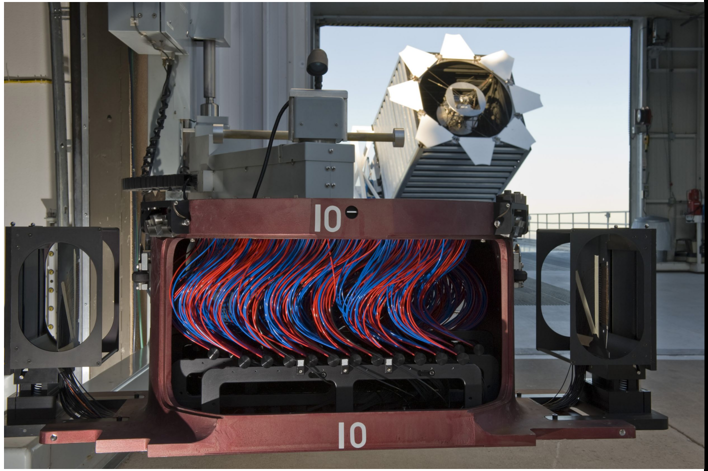
Real-time data analysis challenges for Stage II 21cm

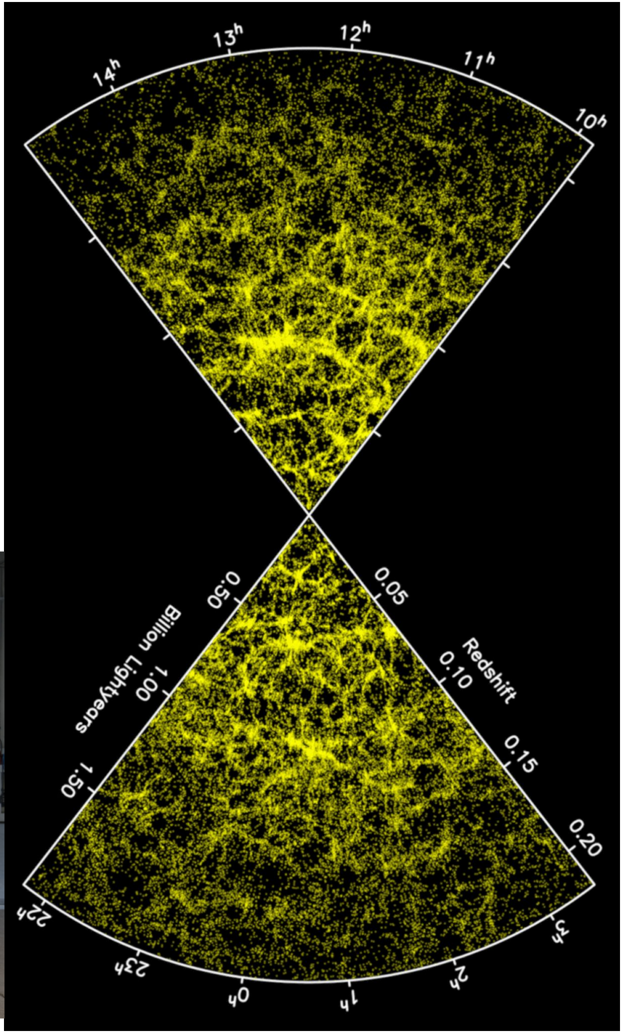
Anže Slosar, BNL

January 2019

Goal: measure structure in the universe

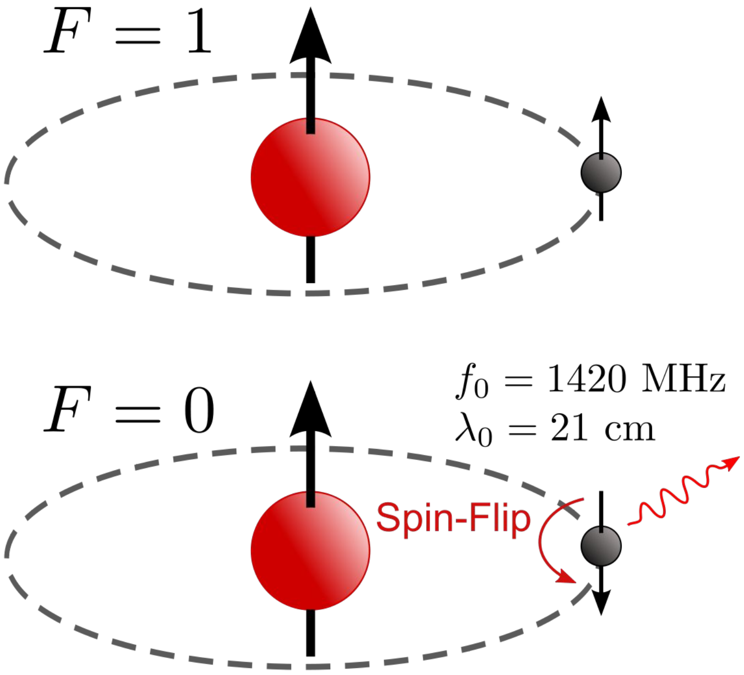
- If we measure structure in the universe, we can learn how it works.
- Traditionally this is done in optical using spectroscopy to take redshift of each individual galaxy
- This is very cumbersome





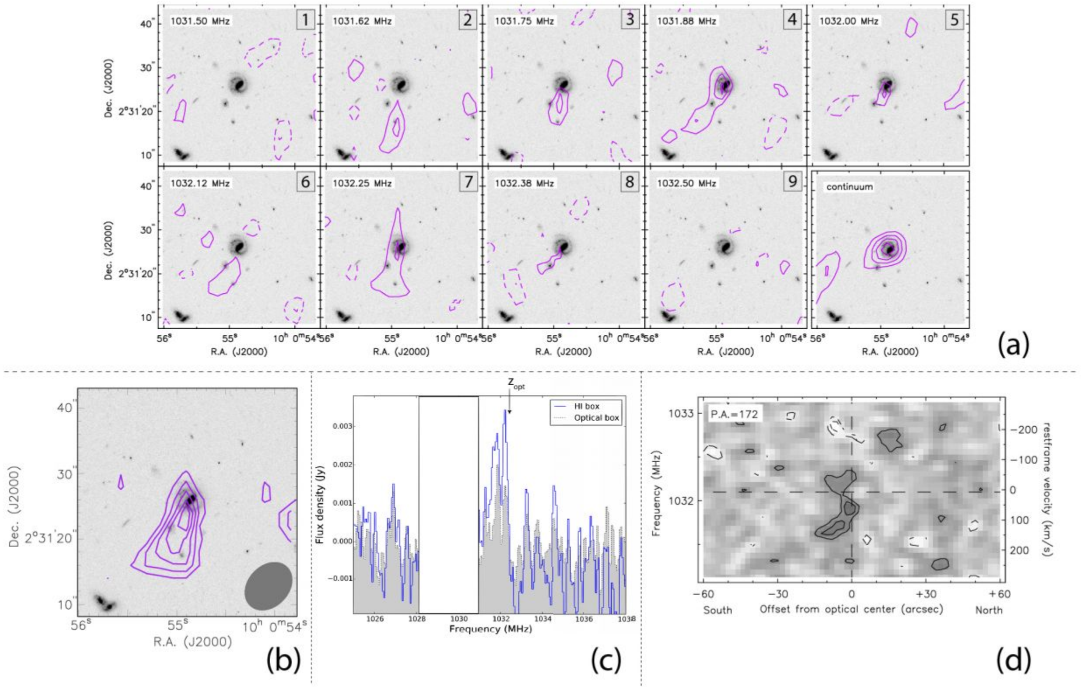
21cm emission

- Hyperfine transition in neutral hydrogen at ν=1420MHz, λ=21.1cm;
- This is the **only** transition around -- if you see a line at 710MHz, it is a z=1 galaxy;
- (not true in optical)
- Universe is mostly hydrogen (75%), but at low redshift we are sensitive to **pockets of neutral hydrogen in galaxies**;
- 21cm surveys are galaxy surveys in radio



Galaxies in 21cm

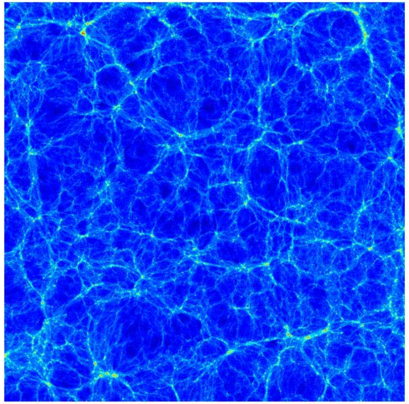
This is a weak transition, 21-cm detection redshift record is z=0.376 using 178 hours of VLA data (Fernández et al, 2016)

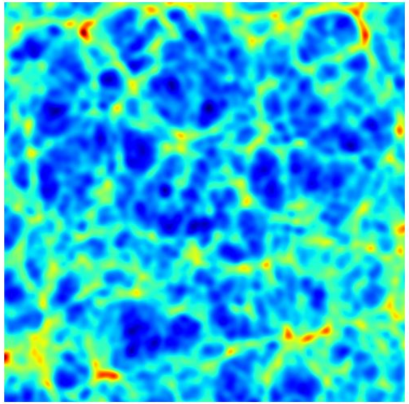


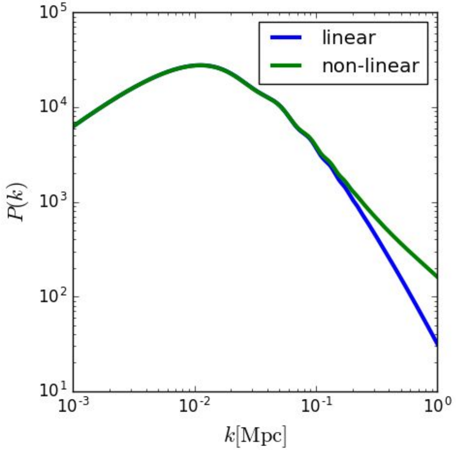
21cm Intensity mapping

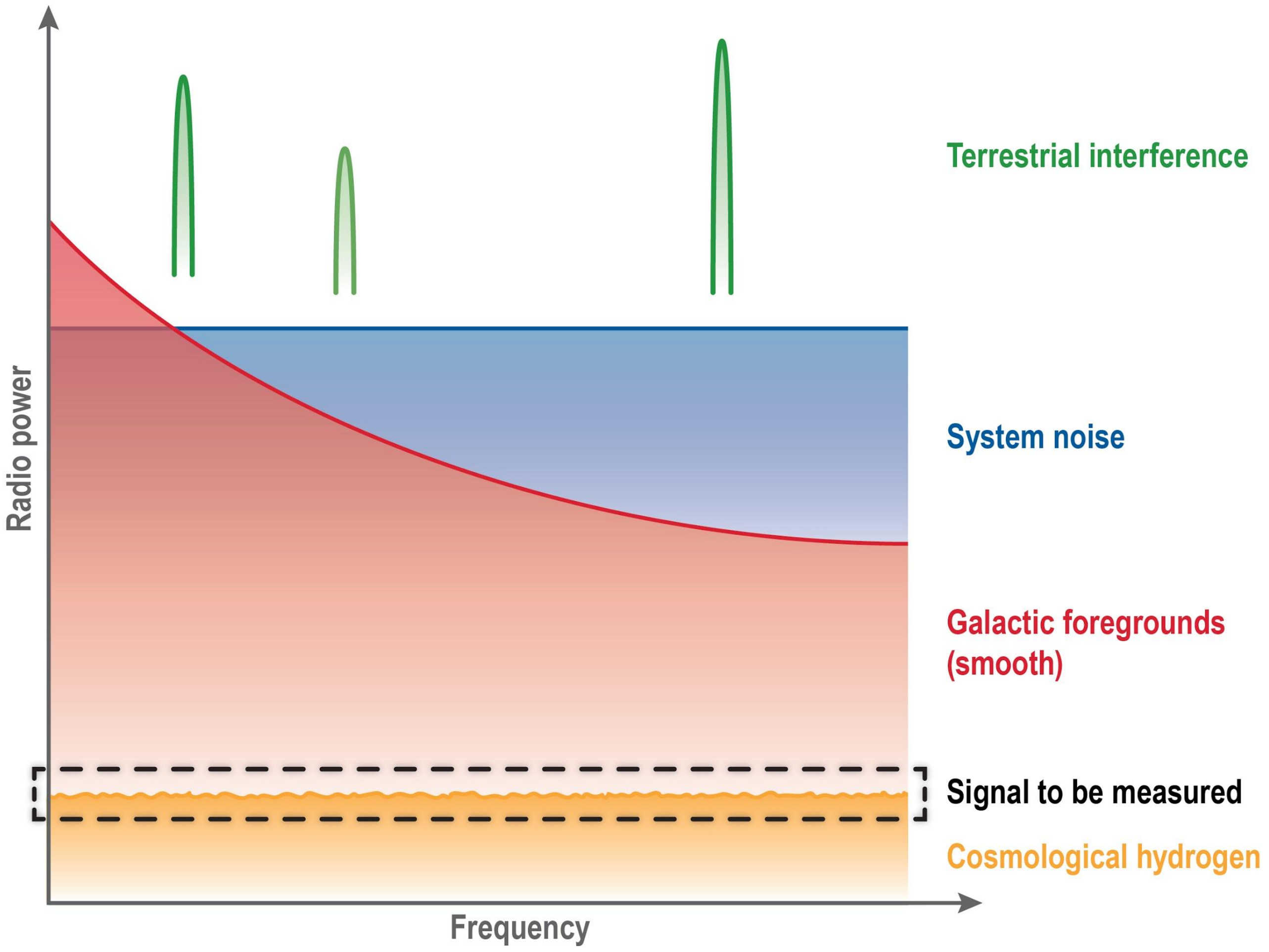
The main idea is to give up on resolving individual galaxies:

- For scales much bigger than individual galaxies, the overall signal will still trace the underlying number density of galaxies
- Put SNR where you really need it -- linear large scale modes
- Signal for galaxies is the only component that is not smooth in frequency



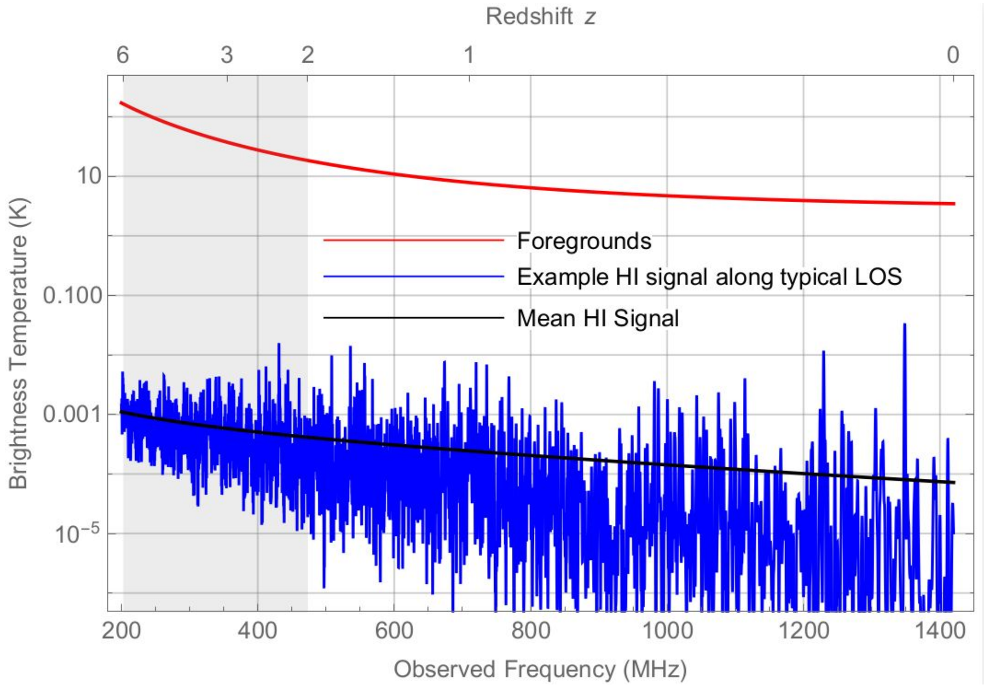






Terrestrial interference
System noise
Galactic foregrounds
(smooth)
Signal to be measured
Cosmological hydrogen
Radio power
Frequency

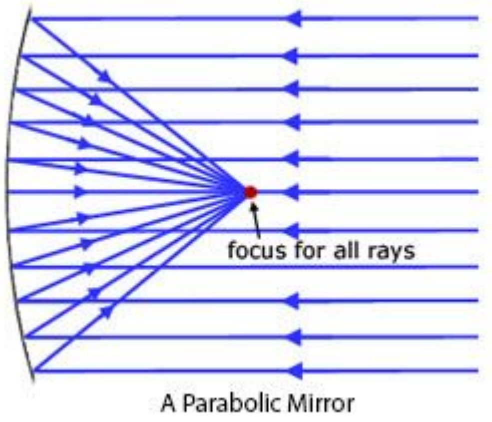
But 21cm is not the only radio signal…

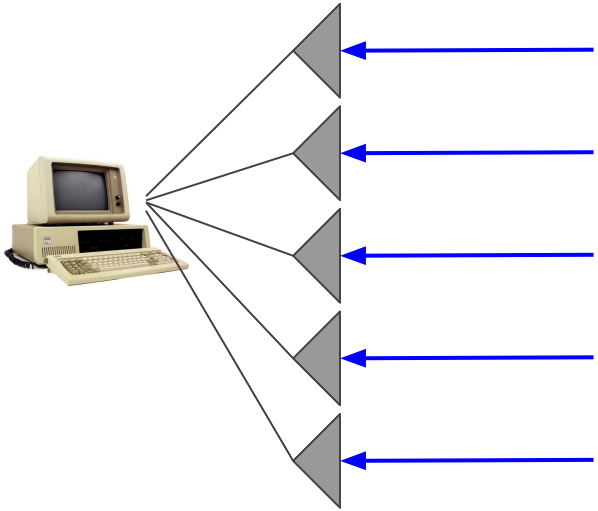


- Signal is subdominant, but the only non-smooth component.
- Of course, instrument can have non-smooth, time-varying response too!

Radio Interferometry

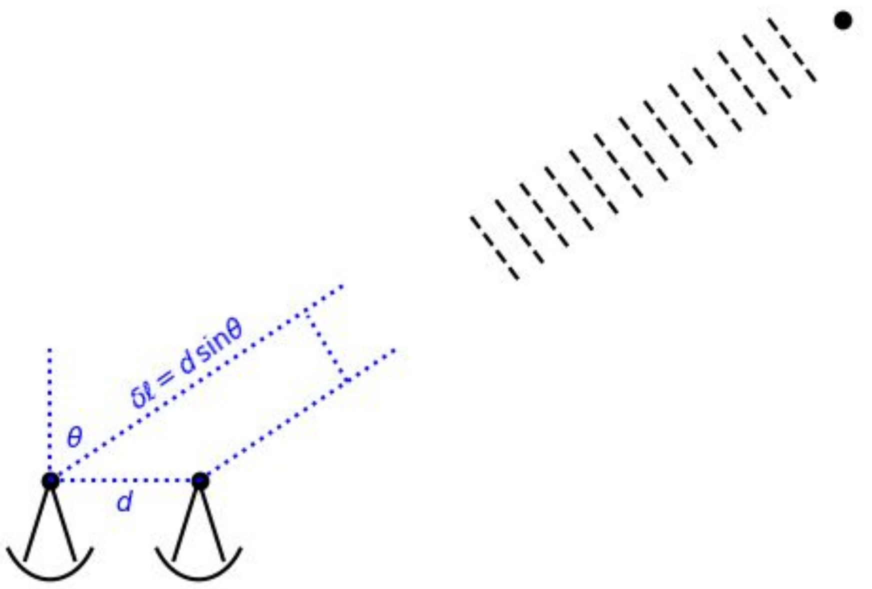
- If you want radio astronomy efficiently, you do interferometry.
- In effect you replace combining the signal in optics with combining them in a electronics.
- Initially, this was to gain resolution
- For us, it is to gain sensitivity and sky coverage
- Massive increases in compute power and telecom advances allow you to do everything digitally.





Interferometer basics

- Each baseline measures a "visibility", V_{ij}.
- Visibility is a Fourier component of the image
- Interferometer measures the sky image directly in the Fourier space
- This aspect leads to both advantages and disadvantages



$$V_{ij} = \langle E_i E_j^\star \rangle_{\mathrm{TIME}}$$

$$= \sum_{\mathrm{SRC}} E^2 \exp\left[2\pi i \hat{n}\cdot d_{ij}/\lambda\right] \mathrm{BEAM}$$

$$= \int d\hat{n}\, I(\hat{n}) B(\hat{n}) e^{2\pi i \hat{n}\cdot u_{ij}}$$

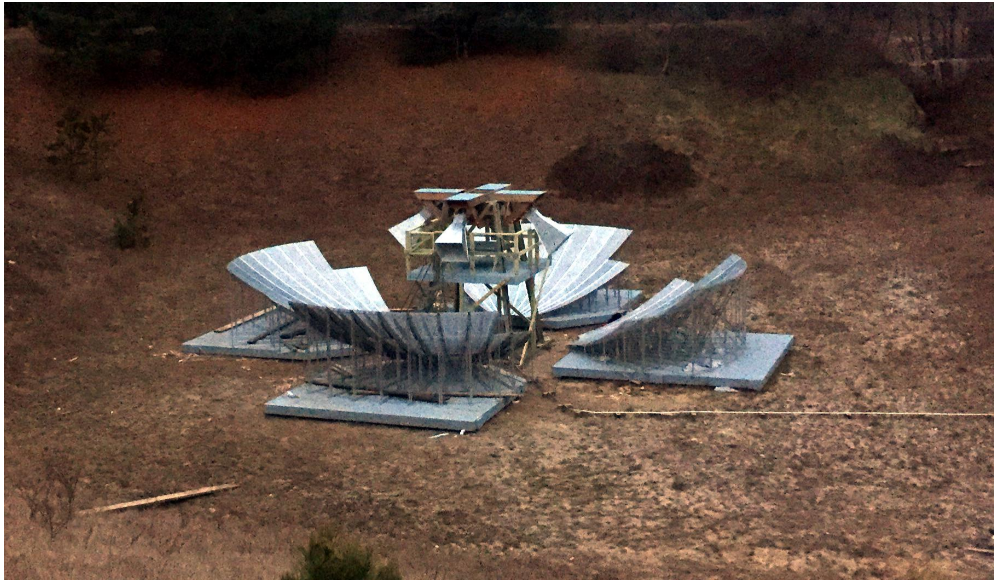
VLA

CHIME

BMX

Challenges:

- BMX is a test-bed, “toy” intereferometer on BNL site
- 4x2 polarizations = 8 channels of 550 MHz BW
- 8 bit sampler running at 1.1GS/s : 8x1.1GS/s = 8 GB/s data data hose
- 64 possible correlations in 4096 frequency channels
- We are calculating just 32 using 2 PCs and 2 GPUs



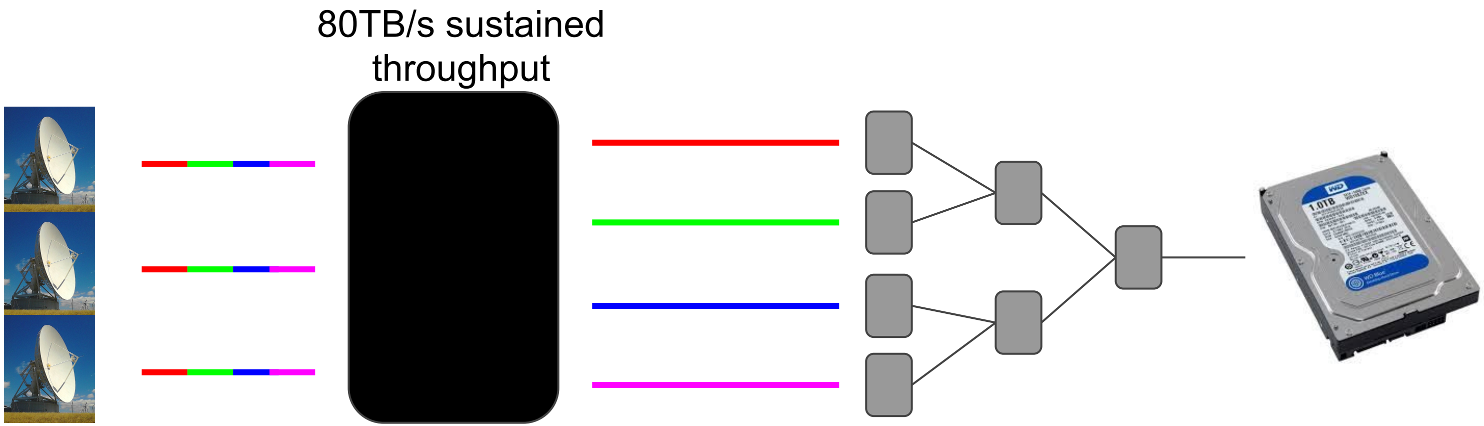
Main team: A Slosar, C Sheehy, P Stankus (Physics), Paul O’Connor (Instrumentation) + engineers and students

Future

- 21cm Stage 2 is an ambitious:
  - 65536 dishes, dual pol., 300MHz BW: 80TB/s raw data rate
- Full N^2 correlation impossible on the full array
- There is an NlogN FFT-based algorithm that is incredibly efficient, but requires perfectly calibrated instrument
- We'll have to switch between partial direct correlations AND full FFT for actual data analysis
- Additional real-time logic required for RFI removal, transient detection, etc.

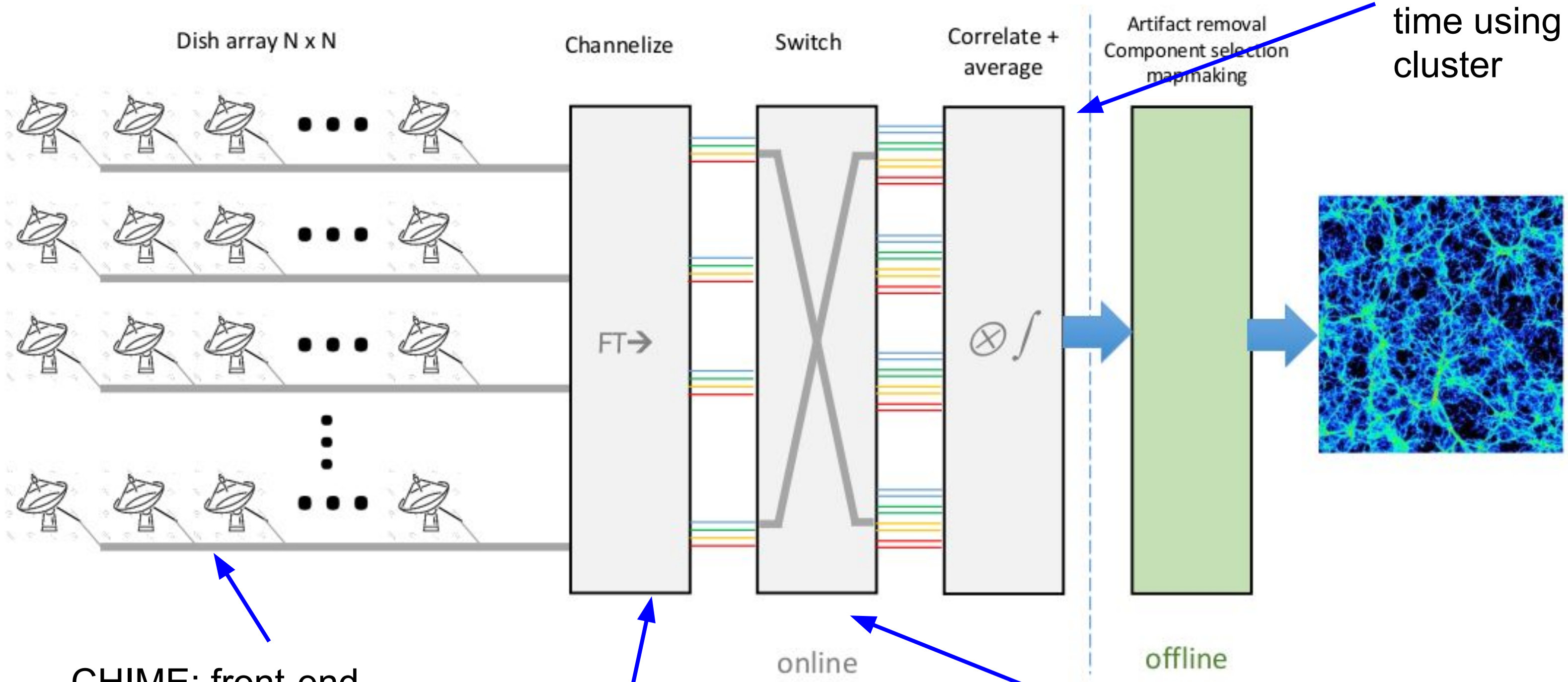
Corner-turning problem

- I suspect this is the biggest problem
- Each antenna gives all frequency channels: 65536 x 1.2GB/s
- What we want is one frequency channel from all antennas: 65536 frequency channels x 1.2GB/s
- 65536 compute nodes do the cross-correlations
- Hierarchical reduce for more advanced analysis/trigger



Where should we be doing stuff?

CHIME: Correlation, RFI removal in real time using GPU cluster



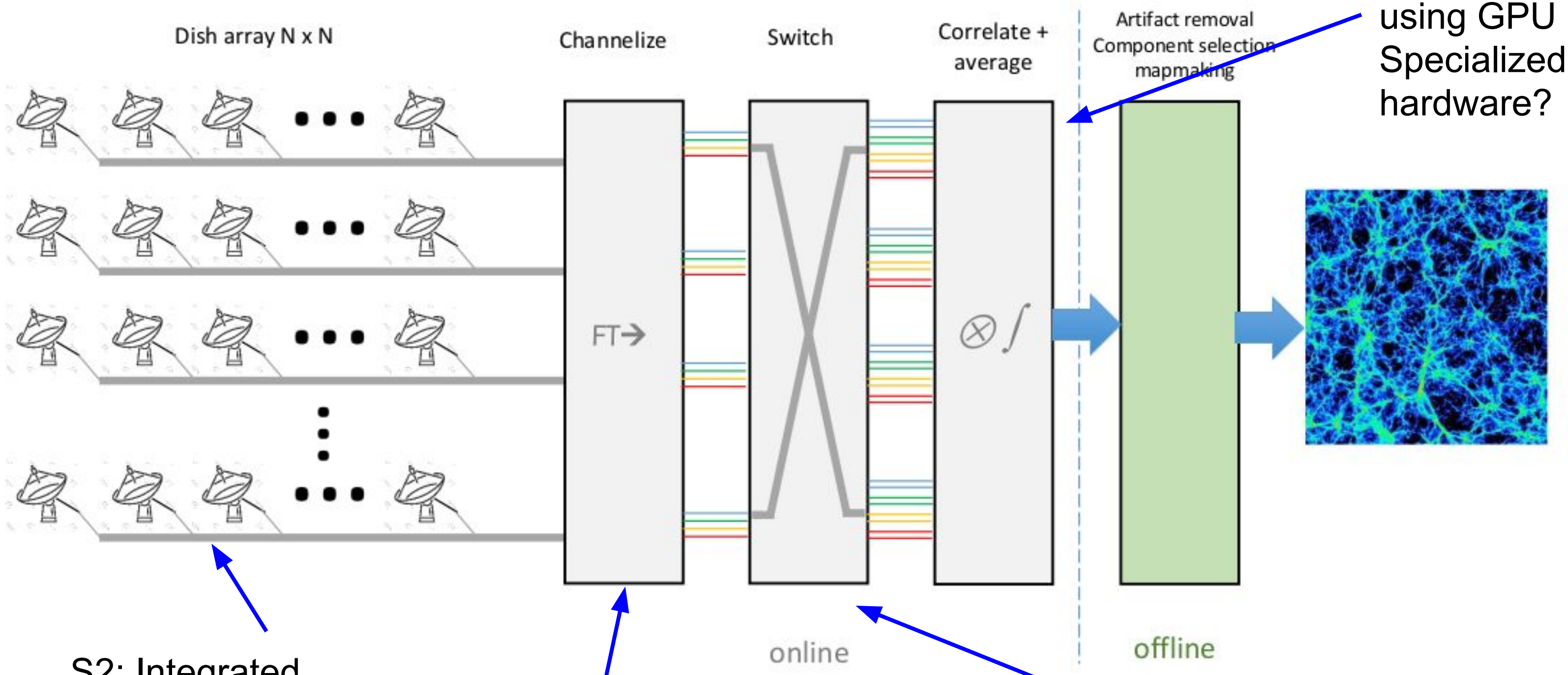
CHIME: front-end amplification on dish

CHIME: Digitization/ channelization in containers next to the array

CHIME:
Corner turning done using custom FPGA based 10GBs network

Where should we be doing stuff?

Correlation, RFI removal in real time using GPU cluster? Specialized hardware?

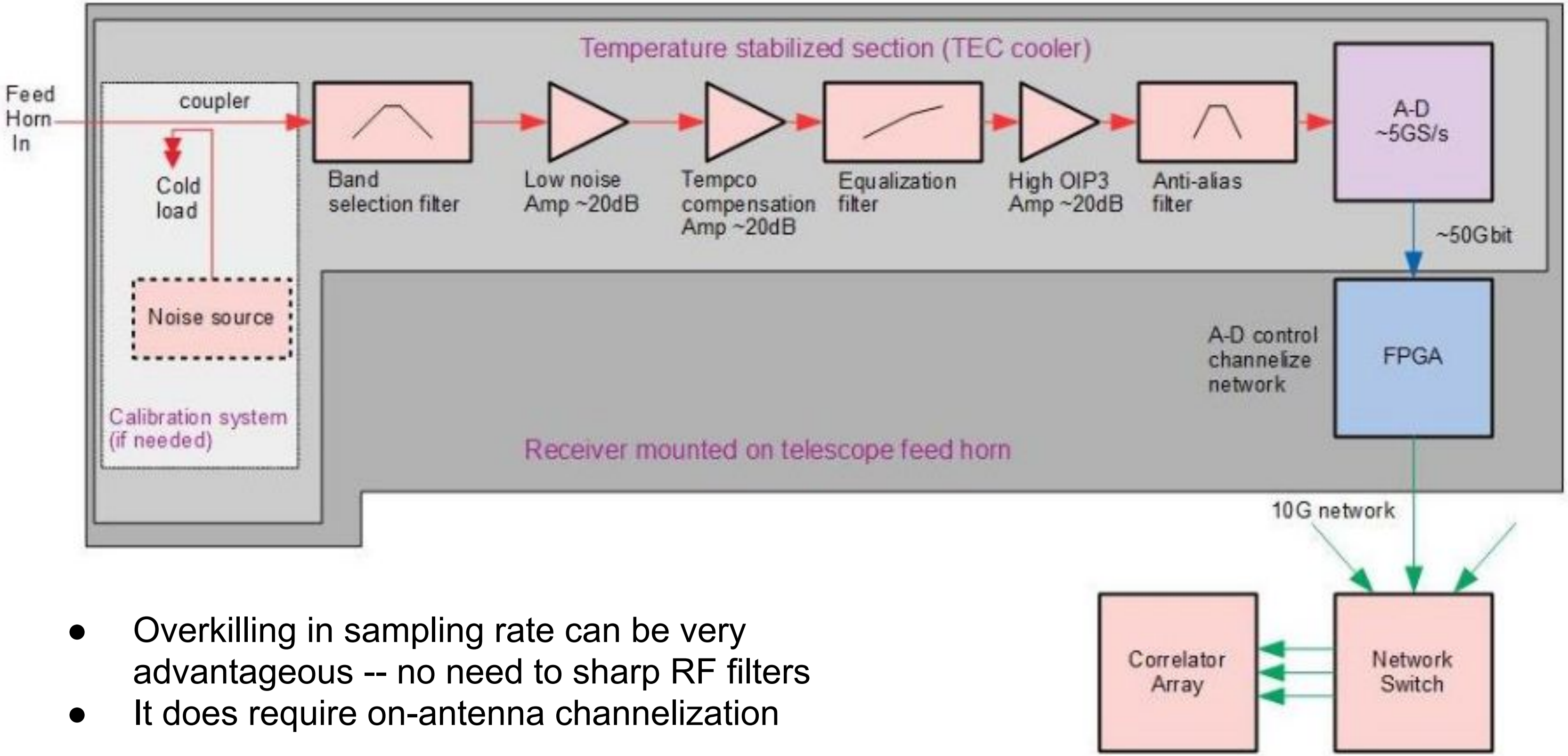


S2: Integrated OMT/amplification and digitization

Channelization on antenna, aggregates or separate?

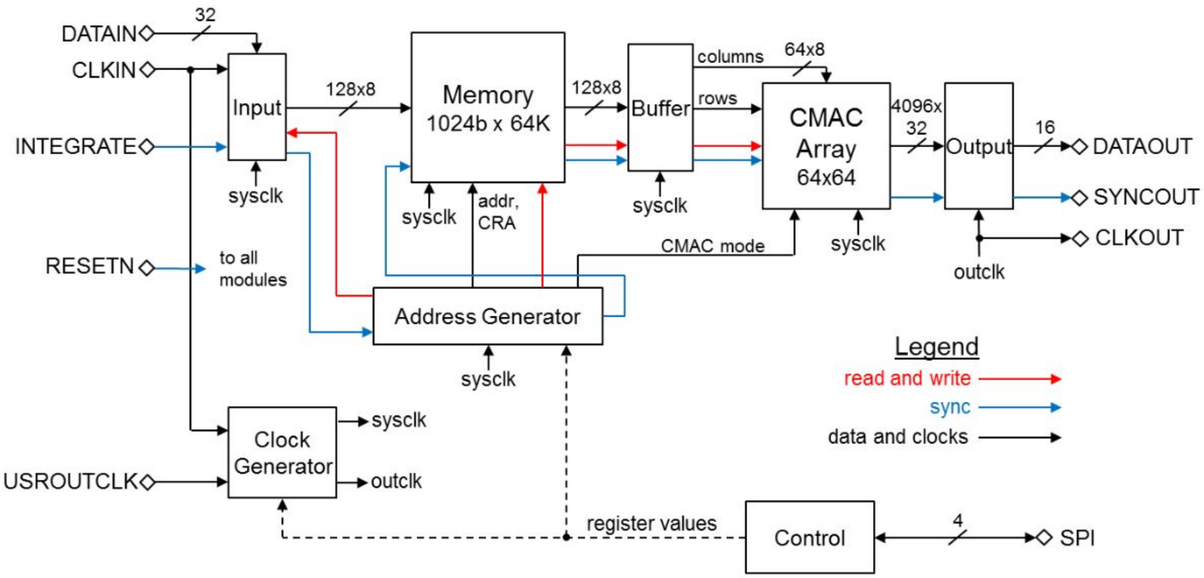
Corner turning done using off-the shelf switching gear?

Possible front-end schematics



- Overkilling in sampling rate can be very advantageous -- no need to sharp RF filters
- It does require on-antenna channelization

Dedicated silicon for correlations:



$$X = \frac{1}{N} \sum_{i=1\ldots N} a_i b_i^*$$

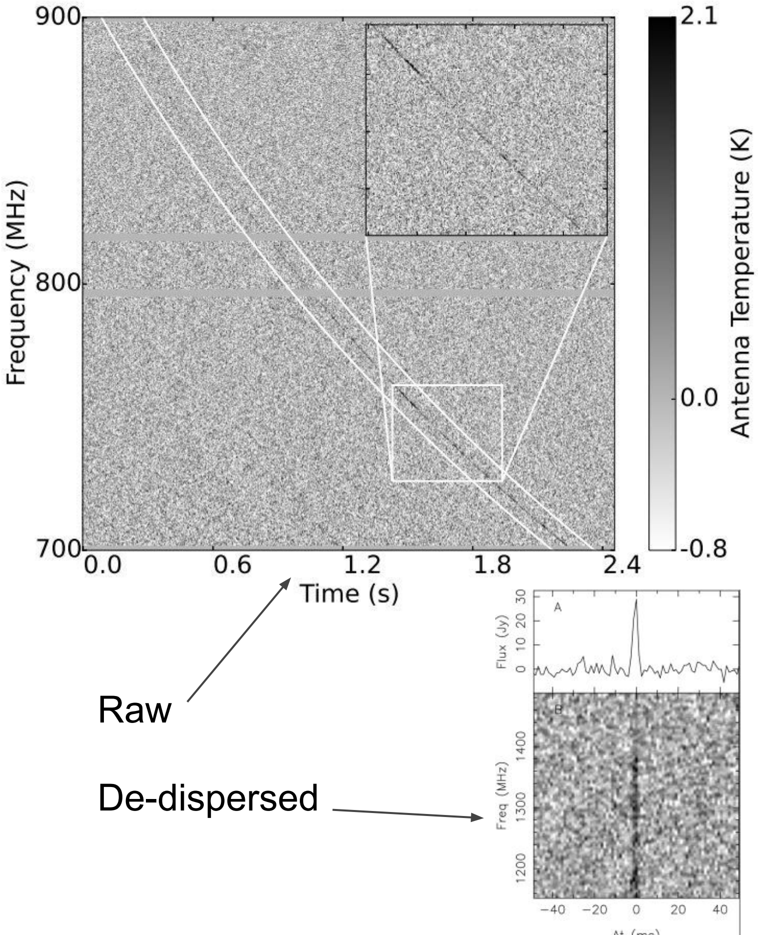
A Low-Power Correlator ASIC
For Arrays With Many Antennas

Larry R. D'Addario and Douglas Wang
Jet Propulsion Laboratory, California Institute of Technology
Pasadena, California, USA
email: ldaddario@jpl.nasa.gov

Extremely lower power designs by d'Addario (JPL).

Fast Radio Bursts

- Example of a real-time analysis you might want to do
- Millisecond bursts that get stretched into several second long signal by interstellar dispersion
- Need a trigger, efficient algorithms developed by Kendrick Smith
- Trigger needs to again combine pixels fron different frequencies (second corner-turning problem)
- You want to de-disperse the signal --
  - Need a ring-buffer stretching back in time
  - Is this feasible for Stage 2?



Conclusions

- Stage 2 offers lots of new problems, which will require both software and hardware solutions
- Hardware:
  - integrated OMT, amplification, digitization;
  - Channelization
  - corner-turning : bespoke clocked networking?
- Software:
  - Very efficient, adaptable GPU computing
  - Real time, robust calibration scheme?
  - Intelligent RFI filtering
  - Triggering on FRB events, dedispering in real time?
- Lots do be done, come speak to us!