Gibbs sampling and Gaussian constrained realisations

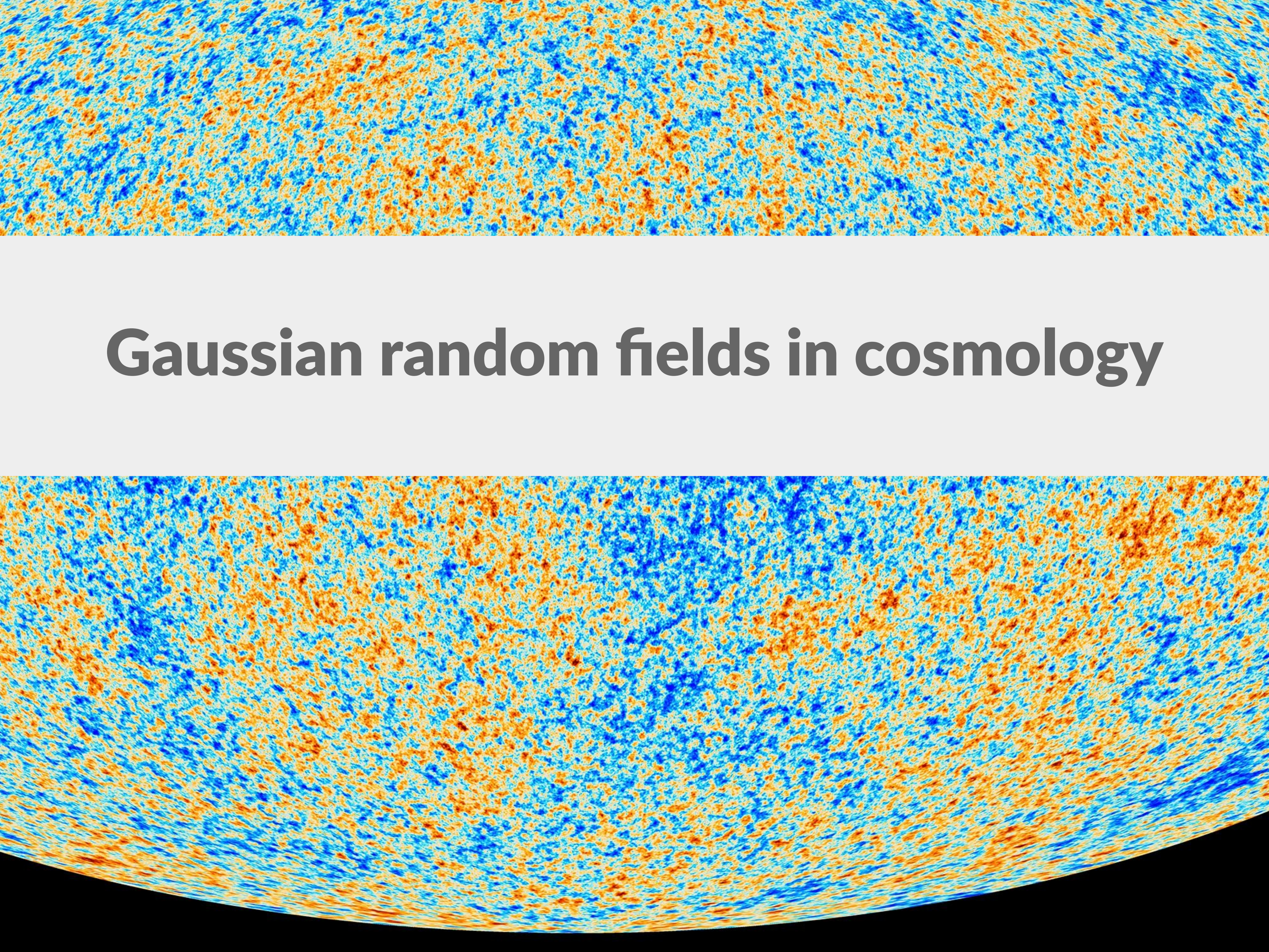
Missing data and high-dimensional inference problems

Phil Bull
QMUL

Overview

1) Gaussian random fields in cosmology
2) Gibbs sampling for high-dimensional problems
3) Gaussian constrained realisations
4) Missing data and the power spectrum

Gaussian random fields in cosmology



Gaussian random fields in cosmology

Many key observables in cosmology probe the distribution of matter (e.g. CDM, baryons, radiation) on large scales

- **Initial conditions** set by seed quantum fluctuations during inflation

    $\rightarrow$ Fluctuations random, statistically homogeneous, Gaussian

Gaussian random fields in cosmology

Many key observables in cosmology probe the distribution of matter (e.g. CDM, baryons, radiation) on large scales

- **Initial conditions** set by seed quantum fluctuations during inflation

  → Fluctuations random, statistically homogeneous, Gaussian

- Physical processes (e.g. gravitational collapse, photon free-streaming) evolve fields from these initial conditions

  → On sufficiently large scales, evolution eqns. can be **linearised**, so the fields remain ~Gaussian

Gaussian random fields in cosmology

Many key observables in cosmology probe the distribution of matter (e.g. CDM, baryons, radiation) on large scales

- **Initial conditions** set by seed quantum fluctuations during inflation

  → Fluctuations random, statistically homogeneous, Gaussian

- Physical processes (e.g. gravitational collapse, photon free-streaming) evolve fields from these initial conditions

  → On sufficiently large scales, evolution eqns. can be **linearised**, so the fields remain ~Gaussian

Key task in cosmology: Measuring the **power spectrum** (Fourier-space covariance) of approx. Gaussian random fields

Cosmic Microwave Background

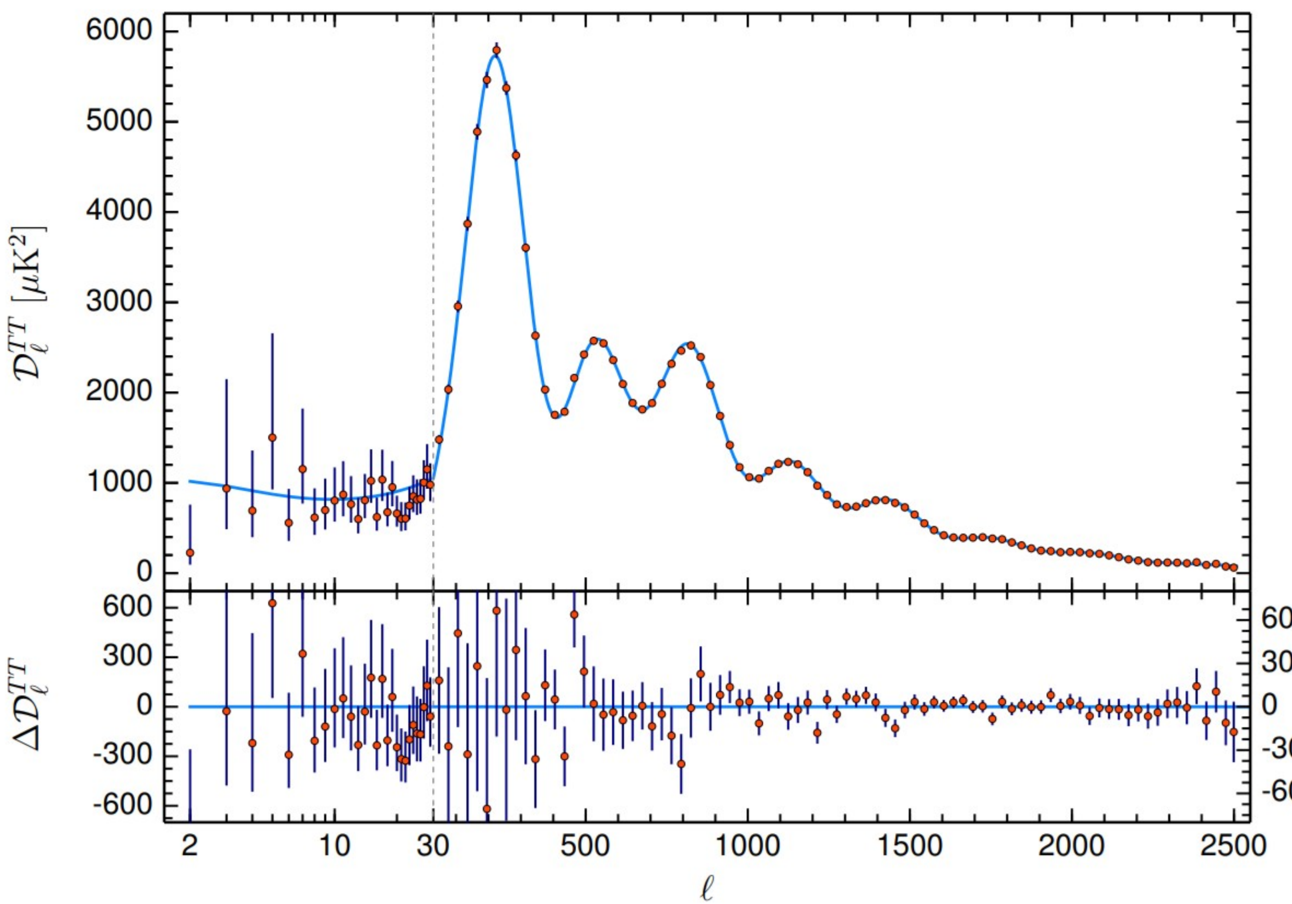
Analysis of fluctuations in the CMB radiation:

- Make map of fluctuations in CMB temperature on the sky
- Transform into spherical harmonic (l,m) basis
- Calculate power spectrum (variance as a function of l)

Cosmic Microwave Background

Analysis of fluctuations in the CMB radiation:

- Make map of fluctuations in CMB temperature on the sky
- Transform into spherical harmonic (l,m) basis
- Calculate power spectrum (variance as a function of l)

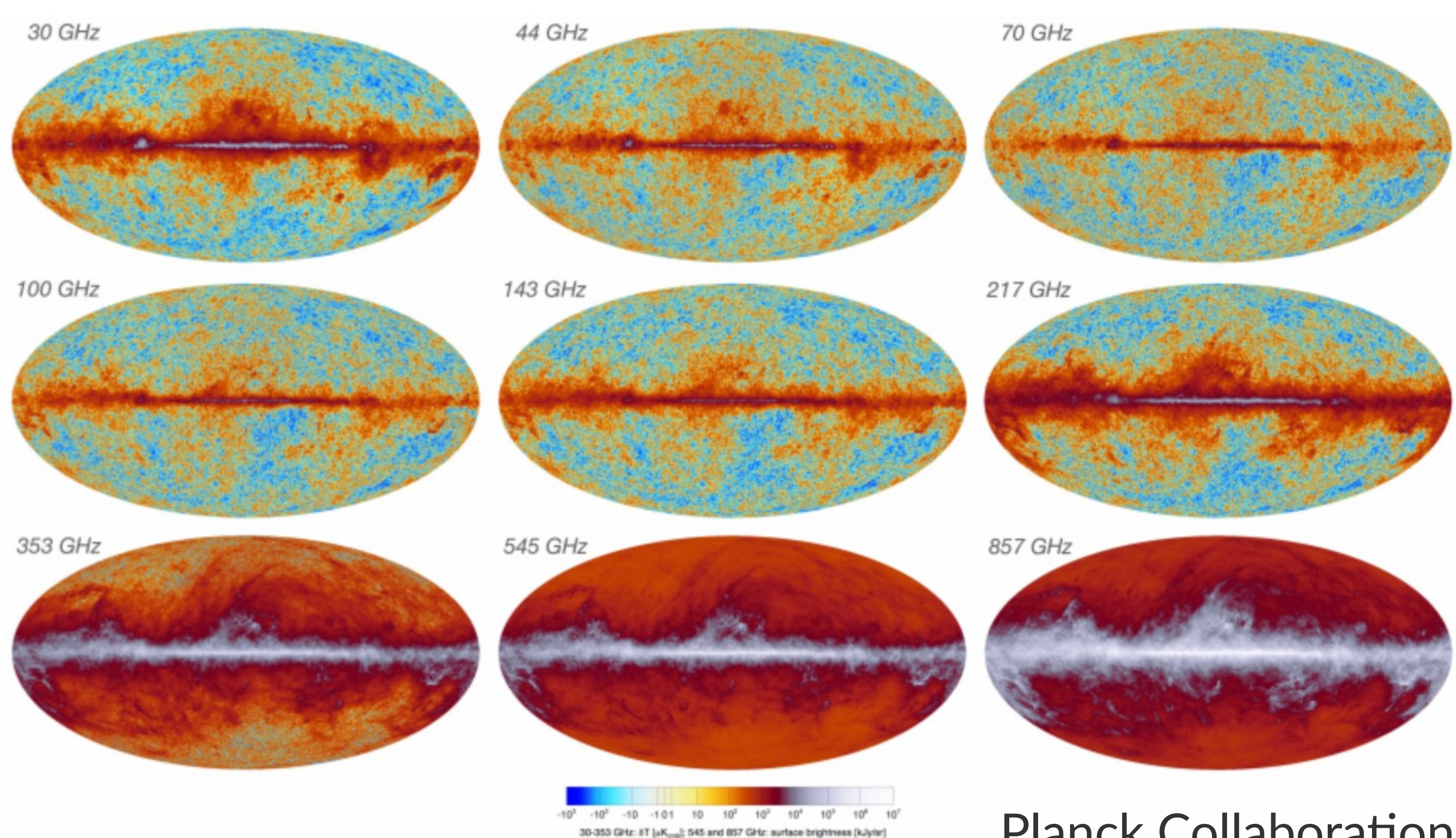


Planck Collab. (2018)

Cosmic Microwave Background

Planck CMB data are microwave maps over the whole sky in multiple frequency bands contaminated by foreground emission.

- 9x frequency bands, 3x polarisations
- Up to 50 million pixels per band per polarisation

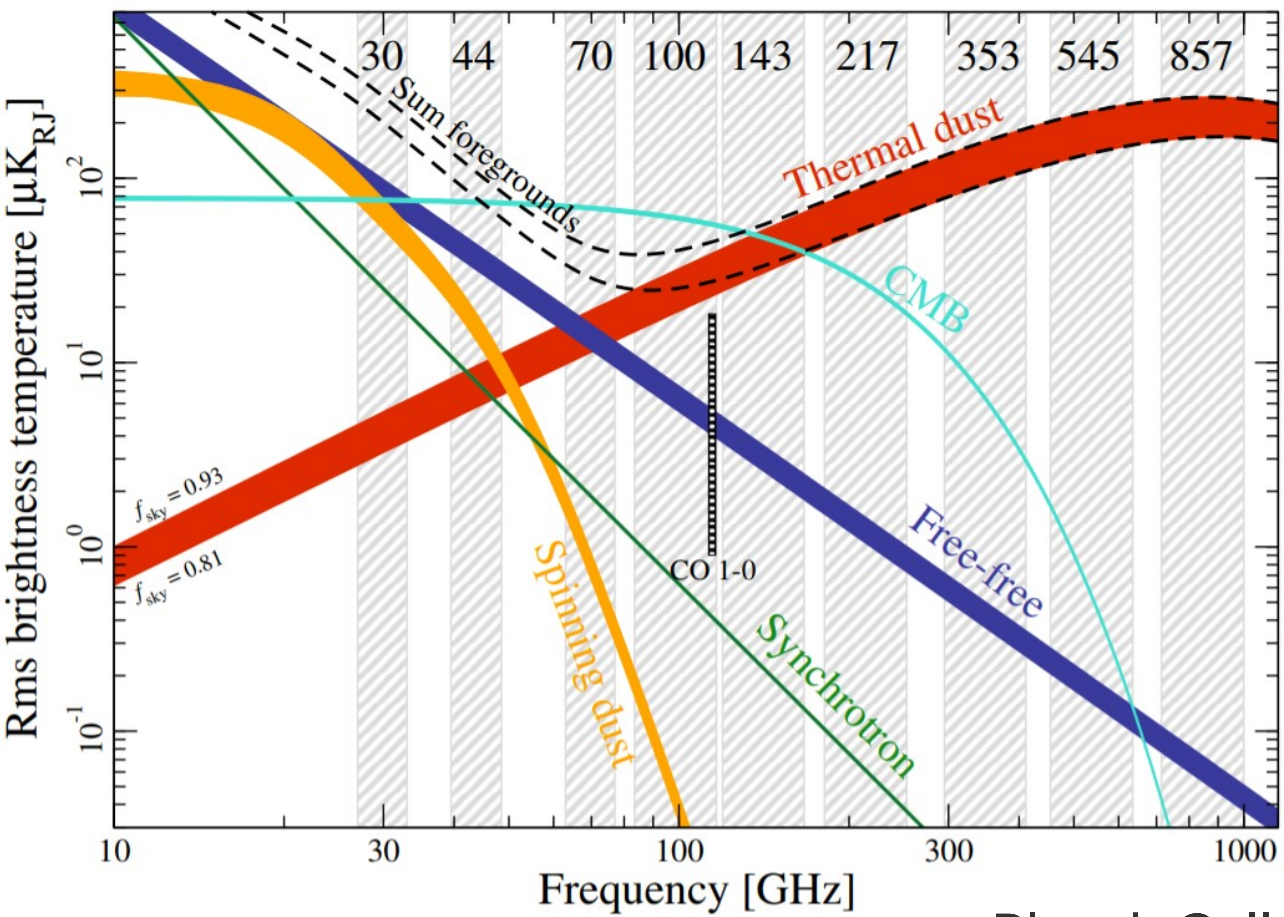


Planck Collaboration (2018)

Cosmic Microwave Background

How to separate the foregrounds from the primary CMB?

- Simple freq.-dependent parametric model for each foreground
- Foreground and CMB parameters vary from pixel to pixel



Planck Collaboration (2018)

CMB data model

Generic data model:

Spatial basis function

Amplitude coefficient

Data

Noise

$$\mathbf{d}(\nu) = \mathbf{B}(\nu) \sum_{i=1}^{N_{\mathrm{comp}}} \mathbf{G}_i(\nu)\mathbf{T}_i\mathbf{a}_i + \mathbf{n}(\nu)$$

Sum over components

Spectral model

CMB data model

Generic data model:

$$\mathbf{d}(\nu) = \mathbf{B}(\nu) \sum_{i=1}^{N_{\mathrm{comp}}} \mathbf{G}_i(\nu)\mathbf{T}_i\mathbf{a}_i + \mathbf{n}(\nu)$$

Data

Spatial basis function

Amplitude coefficient

Noise

Sum over components

Spectral model

Typical component has the following parameters (per pixel):

- 1 amplitude per polarisation
- 1–2 spectral parameters (across all polarisations)

CMB data model

Generic data model:

$$\mathbf{d}(\nu) = \mathbf{B}(\nu) \sum_{i=1}^{N_{\mathrm{comp}}} \mathbf{G}_i(\nu)\mathbf{T}_i\mathbf{a}_i + \mathbf{n}(\nu)$$

- Data
- Spatial basis function
- Amplitude coefficient
- Noise
- Sum over components
- Spectral model

Typical component has the following parameters (per pixel):

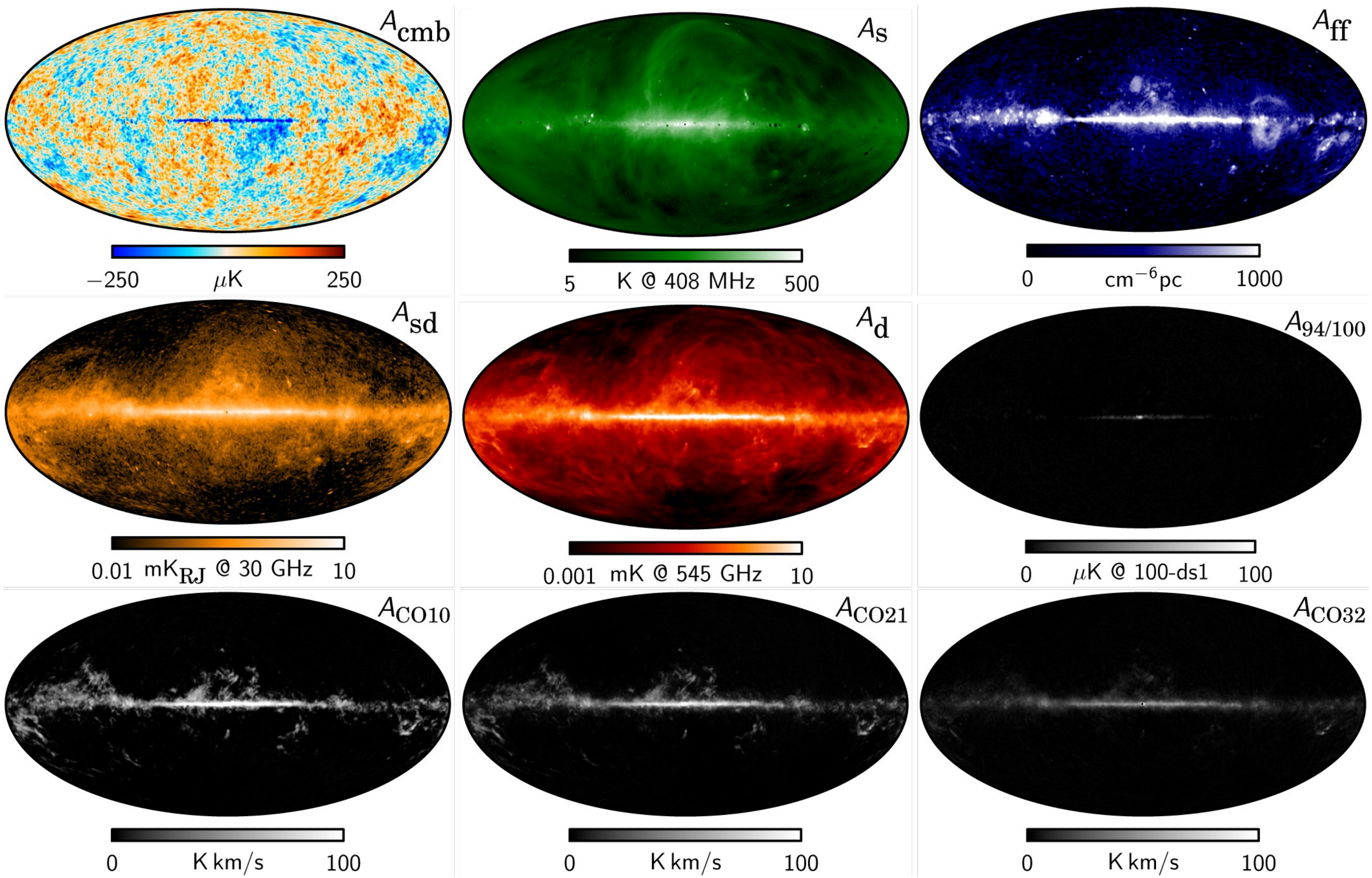
- 1 amplitude per polarisation
- 1–2 spectral parameters (across all polarisations)

Planck analysis: 15 parameters **per pixel**!

Cosmic Microwave Background

One of the major successes of Planck has been creating maps of the foreground components themselves

Planck Collaboration (2015)



CMB data model

How to estimate the posterior of ~15 parameters/px for over ~1 million pixels!?

- Rely on independence of noise between pixels?

  Only approximately true for Planck

- CMB prior is in spherical harmonic space

  Need prior to regularise in case of missing data, but couples pixels at large separations!

CMB data model

How to estimate the posterior of ~15 parameters/px for over ~1 million pixels!?

- Rely on independence of noise between pixels?

  Only approximately true for Planck

- CMB prior is in spherical harmonic space

  Need prior to regularise in case of missing data, but couples pixels at large separations!

This is **not** an embarrassingly parallel problem!

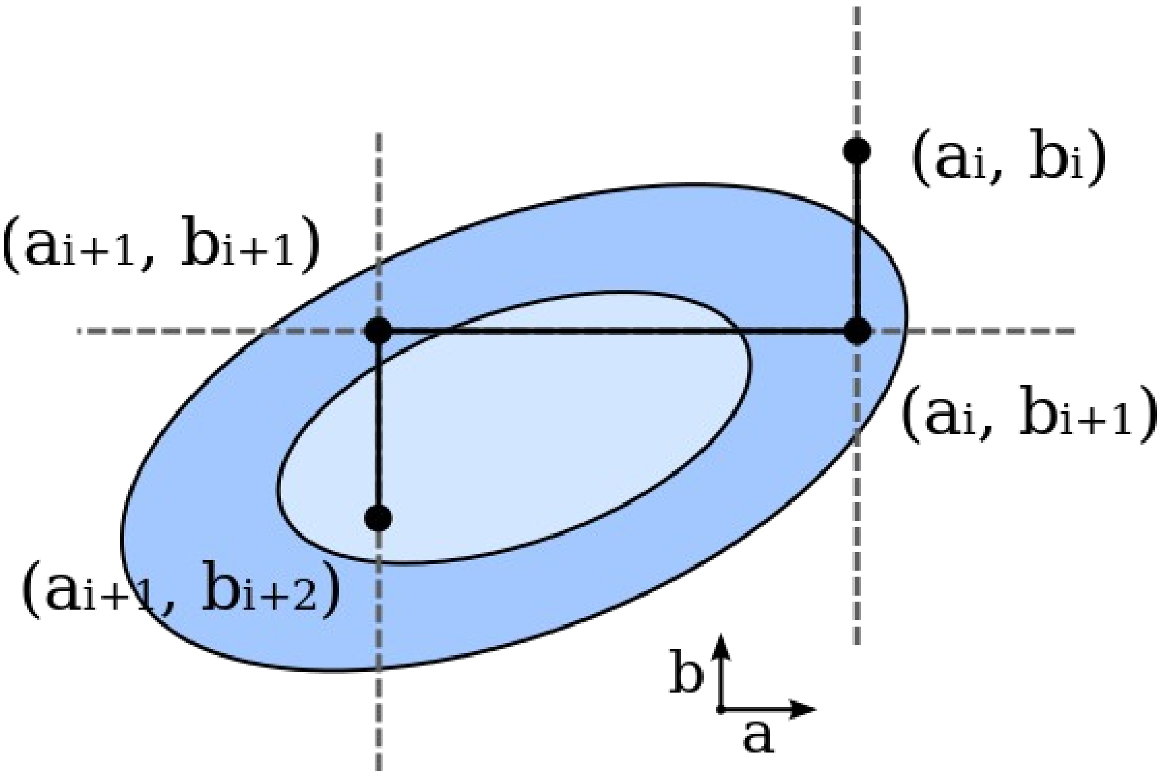
Need a clever way of estimating high-dimensional posterior with non-trivial correlations between parameters

Gibbs Sampling

Gibbs sampling

Sample from **joint posterior** by iteratively sampling from **conditional distributions**

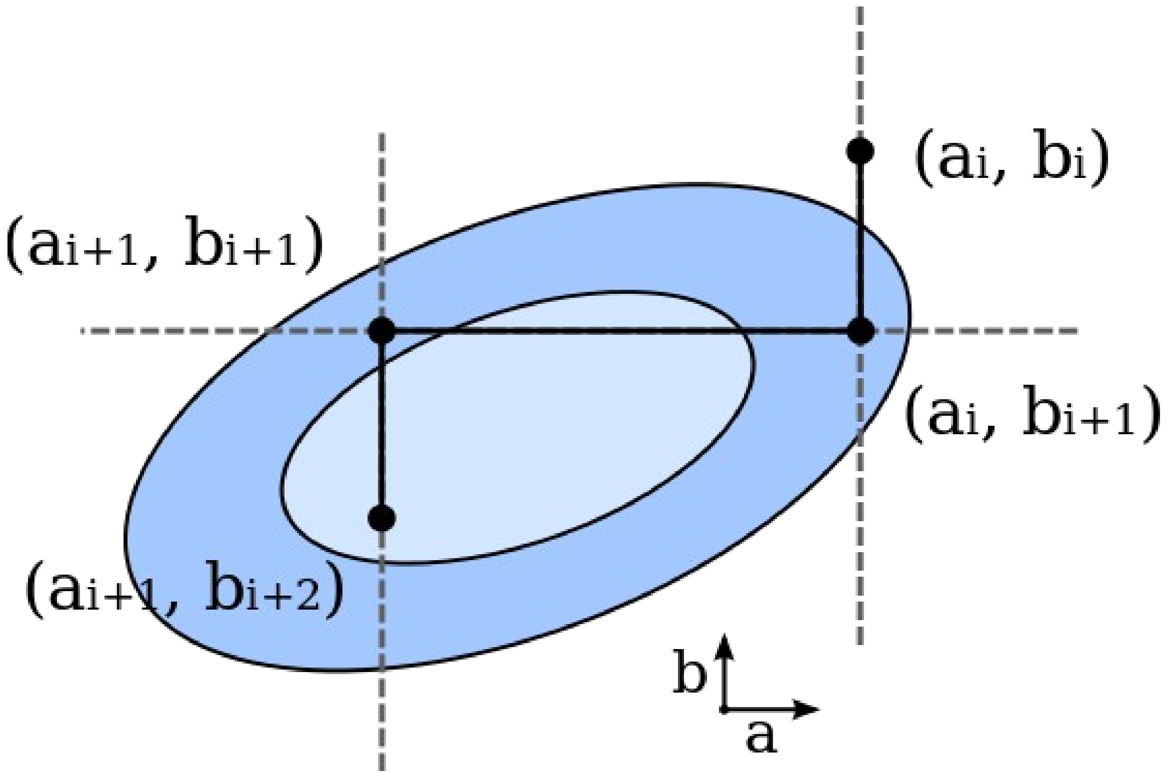
```
P(a,b|d) ~
    Iterate:
      a ← P(a|b,d)
      b ← P(b|a,d)
```



Gibbs sampling

Sample from **joint posterior** by iteratively sampling from **conditional distributions**

```
P(a,b|d) ~
    Iterate:
      a ← P(a|b,d)
      b ← P(b|a,d)
```



Useful if conditionals are **simple** → use **direct** sampling

- Sampling from (e.g.) multivariate Gaussian is "easy"!
- Sampling from general m.v. dists can be **very** hard

Gibbs sampling

Data model:

$$\mathbf{d}(\nu) = \mathbf{B}(\nu) \sum_{i=1}^{N_{\mathrm{comp}}} \mathbf{G}_i(\nu)\mathbf{T}_i\mathbf{a}_i + \mathbf{n}(\nu)$$

Iterations for our problem:

Amplitudes of components $\mathbf{a}^{i+1} \leftarrow P(\mathbf{a}|\mathbf{S}^i, \mathbf{G}^i, \mathbf{T}^i, \mathbf{d})$

Amplitude covariance $\mathbf{S}^{i+1} \leftarrow P(\mathbf{S}|\mathbf{a}^{i+1}, \mathbf{G}^i, \mathbf{T}^i, \mathbf{d})$

Spectral parameters $\mathbf{G}^{i+1} \leftarrow P(\mathbf{G}|\mathbf{a}^{i+1}, \mathbf{S}^{i+1}, \mathbf{T}^i, \mathbf{d})$

Spatial template params $\mathbf{T}^{i+1} \leftarrow P(\mathbf{T}|\mathbf{a}^{i+1}, \mathbf{S}^{i+1}, \mathbf{G}^{i+1}, \mathbf{d})$

Limitations of Gibbs sampling

- Only efficient if conditional distributions are tractable!
- Avoid sampling correlated parameters in separate steps (otherwise MCMC samples are highly correlated)
- Iterative methods can take a long time to converge if starting point is bad
- Generally much heavier than cheating (i.e. using approximate methods)

Constrained Realisations

Constrained realisations

Conditional dist. for **all** amplitude parameters can be written as a single multivariate Gaussian

$$P(\mathbf{a}|\mathbf{d},\mathbf{S},\mathbf{G},\mathbf{T}) \propto P(\mathbf{d}|\mathbf{a},\mathbf{S},\mathbf{G},\mathbf{T})P(\mathbf{a}|\mathbf{S},\mathbf{G},\mathbf{T})$$

$$\propto e^{-\frac{1}{2}(\mathbf{d}-\mathbf{U}\cdot\mathbf{a})^T\mathbf{N}^{-1}(\mathbf{d}-\mathbf{U}\cdot\mathbf{a})} \cdot e^{-\frac{1}{2}\mathbf{a}^T\mathbf{S}^{-1}\mathbf{a}}$$

$$\propto e^{-\frac{1}{2}(\mathbf{a}-\hat{\mathbf{d}})^T\left(\mathbf{S}^{-1}+\mathbf{U}^T\mathbf{N}^{-1}\mathbf{U}\right)(\mathbf{a}-\hat{\mathbf{d}})}$$

(where U = BGT projects amplitudes → map)

Constrained realisations

Conditional dist. for all amplitude parameters can be written as a single multivariate Gaussian

$$P(\mathbf{a}|\mathbf{d},\mathbf{S},\mathbf{G},\mathbf{T}) \propto P(\mathbf{d}|\mathbf{a},\mathbf{S},\mathbf{G},\mathbf{T})P(\mathbf{a}|\mathbf{S},\mathbf{G},\mathbf{T})$$
$$\propto e^{-\frac{1}{2}(\mathbf{d}-\mathbf{U}\cdot\mathbf{a})^T\mathbf{N}^{-1}(\mathbf{d}-\mathbf{U}\cdot\mathbf{a})} \cdot e^{-\frac{1}{2}\mathbf{a}^T\mathbf{S}^{-1}\mathbf{a}}$$
$$\propto e^{-\frac{1}{2}(\mathbf{a}-\hat{\mathbf{d}})^T\left(\mathbf{S}^{-1}+\mathbf{U}^T\mathbf{N}^{-1}\mathbf{U}\right)(\mathbf{a}-\hat{\mathbf{d}})}$$

(where U = BGT projects amplitudes → map)

Sample **directly** from this by solving linear system

Prior info included by conditioning on signal covariance
→ **Constrained realisation**: fills in missing data etc.

Posterior for amplitude parameters

Given (incomplete) data + expected signal and noise covariance, the amplitude parameters follow this posterior pdf:

$$p(\mathbf{x}|\mathbf{S},\mathbf{N},\mathbf{d}) \propto p(\mathbf{d}|\mathbf{x},\mathbf{S},\mathbf{N})\, p(\mathbf{x}|\mathbf{S})$$

$$\propto \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{T}\cdot\mathbf{x})^T\mathbf{N}^{-1}(\mathbf{T}\cdot\mathbf{x}-\mathbf{d})\right)$$

$$\times \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{S}^{-1}\mathbf{x}\right)$$

Posterior for amplitude parameters

Given (incomplete) data + expected signal and noise covariance, the amplitude parameters follow this posterior pdf:

$$
\begin{aligned}
p(\mathbf{x}|\mathbf{S},\mathbf{N},\mathbf{d}) &\propto p(\mathbf{d}|\mathbf{x},\mathbf{S},\mathbf{N})\, p(\mathbf{x}|\mathbf{S}) \\
&\propto \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{T}\cdot\mathbf{x})^T\mathbf{N}^{-1}(\mathbf{T}\cdot\mathbf{x}-\mathbf{d})\right) \\
&\times \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{S}^{-1}\mathbf{x}\right)
\end{aligned}
$$

The max. likelihood solution is the **Wiener filter**:

$$\hat{x} = (\mathbf{S}^{-1} + \mathbf{T}^T\mathbf{N}^{-1}\mathbf{T})^{-1}\,\mathbf{T}^T\mathbf{N}^{-1}\mathbf{d}$$

CR equations

The Wiener filter solution gives us the mean of the target Gaussian distribution:

$$\hat{\boldsymbol{x}} = (\mathbf{S}^{-1} + \mathbf{T}^T \mathbf{N}^{-1} \mathbf{T})^{-1} \, \mathbf{T}^T \mathbf{N}^{-1} \mathbf{d}$$

CR equations

The Wiener filter solution gives us the mean of the target Gaussian distribution:

$$\hat{x} = (\mathbf{S}^{-1} + \mathbf{T}^T\mathbf{N}^{-1}\mathbf{T})^{-1}\,\mathbf{T}^T\mathbf{N}^{-1}\mathbf{d}$$

We can draw a sample **x** from a multivariate Gaussian by solving the following linear system:

$$\mathbf{M} \cdot \boldsymbol{x} = \mathbf{b}$$

where:

White noise vectors

$$\mathbf{M} \equiv \mathbf{S}^{-1} + \mathbf{T}^T\mathbf{N}^{-1}\mathbf{T}$$

$$\mathbf{b} \equiv \mathbf{T}^T\mathbf{N}^{-1}\mathbf{d} + \mathbf{S}^{-\frac{1}{2}}\omega_0 + (\mathbf{T}^T\mathbf{N}^{-1}\mathbf{T})^{\frac{1}{2}}\omega_1$$

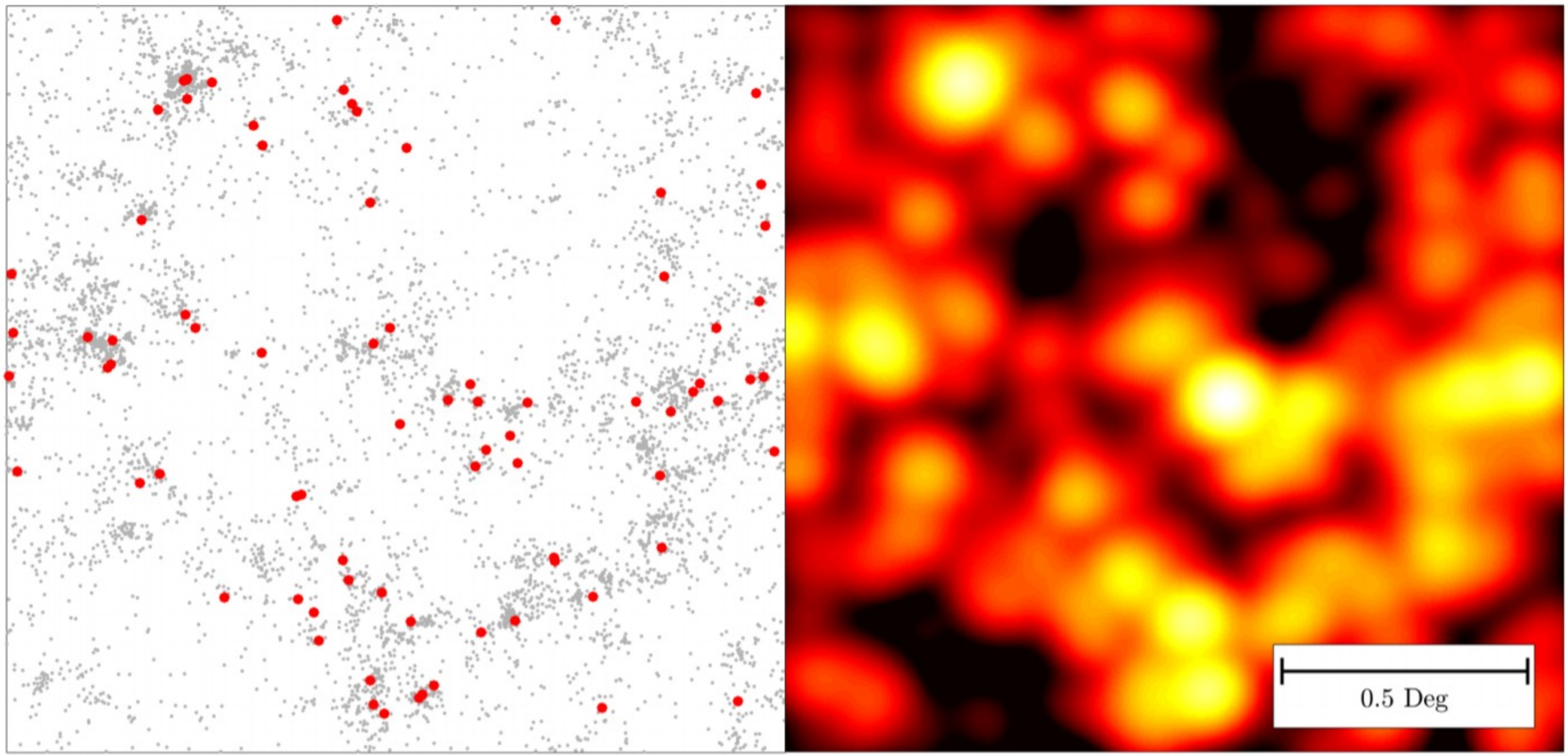
This can be solved for <u>millions</u> of **x** parameters (e.g using CG)

Example application:

21cm intensity mapping

21cm intensity mapping experiments

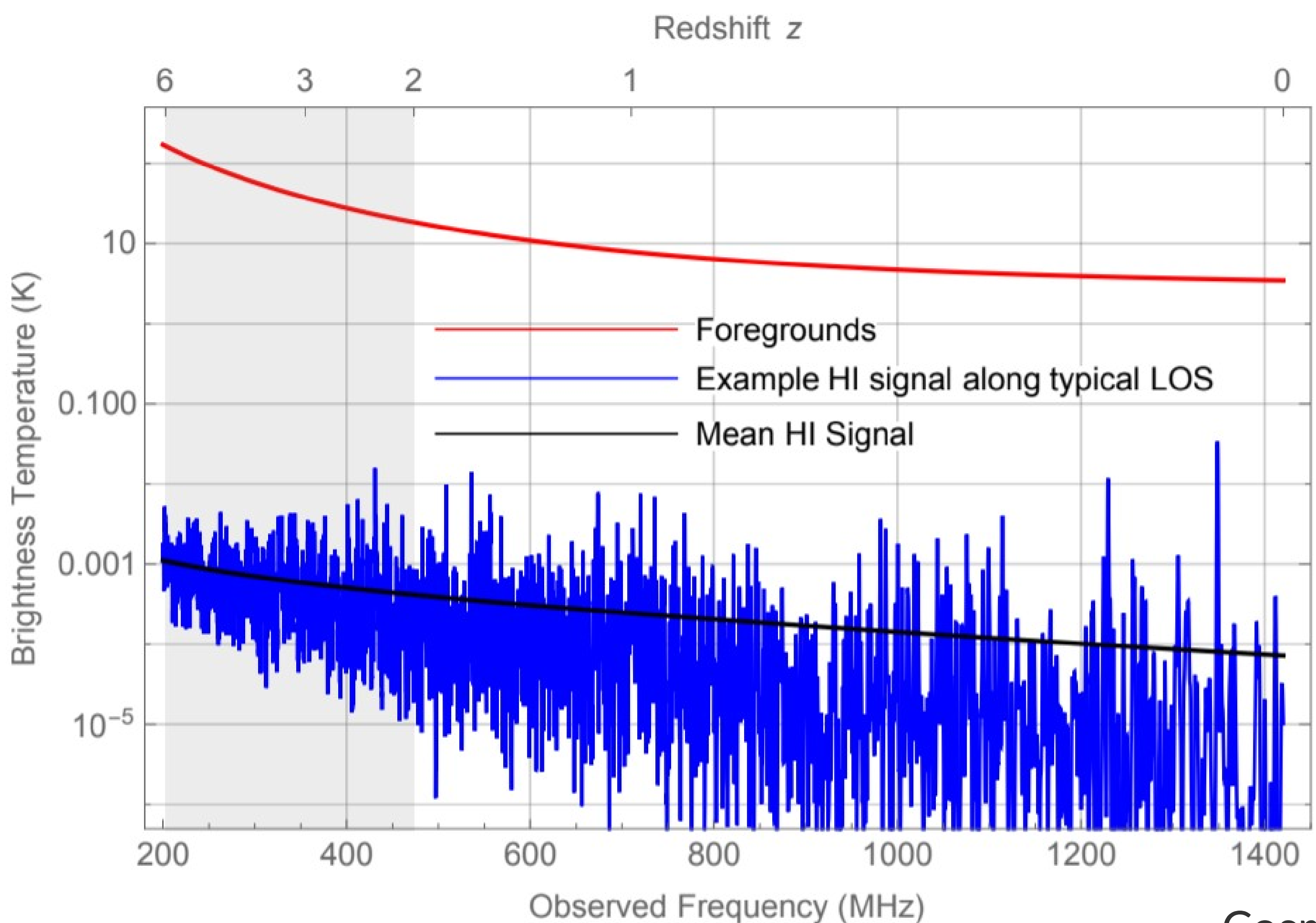
Make **3D maps** of the matter distribution using spectral line emission from galaxies etc.



Kovetz+ (2017)

21cm intensity mapping experiments

Intensity mapping is a **very** high dynamic range problem.
Foregrounds are ~10^5-10^6 larger than the cosmic signal!

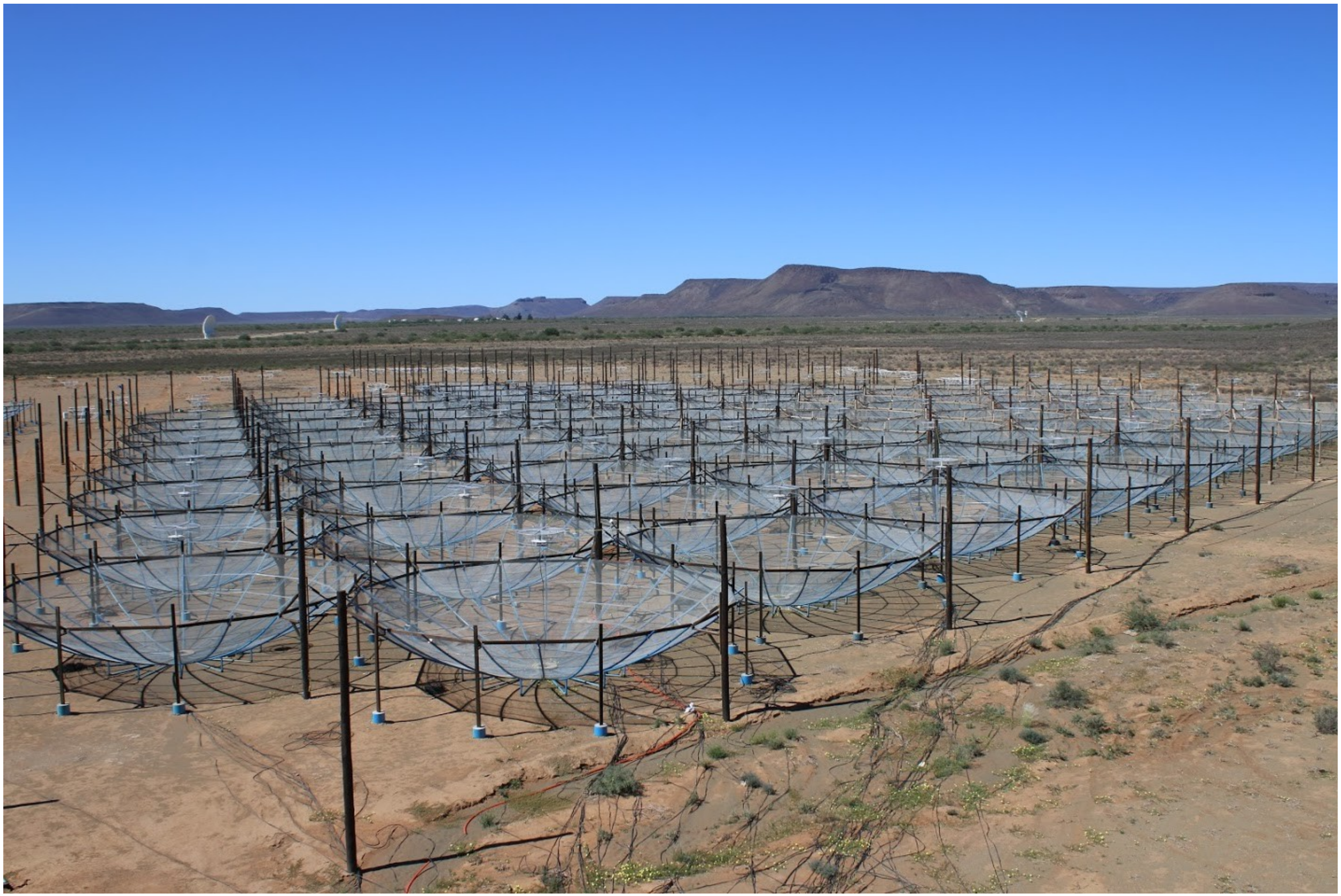


Cosmic Visions 21cm Collab. (2018)

Foregrounds are inherently spectrally smooth, but radio telescopes have a highly chromatic response → imposes spectral structure

21cm intensity mapping experiments

(QMUL just joined the HERA collaboration – ask me for details!)



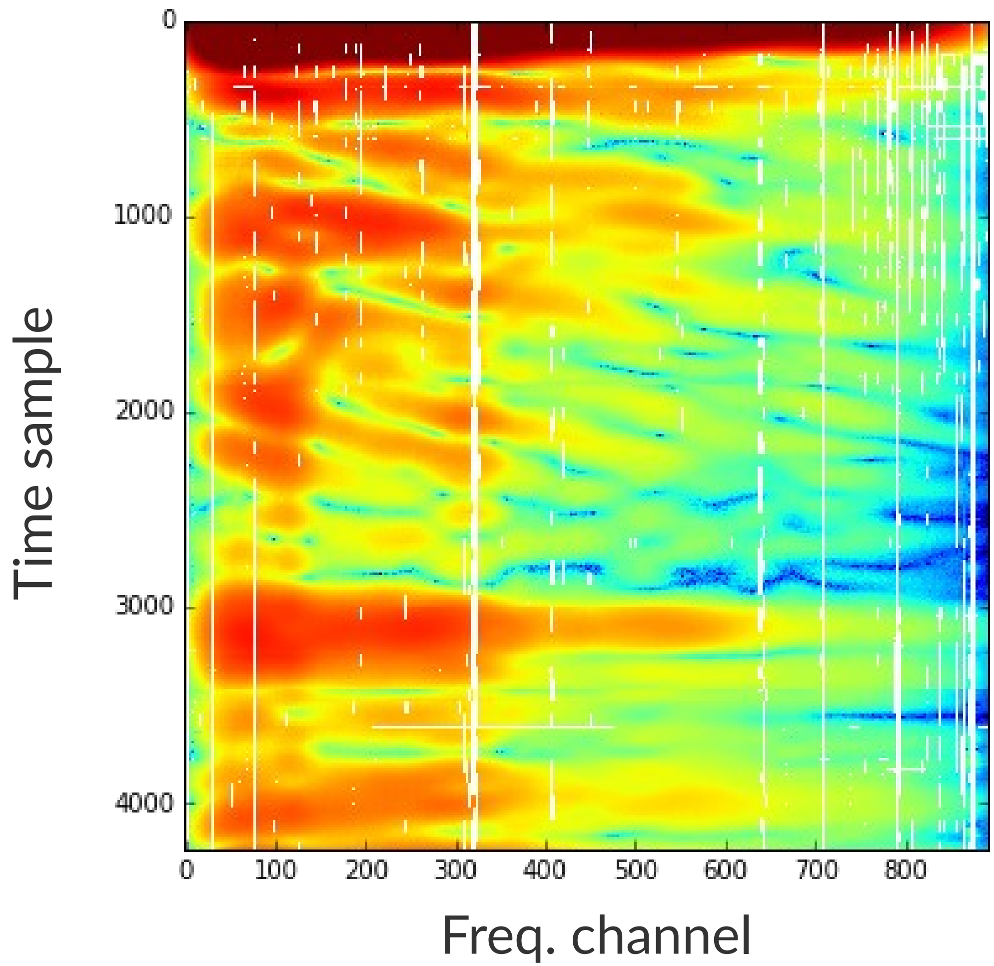
HERA Collaboration / K. Rosie

Masked data

There is always a mask, due to RFI flags and band edges.

- Sharp features break orthonormality of the Fourier basis
- This induces "ringing" and mode-coupling

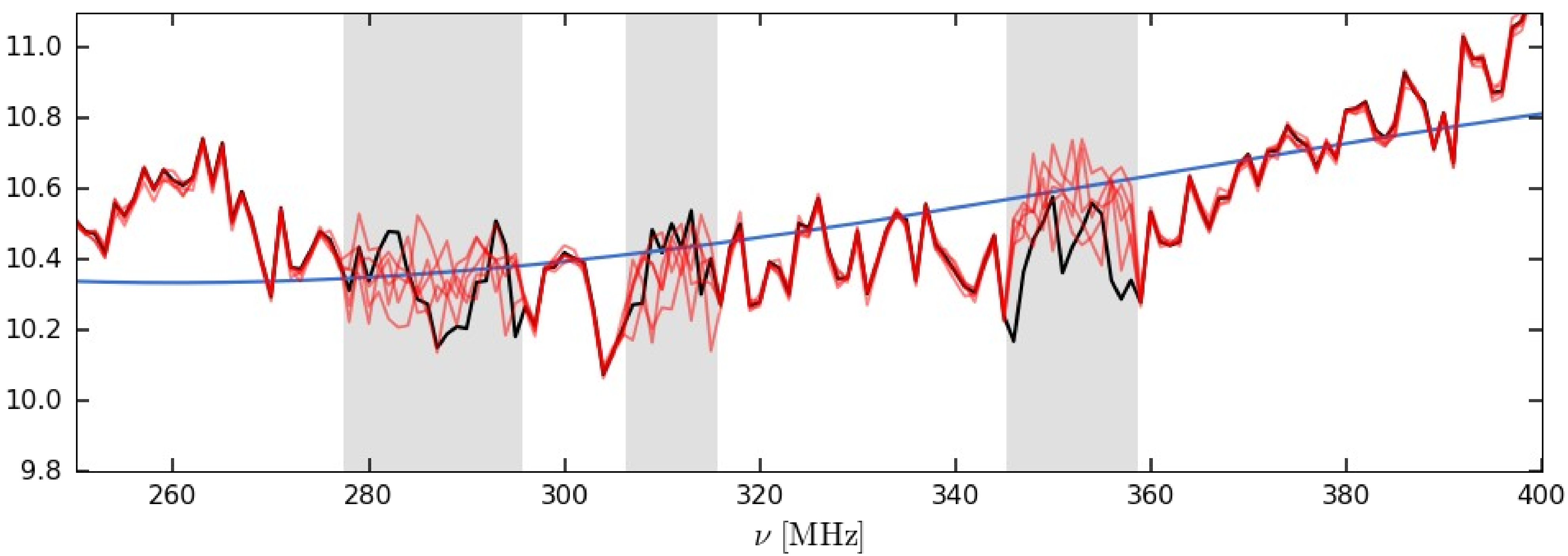
  → Couples bright foregrounds into signal-dominated modes!



Constrained realizations

<u>Data model</u>:

- **Foreground model:** 25-order polynomial
- **Signal model:** Gaussian-distributed random number in each pixel (with some prior on its power spectrum)
- **Noise:** White noise (uncorrelated Gaussian)

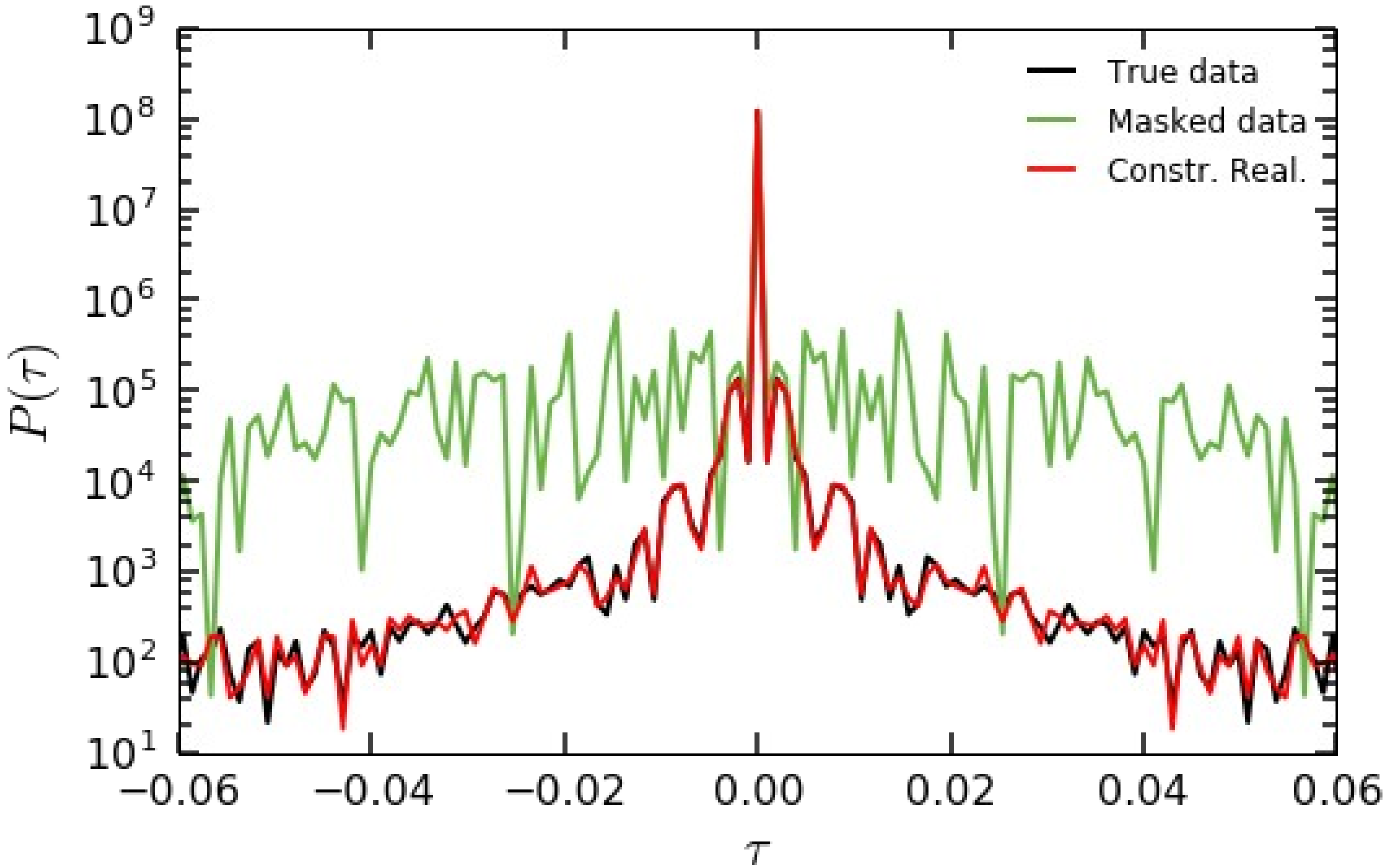


Constrained realizations

- Inspect leakage of bright foreground modes outside of the “smooth” (low Fourier mode) region

    Green: Severe ringing caused by mask

    Red: Constrained realisation inside mask



Summary

- Lots of ~Gaussian random fields in cosmology
- **Gibbs sampling:** Split-up posterior in a clever way to make high-dimensional problems more tractable
- **Gaussian constrained realisations:** elegant solution to estimating power spectra in presence of missing data