RFI Identification and Flagging

Srikrishna (Krishna) Sekhar

NRAO

Outline

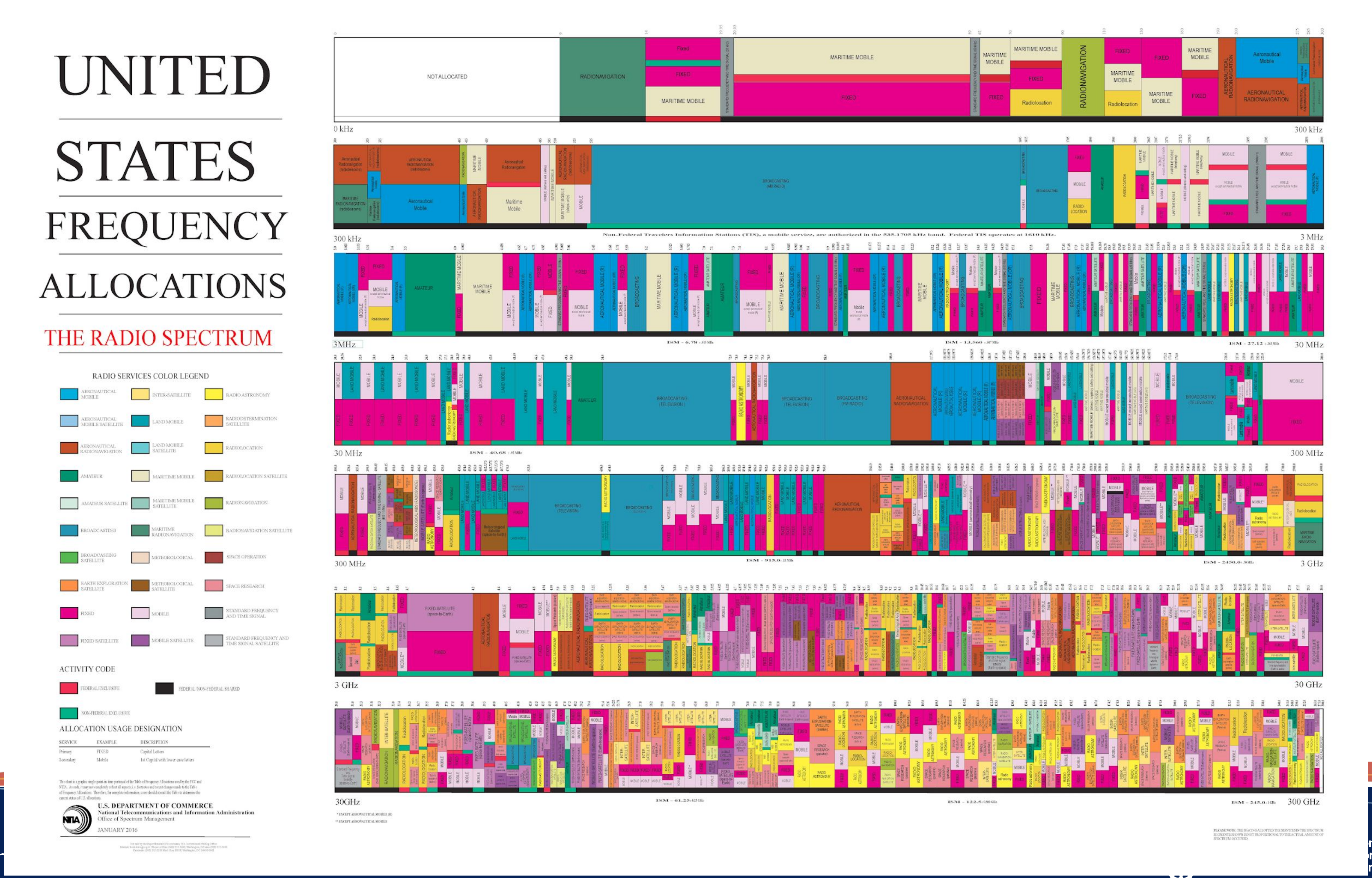
- Introduction to RFI :
  - Sources of RFI
  - Different types of RFI
- RFI Identification
- Handling RFI : Flagging, Mitigation, Excision
- Summary

What is RFI?

- RFI : Radio Frequency Interference
- In general : Any radio signal that is picked up by our telescopes that is **not** astronomical in origin
- Can (and will!) appear in any combination of time and frequency characteristics

Time	**Frequency**
Persistent	Narrowband
Intermittent	Broadband

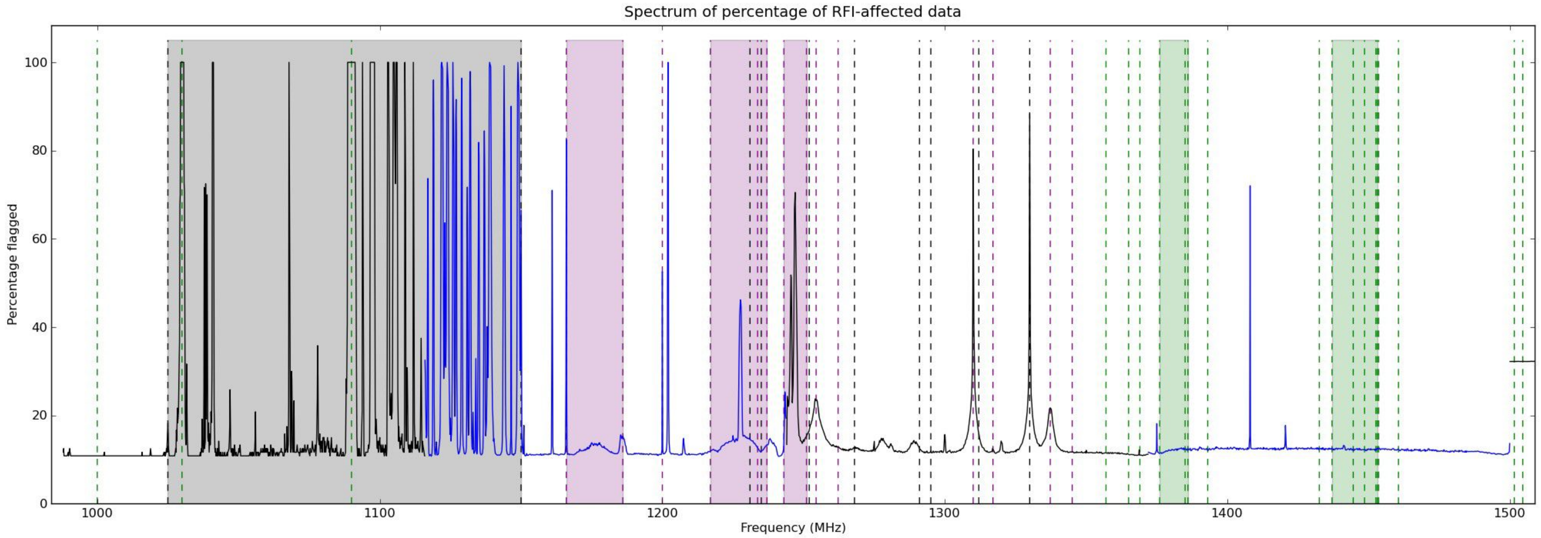
- Narrowband RFI : Satellites, wireless communication etc.
- Broadband RFI : Power lines, motors, spark plugs etc.

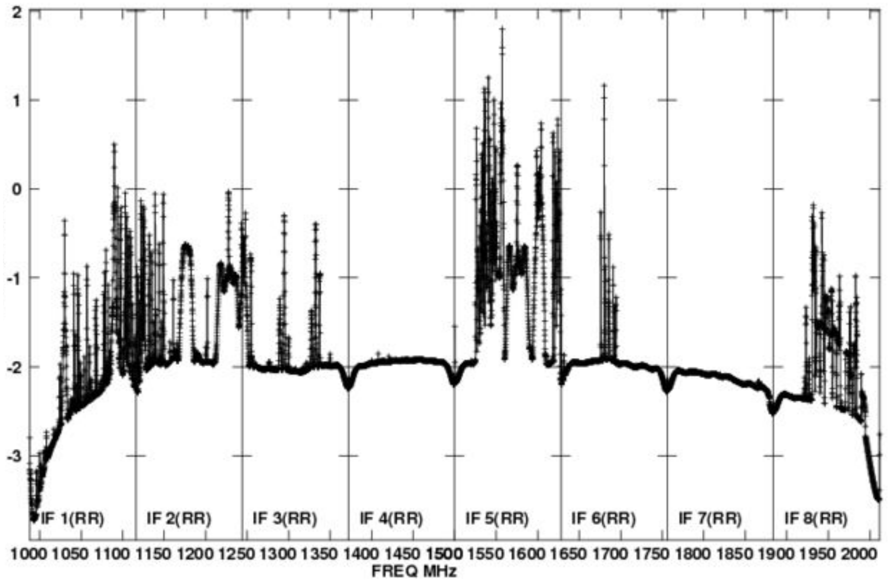


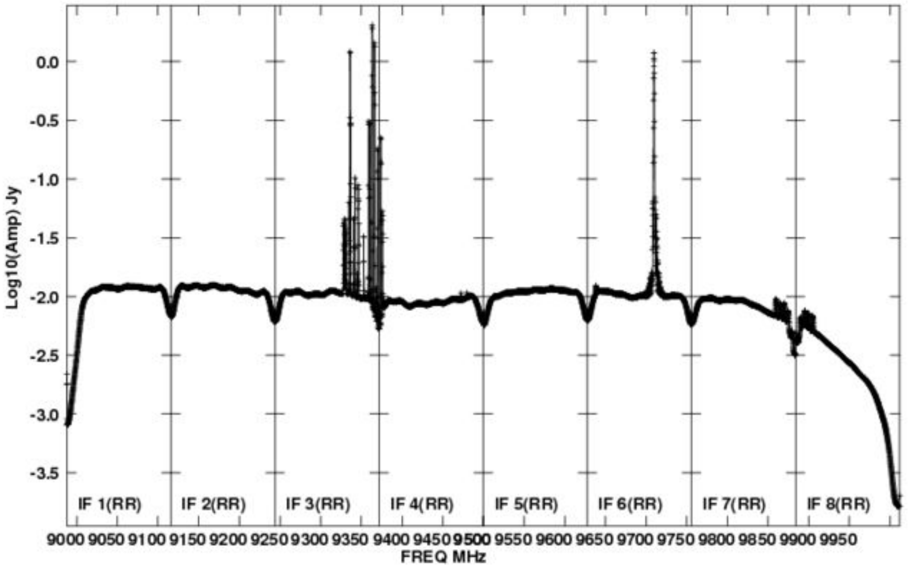
20th

nal Radio
nomy
rvatory

RFI at the VLA







- RFI more prominent at lower frequencies (< 8 GHz) but increasingly showing up at higher frequencies (Starlink!)

Background : Signals & Interferometry

- Measured visibilities are given by (with many simplifications) :

$$V(u,v) = \int I(l,m) e^{-2\pi i(ul+vm)} dldm$$

- In discrete case :

$$V(u,v) = \sum_{l} \sum_{m} I_{lm}(l,m) e^{-2\pi i(ul+vm)}$$

- **Implication** : Signals add **linearly**. Adding multiple sources to your measurement equation simply involves adding their source brightness to the above equation.
- This is also true of non-astronomical sources (RFI; Satellites, communication towers etc)

RFI in Interferometers

- Is RFI different from source signal?
  - **Fringe tracking** : Decorrelation of signals not moving with the phase center (sidereal rate)
  - RFI tends to decorrelate, particularly at longer baselines
  - Decorrelation reduces RFI power in the visibilities, however it can still be significantly stronger than any source of interest.

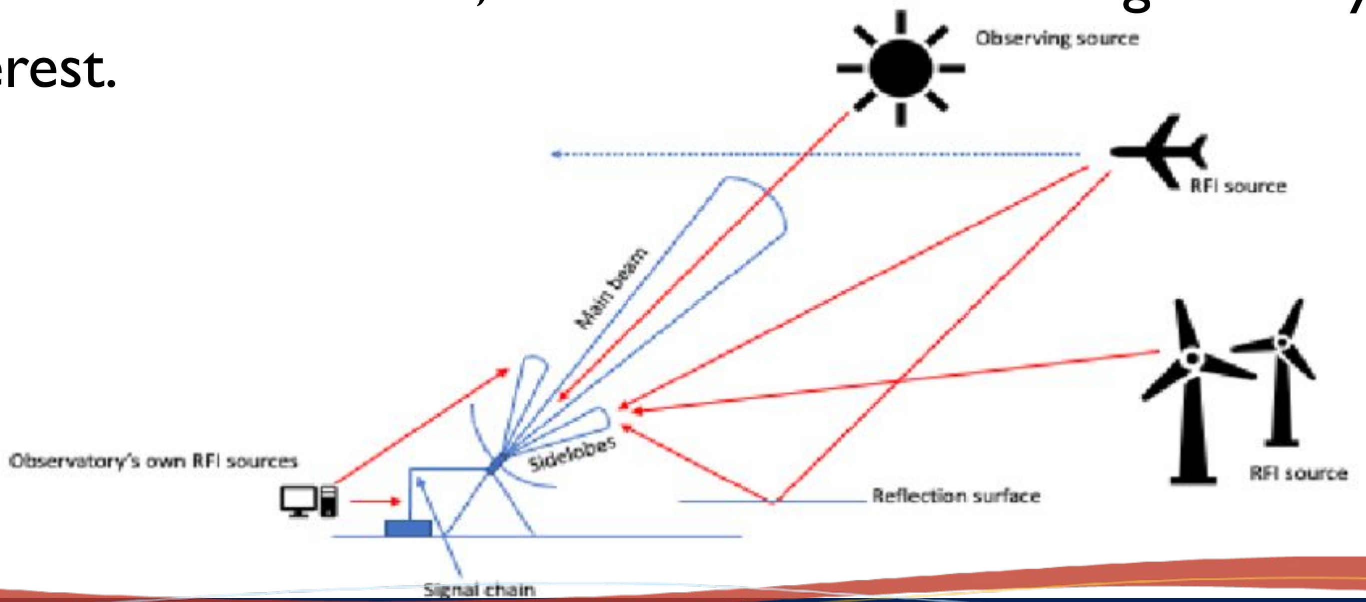
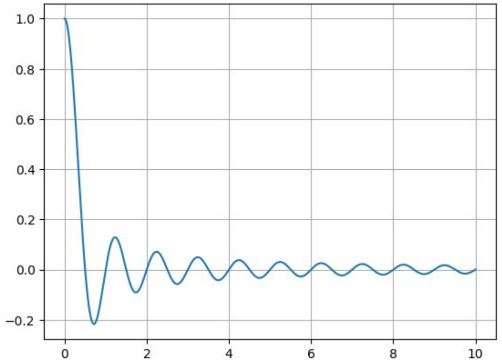


Fig : Kallunki et al. ,2022

Background : RFI Decorrelation in Interferometers

Thomson, Moran, Swenson
Sec. 16.3.2

- Simple case : Source of frequency is stationary relative to array (e.g., cell tower)
  - For full derivations please look at Chap 16 of TMS. These are just the highlights.
- The source then acquires a fringe frequency given by :



$$\nu_f = \omega_e u \cos\delta$$

- The amplitude of the interfering source (averaged over a time τ_a) is then given by :

$$f_1 = \frac{\sin(\pi \nu_f \tau_a)}{\pi \nu_f \tau_a}$$

- **Implication :** RFI tends to have highest power at small U. In general, longer baselines have lower correlated power

Background : RFI Decorrelation in Interferometers

Thomson, Moran, Swenson Sec. 16.3.2

- The amplitude of the interfering source (averaged over a time $\boldsymbol{\tau}_a$) is given by :

$$f_1 = \frac{\sin(\pi \nu_f \tau_a)}{\pi \nu_f \tau_a}$$

- To find correlated power (recast $\nu_f \tau = \psi \Delta u \cos\delta$)

$$\langle f_1^2 \rangle = \frac{2}{\pi} \int_0^{\pi/2} \frac{\sin^2(\pi \psi \Delta u \cos\delta)}{(\pi \psi \Delta u \cos\delta)^2} d\psi \simeq \frac{1}{\pi \Delta u \cos\delta}$$

- **Implication :** RFI tends to have highest power at small U. In general, longer baselines have lower correlated power

Background : Closure Phase

Thomson, Moran, Swenson
Sec 10.3

- For a single baseline, the measured visibilities can be written as $V_{mn}^{obs} = g_m g_n^* V_{mn}^{true}$

- The equivalent relationship for the visibility phases can be written as :

$$\phi_{mn} = \phi_m - \phi_n + \phi_{vmn}$$

- For a triangle of baselines between antennas m, n, and p the phase closure is given by :

$$\begin{aligned} \phi_{c_{mnp}} &= \phi_{mn} + \phi_{np} + \phi_{pm} \\ &= \phi_m - \phi_n + \phi_{vmn} \\ &\quad + \phi_n - \phi_p + \phi_{vnp} \\ &\quad + \phi_p - \phi_m + \phi_{vpm} \end{aligned}$$

Background : Closure Phase

Thomson, Moran, Swenson
Sec 10.3

- For a triangle of baselines between antennas m, n, and p the phase closure is given by :

$$
\begin{aligned}
\phi_{c_{mnp}} &= \phi_{mn} + \phi_{np} + \phi_{pm} \\
&= \phi_m - \phi_n + \phi_{vmn} \\
&\quad + \phi_n - \phi_p + \phi_{vnp} \\
&\quad + \phi_p - \phi_m + \phi_{vpm}
\end{aligned}
$$

$$\phi_{c_{mnp}} = \phi_{vmn} + \phi_{vnp} + \phi_{vpm}$$

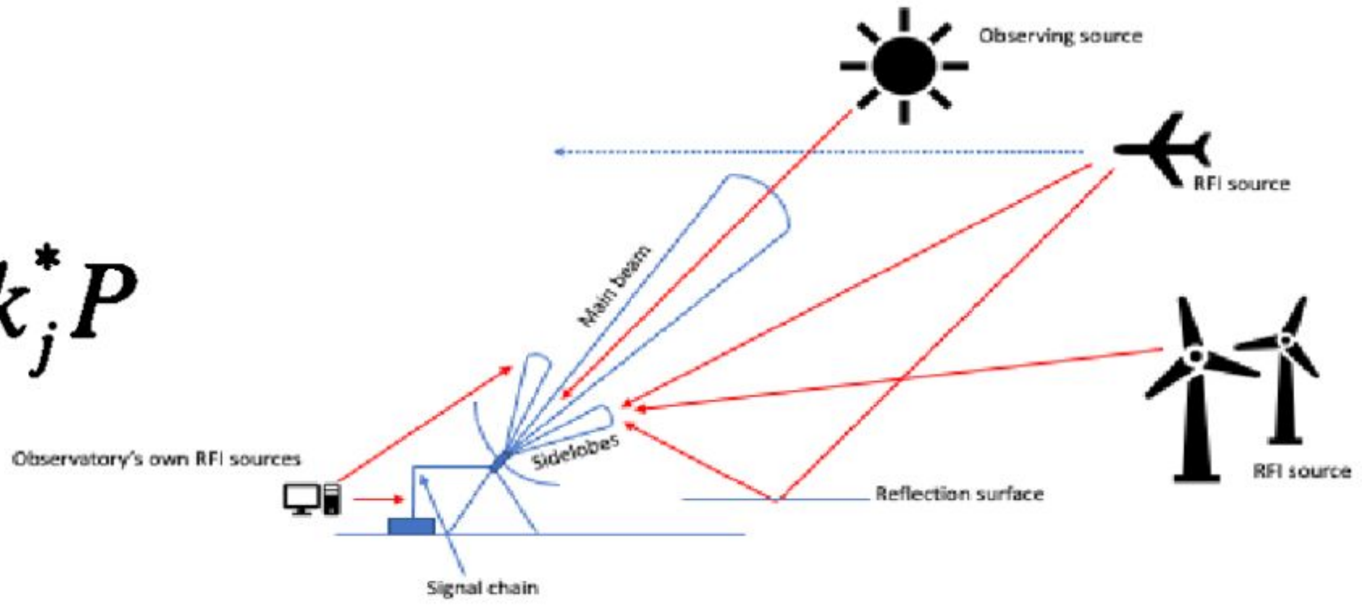
- Simplest case : A point source at phase centre; Visibility phase = 0. Therefore

$$\phi_{c_{mnp}} = 0$$

- In general : RHS a constant (knowable) value

RFI : Closure Errors

$$V_{ij}^{\mathrm{obs}} = g_i g_j^* V^{\mathrm{source}} + a_i a_j^* k_i k_j^* P$$



g : Antenna direction independent gain

a : Antenna primary beam

k : Propagation of interference to antenna

P : Power of interfering source

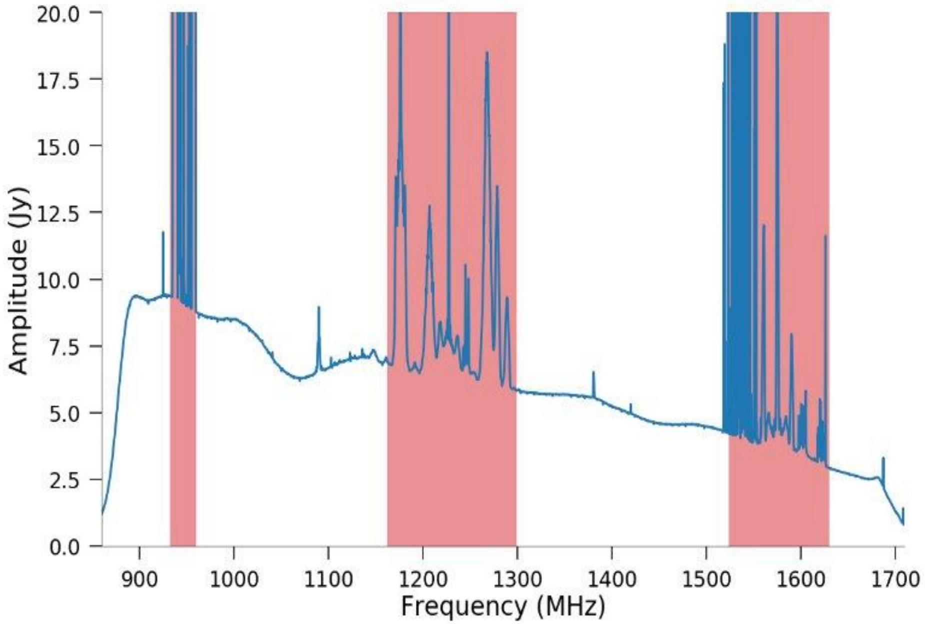
If interfering source **cannot** be cast in this manner, it will violate closure relationships.

(EVLA Memo 86)

RFI : Closure Errors

- Why do we worry about RFI?
  - Typically significantly stronger than our source signal (signals strong enough to saturate the receiver are not uncommon)
  - Can violate closure relationships : **baseline-based closure errors** (EVLA Memo 86)
  - Even on "successful" calibration, prevalent closure errors will result in poor image quality
  - Even if it does not violate closure, it will still dominate calibration solutions
  - Unless RFI is taken care of, further (self-) calibration will not help, limiting achievable sensitivity and dynamic range

RFI Detection



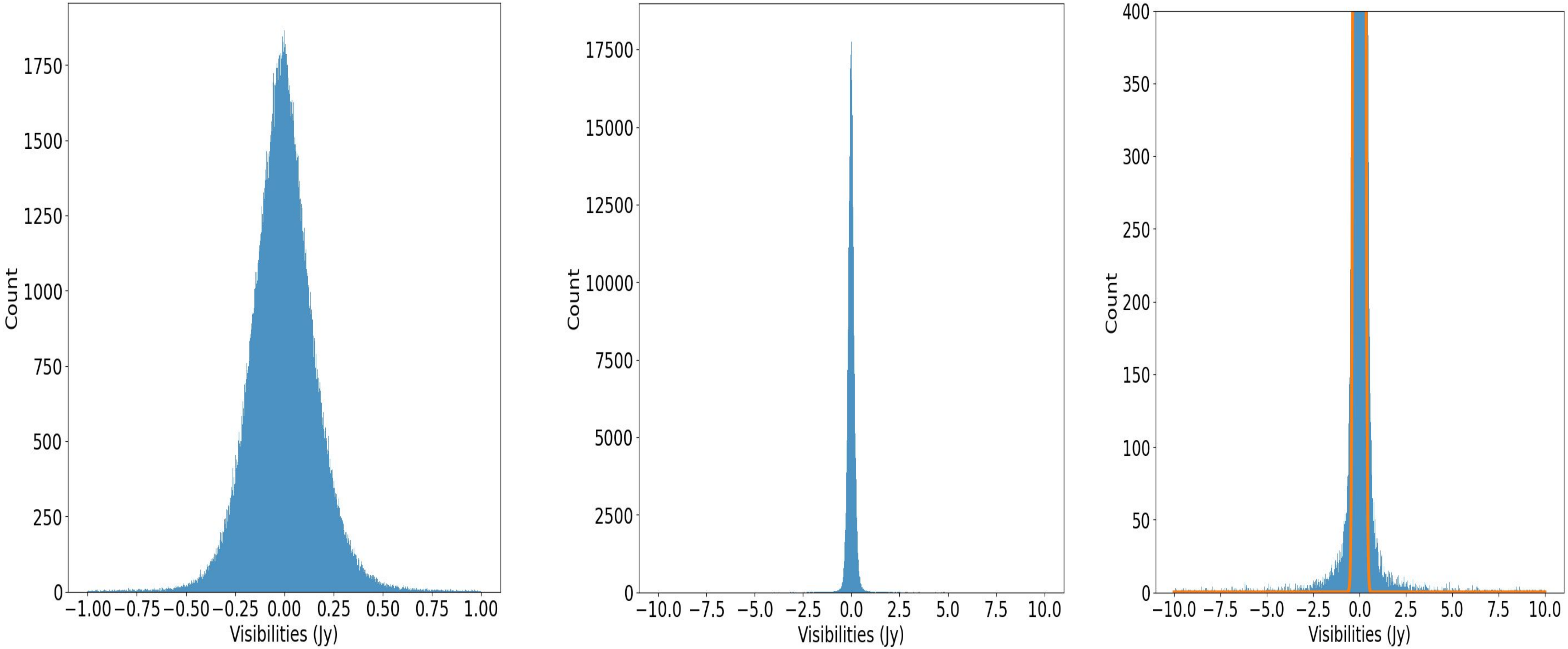
- Detecting strong RFI is "easy" - look for it, throw it away
- Faint RFI is **hard.** Amplitudes similar to source flux but **not stochastic.**
- So longer integrations will not help…
- Faint RFI tends to determine the noise floor even for very long observations if not appropriately taken care of

$$\Delta I_m = \frac{SEFD}{\eta_c \sqrt{n_{\mathrm{pol}} N(N-1) t_{\mathrm{int}} \Delta\nu}}$$

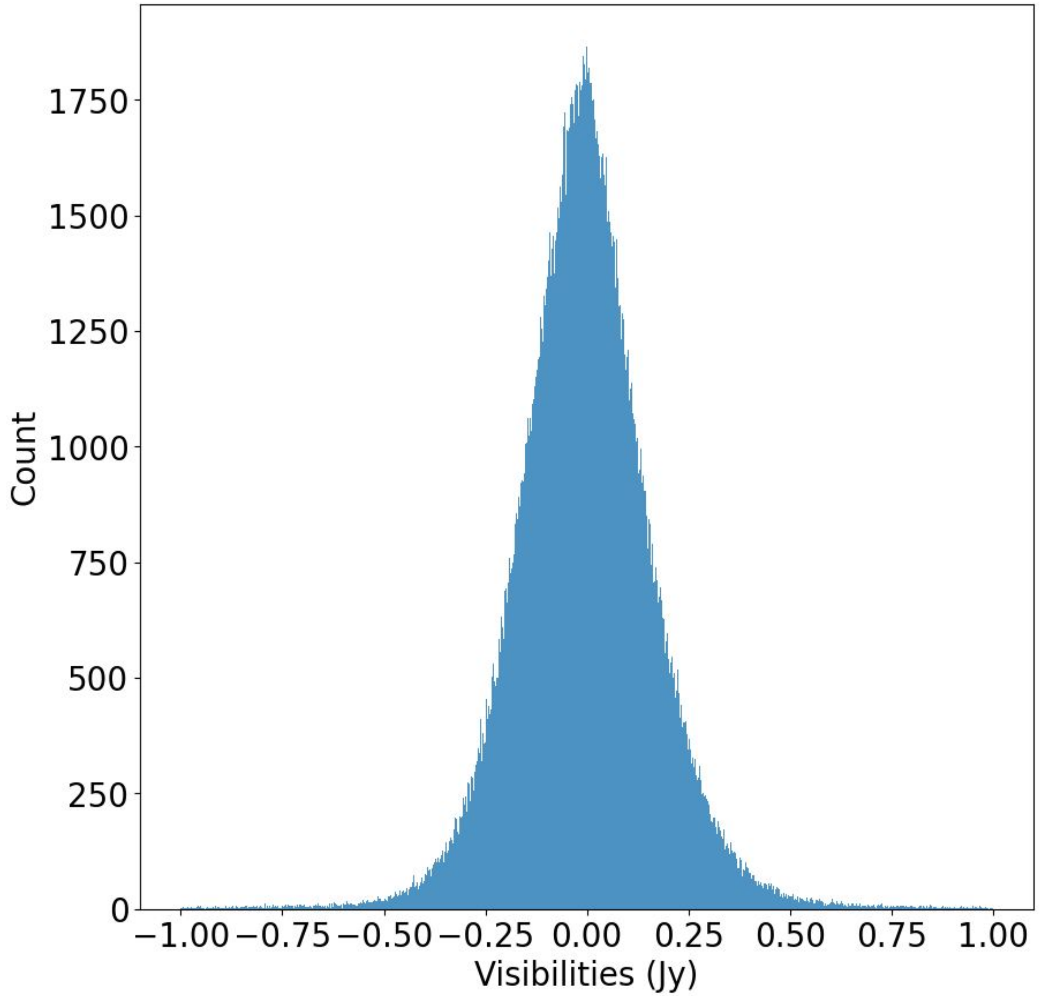
RFI vs Signal : Gaussianity

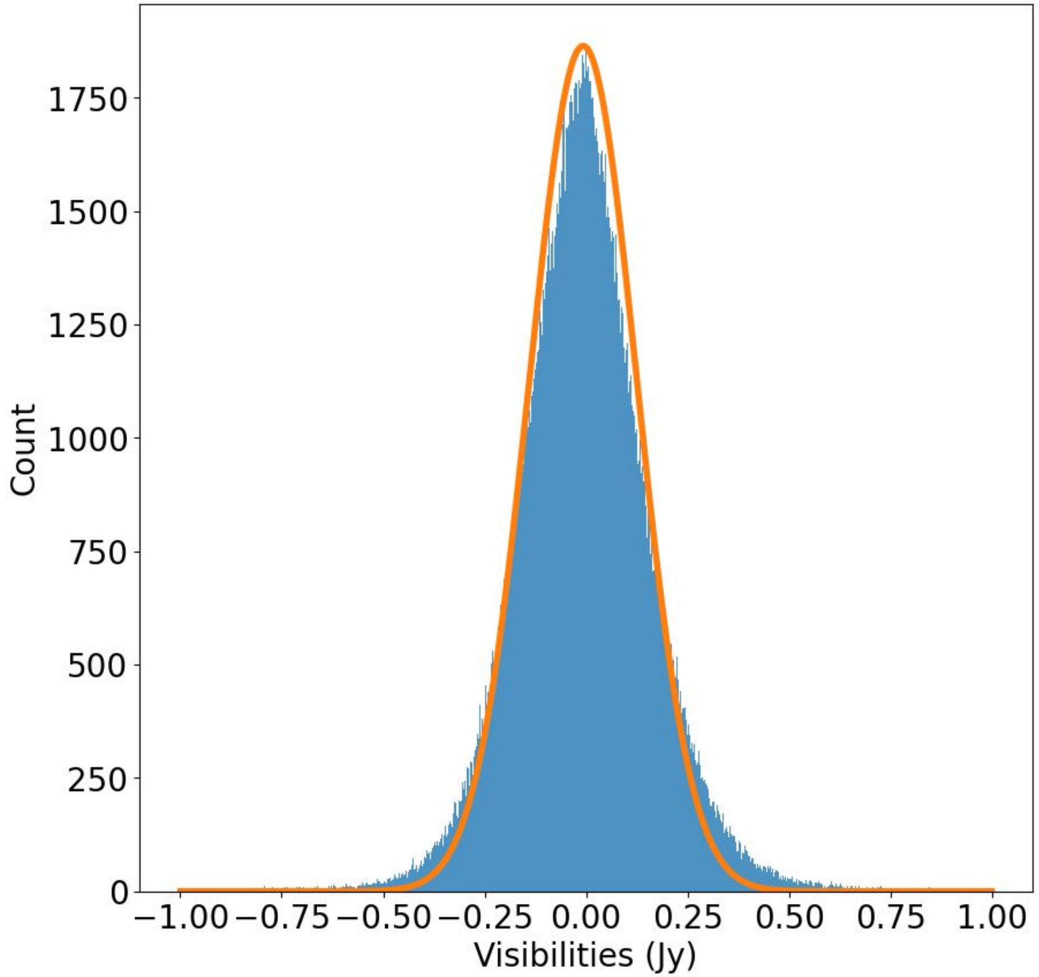
- Is RFI different from source signal?
  - Visibilities from astronomical sources follow a Gaussian distribution; The standard deviation of the Gaussian is determined by instrumental noise + systematics
  - Strong RFI : Contributes long tails in the Gaussian distribution ("excess Kurtosis")
  - Faint RFI : Increases measured standard deviation, but does not necessarily add to the Gaussian tail
- Strong RFI much easier to detect, statistically + visually

RFI vs Signal : Gaussianity



RFI vs Signal : Gaussianity

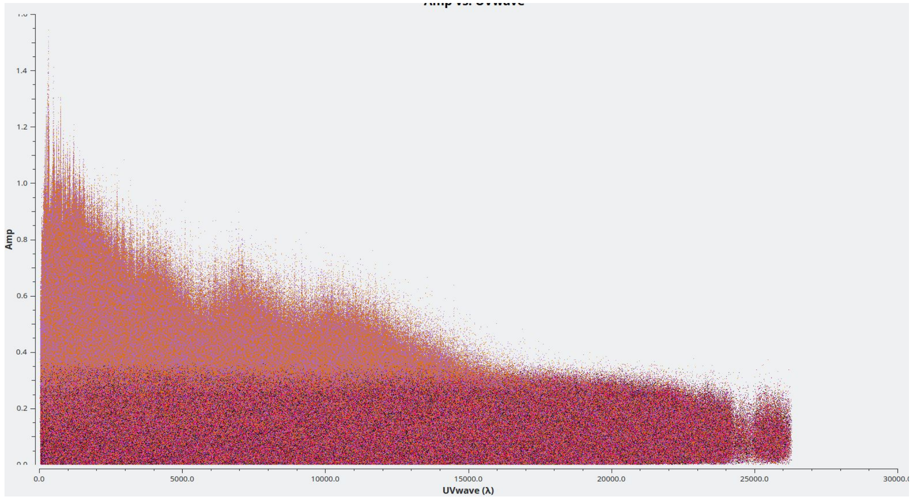




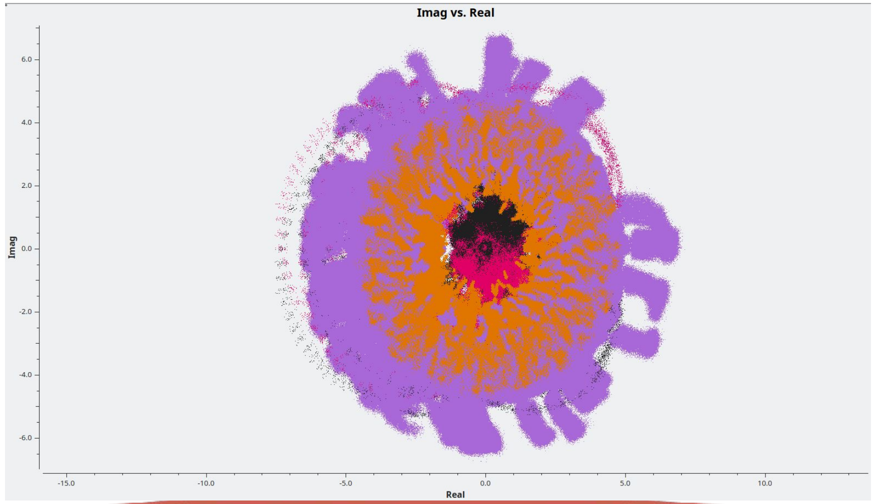
RFI Detection :Visualization

- Different kinds of RFI show up in different ways, so need to visualize data to get a good sense of what's going on
- Plotting data in multiple ways can show how much and what kind of RFI in the data, which will inform flagging strategy
- For example : Plotting amplitude vs. frequency can show which channels/SPWs are affected by RFI
  - Generating these plots for each antenna/baseline/SPW etc., can be instructive.
  - Identifying bad antennas/baselines/SPWs early can be beneficial, allowing automated flaggers to go after fainter RFI

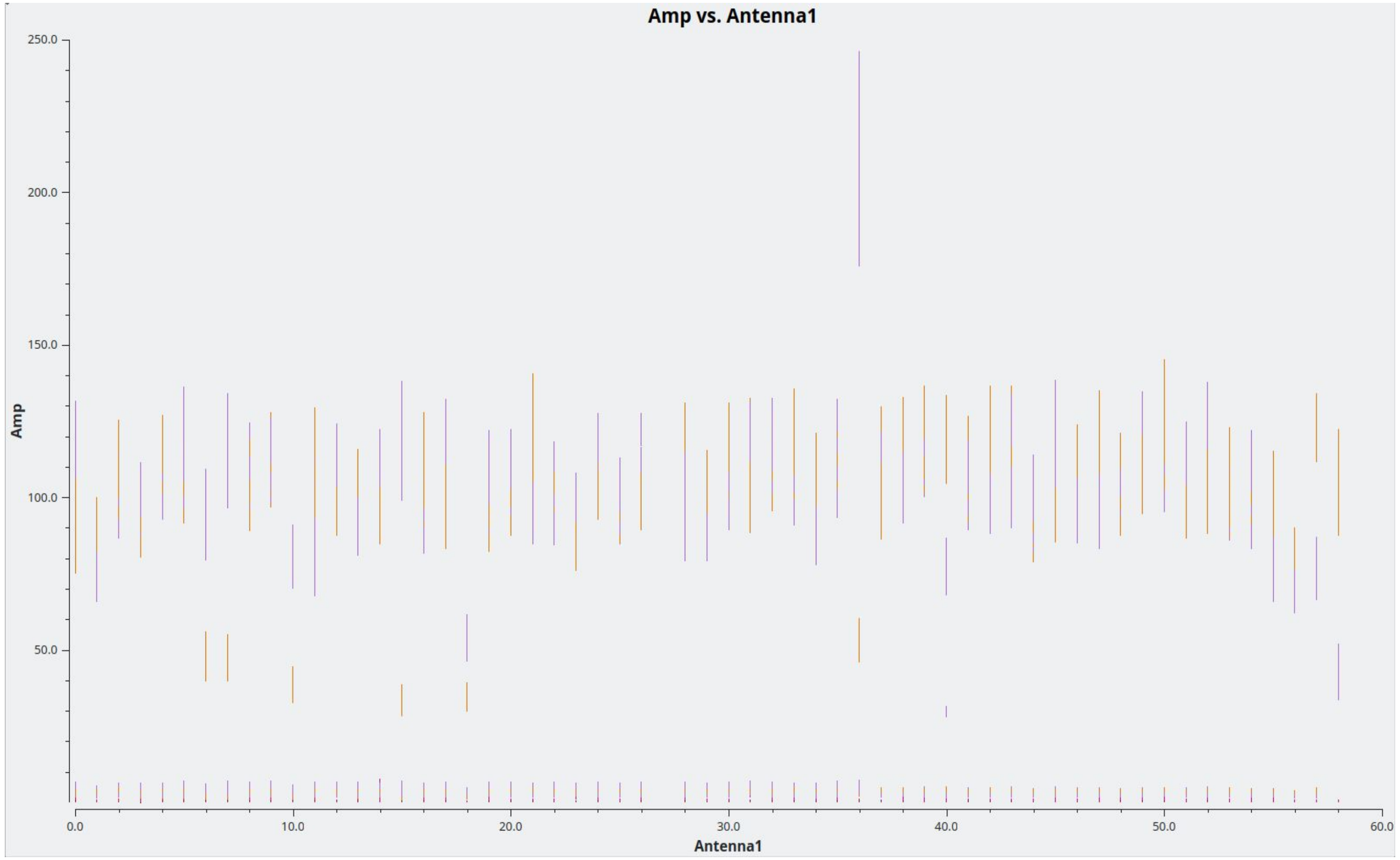
RFI Detection :Visualization



Amplitude vs. UV length



Real v s. Imag



Amplitude vs. Antenna number

RFI Detection :Visualization

- The specifics of the plots need to be informed by the physics of the problem to gain the most insight into data + RFI.
- For example :
  - Point source at phase center has 0 visibility phase. Plotting Re(Vis) vs Im(Vis) should then be a circle centred at (S, 0) — S is the source flux. Outliers on this circle are caused by RFI. Scatter on the circle is the visibility RMS. Doing this for a busy target field will not yield a very informative plot.
  - Shorter UV lengths sensitive to more flux, so simply looking for an increase in visibility flux at low UV lengths is insufficient to find RFI.

RFI Detection : Automation

- We don't want to scrub through ~ terabytes of data manually looking for RFI
- Visualizing data and looking for RFI manually can be useful :
  - When you need to get a sense of how much data is affected
  - Determine what kind of RFI is present in the data - this can inform your flagging strategy
- Flagging RFI manually is **not** recommended because
  - Subjective : Eyes are good at finding patterns, even when they aren't really there
  - Not reproducible : Drawing regions and flagging RFI can be arbitrary. It is better to identify the bad data (antenna, baseline, time, freq etc.) and flag using an automated flagger (e.g., CASA's flagdata)

RFI Detection : Automation

- Use algorithms to detect RFI : In pre-correlation voltage streams, in post-correlation visibilities etc.
- Depending on where in the signal chain, RFI can be subtracted/mitigated/flagged
- Pros :
  - Reproducible
  - Methods used to determine RFI are statistically/physically motivated
- Cons :
  - RFI that violates assumptions of algorithm will not be effectively caught
  - We need multiple algorithms & stages of RFI detection to be more effective
- Almost always more effective (and efficient!) than inspecting by eye

RFI Detection : Mitigation

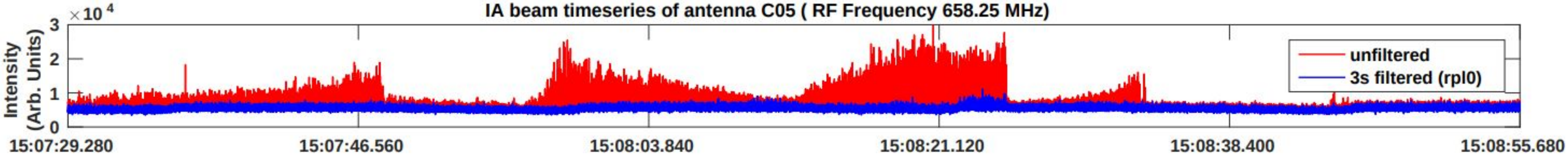
- “Prevention is better than cure” - True for RFI!
- Mitigation : Methods to prevent RFI from affecting your data, prior to storing the data
- RFI can enter data at various points in the signal chain : need multiple approaches to catch it at every stage for most effective removal.
- Some mitigation methods :
  - **Real time pre-correlation flagging of voltage data** (e.g., Buch et al., 2016; Svoboda et al., 2023)
  - Using a reference antenna(s) to measure and subtract RFI (e.g., Hellbourg et al., 2014; Briggs et al., 2000; Barnbaum & Bradley 1998)

RFI Detection : Mitigation

- Problem : Impulsive RFI can occur on very short timescales (~ few ms). Typical integration times are ~ few s. A single occurrence of RFI in the few second window can ruin entire integration.
  - A short burst of RFI in voltage will yield broadband RFI after the correlator
  - Worst case : 1ms impulsive RFI repeats every 1s. Sensitivity of entire observation is degraded because of ~ 0.1% of pre-integration samples
- Solution : Analyze input voltage stream (sampled at few ns to tens of us) and throw away outliers in the voltage stream pre-correlation using some threshold statistics (median absolute deviation, quartiles etc.)

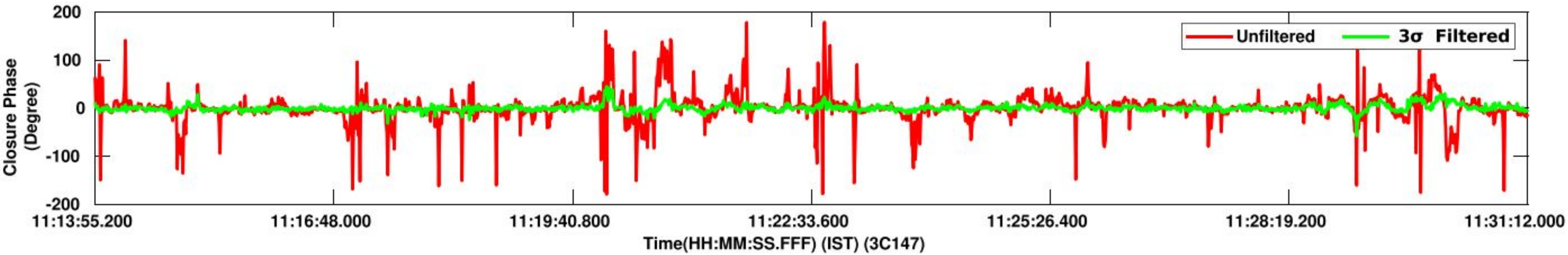
RFI Detection : Mitigation

Buch et al., 2016



Top : Amplitude time series

Bottom : Closure phase time series (different data)



RFI Detection : Excision

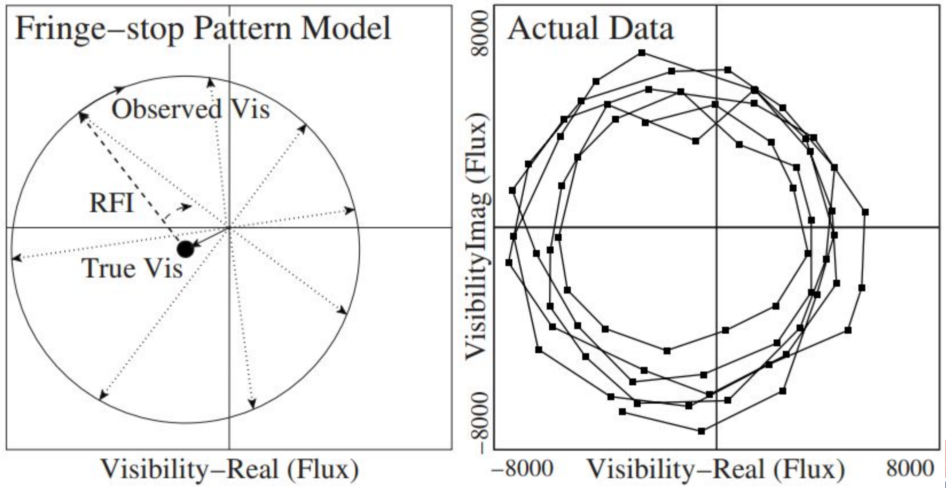
- Excision : Recover underlying visibilities affected by RFI
- Typically "excision" is post-correlation + on the recorded visibilities, whereas "mitigation" is (quasi-) real time.
- Some excision methods :
  - RFI self-partitioning (EVLA Memo 86)
  - **Subtraction of stationary RFI from visibilities** (Athreya 2009; EVLA Memo 146)

RFI Detection : Excision

Athreya 2009
EVLA Memo 146
"CIRC" in task UVRFI in AIPS

- Subtracting (unknown) stationary RFI from correlated visibilities hinges on :
  - A stationary source of RFI will acquire the fringe stop phase $\nu_f = \omega_e u \cos\delta$
  - If sampling is high enough to prevent decorrelation, this causes visibilities to trace a circle in the Re-Im visibility plane as a function of time. Radius of the circle is the correlated power of the RFI
  - Fit for the circle and subtract, recovering original visibilities
  - Re-Im visibility plane ≠ UV plane! Re-Im visibility plane is simply plotting the real vs imag component of the visibilities

Athreya 2009



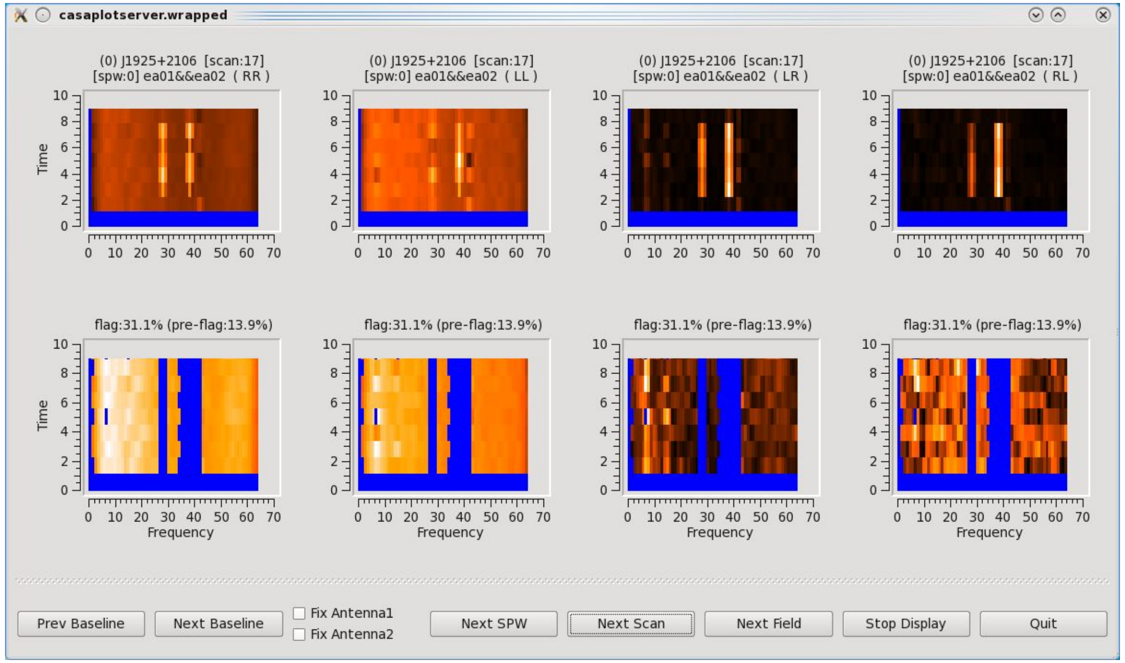
RFI Detection : Flagging

- Remove (“flag”) RFI affected data from data set.
- No “one size fits all” solution. Given the complexity of RFI, typically need multiple algorithms.
- Each algorithm comes with it’s own set of assumptions about RFI behaviour : constant in time, constant in frequency, narrow-band, broad-band etc.
- Therefore multiple algorithms sensitive to different kinds of RFI are likely necessary if your observation is really badly affected.
- Broadly in two categories :
  - Baseline based : tfcrop (Rau 2002), SumThreshold (Offringa et al., 2010), rflag (Greisen 2011) etc.
  - UV plane based : GRIDflag (Sekhar et al., 2018), msuvbinflag (EVLA Memo 198)

RFI Detection : Flagging

- Baseline based flaggers : Look for RFI in the time-frequency plane (per-baseline).
- rflag (operates per baseline) :
  - Calculate local RMS (MAD) within sliding window
  - Calculate median RMS and deviations from median RMS across all sliding windows
  - If local RMS > sigma * (medianRMS + medianDev) flag points within window

VLA flagging CASA guide



RFI Detection : Flagging

- Most baseline based flaggers operate using a similar recipe :
  - Measure smooth background variation
  - Subtract/divide background
  - Identify outliers in this "flat-background" space
  - Use a statistic to identify outliers
  - Flag outliers in original data
- The devil is in the details – Certain assumptions might suit your instrument + observation better than others.

RFI Detection : Flagging

- Background estimation :
  - Fit a 2D "rubber sheet" model – AOFlagger
  - Fit a 2D piecewise polynomial – tfcrop
  - Sliding window median statistic – rflag
- Outlier estimation :
  - Calculate robust mean + std (or median + MAD)
  - Higher order statistics, kurtosis
- **No "one size fits all"** – think about your problem, kinds of RFI you might face, and which algorithm might suit it best.
- Some statistics might work better for certain kinds of RFI
  - e.g., median filters work very well for narrow-band RFI, but less well for broad-band

RFI Detection : Flagging

- UV plane based flaggers : Construct gridded UV plane from visibilities, look for outliers in local neighbourhood
- Details here are as malleable as baseline-flaggers :
  - Accumulate statistics across entire UV grid? A local neighbourhood? A single pixel?
  - Flag all points within a pixel? Just a subset?

- Notice : We expect more RFI at small U, and that is what we see. The math works!

$$f_1 = \frac{\sin(\pi \nu_f \tau_a)}{\pi \nu_f \tau_a}$$

$$\nu_f = \omega_e u \cos\delta$$

RFI Detection : Flagging

- Implementing your own flagger using the CASA toolkit for data I/O + Python for the statistics is **very achievable**!
- Sometimes "off-the-shelf" software will not work for your niche use-case. Don't be afraid to get your hands dirty and try different things.
- If you can calculate statistics on NumPy arrays, you can write a custom flagger.

See also : RFI Flagging Tutorial

Summary

- The big picture here is - RFI will show up in every possible way in your data. The specifics of the RFI and your instrument dictate your flagging strategy.
- Understanding a little about how these different flagging algorithms work can provide you a great place to start, since the "ready-to-eat" recipes you find online may not always work for you.

Resources:

- Interferometry and Synthesis in Radio Astronomy - Thomson, Moran, and Swenson 2017
  - Open access book available via Springer : https://link.springer.com/book/10.1007/978-3-319-44431-4