CMB
Angular Power Spectra and their Likelihoods: in Theory and in (Planck) Practice

E. Hivon & S. Galli

Plan

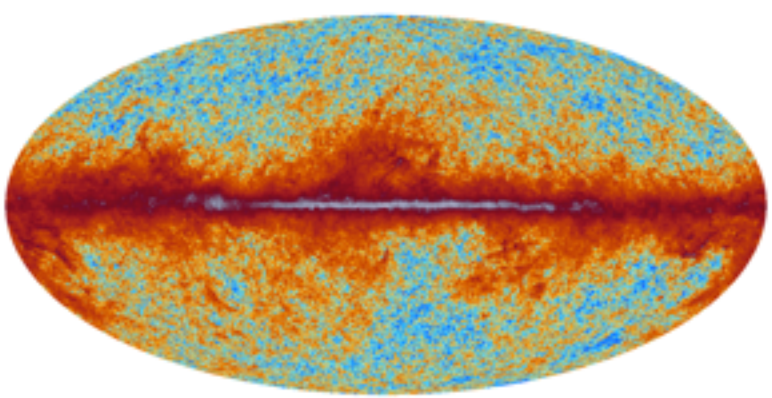
Angular Power Spectra (C_ℓ) and Likelihood

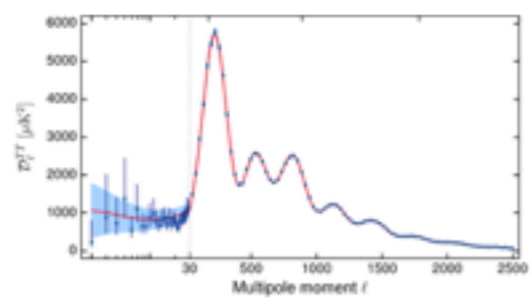
- In theory
  - What are they ?
  - Why do we need them ?
  - How to generate them fast enough on large data sets ?
- In practice
  - How to apply them to real data (Planck in particular) ?

The goal

- Goal: we want to measure parameters (cosmological+others) from the data.
  We need to evaluate the posterior distribution of the parameters $\{\Omega\}$ given the data d, $P(\{\Omega\}|d)$,
  where d is the map or the angular power spectrum C_ℓ

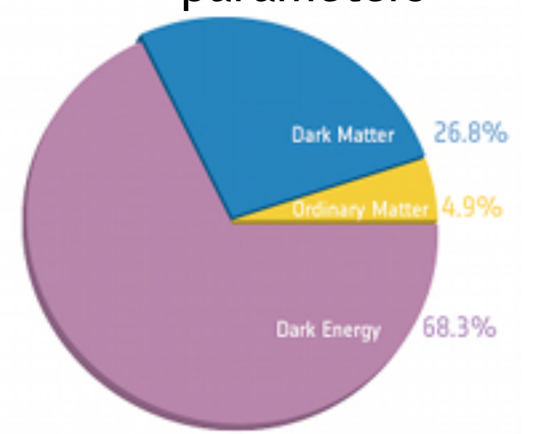
Map: 50 10^6 pixels





C_ℓ: 2500 multipoles

Cosmological (+ nuisance) parameters



Bayes theorem

- To relate the posterior of the parameters given the data *P*({Ω}|*d*) to the probability of the data given the parameters *P*(*d*|{Ω}) (the likelihood), use Bayes theorem:

$$P(d, \{\Omega\}) = P(d|\{\Omega\})P(\{\Omega\}) = P(\{\Omega\}|d)P(d)$$

$$\Longrightarrow P(\{\Omega\}|d) = \frac{P(d|\{\Omega\})P(\{\Omega\})}{P(d)}$$

Prior

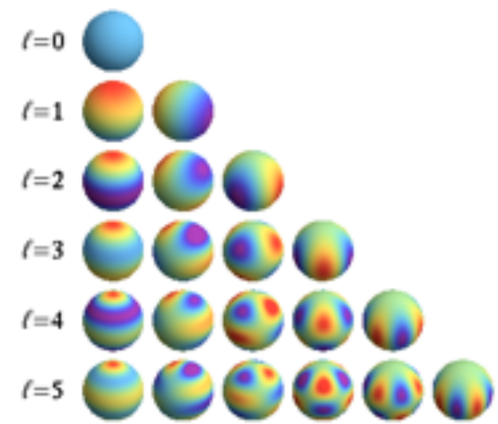
Posterior

Likelihood

So, what is the likelihood for CMB data?

Spherical Harmonics

- We can decompose the temperature maps in Spherical Harmonics (see E. Komatsu lecture #2)
  - ✦ eigenfunctions of the angular part of the Laplace operatorin spherical coordinates. $\nabla^2 Y_\ell^m = -[l(l+1)]Y_\ell^m$



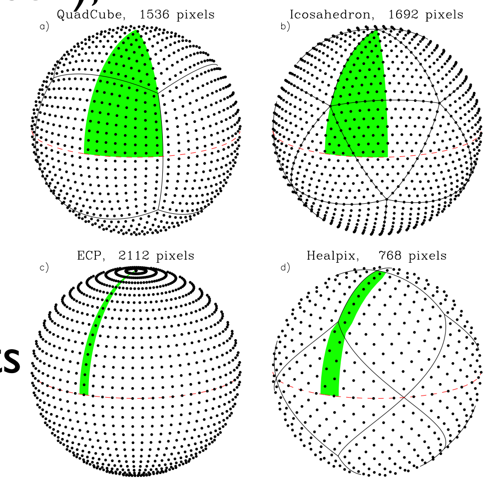
  - ✦ They form a ortho-normal and complete basis.

$$\int d\hat{\mathbf{n}} Y_\ell^{m*}(\hat{\mathbf{n}}) Y_{\ell'}^{m'}(\hat{\mathbf{n}}) = \delta_{\ell\ell'}\delta_{mm'}$$

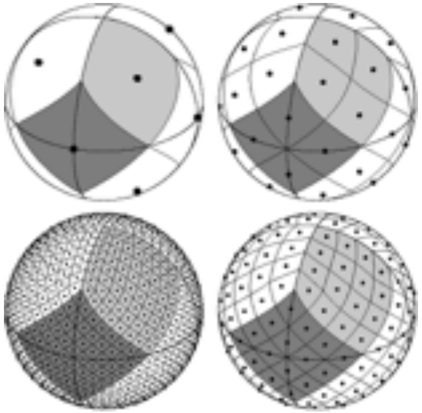
$$\sum_{\ell m} Y_\ell^{m*}(\hat{\mathbf{n}}) Y_\ell^{m}(\hat{\mathbf{n}}') = \delta(\phi - \phi')\delta(\cos\theta - \cos\theta')$$

  - ✦ Complex representation: $Y_\ell^{m*} = (-1)^m Y_\ell^{-m}$
  - ✦ Parameterized by the multipole (degree?) ℓ and (order?) m (cf, eigenvalues of kinetic moments L^2 and L_z in Quantum Mechanics)
    - $\ell \sim \pi / \theta$, with the θ angular separation in the sky,
    - For each ℓ, $-\ell \leq m \leq \ell$. There are $2\ell+1$ m-modes for each ℓ,
    - The projection on the m-modes depends on the reference system.
- For polarization (spin ±2 quantity), use spin-weighted SH (= second derivatives of scalar SH, see E. Komatsu lecture #5)
  - ✦ Zaldarriaga & Seljak (1997), Kamionkowski et al (1997)

Sphere Pixelisation



- To allow numerical treatment, data have to be discretised (pixelised), with $N_{pix} \sim 10^6 - 10^8$
- For each pixel q, one computes $Y_{\ell m}(q) = N_{\ell m} P_{\ell m}(\theta_q) \exp(i\, m\, \varphi_q)$ where Legendre Polynomials $P_{\ell m}(\theta_q)$ require a costly recursion
- Iso-latitude layout of map pixels allows a faster calculation of $Y_{\ell m}(q)$ by a factor $\sqrt{N_{pix}}$ compared to more traditional layouts
    - ECP (Mucaccia et al, 1997), HEALPix (Gorski et al, 2005), Igloo (Crittenden et al, 1998), GLESP (Doroskevich et al, 2003)



- Extra requirements for astrophysics
    - Equal pixel area: easier pixel value $\longleftrightarrow$ flux density
    - Hierarchical: easier change of resolution
- HEALPix (Hierarchical, Equal Area, iso-Latitude Pixelisation of the sphere)
    - used in WMAP, Planck, GAIA, Euclid, …
    - $N_{pix} = 12\, N_{side}^2$, $\quad \ell_{max} = 2 - 3\, N_{side}$

CMB signal

- CMB anisotropies are expected in the simplest inflation models (and observed by Planck, see S. Matarese lectures) to be distributed as a Gaussian random field.

- We cannot theoretically predict the value of the temperature in the pixels, but only predict their statistical properties.

- A Gaussian distribution is fully characterized by a mean (m) and a variance (σ^2).
  - ✦ All higher odd moments are 0,
  - ✦ even moments can be written in terms of the variance (Wick's theorem)

Decomposition in SH

- Decompose the fractional temperature variation in spherical harmonics

$$\Delta T(\mathbf{n}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\mathbf{n})$$

- Applying the orthogonality of spherical harmonics:

$$a_{\ell m} = \int_{4\pi} d\mathbf{n}\ \Delta T(\mathbf{n})\ Y^*_{\ell m}(\mathbf{n})$$

$$\simeq \sum_{p=1}^{N_{\rm pix}} \Omega_{\rm pix}\ \Delta T(p)\ Y^*_{\ell m}(p)$$

- As far as we know, ΔT(**n**) is a Gaussian random field, with isotropic statistical properties (ie $\langle \Delta T(\mathbf{n})^2 \rangle = \langle \Delta T^2 \rangle$).
  Then, $a_{\ell m}$ are statistically independent and randomly distributed,
  each described by a Gaussian distribution.

Angular power spectrum C_ℓ

- To characterise the statistical properties of a Gaussian random field, we can calculate the mean and the variance of the field. For the CMB, the mean of the anisotropies is zero (by definition).
  The variance can be calculated either as the 2-point correlation function in real space, or equivalently, as the **angular power spectrum in harmonic space**, which describes the variance of the anisotropies as a function of scale

- $\langle a_{\ell m} \rangle = 0 \qquad \langle a_{\ell m}\, a_{\ell' m'} \rangle = \delta_{\ell\ell'}\, \delta_{mm'} C_\ell$
  $\langle\rangle$ are ensemble averages over many realisations of the sky.
  But, we have only one sky available !
  (CMB C_ℓ predicted by theory, and computed by Boltzmann code (eg, CAMB, CLASS))

- Because of statistical isotropy, $a_{\ell m}$ with same ℓ and different m are extracted from Gaussian distribution with the same variance C_ℓ
  so an estimator of C_ℓ is

$$\widehat{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{m=\ell} a_{\ell m} a^*_{\ell m}$$

  and

$$\langle \widehat{C_\ell} \rangle = C_\ell$$

Cosmic Variance

- So we have an estimator $\widehat{C_\ell}$ whose expectation value is $\langle\widehat{C_\ell}\rangle = C_\ell$
- Since we only have $2\ell+1$ samples for each ℓ, there is an intrinsic uncertainty!

$$\frac{\sigma^2_{C_\ell}}{C_\ell^2} = \frac{\langle(\hat{C}_\ell - C_\ell)(\hat{C}_\ell - C_\ell)\rangle}{C_\ell^2} = \frac{\langle\hat{C}_\ell\hat{C}_\ell\rangle - C_\ell^2}{C_\ell^2}$$

$$= \frac{1}{(2\ell+1)^2 C_\ell^2}\left\langle\sum_{mm'} a_{\ell m}^* \, a_{\ell m} \; a_{\ell m'}^* \, a_{\ell m'}\right\rangle - 1$$

$$= \frac{1}{(2\ell+1)^2}\sum_{mm'}(\delta_{mm'} + \delta_{m-m'}) = \frac{2}{2\ell+1}$$

For a gaussian field, Wick's theorem says that any N-point (N even) statistics can be written as a function of the 2-point correlation

$$\boxed{\sigma^2_{C_\ell} = \frac{2}{2\ell+1} C_\ell^2}$$

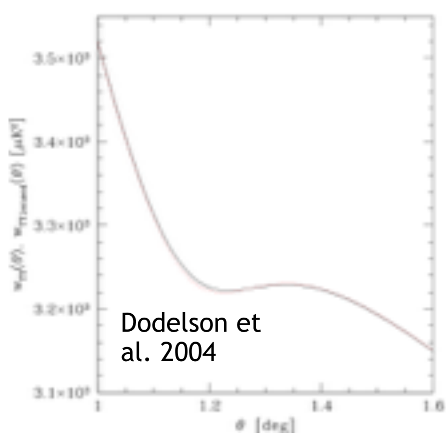
C_ℓ's and 2-point correlation function

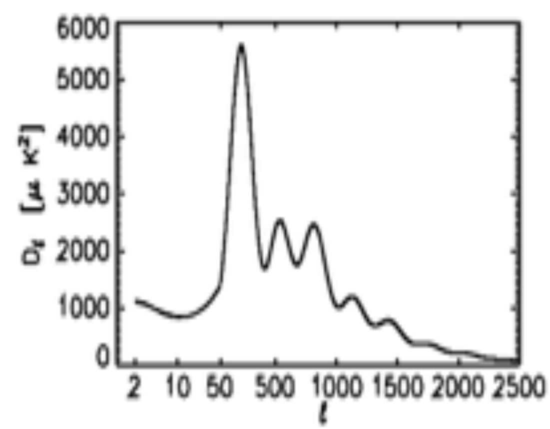
- We can relate the angular power spectrum to the 2-point correlation function in real space using the Legendre polynomials and the addition theorem:

$$\sum_m Y^*_{\ell m}(\mathbf{n}_i) Y_{\ell m}(\mathbf{n}_j) = \frac{2\ell+1}{4\pi} P_\ell(\hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_j) \qquad \langle \Theta_i \Theta_j \rangle = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell P_\ell(\hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_j)$$

- Because of isotropy, the two-point correlation function depends only on the angular separation in the sky θ, not on the orientation of the separation.

- $C(\theta)$ is much less showed than C_ℓ because
  - it lacks features (tale-telling 'acoustic peaks'),
  - has correlated errors
  - but can be convenient for some calculations:
    a product of 2 $C(\theta)$ can replace the 'convolution' of 2 C_ℓ (e.g. Chon et al, 2004)





Noise and cross-power-spectrum

- In the presence of (instrumental) noise or contaminant n : $d = s + n$, $d_{\ell m} = s_{\ell m} + n_{\ell m}$
  The <u>auto power spectrum</u> is
  $\langle \hat{C}_\ell \rangle = S_\ell + N_\ell$ with variance $\langle \Delta\hat{C}_\ell^2 \rangle = 2(S_\ell + N_\ell)^2 / (2\ell+1)$
  → Noise bias and increased variance

- For 2 data-sets with the same signal but un-correlated noises
  $d_1 = s + n_1$, $d_2 = s + n_2$, the <u>cross power spectrum</u> $\widehat{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{m=\ell} d_{1,\ell m}\, d^*_{2,\ell m}$
  has $\langle \hat{C}_\ell \rangle = S_\ell$ with variance $\langle \Delta\hat{C}_\ell^2 \rangle = (S_\ell^2 + (S_\ell + N_{1,\ell})(S_\ell + N_{2,\ell})) / (2\ell+1)$
  → No noise bias, but even larger variance

  ✦ 2 data-sets with un-correlated noises and $N_1 = N_2 = N$
  (biased) <u>auto spectrum</u> of $d = (d_1+d_2)/2$ has variance $(2S_\ell^2 + N_\ell^2/2 + 2S_\ell N_\ell) / (2\ell+1)$,

  (unbiased) <u>cross spectrum</u> of d_1, d_2 has variance $(2S_\ell^2 + N_\ell^2 + 2S_\ell N_\ell) / (2\ell+1)$.
  In spite of larger variance when $N_\ell \geqslant S_\ell$, cross-spectrum is often preferable because it is un- (or less) biased, and does not mixes up systematics

- N_d data-sets:
  - a single <u>auto-spectrum</u> of bias N_ℓ / N_d and variance $2 N_\ell^2 / N_d^2$
  - vs $N_d (N_d-1)/2$ un-biased <u>cross-power spectra</u>, each of variance N_ℓ^2
    average of cross-spectra has ~ same variance as auto-spectrum

Likelihood of Gaussian CMB maps

- CMB maps **m**=(I,Q,U) have Gaussian fluctuations with

$$\mathcal{L}(C_\ell) = \mathcal{P}(\boldsymbol{m}|C_\ell) = \frac{1}{|2\pi \mathrm{M}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{m}^{\mathrm{T}}\,\mathrm{M}^{-1}\boldsymbol{m}\right)$$

- **m** is the *data vector* of length N_{pix} containing the pixels of the map

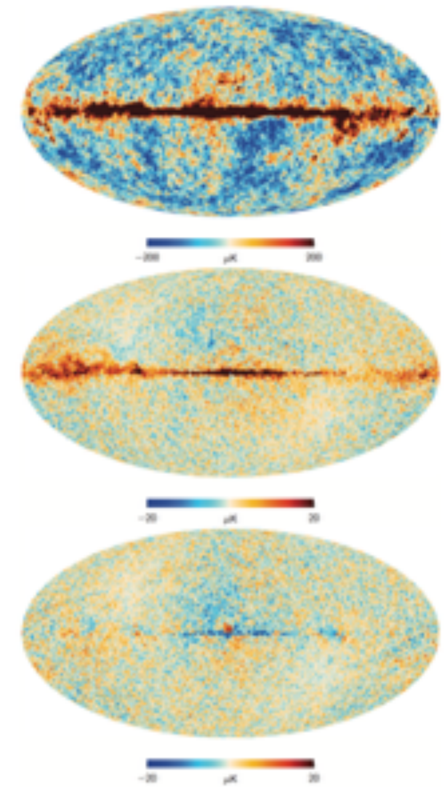
- **M** is a N_{pix}*N_{pix} *covariance matrix*.
  M_{ij} tells us how much pixels *i* and *j* are correlated:
  $\mathbf{M}_{ij} = \langle \mathbf{m}_i . \mathbf{m}_j^{T} \rangle = \mathbf{S}(\theta_{ij}\,; C_\ell(\{\Omega\})) + \mathbf{N}(\gamma_i, \gamma_j)$

  - ✦ $\mathbf{S}(\theta_{ij}\,; C_\ell(\{\Omega\}))$ is the signal covariance related to the (theoretical) power spectrum through (for T)

$$S(\theta_{ij}; C_\ell(\{\Omega\})) = \langle \Delta T_i \Delta T_j \rangle$$
$$= \frac{1}{4\pi}\sum_\ell (2\ell+1) C_\ell(\{\Omega\}) P_\ell(\theta_{ij})$$

    (assuming isotropy of the signal)

  - ✦ $\mathbf{N}(\gamma_i, \gamma_j)$ is the pixel space noise covariance matrix (diagonal only for white noise)



Pixel-based Likelihood

$$\mathcal{L}(C_\ell) = \mathcal{P}(\boldsymbol{m}|C_\ell) = \frac{1}{|2\pi \mathrm{M}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{m}^{\mathrm{T}}\,\mathrm{M}^{-1}\,\boldsymbol{m}\right)$$

- Working in pixel-space has advantages.
  E.g. dealing with masks is easy.
- BUT! In order to calculate the likelihood for each set of parameters $\{\Omega\}$ we need to
  - ✦ calculate the *determinant* of **M** and inverse **M** (or rather solve **M** . **x** = **m** when applicable) → goes like $\mathcal{O}(N_{pix}^3)$,
  - ✦ **M** = **S** + **N**, for masked sky and/or non-white noise **S** and **N** not diagonal in neither pixel- nor multipole- spaces.
- It can be used at low resolutions, with variants eg, NRML(Bond et al, 1998), QML(Tegmark, 1998, Gruppuso et al, 2009), TEASING (Benabed et al, 2009),
  but Planck HFI maps have $O(10^6)$ pixels! Need other solutions!

Multipole-based Likelihood

- One writes the Likelihood in terms of $\hat{C}_\ell$ (= square of Gaussians) instead: since for the full sky $\mathcal{L}(a_{\ell m}|C_\ell) = (2\pi C_\ell)^{-1/2} \exp(-a_{\ell m}\, a_{\ell m}^* / 2C_\ell)$
  $\Rightarrow \mathcal{L}(\hat{C}_\ell|C_\ell) \propto 1/\hat{C}_\ell \; [\hat{C}_\ell/C_\ell \exp(-\hat{C}_\ell/C_\ell)]^{(2\ell+1)/2}$

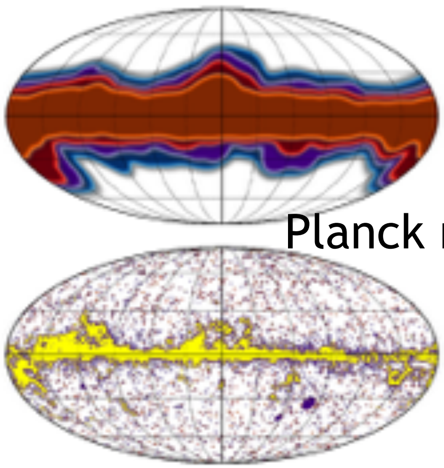
  each multipole treated separately (on the full sky !).

  - ✦ $(2\ell+1)\hat{C}_\ell/C_\ell$ has a χ^2 distribution with $2\ell+1$ degrees of freedom (mean=$2\ell+1$, variance=$2(2\ell+1)$)
    Wishart distribution when dealing with $(C_\ell^{TT}, C_\ell^{TE}, C_\ell^{EE})$
  - ✦ $\hat{C}_\ell/C_\ell$ has Γ distribution with $2\ell+1$ dof (mean=1, variance= $2/(2\ell+1)$)
  - ✦ $C_\ell/\hat{C}_\ell$ has inverse Γ distribution with $2\ell+1$ dof
    inverse Wishart distribution when dealing with $(C_\ell^{TT}, C_\ell^{TE}, C_\ell^{EE})$

- Note that when $2\ell+1 >> 1$, all of these distributions have a Gaussian like shape around their peak, see e.g. Percival & Brown (2006)

Life is more complicated…

- Many things complicate the calculation of the likelihood, starting with the necessity of masking the most foreground-contamined pixels
- When masking the sky: ΔT(**n**) → ΔT(**n**) W(**n**) (EH et al, 2002)



Planck masks

$$\begin{aligned}\tilde{a}_{\ell m} &= \int d\mathbf{n} \Delta T(\mathbf{n}) W(\mathbf{n}) Y^*_{\ell m}(\mathbf{n}), \\ &= \sum_{\ell' m'} a_{\ell' m'} \int d\mathbf{n} Y_{\ell' m'}(\mathbf{n}) W(\mathbf{n}) Y^*_{\ell m}(\mathbf{n}) \\ &= \sum_{\ell' m'} a_{\ell' m'} K_{\ell m l' m'}[W],\end{aligned}$$

$\tilde{a}_{\ell m}$ are still Gaussian, but not independent since they all depend on the sum of $a_{\ell' m'}$. $\tilde{a}_{\ell 1 m 1}$ and $\tilde{a}_{\ell 2 m 2}$ are correlated.

The smaller the observed patch, the poorer the spectral resolution!

$\tilde{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |\tilde{a}_{\ell m}|^2$ correlated for different ℓs, and their variances get mixed up

=> all ℓ now coupled in Likelihood (no more χ^2 nor Wishart)

Unbiased pseudo-C_ℓ estimator

- Since, the "pseudo C_ℓ" of the masked sky is related (in average) to the true C_ℓ via

$$\langle \widetilde{C}_\ell \rangle = \sum_{\ell'} M_{\ell\ell'}[W] \langle C_{\ell'} \rangle \qquad \sim f_{sky} \langle C_\ell \rangle$$

one defines a new estimator

$$\widehat{C}_\ell = \sum_{\ell'} M^{-1}_{\ell\ell'}[W] \widetilde{C}_{\ell'}$$

such that it is unbiased: $\langle \hat{C}_\ell \rangle = \langle C_\ell \rangle$, with

$$M_{\ell\ell'}[W] = \frac{2\ell'+1}{4\pi} \sum_{\ell''} (2\ell''+1) \mathcal{W}_{\ell''} \begin{pmatrix} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{pmatrix}^2$$

Mask APS

Wigner 3J symbol ~ Clebsch-Gordon coeff.
= coupling of 3 $Y_{\ell m}$

- implemented in eg, MASTER (EH et al, 2002), PolSpice (Chon et al, 2004), Xspect (Tristram et al, 2005), Romaster (Polenta et al, 2005), XFaster (Rocha et al, 2009).

- Bin the C_ℓ to lump $\Delta\ell$ multipoles that are tightly correlated because of the mask.

- The covariance $\langle \Delta\hat{C}_\ell \Delta\hat{C}_{\ell'} \rangle$ can also be estimated (Efstathiou, 2004) if the mask is smooth enough (narrow enough ℓ band-width).
  → extra care required for Point Sources mask

Let's approximate: Gaussian!

- For large degrees of freedom $\nu=(2\ell+1)\Delta\ell f_{sky}$, the distribution of the $\hat{C}_\ell \rightarrow$ Gaussian distribution (central limit theorem).

$$-\ln \mathcal{L}(\widehat{\mathbf{C}}|\mathbf{C}(\{\Omega\})) = \frac{1}{2}\left[\widehat{\mathbf{C}} - \mathbf{C}(\{\Omega\})\right]^T \mathbf{M}^{-1}\left[\widehat{\mathbf{C}} - \mathbf{C}(\{\Omega\})\right] + \text{const.}$$

data (debiased pseudo C_ℓ vector)

Model (that depends on the parameters we want to determine)

$\langle \Delta\hat{C}_\ell \, \Delta\hat{C}_{\ell'} \rangle$ covariance matrix (can be estimated with a fixed fiducial set of parameters, often validated with Monte-Carlo simulations)

But! It works only at high-ℓ (large dof)

See e.g. Hamimeche & Lewis 2008, 2009 for discussion and improvement.
Validated by comparison with more sophisticated approximation, including on actual Planck data (Planck 2013-XV, 2014)

Effect of beam

- All experiments have a finite angular response:
  the optical beam, possibly convolved with the instrumental time response in case of scanning

- For a beam of power spectrum B_ℓ

  $$\hat{C}_\ell \rightarrow \hat{C}_\ell \, B_\ell$$

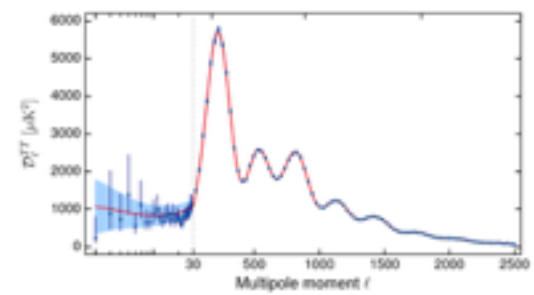
  if, and only if

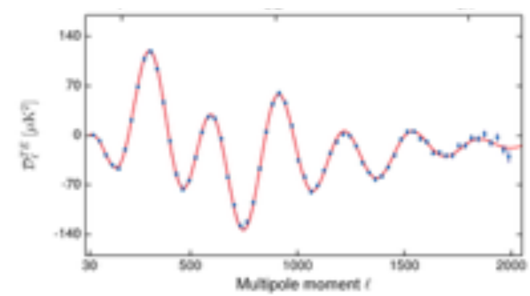
  - *either* the beam is circular, whatever is the scanning,
  - *or* the beam is arbitrary, but remains parallel to itself (raster scan)

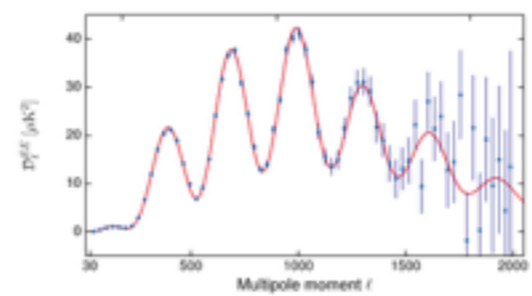
- In all other cases, and especially for polarisation,
  things get more complicated.
  See 2nd lecture

Planck likelihood: A hybrid approach

- Low-ℓ ($\ell < 30$):
  - **TT**: Pixel-based approach based on $N_{side}=16$ Commander component separated map, 92% sky, all Planck frequencies used+WMAP+Haslam
  - **TE and EE**: Pixel based approach based on Planck LFI 70GHz map, 46% of the sky. 30 GHz and 353GHz used for foreground cleaning.







- High-ℓ ($30 < \ell < 2500$):
  - **TT**: Gaussian likelihood based on HFI 100, 143, 217GHz at (70, 60, 50% sky)
  - **TE,EE**: Gaussian likelihood, HFI 100, 143, 217GHz at (70, 50, 40% sky).

Planck 2015 results. XI.

Recap

We have seen

- Why we need the C_ℓ and their likelihood.
- How to compute them
  - ✦ when the foregrounds can be removed by smooth (apodized) masks,
  - ✦ when the instrument is perfectly known and well-behaved:
    - friendly noises, non-correlated between detectors,
    - instantaneous measurements,
    - well measured circular beams,
    - constant instrumental responses.
- All residual systematics are well below the instrumental noise.

Recap

We have seen

- Why we need the C_ℓ and their likelihood.
- How to compute them for Planck
    - ~~when the foregrounds can be~~ … ~~smooth (apodized) masks,~~
    - ~~when the instrument is p~~ … ~~n and well-behaved:~~
        - ~~friendly noises, n~~ … ~~between detectors,~~
        - ~~instantanee~~ … ~~ents,~~
        - ~~well m~~ … ~~ar beams,~~
        - ~~co~~ … ~~ental responses.~~
- ~~All re~~ … ~~tics are well below the instrumental noise.~~

Tomorrow

Bibliography

Benabed et al, 2009 (Teasing) 2009MNRAS.400..219B

Bond et al, 1998 (NRML) 1998PhRvD..57.2117B

Chon et al, 2004 (PolSpice) 2004MNRAS.350..914C

Crittenden & Turok 1998 (Igloo) 1998astro.ph..6374C

Doroshkevich et al 2003 (GLESP) 2003astro.ph..5537D

Efstathiou 2004 (Hybrid) 2004MNRAS.349..603E

Gorski et al, 2005 (Healpix) 2005ApJ...622..759G

Gruppuso et al, 2009 (BolPol) 2009MNRAS.400..463G

Hamimeche & Lewis, 2008 2008PhRvD..77j3013H

Hamimeche & Lewis, 2009 2009PhRvD..79h3012H

Hivon et al, 2002 (Master) 2002ApJ...567....2H

Kamionkowski et al, 1997 1997PhRvD..55.7368K

Mucaccia et al,1997 (ECP) 1997ApJ…488L..63M

Percival & Brown, 2006 2006MNRAS.372.1104P

Planck 2013-XV, 2014 2014A&A…571A..15P

Planck 2015-XI, 2016 2016A&A…594A..11P

Polenta et al, 2005 2005JCAP...11..001P

Rocha et al, 2009 (Faster) 2009arXiv0912.4059R

Tegmark, 2009 (QML) 2009PhRvD..55…10T

Tristram et al, 2005 (XSpect) 2005MNRAS.358..833T

Zaldarriaga & Seljak, 1997 1997PhRvD..55.1830Z

Back to page 1