National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Machine Learning for Transient Detection with Radio Arrays

David R. Thompson[1], Kiri L. Wagstaff[1], Randall Wayth[2], Adam Deller[3], and Steven Tingay[2]
1: Jet Propulsion Laboratory, California Institute of Technology;
2: ICRAR/Curtin University of Technology;
3: National Radio Astronomy Observatory
Contact: `david.r.thompson@jpl.nasa.gov`

Summary: We investigate the use of machine learning and data analysis methods in support of radio astronomy investigations. In this work, we focus on a data-driven approach to detecting anomalous transient pulses, by modeling actual observations and estimating how unusual a new observation is. For radio arrays such as the VLBA, this approach allows us to distinguish remote radio sources (which appear in all antennas) from local radio frequency interference (RFI, which appears in only a single antenna). Ultimately, these techniques could be useful for other arrays with long baselines, such as the Square Kilometer Array (SKA).

Very Long Baseline Array (VLBA)
10 radio antennas

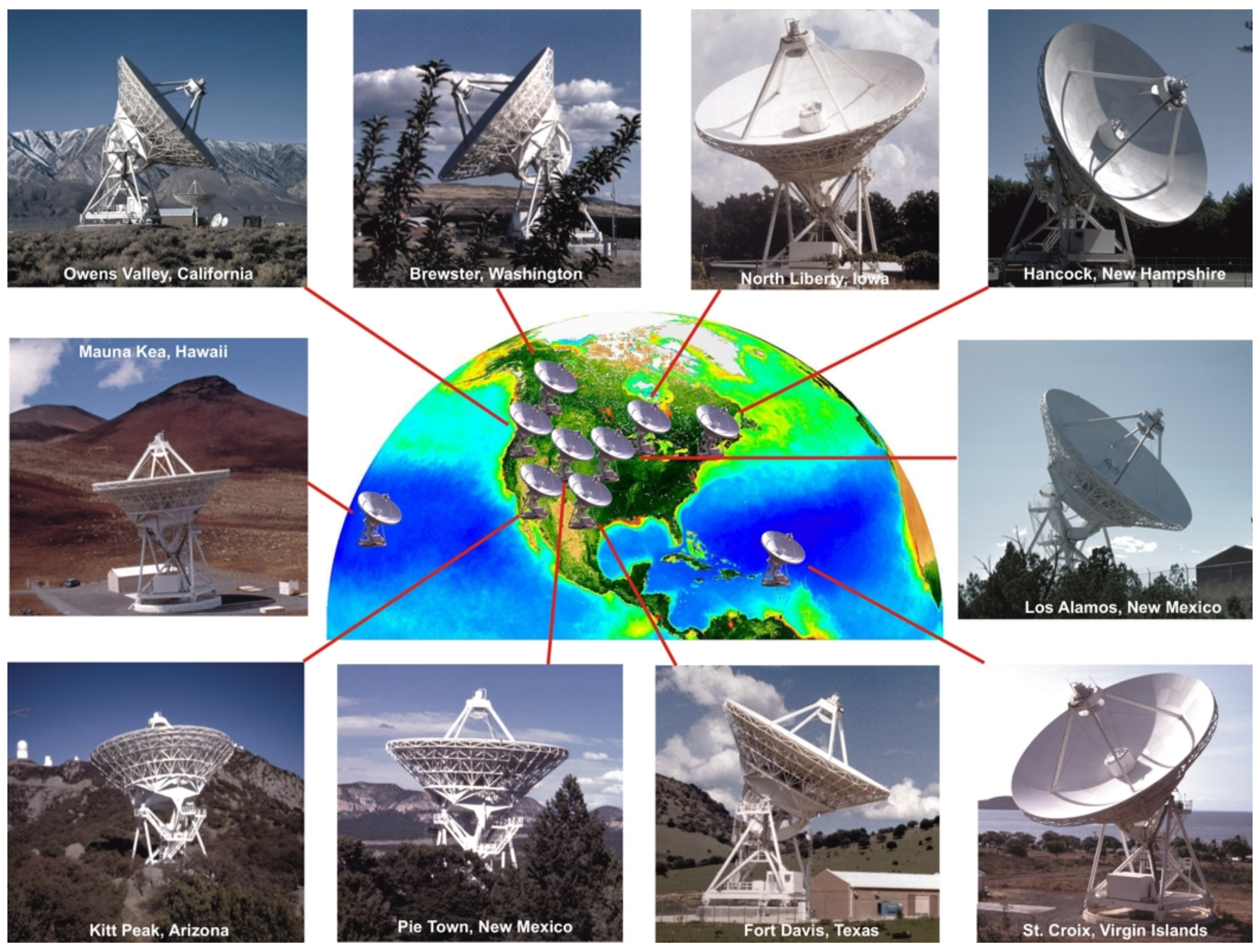
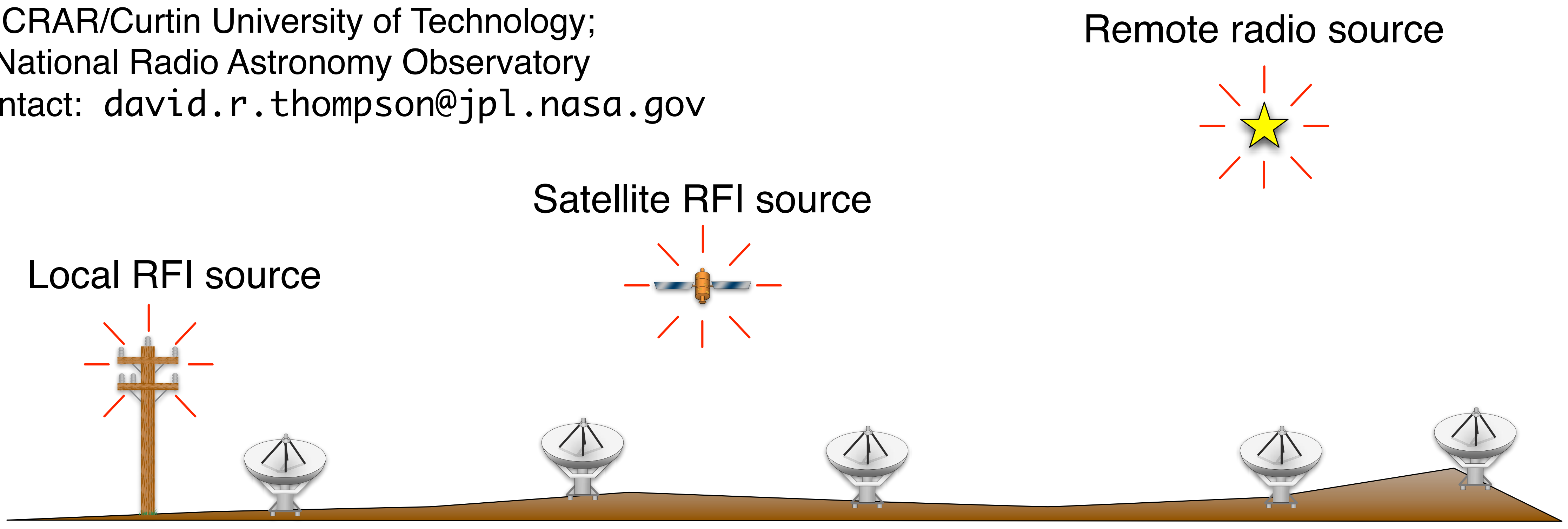


Image courtesy of NRAO/AUI and Earth image courtesy of the SeaWiFS Project NASA/GSFC and ORBIMAGE.



Remote radio source

Satellite RFI source

Local RFI source

Goal: Detect remote transient radio sources, while excluding local radio-frequency interference (RFI).

Solutions

Key insight: RFI manifests locally (at a single antenna), while remote radio sources are observed by all antennas.
Compute a detection statistic and if it exceeds θ, flag a detection.

$$\begin{aligned} A &= \text{Number of streams (antennas)} \\ DM &= \text{Dispersion measure} \\ S_a(DM) &= \text{Strength of response for stream } a \text{ de-dispersed using DM } DM \end{aligned}$$

Note: equations assume antenna streams are sorted in ascending strength.

State of the art:
Average all streams (incoherent addition) and take the max over all DMs.

$$\boxed{\max_{DM} \frac{1}{A} \sum_{a=1}^{A} S_a(DM)}$$

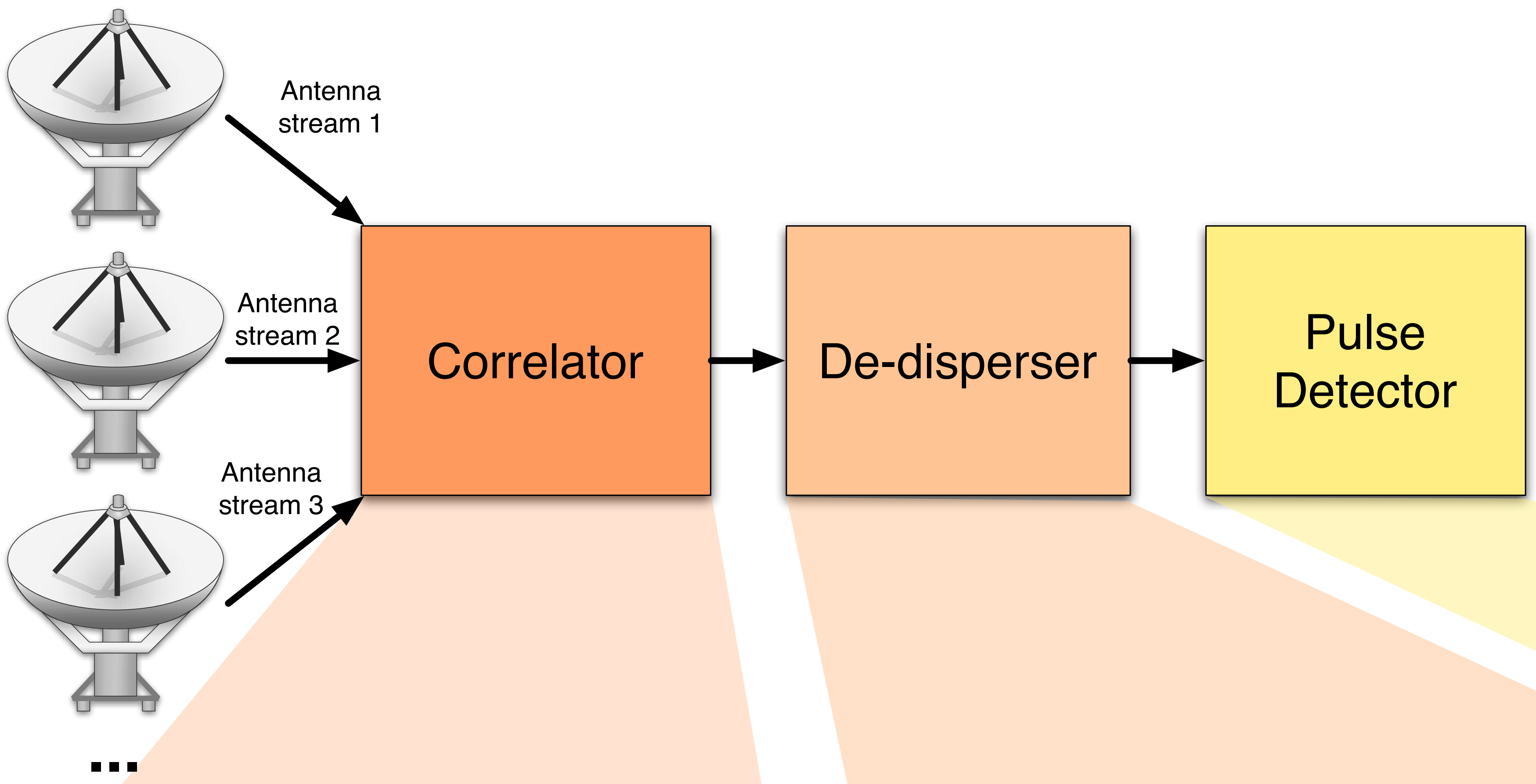
1) Trimmed (Robust) Estimator:
Exclude max stream, average the rest, and take the max over all DMs.

$$\boxed{\max_{DM} \frac{1}{A} \sum_{a=1}^{A-1} S_a(DM)}$$

2) Ensemble CDF Estimate:
Model prior observations, per stream, and compute the probability that the new observation is stronger than any randomly drawn X from all observations, then take the max over all DMs.

$$\boxed{\max_{DM} \frac{1}{A} \sum_{a=1}^{A} P(X \leq S_a(DM))}$$

VLBA Data Processing Path



Antenna stream 1

Antenna stream 2

Antenna stream 3

...

Correlator

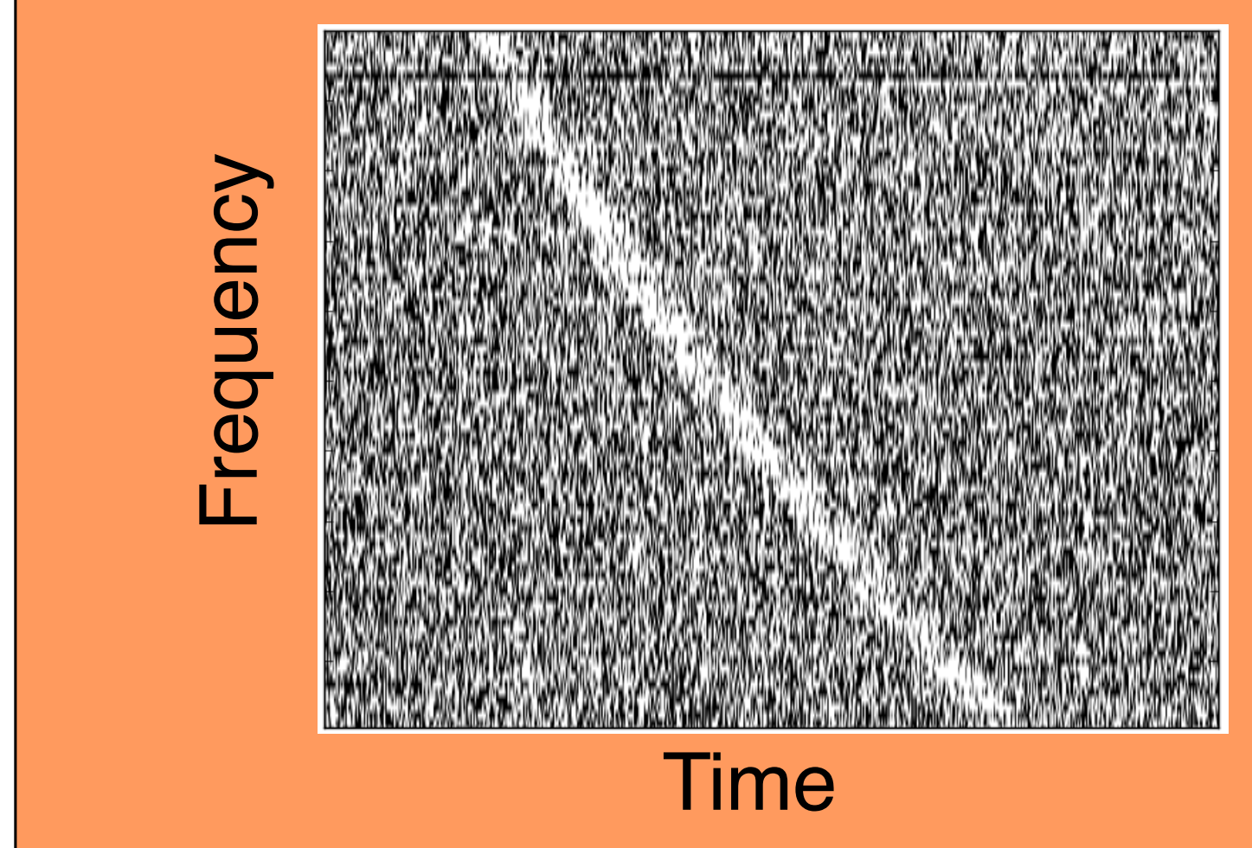
De-disperser

Pulse Detector

Correlator Output

The correlator computes all pairwise correlations. We use only the auto-correlations of each stream with itself, yielding the radio-frequency power output of each antenna as a function of time and frequency.

Example result for one antenna:
Pulse is dispersed in time
due to interstellar medium.

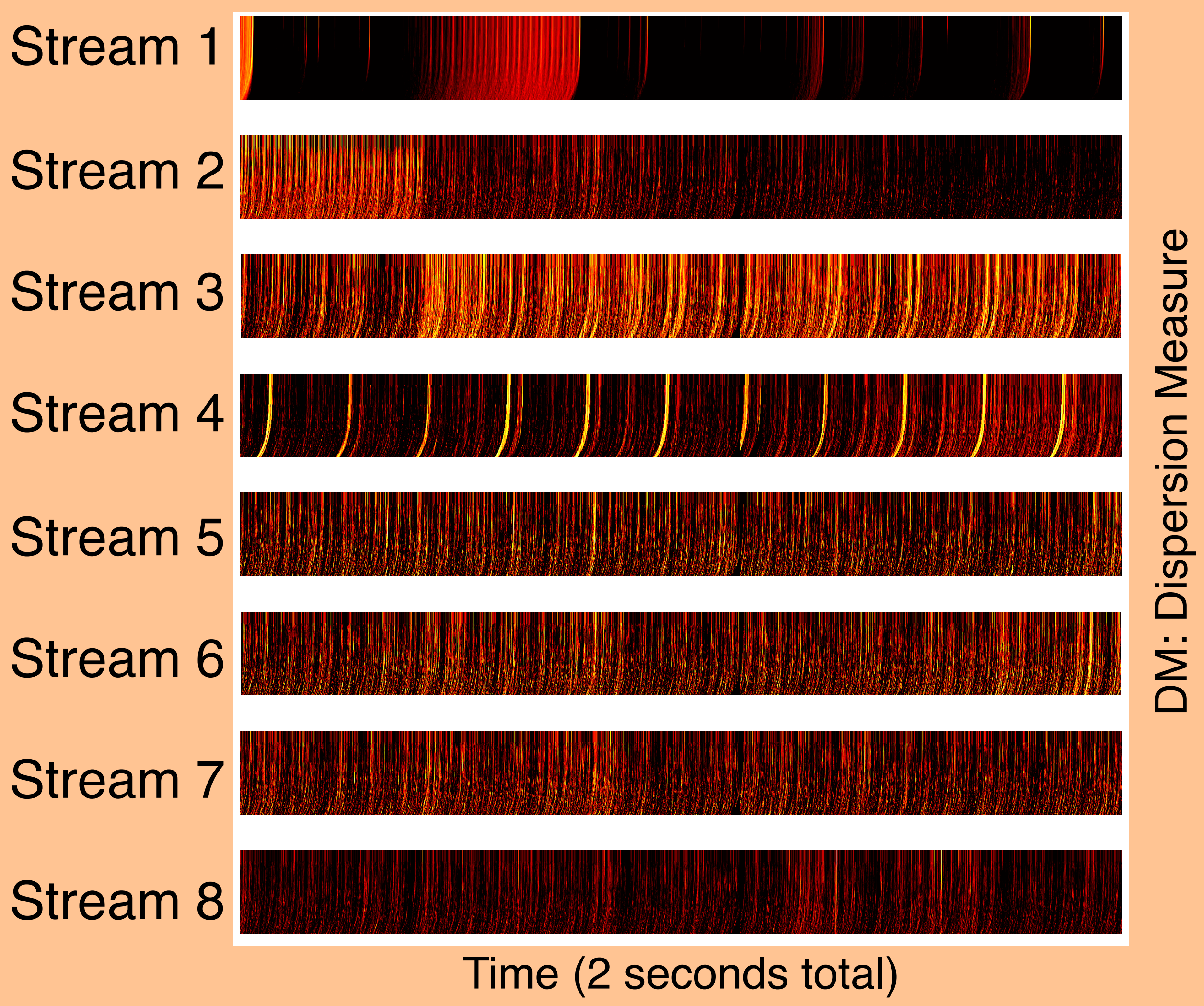


Example data from Parkes Observatory
(Edwards et al., 2001)

De-dispersion Output

De-dispersion reverses cosmic dispersion effects via an approximation to a matched filter, for a specified DM (dispersion measure). Since the true DM is not known, we search over many DMs, yielding a time-DM data set for each stream.

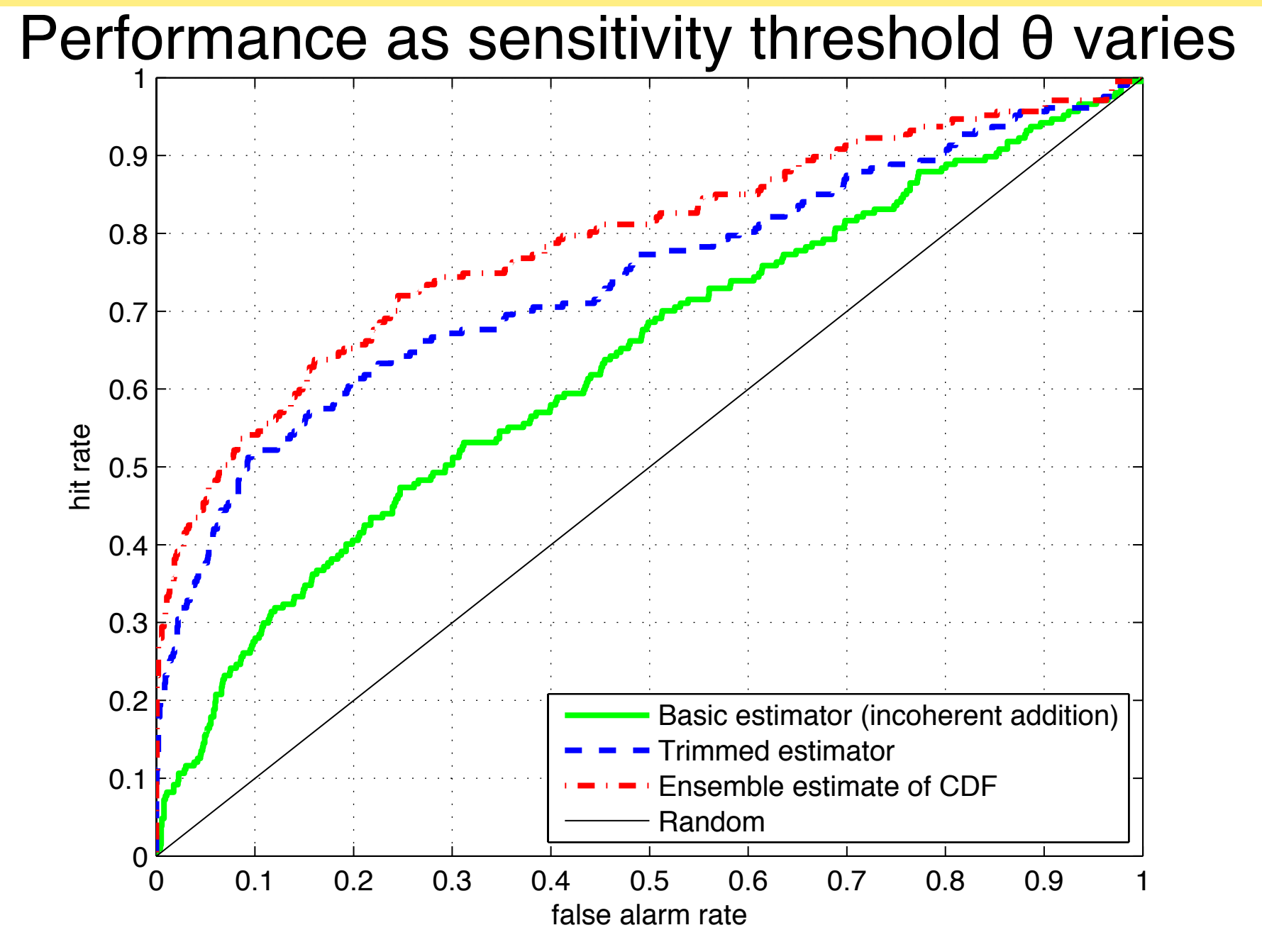
Intensity of de-dispersed data, DM 0 to 400



Example VLBA data, from 8 stations

Pulse Detector Results

We injected 400 synthetic pulses at random intervals in ~160 seconds of otherwise raw VLBA data (using 8 antennas). This data includes typical VLBA noise, RFI, and other confounding effects. Each synthetic pulse had a randomly chosen SNR from 5 to 10 and a true DM from 0 to 400. We then split the data into 0.1-second windows and analyzed each one with the three methods above.



Conclusions: The best performer was the ensemble CDF estimator, which uses a data-driven model of signal strength to detect pulses. It achieved a higher detection rate with fewer false detections than the standard incoherent summing approach or a basic robust estimator.

We gratefully acknowledge Walid Majid, Walter Brisken, Dayton Jones, Bob Preston, and the entire CRAFT (Commensal Real-time ASKAP Fast Transients) team.
This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.