A New Technique for Ultra-Faint RFI Mitigation

Mike Wilensky, University of Washington
Advisor: Miguel Morales
January 9th, 2019
URSI, Boulder, CO

Talk Overview

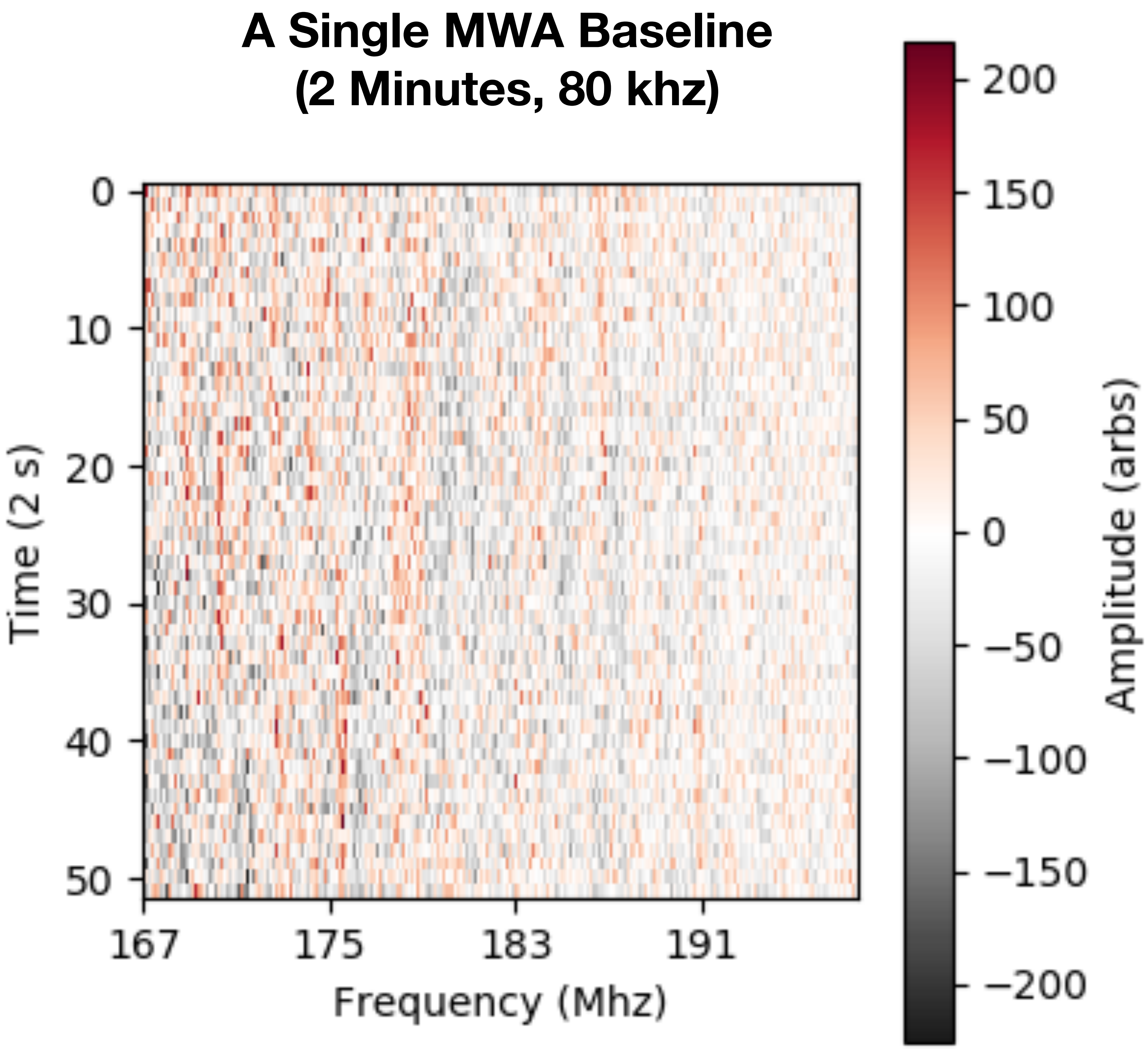
- Introduce the problem and method
  (Sky-Subtracted Incoherent Noise Spectrum - SSINS)

- Demonstrate with the Murchison Widefield Array (MWA)

- Show some examples with the Hydrogen Epoch of Reionization Array (HERA)

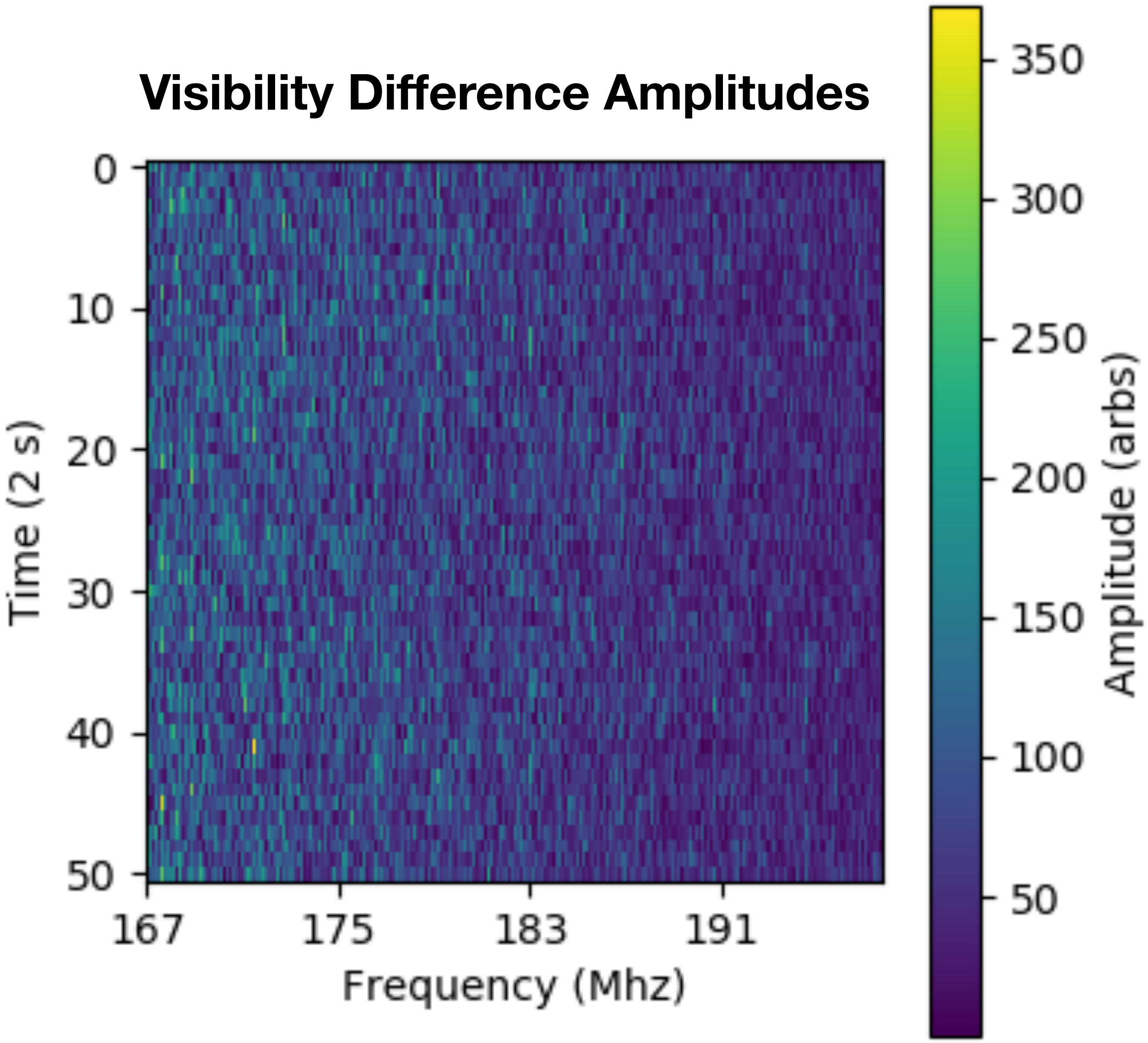
- Summarize the results

One Baseline is not Sensitive Enough



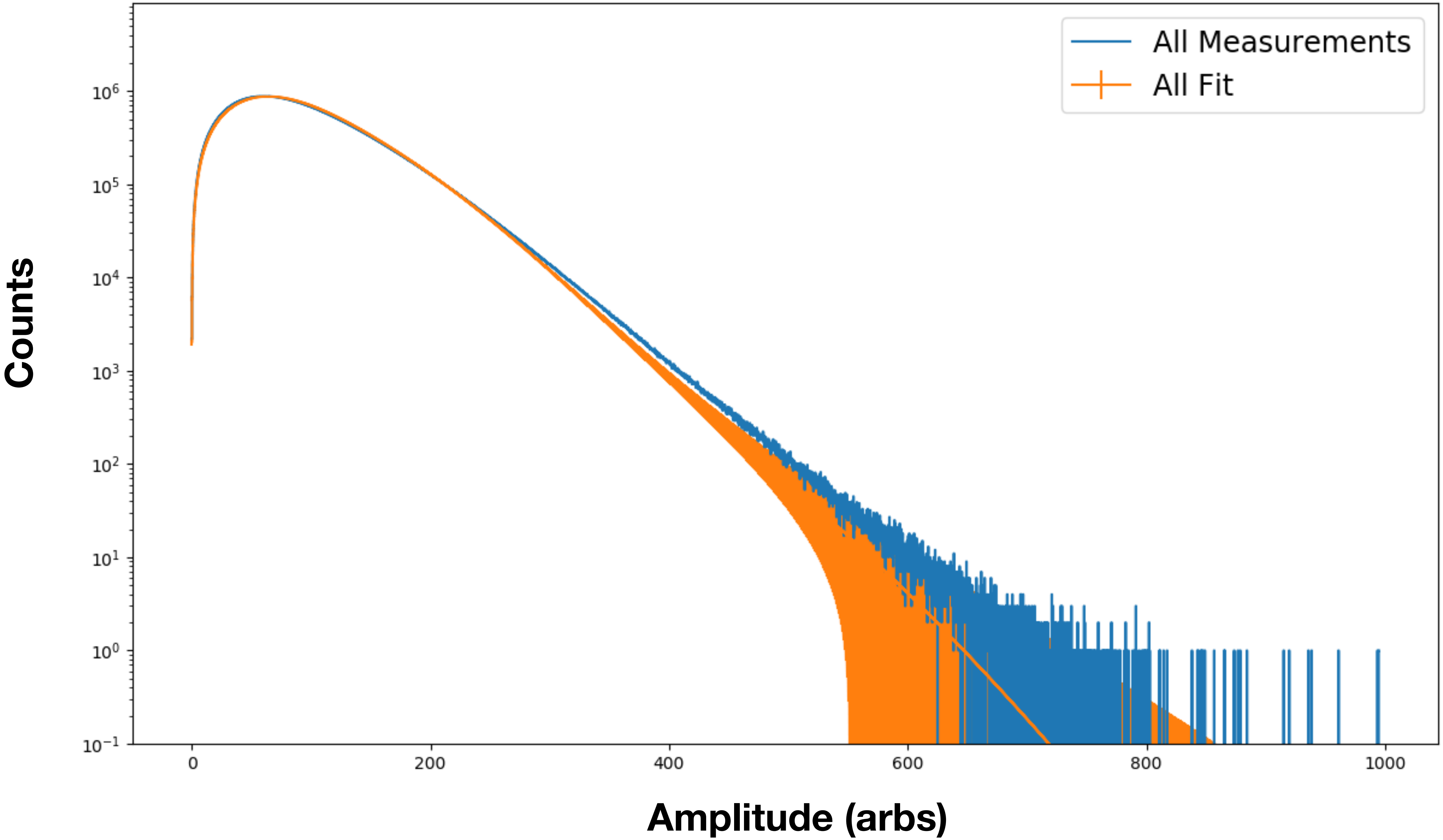
- Current RFI flaggers operate on a single baseline at a time
- Subthermal RFI may be lurking
- Three components: sky signal, thermal noise, RFI
- "Sky signal" here refers both to foregrounds and cosmology
- Need a way to disentangle the three components

Sky-Subtraction Ensures we Are Dealing with Only RFI and Background (Thermal Noise)



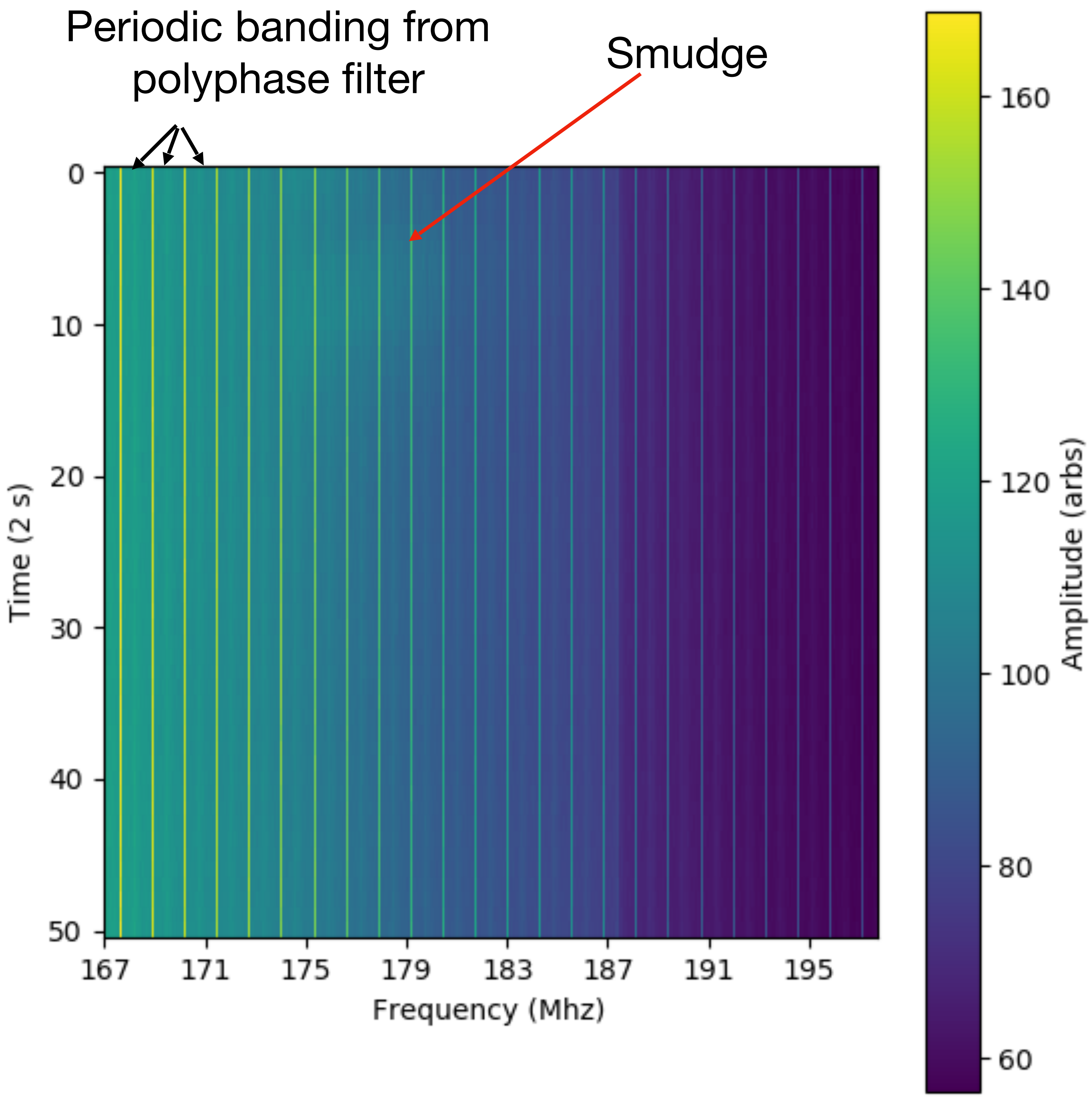
- Sky signal varies slowly in time (relevant time scale is 2s here)
- Taking the difference of the visibilities in time should subtract out the sky signal
- What remains is thermal noise and potentially RFI
- Pictured here are the amplitudes of those differences

Visibility Difference Amplitude Histogram



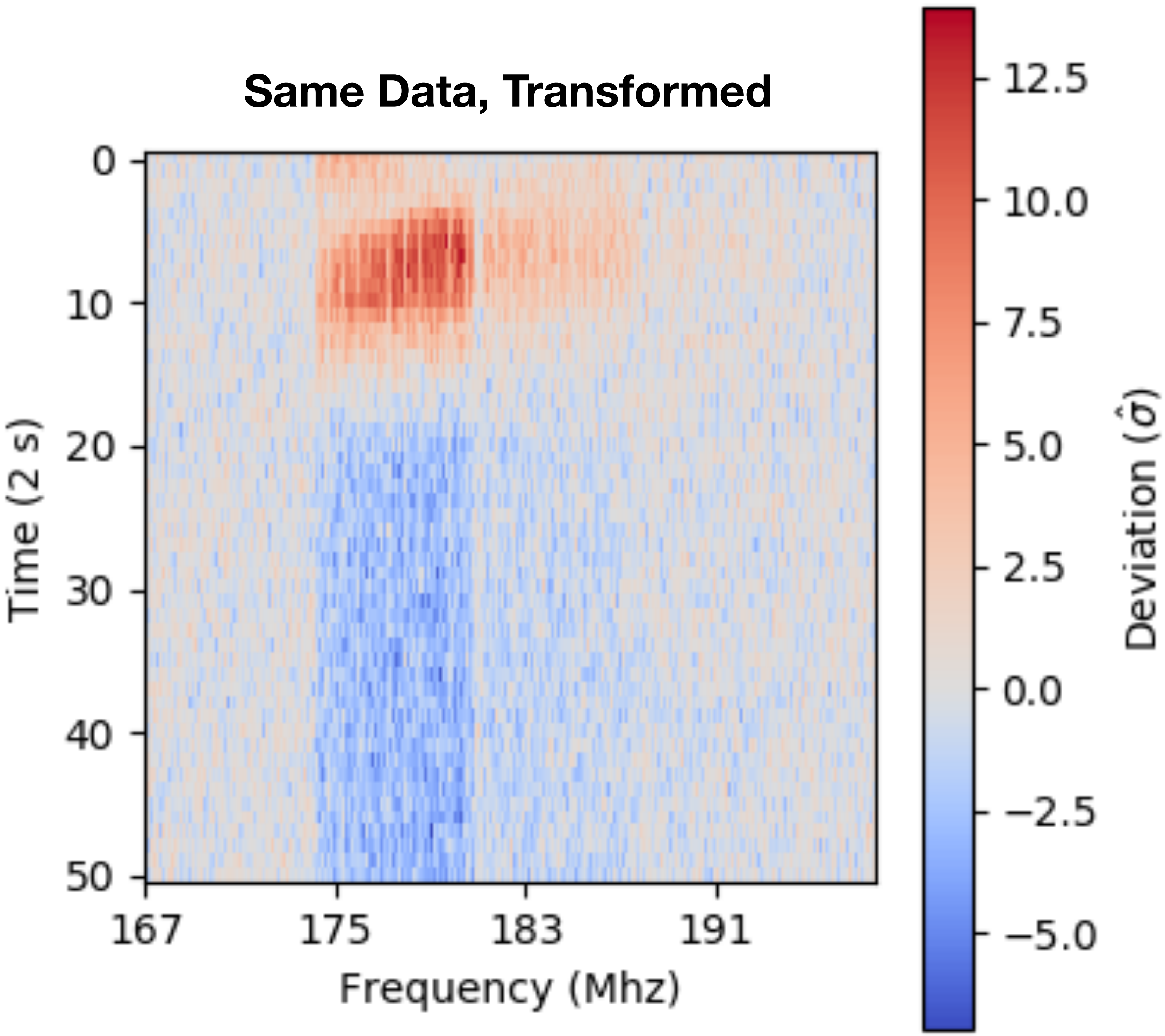
- The fit looks good for the most part, but this does not mean the observation is clean!
- Need to boost RFI-to-noise ratio!

Incoherent Average to Boost RFI-Noise Ratio



- Boost RFI to noise by averaging visibility difference amplitudes across all baselines
- Result is the sky-subtracted incoherent noise spectrum (SSINS)
- ~8000 baselines increases sensitivity by a factor of ~90
- Averaged over ~8000 baselines so the background is now gaussian distributed (Central Limit Theorem)
- Need to boost contrast of smudges

Mean-Subtraction Boosts Contrast



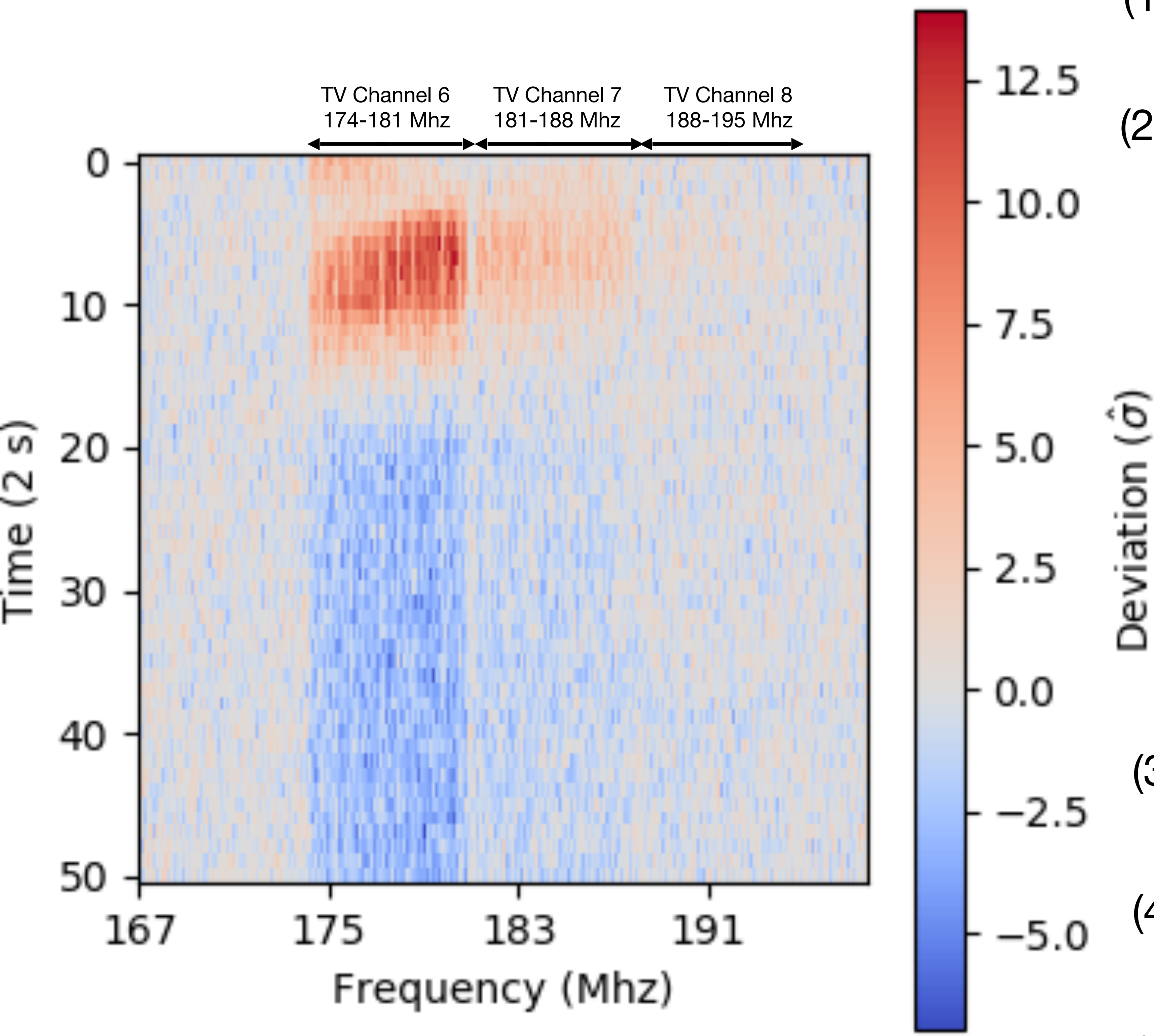
(1) Estimate a mean per frequency

(2) Estimate variance from mean (original background had a single parameter per frequency)

(3) Subtract mean and normalize w/respect to variance for each data point

- **Now the data is in units of $\hat{\sigma}$**
- Now we can see that the little smudge has outliers as strong as 13 $\hat{\sigma}$

The Match-Frequency Filter



(1) Pass known RFI frequencies (shapes)

(2) Find the strongest outlying time of each shape

(a) What remains should be zero mean (by construction)

(b) Average over frequencies of shape at each time to calculate shape deviations

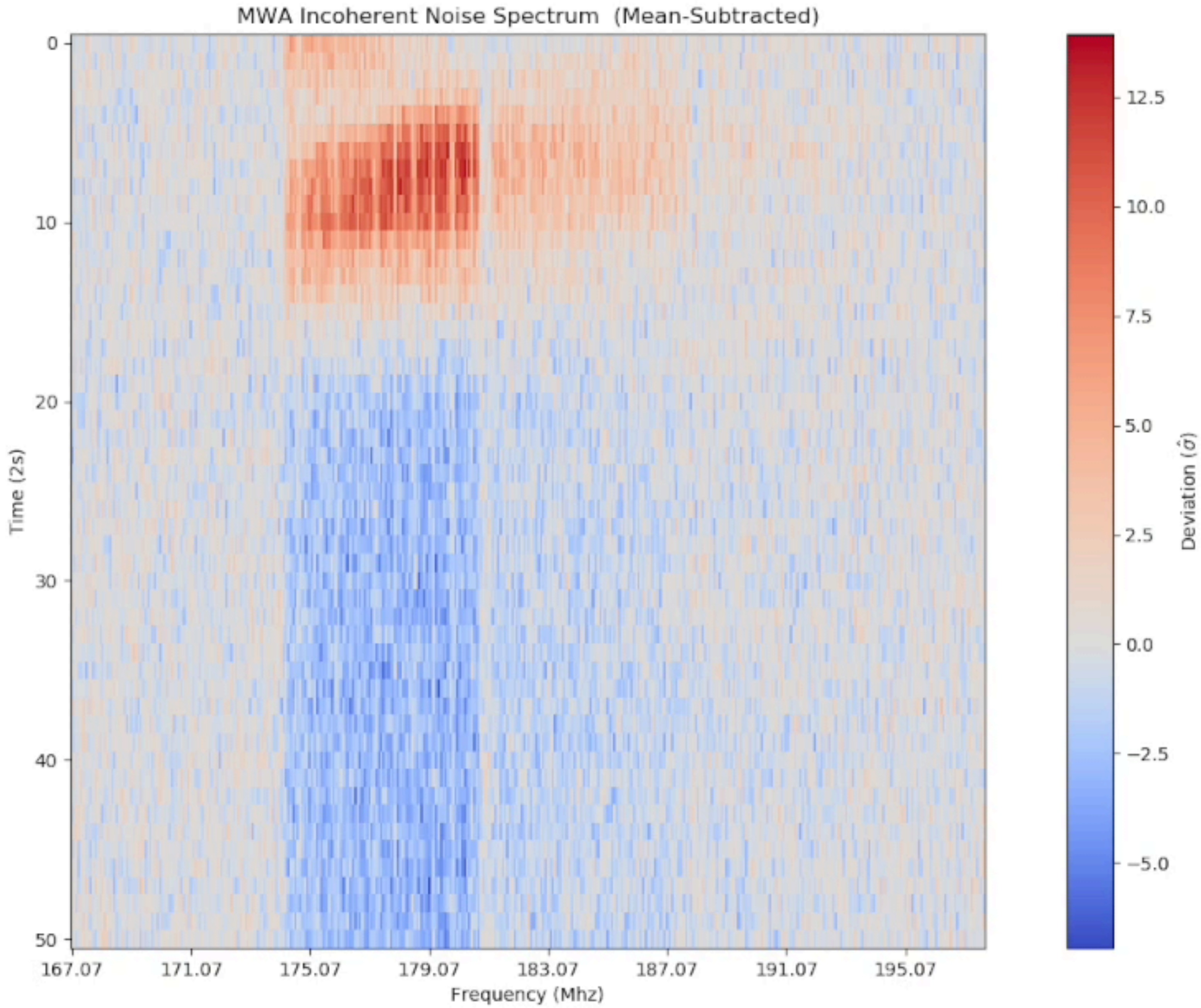
(c) Shape sensitivity boosted by $\sqrt{N_{\mathrm{chan}}}$ relative to single frequency

(3) Find the strongest outlying shape/time of them all

(4) Flag the strongest outlying time for that shape

(5) Recalculate the spectrum

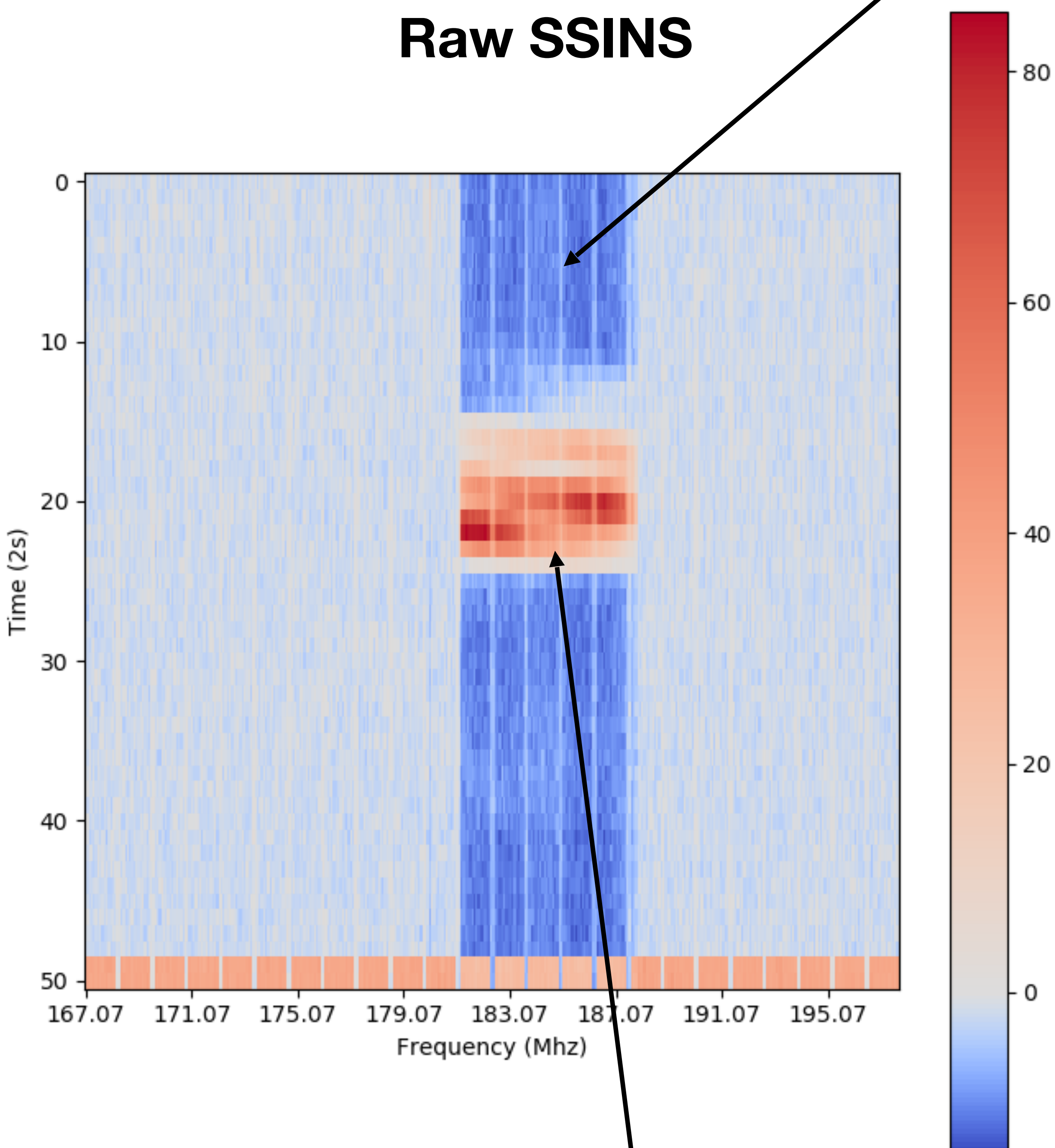
(6) Repeat until clean up to some significance threshold


MWA Incoherent Noise Spectrum (Mean-Subtracted)
Time (2s)
0
10
20
30
40
50
167.07
171.07
175.07
179.07
183.07
187.07
191.07
195.07
Frequency (Mhz)
12.5
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
Deviation ($\hat{\sigma}$)

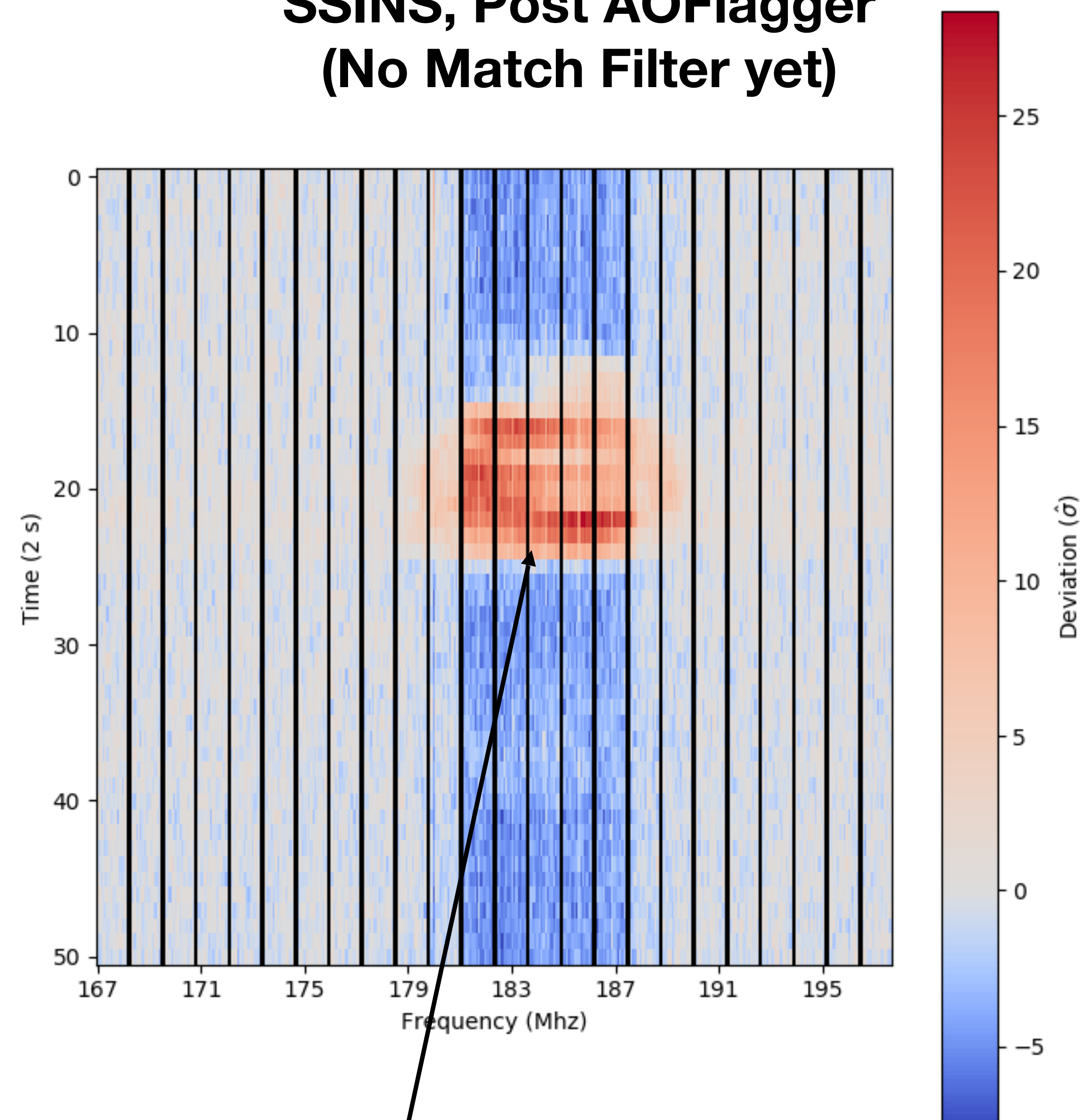
Comparison with AOFlagger

Calibrate on clean part

Raw SSINS

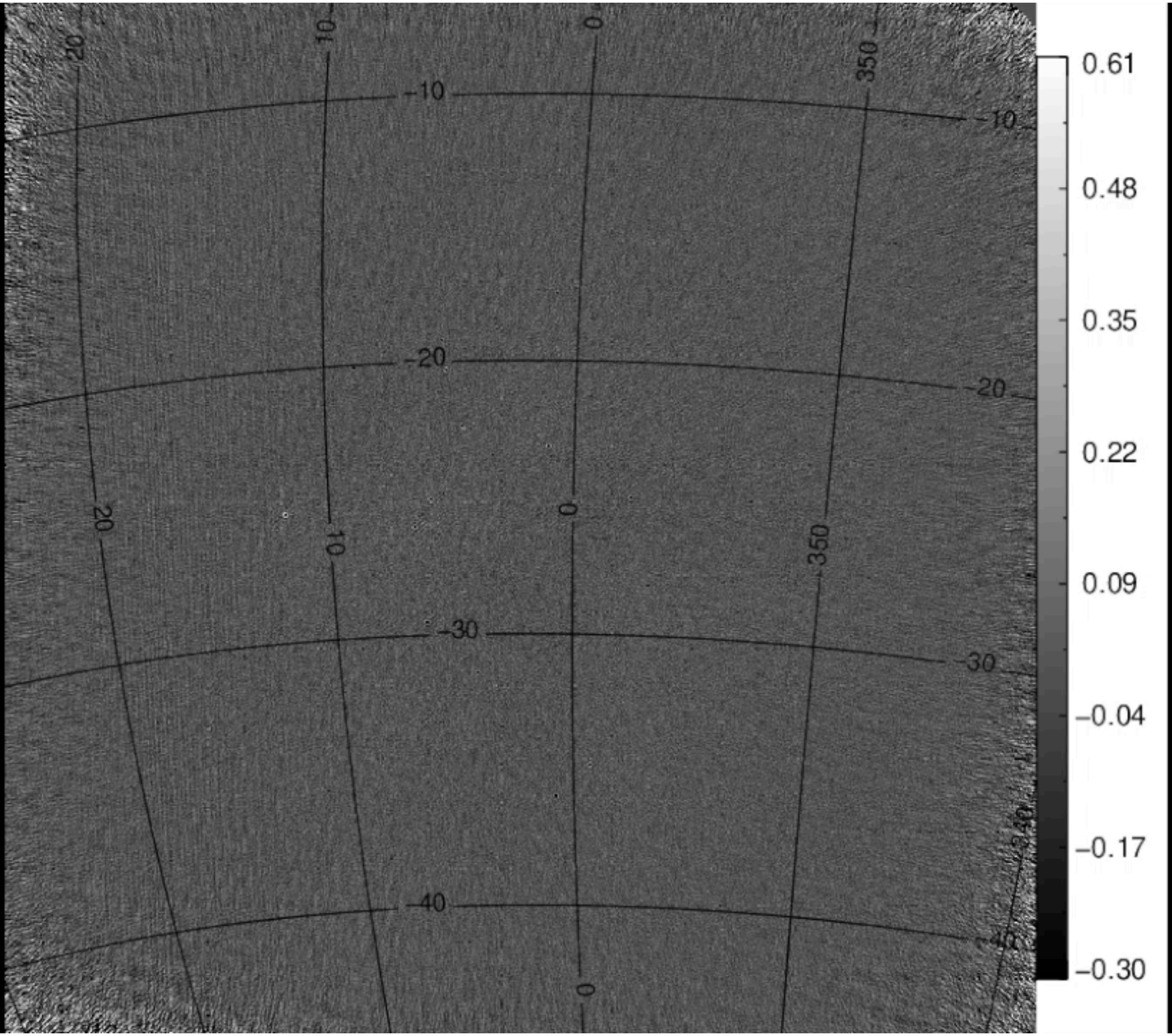


SSINS, Post AOFlagger (No Match Filter yet)



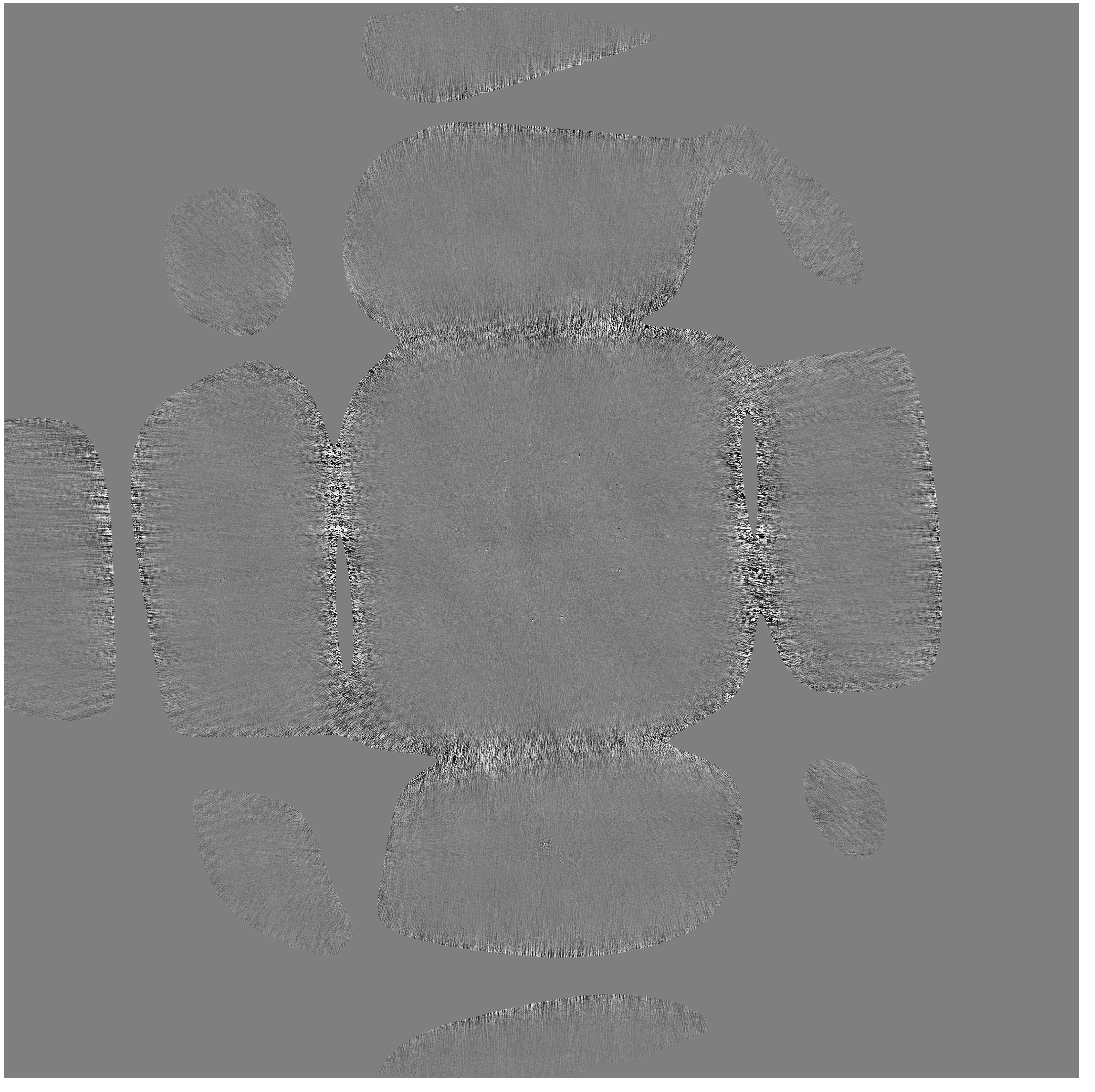
Make snapshots of contaminated part

Leftover RFI
(New match filter snuffs this out easily)



-10
-20
-30
-40
20
10
0
350
340
0.61
0.48
0.35
0.22
0.09
−0.04
−0.17
−0.30

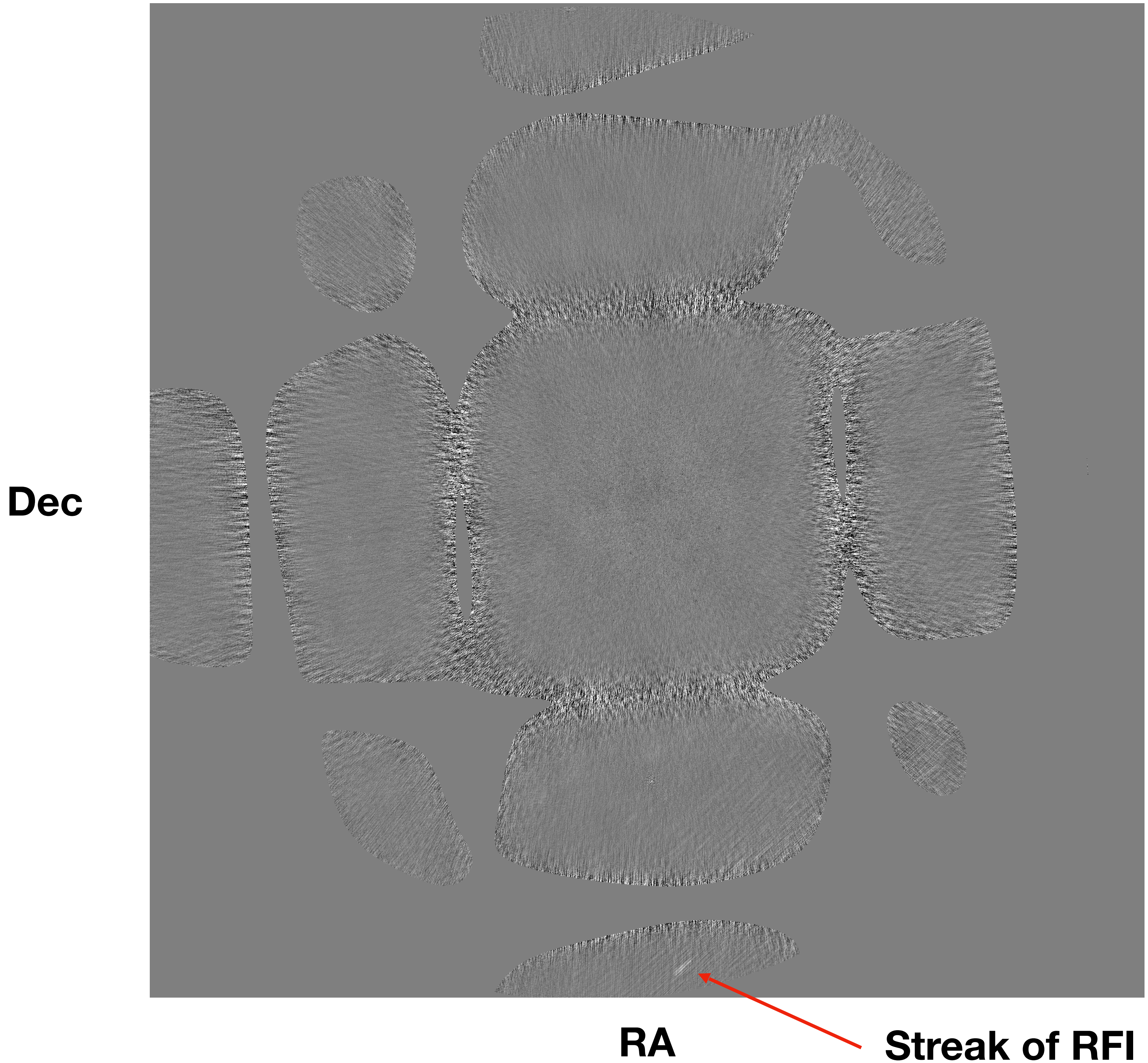
Calibration Image



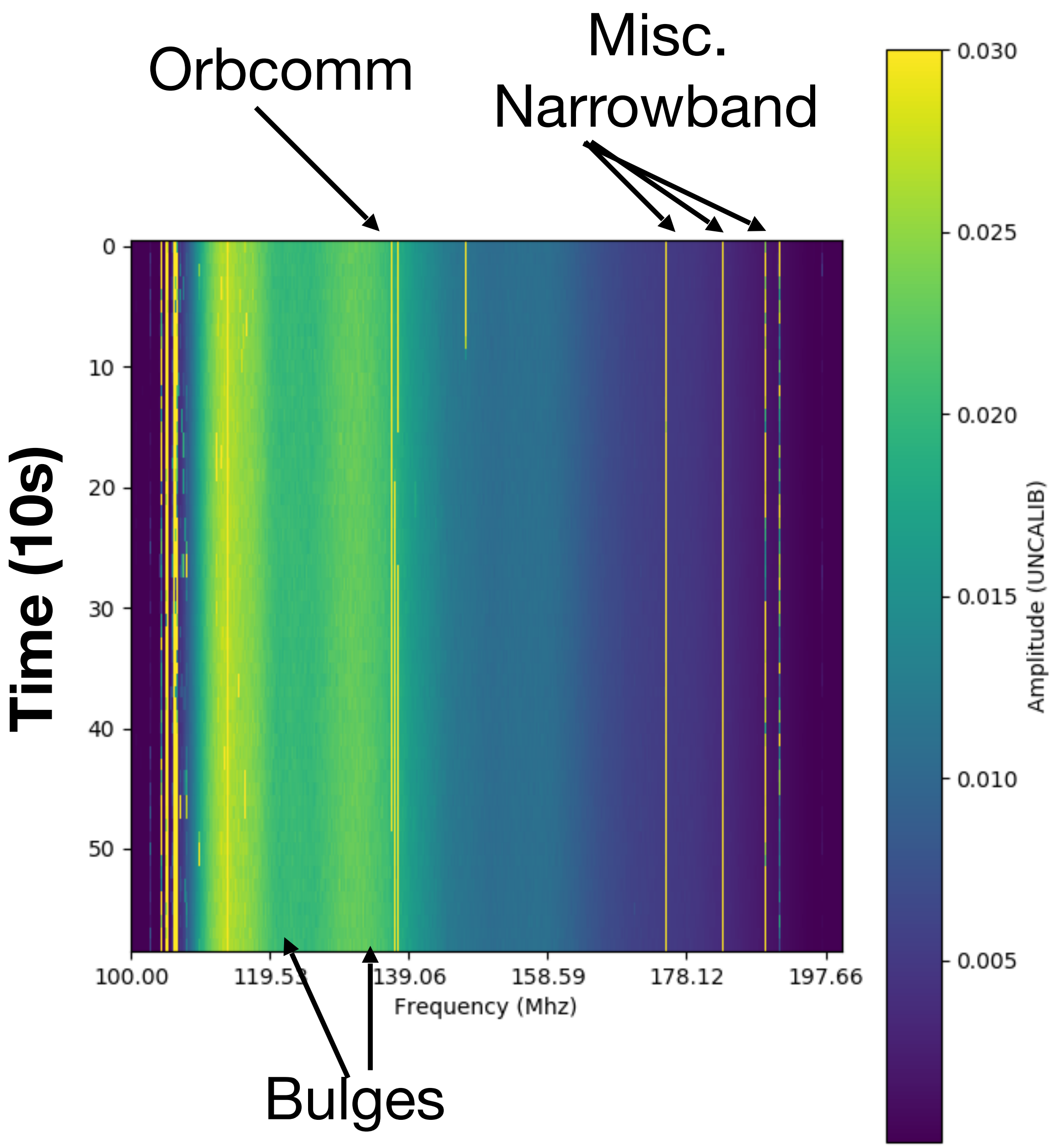
Dec

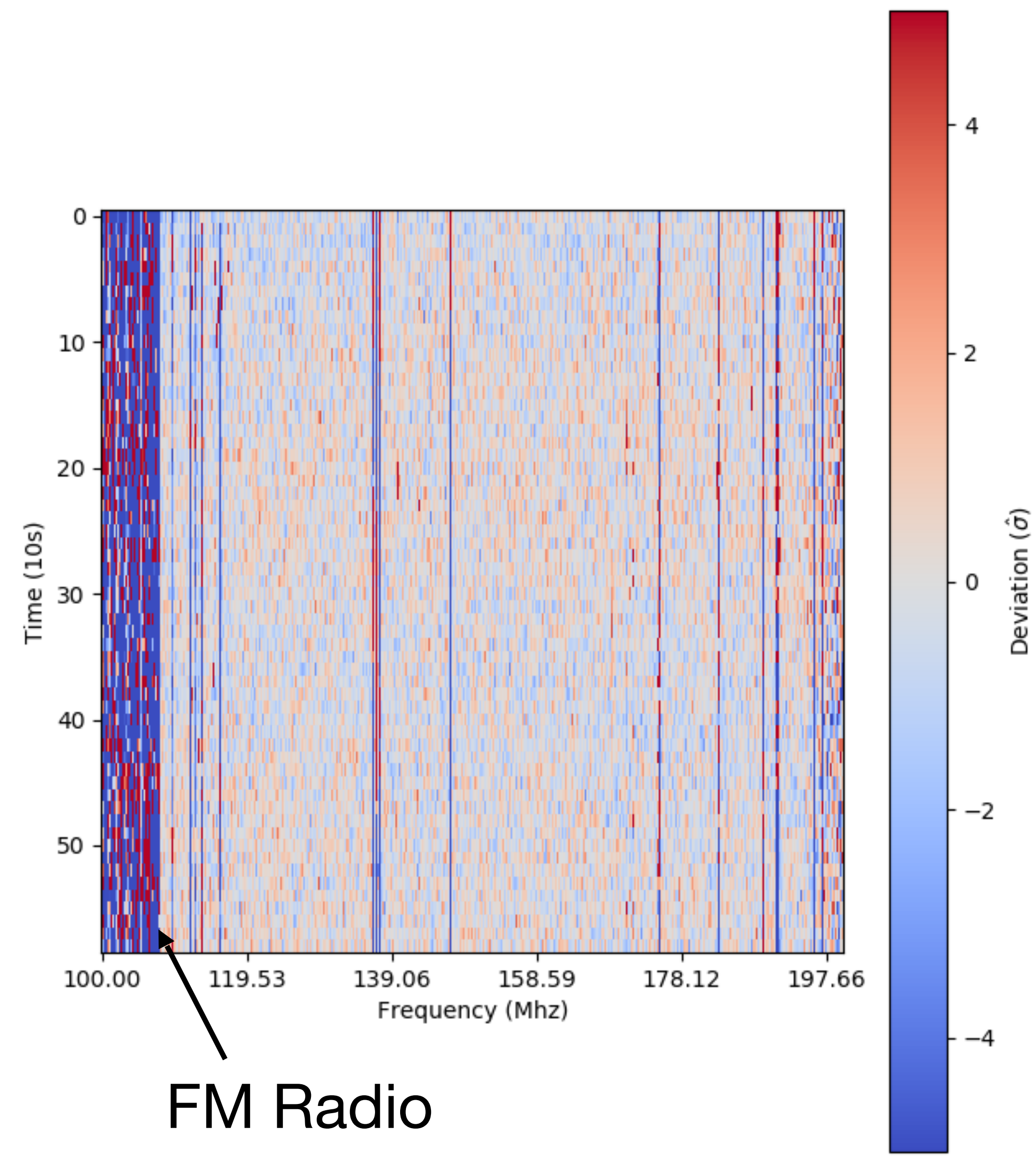
RA

Contaminated Image



Relatively Clean HERA Example

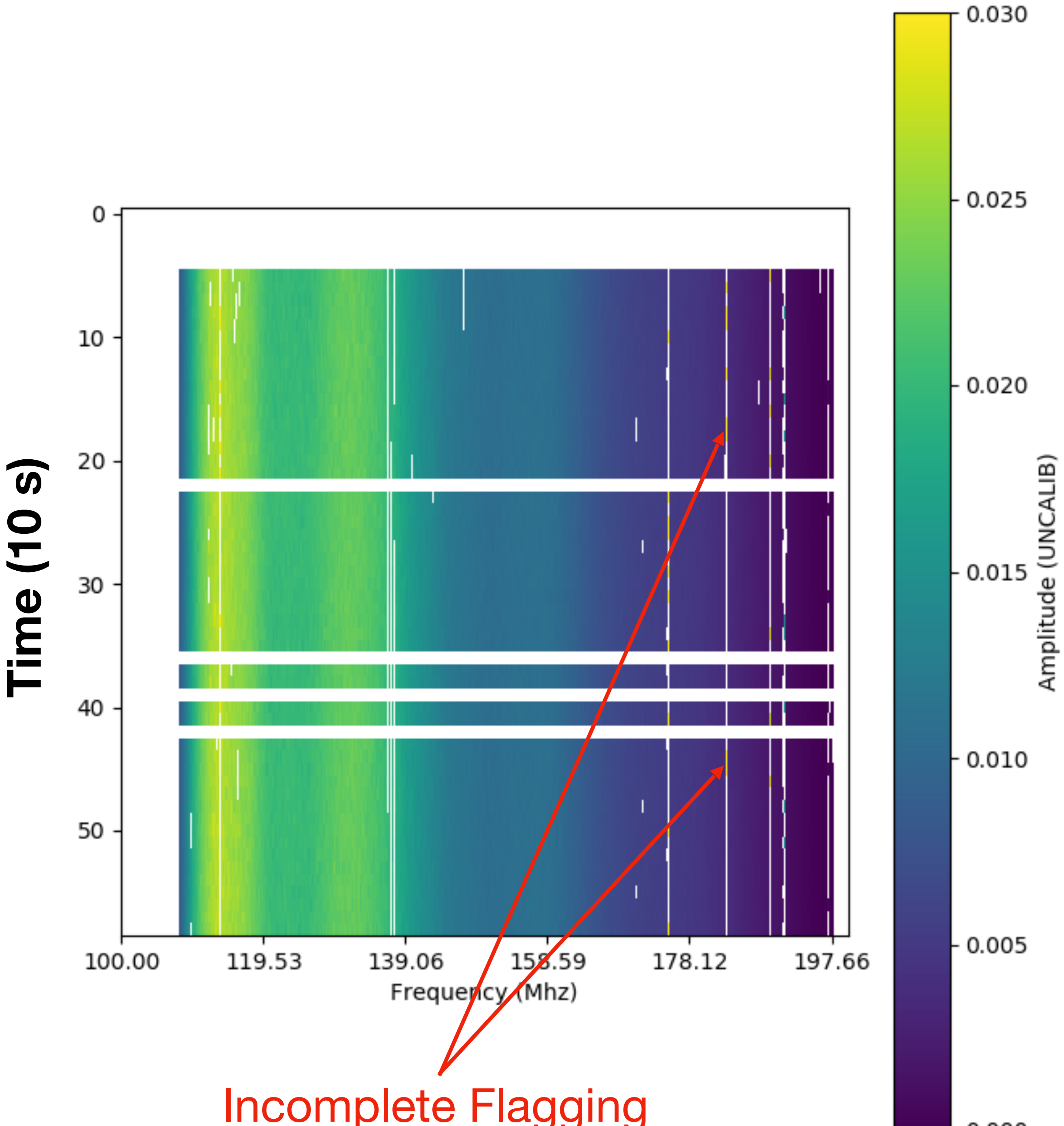




- Orbcomm and FM radio are ever-present in HERA data
- Bulges in lower part of band are some combination of antenna response and sky-noise shape

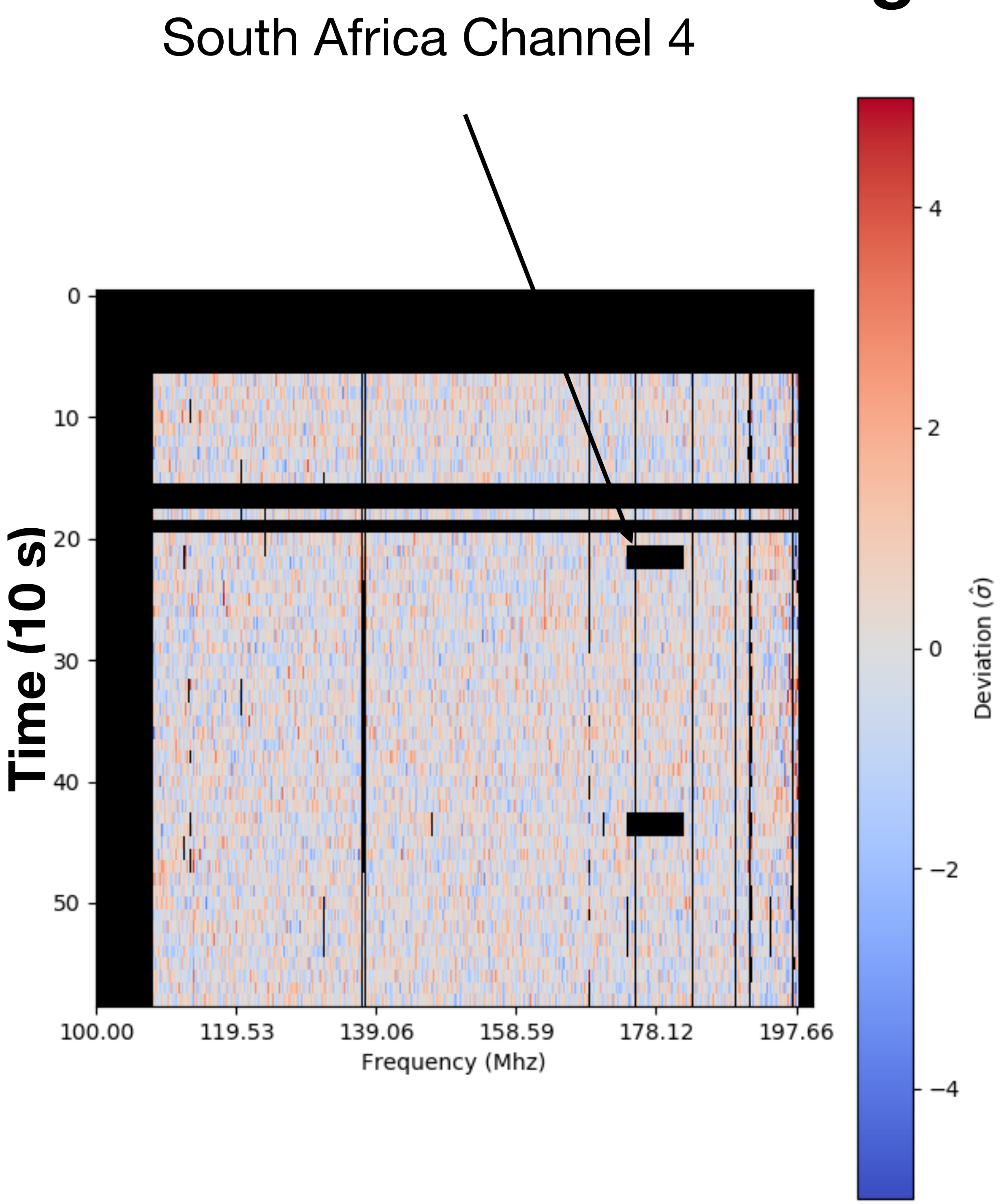
- Typically there are some miscellaneous narrowband transmitters in the high band.

Steady RFI can elude the match filter



- RFI which persists through the entire observation affects the mean no matter how far we iterate
- After substantial flagging, few samples enter the estimation of the mean - the uncertainty in the estimator is high
- At this point, the estimator can no longer be trusted, and it is also unlikely that any clean data remains in that channel
- Just set a threshold to flag what remains
- Sufficiently steady RFI (i.e. as steady as the instrument) will be entirely camouflaged in the mean-subtracted spectrum

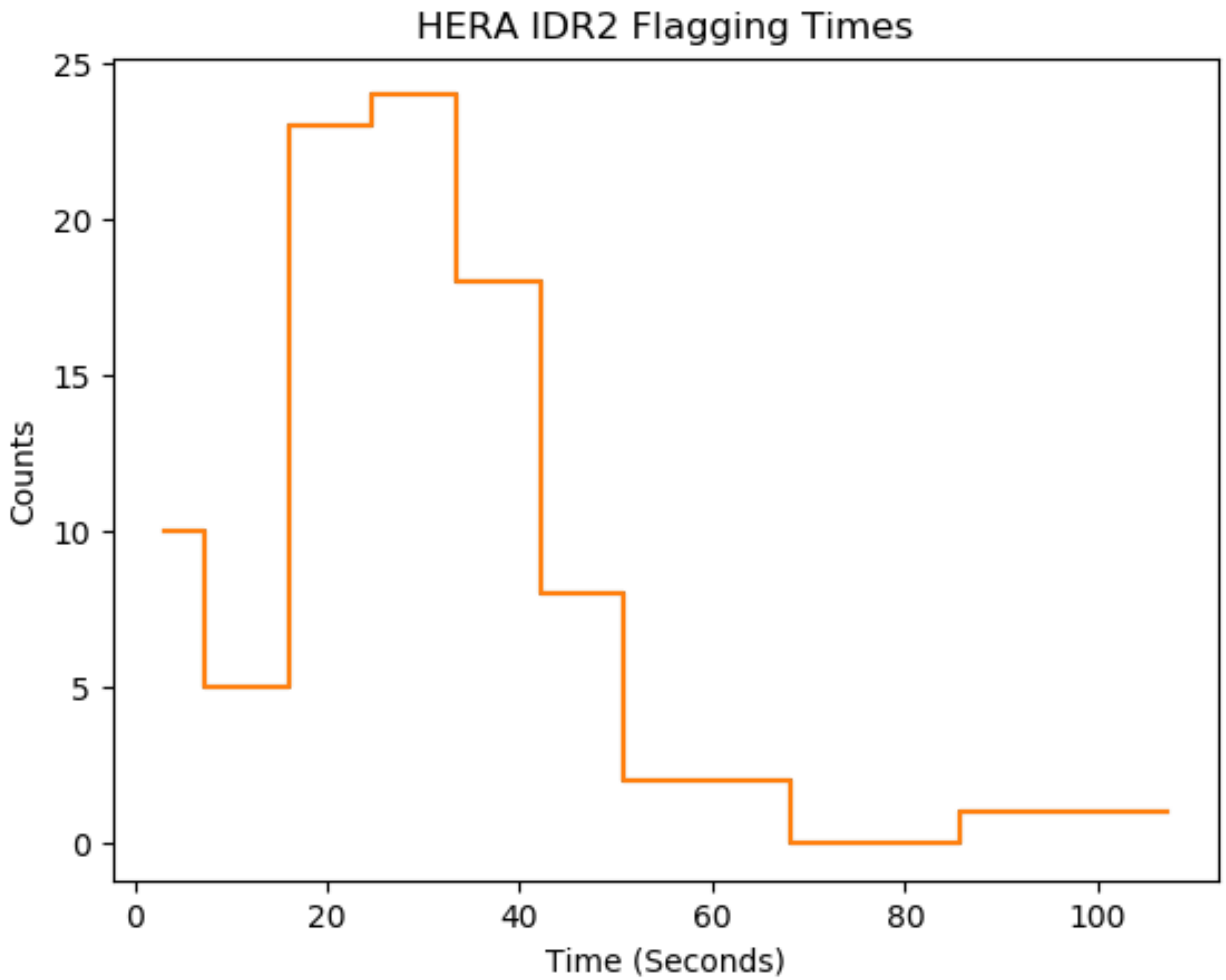
Digital TV in HERA



- Digital TV in South Africa starts at 174 Mhz as well
- 8 Mhz wide instead of 7
- Several adjacent channels
- Match Flagger snags it

Summary/Conclusions

- Introduced SSINS, which can dig deep enough to make ultra-faint RFI obvious, and the SSINS match filter can surgically identify and excise this RFI from our observations
- We verified the efficacy of SSINS by carefully imaging RFI in MWA Observations
- It's really fast! The SSINS of a typical MWA EoR observation can be flagged in less than a second! (HERA observations typically take between 20 and 60s).
- This worked on HERA data, and can work for your telescope if the data is readable by the pyuvdata software package: https://github.com/RadioAstronomySoftwareGroup/pyuvdata
- SSINS is available on GitHub at https://github.com/mwilensky768/SSINS.git
- You can also reach me at mjw768@uw.edu
- This work is funded by NSF Enhancing Access to the Radio Spectrum Grant #1643011



- 95 10-minute Observations
- Mean Flagging Time: 34.23s

- Outliers contained steady narrowband interference over many different channels
- MWA observations (2 minutes) are significantly less contaminated and smaller after baseline averaging - often flagged in less than 1s