in units of pc/cm^3.

In the datasets I have run through, there have been at most 10 channels spaced by 25KHz each. At a center frequency of 75MHz, this would result in a time-difference of 280 ms. Since our characteristic broadening time could be as low or lower than 400 ms, dedispersion will probably be necessary to detect a CGP.

There is also a question of just how much bandwidth is needed to give us the required sensitivity to detect a CGP. Ellingson's paper suggests that, at 76MHz, you expect to see one CGP per hour with peak intensity 1000 Jy or larger. They use a (seemingly arbitrary) value of 4.4 sigma as their threshold for a significant detection. They also quote a beam SEFD of 13.7 kJy at 76 MHz.

This in mind, we can calculate the amount of bandwidth we need for a 4.4-sigma detection as

sqrt(Bandwidth) = 4.4 * SEFD / (Peak Intensity * sqrt(integration time))

Using SEFD=13.7 kJy, Peak Intensity = 1000 Jy, and integration time = 50 ms, this gives a bandwidth requirement of 73 kHz to see one CGP per hour, which is well within our capabilities.

However, that number does not align at all with Ellingson's results, which claim the sensitivity at 76 MHz with a bandwidth of 14.07 MHz is 656 Jy; they use a higher bandwidth yet find a much higher threshold for detection. The cause of this discrepancy is very unclear, though it may have something to do with what they describe as the "effective signal-to-noise" ratio. This still doesn't fully explain the discrepancy, as their effective signal-to-noise ratio is expected to be higher than the peak SNR, yet they find a higher effective noise. Basically, take this bandwidth-requirement estimate with a grain of salt, because there might be something we're missing.

Data Processing Methods

One fact which significantly limits the scope and difficulty of looking for CGPs is that we know exactly where a CGP will happen. In principle, given an isolated time vs. luminosity sequence looking directly at Crab's location, finding a CGP is as simple as looking for statistically significant changes in flux at that single point (providing we have successfully dedispersed and combined the separate frequency bins).

Unfortunately, the data contain radio-frequency interference (RFI). It is very common to see bright flashes in the field of view of the LWA, typically on the horizon (indicating they originate on the Earth, from human-made structures). Although they are often localized point sources, these flashes are so bright that their sidelobes illuminate the entire field-of-view, masking any real signal temporarily. This illumination triggers a false positive in the simple detection method of looking for statistically significant changes at Crab's location. So, we require a method to flag these false positives so that we can ignore them while we look for CGPs.

The aim of my attempt at filtering was to detect times at which RFI was occuring, and flag these times as false positives. This came with the additional constraint that it must run relatively efficiently, hopefully better than real-time.

The algorithm I used relies on data from 4 nearby comparison pixels. I chose to use the four pixels that were 3 pixel-widths North, South, East, and West of the target pixel. I chose to use 3 pixel-widths because the images generated by EPIC are calibrated such that the beam is about the size of one pixel, and so I reasoned that 3 pixel-widths was far enough away from the target to be statistically isolated from its beam's sidelobes.

The relative luminosity of each comparison pixel (and the target pixel as well) is stored as a timestream in memory, for the past 200 time samples. This is simply so that we have statistics about each pixel. Then, on each frame of the data, z-scores are calculated for each of those pixels. The z-score is calculated as (current value - median value of last 200 samples) / (standard deviation of last 200 samples); in English, the z-score is simply how many standard deviations above the median the pixel's current luminosity is. The median was used rather than the mean, because the mean could vary wildly in the presence of RFI. For similar reasons, one may wish to use the Median Absolute Deviation (MAD) in place of the standard deviation.

If the sum of the z-scores of the four comparison pixels is greater than 7, the frame is flagged as a false positive (the assumption here is that such a statistically significant increase in the comparison pixels' z-scores must be the result of RFI, or, at the very least, not indicative of a CGP). This is unless the z-score of the target pixel is larger than any of the comparison pixels', in which case the frame is flagged as a true positive (because, in this case, it may be that the Crab is so bright that its sidelobes illuminate the comparison pixels as well--though, I do not think CGPs are that bright).

The choice of 7 standard deviations was somewhat arbitrary. Fiddling with how many standard deviations constitutes a false positive, and how many times larger Crab must be than the comparison pixels to warrant flagging as a true positive, could both significantly reduce the amount of data which needs to be searched.

With the numbers as they were, there were still a number of instances where RFI was occurring, and the target pixel was visibly increased, but the frame was flagged as a true positive rather than a false positive. Adam and I suspect this is likely due to uncommon occurrences where the comparison pixels are in the nulls between the sidelobes of the RFI sources.

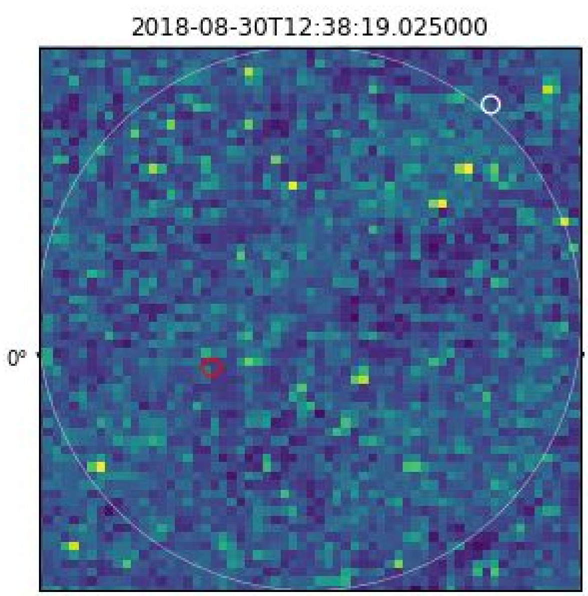
Catalog of Observations

I ran through a few different datasets, all on the Enterprise cluster. I'll refer to them by their respective directories, which should be public to anyone with access to Enterprise. My aim here is to catalog things about the data which were glitchy or interesting, in the hopes that they will be less mysterious to someone investigating these datasets in the future.

Data Location: /data5/LWA_SV_data/CGP_180830/
Date: 30 August 2018
Center Frequency: 73.6MHz
Observing Duration: 1066s (~18m)

This data was taken with the hopes of looking at Crab. Unfortunately, the data seems unrecoverably broken.

Here is an example image from the dataset:



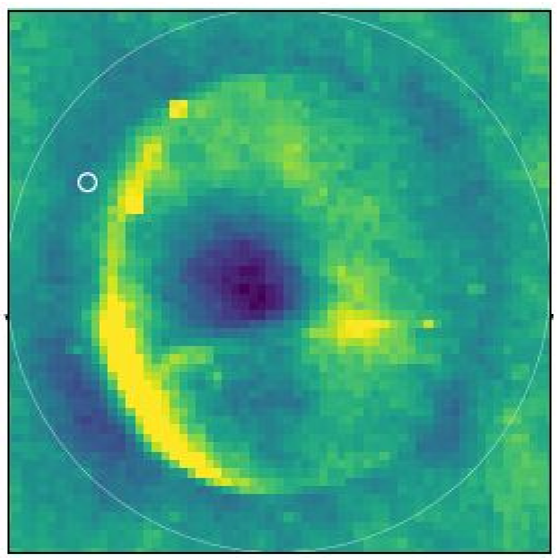
In theory, the Crab Pulsar (and the Crab Nebula, which is a visibly bright radio source) is circled in red, and Cygnus A is circled in white. In practice, there is nothing visible there. Although the data appears, from a static observation, to be qualitatively noisy, looking at a timestream of the data shows that the brightest points in the datastream stay consistently bright, even though they do not seem to correspond to any physical source. I suspect there was an error in the way the data was initially processed or written by EPIC.

Data Location: /data5/LWA_SV_data/Chan2_180831/
Date: 31 August 2018
Center Frequency: 55.3MHz and 43.6MHz (see below)
Observing Duration: 7741s (~129m)

This dataset changes from 55.3MHz to 43.6MHz at frame 100000, or 5000 seconds into the observation.

Unfortunately, Crab is not visible throughout the duration of this observation.

A glitch in the version of EPIC used to create this data resulted in the horizon being transformed from what was expected. Here is a frame from the 55.3MHz section of the data:



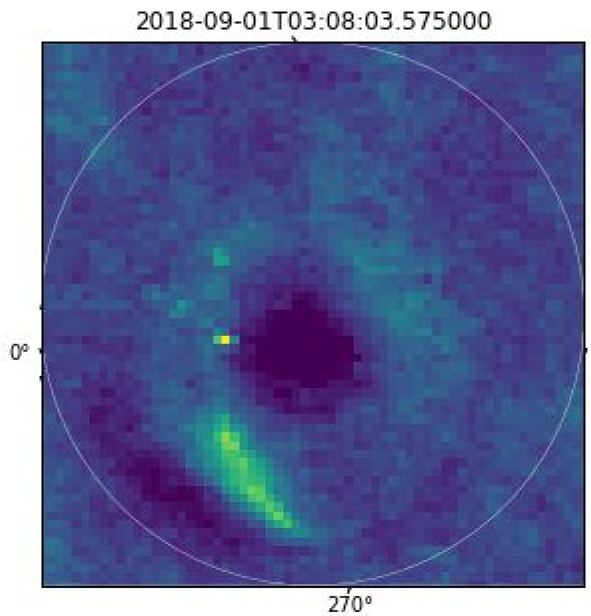
As can be seen, the horizon of the telescope's field of view is shrunk inwards from where it should be: it should match the large gray circle. This also causes problems for the coordinate-tracking software: this is visible from the white circle, which should be the location of Cygnus A, being removed radially outward from the bright point which actually represents Cygnus A in this image. The coordinate-tracking can be hack-fixed by using the following lines of code when the data is read in:

```
hdu.header['CDELT1'] *= 74/55.25
hdu.header['CDELT2'] *= 74/55.25
```

In this case, 74/55.25 is the ratio between the frequency the software expected the data to be at, and the actual frequency at which it was taken.

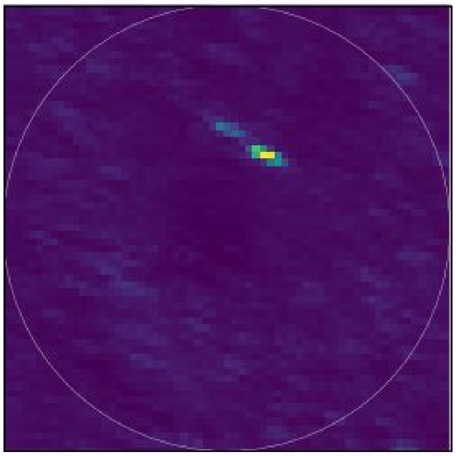
This fix does not change the position of the horizon in the actual resulting image.

This dataset also has a horizon displacement issue in the 44.6MHz portion. In addition to being shrunk, the horizon is displaced southward:



In addition, the hack-fix used previously does not work for this part of the dataset to fix coordinate tracking.

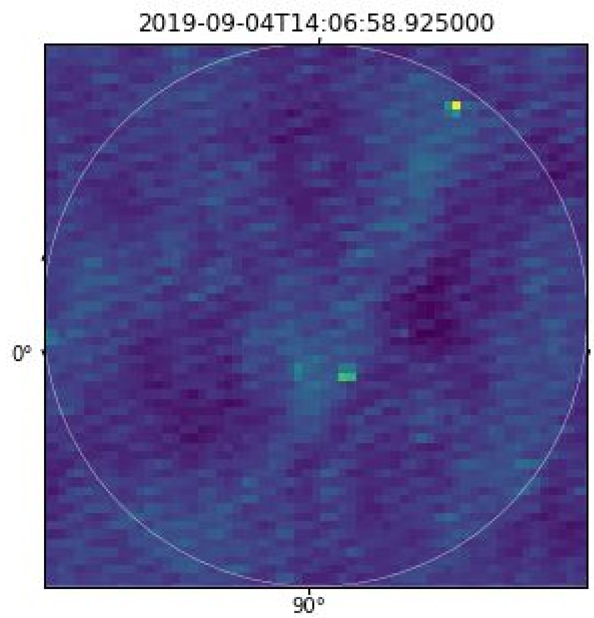
This section of the dataset exhibits an interesting transient from frames 115250 to 115300:



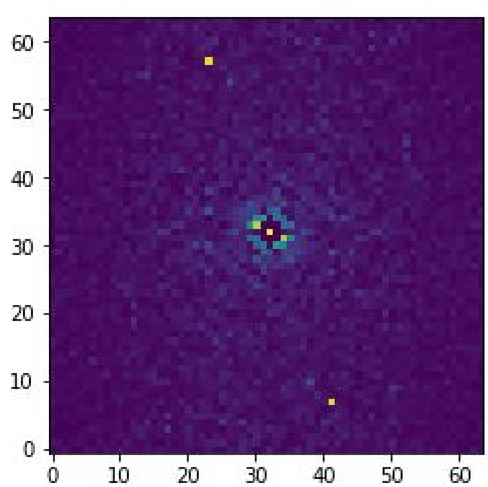
We initially assumed this to be a streaky sky-based source, perhaps a meteor. However, after discovering how shifted the horizon is in this section of the dataset, we think it is more likely ground-based RFI. The source is observed to “flicker” on and off for around 2 seconds before disappearing. For a few frames, the source even appears to shift position slightly. It is unclear why the shape is so streak-like. There is some speculation it could be lightning. Historical weather data from Albuquerque lists only partial cloud cover around the time this data was taken, though Albuquerque is some distance from the LWA.

Data Location: /data5/LWA_SV_data/20190904_131650_7200s/
Date: 4 September 2019
Center Frequency: 73.6MHz
Observing Duration: 5998s (~100m)

This dataset has the Crab Pulsar pass through its highest point, and so would be a fantastic candidate for searching for CGPs. Unfortunately, there is some streaky distortion to the data:



We converted a frame of the data to the uv-plane by performing a 2d Fourier Transform on it. This shows that there seems to be a single baseline which is overcorrelated:



The yellow dot in the top-left and bottom-right represent a pair of antennae which are overcorrelated with each other, due to some error in how signals are being transferred or the data is being processed. Any other frame in the data produces a roughly similar uv-plane image. This dataset can be cleaned up somewhat by FFTing each frame of the data, zero-ing the two offending pixels, and FFTing back to an image. I have not tested whether it is feasible to do this for every single frame of the dataset.

Conclusion

Ultimately, I think a big takeaway from this project is that it is quite difficult to work around messy data taken by EPIC. While removing offending antennae and baselines is a fairly straightforward post-processing step in traditional radio astronomy, the fact that there is no way to work backwards to individual baselines and antennae with EPIC data means that all mis-behaving antennae need to be known before an observing run, rather than after.

Of the two datasets we had that actually observed the Crab Pulsar, one appeared to be total nonsense, and the other had not only significant streakiness due to a single overcorrelated baseline, but other noise due to other misbehaving antennae as well.

A secondary concern is the current lack of bandwidth that EPIC is working with, due to the limitations of running on a single GPU. We have barely-sufficient SNR to detect a CGP with our current bandwidth. More bandwidth would definitely be ideal. Additionally, the previous LWA detection of CGPs notes that they blank out entire bands which exhibit frequent RFI; because EPIC has very few bands to work with, we don't have this luxury as an option to deal with false positives caused by RFI.

Dedispersion will likely be necessary for a significant detection. All my work thus far has been averaging across frequency bands, which is questionable even with our very small bandwidth, and probably completely useless when considering bandwidths on the order of 1 MHz or more.

Finally, I'll provide here a brief index of problems you may encounter while continuing this project, and where in this document you can read more about them:

Issue	**Page**
Streaky image, over-correlated antennae	7-8
Displaced or shrunken horizon	5-6
Stationary noise, poor image	4
Inaccurate coordinate tracking	5-6
RFI appearing in the sky	6
Multiple frequencies in one dataset	5