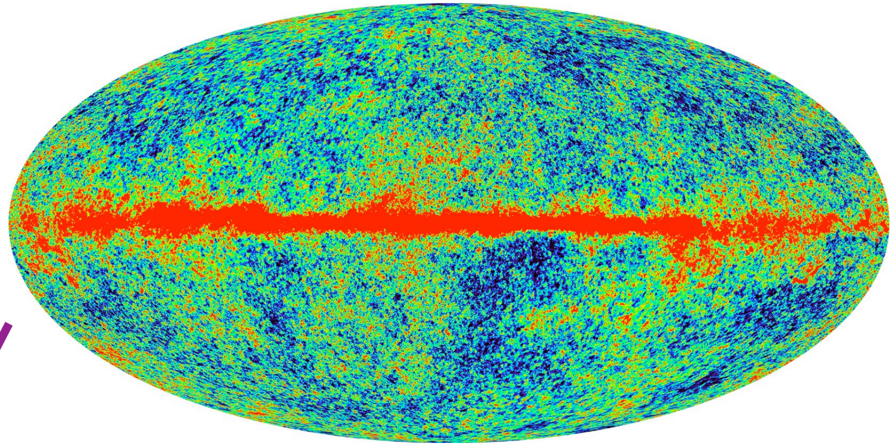
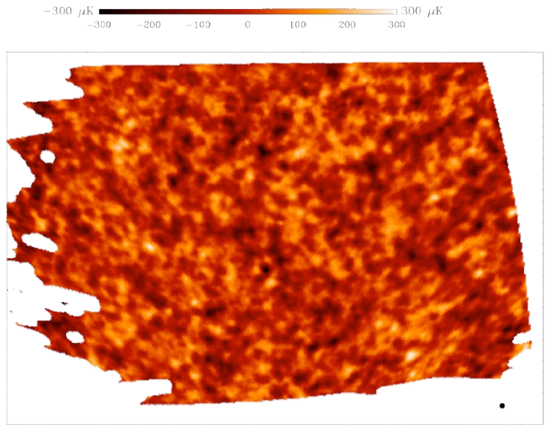
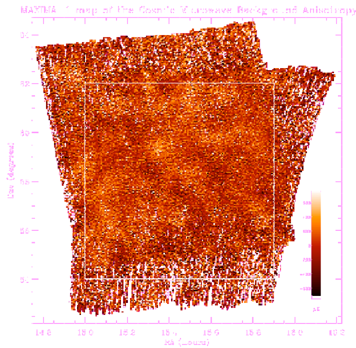
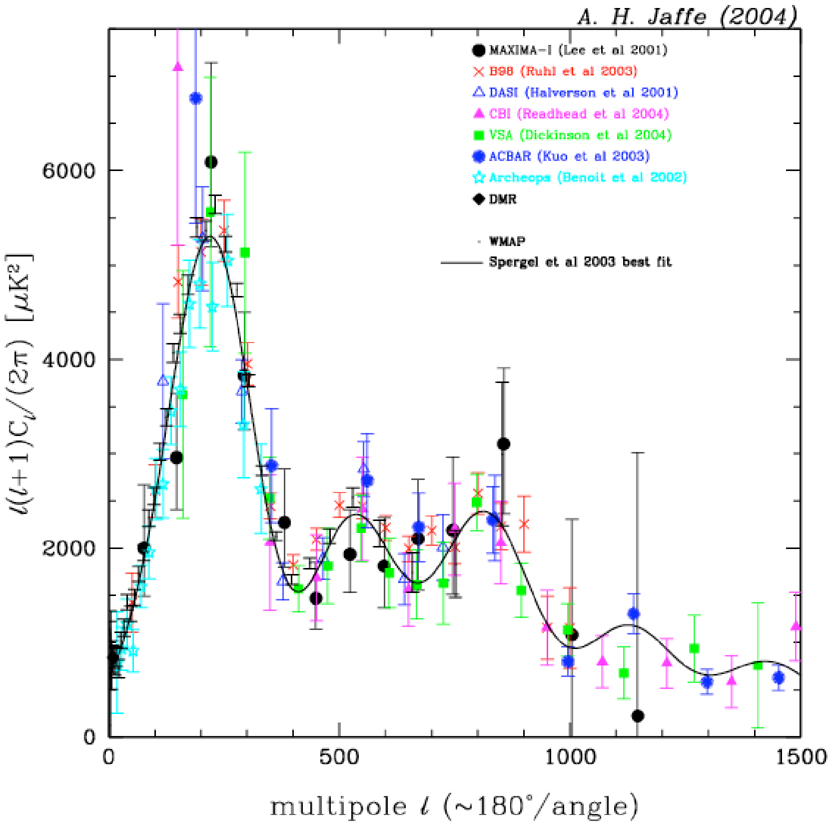






Estimating the power spectrum of the CMB



Andrew Jaffe

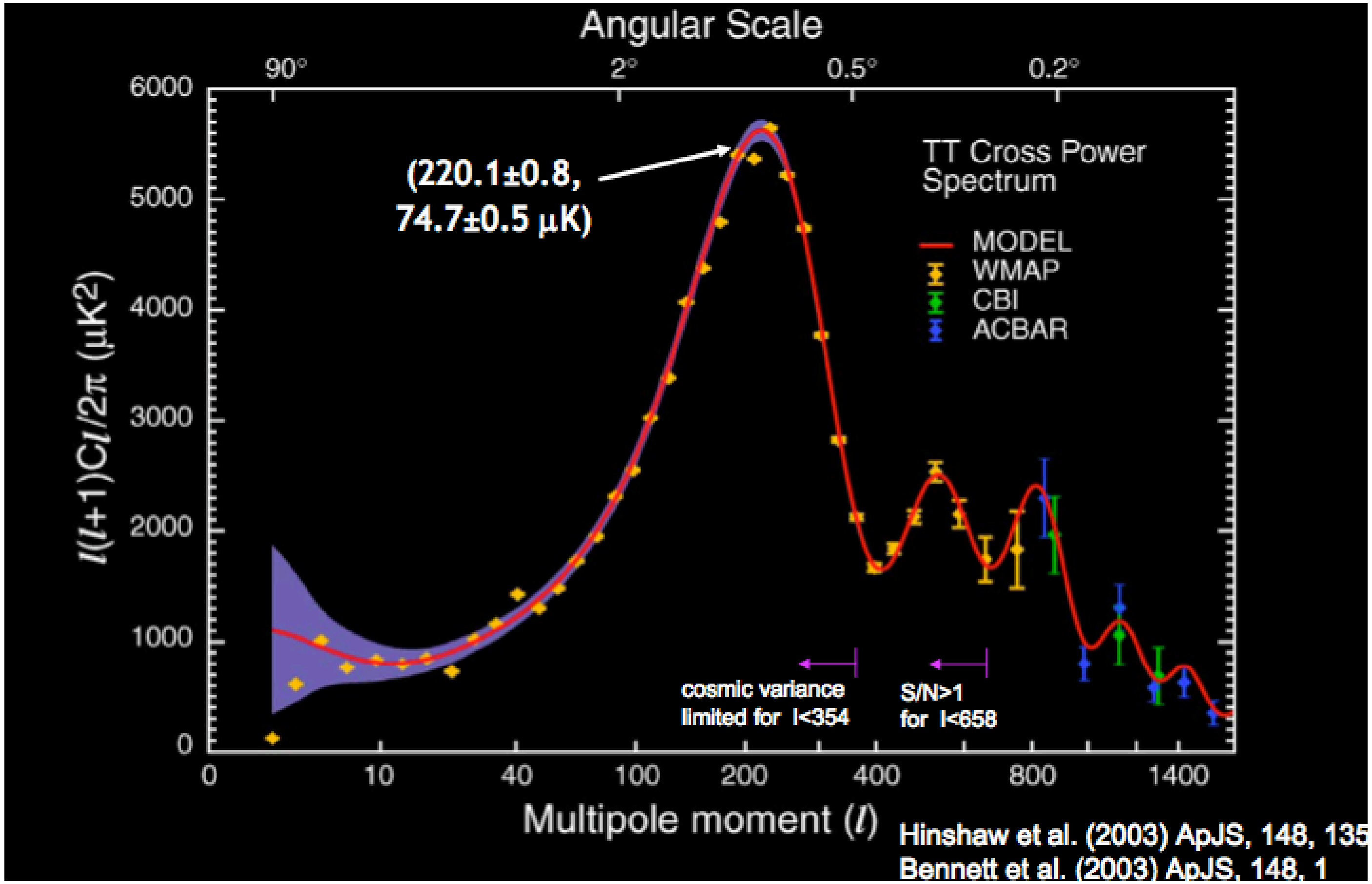
Imperial College

XX IAP:

CMB Physics & Observation

Estimating the Power Spectrum of the CMB

- Philosophy — What is C_l?
  - Bayesian/Frequentist
- History
- Practice
  - — Methods
- Future?



What is C_l?

- Sky average $\frac{1}{2\ell+1}\sum_{\ell}\left|a_{\ell m}\right|^2$
- "Ergodic" (cosmic) average $\left\langle \left|a_{\ell m}\right|^2 \right\rangle$
- Variance of a Gaussian distribution
- Variance of some other distribution

CMB Data

- data = signal + noise
- $d_p = s_p + n_p$ (*p*=pixel number), correlations:

$$\langle s_p s_{p'} \rangle = S_{pp'} = \sum_\ell \frac{2\ell+1}{4\pi} B_\ell^2 C_\ell$$

(scanning temperature experiments)

$$\langle n_p n_{p'} \rangle = N_{pp'}$$

- Polarization: $S_{pp'}$ is linear combination of $C_l^{XX'}$
- Task: measure C_l (or bandpowers — Bond) **and preserve *all* sky information for parameter estimation**

Probability distributions

- Likelihood function $\mathrm{P}(d_p \,|\, C_l\, N_{pp'}, I)$
  - probability density of data data given signal and noise variances (& information *I*)
- Frequentist:
  - underlying physical mechanism responsible for "long-run" frequency distribution of data
- Bayesian:
  - encodes information I (which may be that same physical mechanism)
- e.g., Gaussian Signal + Noise:

$$P(d \,|\, SNI) = \frac{1}{\left|2\pi(S+N)\right|^{1/2}} \exp\left[-\tfrac{1}{2} d^T (S+N)^{-1} d\right]$$

Frequentist methods

- Devise an "estimator" $E_l[d]$ such that $E_l[d]$~(input C_l)
- e.g., unbiased:

$$\langle E_\ell \rangle \equiv \int d^n d_p \, E_\ell[d] P(d_p | C_\ell N_{pp'} I) = C_\ell$$

- depends on likelihood as function of varying data for fixed (fiducial) C_l
- in practice, "quadratic estimators"
  - $E_l[d] = Q_l[d] = d^T Q_l d - b_l$
  - $\langle d^T Q_l d \rangle = \mathrm{Tr}[(S+N)Q_l] = \sum C_l M_{ll'} F_l B_l^2 + b_l$ in simple Gaussian case

Frequentist Methods (II)

- Quadratic form:
  - $E_l[d] = Q_l[d] = d^T Q_l d - b_l$
  - $\langle d^T Q_l d \rangle = \mathrm{Tr}[(S+N)Q_l] = \sum C_l M_{ll'} F_l B_l^2 + b_l$ in simple Gaussian case
- estimate is $E_l[d] \pm \sigma_l[d]$
  - with σ_l from diagonal elements of $V_{ll'}[d] = \langle E_l E_{l'} \rangle - \langle E_l \rangle \langle E_{l'} \rangle$
- How do we use $E_l \pm \sigma_l$ for parameter estimation?
  - full frequentist parameter estimation hard/ill-defined (Abroe et al, Schaefer & Stark)

Bayesian methods

- Characterize likelihood function $\mathrm{P}(d_p \mid C_l\, N_{pp}, I)$ as function of C_l for fixed (observed) data.
- depends on *use* of estimate:
  - for actual "C_l estimate":
    - assign prior $\mathrm{P}(C_l|I)$, use Bayes's theorem:

$$P(C|dNI) = \frac{P(C \mid I)P(d \mid CNI)}{P(d \mid NI)}$$

    - report, e.g., mean, variance
  - for further parameter estimation, need full shape of $L(C_l)$= $\mathrm{P}(d_p|C_l NI)$ for use in Bayes's theorem estimation of parameters
    - C_l prior doesn't enter — "hierarchical model"

Probabilities and Entropy

- Bayesian: probabilities are primarily about information, and only secondarily about frequency
  - How do we assign a distribution based on our information?
- Entropy — maximize subject to constraints
  - Gaussian has maximum entropy for given covariance
  - Uncorrelated Gaussian has maximum entropy for given variances (diagonal elements, σ_i^2)
    - e.g., σ_i^2 is marginalized variance irresp. of off-diag terms
    - Gaussianity is conservative choice!

Bayesian methods: hierarchical models

- Timestream (d_t)
  - $\Rightarrow$ Map ($T_p \sim d_p$)
  - $\Rightarrow$ Spectrum ($C_l \sim d_l$)
  - $\Rightarrow$ cosmology
- without loss of information?

- $P(\text{Cosmology}|d_t N_{tt'}) = P(\text{Cosmology}|d_p, N_{pp'})$
  $\approx P(\text{Cosmology}|d_l N_{ll'}, x_l)$
- assume that we can calculate $P(\text{Cosmology}|d_l N_{ll'}, x_l)$ even from non-Bayes estimators

nb. Wiener filter from $P(d_p|C_l)$

- e.g., post hoc polzn separation, prediction

Bayesian/Frequentist Correspondance

- Why do both methods seem to work?
- frequentist mean ~ likelihood maximum
  frequentist variance ~ likelihood curvature
- Correspondance is *exact* for
  - linear gaussian models (mapmaking)
  - variance estimation with no correlations and "iid" noise — simple version of C_l problem
    - e.g., all sky, uniform noise
    - likelihood only function of d_{lm}^2
    - breaks down in realistic case of correlations, finite sky, varying noise
  - "asymptotic limit"
    - ~ high l iff noise correlations not "too strong"

Expected errors

- Knox 95, Hobson & Magueijo 96

$$\left(\delta C_\ell\right)^2 \cong \frac{2}{(2\ell+1) f_{\rm sky}} \left(C_\ell + N_\ell\right)^2 \qquad N_\ell \approx w^{-1} = \left(\theta_p \sigma_p\right)^{-2}$$

# of modes

Sample Variance

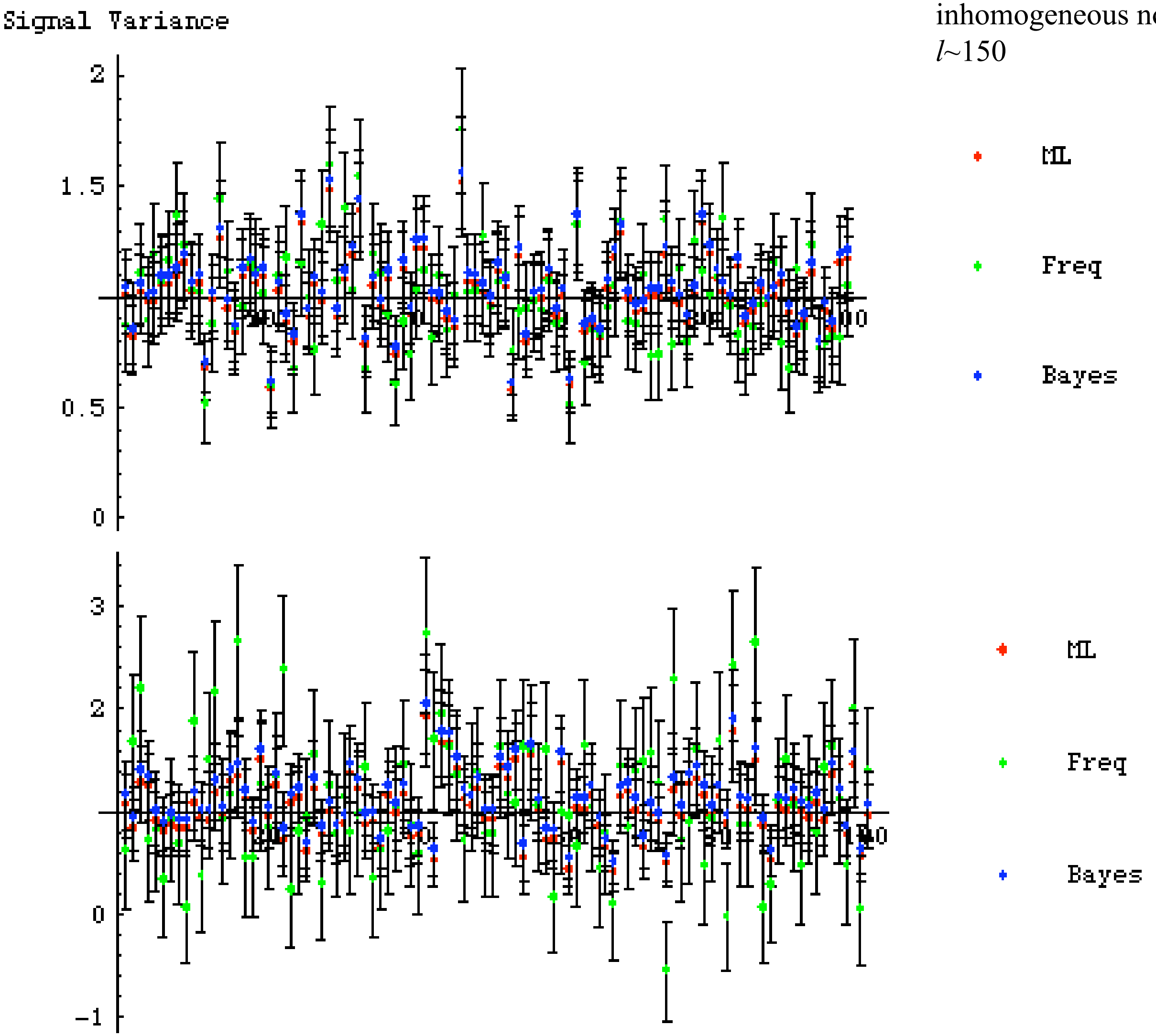
Noise variance

~(weight per sr)$^{-1}$

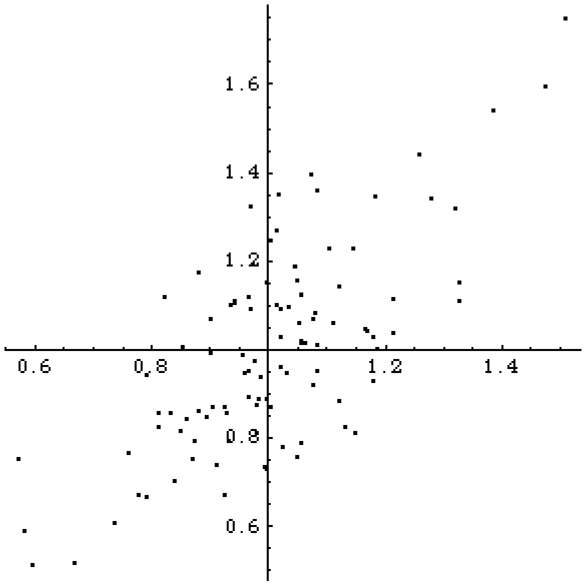
Case study

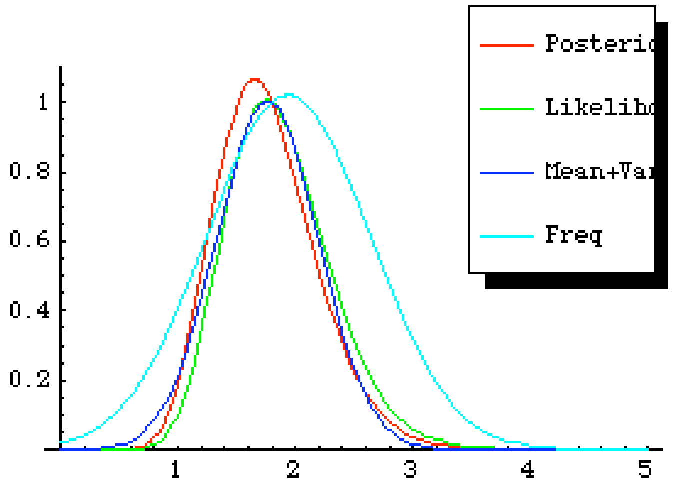
- Toy version of a single l $(m = -l, \ldots, +l)$
  - $d_m = a_m + n_m$ $\langle a_m a_{m'} \rangle = C\delta_{mm'}$, $\langle n_m n_{m'} \rangle = N_{mm'}$
- Naïve Quadratic estimator $Q = \sum_m |d_m|^2 - b$
- Likelihood Maximum, curvature
- Posterior mean, variance
  [with "Jefferys Prior" $P(C|I) \propto 1/C$]

toy model — inhomogeneous noise

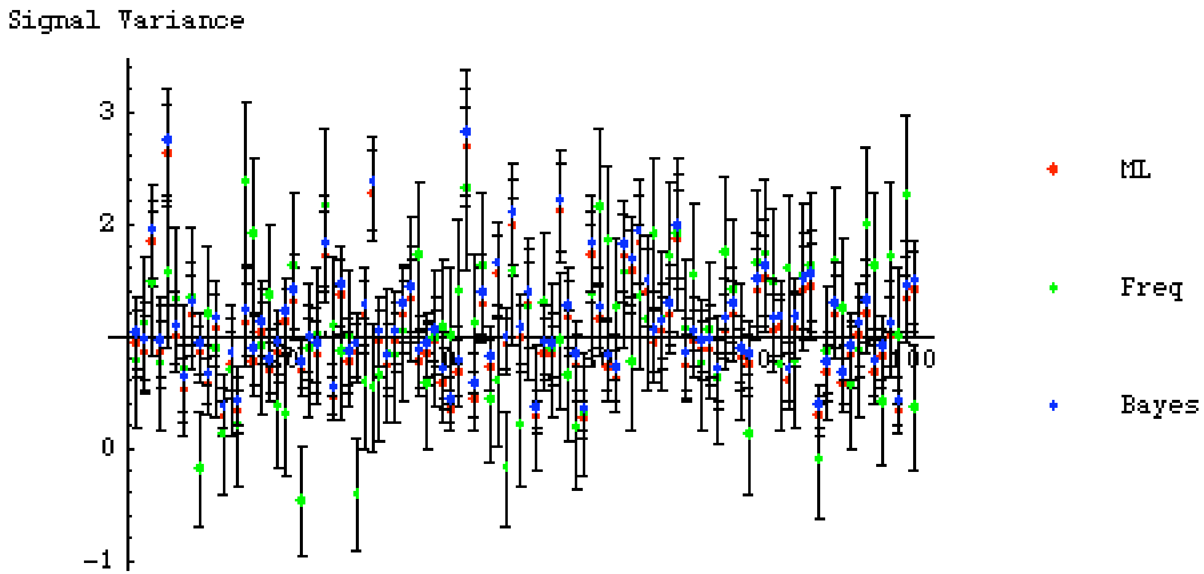


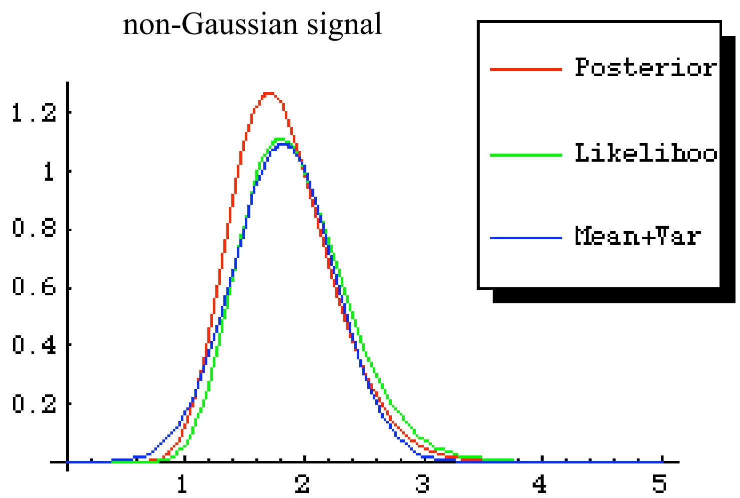
inhomogeneous noise
l~150



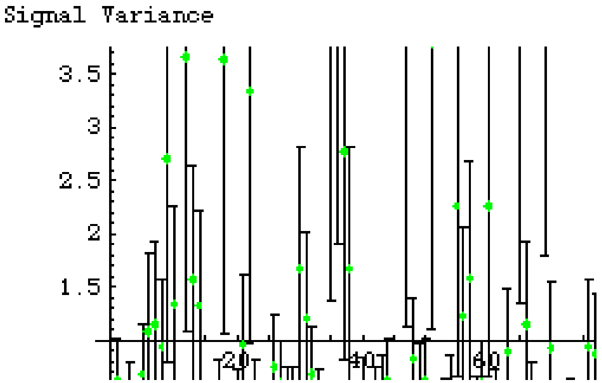


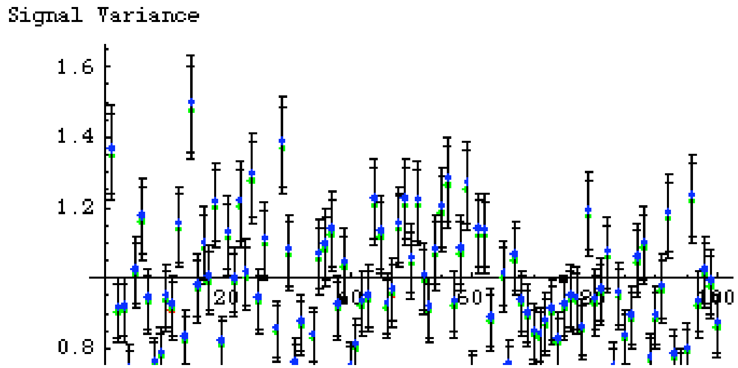
toy model — non-Gaussian signal

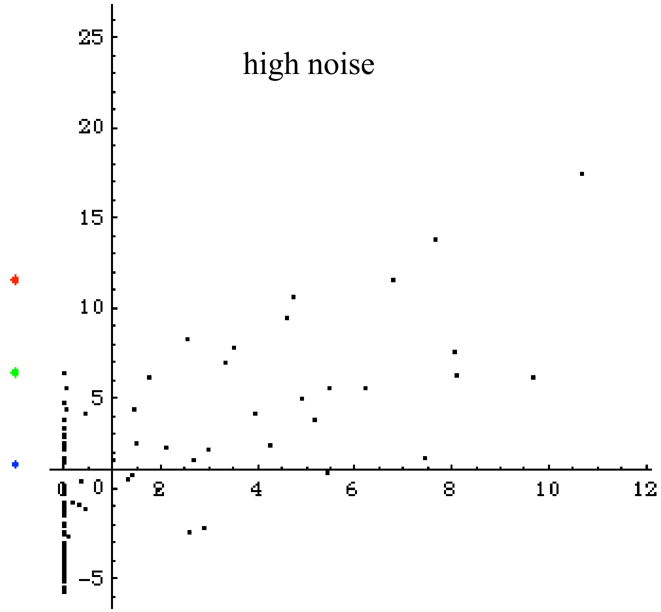




low noise

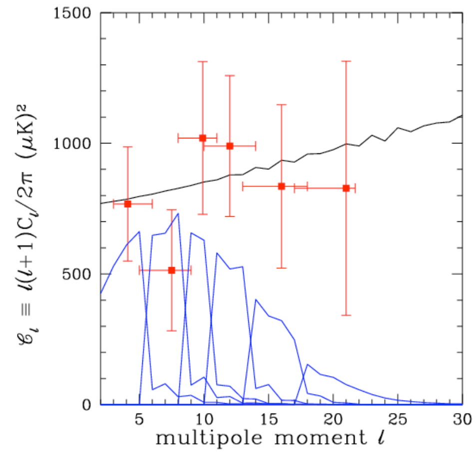






History

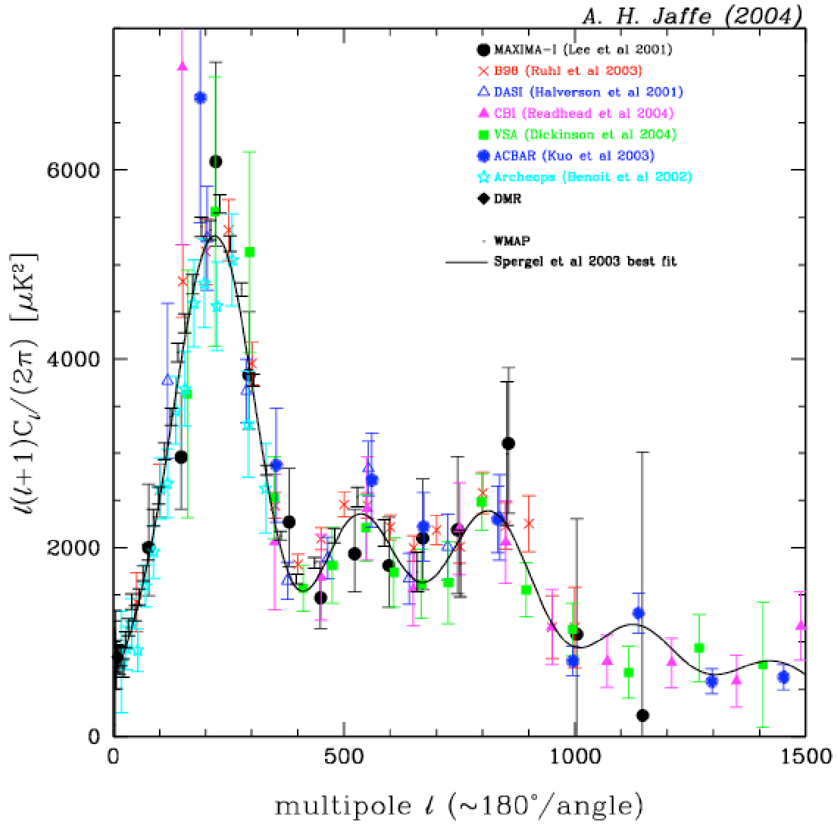
- Galaxy surveys — correlation functions [e.g., Peebles], P(k) [e.g., Feldman, Kaiser & Peacock]
- DMR
  - $C(\theta)$ estimation; Boughn-Cottingham; ($Q_{rms\text{-}PS}$,n)
  - Likelihoods
    - Seljak & Bertschinger
    - Tegmark & Bunn
    - Bond — forecasts, likelihoods and esp. "bandpowers"
    - Gorski



- CMB upper limits (GACF — Gaussian autocorrelation function); first post-DMR experiments
  - Bandpowers: e.g., Crittenden, Bond et al (SP); Netterfield (SK)
- param. forecasts — Jungman et al; Bond, Efstathiou, Tegmark

“Modern” methods

- “Optimal Quadratic” (Tegmark)
- Newton-Raphson Iteration to Likelihood Max
  = Iterated optimal quadratic [BJK 98]
  - MADCAP (Borrill &c)
  - Interferometers
    (e.g. VSA: Maisinger, Hobson, et al)
- OSH — Monte Carlo methods
- pseudo-C_l methods
  - MASTER
  - Gabor transforms
- SPICE
- WMAP



Bayesian methods: MADCAP/MADspec

- (quasi-)Newton-Raphson iteration to Likelihood maximum
- Algorithm driven by matrix manipulation (iterated quadratic):

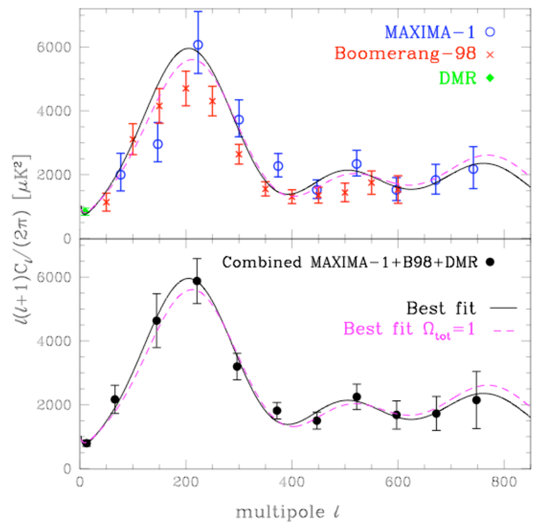
$$\delta C_\ell = \tfrac{1}{2} F^{-1}_{ll'} \,\mathrm{Tr}\left[\left(dd^T - C\right)\left(C^{-1}\frac{\partial C}{\partial C_\ell}C^{-1}\right)\right]$$

$$F_{ll'} = \tfrac{1}{2}\mathrm{Tr}\left[C^{-1}\frac{\partial C}{\partial C_\ell}C^{-1}\frac{\partial C}{\partial C_\ell}\right] \quad \text{Fisher matrix}$$

$$\mathrm{C} = \mathrm{S} + \mathrm{N}$$

- Fisher = approx. Likelihood curvature
- full polarization: signal matrix $S^{xx'}_{pp'}$
- Arbitrary (precomputed) noise spectrum
- Arbitrary linear filters
  - Stompor et al; Jaffe et al; Slosar et al

- $O(N^3)$ operations naïvely (matrix manipulations), speedup to ~$O(N^2)$ for spectrum estimates (potentially large prefactor)
  - Fully parallelized (MPI, SCALAPACK)
  - do calculations in the natural basis
  - no explicit need for full $N_{pp'}$ matrix in pixel basis (just noise spectrum or autocorrelation)
- e.g., MAXIMA, BOOMERANG



BJK 98

Borrill, Cantalupo, Stompor et al

Monte Carlo methods: MASTER, SPICE &c

- MASTER: quadratic pseudo-C_l estimate

$$d_{\ell m} = \sum_p d_p w_p \Omega_p Y_{\ell m}(\hat{x}_p)$$

$$\hat{C}_\ell = \frac{1}{2\ell+1} \sum_m \left| d_{\ell m} \right|^2$$

$$\left\langle d_{\ell m} d_{\ell' m'} \right\rangle = \sum_\ell C_\ell M_{\ell\ell'} F_\ell B_\ell^2 + N_\ell$$

takes advantage of fast SHT

(Hivon et al)
e.g., B98

- SPICE: transform of correlation function estimate

(Szapudi et al; Fosalba talk)

- Gabor transform: (apodized) quadratic
+ pseudo-ML for inverting Kernel

(Hansen et al)

- Issues: filters, weights, noise estimation/iteration, input maps — optimal or naïve?

Hybrid Methods: FASTER

- Key insight: MASTER covariance formalism allows calculation of diagonal part of pseudo-a_{lm} covariance — use for likelihood maximization
  - (nb. this has maximum entropy and so is conservative!)
  - *Diagonal* likelihood:

$$P(d_{\ell m} \mid C_\ell I) = \frac{1}{\left[2\pi\left\langle \hat{C}_\ell + N_\ell \right\rangle\right]^{1/2}} \exp\left[-\frac{1}{2}\frac{\left|d_{\ell m}\right|^2}{\left\langle \hat{C}_\ell + N_\ell \right\rangle}\right]$$

  - MC evaluation of means;
  - Newton-Raphson iteration towards maximum
  - Easy calculation of Likelihood shape parameters

B98, CBI; Contaldi et al

(related suggestions from Delabrouille et al)

WMAP: Cross-correlations

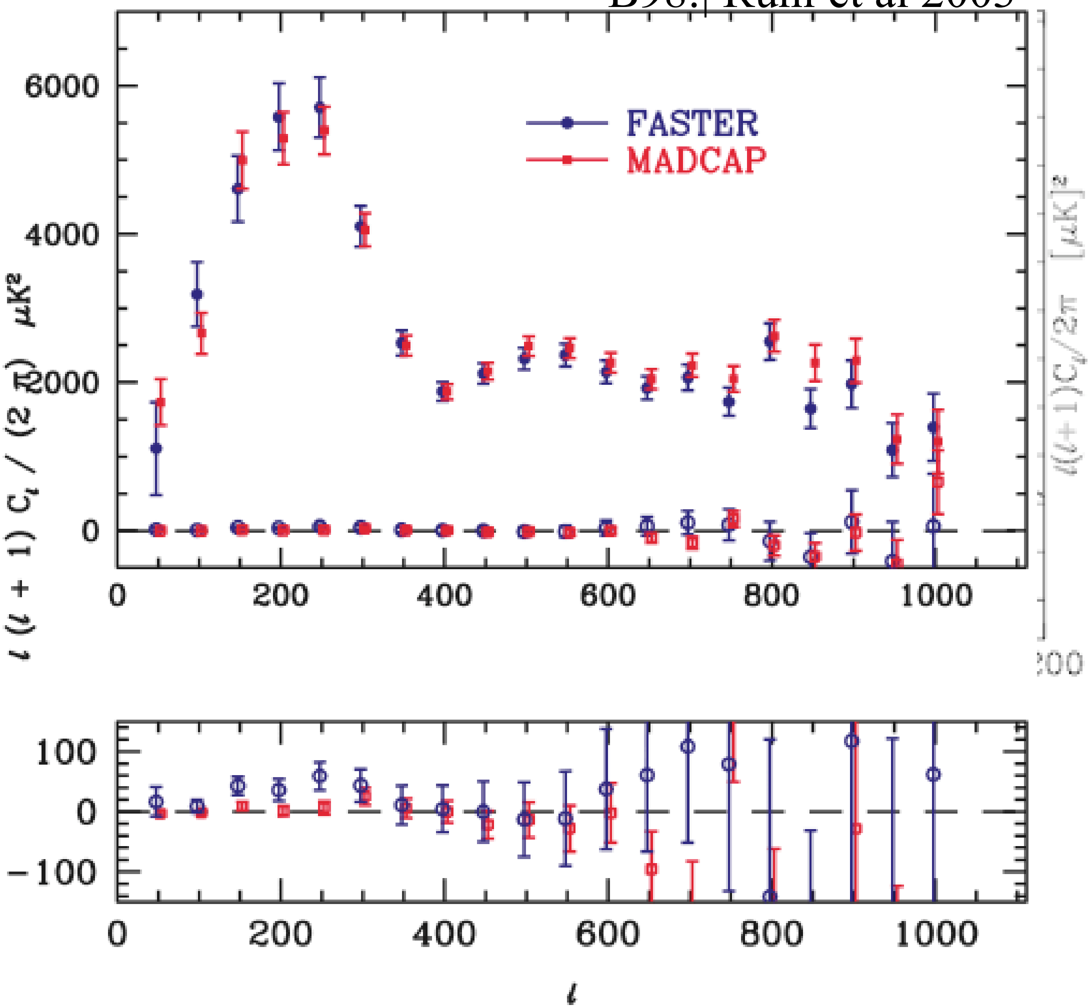
- Take advantage of uncorrelated noise between different detectors
  $$\left\langle d_p^1 d_{p'}^2 \right\rangle = \left\langle \left(s_p^1 + n_p^1\right)\left(s_{p'}^2 + n_{p'}^2\right) \right\rangle = S_{pp'}^{12} + N_{pp'}^{12} = S_{pp'}$$
  Monte Carlo method — without need for noise bias removal
- (also Archeops—XSPECT; Polenta et al)

Method Miscellenea

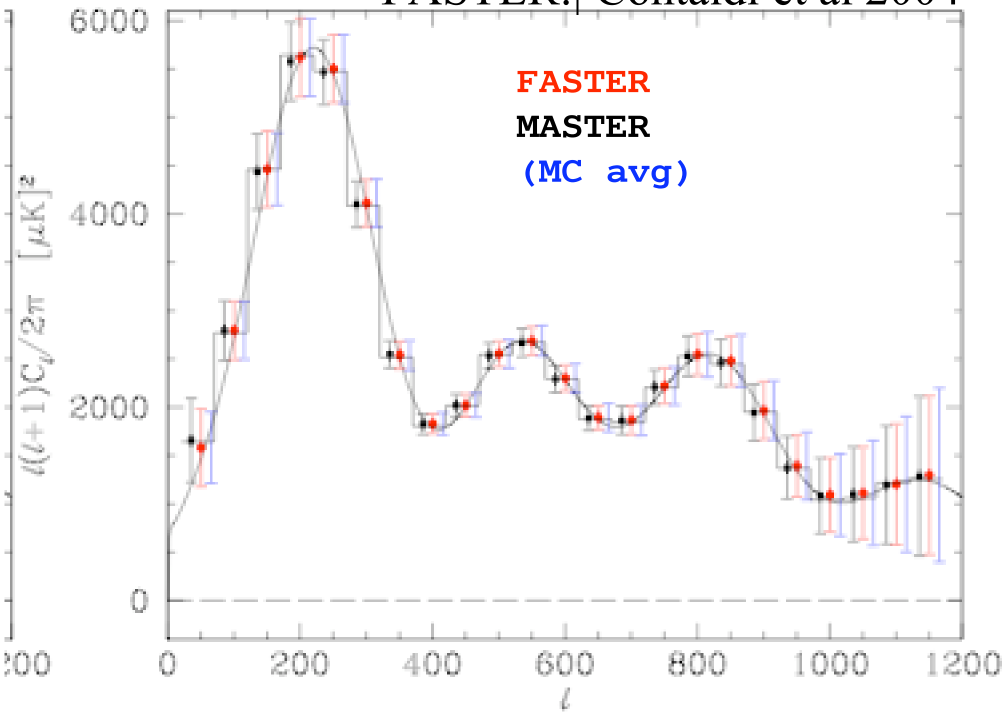
- Efstathiou: Bayes/Frequentist discrepency potentially largest at low *l* — Bayes for low *l*, MC for high *l*
- Knox/Dore/Peel — hierarchical quadratic estimator
- Ring/Harmonic Methods
  - Wandelt et al — full pseudo-C_l likelihood
  - Challinor, van Leeuwen et al
- MCMC search for C_l (Wandelt)

Comparisons

B98:| Ruhl et al 2003



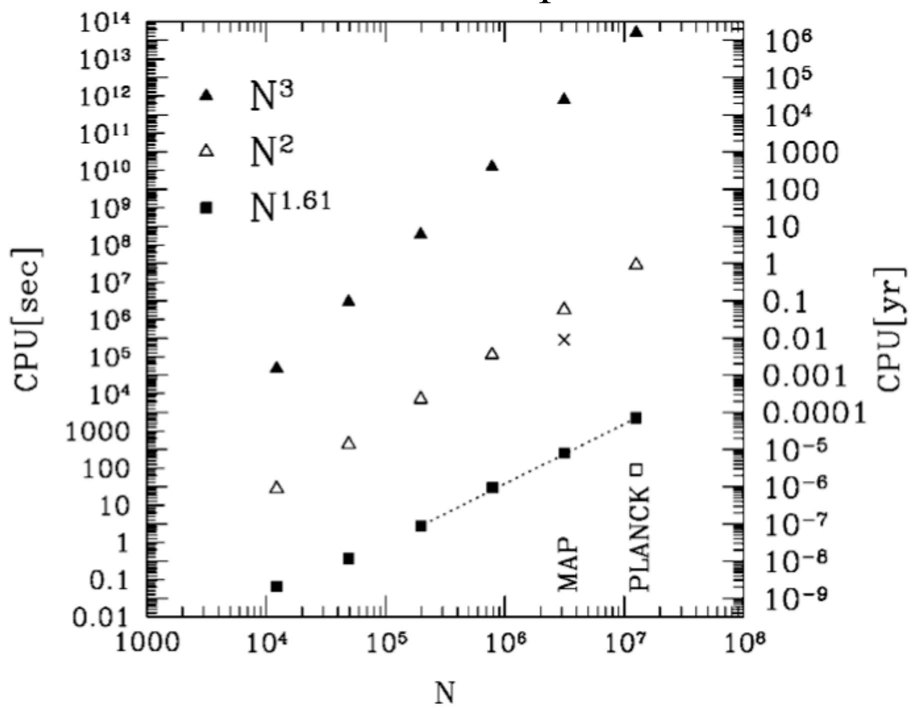
FASTER:| Contaldi et al 2004



Timing and efficiency

- time
  - optimal/bayes: N_p^3
  - monte carlo: $N^{1.5}$
  - prefactors: N_{MC}, N_{bin}, …
- Space
  - TOI: 50 GB/yr @200Hz
  - maps: 384 Mb @ N_{side}=2048
  - noise matrix: $N^2/2$ entries
    ~9 petabytes @ N_{side}=2048

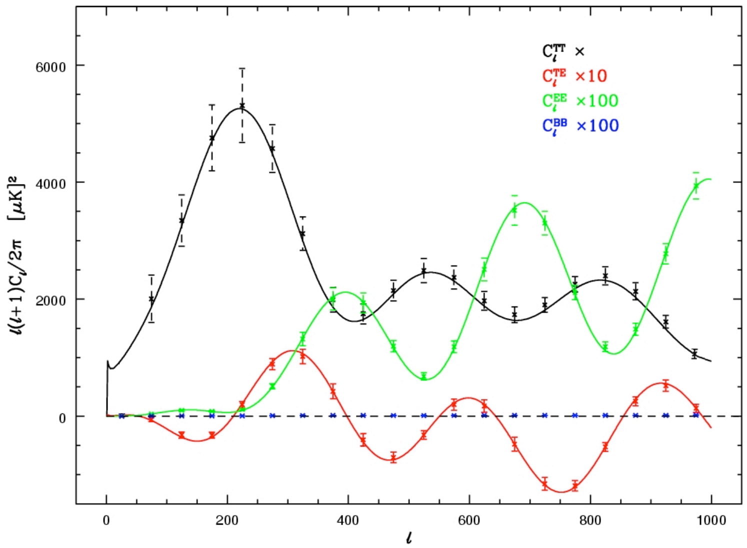
SPICE: Szapudi et al



resource management will become an issue even for cheapest methods

Polarization

- Formally the same problem:
  - $d_p \Rightarrow (i,q,u)_p = d_{i,p} = d_q$
  - $\langle d_q d_{q'} \rangle = N_{qq'} + S_{qq'}$
    - low S/N, large systematics
    - complicated correlations:
      - $N_{qq'}$: pixel differences
      - $S_{qq'} = S^{ij}_{qq'}$: linearly dependent on all of $C_l^{XX'}$ (X=T,E,B)
  - e.g., Seljak, Zaldarriaga; Kamionkowski, Kosowsky, Stebbins; &c.



- E/B leakage (= T/E/B correlation)
  - in principle, don't need extra separation step if full correlations/distributions is known
  - in practice, E/B characteristics impose specific correlation structure — easier to "separate"
    - e.g., Lewis talk — separate at map or C_l?
    - Wiener filter for map from C_l.

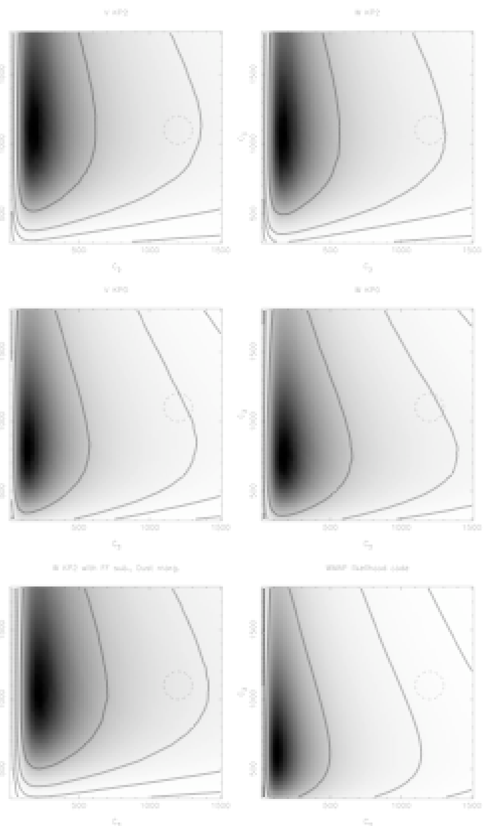
Interferometers

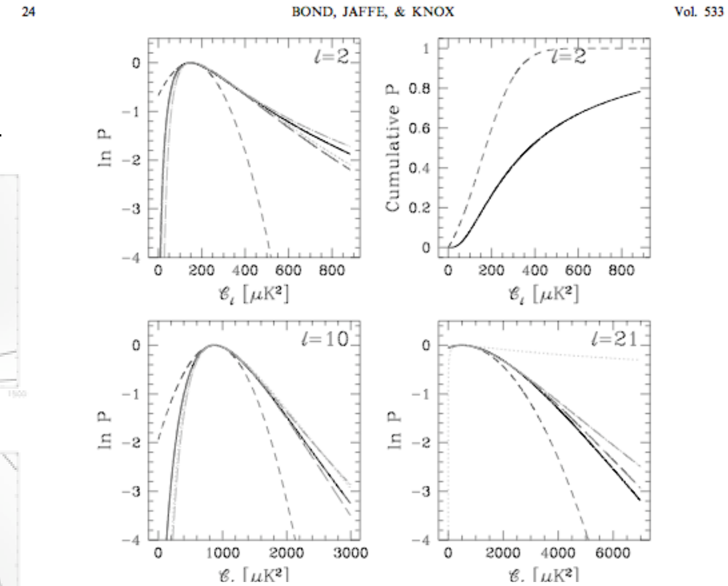
- ~Direct measurement of binned spherical harmonic components
  - great simplification: noise and signal correlations simple in the same basis
- CAT: bandpower likelihoods
- DASI, Hobson & Maisinger/VSA: Likelihood/Bayesian methods

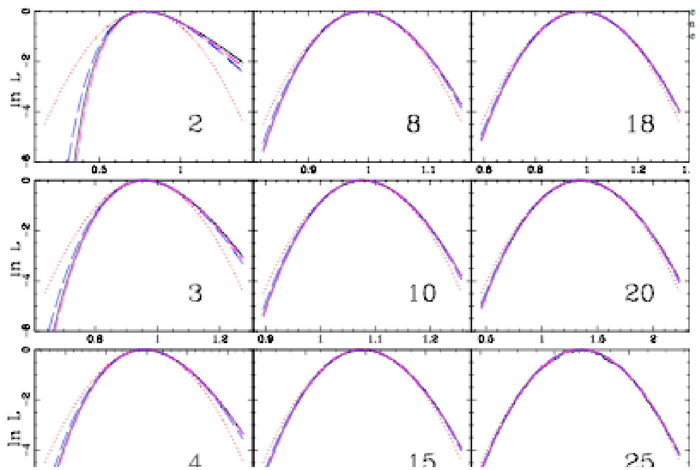
Parameter estimation from C_l

- $P(d|\theta I) = P(d|C_l[\theta])$ [Bayes]
  - explore w/ grids or MCMC (Knox et al; CosmoMC; Dunkley et al; WMAP)
  - no freq. alternative?
- shape of likelihood $L(C_l)$.
- BJK 00 &c: offset lognormal distribution/eq. var. approx.
- WMAP approx — but e.g., Slosar et al
  - Potentially breaks down
    - in tails (should be power law ~ 1/Cn)
    - in presence of correlations
- Bandpowers: window & filter fns (Knox)

Slosar et al 2004



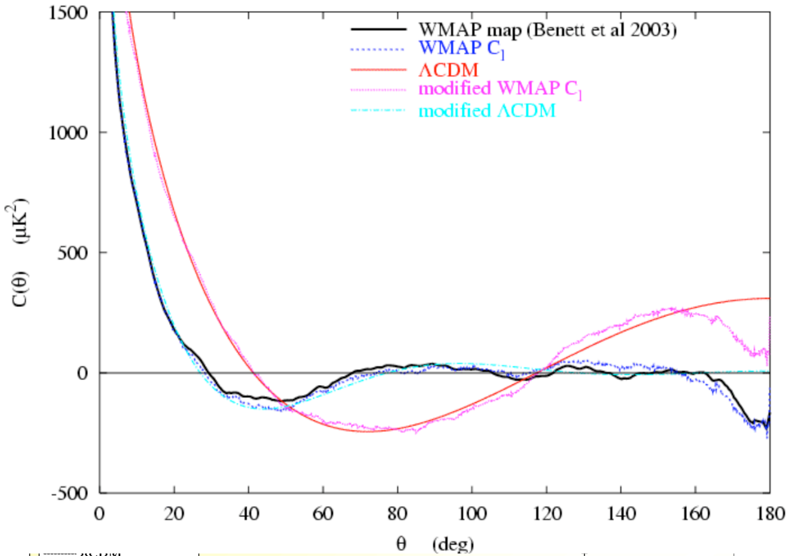


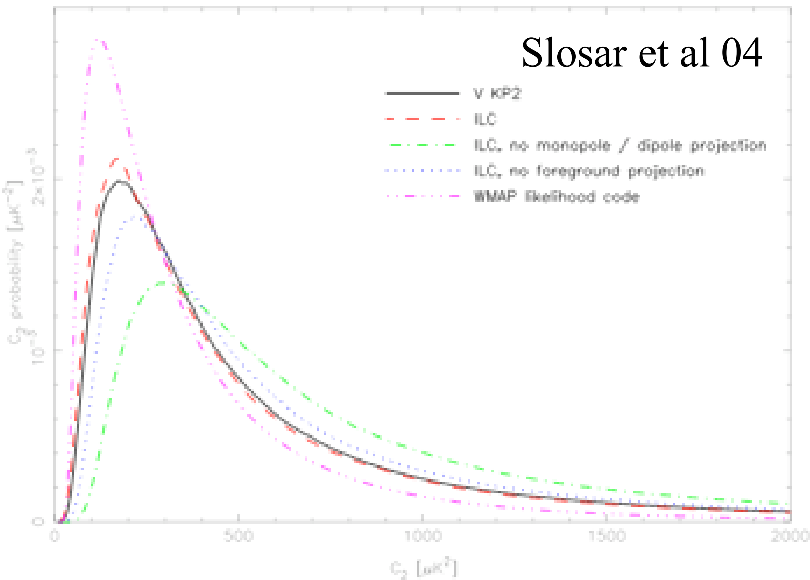


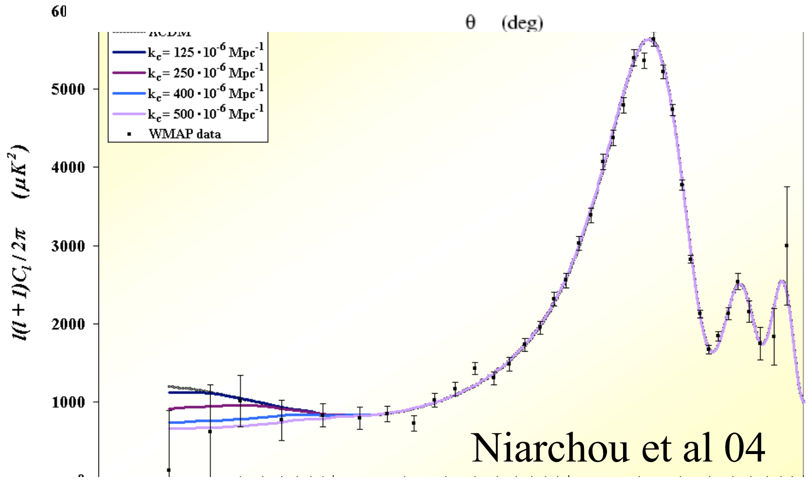
FASTER: Contaldi et al 04

The WMAP quadrupole

- *l*=2,3: only 2.7σ (Bayes model comparison; see Liddle talk)
  - quadrupole, octopole alignment?
- other "anomolies"
  - l~25; first peak; C_7^{TE}







The future of C_l

- Extensions:
  - C_l does assume *isotropy*
- Propagating noise
  timestream ⇒ maps ⇒ C_l ⇒ cosmology
  - statistics and systematics
    - MADCAP: use N(t-t'); Stompor & White; Ashdown et al: rings (Planck)
- Asymmetric beams/beam errors
- Combining results — after the fact or before
- Noise estimation and errors
- Details: likelihood shape; window functions; beam/calibration error,…

Conclusions

- a dozen methods out there
  - Bayes/freq, Monte Carlo, correlation function, apodization, …
  - all approximations to ‘optimal’ Bayesian method
  - all agree (in simple cases)
- for Precision Cosmology
  - compare with exact/optimal in more complicated cases
  - requires wider tests & comparisons — correlations, non-Gaussianity, etc.