Real-Time RFI Mitigation using Spectral Kurtosis

Anastasia Kuske[1], Luke Hawkins[2]

Other Credit: Aman Kar[3], Richard Prestage[4], Evan Smith[4]

[1]Franklin & Marshall College, [2]Green Bank Observatory,
[3]University of Wyoming, [4]West Virginia University

National Radio Science Meeting, January 9, 2019

Background: SERAproject

- Software Enabled Radio Astronomy
- Pilot Project: Pulsar Timing Residuals
- Success: Improvement in residuals
- Runs much slower than real-time

Theory of Spectral Kurtosis (SK)

- The SK Estimator was proposed by *Nita et al.* (2007) as a real-time RFI mitigation technique
- SK measures the randomness of a data set
  - Can identify signals that deviate from Gaussian statistics
  - Sensitive to man-made signals such as cellphones and radar

The SK Estimator

- $\widehat{SK} = \frac{MNd + 1}{M - 1}(\frac{MS_2}{{S_1}^2} - 1)$
- $S_1 = \sum_{i=1}^{M}(\sum_{j=1}^{N} P_j)_i$
- $S_2 = \sum_{i=1}^{M}(\sum_{j=1}^{N} P_j)_i^2$
- P is the power spectrum
- *M* is the integration time
- *N* is the number of channels
- *d* is the shape parameter

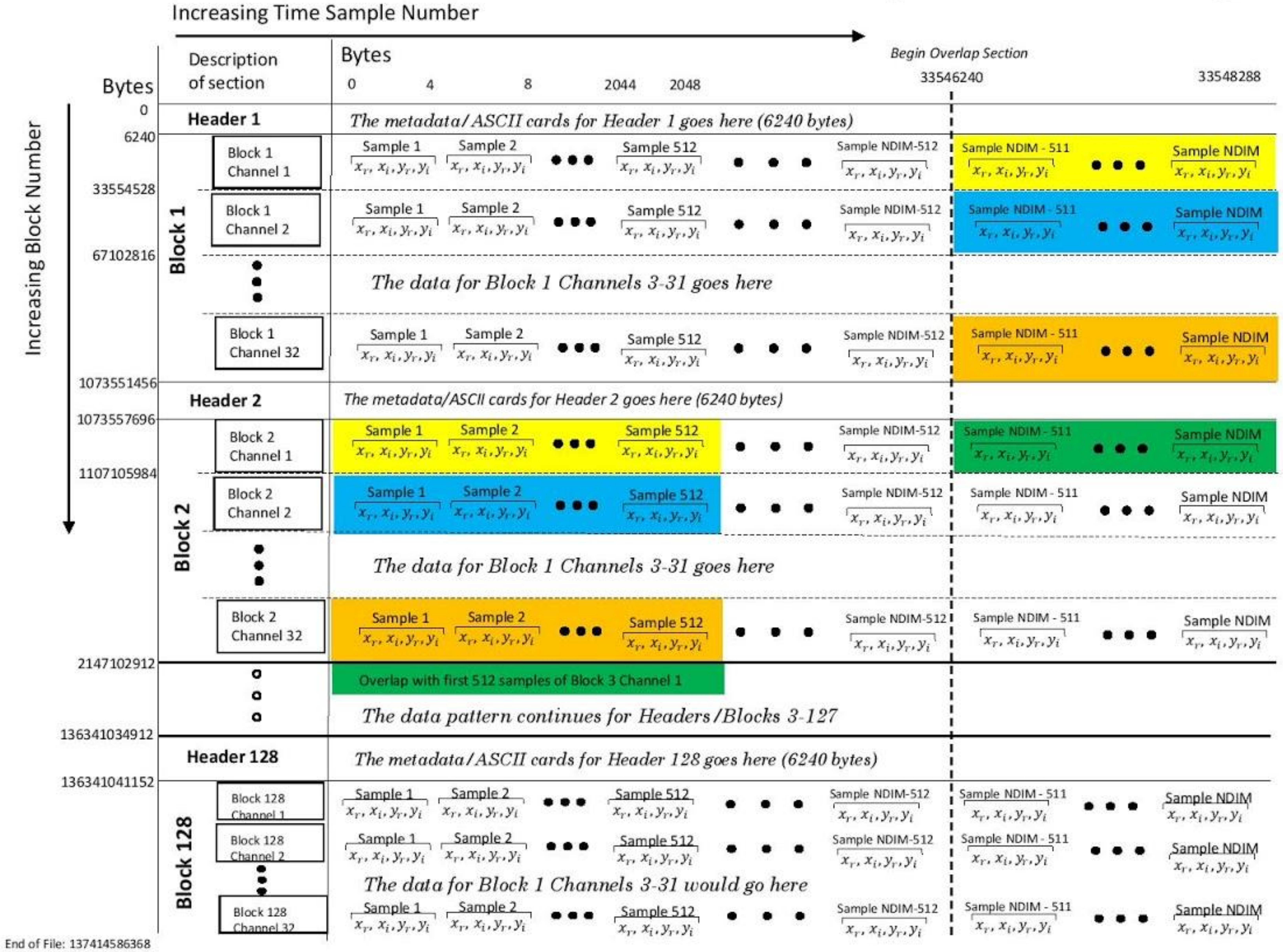
The General SK Estimator

- The original SK estimator was limited by existing hardware
- Requires instantaneous PSD estimates
- Hardware on existing instruments compute only the averaged power spectrum
  - No capability to obtain instantaneous spectra
  - Nita & Gary (2010b) proposed a generalized SK estimator to overcome the hardware limitation

Theory of Mitigation

- Natural astrophysical phenomena produce additive white Gaussian noise
- RFI skews the data can be identified using SK
- Used GUPPI raw files from the GBT

GUPPI Raw Data Format – Specific Design



For argument (and this is reasonable based on the values specified in the header sections of GUPPI files), we assume there are 128 Blocks, the header file is 6240 bytes, OBSNCHAN = 32, NPOL = 4, NBITS = 8, BLOCSIZE = 1073545216, and OVERLAP = 512.

Additionally, NDIM is the number of time samples per channel (per block). The value is calculated by the following:

$$NDIM = \frac{BLOCSIZE}{OBSNCHAN * NPOL * \left(\frac{NBITS}{8}\right)}$$

In our argument, NDIM = 8387072.

The total bytes in a common Lustre GUPPI file is 137,414,586,368. Thus, the size is approximately 128GB (check byte conversions online if this is not readily apparent).

Interpreting the GUPPI Data File

Recall that GUPPI data files are .raw (i.e. binary) files. The data is a series of bytes (8-bit, 1s and 0s). However, the file is formatted in a very precise manner. The bytes are ordered: Header 1, Block 1, Header 2, Block 2, . . . , Header 128, Block 128. For each block, the bytes are ordered Channel 1, Channel 2, . . ., Channel 32. For each channel, the bytes are ordered Sample 1, Sample 2, . . ., Sample NDIM. For each time sample, the bytes are ordered x polarization real, x polarization imaginary, y polarization real, y polarization imaginary.

Because of the strict requirements of the byte ordering, we can think of the stream of bytes as an array of bytes. By imagining an array, not only can we gain a more intuitive feel of the data, but also we can create the visual representation on the left.

Understanding the Visual Representation

First, let's identify the axes. Our axis dimensions are in bytes. From left to right, we indicate the number of bytes as our time sample increases. From top to bottom, we indicate the number of bytes as our channel/block number increases.

The bytes of data in the .raw file correspond to the following path:

1. Start at Byte 0. The next 6240 bytes correspond to Header 1.
2. Arrive at Block 1: The first four bytes indicate the two voltage values (real and complex) for each polarization (x and y) of time sample 1 of Channel 1. The next four bytes indicate the two voltage values (real and complex) for each polarization (x and y) of time sample 2 of Channel 1. Continue through until time sample NDIM. Then, Channel 2 begins, and the pattern continues until the end of Channel 32.
3. Now, byte 1073551456 begins Header 2.
4. The previous steps are repeated until the end of the file (i.e. Block 128)

Understanding the Overlap (colored)

The highlighted sections in the data blocks (yellow, blue, orange, and green) indicate overlapping regions. For example, examine Block 1 Channel 1. There are 512 samples highlighted in yellow. At the beginning of Block 2 Channel 1, 512 samples are highlighted in yellow as well. This overlap simply means that the last 512 samples of Block 1 Channel 1 are the first 512 samples of Block 2 Channel 1. Analogous argument for the other colors.

Notes

Essentially, the string of binary data can be tracked by reading this visual representation like a book. Begin at the top left and read from left-to-right. The bytes along the top indicate how many bytes have been passed during that channel, and the bytes along the side indicate, collectively how many bytes have been passed. Thus, the byte number of any location can be determined by adding the two axis values.

Breakthrough Listen

The visual representation can be understood in the same fashion as described for GUPPI. The only adjustment is to ignore the highlighted regions. There is no overlap in this format. The other difference would be if DIRECTIO was specified to a non-zero value, as described in the previous slides.

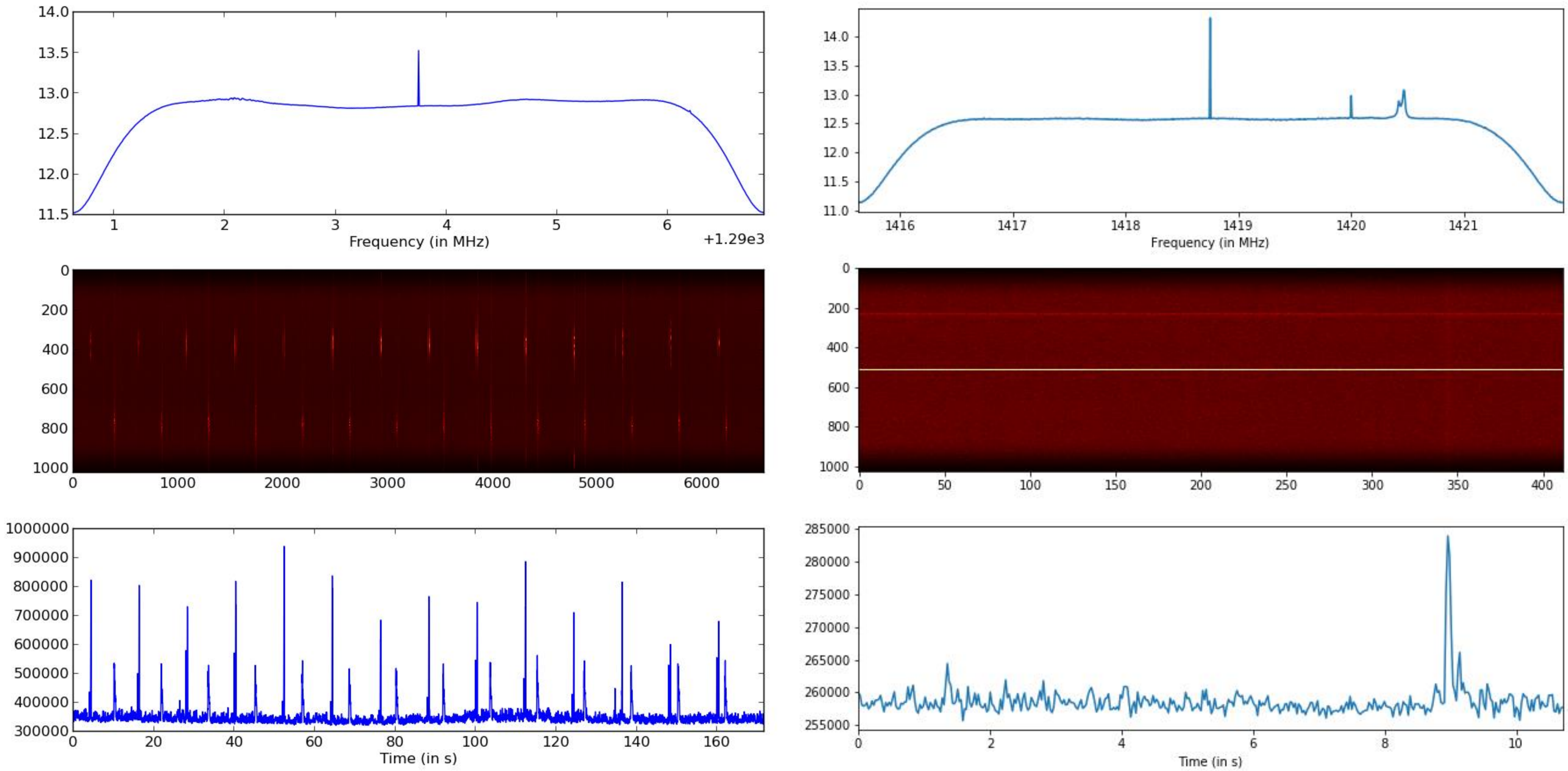
The Useful Programs

- gbtraw.py – (designed by Richard Prestage) offers ways to manipulate GUPPI raw data files
- plot.py – scans through raw data files and looks for potential locations of RFI
- readwrite.py – simulates real-time processing for SK
- sk_data.py – (designed by Evan Smith) contains all the computational functions that are needed to create the SK Estimator

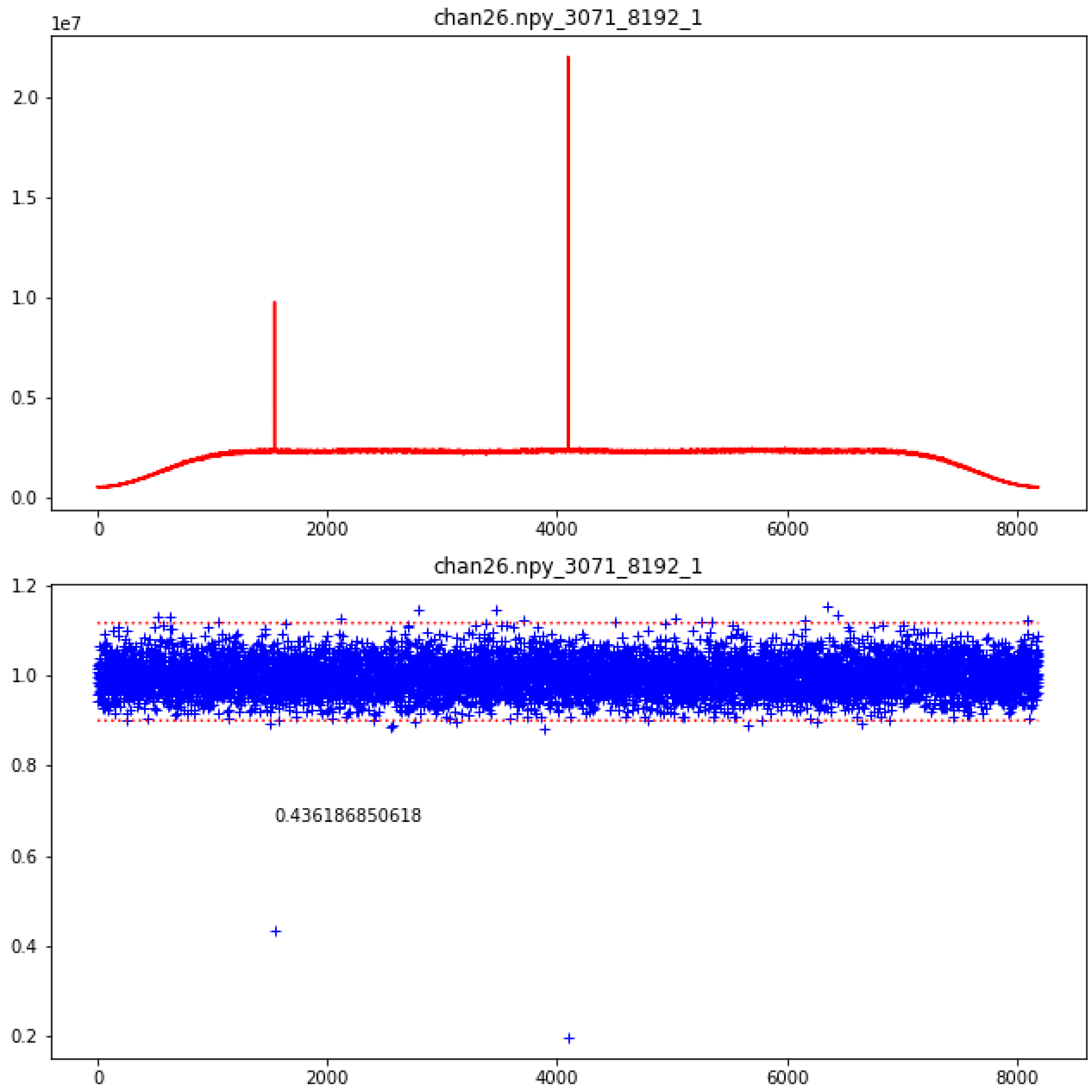
The Final Program:

- Read raw data from file
- Mitigate RFI:
  - FFT
  - Spectral Kurtosis algorithm
  - IFFT
- Write modified data to raw file

Setting the Stage: Spectrograms



Identifying RFI



Future Endeavors

- Polish SK Estimator
- Implement Kurtosis
- Mitigate RFI
- Test Results
- Implement on VEGAS
- Extend to other areas

The Green Bank Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities Inc.

The Green Bank Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.