Data-mining 21cm data sets

Peng Oh (UCSB)

Astro 1 Multiple Choice Final Exam

1. When voting in a US presidential election, which of the following is the most important consideration?

A. Charisma.
B. The Iraq war.
C. Increasing funding to the NSF-AST division.

2. Which of these best describes the reionization of the universe?

A. A riddle inside a mystery, wrapped in an enigma.
B. A process whose properties can be predicted from first principles in numerical simulations, just like galaxy formation
C. A somewhat over-rated "landmark" event in the history of the universe whose importance is surprisingly difficult to justify to non-astronomers.

courtesy S. Furlanetto

Advertising: 21cm Review Article

ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Physics Reports ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮

PHYSICS REPORTS

www.elsevier.com/locate/physrep

Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe

Steven R. Furlanetto[a,*], S. Peng Oh[b], Frank H. Briggs[c]

[a] *Department of Physics, Yale University, PO Box 208121, New Haven, CT 06520, USA*

[b] *Department of Physics, University of California, Santa Barbara, CA 93106, USA*

[c] *Research School of Astronomy and Astrophysics, The Australian National University, Mount Stromlo Observatory, Cotter Road, Weston, ACT 2611, Australia*

Accepted 1 August 2006

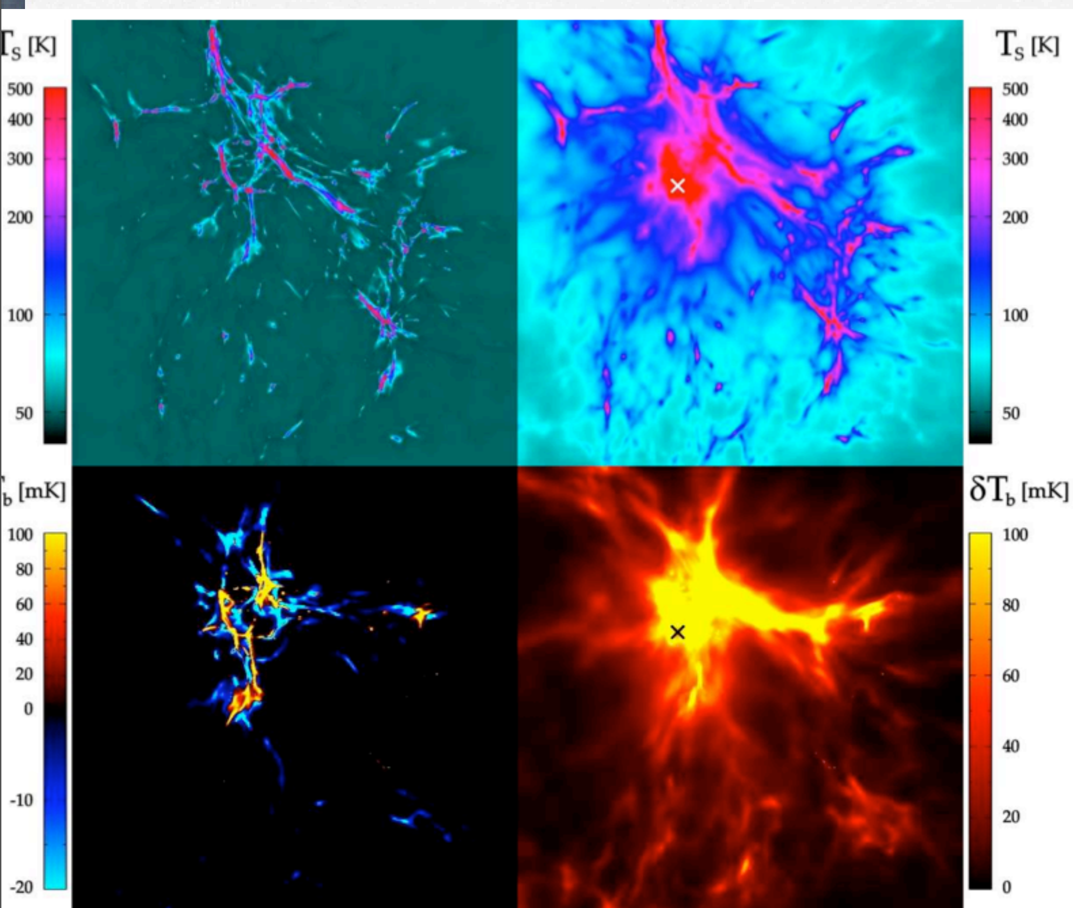
Furlanetto, Oh & Briggs 2006 (FOB06)

Physics Reports

Top Google hit for '21cm transition'

But if you Google '21cm line' instead you get...

21cm observations will revolutionize the field



Kuhlen & Madau 06

See 21cm emission from IGM in absorption or emission against CMB

Couple spin and kinetic temperatures by collisions or Wouthuysen-Field effect

Probe both Dark Ages and First Light

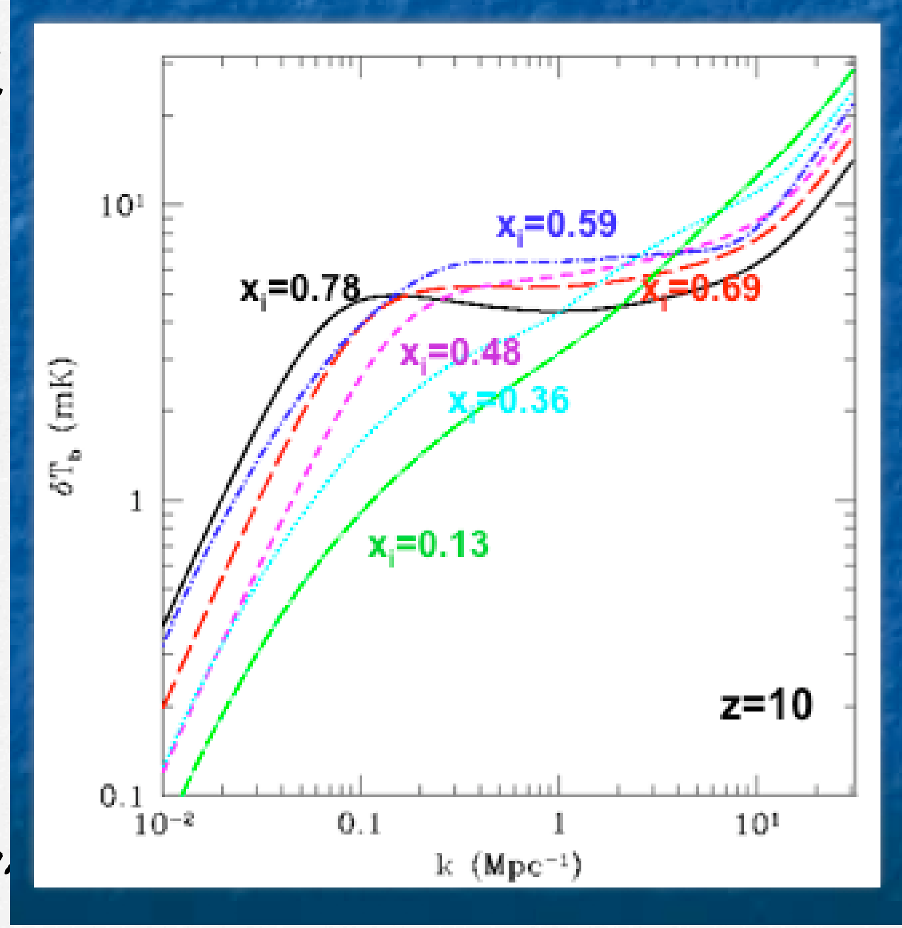
MWA---Western Australia

LOFAR--Netherlands

SKA--??

21cm Power Spectrum

- Language generally used for 21cm fluctuations
- Tools developed for CMB/ galaxy surveys
- Natural language for interferometer
- Good choice for Dark Ages, before ionizing sources turn on. But after that...

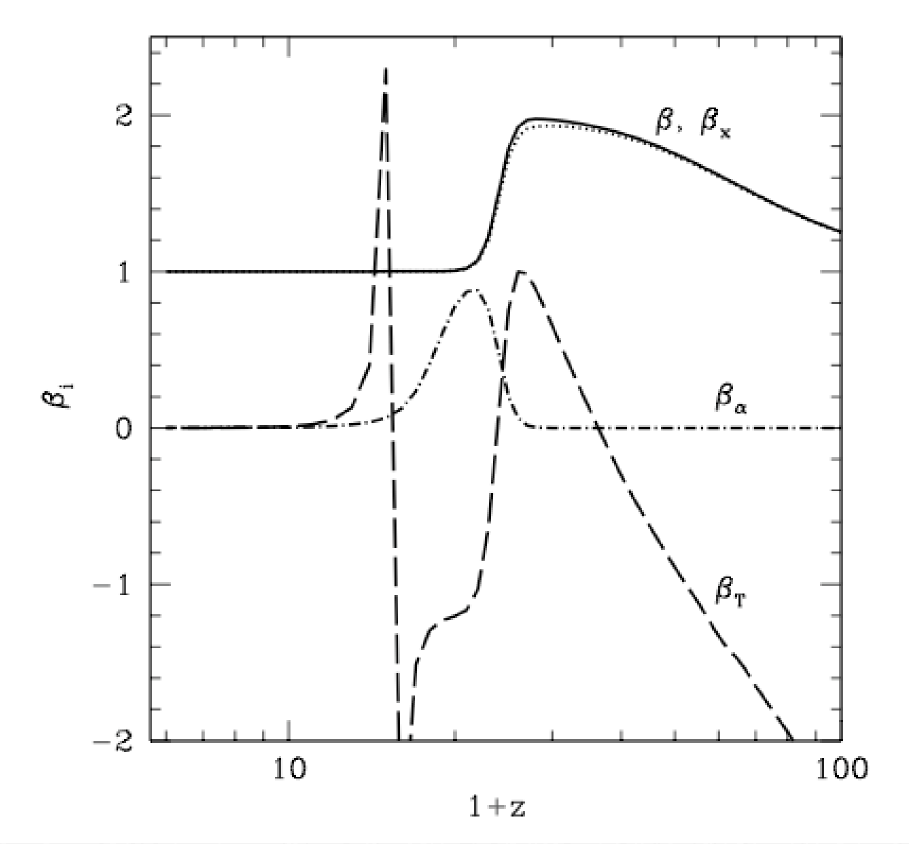


FOB06

...many effects contribute

Fluctuations in...

- ☐ density (Gaussian)
- ☐ Ly-alpha flux
- ☐ ionization state
- ☐ temperature
- ☐ velocity gradients

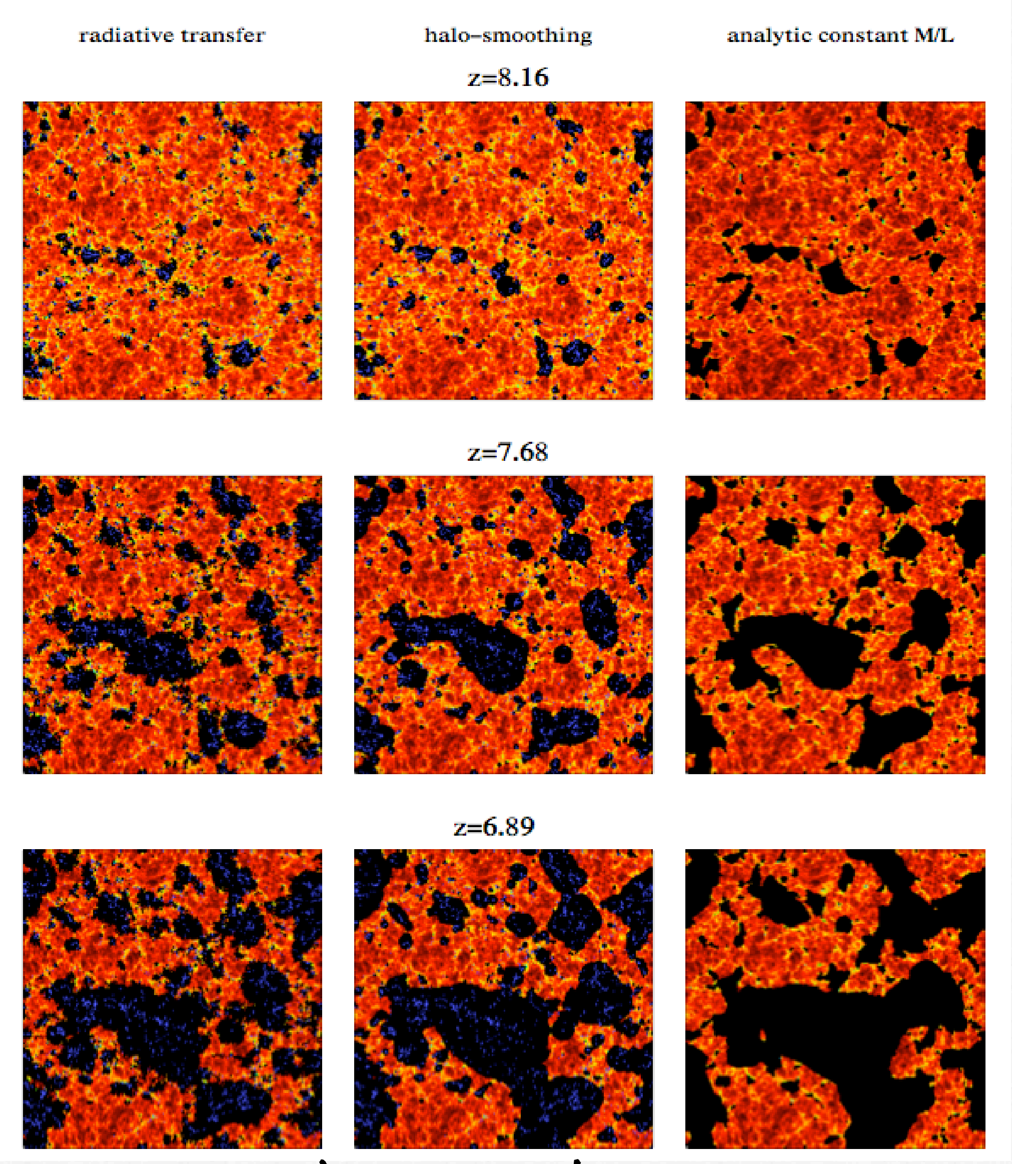


FOB06

Many likely to be correlated

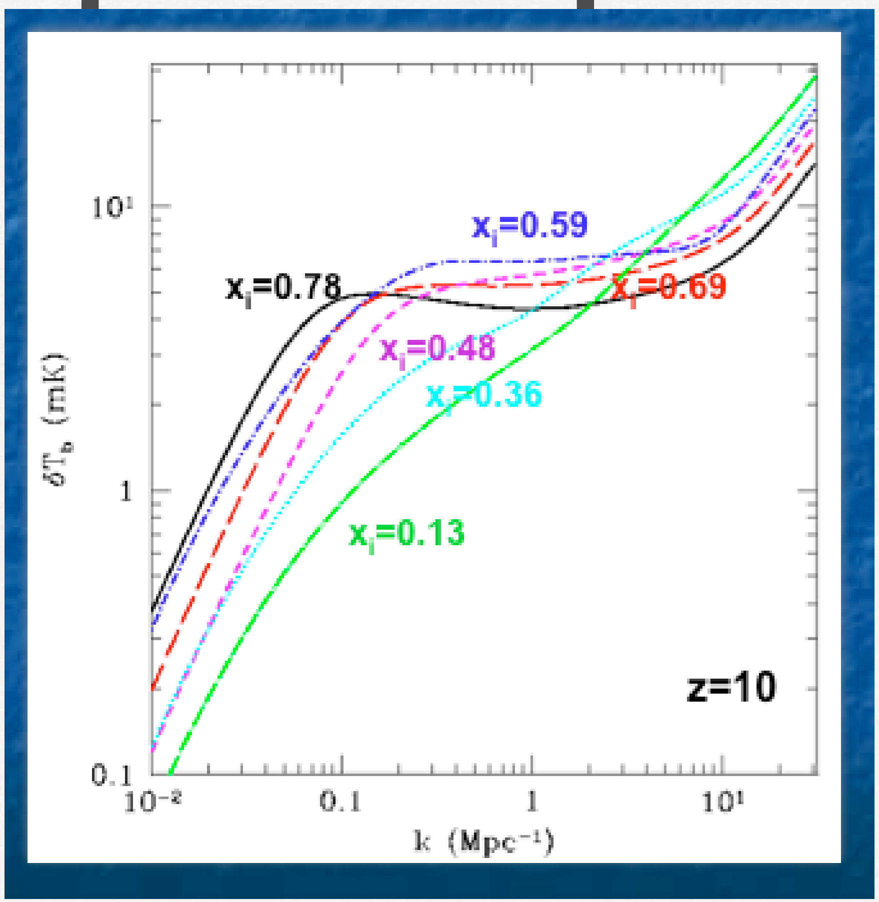
...it's a highly no

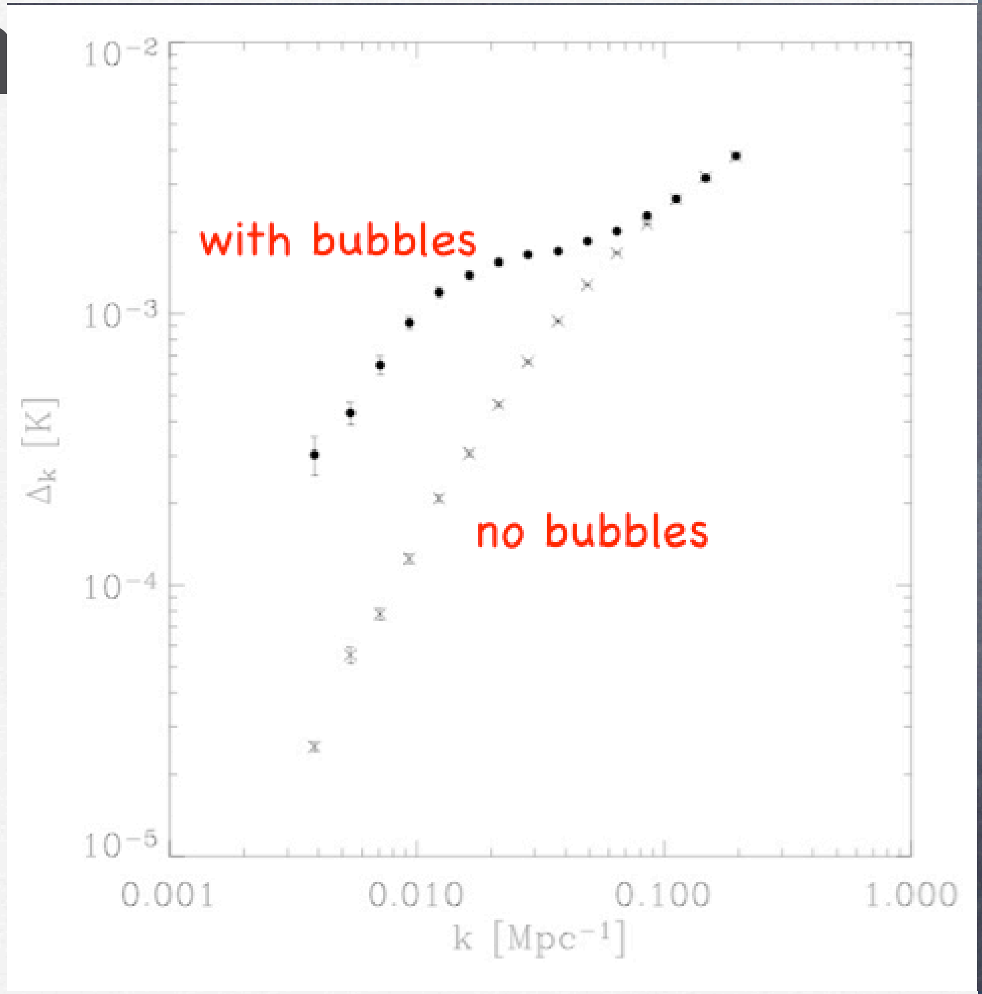
If we want to study growth and topology of reionization, we should focus on the bubbles



Zahn et al 2006

...bubbles DO strongly affect power spectrun

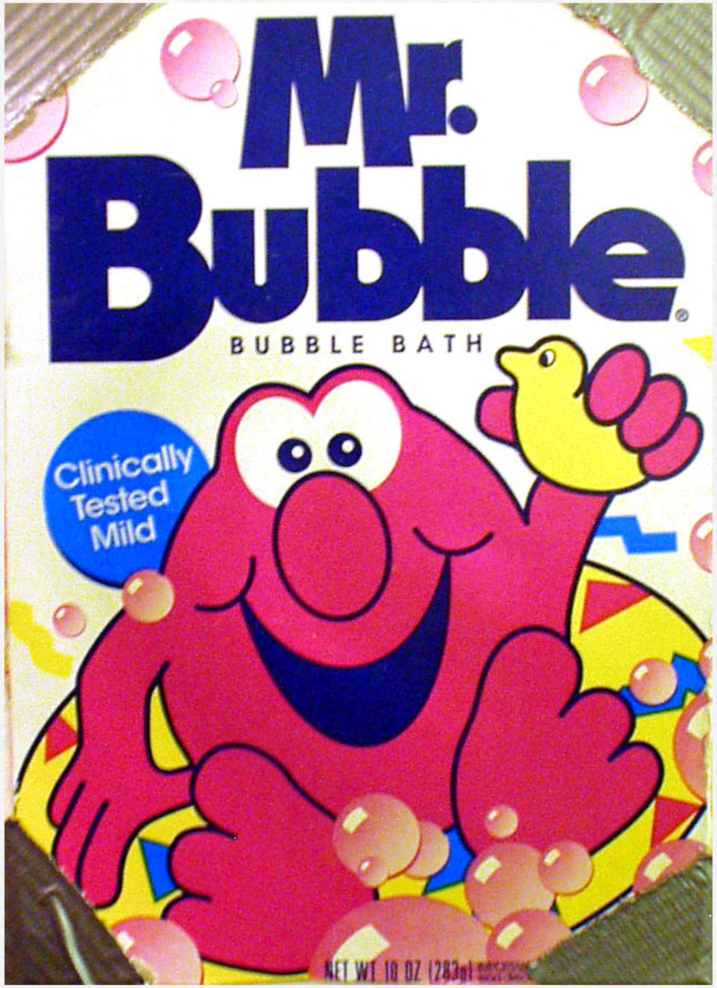




...but quantifying this will be model-dependent

Bubbles are your Friend

- Probe of ionizing source population (supposed to be big)
- Directly extract HII filling factor
- Foreground calibrator:
  - Measure mean temperature T(z)
  - Remove long wavelength artifacts from foreground removal



Direct Imaging

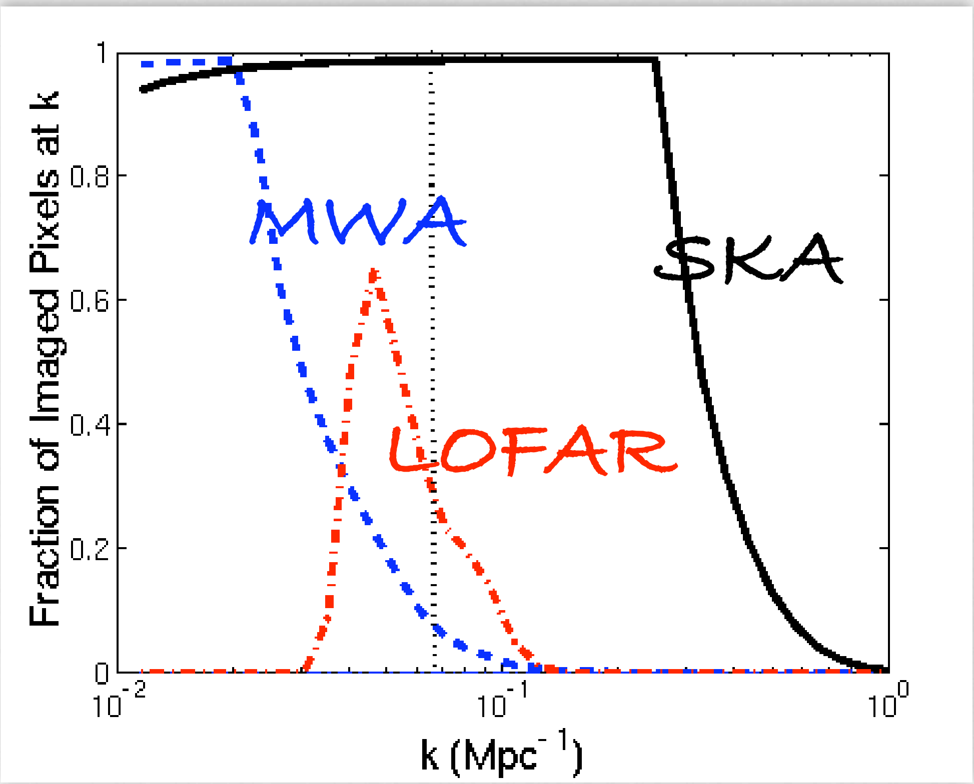
S/N high only on largest scales, need R~20 Mpc

Rare bright quasars (or clustered galaxies)

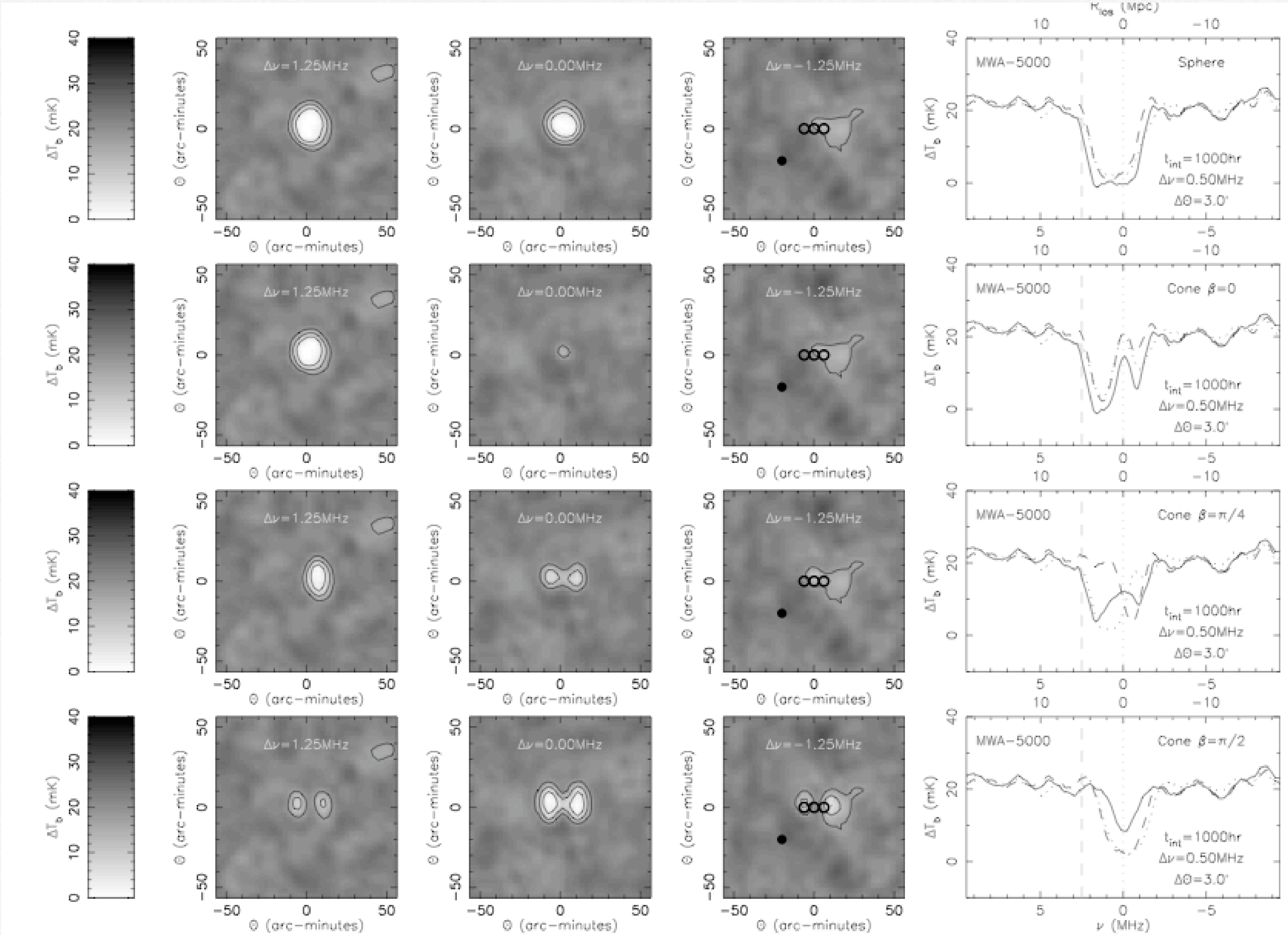
BUT: survey volume is HUGE!

Expect 1 active/fossil HII region in every MWA FOV with R > (24,40) Mpc at z=7

(Wyithe, Loeb & Barnes 2004)



McQuinn et al 2006



Wyithe, Loeb & Barnes 2004

...what do we get?

- -- $\delta T_b(z)$ X-rays, fossil HII
- Foreground calibrator
- Size, shape of HII region--> QSO properties
- Discover QSOs? (though mostly their fossils)
- Try to cross-correlate with galaxy population

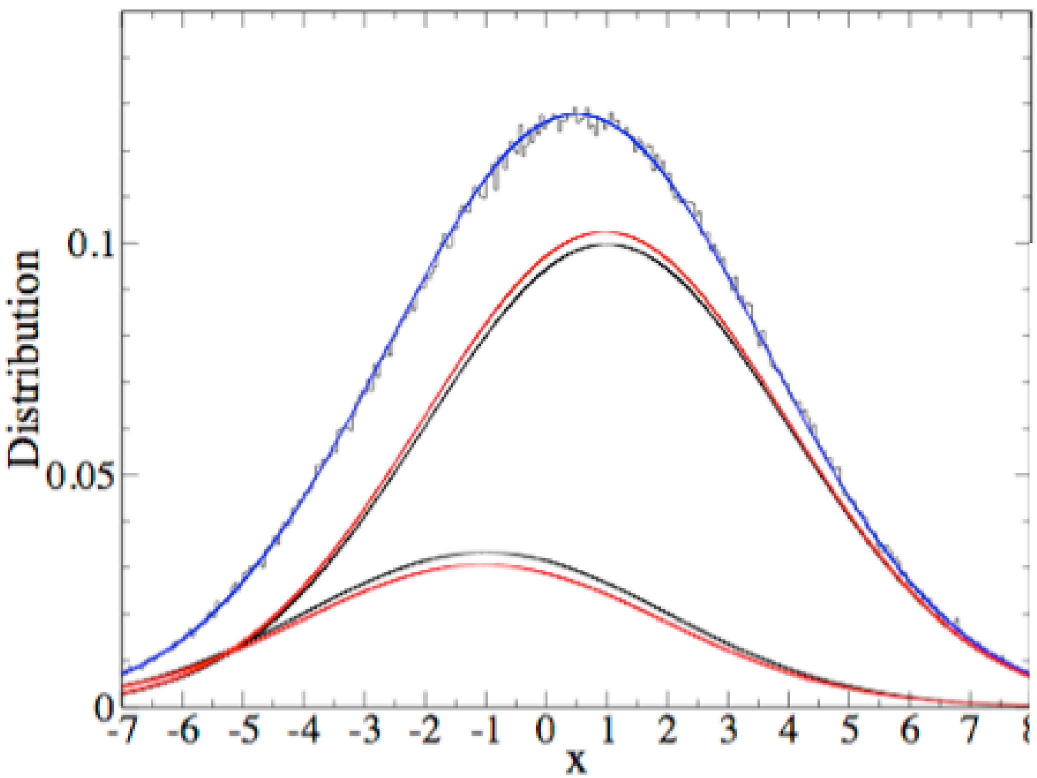
But can we see the smaller bubbles and get Q_HII(z)?

Back to Basics: One Point Statistics

Hansen, Oh & Furlanetto
(2007, in prep)

One Point Statistics

Bubbles create bimodality in the PDF



Can we pick it out?

Directly tells us Q_HII(z)!!!

DETECTING BIMODALITY IN ASTRONOMICAL DATASETS

KEITH M. ASHMAN AND CHRISTINA M. BIRD
Department of Physics and Astronomy, University of Kansas, Lawrence, Kansas 66045-2151
Electronic mail: ashman@kusmos.phsx.ukans.edu, tbird@kula.phsx.ukans.edu

STEPHEN E. ZEPF[1]
Department of Astronomy, University of California, Berkeley, California 94720
Electronic mail: zepf@astron.berkeley.edu
Received 1994 April 13; revised 1994 July 29

ABSTRACT

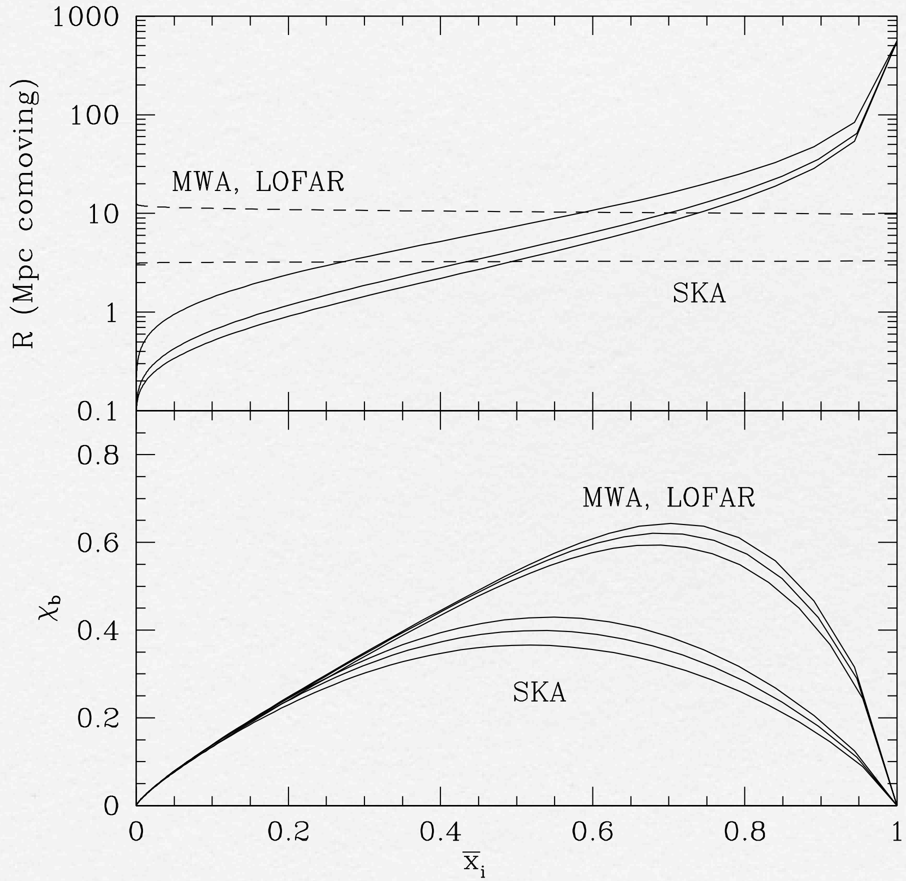
We discuss statistical techniques for detecting and quantifying bimodality in astronomical datasets. We concentrate on the KMM algorithm, which estimates the statistical significance of bimodality in such datasets and objectively partitions data into subpopulations. By simulating bimodal distributions with a range of properties we investigate the sensitivity of KMM to datasets with varying characteristics. Our results facilitate the planning of optimal observing strategies for systems where bimodality is suspected. Mixture-modeling algorithms similar to the KMM algorithm have been used in previous studies to partition the stellar population of the Milky Way into subsystems. We illustrate the broad applicability of KMM by analyzing published data on globular cluster metallicity distributions, velocity distributions of galaxies in clusters, and burst durations of gamma-ray sources. FORTRAN code for the KMM algorithm and directions for its use are available from the authors upon request.

...partially ionized boundary pixels create complications

$$f_{bd} \approx 3\frac{r_{pix}}{R_{bub}}Q_{\rm HII}$$

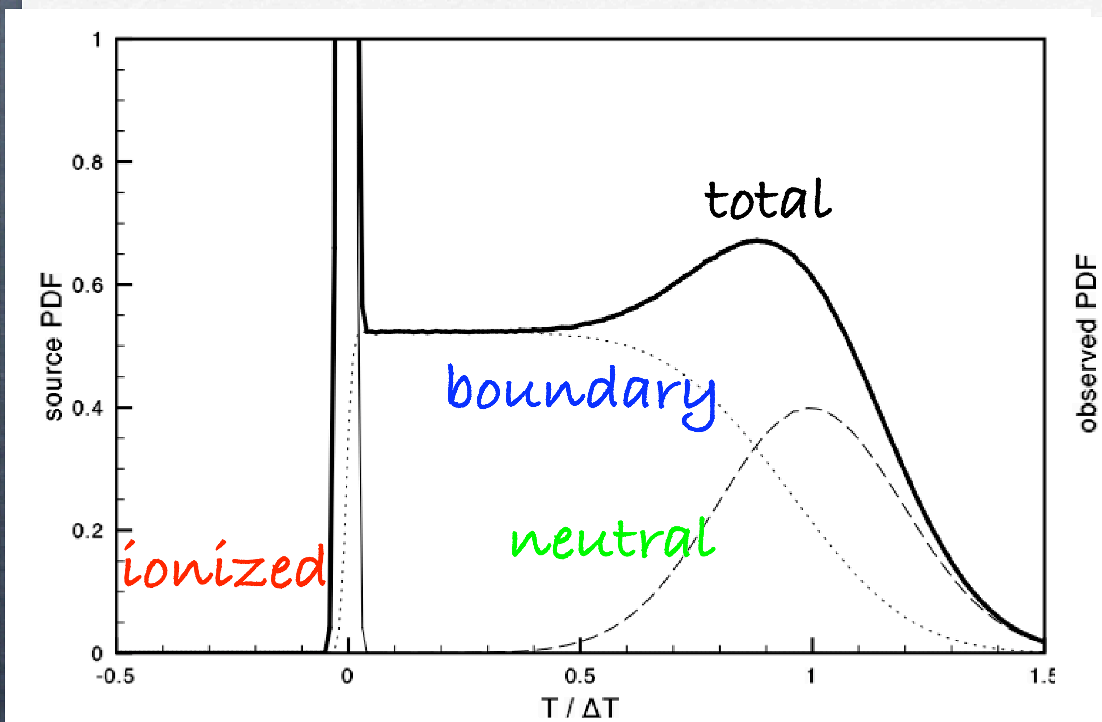
Can be ~10-70% of pixels

Dependent on telescope resolution+bubble size

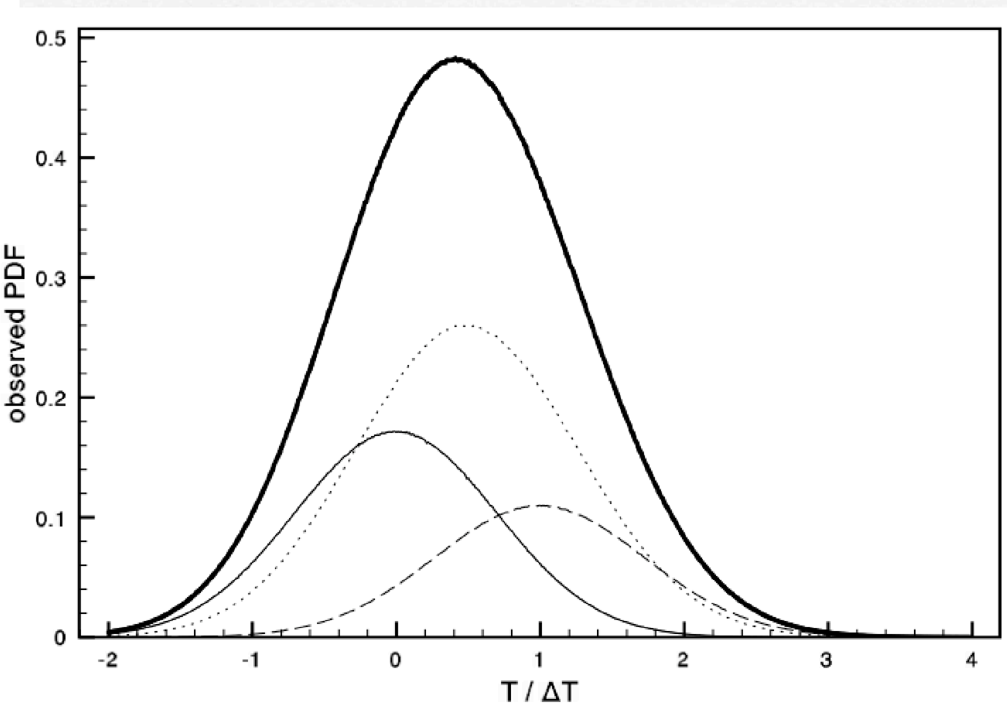


Also partially ionized pixels from X-rays, fossil regions

...here's a more realistic PDF



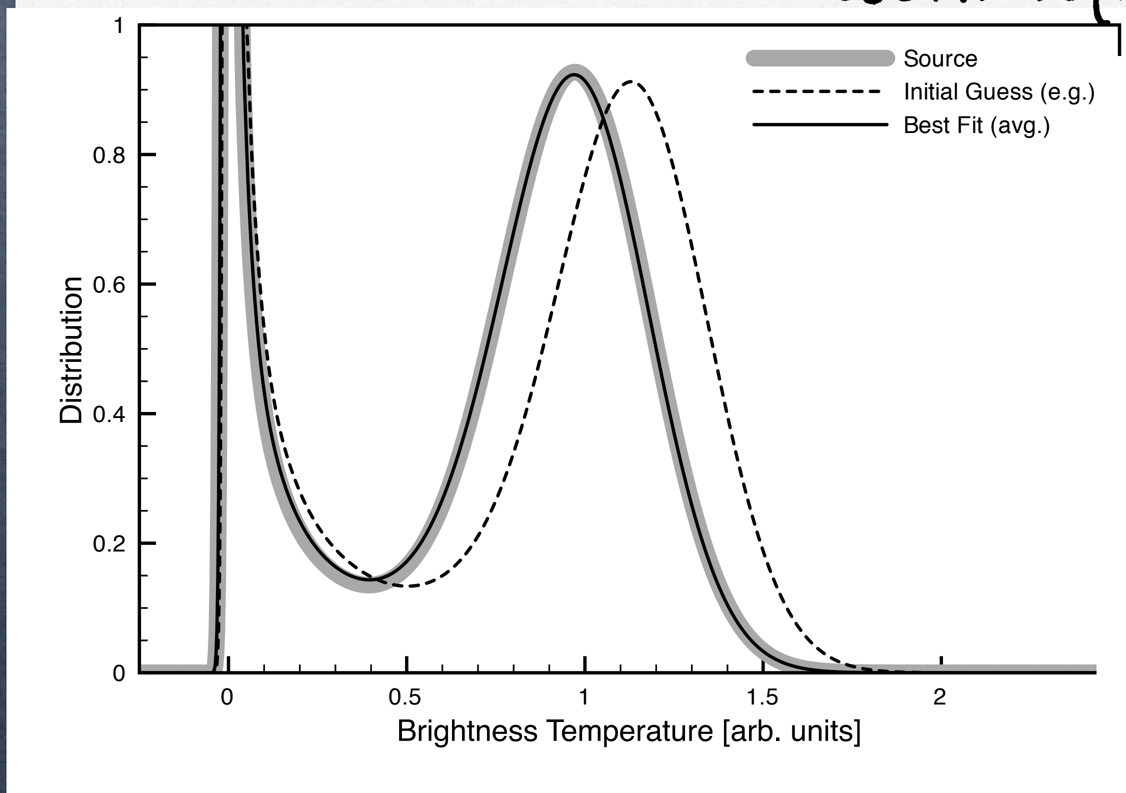
no noise

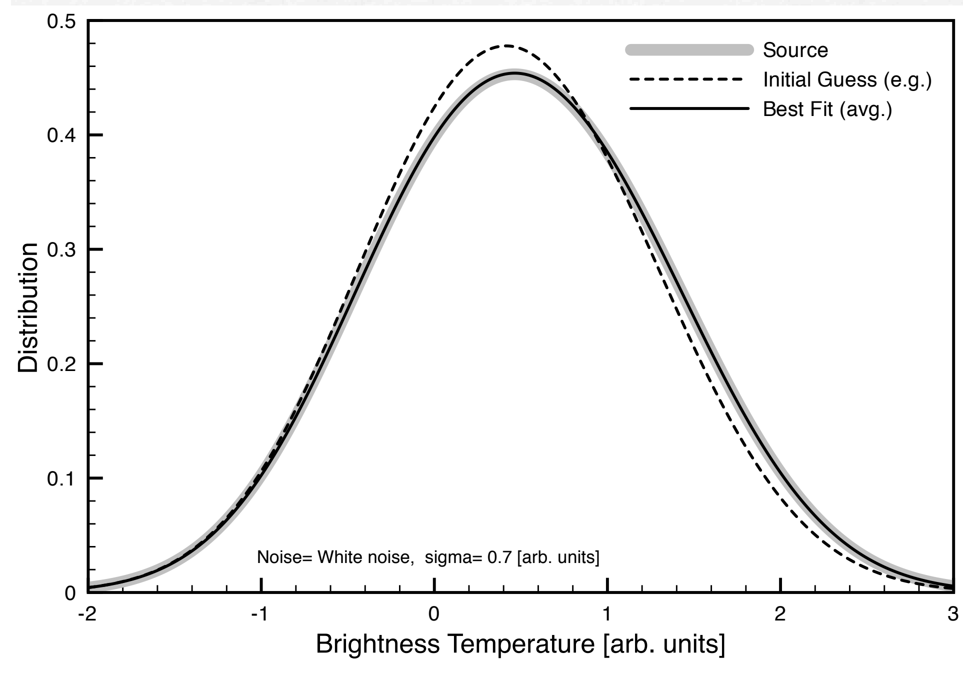


with noise

It works!

Solve for populations via iterative Max-likelihood technique



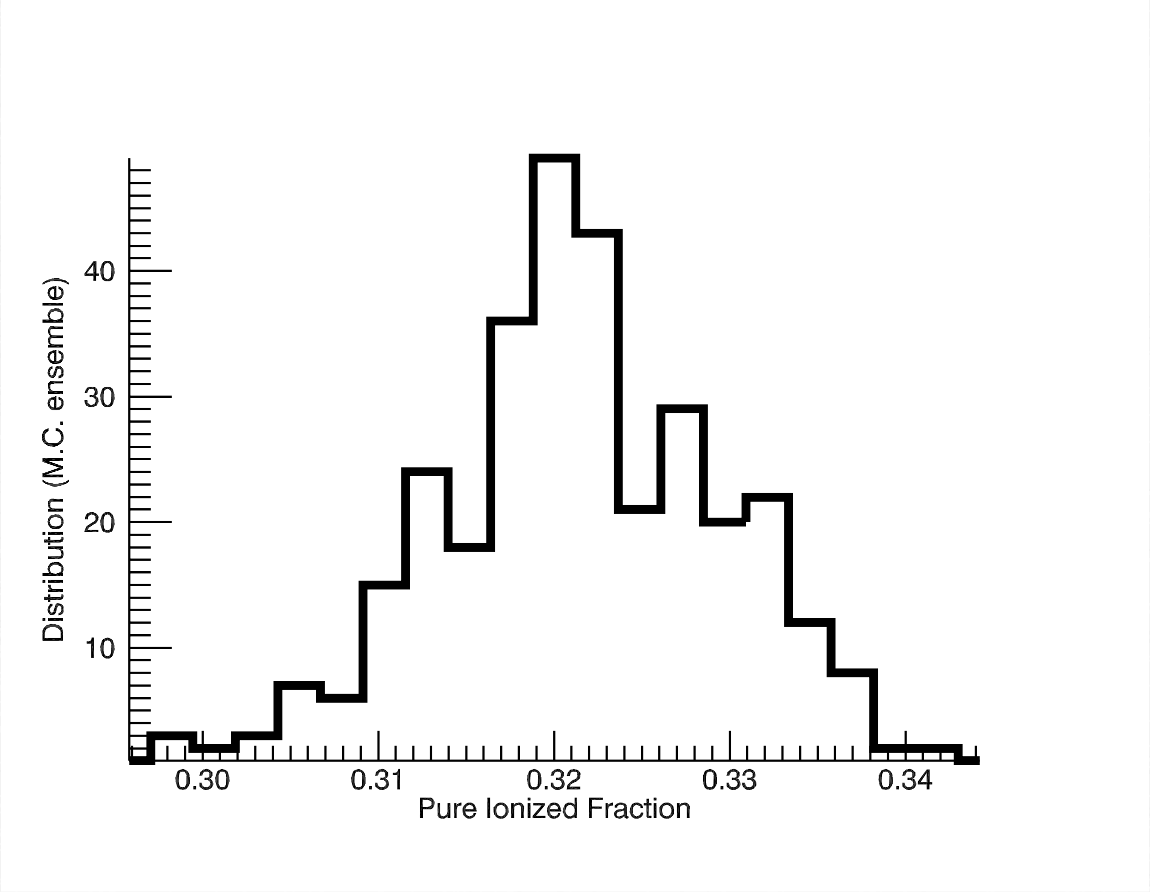


original distribution

with noise

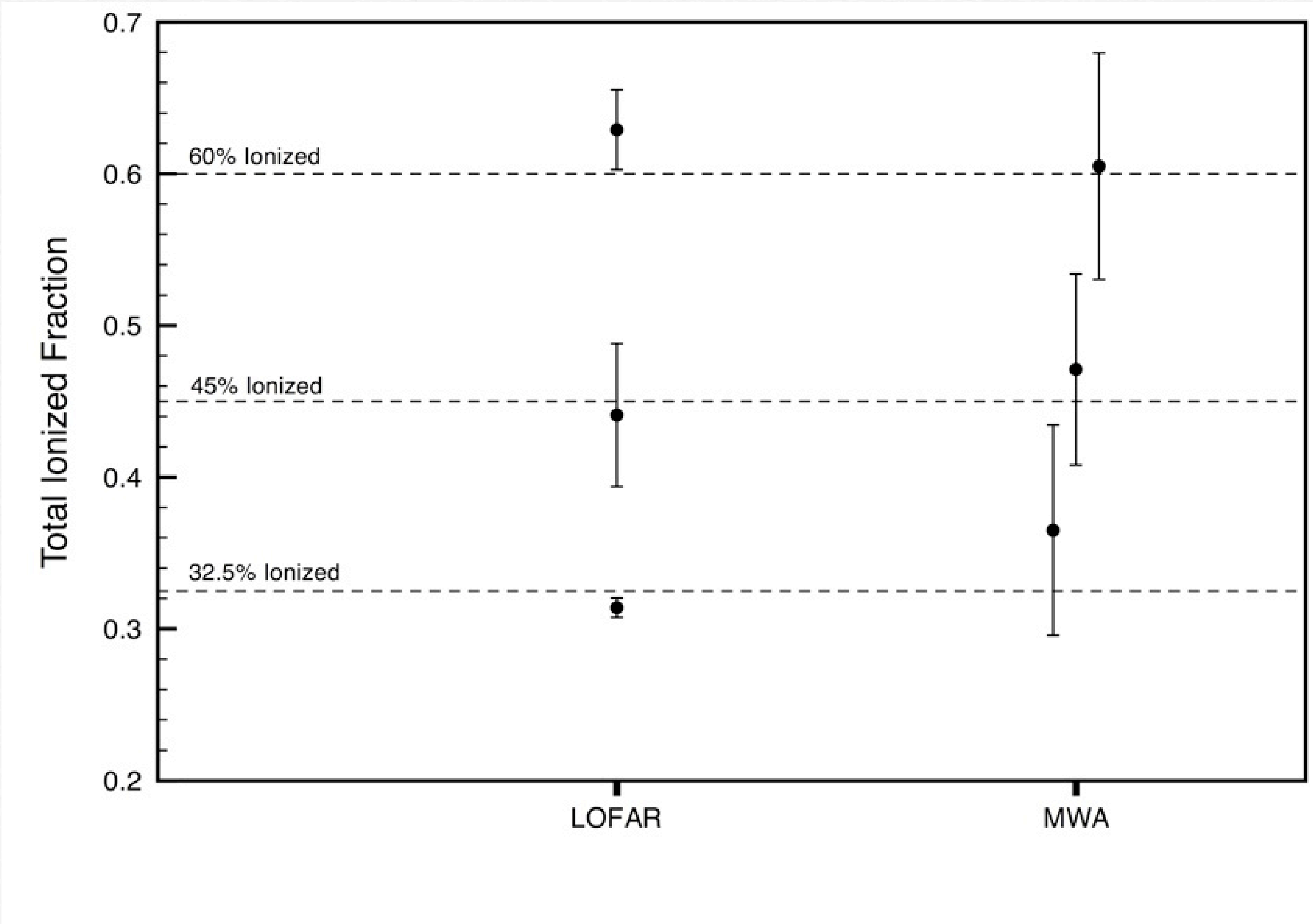
Leverage comes from having many pixels

Monte Carlo Errors agree with Fisher Matrix estimates



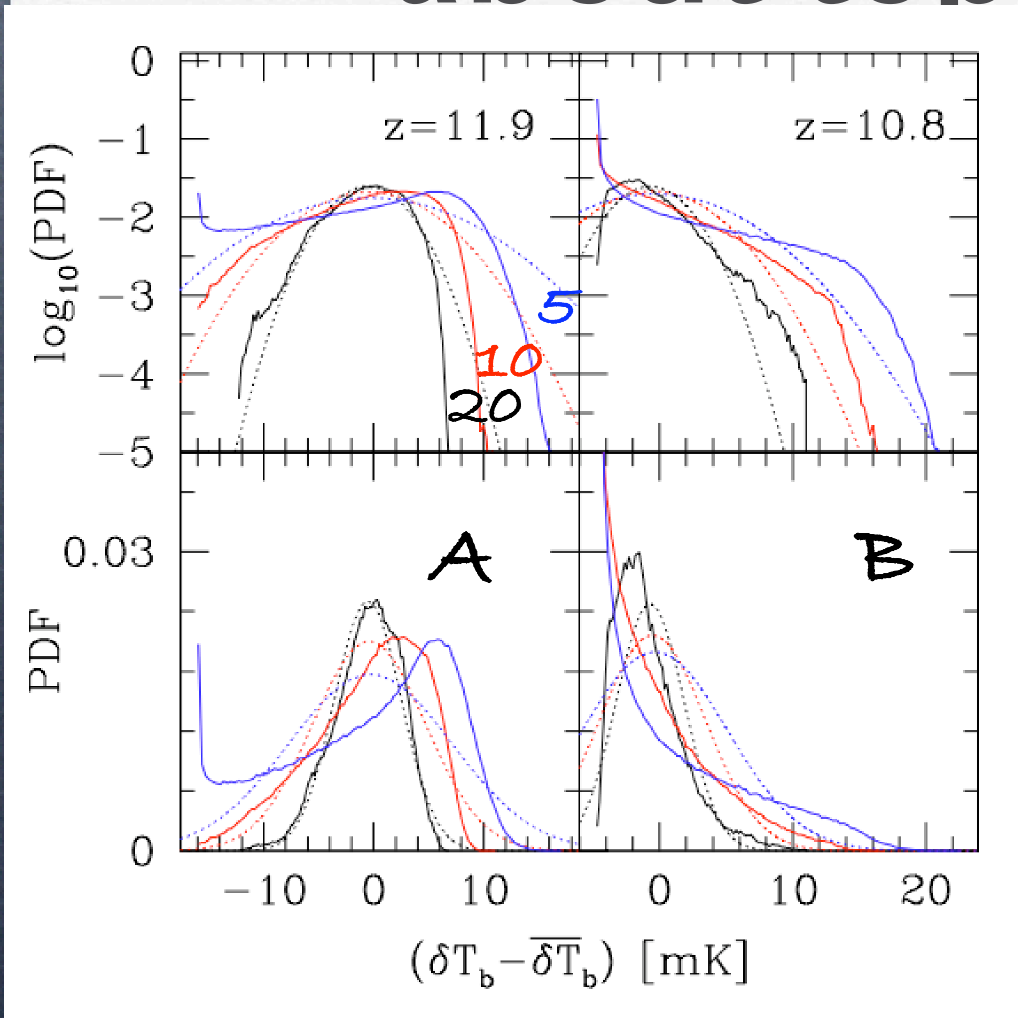
An idealized case, but results are very encouraging...

Results...



Looks very promising...

In principle, PDF has info about topology too



Mellema et al 2006

A: Cutoff at high Tb-- inside-out reionization

B: Tail at high Tb-- islands of neutral gas inside HII regions

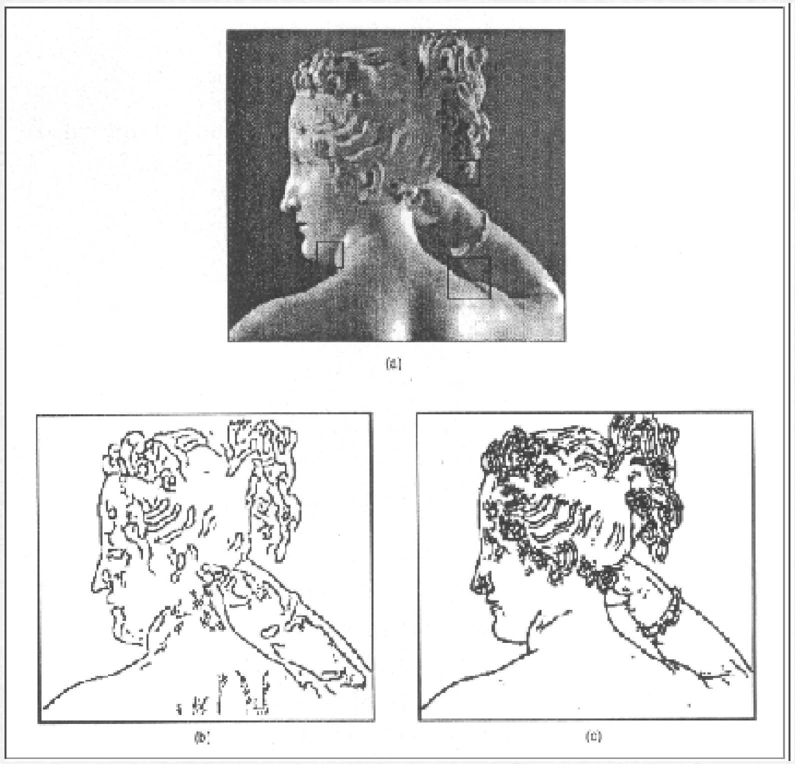
Note: distribution narrows as smoothing scale increases

Direct Bubble Detection: The Canny Algorithm

Phelps et al (2007, in prep)

How to detect bubbles directly?

Look for edges in noisy background: classic image processing problem



Canny Algorithm:

--optimal edge detector

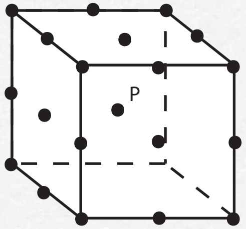
--looks for maxima in derivatives of smoothed image

1. Apply Gaussian filter

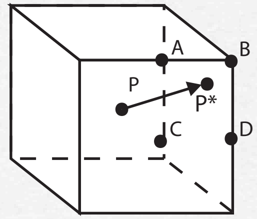
--need to do this several times at different scales

2. Find edge pixels

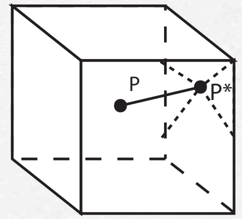
-- find maximum in 3D spatial gradient



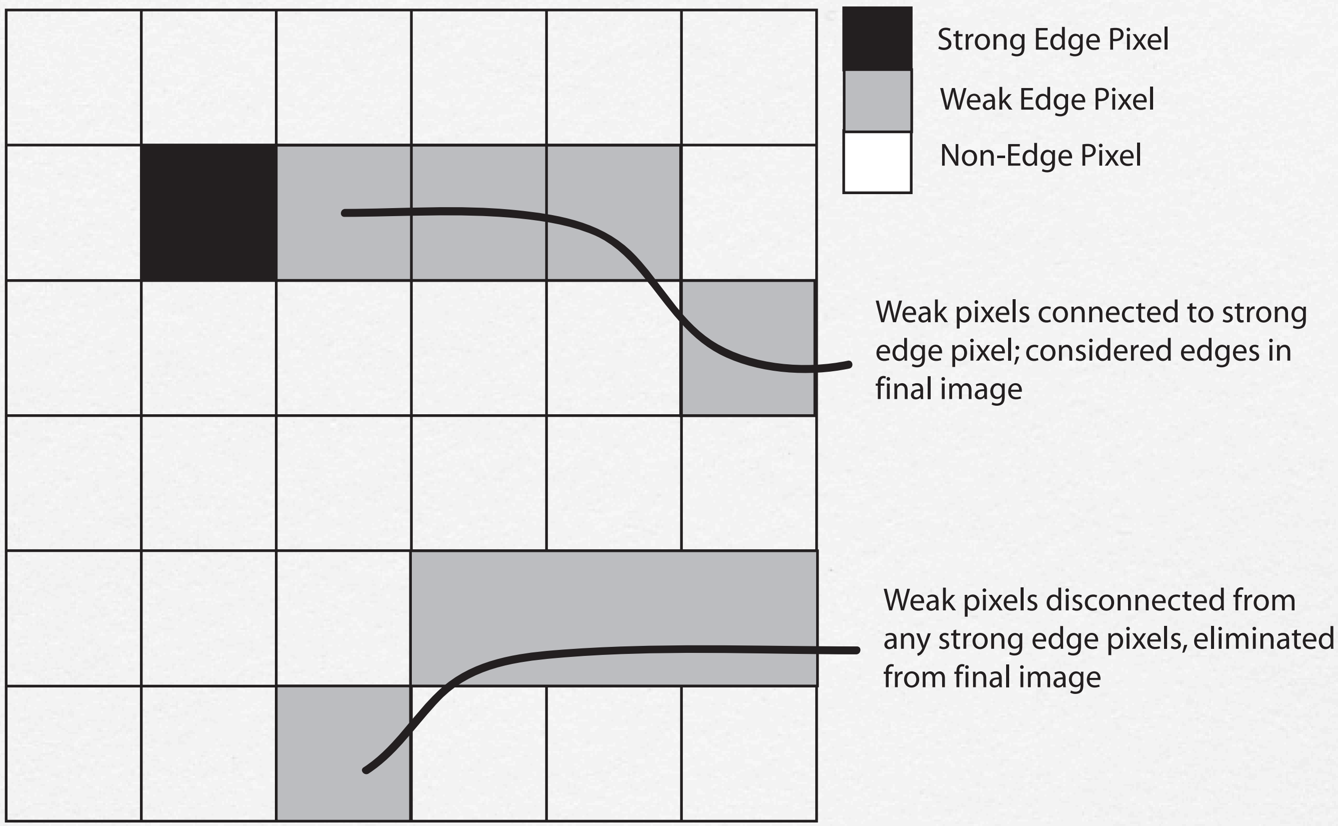
We want to determine whether pixel P (in the center of the cube) is a local maximum of the signal



By using the gradient information, we are able to determine that the gradient of P intersects the cube at point P*, which lies closest to pixels A, B, C, and D.



Interpolating the gradient magnitudes of the signal at pixels A, B, C, and D, we are able to determine the gradient magnitude of P*. If this value, and the corrolating value in the opposite direction, are less than the gradient magnitude at P*, P* is indeed a maximum.

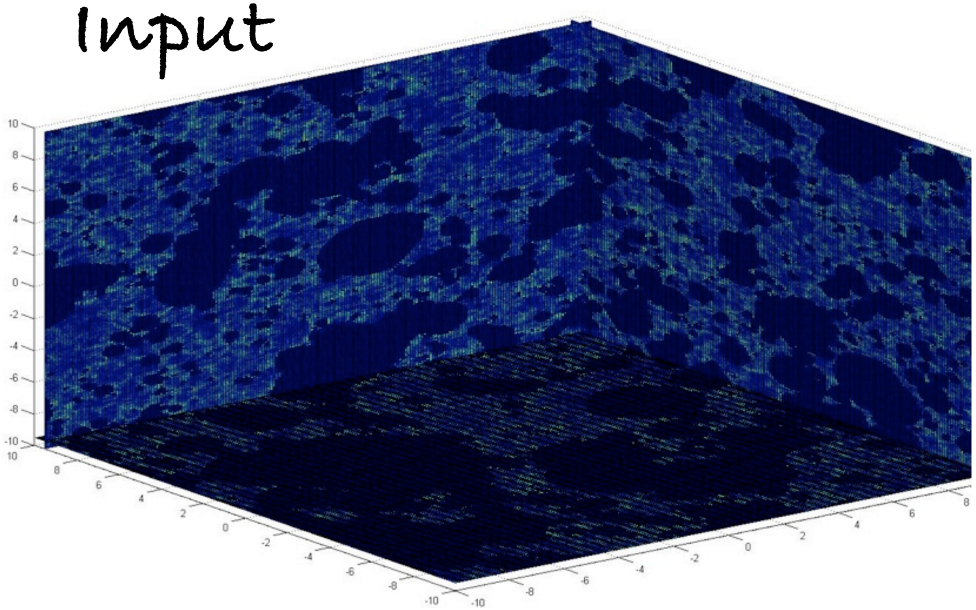


3. Apply thresholding with hysteresis

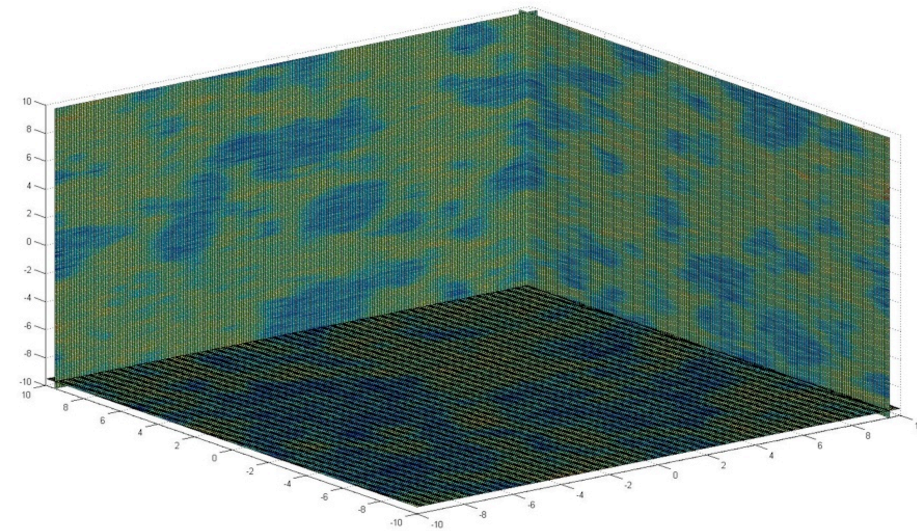
--strong pixel: automatically part of edge

--weak pixel: only part of edge if connected to strong pixel

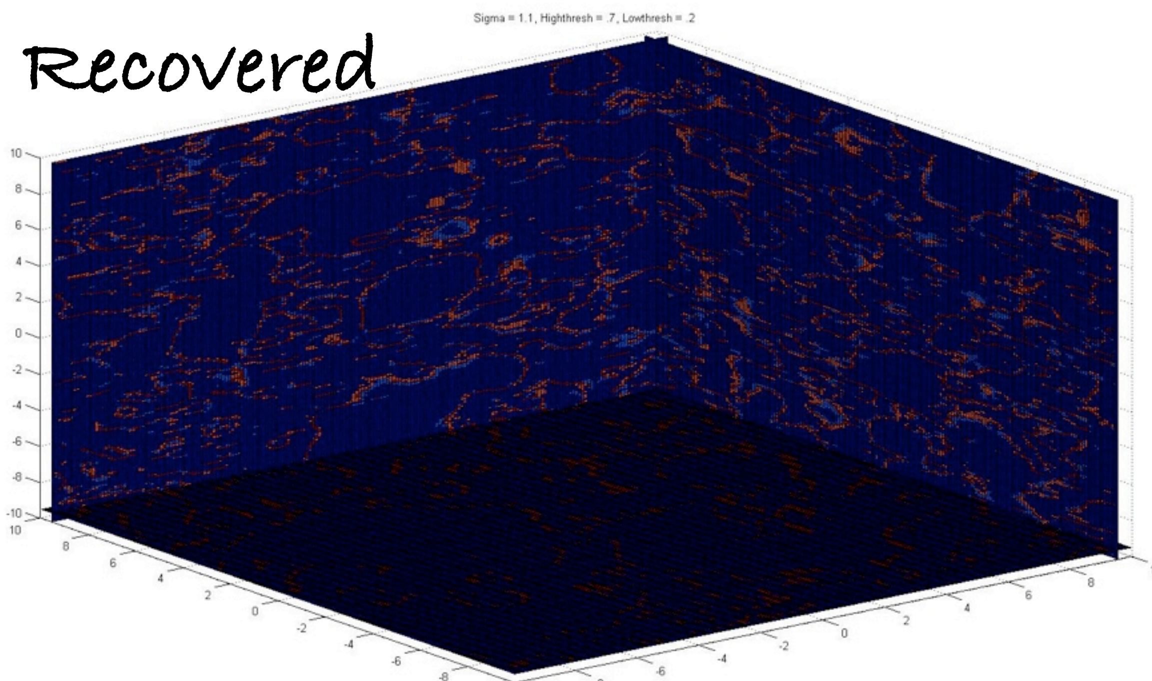
Input



with noise



Recovered



Input box from Mesinger & Furlanetto 2007

A few words about foregrounds…

Continuum foregrounds

Signal: ~10 mK

Noise: 1) Galactic foreground:

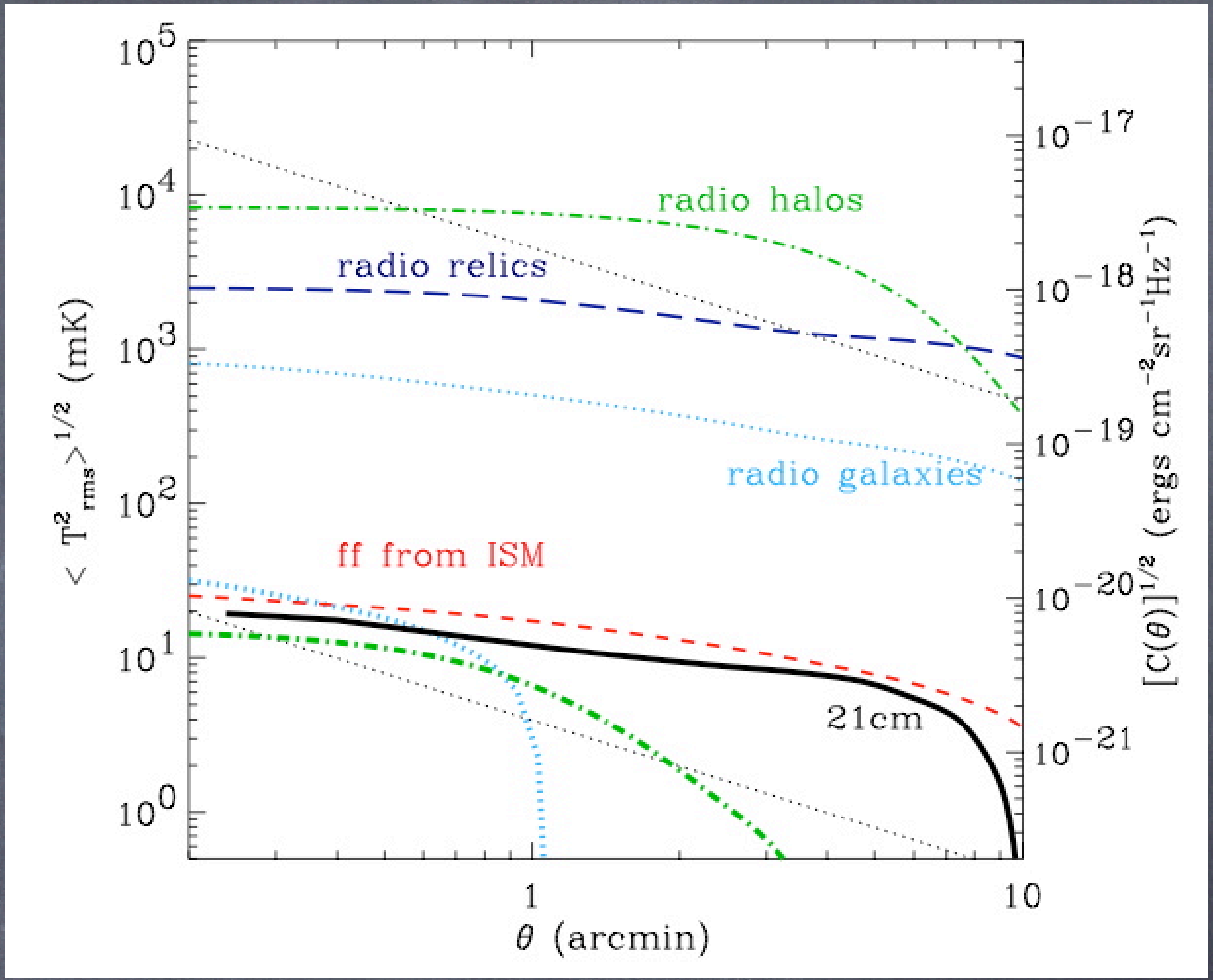
~250K at 150 MHz

2) Associated telescope noise:

$$\Delta T = 7.5 \left(\frac{1.97}{C_{\rm beam}}\right) {\rm mK} \left(\frac{A}{A_{\rm LFD}}\right)^{-1} \times \left(\frac{\Delta\nu}{1{\rm MHz}}\right)^{-1/2} \left(\frac{t_{\rm int}}{100{\rm hr}}\right)^{-1/2} \left(\frac{\Delta\theta_{\rm beam}}{5'}\right)^{-2}.$$

3) Extragalactic radio sources: DC noise~ 30K at 150 MHz

AC noise: Angular Brightness temperature fluctuations swamp 21cm signal

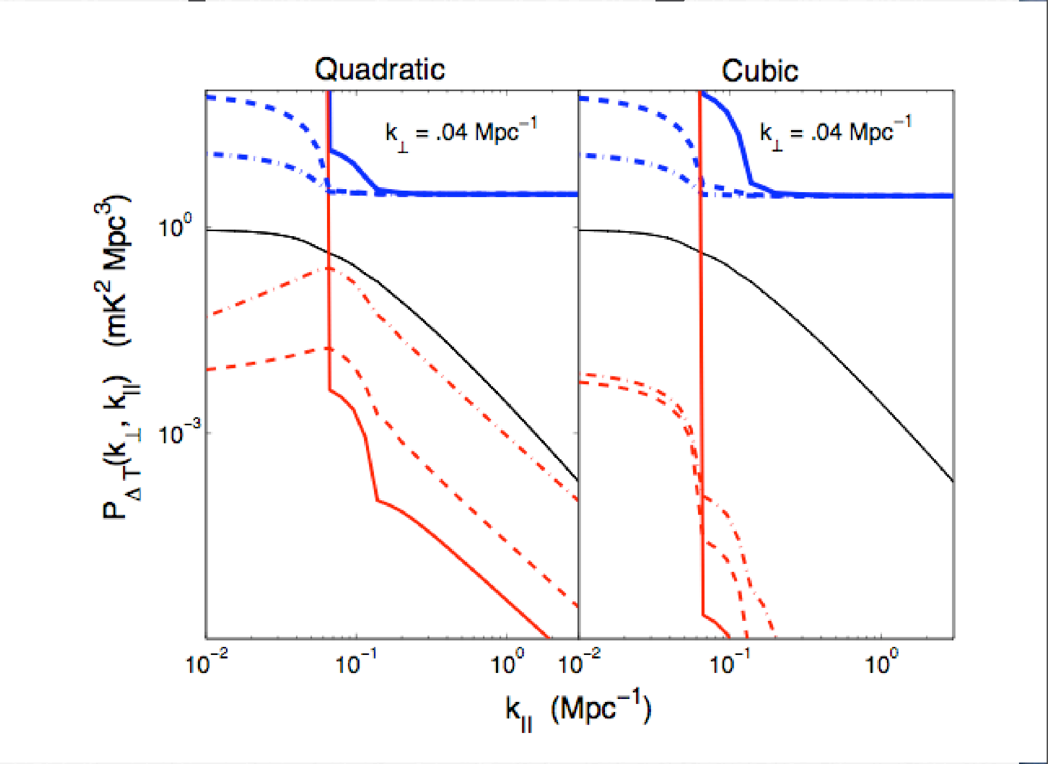


di Matteo et al (2004)

On large scales, dominated by clustering of sources
Try to reduce by point source removal....

Continuum is spectrally

- frequency channels are highly correlated
- apply trend removal: fit and subtract smooth function to data



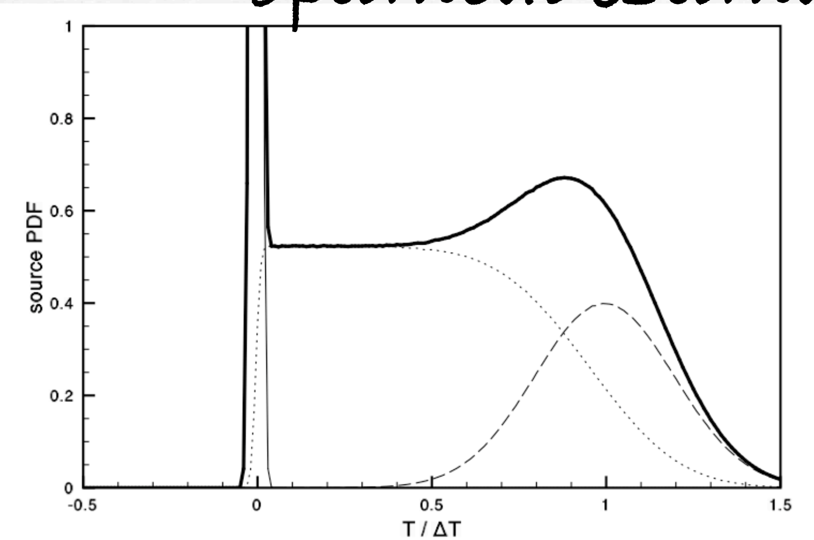
McQuinn et al 2006

But this also removes large scale power (esp for high-order fit to small length)

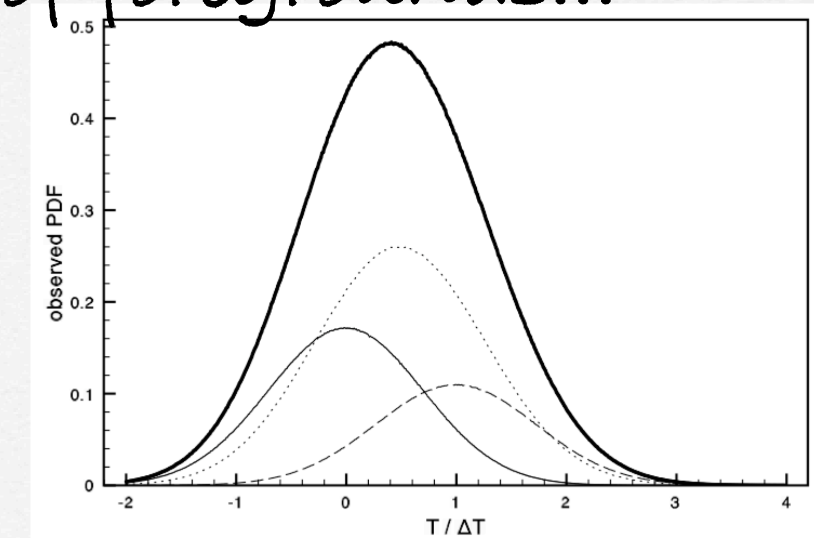
But wait! Is doing a least

Treat foreground as 'signal', 21cm + background as 'noise'....

But 'noise' PDF is NOT gaussian--least squares not an optimcal estimator of foregrounds...



no noise

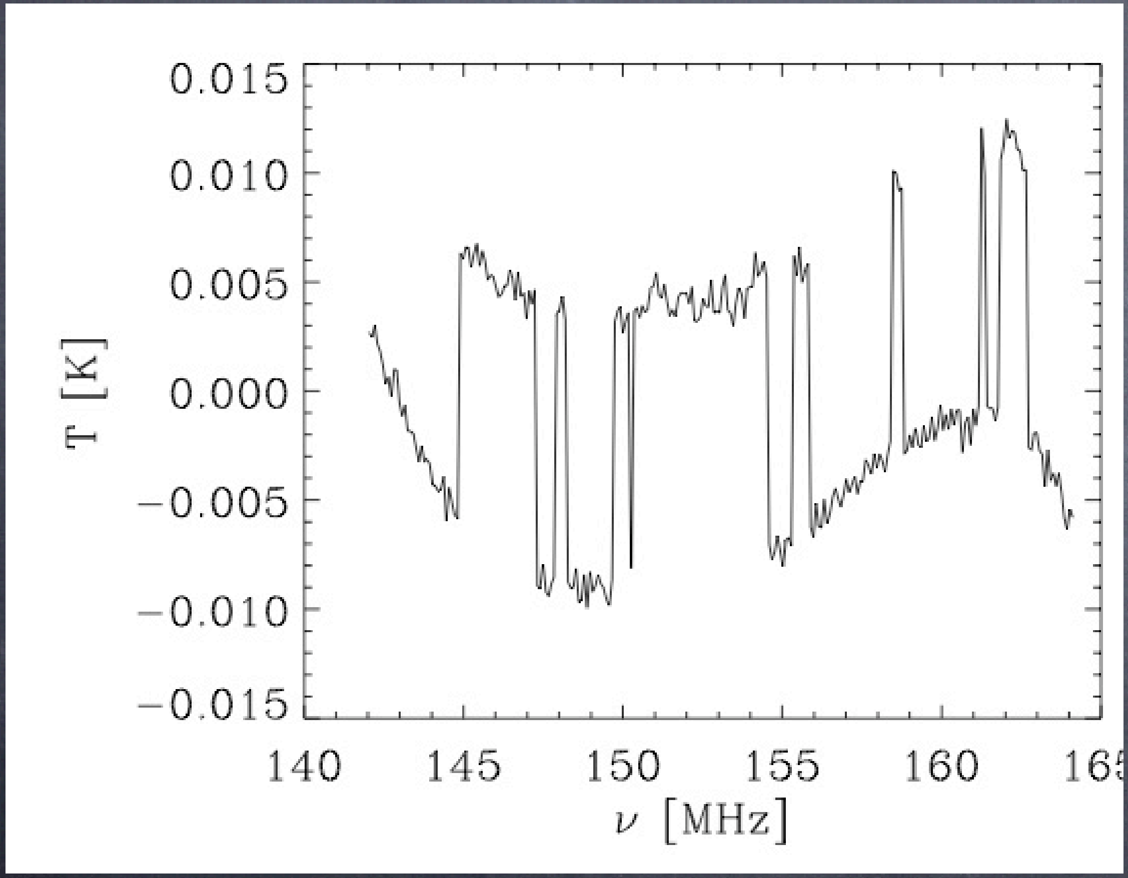


with noise

Need to use robust

- Fisher matrix error bars are underestimates
- Most apparent in high S/N data
- Optimal estimator uses prior information about PDF-- Expectation Maximization algorithms (Nada Petrovic & Oh, 2007 in prep)

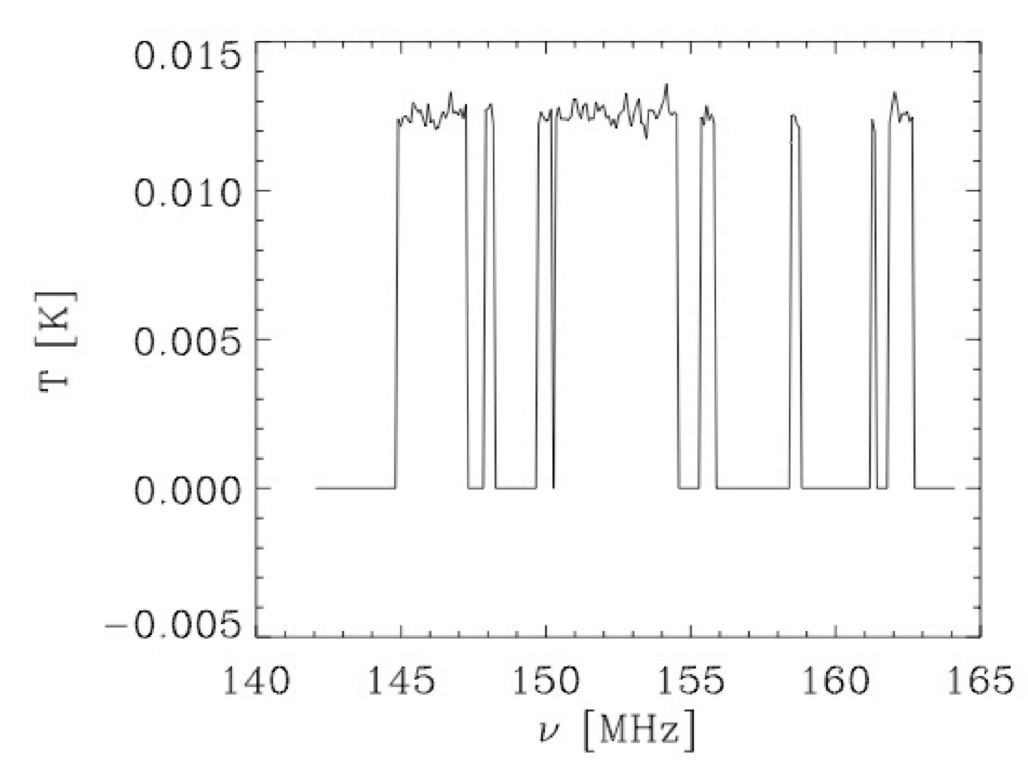
…another possible solution: Detrending with large



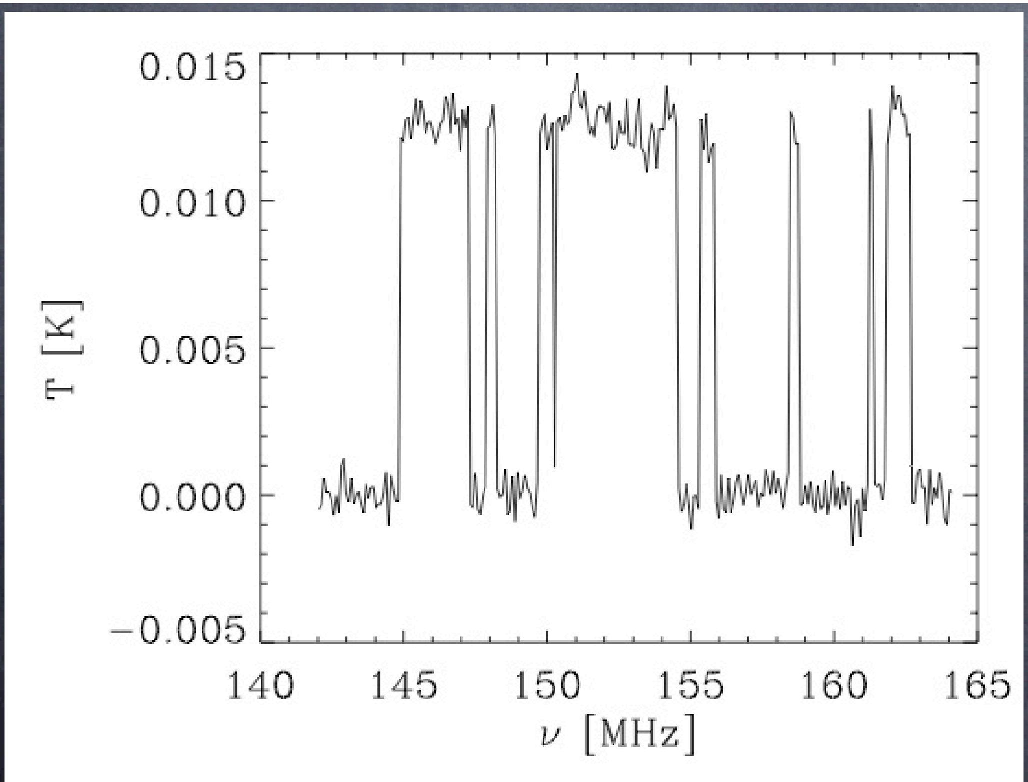
Recovered after continuum subtraction

Chang & Oh (in prep)

Bubbles are foreground only….so use the minima of recovered spectra to normalize no 21cm baseline

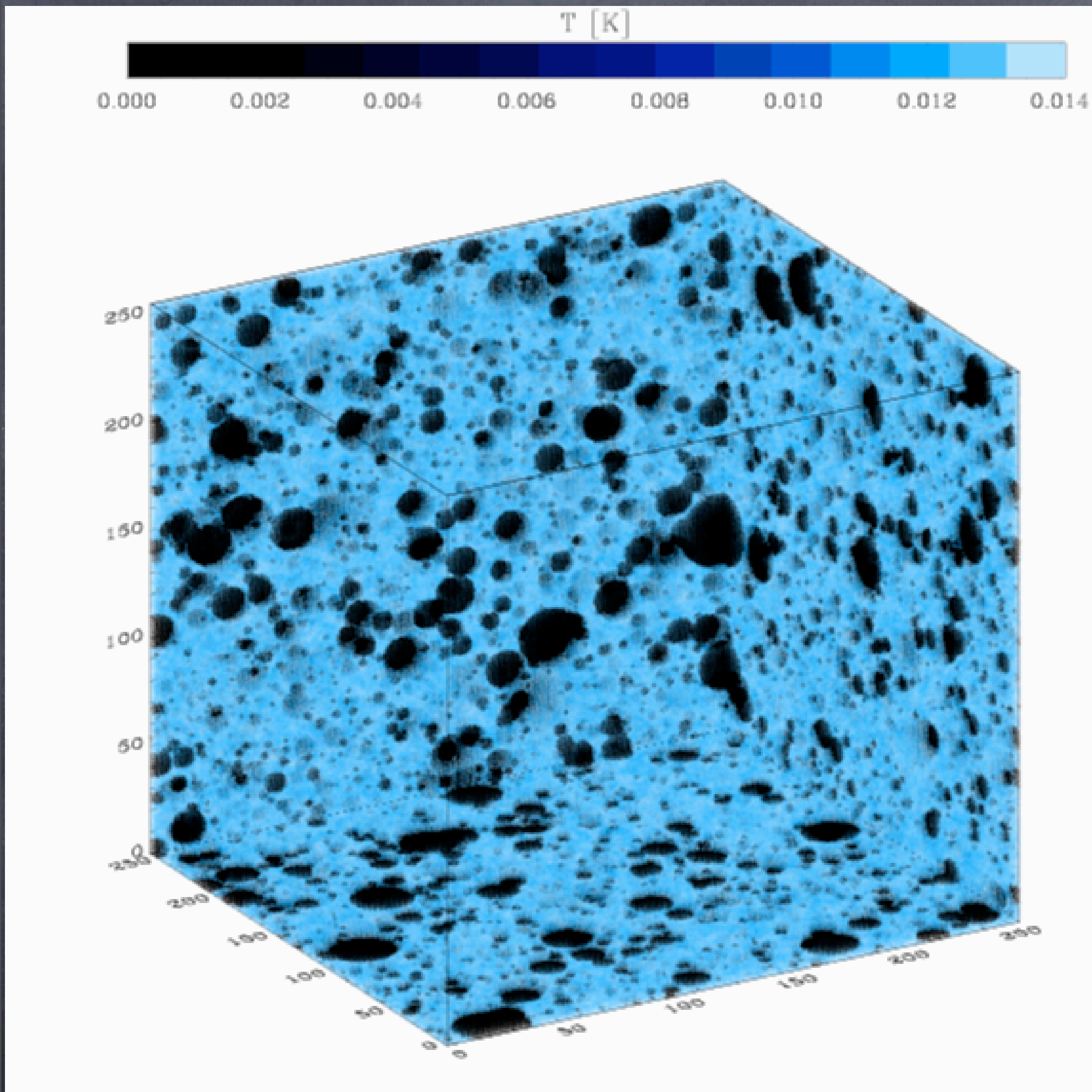


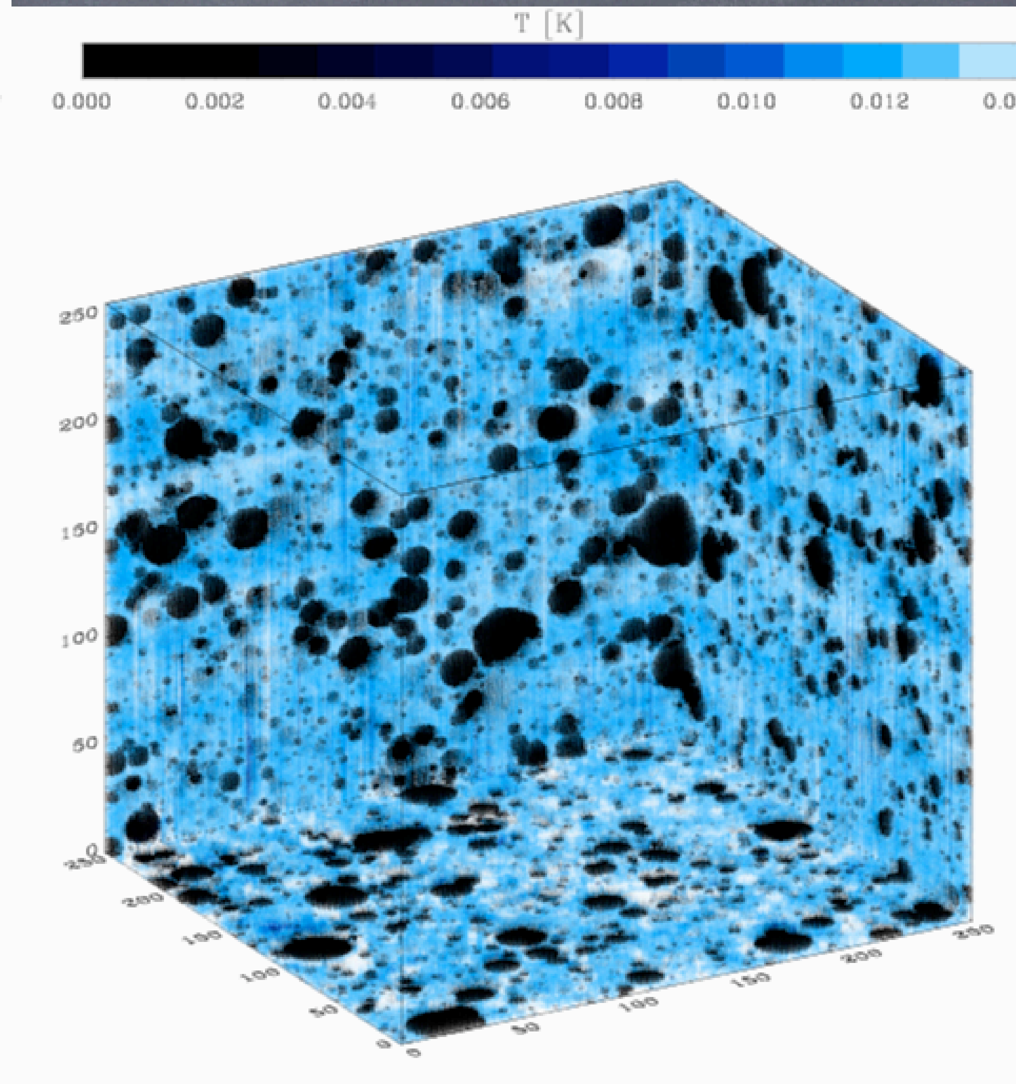
True (input) spectrum



Bubble Detrended spectrum

For an SKA type instrument...





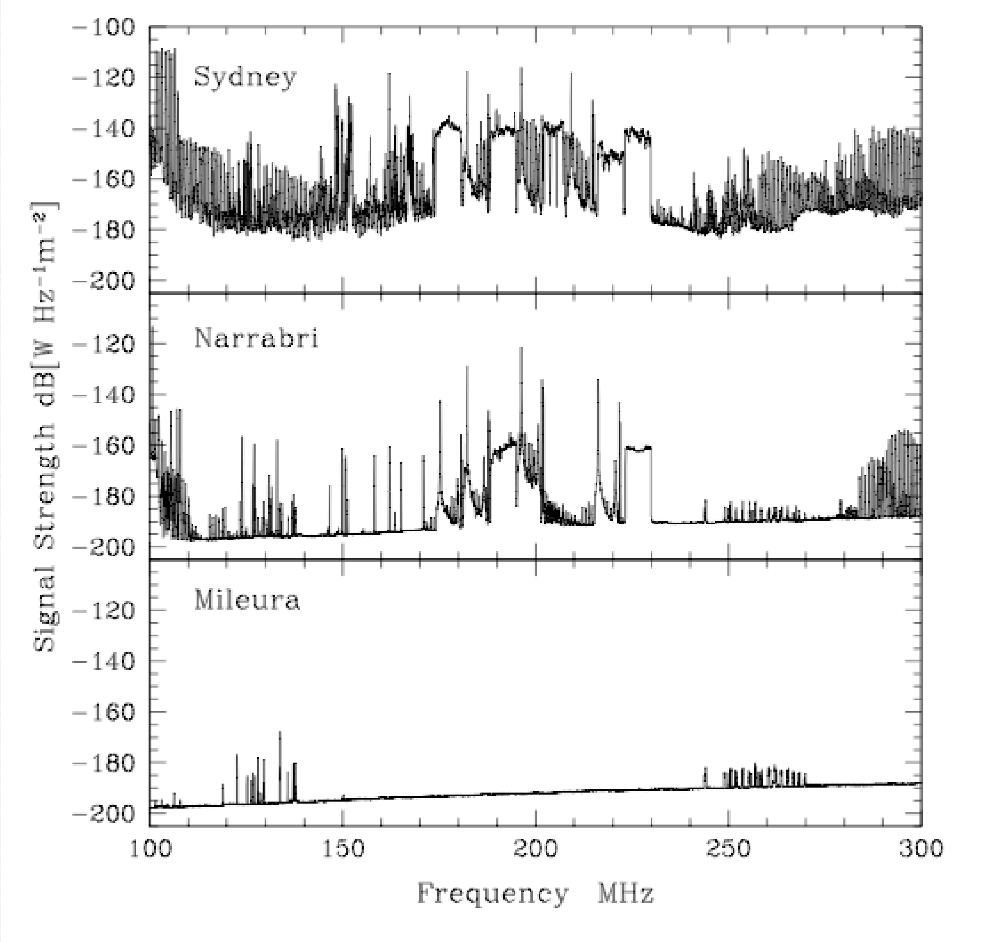
Input

Recovered

...extremely good imaging is feasible!

Much scarier: spectral

- Man-made interference
- Ionosphere
- Radio recombination lines
- Polarization/Faraday rotation
- Frequency-dependent side-lobes



FOB06

Bottom Line

- HII Bubbles are main feature (holes in 21cm emission) after first sources light up
- Much needed foreground calibrators
- Can only directly image biggest ones: sharpen detection w/ Canny algorithm
- If can detect statistically, obtain Q_HII(z)
- More work needed!