An introduction to CMB data analysis

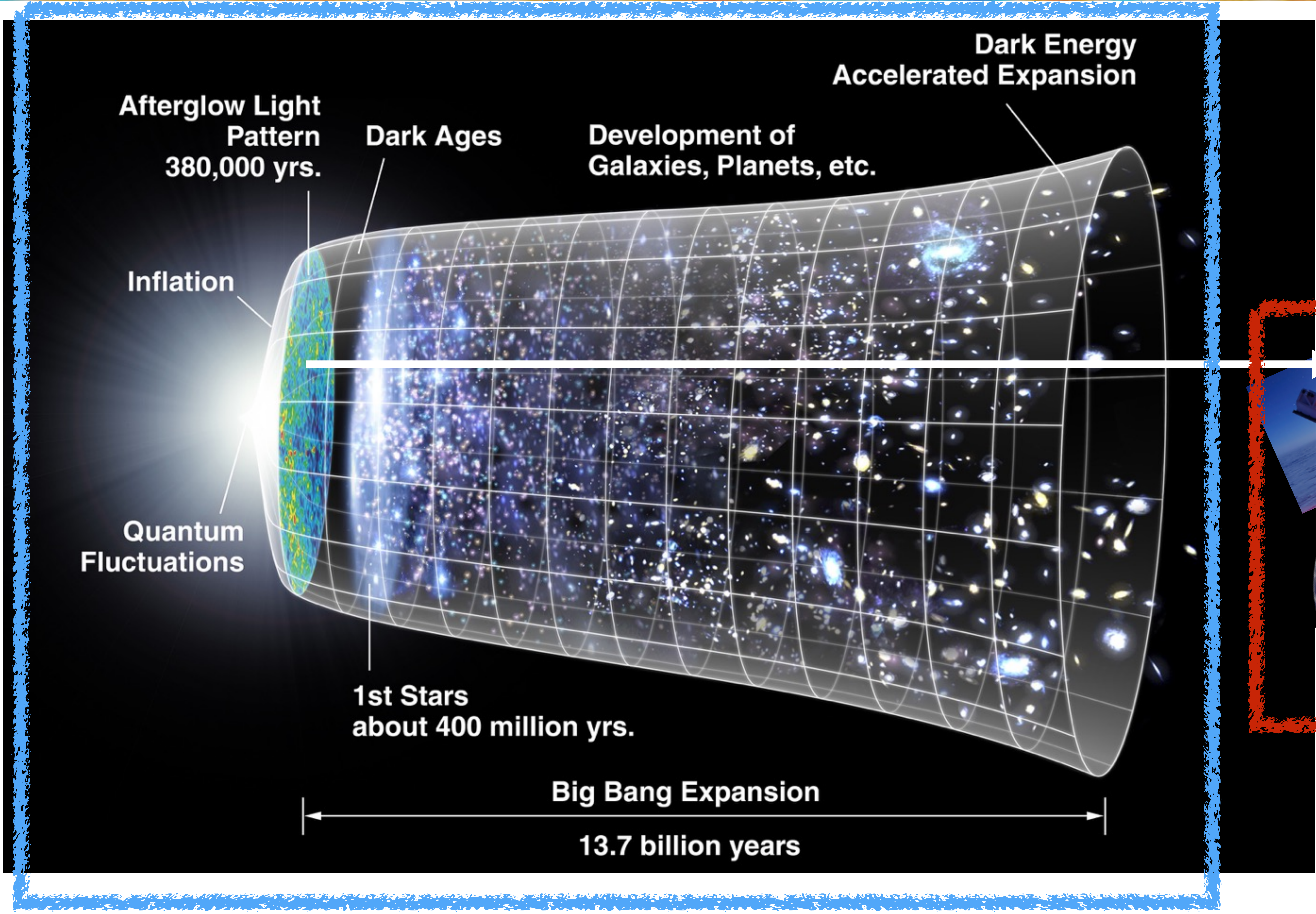
PhD course

Linear Cosmological Perturbations and CMB anisotropies

Davide Poletti and Nicoletta Krachmalnicoff

01/03/2017

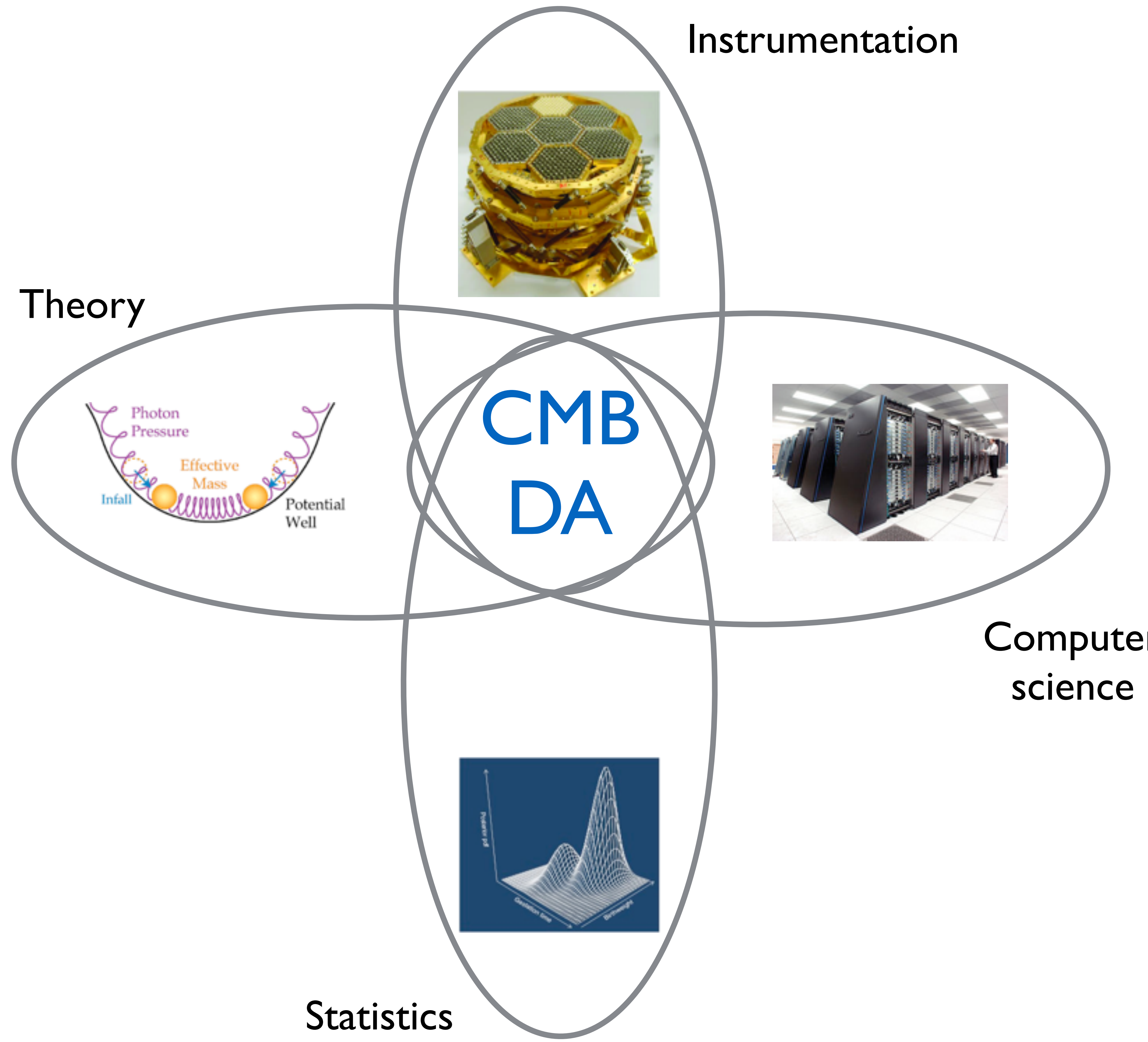
SISSA



This lecture

Elements of CMB data analysis:

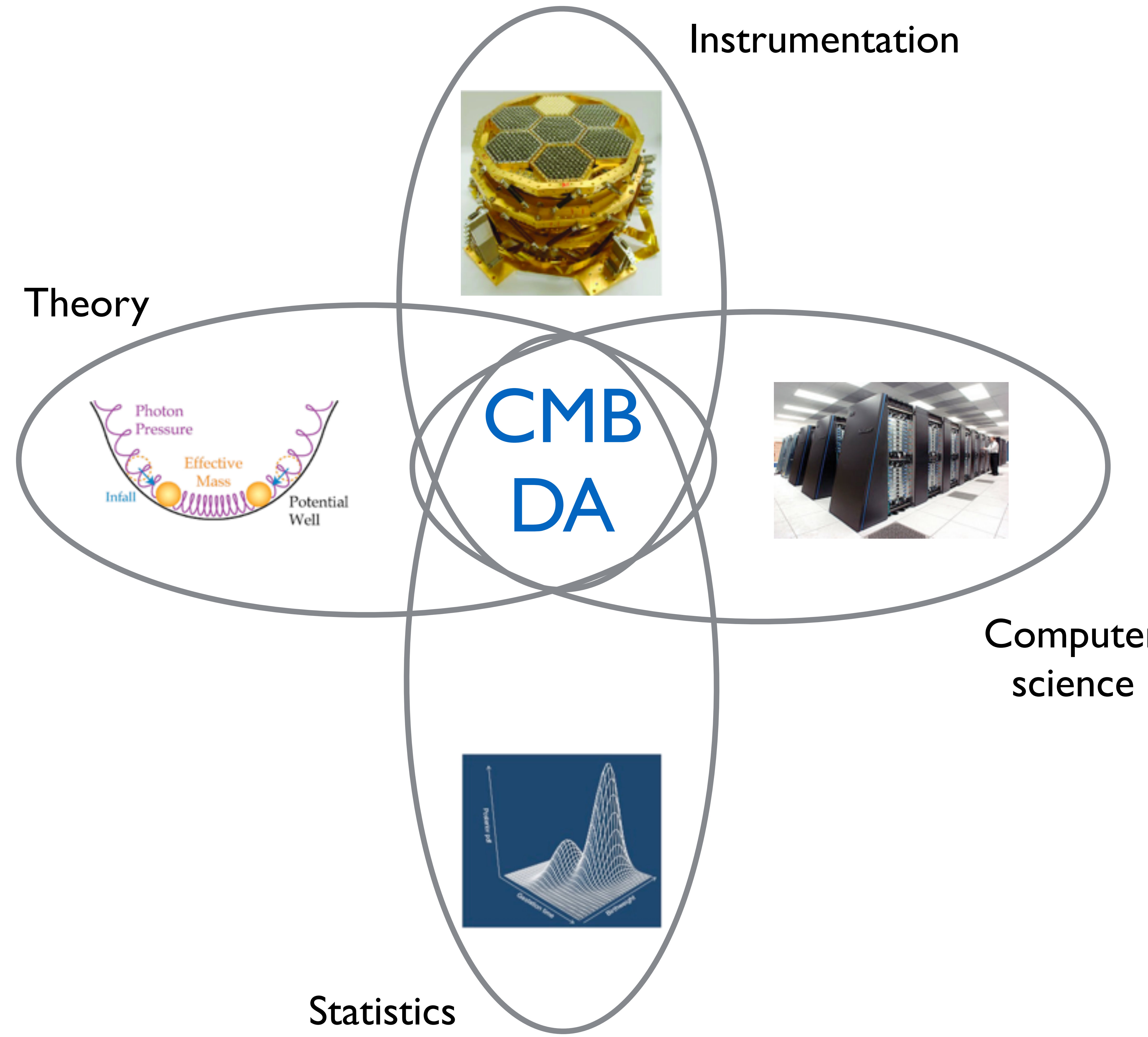
- Instrument and observations
- Time-Ordered data
- Map-making
- Component separation
- Power spectrum estimation
- Parameter estimation



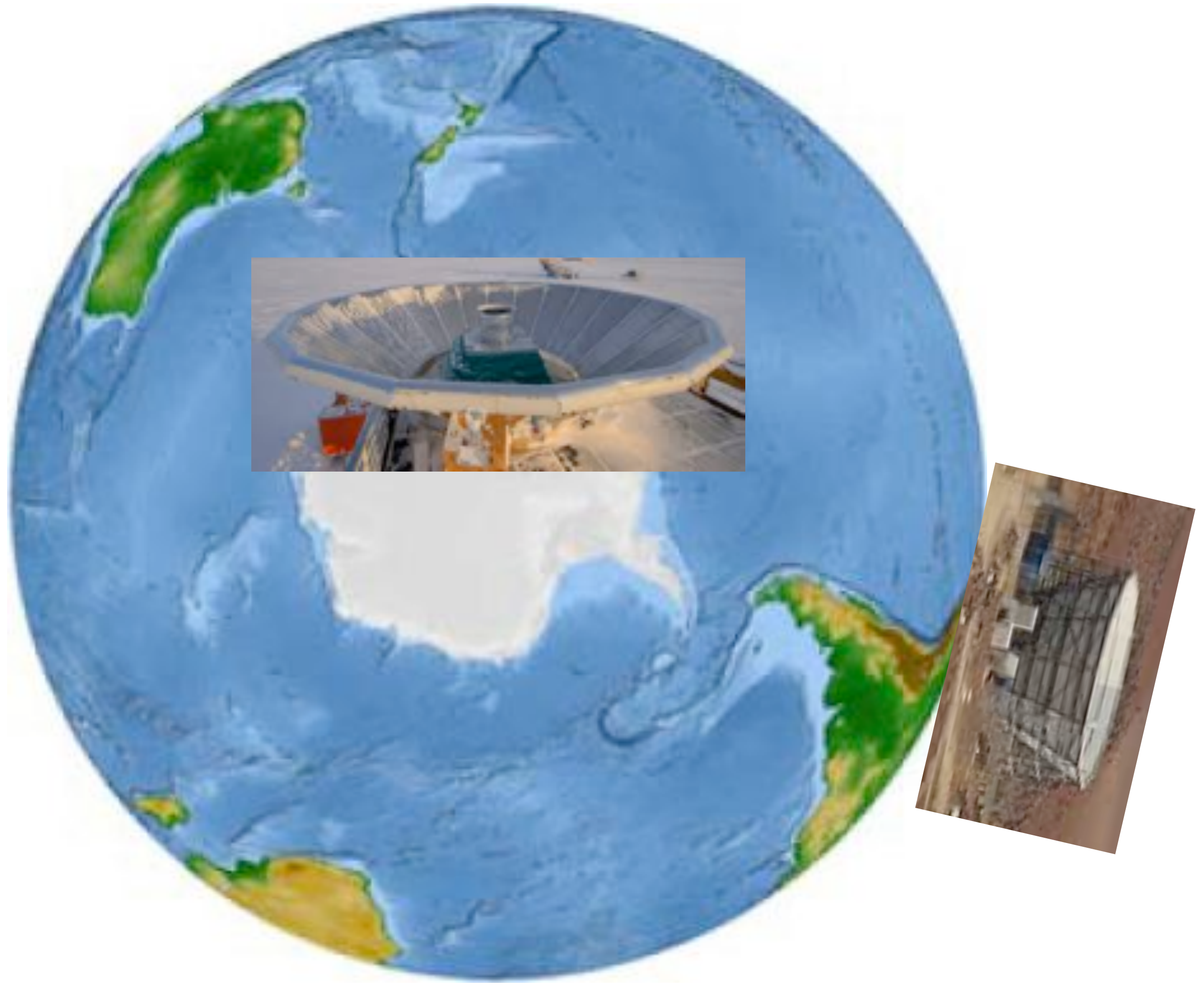
Overview

Elements of CMB data analysis:

- **Instrument and observations**
- **Time-Ordered data**
- Map-making
- Component separation
- Power spectrum estimation
- Parameter estimation

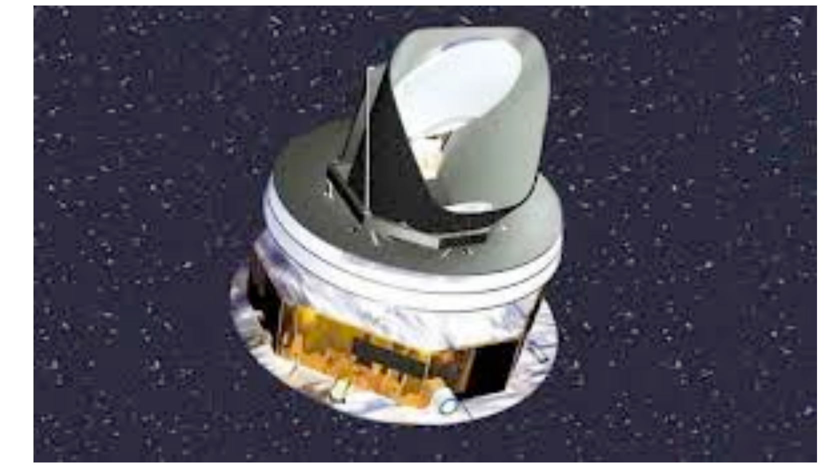


Where do the data come from?



Ground

- Heavy hardware
  - large telescopes (resolution)
  - large receivers (sensitivity)
- Environment contamination (atmosphere, ground…)
- Cutting-edge technology
- Maintenance possible
- <10y to deploy
- ~ 10 M€

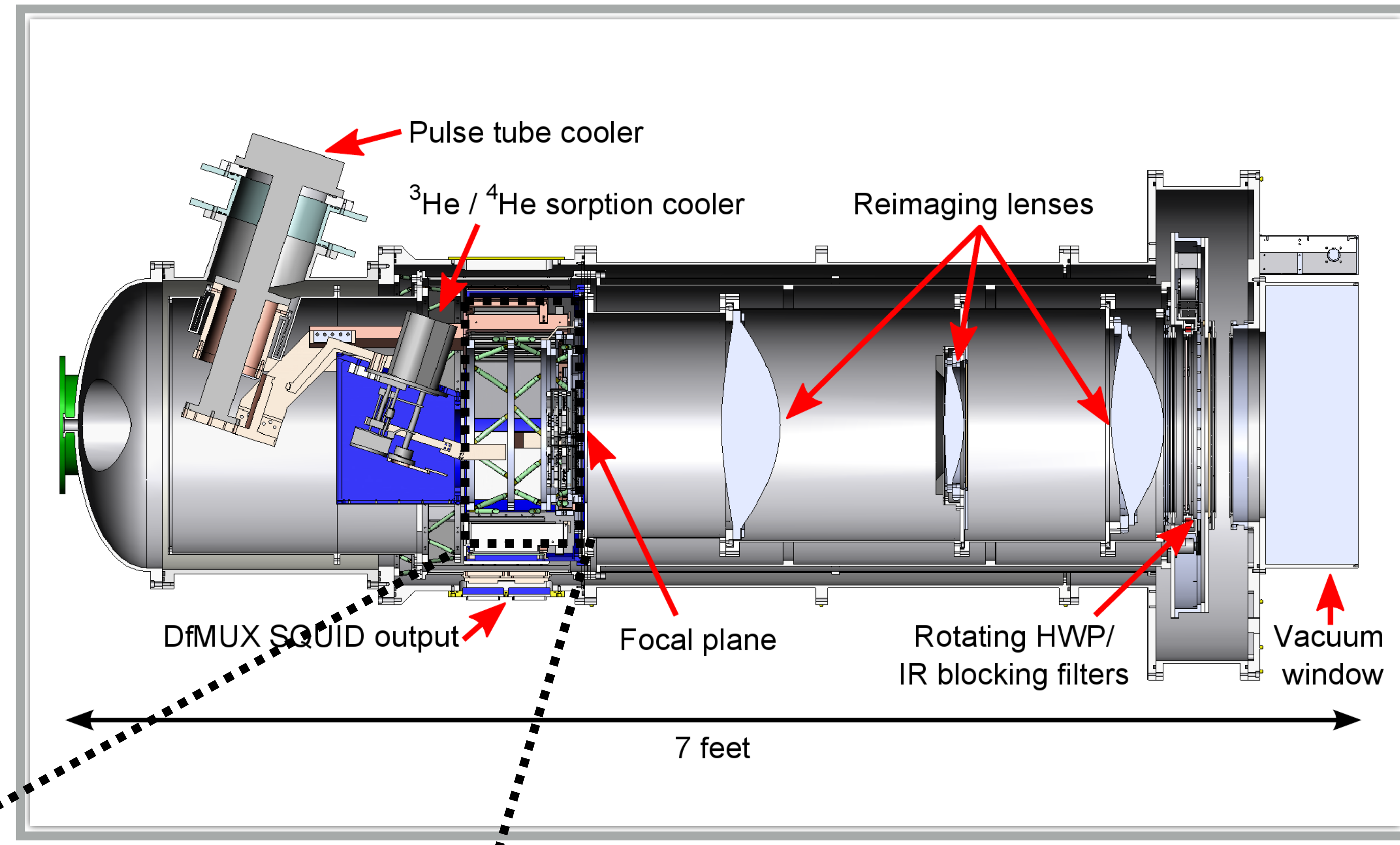


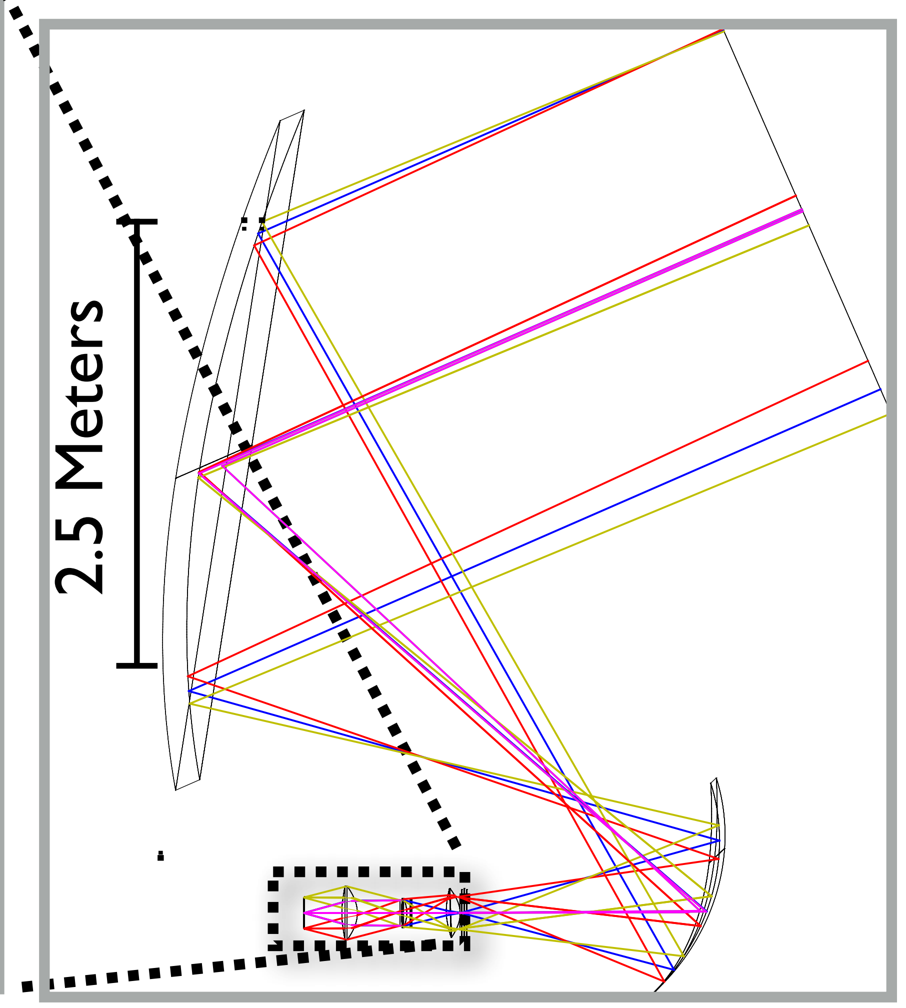
Space

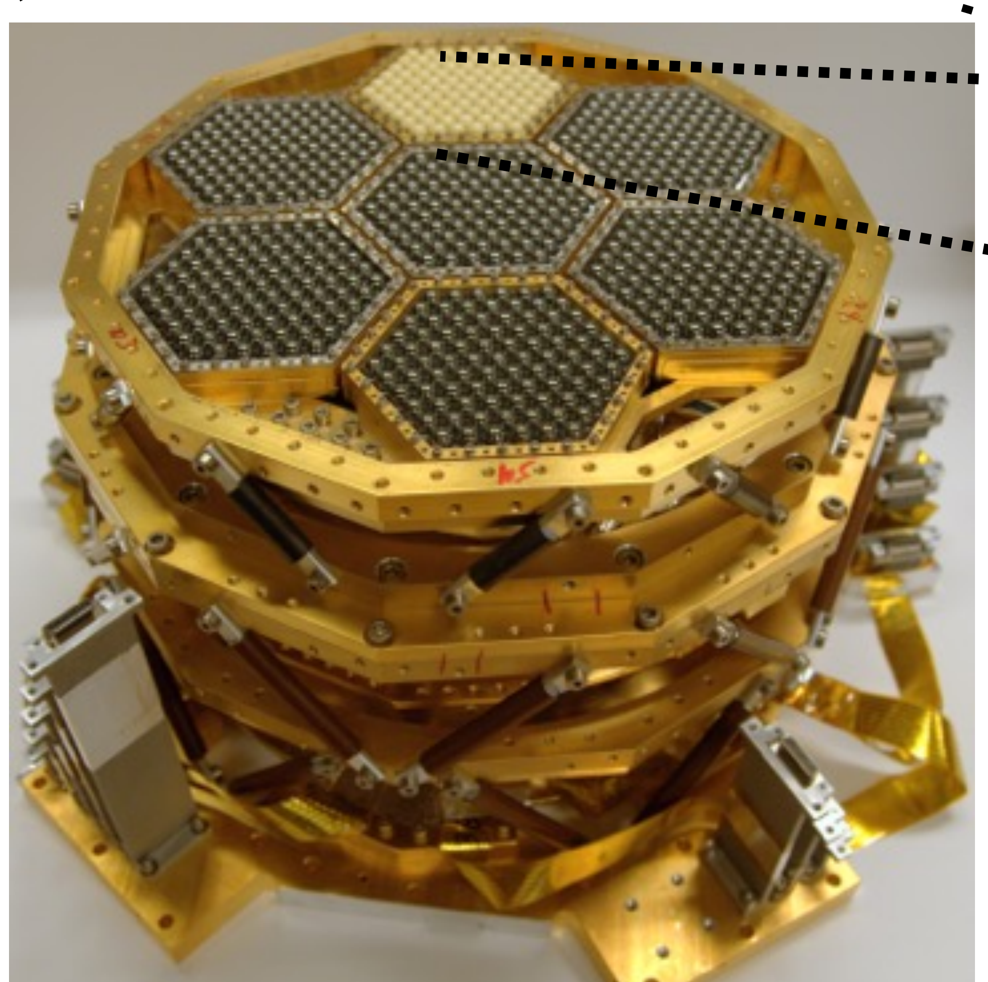
- Light hardware
- Extremely reliable technology
- Stable environment
- ~20y to deploy
- around G€
- full sky
- ~Only galactic contamination

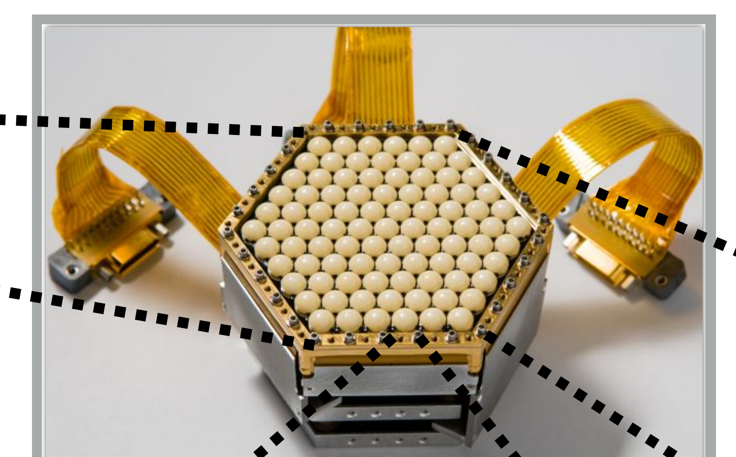
Balloon are midway:
(notably, limited atmosphere)

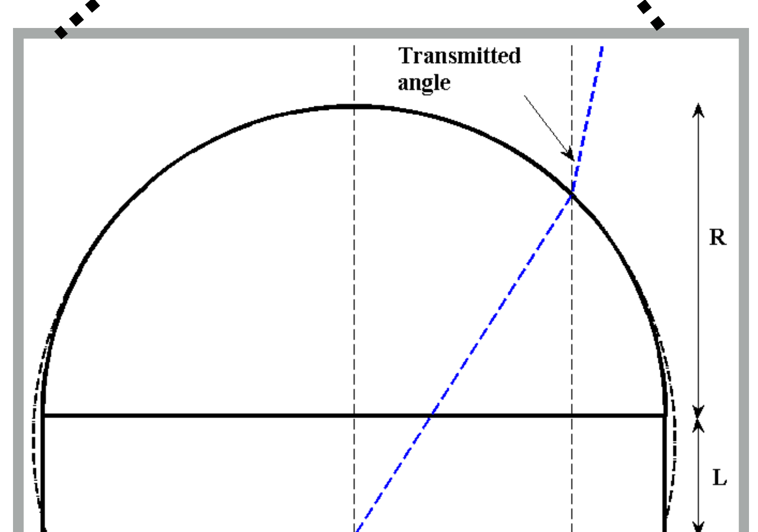
An example of CMB imager (POLARBEAR)

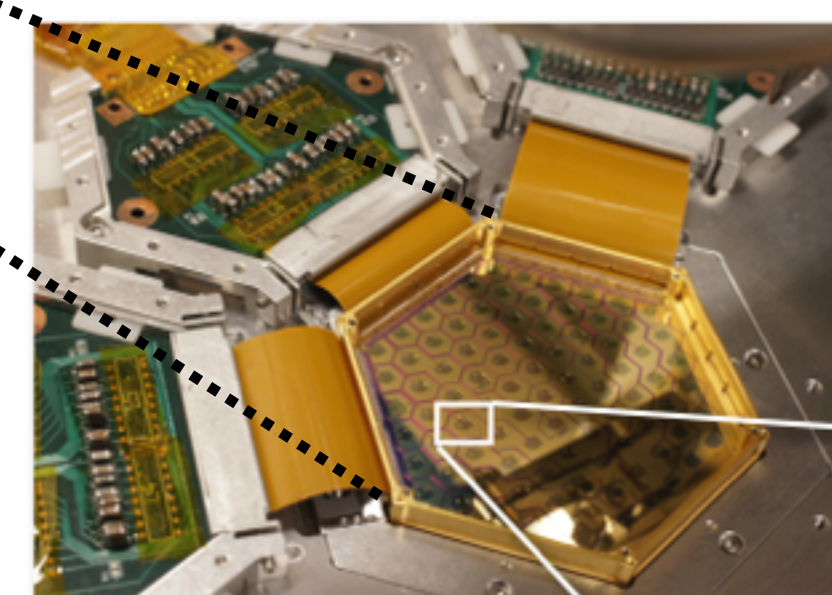


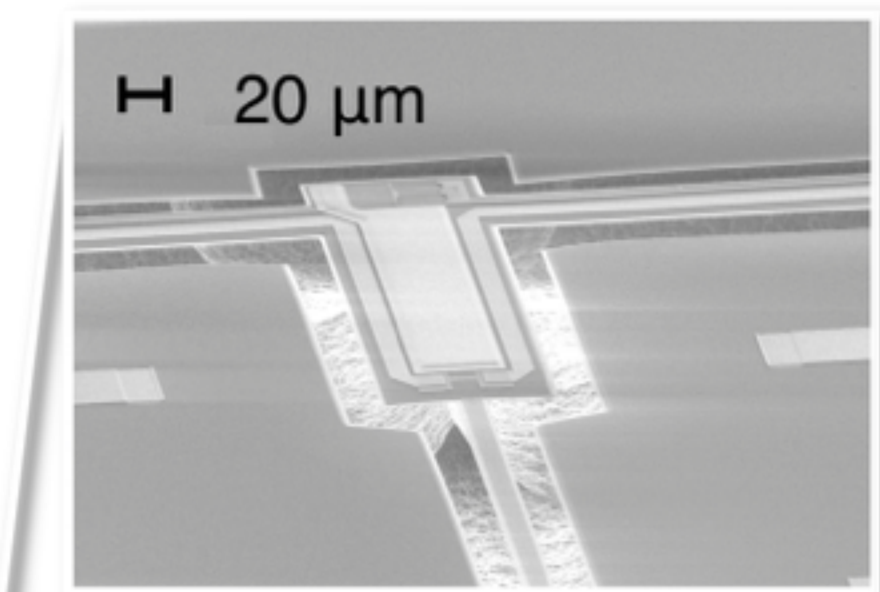


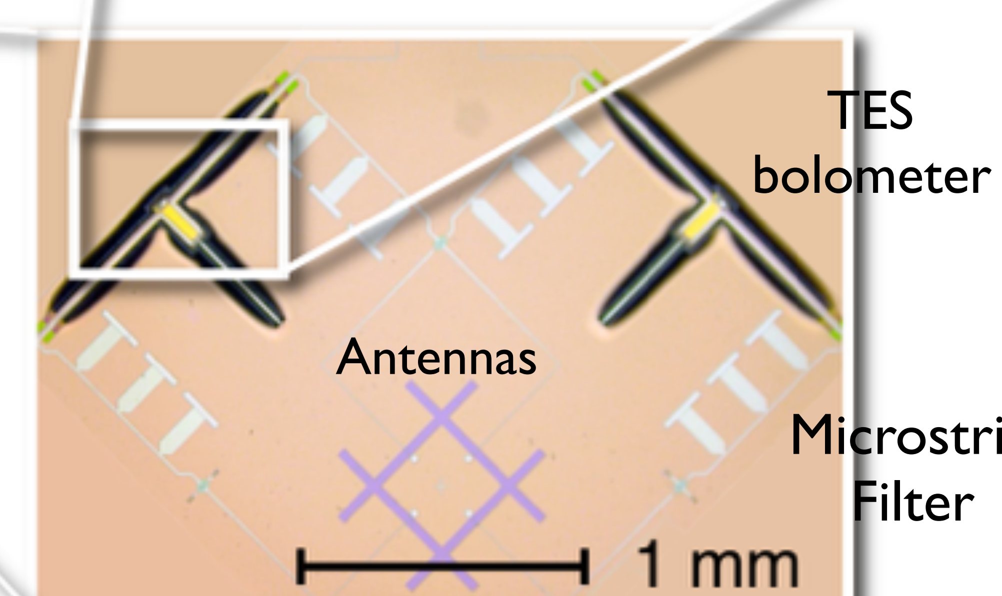






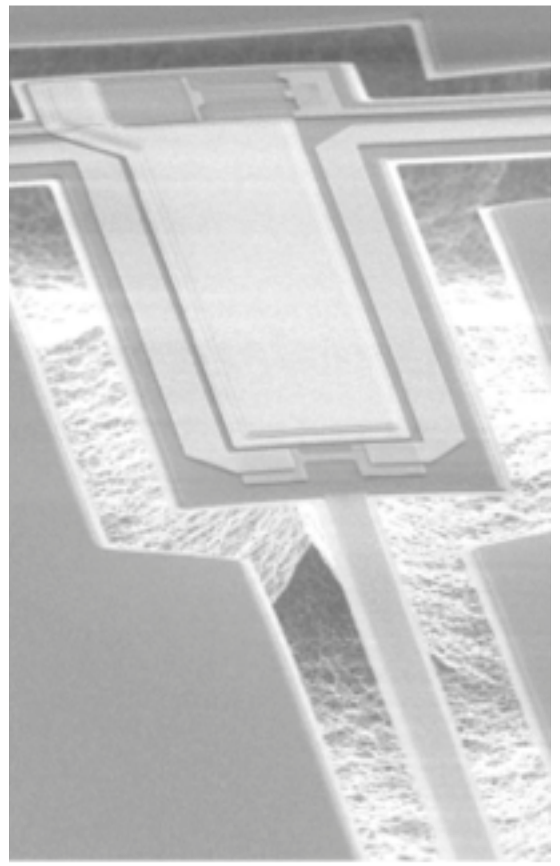


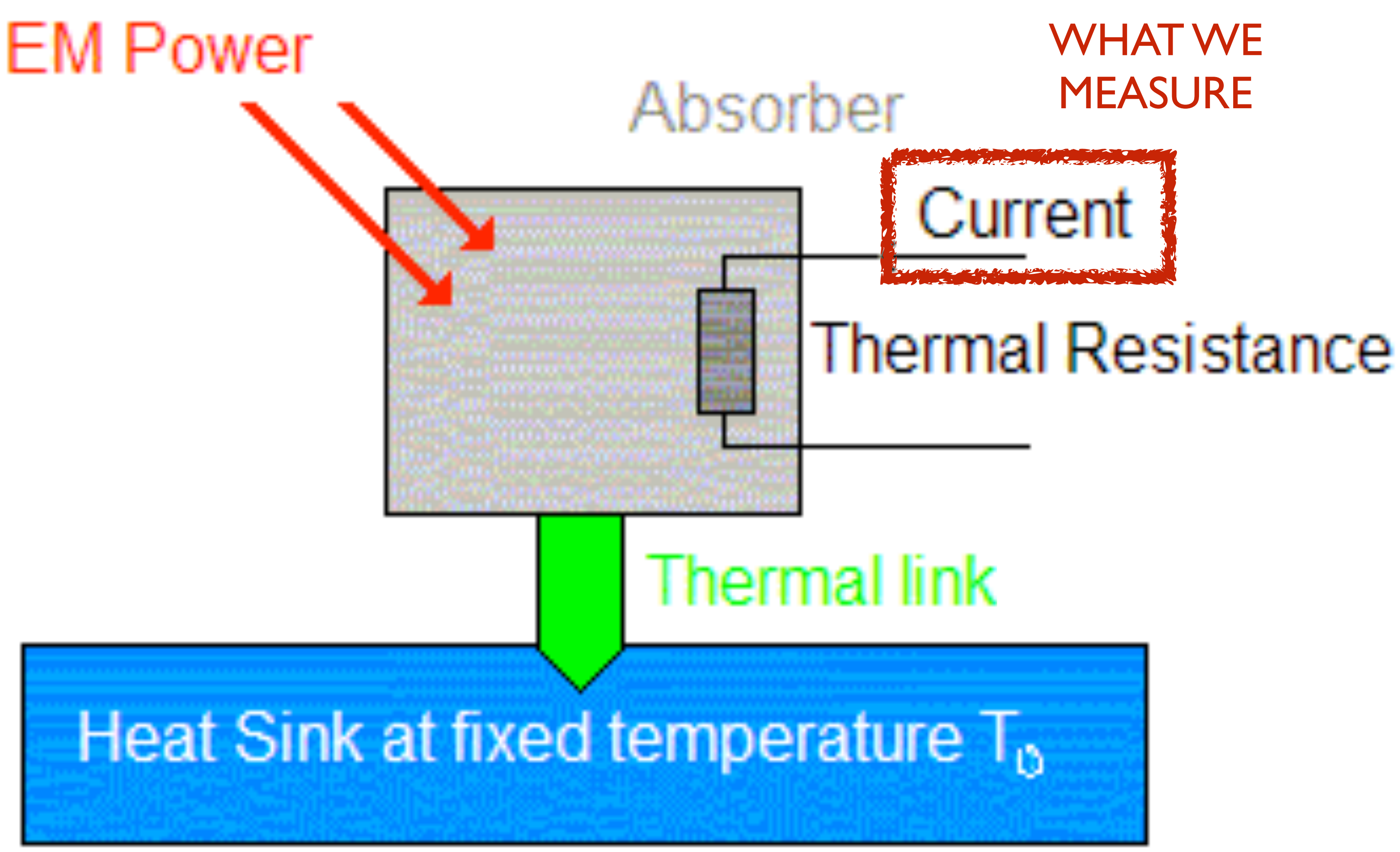




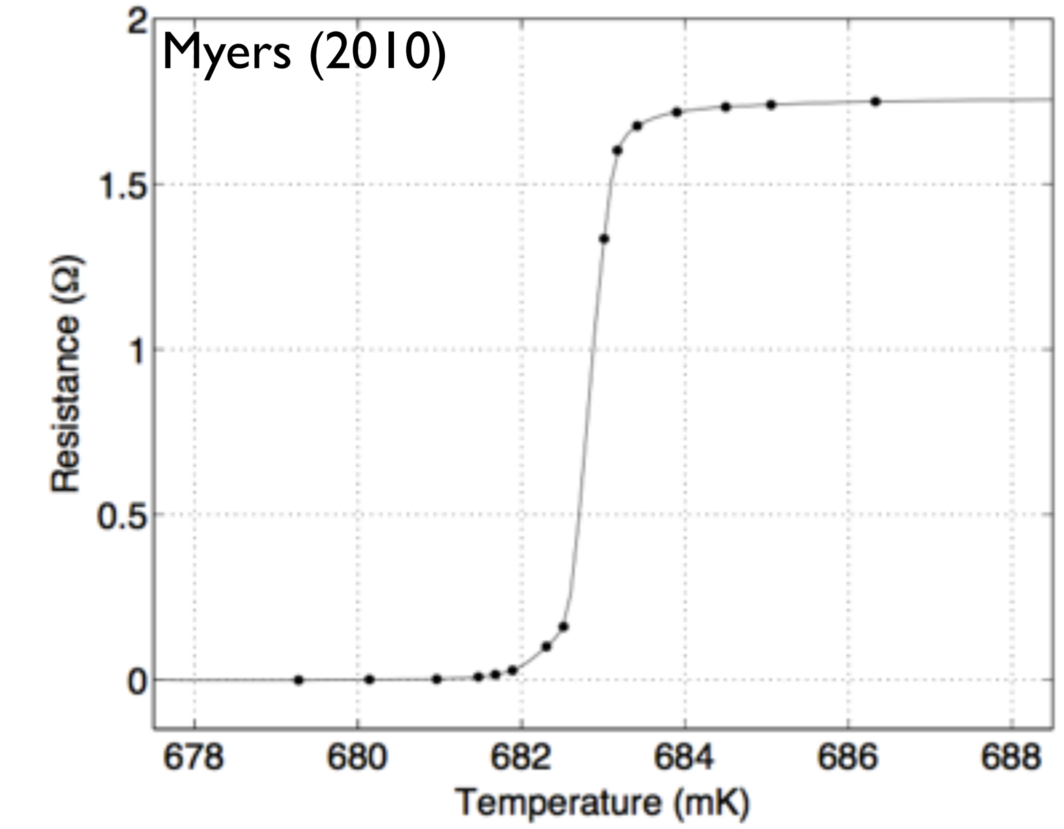
Read at ~200 Hz: Time Ordered Data

The detector





Credits: Lloyd Watkin



Superconducting transition:
➡ small change in temperature causes big change in current

Optical power on antennas
↕
Absorber temperature
↕
Resistance
↕
Current

CALIBRATION

However,

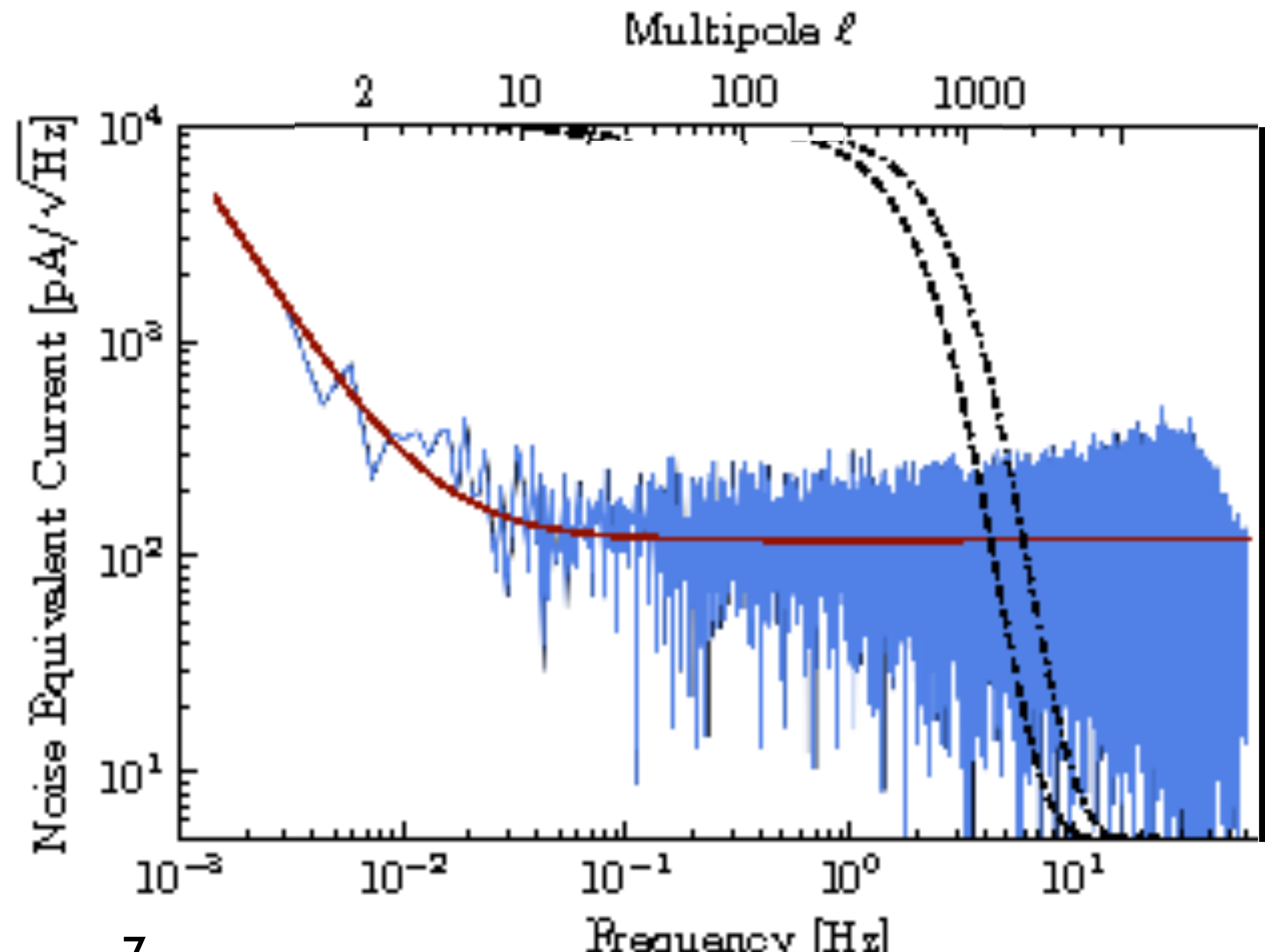
- drifts in the system
- 1/f noise
- …

➡ limit to the low frequency sensitivity

- response time

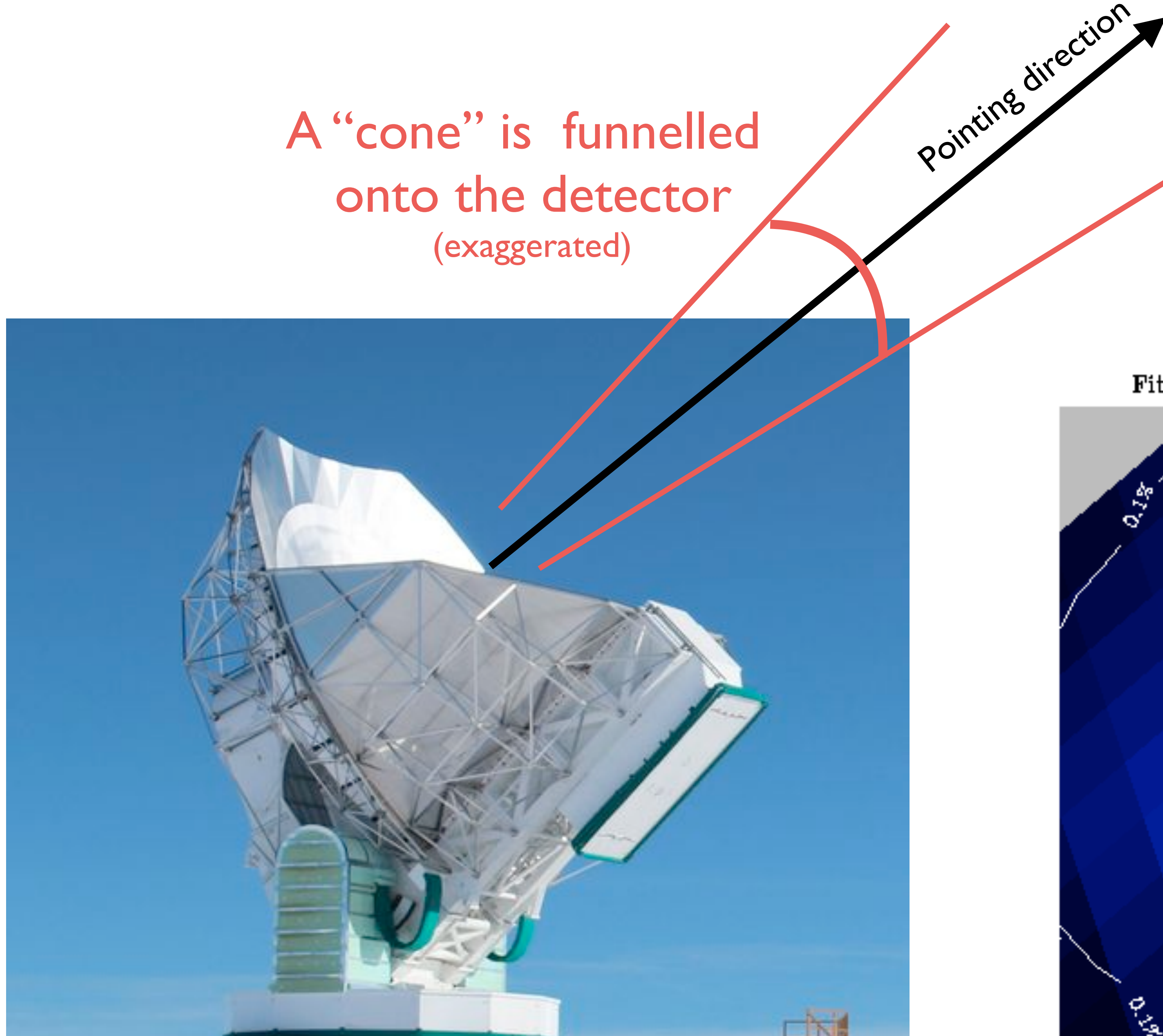
➡ limit to high frequency sensitivity

Telescopes scan the sky.
Scanning speed defines a correspondence between **frequency** and **angular scales**



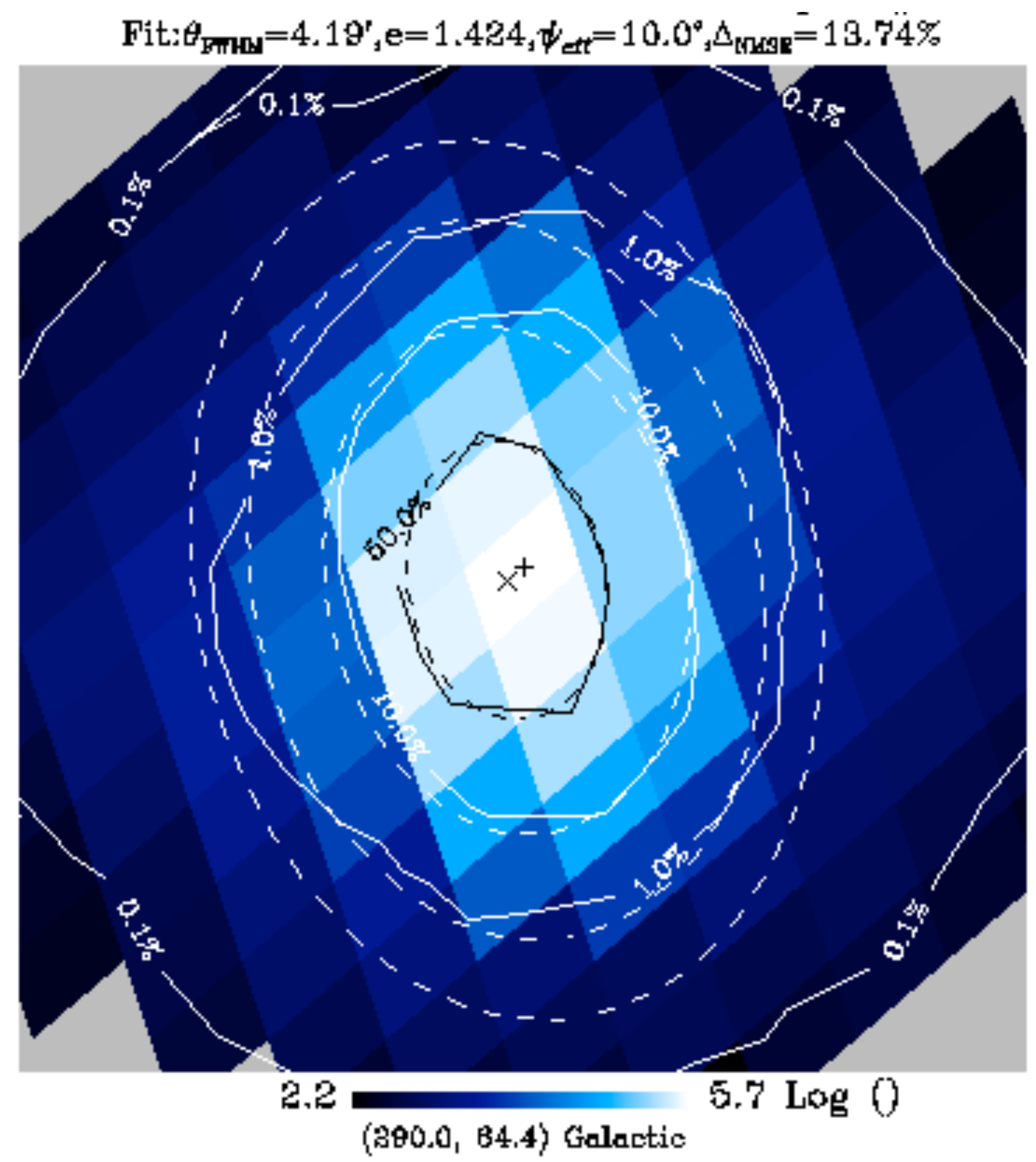
Adapted from Rahlin et al. (2014)
(multipole axis modified)

The beam



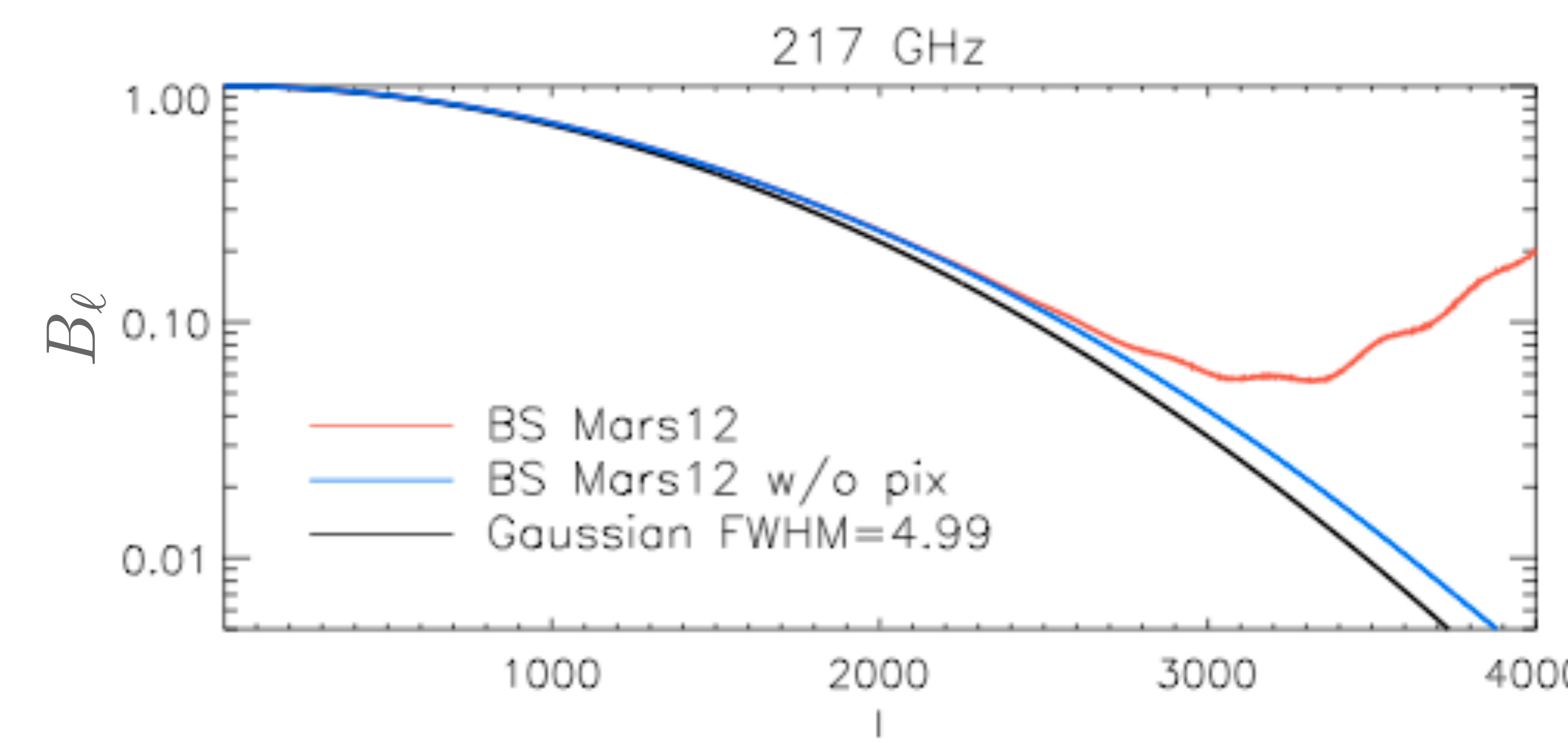
For given pointing direction the sky signal is convolved with the **beam function.**

Planck 217 GHz



$$\mathrm{FWHM} \propto \frac{\mathrm{wavelenght}}{\mathrm{primary\ \ diameter}}$$

It is approximated as a function of the angle from the center $B(\theta)$, which can be expanded in multipoles: B_ℓ



Given a sky $T(\hat{\boldsymbol{n}}) = \sum_{lm} a_{lm}\, Y_{lm}(\hat{\boldsymbol{n}})$ the actual sky seen by the detectors is $T(\hat{\boldsymbol{n}}) = \sum_{lm} a_{lm}\, Y_{lm}(\hat{\boldsymbol{n}})\, B_\ell$

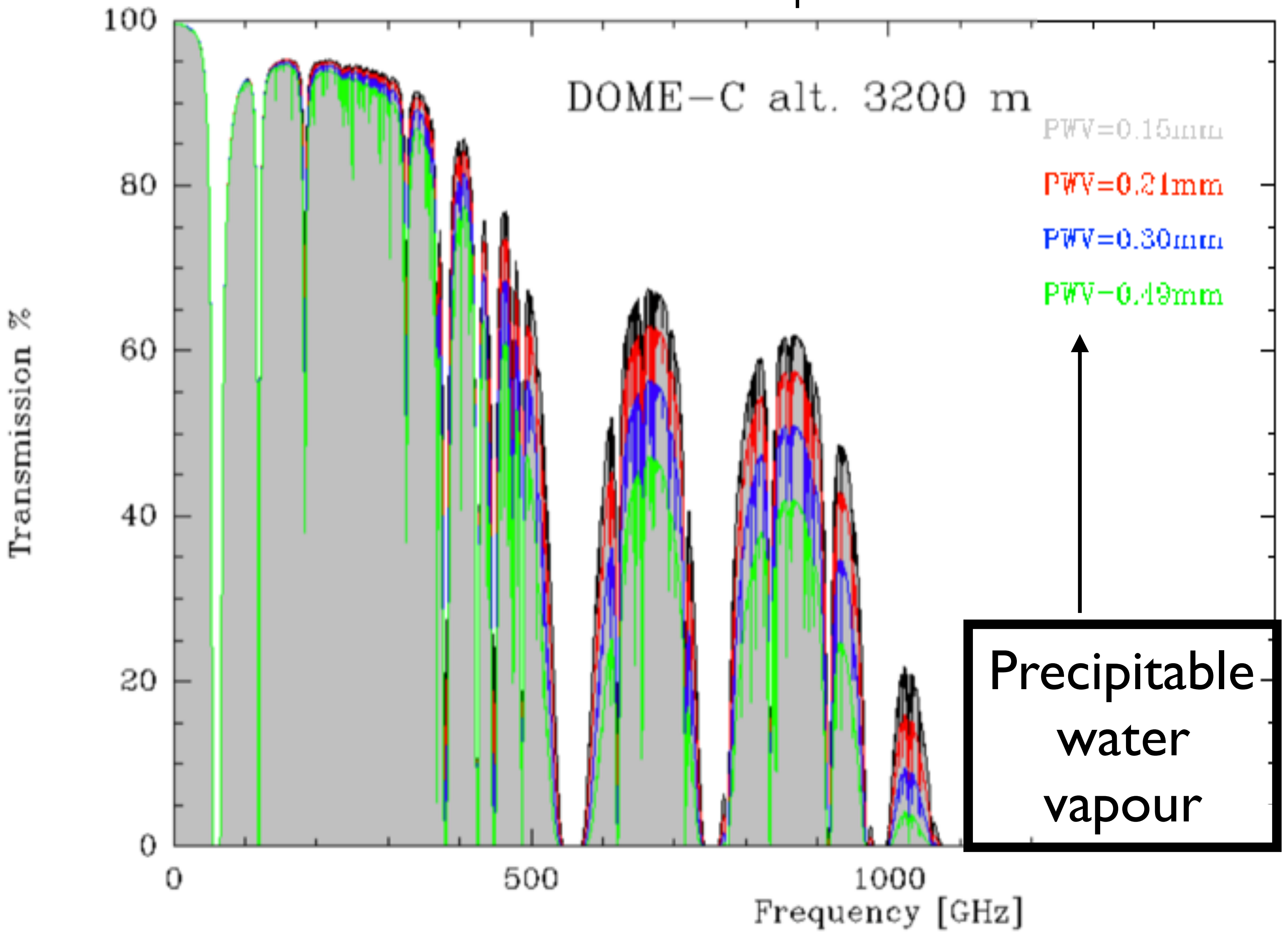
➡ the beam suppresses small scales power

TODOs are not just CMB

ATMOSPHERE

- **obstacle for high frequencies**

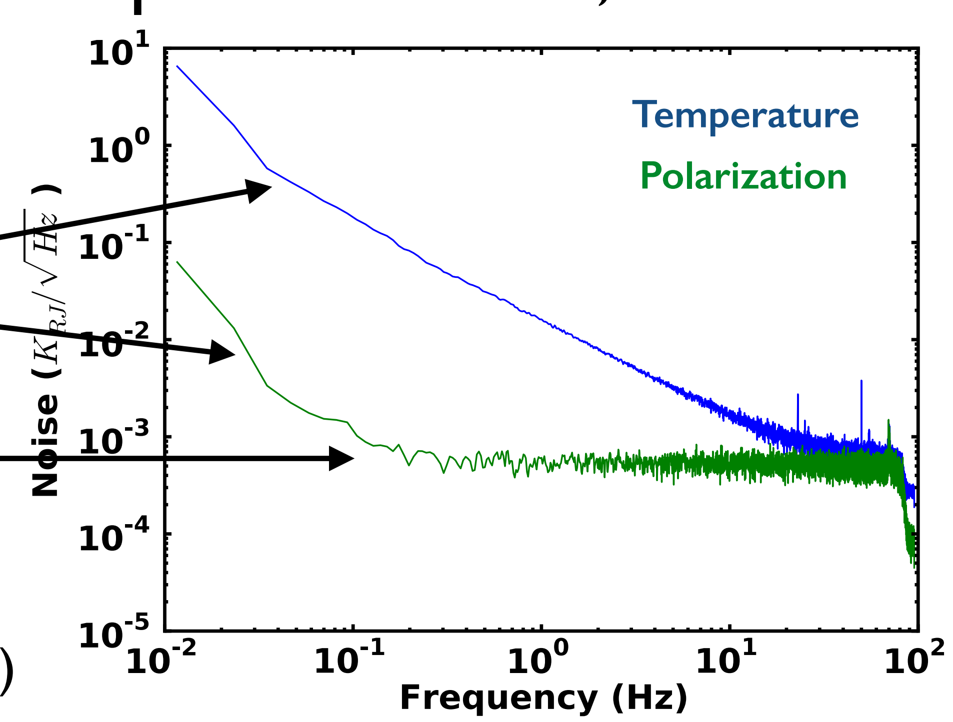
Adapted from Tremblin et al. 2012



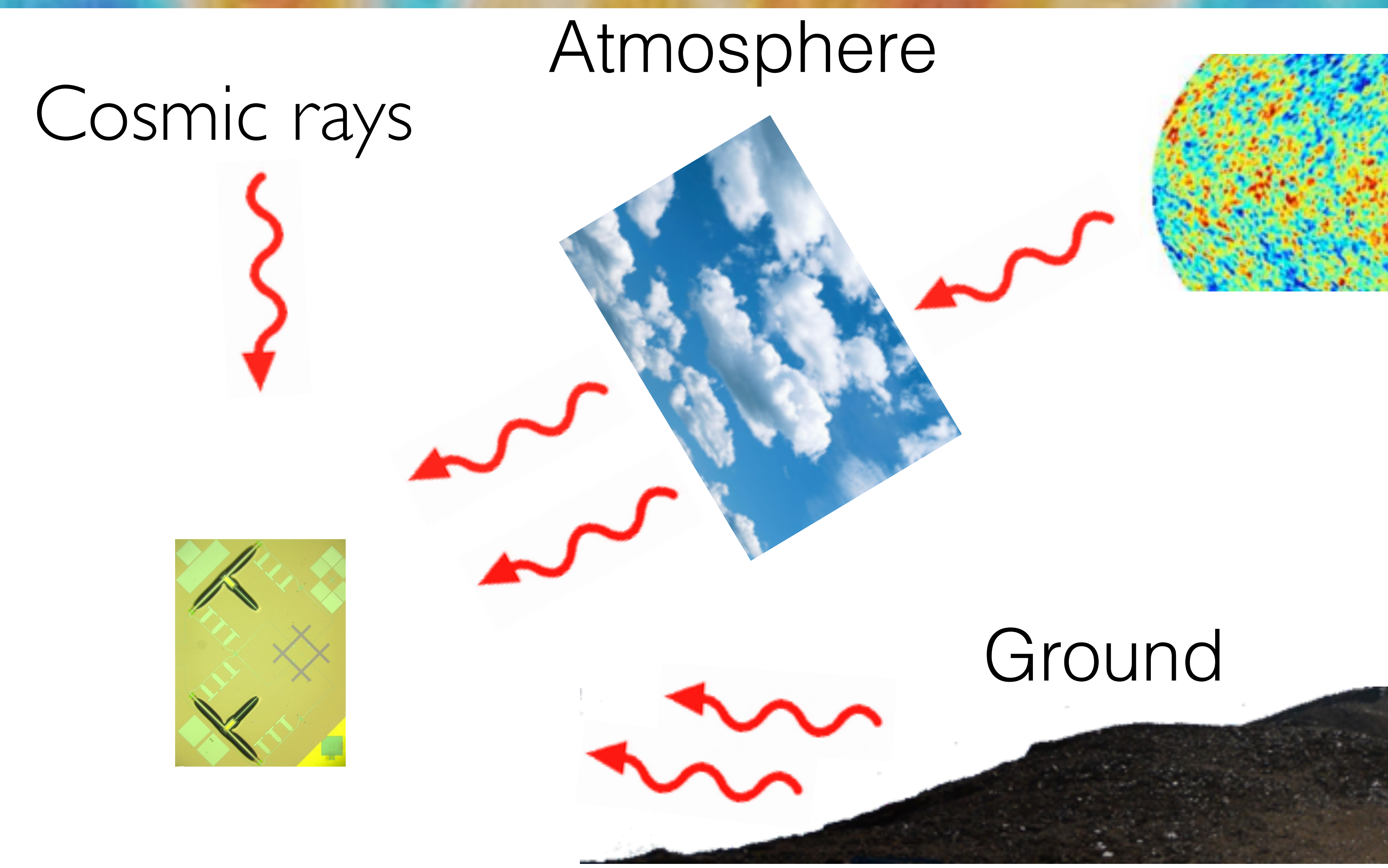
- not (significantly) polarized
- but atmospheric fluctuations behave like **1/f noise** and can leak to polarization (e.g. bandpass mismatch, instrumental polarization…)

Atmospheric noise

Detector noise



Adapted from Kermish et al. (2012)

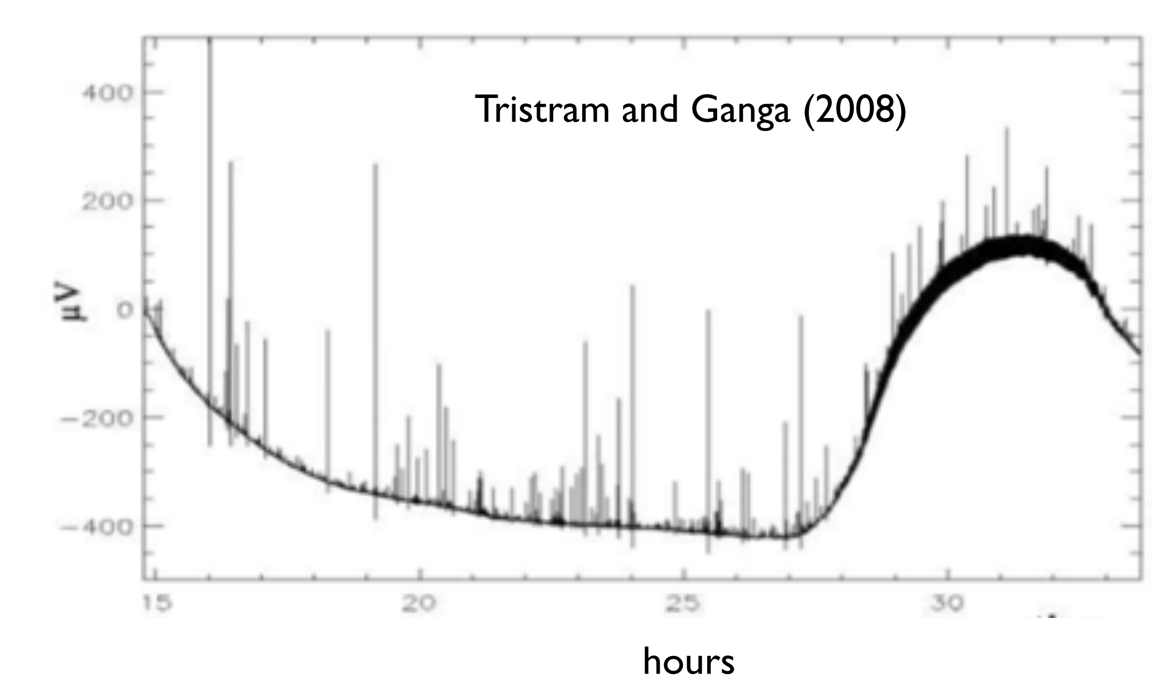


GROUND pickup:

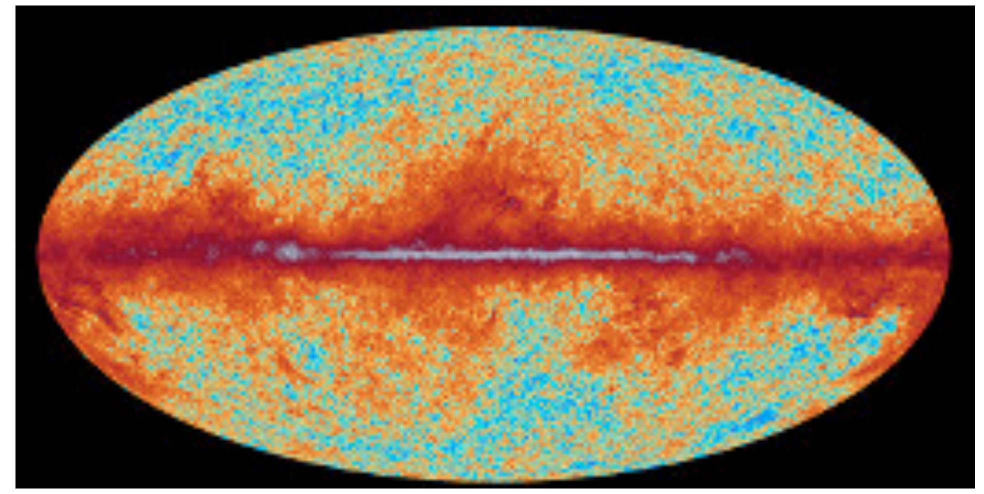
- The beam has sidelobes. They have very low amplitude but the ground is very bright
- Other effects (e.g., local magnetic field)
- ➡ ground-synchronous signal

AND MORE:

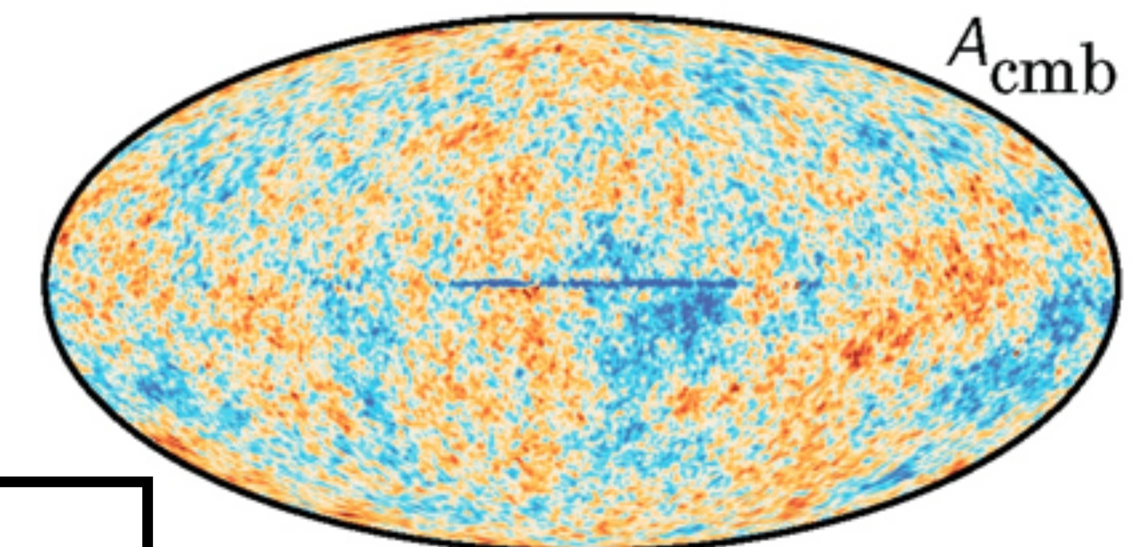
- cosmic rays,
- instrumental glitches,
- …
- ???



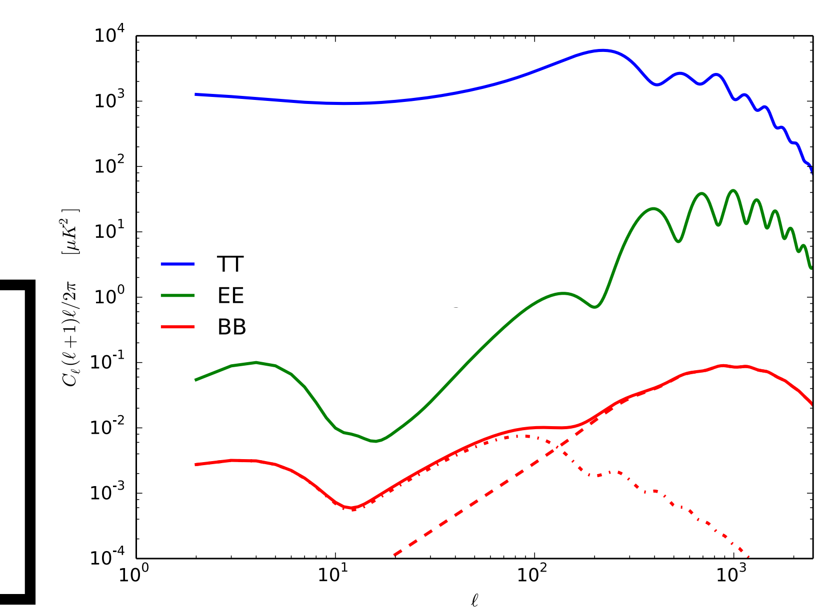
Timestream



Map-making



Component separation



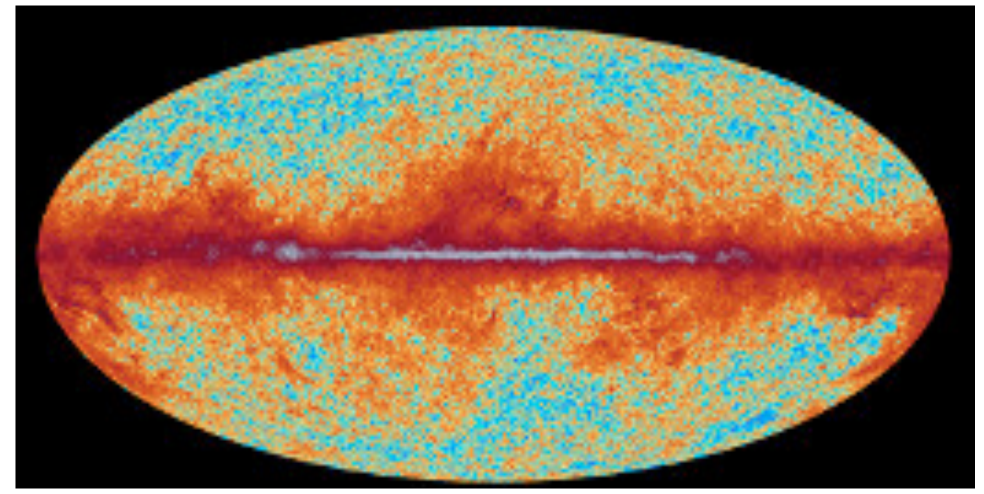
Power spectrum estimation

$\Omega_b, \Omega_m, \sigma_8, \ldots$

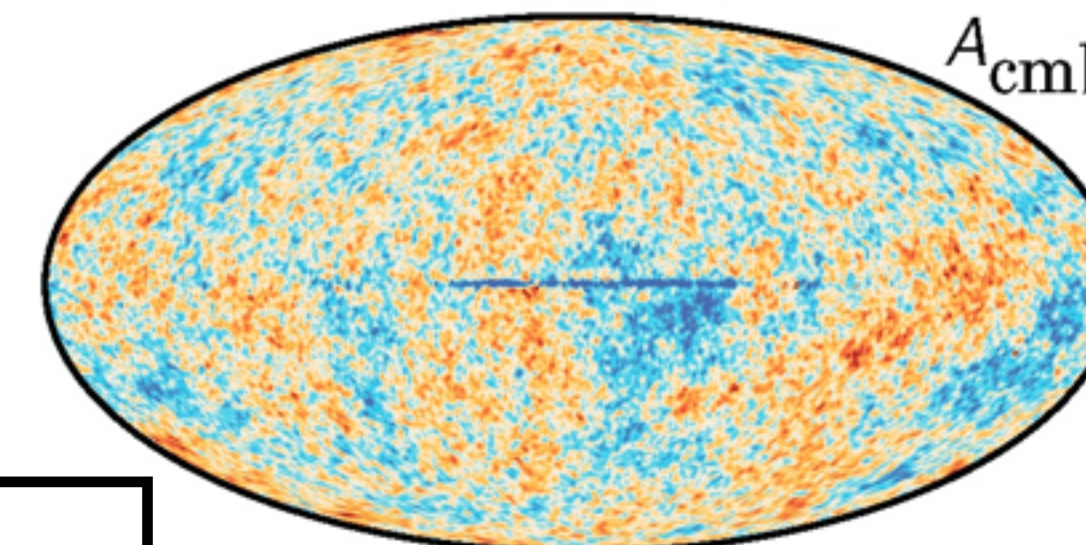
Parameter estimation

Overview

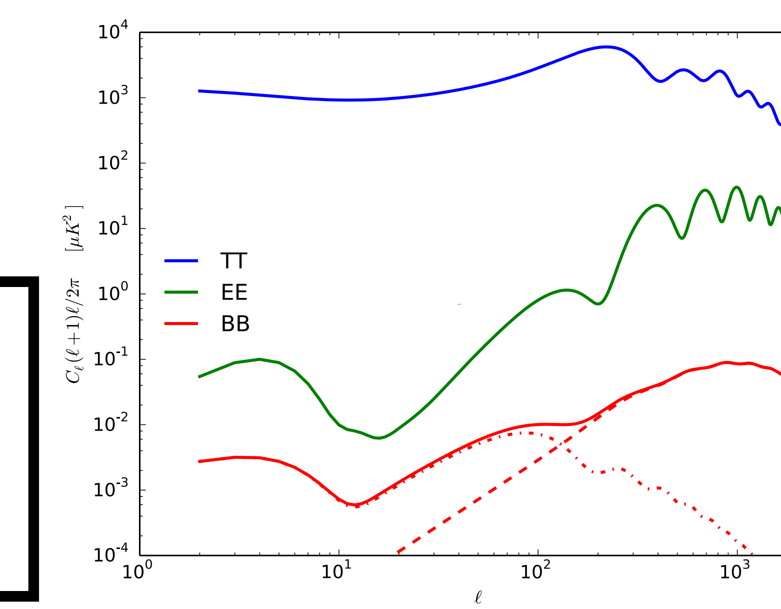
Timestream



Map-making



Component separation



Power spectrum estimation

$\Omega_b, \Omega_m, \sigma_8, \ldots$

Parameter estimation

CMB Data analysis, quantitatively

- **Time-Ordered data:**
  Polarbear:
  Volume = *sampling rate* x *detector number* x *observation time*
  = *~100 Hz x ~1000 x ~10^7 s = ~$\mathbf{10^{12}}$ samples (~10 TB)*

- **Map-making**
  Planck HF maps: 1.7 arcmin resolution, full sky: 5 x $\mathbf{10^7}$ pixels
  Polarbear: 1.7 arcmin, 0.1% sky: 5 x $\mathbf{10^4}$ pixels

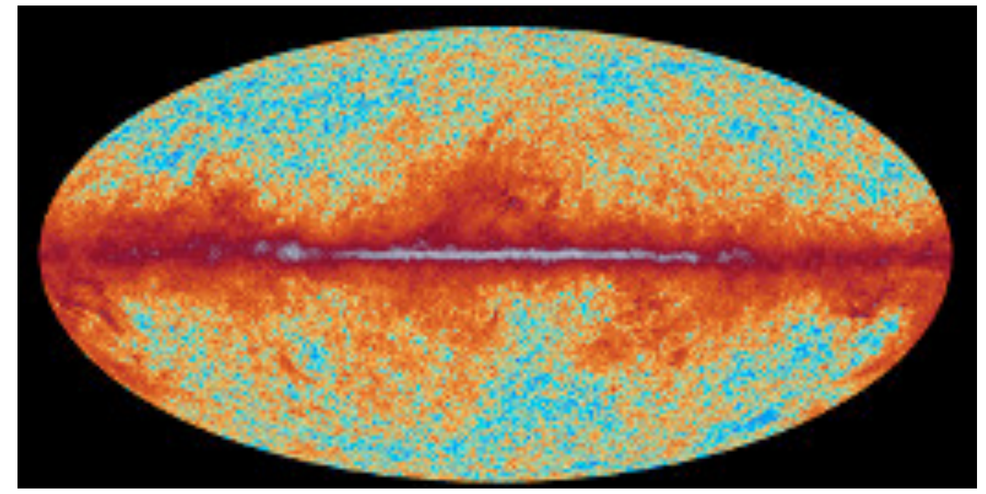
- **Component separation**
  Typically, information compression of *O(1)*

- **Power spectrum estimation**
  Typically *O(**10**)-O(**100**)* power spectrum points

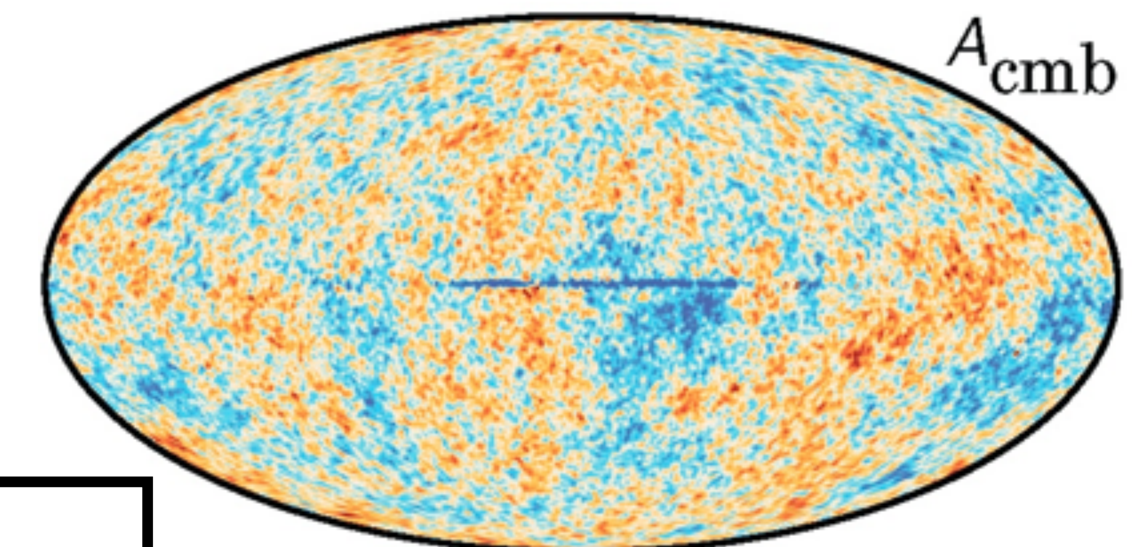
- Estimation of *O(**1**)-O(**10**)* **cosmological parameters**

Compression has to be efficient and effective
➡ computer science and statistics play important roles

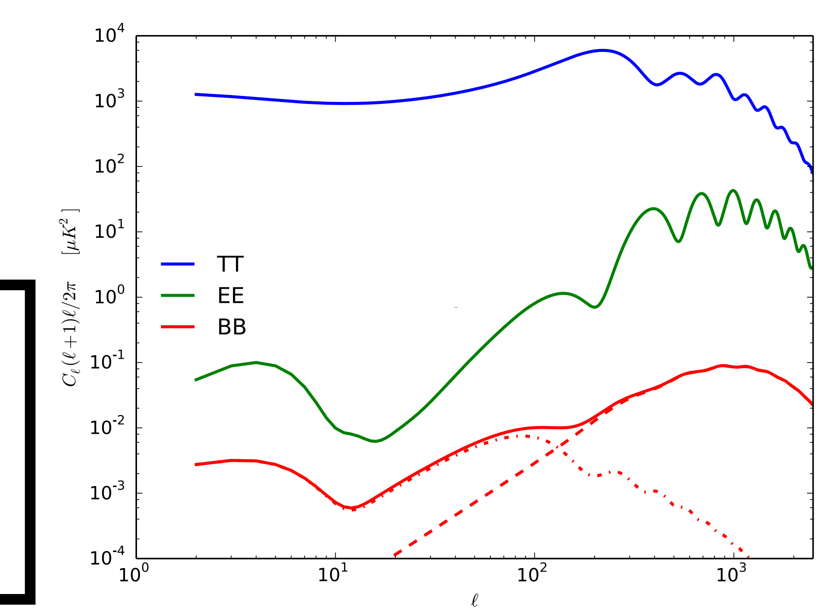
Timestream



Map-making



Component separation



Power spectrum estimation

$\Omega_b, \Omega_m, \sigma_8, \ldots$

Parameter estimation

Map-making

$$d_t = I_{p_t} + \cos(2\varphi_t)Q_{p_t} + \sin(2\varphi_t)U_{p_t} + n_t$$

Noise contribution

Samples of the TOD recorded at time *t*

Pixelized maps of the Stokes parameters

Orientation of the detector projected on the sky

Sky pixel observed

The complete time stream ➡

$$\mathbf{d} = \mathbf{A}\mathbf{s} + \mathbf{n}$$

$\mathbf{A}$ = Pointing matrix

$\mathbf{s}$ = sky signal

$\mathbf{n}$ = noise with covariance $\mathbf{N}$

Generalised
Least
Squared
estimator

$$\hat{\mathbf{s}} = (\mathbf{A}^\top \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{W} \mathbf{d}$$

$$\hat{\mathbf{s}} = \boxed{\mathcal{N}_p \times \mathcal{N}_p}^{-1} \cdot \boxed{\mathcal{N}_p \times \mathcal{N}_t} \cdot \boxed{\mathcal{N}_t \times \mathcal{N}_t} \cdot \boxed{\mathcal{N}_t}$$

Minimize the "chi-square" $(\mathbf{d} - \mathbf{A}\mathbf{s})^\top \mathbf{W} (\mathbf{d} - \mathbf{A}\mathbf{s})$

$\mathbf{W}$ can be any symmetric positive definite matrix.

Minimum variance when $\mathbf{W} = \mathbf{N}^{-1}$
➡Noise characterization important

Map-making challenges

We saw earlier that:

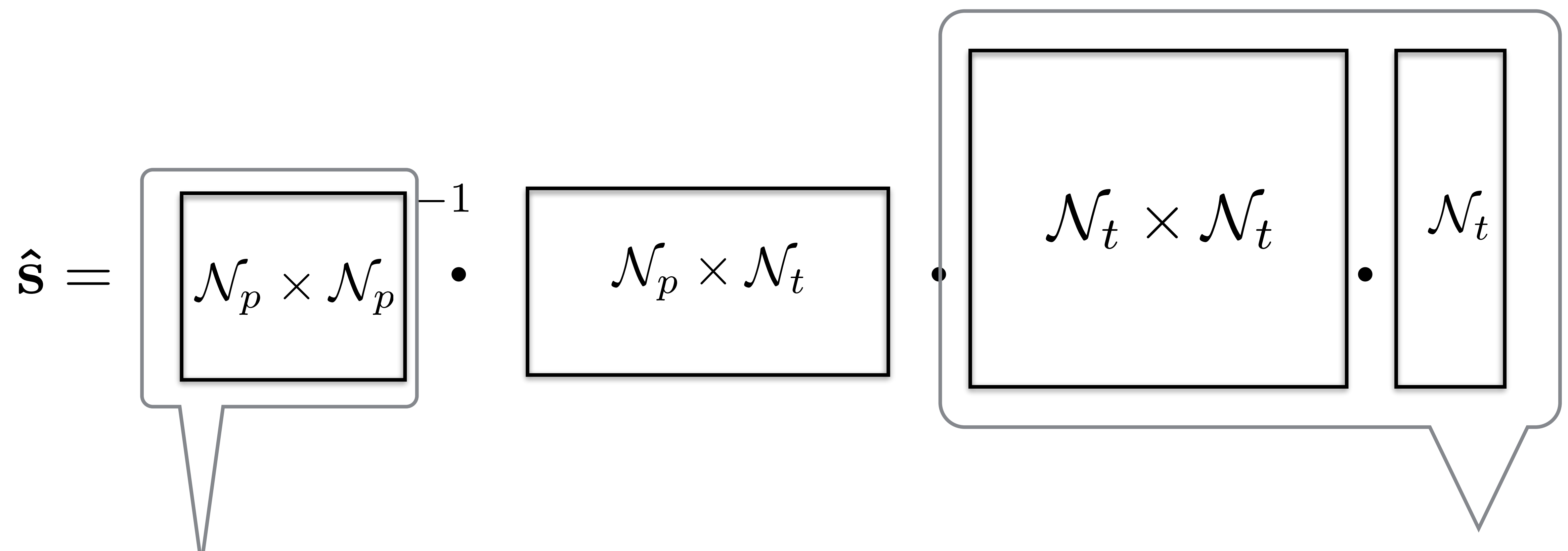
- Noise is correlated
  - ➡ The optimal $\boldsymbol{W}$ is <u>not</u> diagonal
- Data are not just CMB and noise: $\mathbf{d} = \mathbf{As} + \mathbf{Ty} + \mathbf{n}$
  $\mathbf{Ty}$ models the contaminating signals (known templates with unknown amplitude)
  ➡ The GLS solution becomes $\hat{\mathbf{s}} = (\mathbf{A}^\top \mathbf{F_T A})^{-1}\mathbf{A}^\top \mathbf{F_T d}$
  $\mathbf{F_T} \equiv \mathbf{W} - \mathbf{WT}(\mathbf{T}^\top \mathbf{WT})^{-1}\mathbf{T}^\top \mathbf{W}$ is a filtering operator: $\mathbf{F_T T} = 0$

Remind, sensible values are

$\mathcal{N}_p = 10^6$

$\mathcal{N}_t = 10^{12}$

Speed of processors is $\sim 10^9$ operations/sec



Mostly determined by the scanning strategy
➡ A priori dense
➡ Challenging inversion

Careful choice of $\mathbf{W}$ and $\mathbf{T}$ to make this feasible

Map-making: solving a (too) large inverse problem

How to invert this $\mathcal{N}_p \times \mathcal{N}_p$ matrix? Inversion requires $\mathcal{N}_p^3 \sim 10^{18}$ operations (100 cpu y)

➡ Find approximate solution without explicit inversion using the **P**reconditioned **C**onjugate **G**radient technique

Solve $\mathbf{B}\,\mathbf{x} = \mathbf{b}$ with $\mathbf{B}$ symmetric positive definite.

Idea:

- use $\mathbf{B}$ as scalar product, given a search direction $\hat{\mathbf{p}}$ (with $\hat{\mathbf{p}}^t\,\mathbf{B}\,\hat{\mathbf{p}} = 1$), the projection onto it

  $$\hat{\mathbf{p}}(\hat{\mathbf{p}}^t\,\mathbf{B}\,\mathbf{x}) = \hat{\mathbf{p}}(\hat{\mathbf{p}}^t\,\mathbf{b})$$

  can be interpreted as an approximate solution.
- project the solution on an increasingly larger subspace until the approximate solution is "good enough": e.g., $|\mathbf{B}\,\mathbf{x} - \mathbf{b}|/|\mathbf{b}| < 10^{-6}$

$$\mathcal{N}_p^2 \times \mathcal{N}_{\mathrm{iter}} \sim 10^{12} \times [O(10) \text{ to } O(1000)] \sim 0.1 \text{ cpu y (optimistic)}$$

It assumes $\mathbf{B}$ precomputed and stored (often it isn't)

It can be reduced by proper preconditioning

It assumes 100% cpu (never achieved)

High performance computing

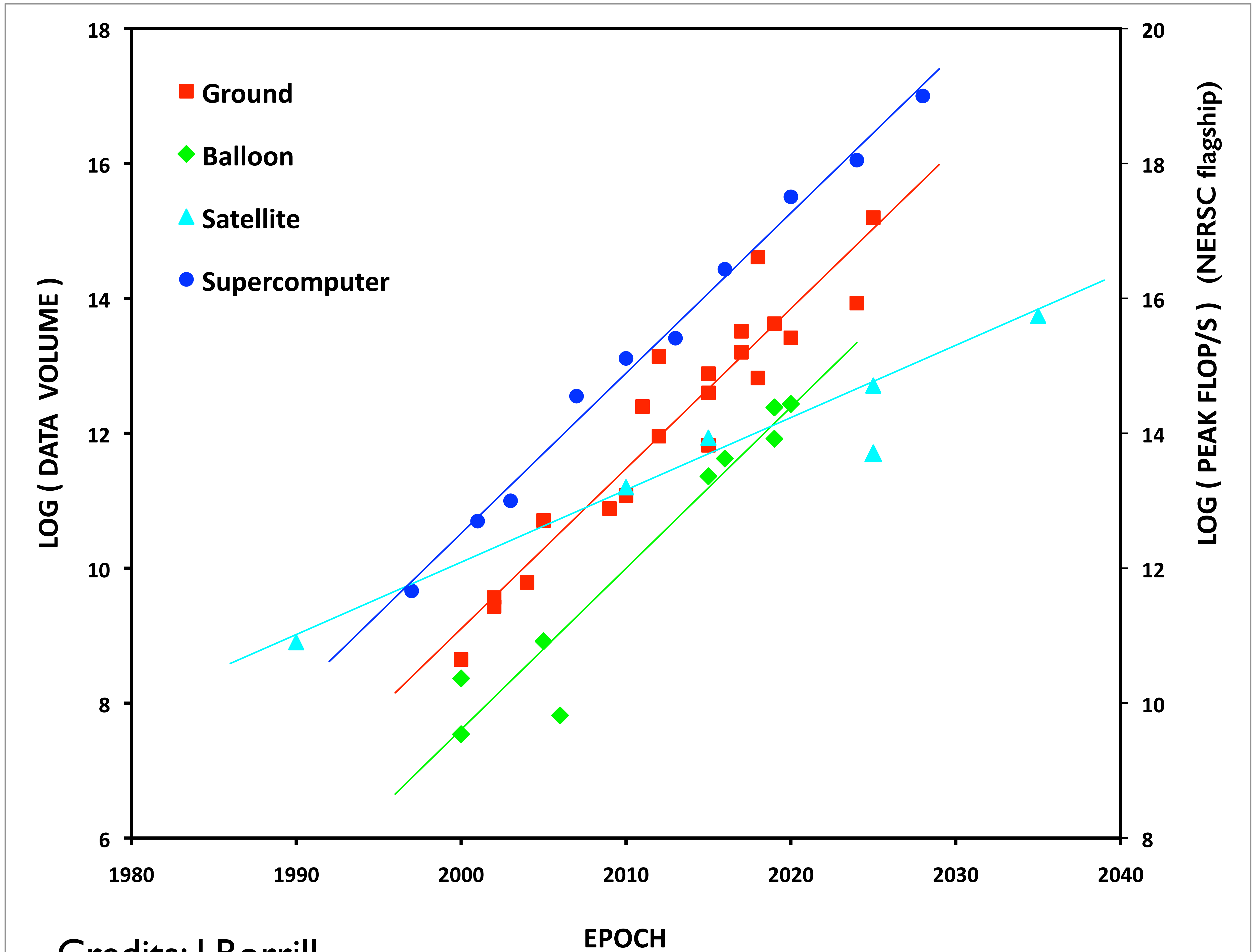
- Large computational cost requires careful implementation
  - ✓ efficiency
  - ✓ massive parallelism

Supercomputers are required

Example: Edison (NERSC)

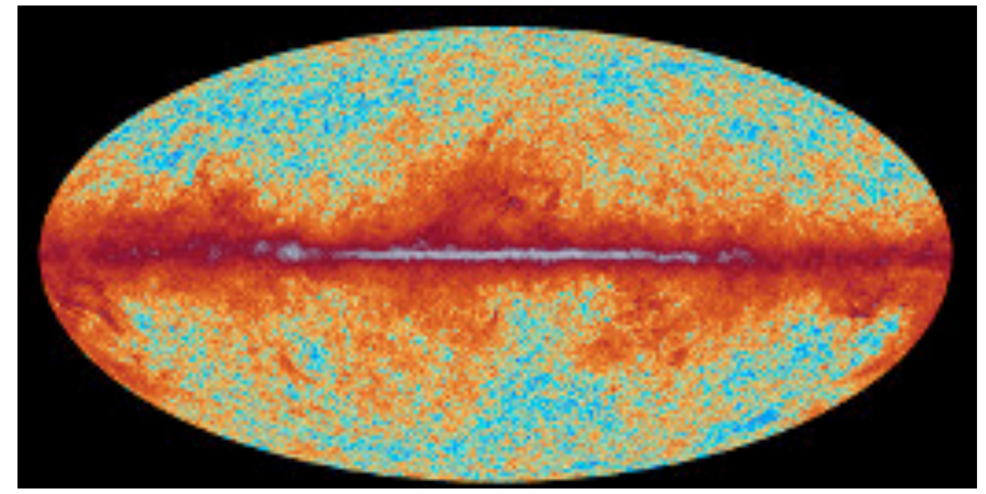
- 5586 nodes
- 24 cores/node at 2.4 GHz
- 64 GB/node RAM

Data volume rapidly increasing but exploiting the growing computational power is a challenge

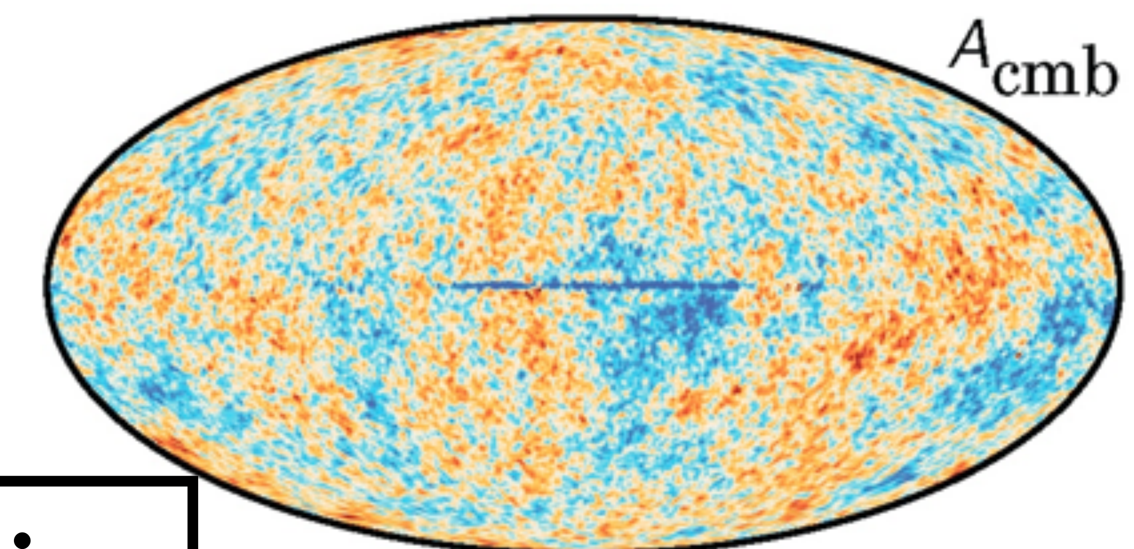


Credits: J.Borrill

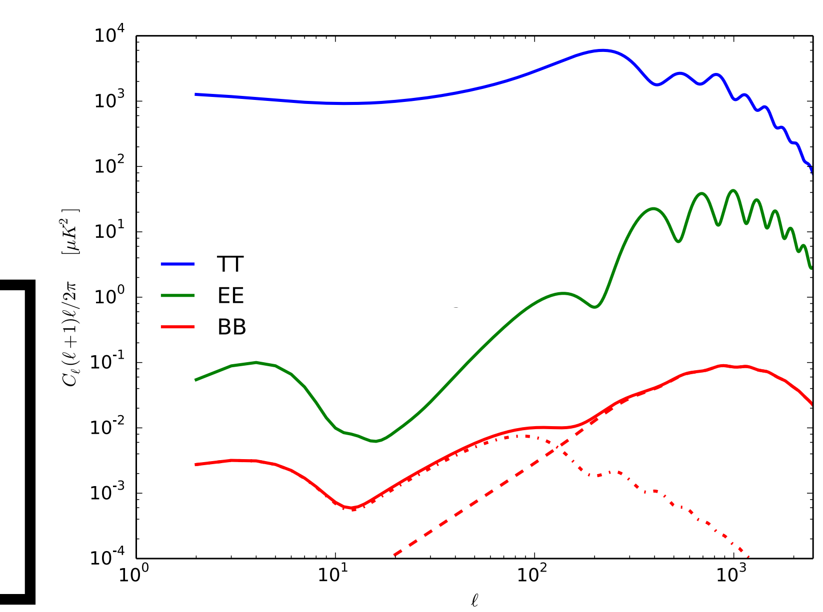
Timestream



Map-making



Component separation

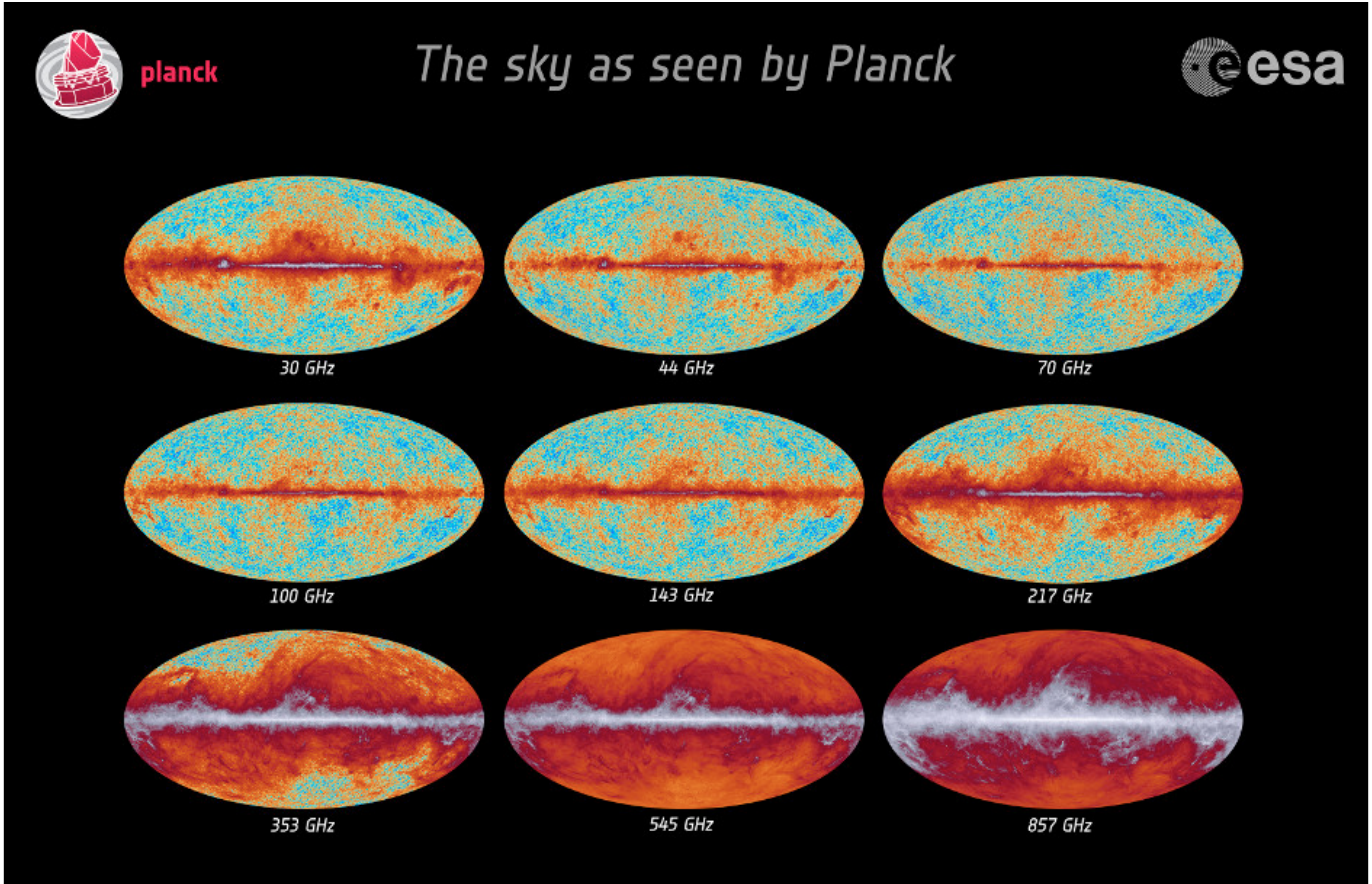


Power spectrum estimation

$\Omega_b, \Omega_m, \sigma_8, \ldots$

Parameter estimation

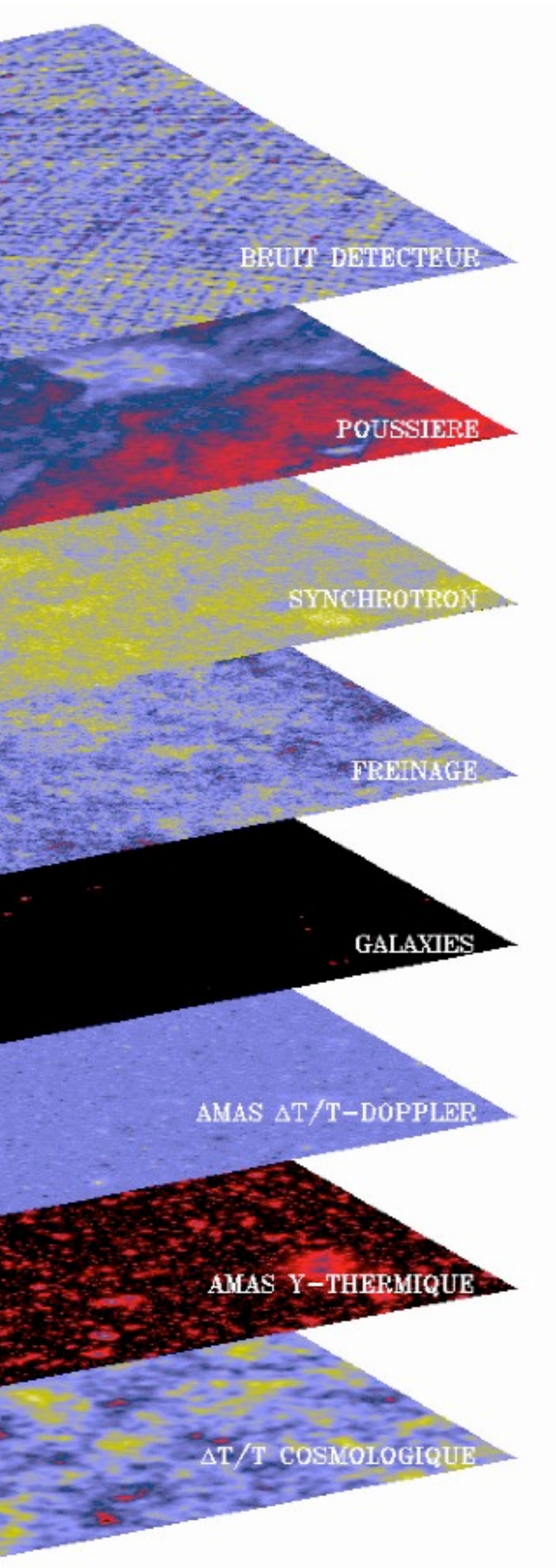
Component separation



For B-modes science, **no sky region** can be considered **foreground-free**

The **frequency dependence of the emission law** is different for different components

Component separation: from frequency maps to component maps



F.R. BOUCHET & R. GISPERT 1996

For a given direction in the sky:

Frequency dependent noise

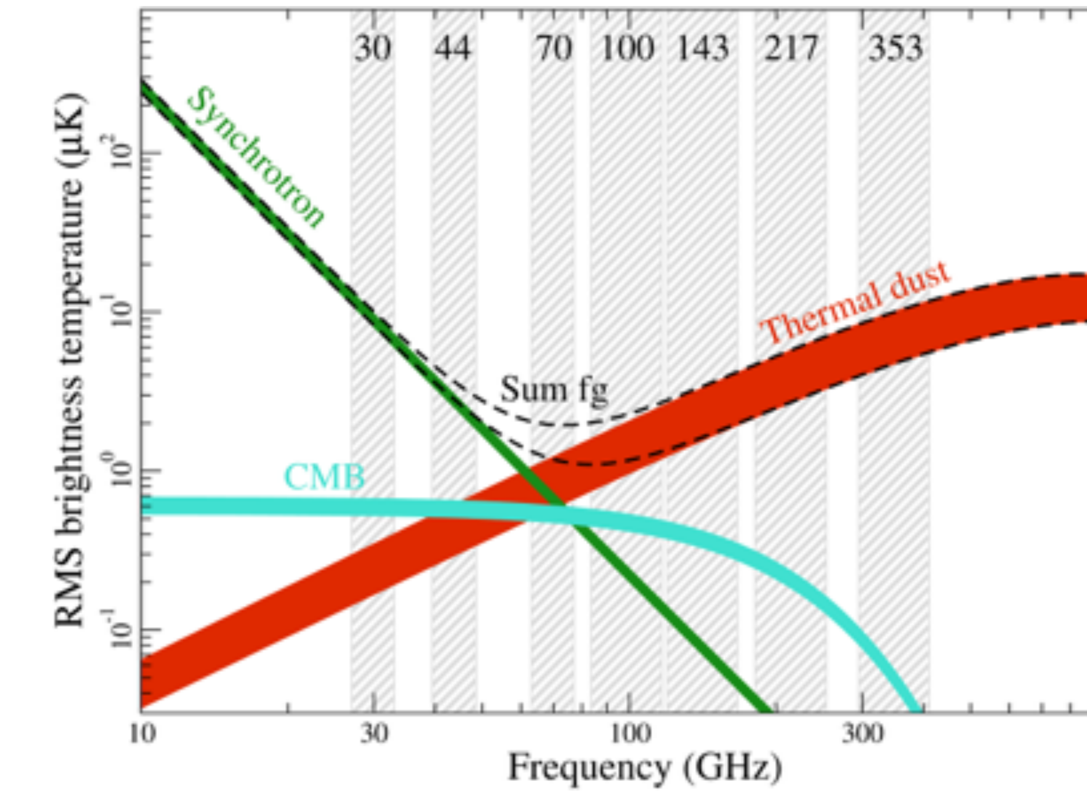
Frequency data

$$d_\nu = \sum_c a_{c,\nu} s_c + n_\nu$$

Components

Component-specific emission law

Amplitude of the component c



Equivalently

Frequency data vector

$$\mathbf{d} = \mathbf{A}\mathbf{s} + \mathbf{n}$$

Noise vector

Mixing matrix

Vector of the components

Different assumptions, different level of blindness
Most of them are two step processes:

- Use many "pixels" to constrain a distinctive property of the components
- Use this property to separate the components

Examples

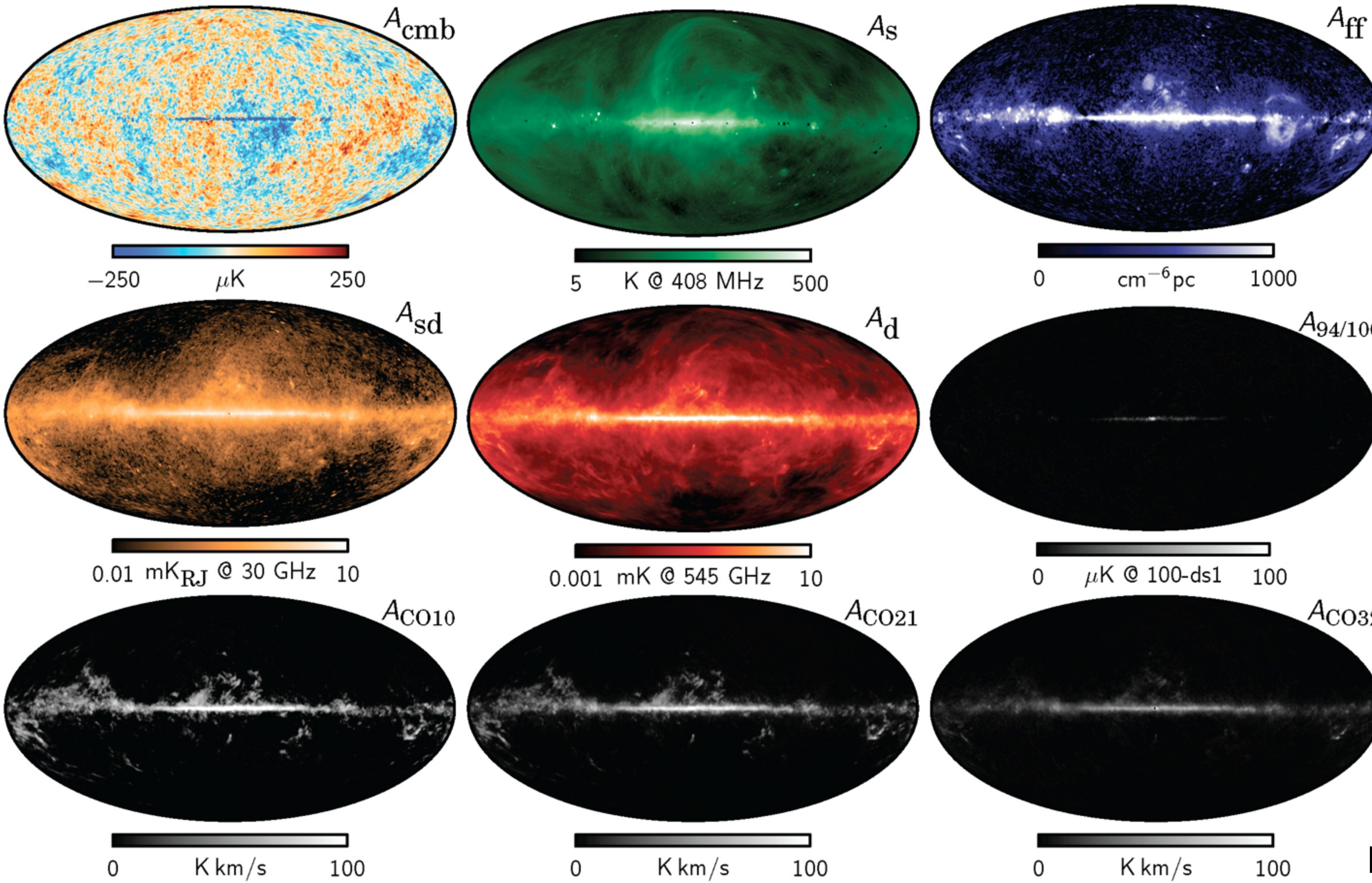
- Assume perfect knowledge of A ➡ GLS solution
- Assume knowledge of A up to some free parameter β. Minimize $(\mathbf{d} - \mathbf{A}(\beta)\,\mathbf{s})^\top \mathbf{W}(\mathbf{d} - \mathbf{A}(\beta)\,\mathbf{s})$ with respect to both s and β.
  Maximum likelihood parametric fitting. e.g. Stompor et al. (2008)
- Assume the scaling law of CMB, $\boldsymbol{a}$. Estimate the CMB as a linear combination, minimizing the variance of the output

  $$\hat{s}_{\mathrm{CMB}} = \frac{\boldsymbol{a}^t\,\widehat{\mathrm{R}}^{-1}}{\boldsymbol{a}^t\,\widehat{\mathrm{R}}^{-1}\,\boldsymbol{a}}\,\mathbf{d}$$

  $\widehat{\mathrm{R}}$ is the empirical covariance matrix

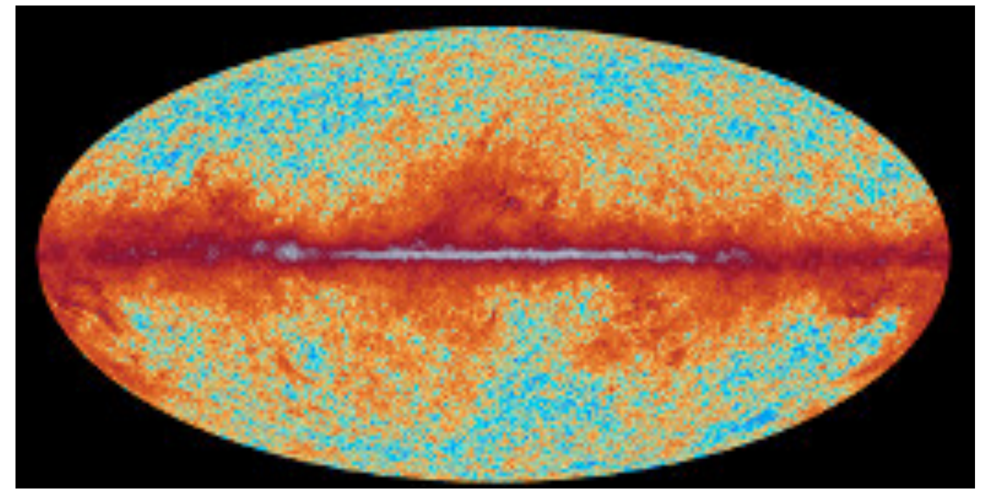
  Internal linear combination. e.g. Delabrouille et al. (2009)
- *Many more (Independent component analysis, template fitting…), not necessarily in pixel domain (but, e.g., harmonic, needlet…)*

CMB and foregrounds maps

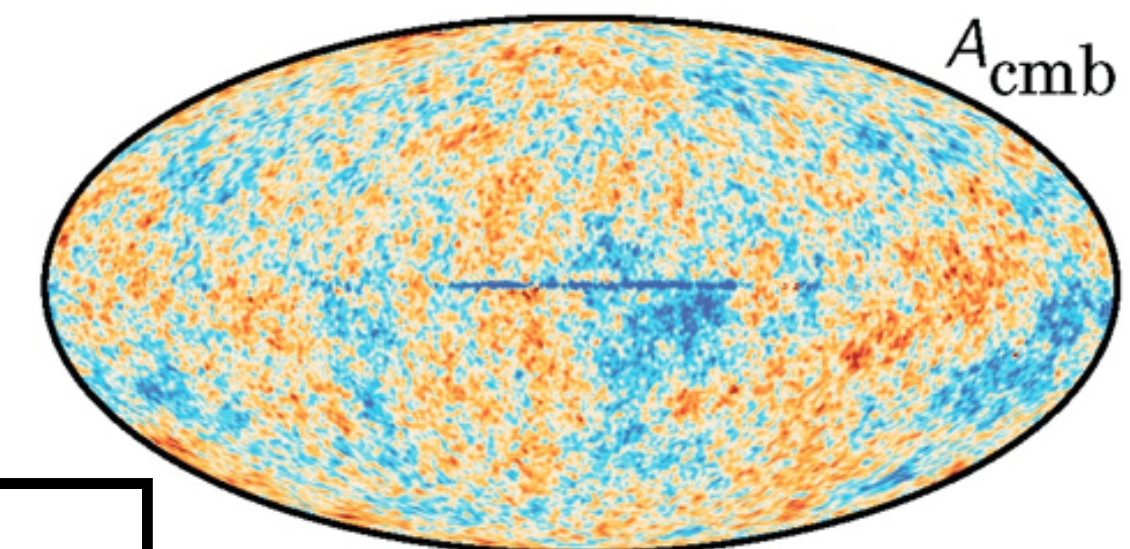


Planck I (2015)
(combined with WMAP)

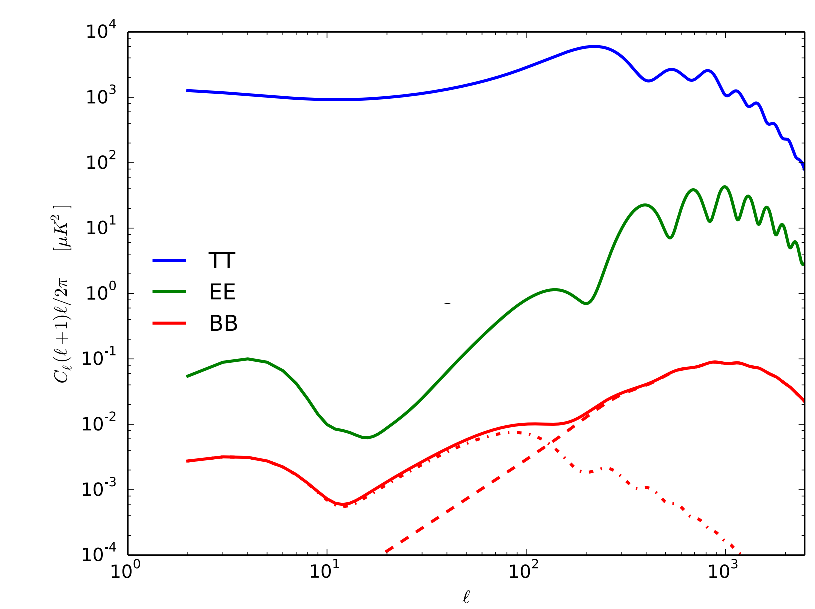
Timestream



Map-making



Component separation



Power spectrum estimation

Ω_b, Ω_m, σ_8, …

Parameter estimation

Angular power spectrum estimation: very basics

$C(\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}') \equiv \langle X^*(\hat{\boldsymbol{x}}) X(\hat{\boldsymbol{x}}') \rangle$ Two point correlation function

Assuming statistical isotropy

$$C(\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}') \equiv C(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}}') \equiv \frac{1}{4\pi} \sum_{\ell=0}^{\infty} (2\ell + 1) C_\ell P_\ell(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}}') \quad \Rightarrow \quad \langle a^*_{\ell m} a_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m m'} C_\ell$$

Expansion on the Legendre polynomials

Simplest power spectrum estimator

$$\hat{C}_\ell = \sum_{m=-\ell}^{\ell} \frac{a^*_{\ell m} a_{\ell m}}{2\ell + 1} \qquad \langle \hat{C}_\ell \rangle = C_\ell \qquad \mathrm{Var}(\hat{C}_\ell) = \frac{2}{2\ell + 1} C_\ell^2 \frac{1}{f_{sky}}$$

Computational cost (due to the SHT) $\mathcal{N}_p^{3/2}$

Cosmic variance

Partial sky coverage and (inhomogeneous) noise

CMB map $\hat{\mathrm{s}} = \boldsymbol{s} + \boldsymbol{n}$ covering only a fraction of sky, $\boldsymbol{n}$ Gaussian but not isotropic

Covariance of the map $\boldsymbol{N} + \boldsymbol{C}$

Not a function of $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}}'$

$$C(\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}') \equiv \frac{1}{4\pi} \sum_{\ell=0}^{\infty} (2\ell+1) C_\ell P_\ell(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}}')$$

Maximize the likelihood

$$P(\boldsymbol{x}|C_\ell) = \mathcal{L}(C_\ell) = \frac{\exp\{\frac{1}{2}\hat{\mathrm{s}}^\top \boxed{(\boldsymbol{N}+\boldsymbol{C})^{-1}} \hat{\mathrm{s}}\}}{\sqrt{(2\pi)^{\mathcal{N}_p} \boxed{\det(\boldsymbol{N}+\boldsymbol{C})}}}$$

$\mathrm{cost} \sim \mathcal{N}_p^3 \sim \ell_{\mathrm{max}}^6$

➡ Doable only for large scales (ell < few tens)

➡ Small scales: keep the $\mathcal{N}_p^{3/2}$ scaling (with large or very large pre-factor)

- Pseudo-power spectrum estimators
- Gibbs sampling

Pseudo-power spectrum estimators

Problem with $\hat{C}_\ell = \sum_{m=-\ell}^{\ell} \frac{a^*_{\ell m} a_{\ell m}}{2\ell+1}$: a_{lm} are not available (partial sky coverage)

C_l is estimated by

- sampling the power of s along functions in the l subspace (the $\{Y_{lm}\}_{m\in\{-l\ldots l\}}$ functions)
- averaging (the expected power on each function is C_l)

In this case the functions are orthogonal but don't have to be

If a function f straddles multiple l subspaces, the expected power of s along f is a **(known) linear combination of C_l**

Define a pseudo-basis of such f functions: $\tilde{Y}_{\ell m}$ (any set!) and sample s along them. $\tilde{a}_{\ell m} \equiv \int \tilde{Y}^*_{\ell m} s \mathrm{d}\Omega$

$\tilde{C}_\ell = \sum_{m=-\ell}^{\ell} \frac{\tilde{a}^*_{\ell m} \tilde{a}_{\ell m}}{2\ell+1}$ Pseudo-power spectrum

$$\langle \tilde{C}_\ell \rangle = \sum_{\ell'} M_{\ell\ell'} C_{\ell'}$$

Compute and "invert" the M_{ll} and and you are done

Notable example: $\tilde{Y}_{\ell m} \equiv W Y_{\ell m}$

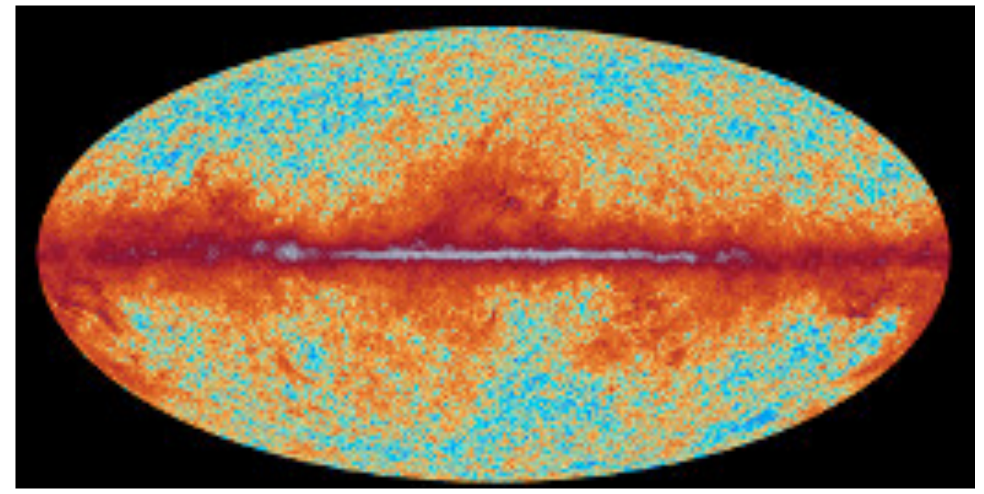
W is the inverse noise

- handle cut sky
- handle inhomogeneous coverage
- Computing the M_{ll} scales as $N_p^{3/2}$

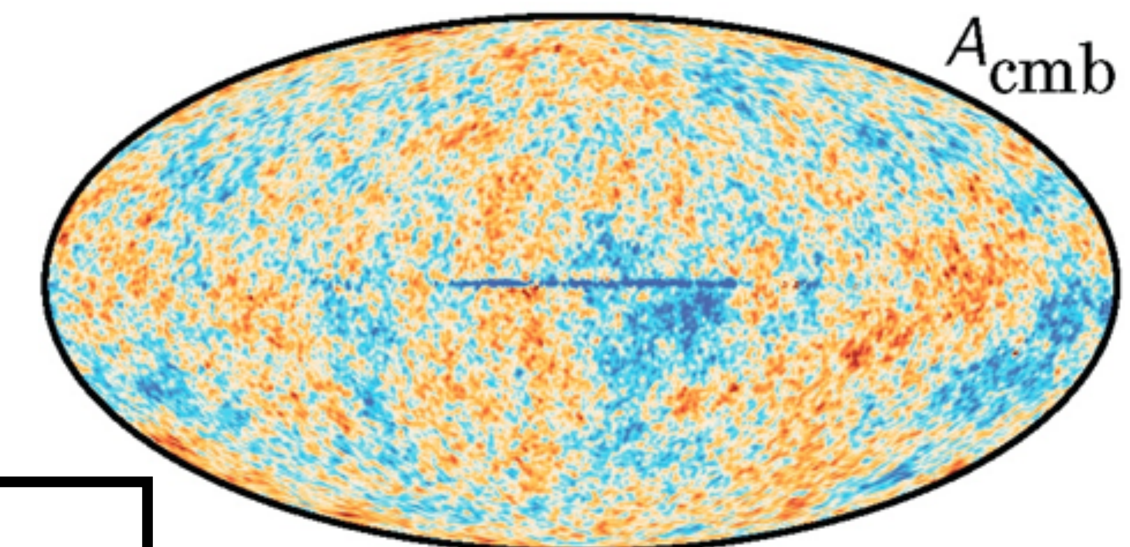
Hauser and Peebles (1973), Hivon et al. (2002), Kogut et al. (2003)

Overview

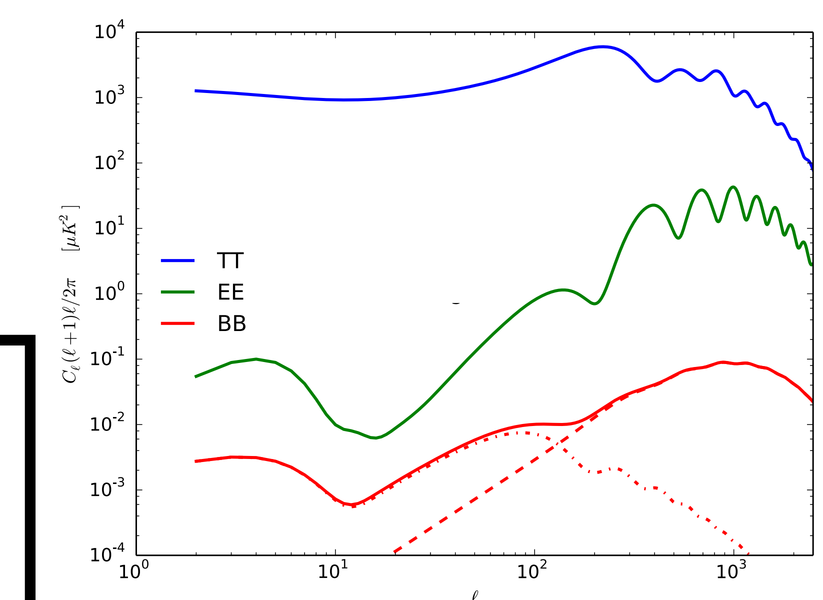
Timestream



Map-making



Component separation



Power spectrum estimation

Ω_b, Ω_m, σ_8, …

Parameter estimation

Cosmological parameter estimation

Likelihood of the *cosmological parameters* θ

- $C_\ell(\theta)$, so $\mathcal{L}(C_\ell)$ is actually $\mathcal{L}(\theta)$

  - CAMB
  - CLASS
  - …

  - previous likelihood
  - approximations
  - …

No analytical form.

- How to maximize it?
- What shape? (uncertainties, correlations, degeneracies…)

Evaluate on grid of θ ? Impossible or inefficient

MCMC algorithms sample from the θ distribution $\mathcal{L}(\theta)$

➡ chain of values of θ whose density is proportional to $\mathcal{L}(\theta)$

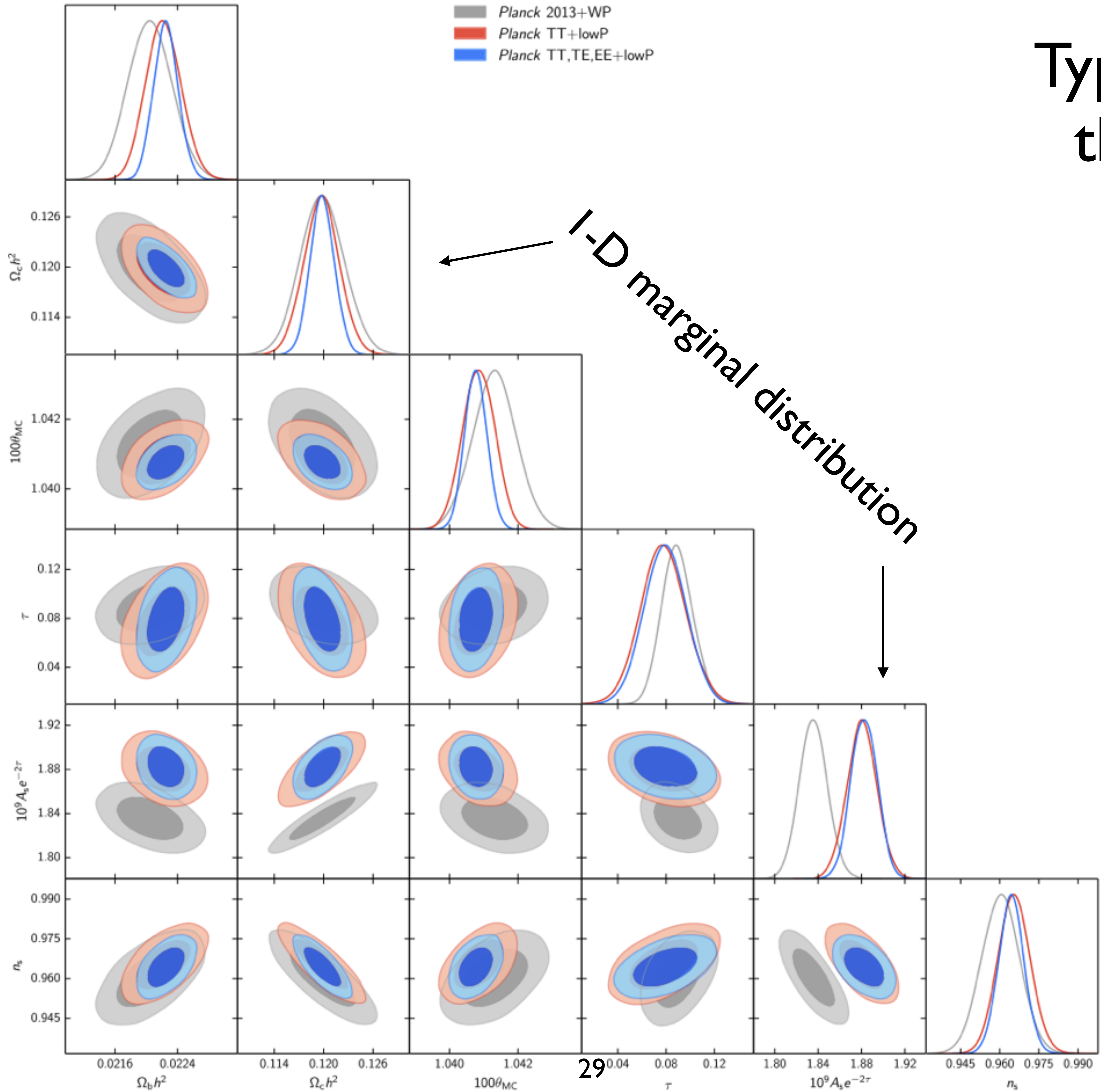
- θ_0
- rule for θ_{i+1} from θ_i
  - ➡ rejection criterion based on $\mathcal{L}(\theta_i)$ and $\mathcal{L}(\theta_{i+1})$

Challenge: fast but accurate evaluation of $\mathcal{L}(\theta)$, low rejection rate

Typical analysis of the MC chains

1-D marginal distribution

2-D marginal distribution

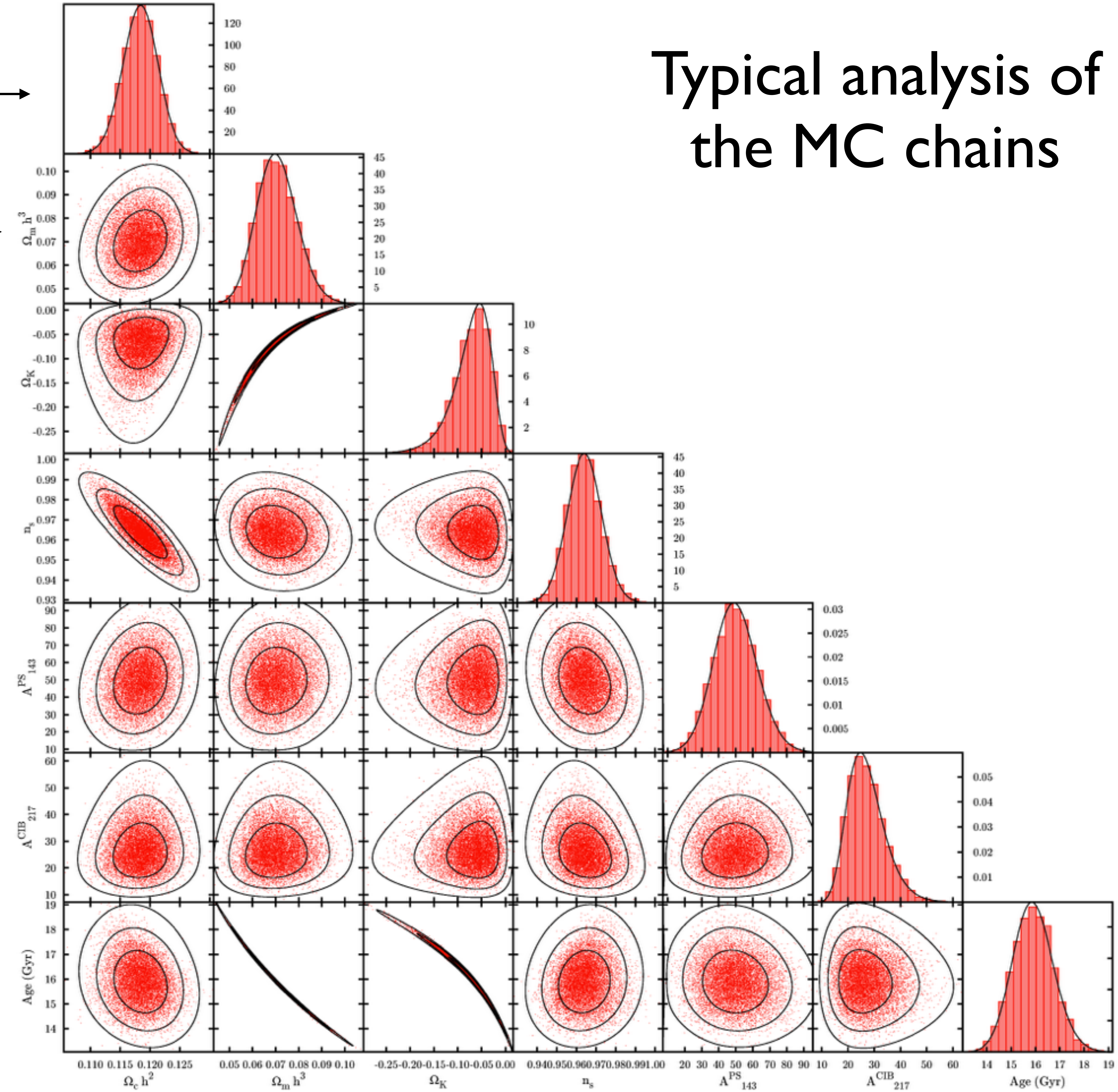


Planck XI (2015)

Cosmological parameter estimation

1-D marginal distribution

2-D marginal distribution



Typical analysis of the MC chains

Schuhmann et al. (2016)

Note: actually in this example the contours are not derived from the points

Timestream

Map-making

Wrap up

Component separation

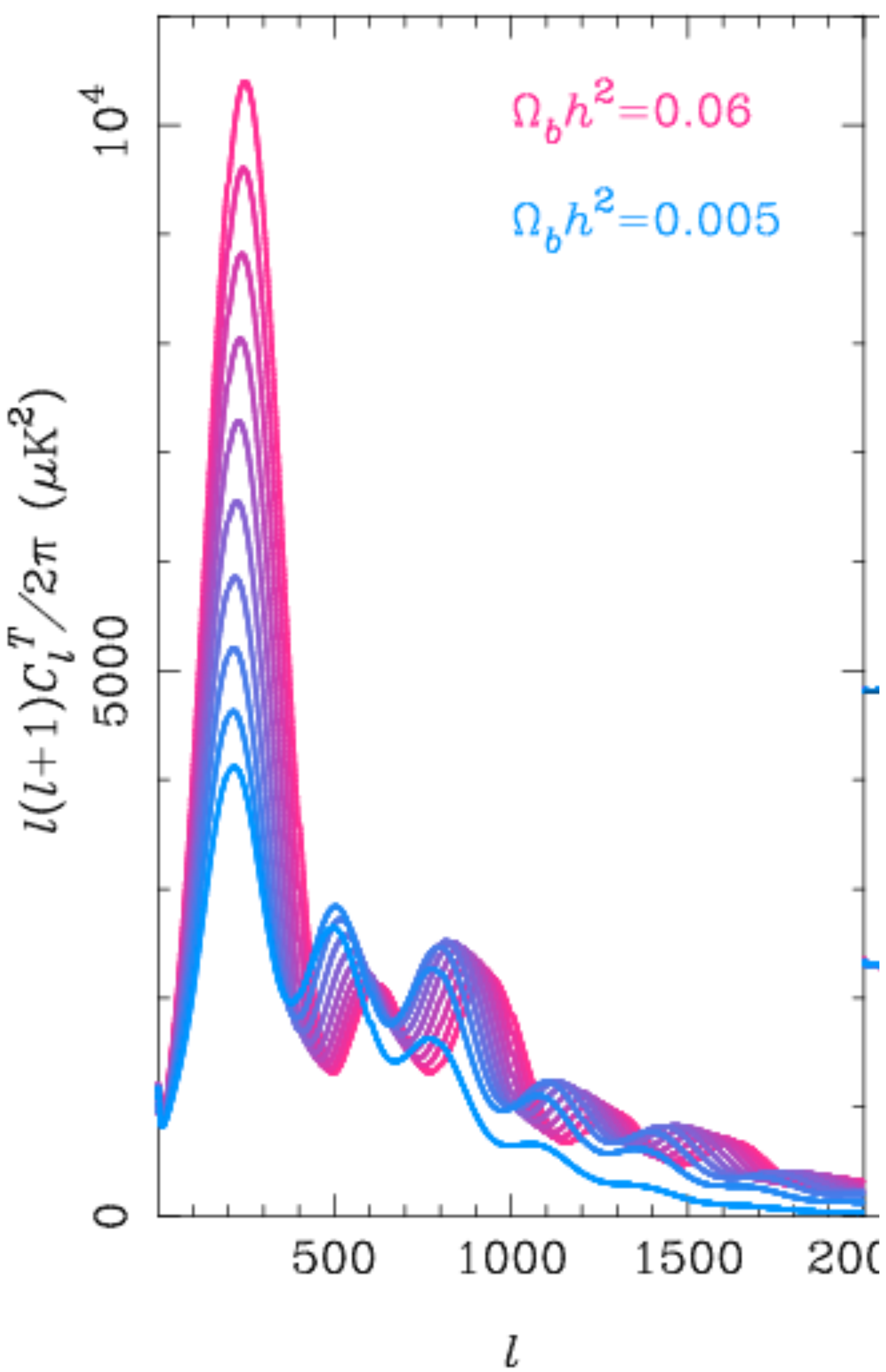
Power spectrum estimation

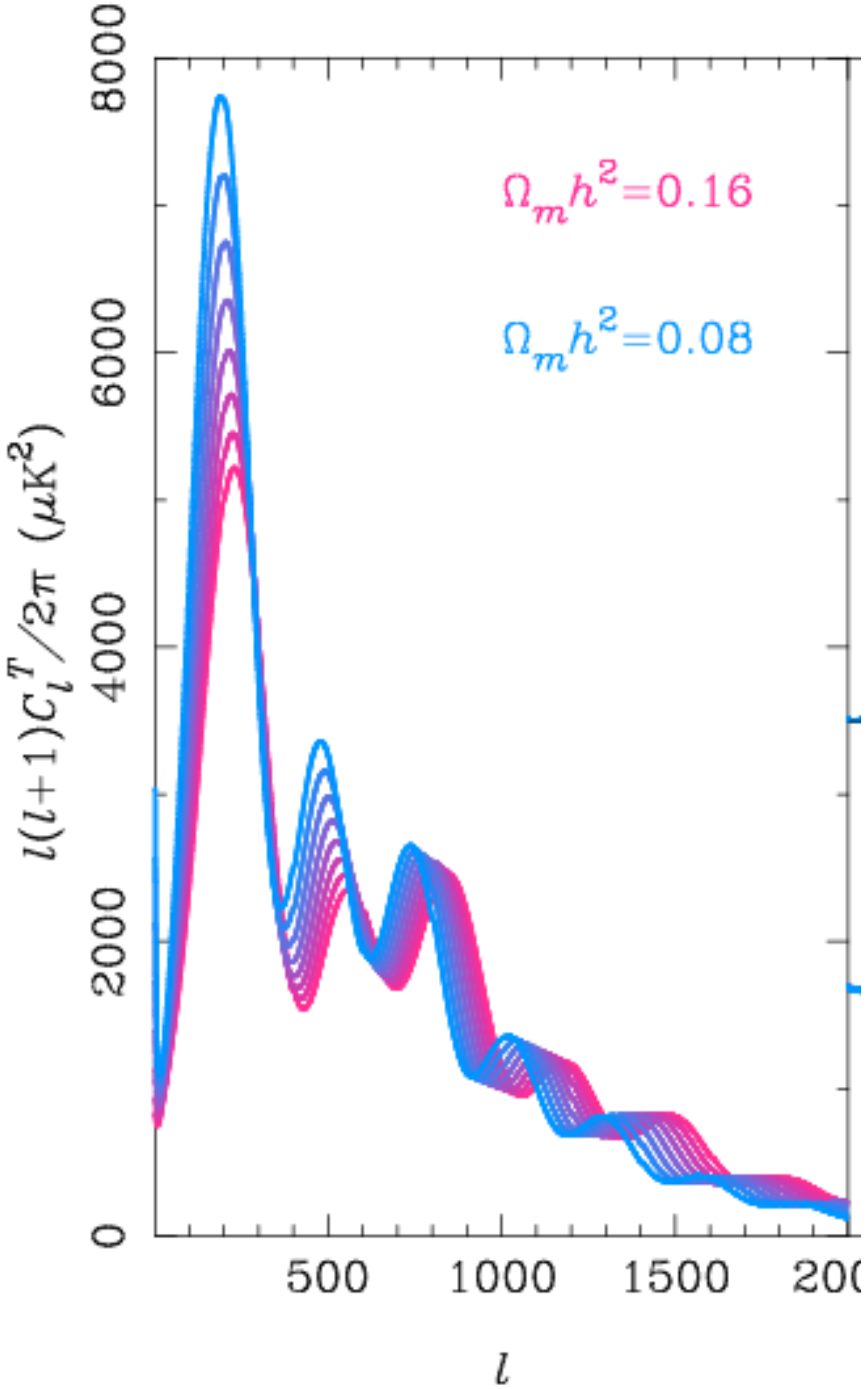
$\Omega_b, \Omega_m, \sigma_8, \ldots$

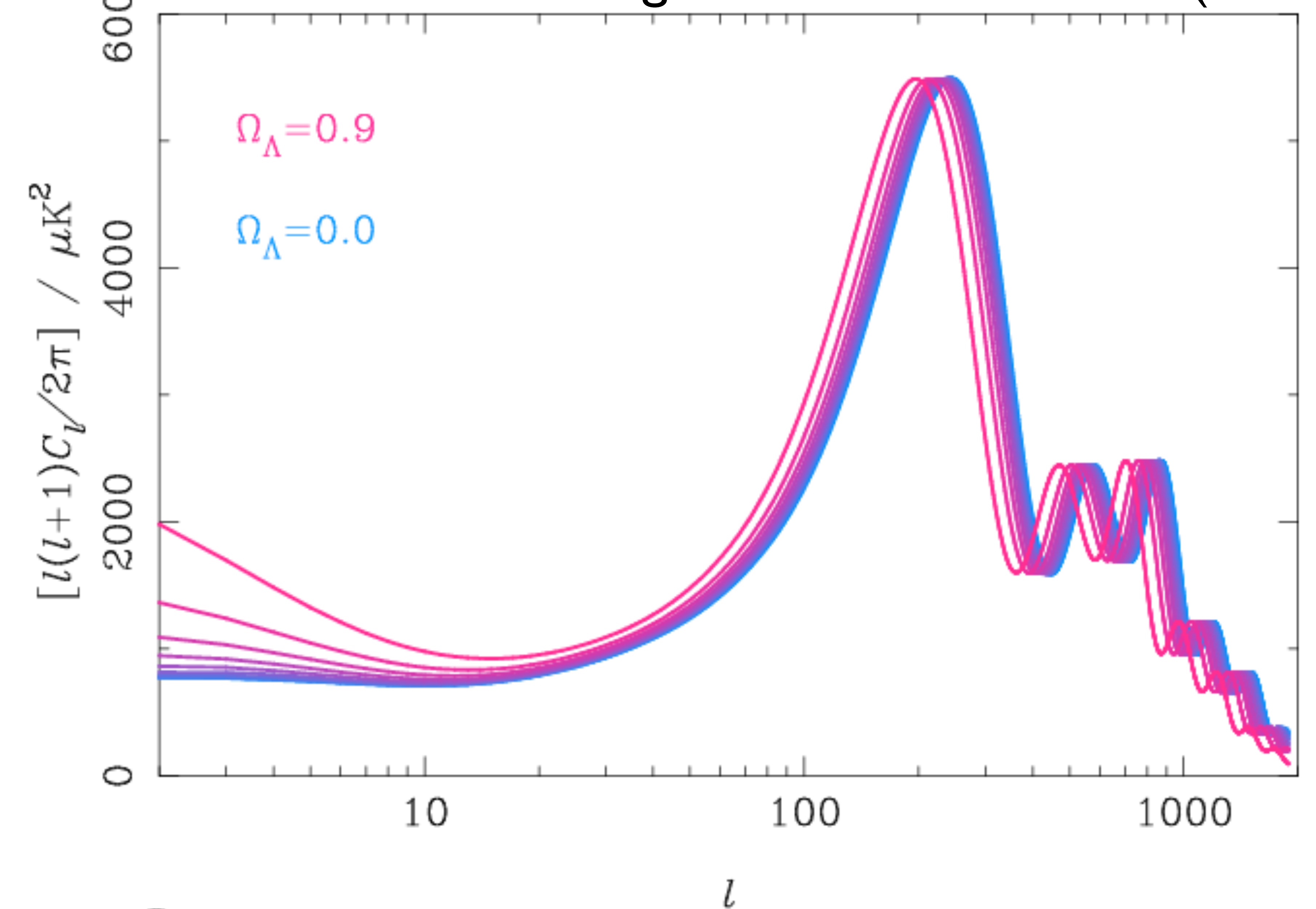
Parameter estimation

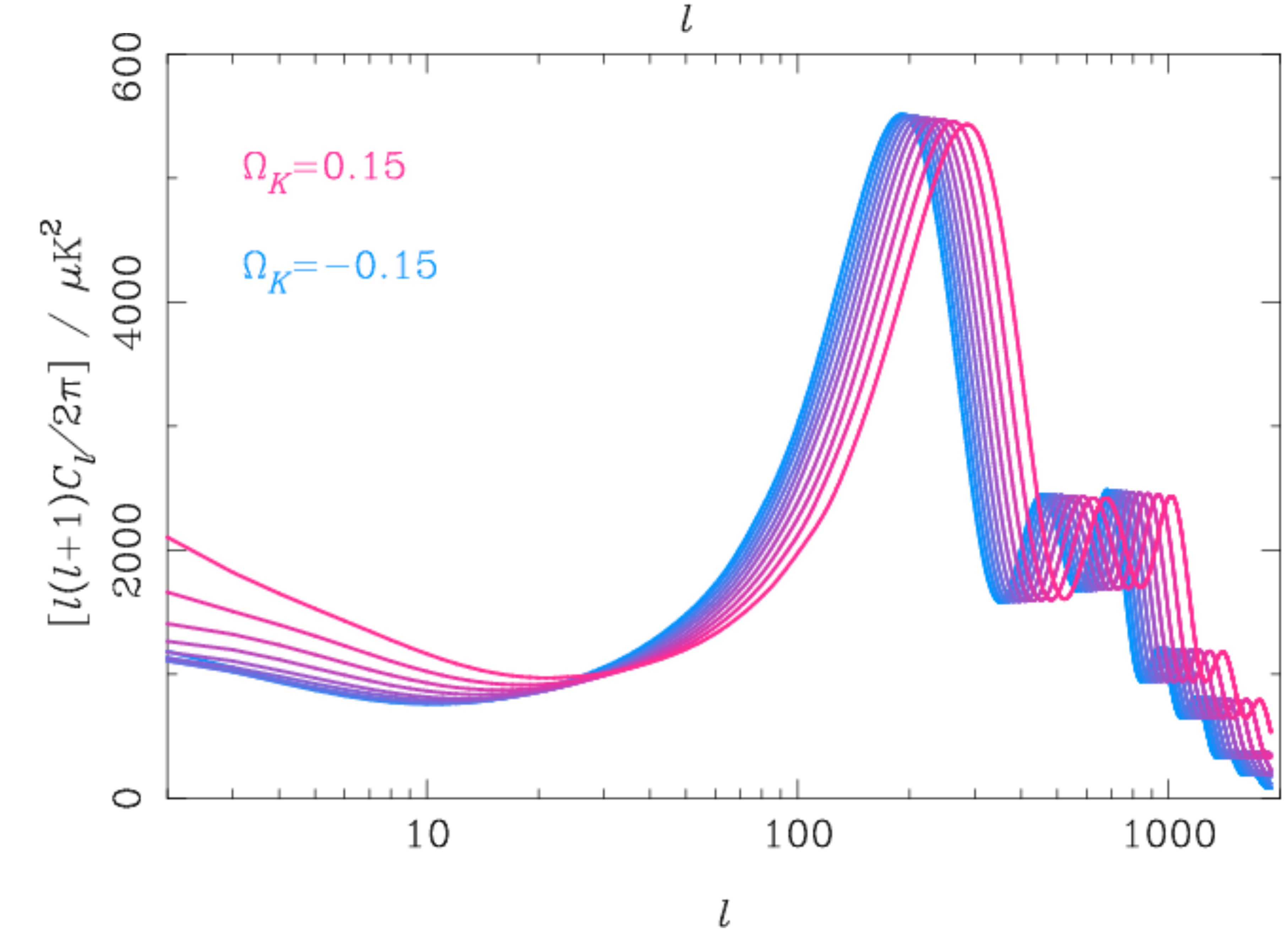
Recap: amazing physics with the CMB angular power spectrum

All figs: Challinor and Peiris (2009)



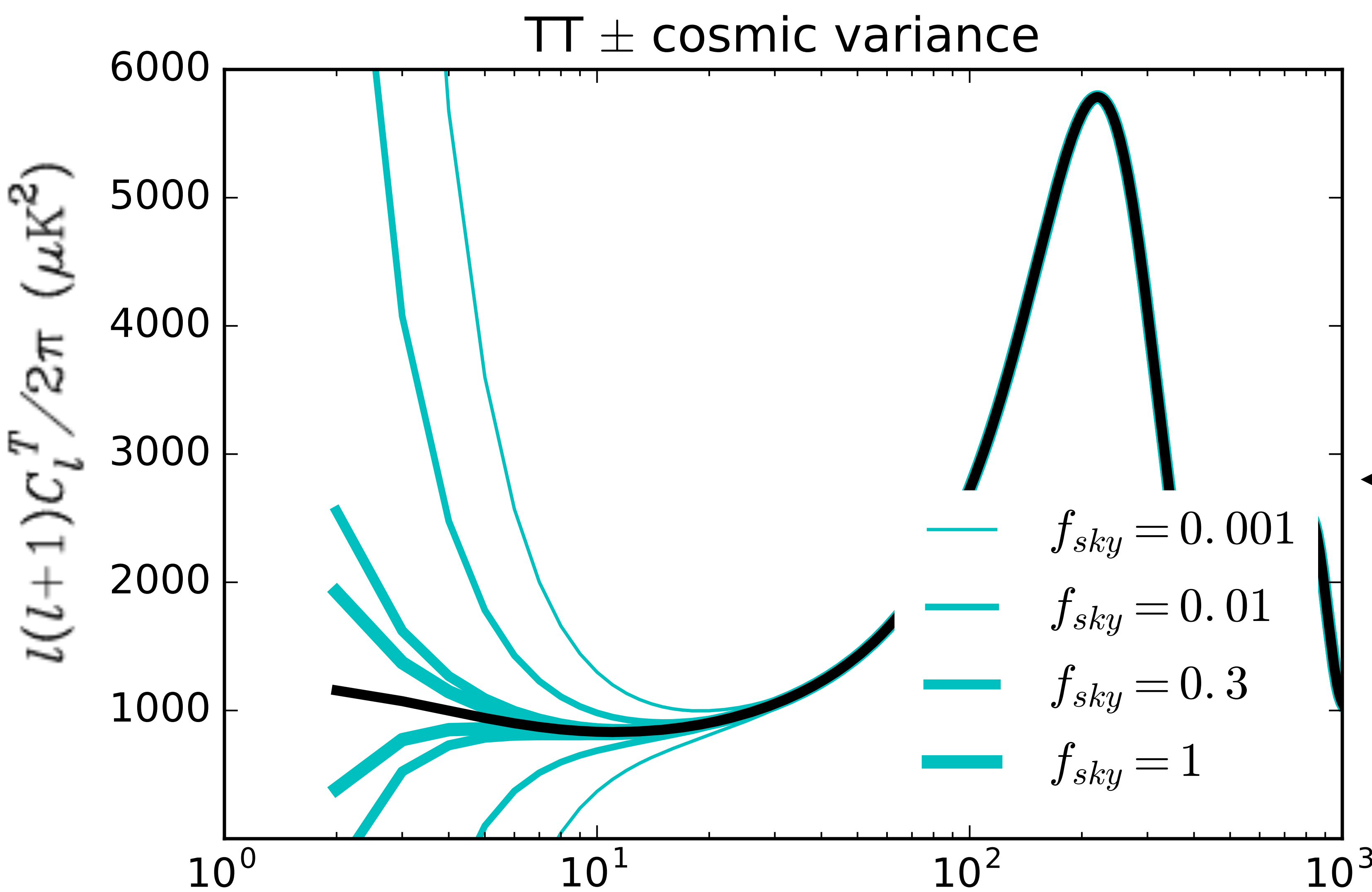


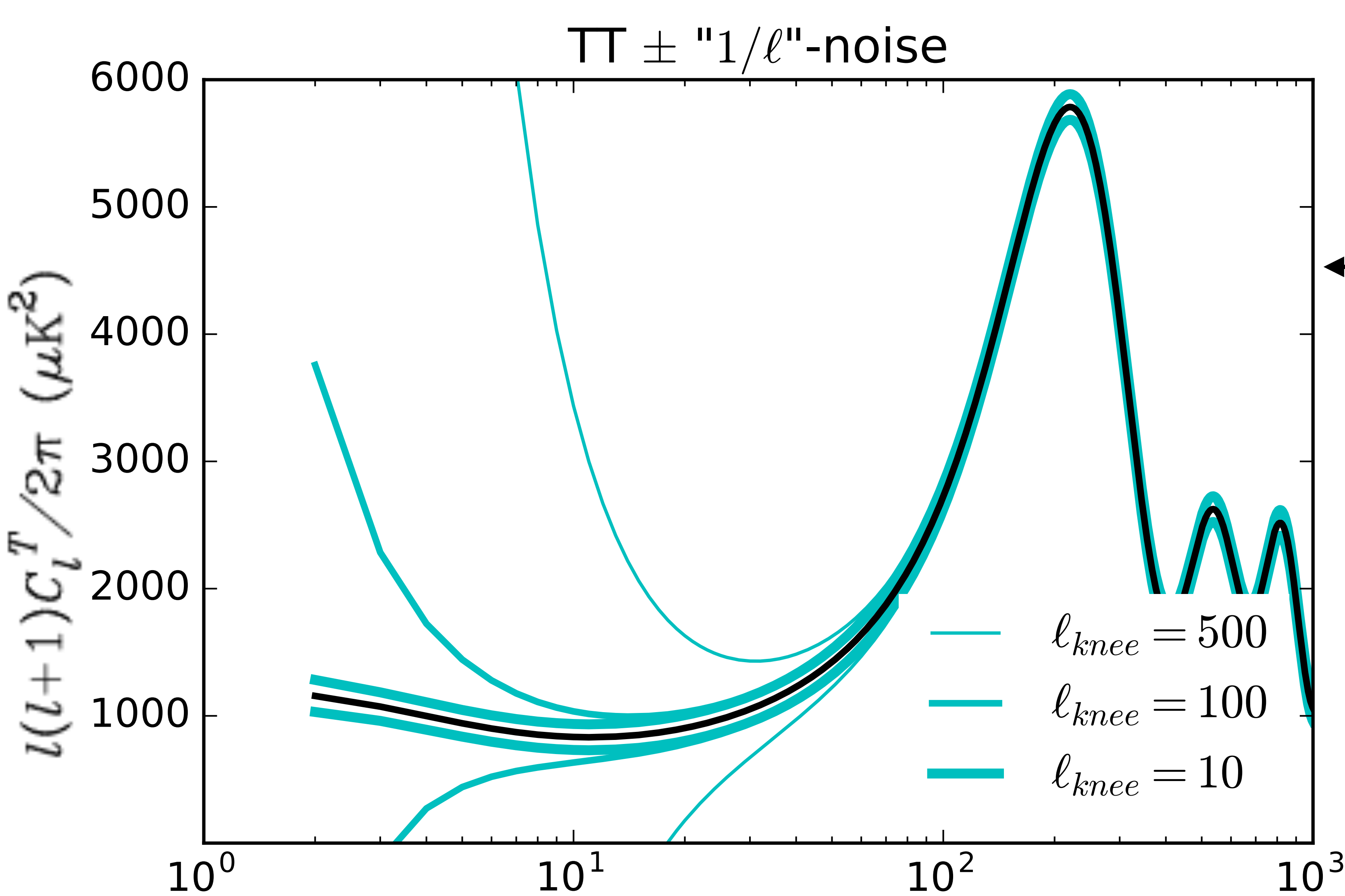




Accurately measure crucial cosmological parameters

Recap: angular power spectrum main uncertainties





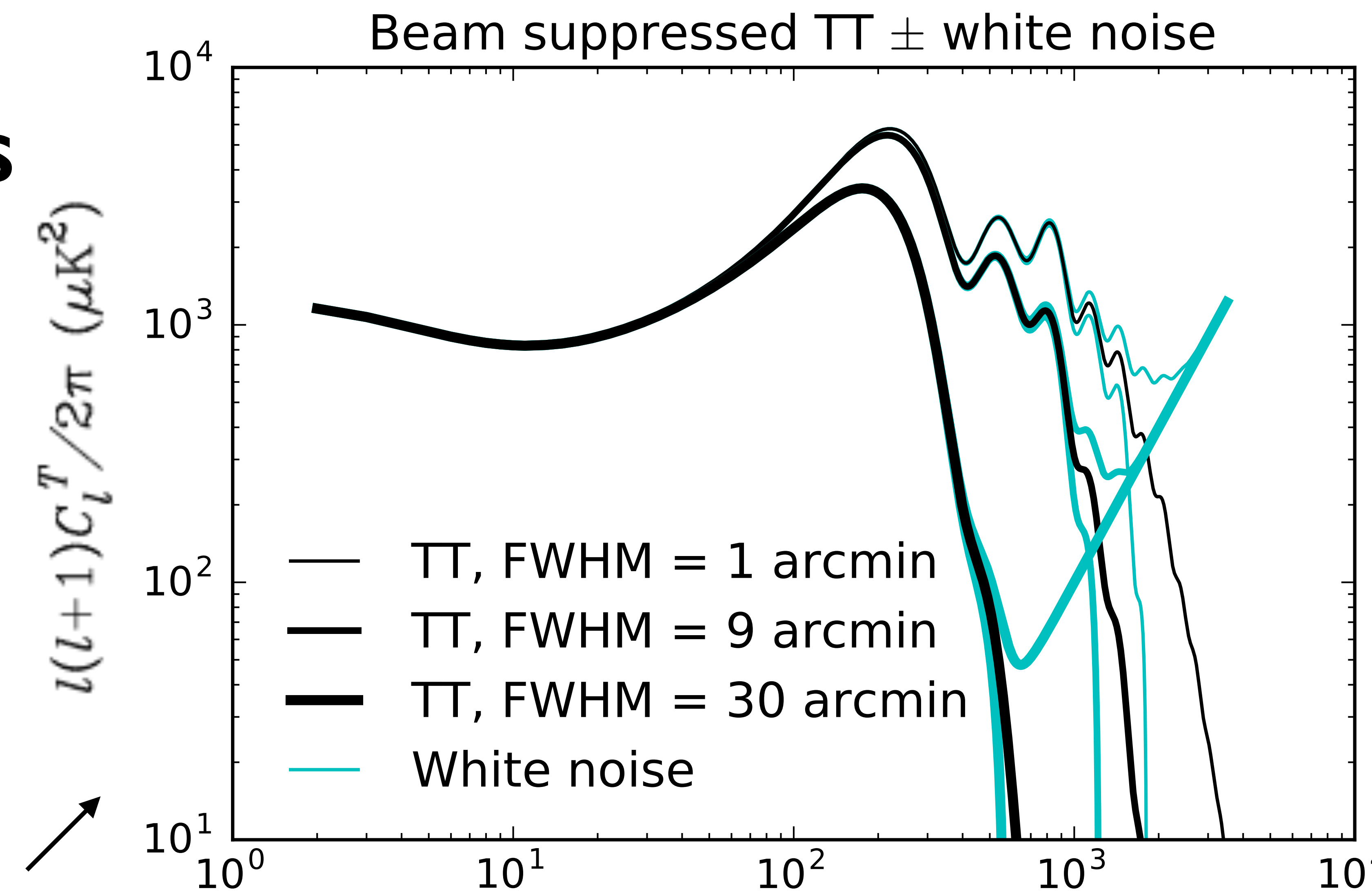
Large Scales

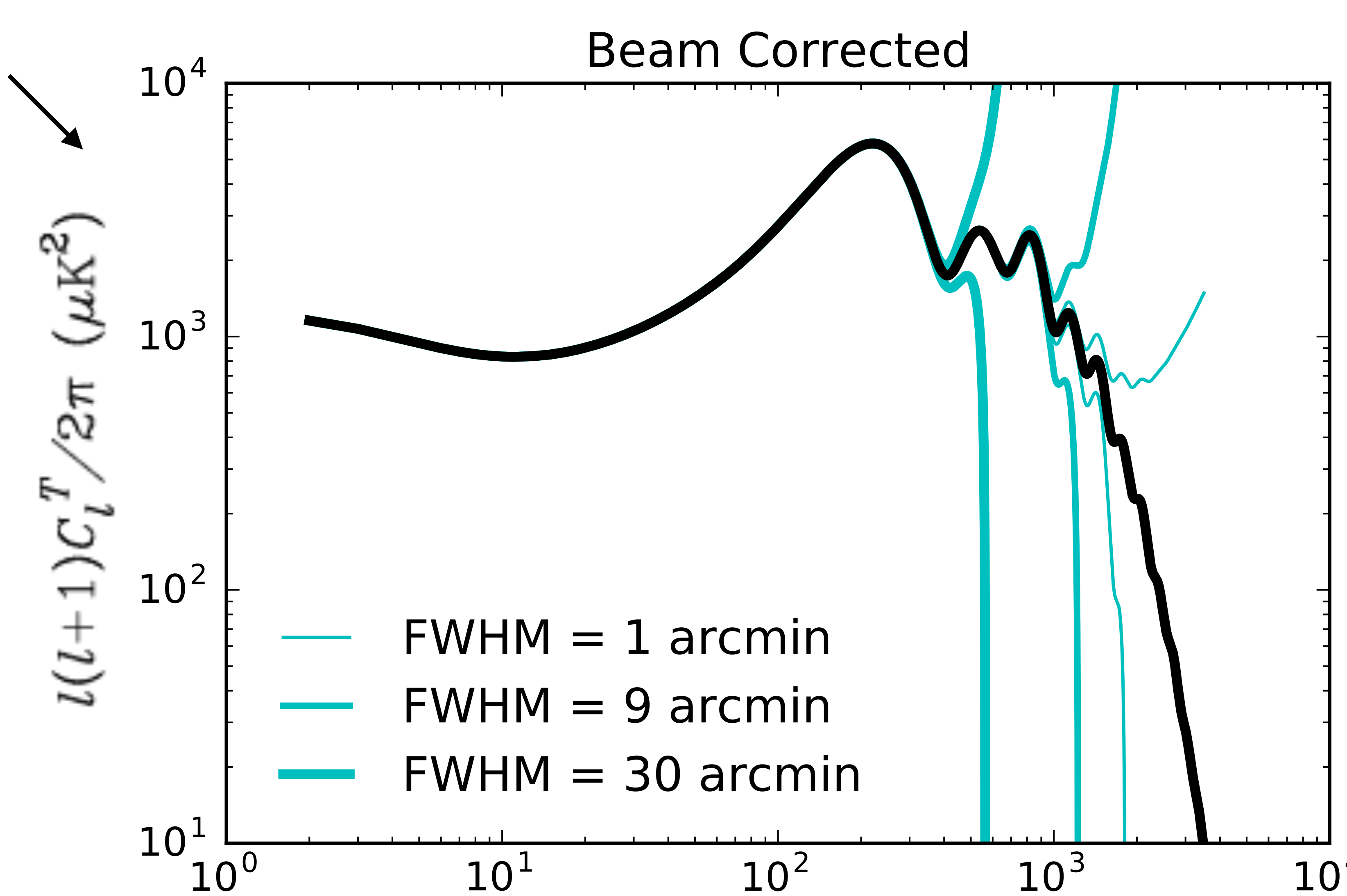
Cosmic variance

Correlated noise

Small Scales

White noise





Foregrounds

Galactic

Extra-galactic

Final remarks

CMB science is much more than its power spectrum

- Sunyaev-Zeldovich effect
- Lensing
- Cross-correlations
- Spectral distortions
- Non gaussianities
- l.o.s integrated effects (e.g., cosmic birefringence)
- …

Far infrared and microwave observations are not only CMB

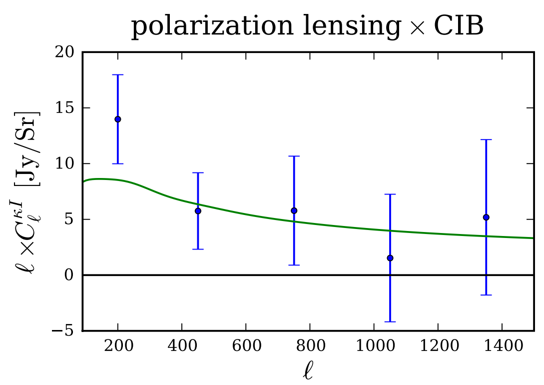
- Cosmic infrared background
- Synchrotron
- Interstellar medium
- Galactic magnetic field
- …

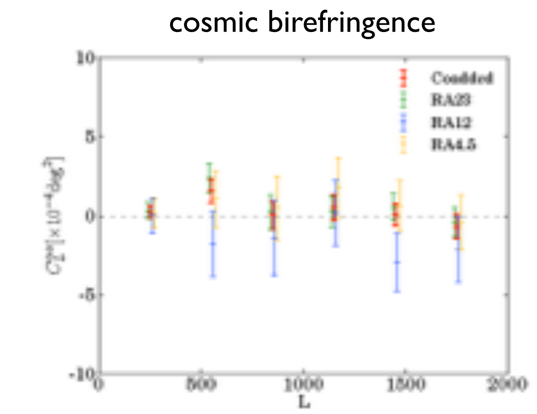
CMB data analysis (and instrumentation) is much more than the overview proposed

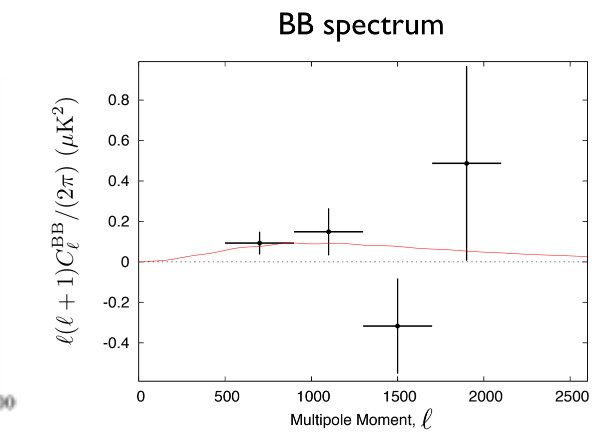
Join the force

The SISSA group is at the forefront of the CMB effort, with POLARBEAR/Simons Array.

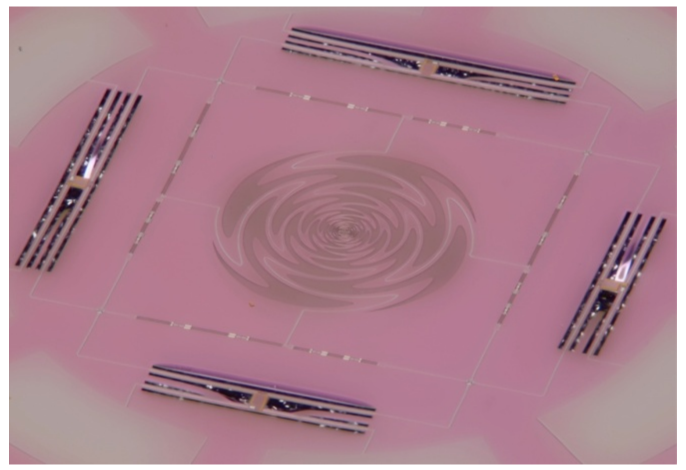
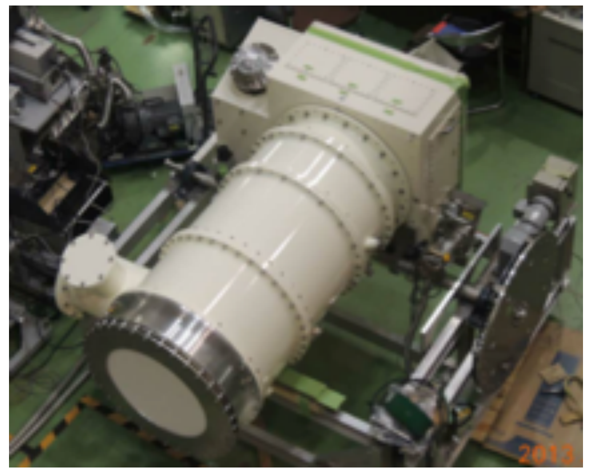
- **POLARBEAR**
  - Lensing reconstruction Phys. Rev. Lett. 112, 131302 (2014); Phys. Rev. Lett. 112, 131302 (2014)
  - BB spectrum Astrophysical J. 794, 171 (2014)
  - Cosmic birefringence Phys. Rev. D 92, 123509 (2015)



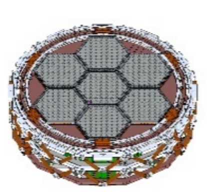


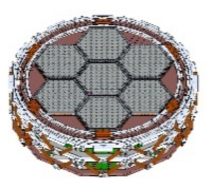


- **POLARBEAR2** (Summer 2017)
  - new telescope and receiver
    - 7,588 detectors
    - Multichroic pixels (95/150 GHz)

- **Simons Array** (Early 2018)
  new telescopes, 2 new PB2-like receivers
  - 22,764 detectors
  - 95/150/220 GHz channel

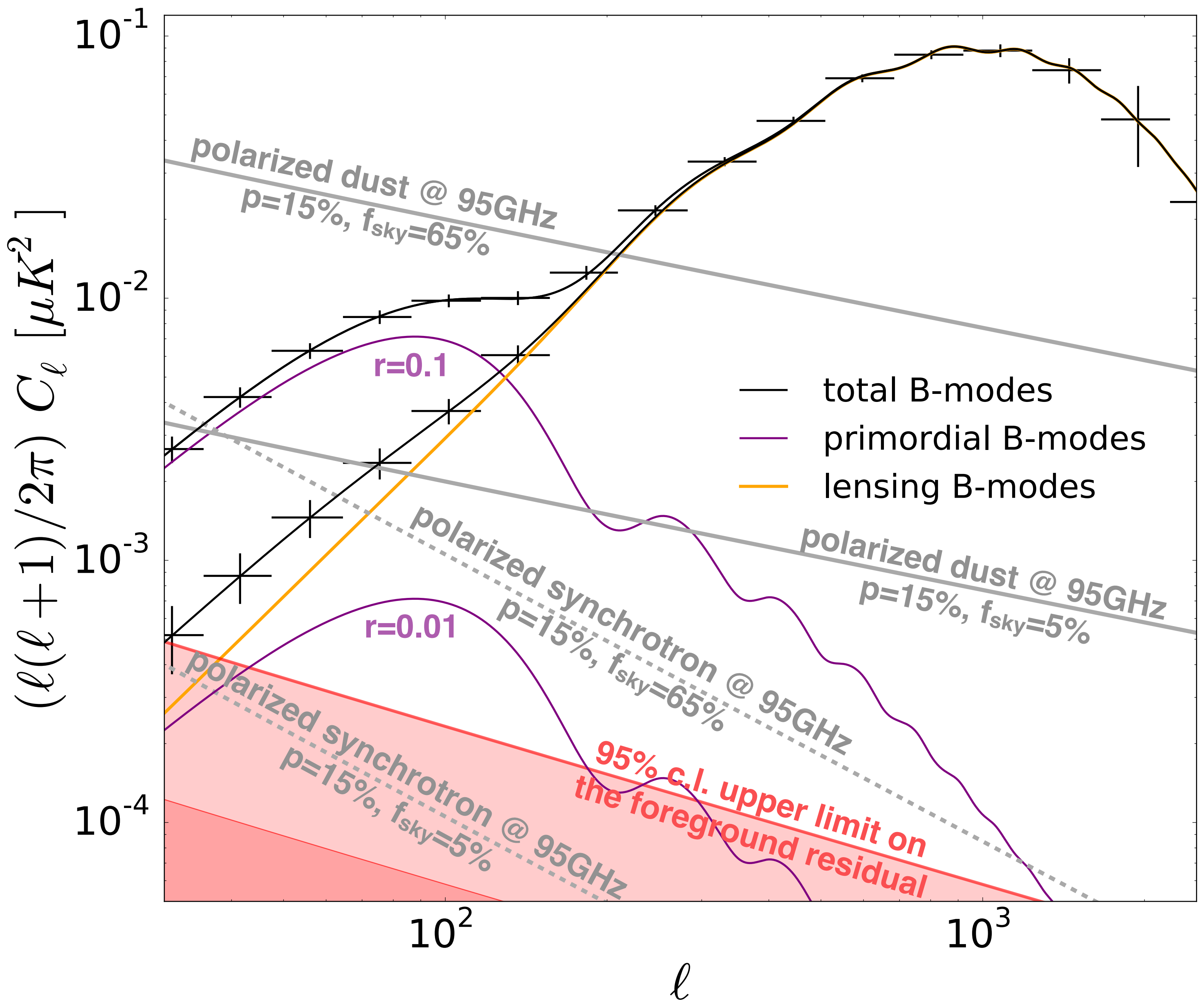
2x

95/150

1x

150/220

$$\sigma(r=0.1) = 6\cdot 10^{-3} \; (4\cdot 10^{-3})$$

$$\sigma(\Sigma m_\nu) = 40 \text{ meV} \; (19 \text{ meV})$$

Combined combined with
Planck and C-Bass with DESI BAO



- **Simons Observatory**