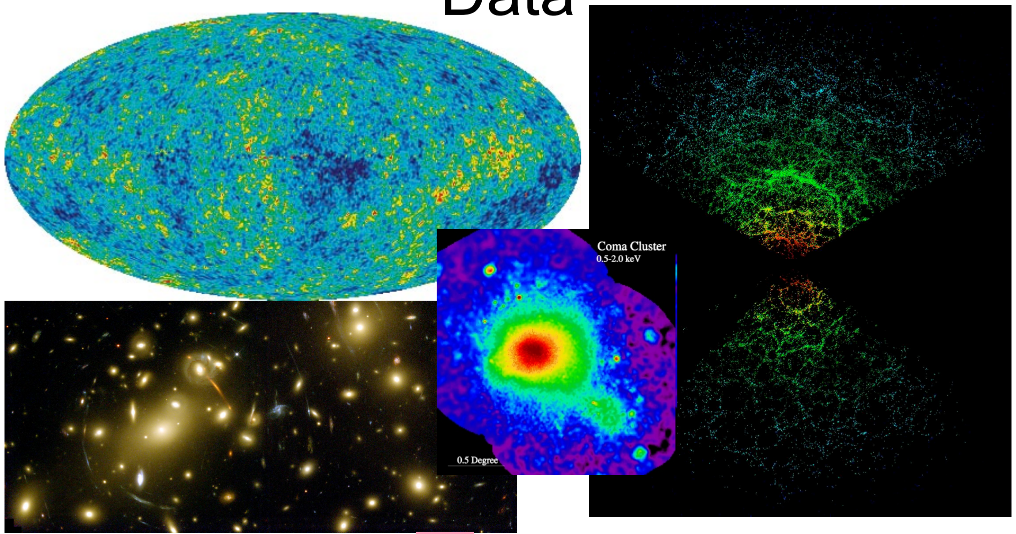
Non-Gaussian errors of large-scale structure probes

Masahiro Takada (IPMU)

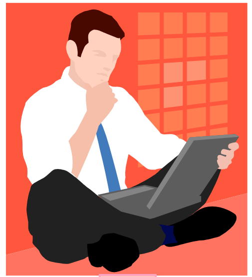
INSTITUTE FOR THE PHYSICS AND MATHEMATICS OF THE UNIVERSE

Cosmology vs. Statistics

Data



Theory



C_l^{obs} ⟷ Statistical quantities: e.g., $C_l^{\mathrm{model}}(\boldsymbol{p})$

Parameter estimation

- Cosmological model predicts statistical properties of the Universe as a function of a handful set of parameters
- Statistics is important in several steps of this procedure

Parameter estimation in cosmology: maximum Likelihood analysis

- Bayes' Theorem

Likelihood function of data

Priors on parameters

$$P(\boldsymbol{p} \mid \boldsymbol{C}^{\mathrm{obs}}) \propto \underbrace{L\left(\boldsymbol{C}^{\mathrm{obs}} \mid \boldsymbol{C}^{\mathrm{model}}(\boldsymbol{p})\right)}_{\text{Likelihood function of data}} \underbrace{P(\boldsymbol{p})}_{\text{Priors on parameters}}$$

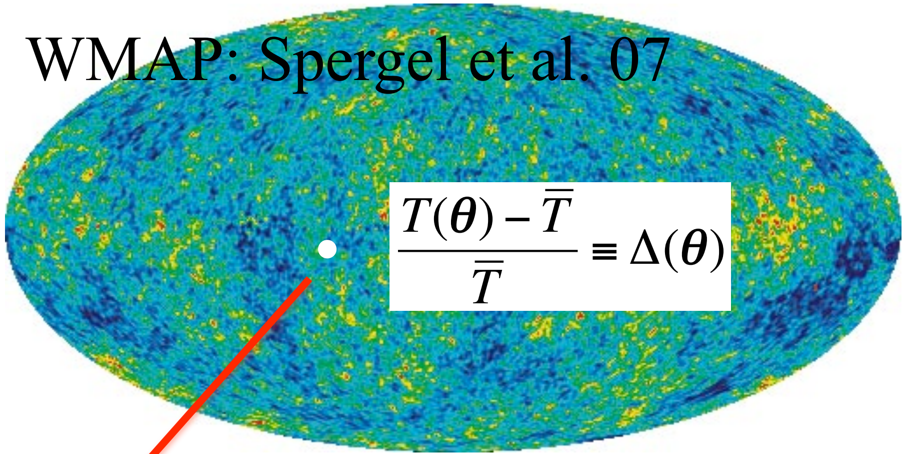
$\boldsymbol{C}^{\mathrm{obs}}$: data vector (observables: e.g. power spectrum)

$\boldsymbol{C}^{\mathrm{model}}$: model vector of the observables

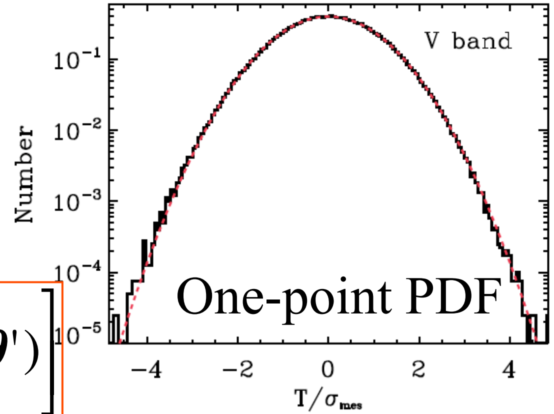
$\boldsymbol{p}$: model parameters

- In cosmology, the likelihood function of data can be fairly well modeled
- Cosmological parameters are estimated according to the maximum likelihood method

The Case of CMB: Gaussian fluctuations



- The observed CMB temperature fluctuation field is consistent with a Gaussian field (also as predicted by inflationary scenario)
- The statistical properties of Gaussian field is fully characterized by the two-point correlation function



$$L(\Delta(\boldsymbol{\theta})\,|\,\boldsymbol{C}) \propto \frac{1}{\sqrt{\det \boldsymbol{C}}} \exp\left[-\frac{1}{2}\oint\frac{d\boldsymbol{\theta}}{4\pi}\oint\frac{d\boldsymbol{\theta}'}{4\pi}\Delta(\boldsymbol{\theta})\boldsymbol{C}^{-1}(|\boldsymbol{\theta}-\boldsymbol{\theta}'|)\Delta(\boldsymbol{\theta}')\right]$$

$\Delta(\boldsymbol{\theta})$: observed temperture field

$\boldsymbol{C}(|\boldsymbol{\theta}-\boldsymbol{\theta}'|) = \langle \Delta(\boldsymbol{\theta})\Delta(\boldsymbol{\theta}')\rangle$: 2pt correlation func.

$$= \frac{1}{4\pi}\sum_l (2l+1) C_l^{\rm model} P_l(\hat{\boldsymbol{\theta}}\cdot\hat{\boldsymbol{\theta}}')$$

C_l: power spectrum, given as a function of cosmo paras

CMB: Gaussian Field (contd.)

e.g. Bond et al. 01; Verde et al. 03

- Likelihood function

$$L\left(\Delta(\boldsymbol{\theta})\,|\,\boldsymbol{C}\right) \propto \frac{1}{\sqrt{\det \boldsymbol{C}}}\exp\left[-\frac{1}{2}\oint\frac{d\boldsymbol{\theta}}{4\pi}\oint\frac{d\boldsymbol{\theta}'}{4\pi}\Delta(\boldsymbol{\theta})\boldsymbol{C}^{-1}(|\boldsymbol{\theta}-\boldsymbol{\theta}'|)\Delta(\boldsymbol{\theta}')\right]$$

$$\Delta(\boldsymbol{\theta}) = \sum_{l=1}^{\infty}\sum_{m=-l}^{l} a_{lm}Y_{lm}(\boldsymbol{\theta})$$

$$C_l^{\rm obs} = \frac{1}{2l+1}\sum_{m=-l}^{l}|a_{lm}|^2$$

+ the orthogonal relation of the spherical harmonics Y_lm

- Can reduce to the likelihood function of the estimated power spectrum

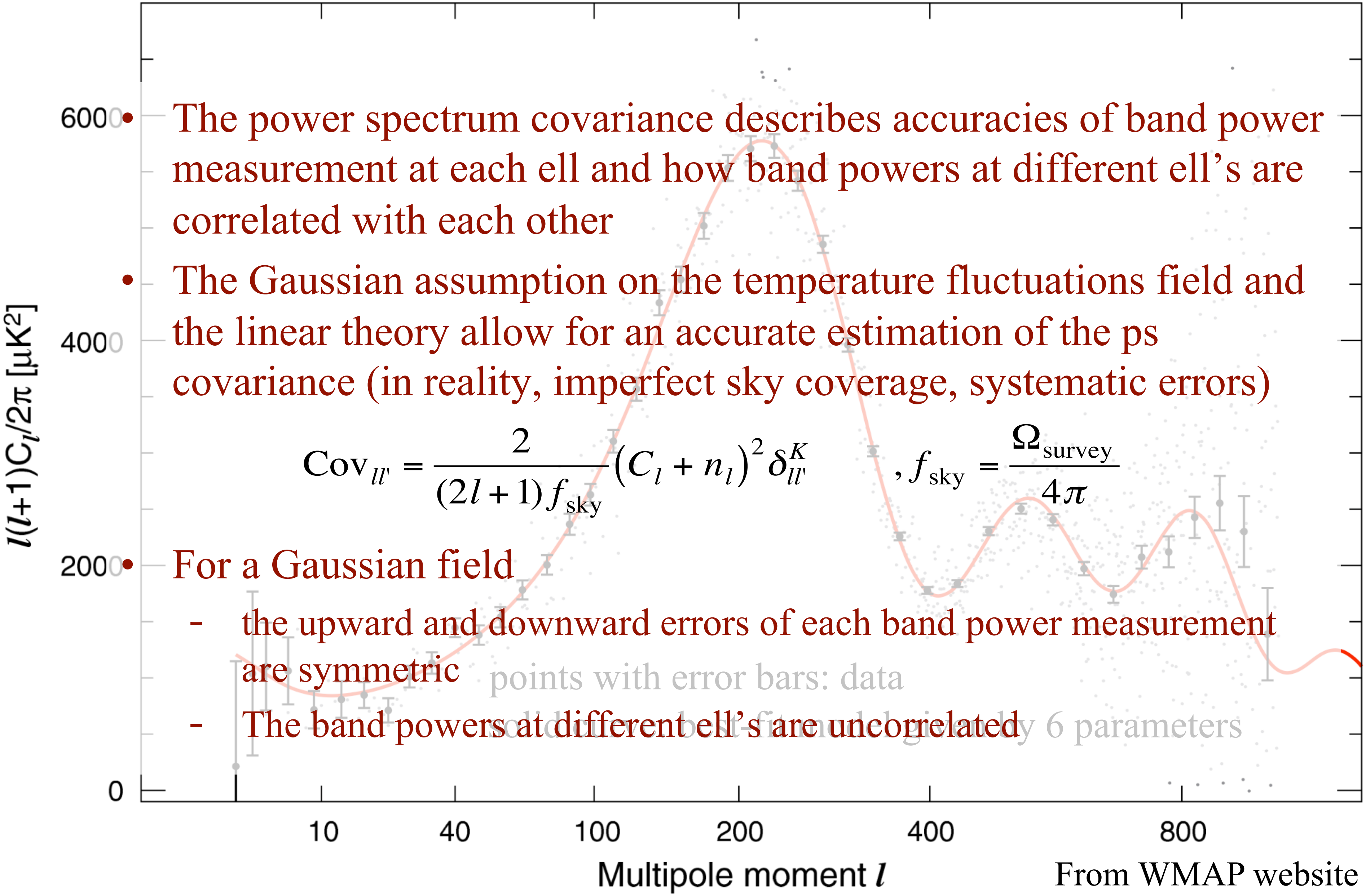
$$-2\ln\left[L\left(C_l^{\rm obs}\,|\,C_l^{\rm model}(\boldsymbol{p})\right)\right] \propto \sum_l (2l+1)\left(\ln\frac{C_l^{\rm model}}{C_l^{\rm obs}} + \frac{C_l^{\rm obs}}{C_l^{\rm model}} - 1\right)$$

- At large *l* limit the likelihood can be approximated as a Gaussian form:

$$\ln\left[L\left(C_l^{\rm obs}\,|\,C_l^{\rm model}(\boldsymbol{p})\right)\right] \propto -\frac{1}{2}(C_l^{\rm obs} - C_l^{\rm model})[\mathbf{Cov}]_{ll'}^{-1}(C_{l'}^{\rm obs} - C_{l'}^{\rm model})$$

PS covariance (needs to be modeled): $[\mathbf{Cov}]_{ll'} = \left\langle C_l C_{l'}\right\rangle - C_l^{\rm model}C_{l'}^{\rm model}$

CMB (contd.): theory vs. data



- The power spectrum covariance describes accuracies of band power measurement at each ell and how band powers at different ell's are correlated with each other
- The Gaussian assumption on the temperature fluctuations field and the linear theory allow for an accurate estimation of the ps covariance (in reality, imperfect sky coverage, systematic errors)

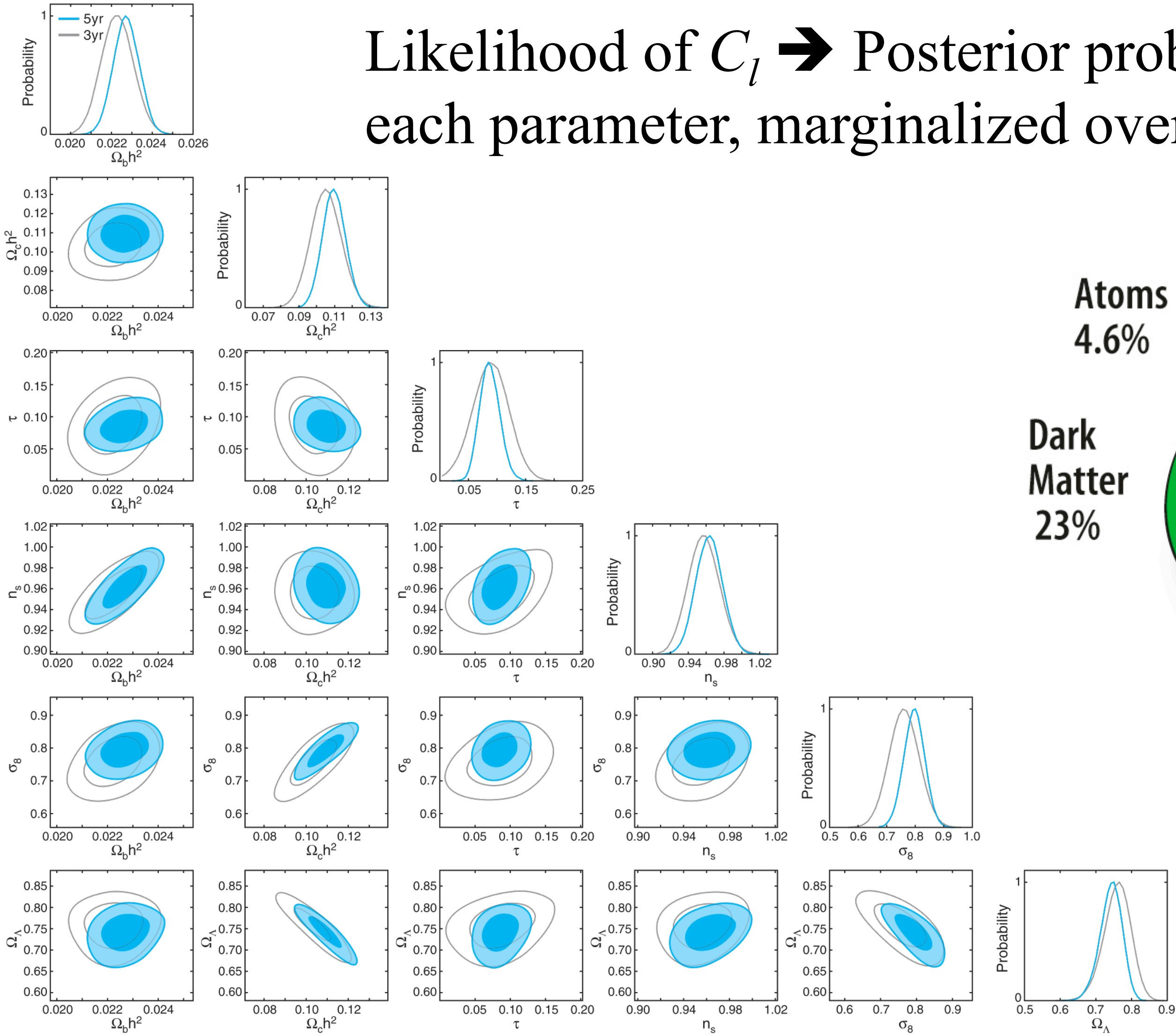
$$\mathrm{Cov}_{ll'} = \frac{2}{(2l+1) f_{\mathrm{sky}}} (C_l + n_l)^2 \delta^K_{ll'} \quad , f_{\mathrm{sky}} = \frac{\Omega_{\mathrm{survey}}}{4\pi}$$

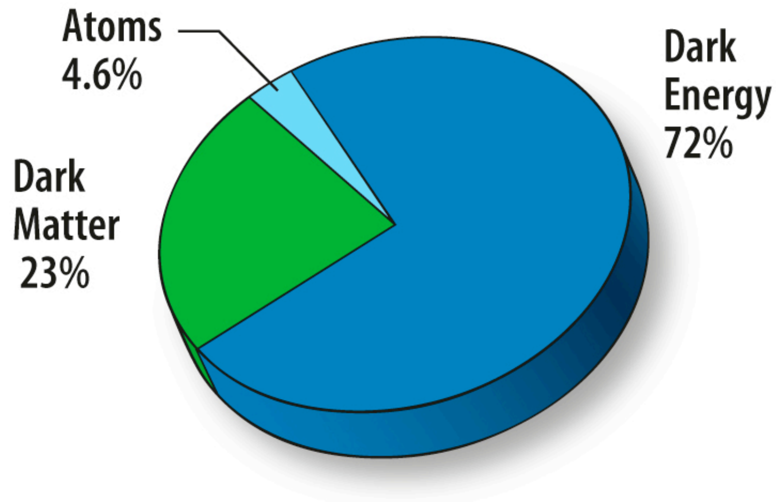
- For a Gaussian field
  - the upward and downward errors of each band power measurement are symmetric
  - The band powers at different ell's are uncorrelated

From WMAP website

CMB (contd.): parameter estimation

Likelihood of C_l ➔ Posterior prob. distribution of each parameter, marginalized over other parameters





Dark matter: 23%

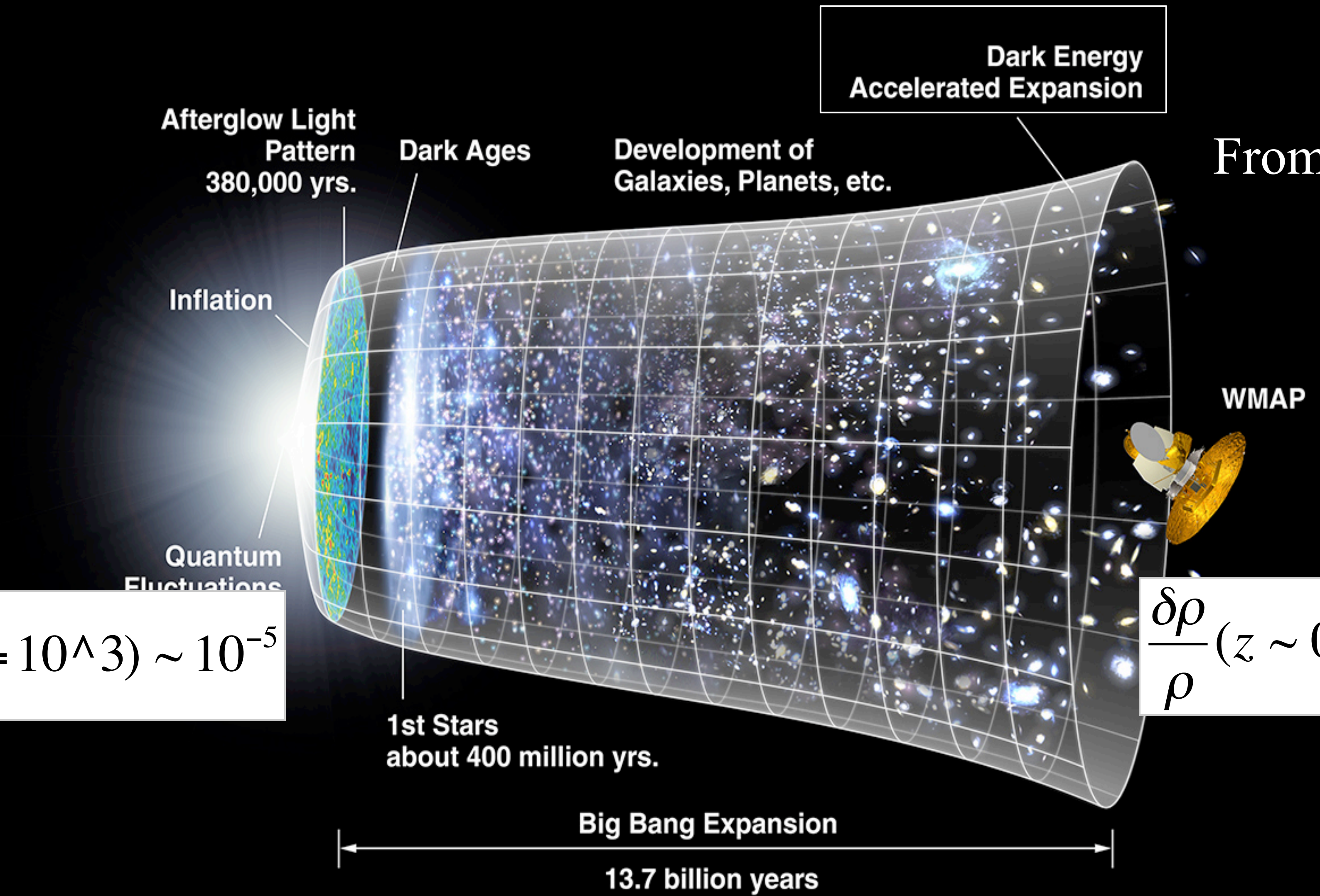
Dark energy: 72%

Baryons: 4.6%

Neutrinos:<1%

WMAP team: Dunkley et al. 09; Komatsu et al. 09

Large-scale Structure Probe
- Growth of density fluctuations -

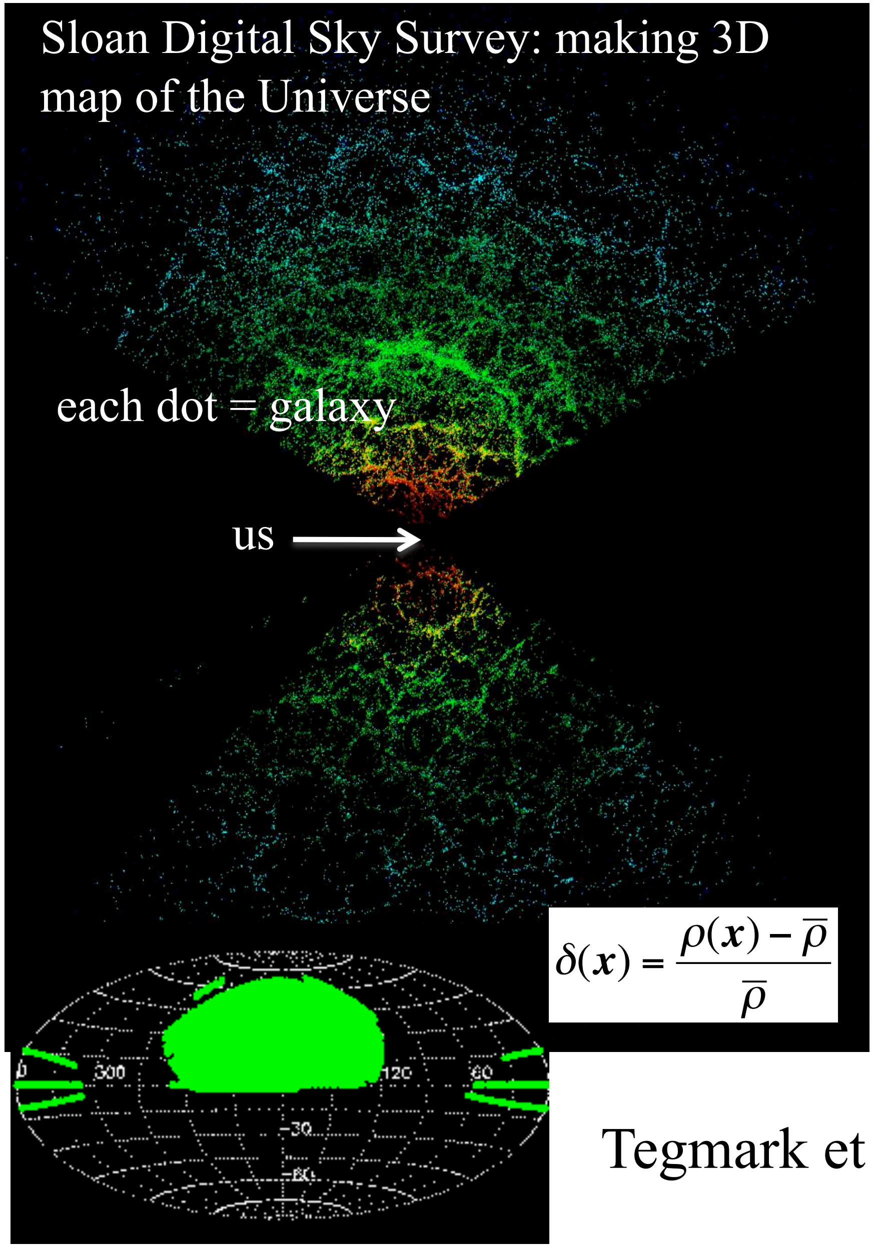


From WMAP

$$\frac{\delta T}{T}(z = 10\wedge 3) \sim 10^{-5}$$

$$\frac{\delta\rho}{\rho}(z \sim 0) \sim 1$$

Large-scale Structure Formation



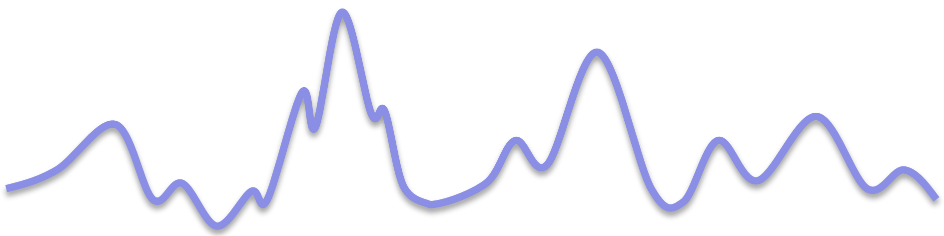
$$\delta(\boldsymbol{x}) = \frac{\rho(\boldsymbol{x}) - \bar{\rho}}{\bar{\rho}}$$

Tegmark et al. 03

- In the present-day universe, *invisible* dark matter plays a major role in structure formation
- Need to infer clustering strengths of dark matter distribution from visible/astronomical objects
  - Galaxy distribution
  - Lensing effects on distant galaxies
  - Clusters of galaxies: counting statistics & clustering analysis
  - Intergalactic medium: Hydrogen distribution (Lyman-alpha: 21cm cosmology)

LSS (contd.)

Tiny density fluctuations at z~1000: δ_m~10^-3

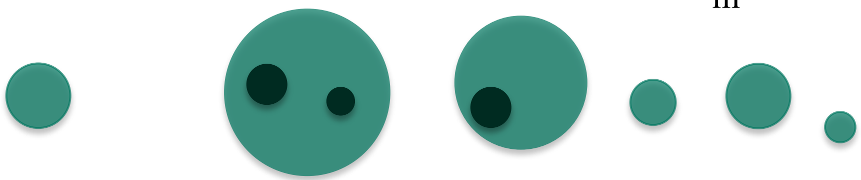


Gravitational instability

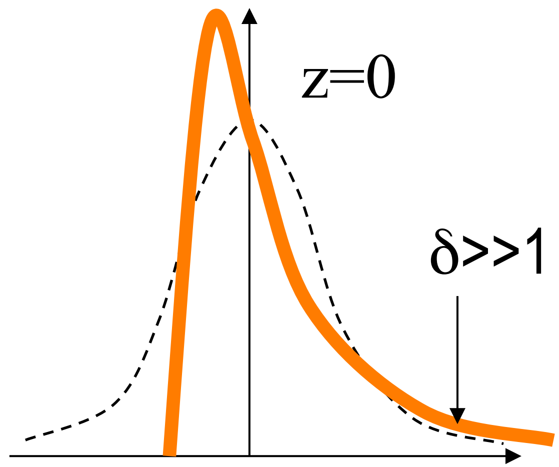
(gravity ⇔ cosmic expansion)

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G\bar{\rho}_m\delta_m = 0$$

Dark matter halo formation at z~0: δ_m>>1



- The seed tiny fluctuations grow up to the nonlinear regime ($\delta \geq O(1)$) by gravitational instability
- Due to the nonlinearity of gravity, the present-day field is non-Gaussian



Nonlinear clustering regime: mode-coupling

- Linear regime O(δ)<<1; all the Fourier modes of the perturbations grow at the same rate; the growth rate *D*(z)
  - The linearized perturbation theory (FRW + GR) gives an accurate modeling

$$\delta_{\boldsymbol{k}}(z) = D(z)\delta_{\boldsymbol{k}}(z=1000)$$

- Mildly non-linear regime O(δ)~1; a mode coupling between different Fourier modes is induced
  - The higher-order perturbation theory predicts the mode-coupling btw different modes arise: different wavenumber modes are not independent

$$\delta(z) = \delta^{(1)} + \delta^{(2)} + \delta^{(3)} + \cdots$$

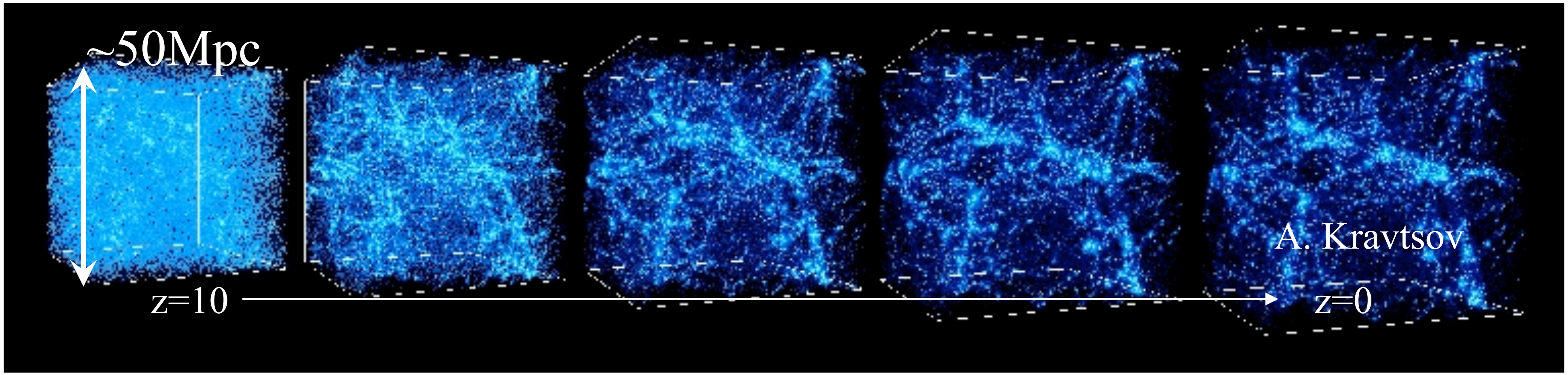
$$\delta_{\boldsymbol{k}}^{(2)} = \int d^3\boldsymbol{k}_1 \int d^3\boldsymbol{k}_2 F(\boldsymbol{k}_1, \boldsymbol{k}_2)\delta_{\boldsymbol{k}_1}^{(1)}\delta_{\boldsymbol{k}_2}^{(1)}\delta(\boldsymbol{k} - \boldsymbol{k}_1 - \boldsymbol{k}_2)$$

$$\Rightarrow \left\langle \delta^3 \right\rangle \propto \left\langle (\delta^{(1)})^2 \delta^{(2)} \right\rangle \propto \left\langle (\delta^{(1)})^4 \right\rangle \neq 0$$

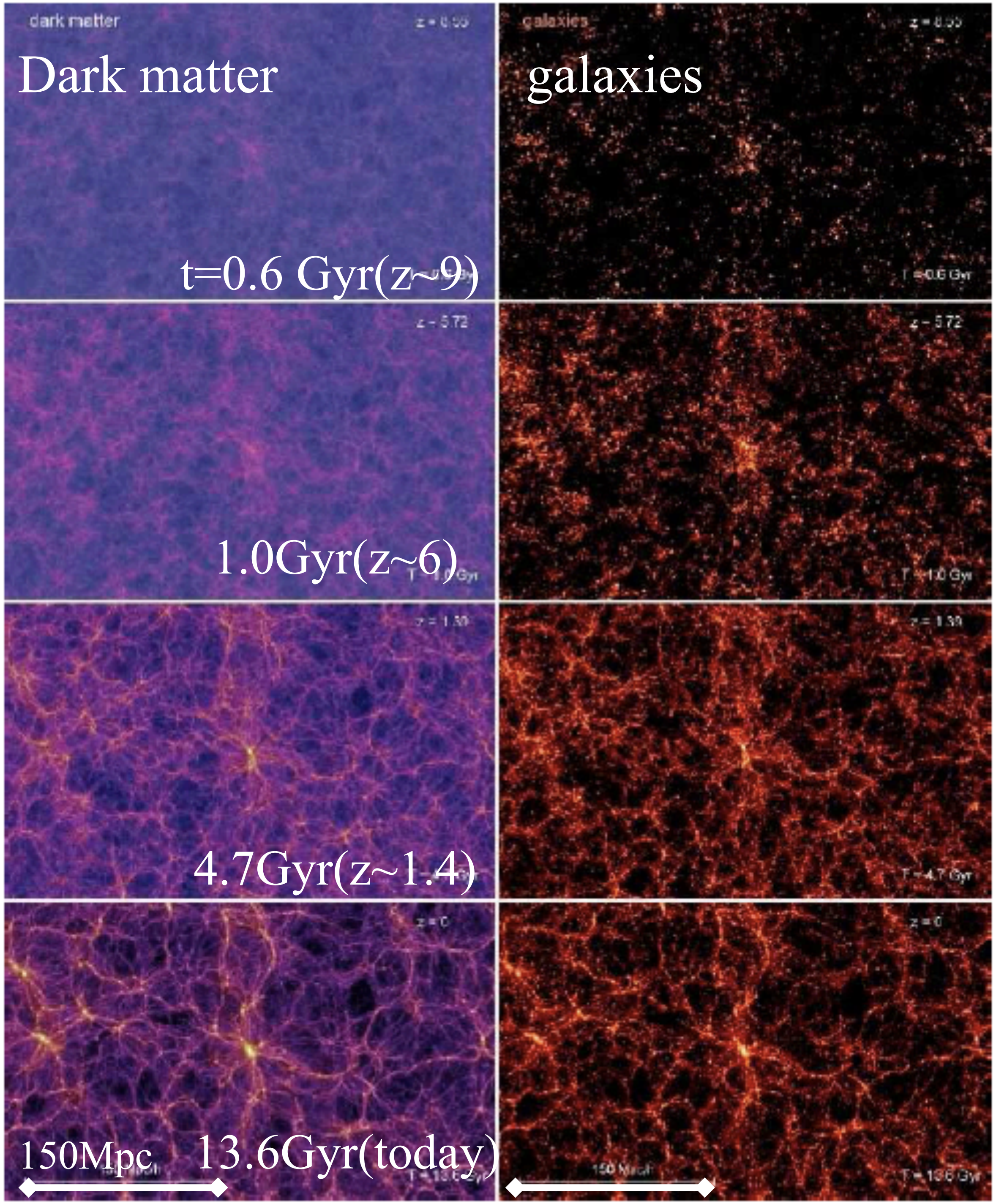
- Highly non-linear regime (δ>>1); a more complicated mode coupling
  - N-body simulation based methods or an empirical semi-analytic method

Modeling nonlinear LSS formation - N-body simulations -

- The initial conditions of SF is now well constrained by CMB
- The cold dark matter (CDM model) dominated SF scenario has been remarkably successful in explaining various observations
- In a CDM model, gravity due to dark matter distribution plays a major role
- N-body simulation based method is becoming most powerful method to follow nonlinear clustering processes in SF
  - N-body particle = DM super particle; e.g. each N-body particle = 10^11 M_sun = 10^50 DM particles
- Simulations have been used in various cosmological studies



CDM dominated structure formation scenario

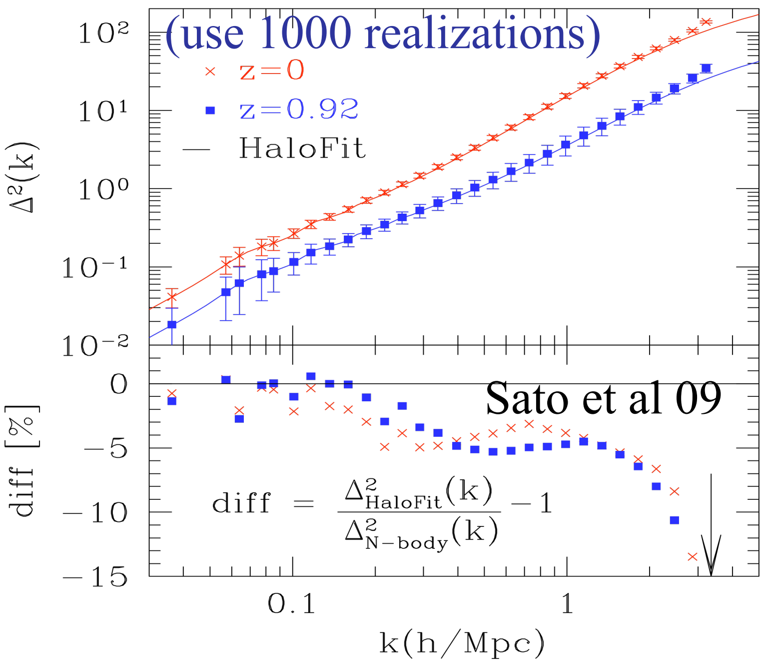


Springel+(including N.Yoshida)05, Nature

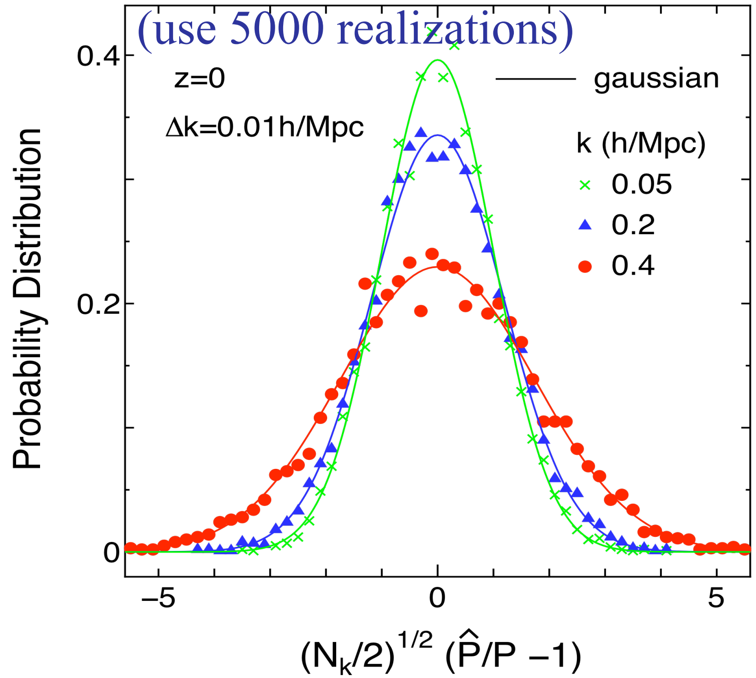
Cosmic calibration

- With the advent of high-performance computer resources, it is becoming possible to make an end-to-end simulation of cosmological observables
- Possible to simulate the likelihood function of PS band powers at each wavelengths (e.g. use 10^3-10^4 realizations), *but only for one cosmological model (even harder if including gastrophysics)*
- Still challenging to perform a full likelihood analysis of LSS probes (usually assume the Gaussian error assumption or the χ^2–fitting)

DM power spectra at z=0, 1 (use 1000 realizations)

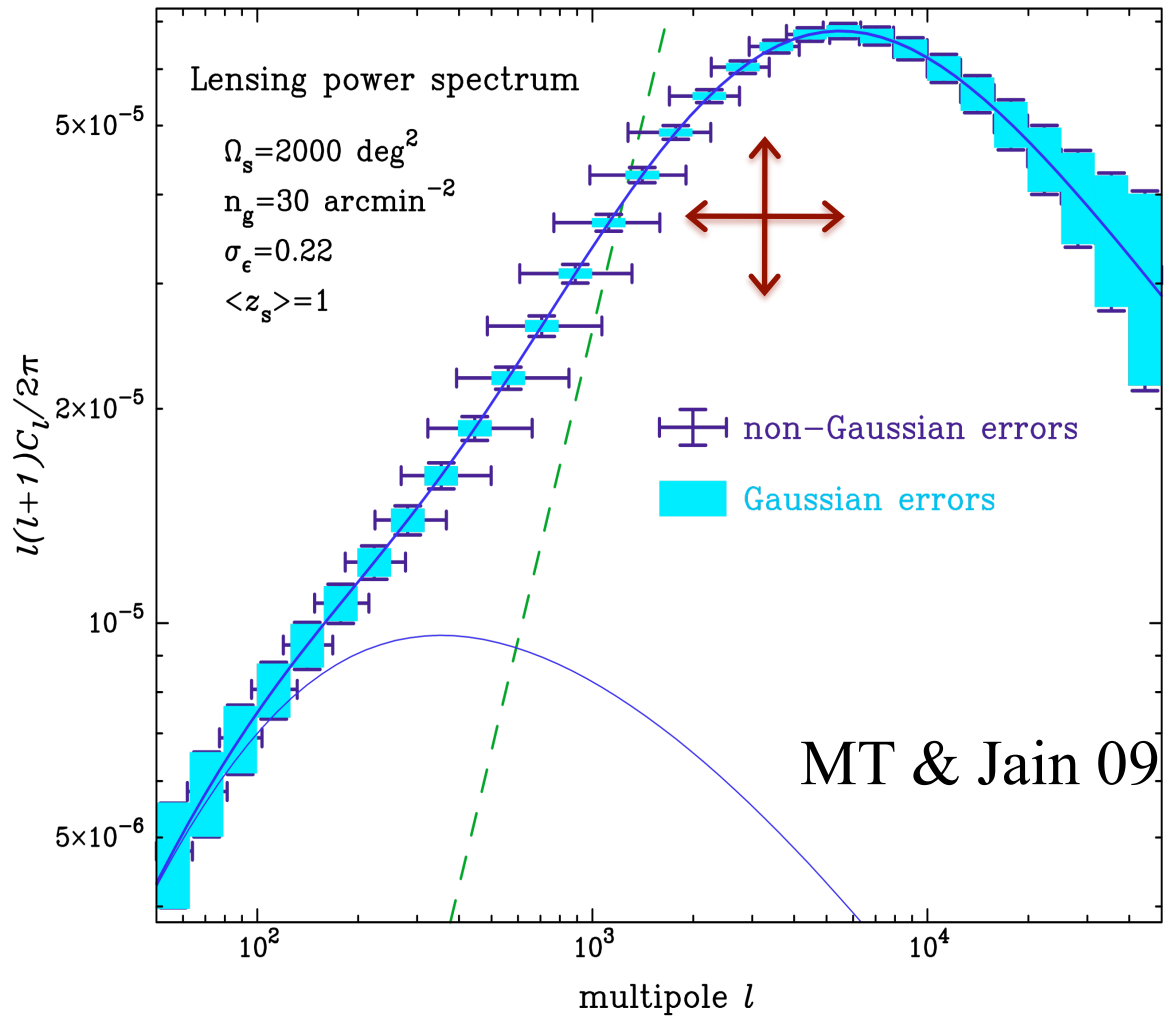


Likelihood of DM PS band powers (use 5000 realizations)



Non-Gaussian errors

Weak lensing power spectrum as an example



- Power spectra of LSS probes are non-Gaussian on relevant scales
- Nonlinear clustering causes…
  - Increase uncertainties in measuring band power at each wavenumber
  - Band powers at different wavenumbers are correlated
  - Upward and downward error bars are in general asymmetric
- The cosmological error bars (the likelihood of C_l) are predictable

Issues needs to be clarified towards precision cosmology

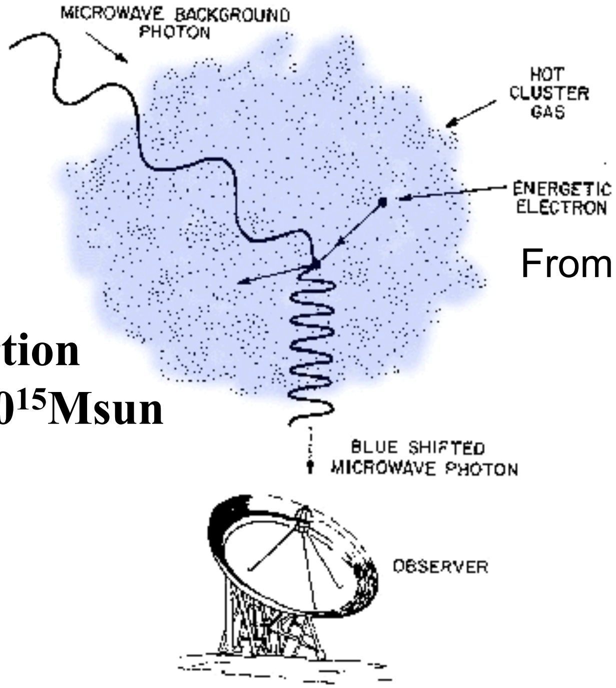
- The impact of non-Gaussian errors?
- The necessity of "non-Gaussian" likelihood analysis for large-scale structure probes, especially in preparation for ongoing and future surveys
- Need to quantify the degree of "bias" in parameter estimation
- The Gaussian type (χ^2 type) likelihood analysis can in practice be a good approximation? If so, need to take into the full power spectrum covariance including the non-Gaussian contributions?

$$\chi^2 = (C_l^{\rm obs} - C_l^{\rm model})[\mathbf{Cov}]^{-1}_{ll'}(C_{l'}^{\rm obs} - C_{l'}^{\rm model})$$

A working example: the Sunyeav Zel'dovich Effect

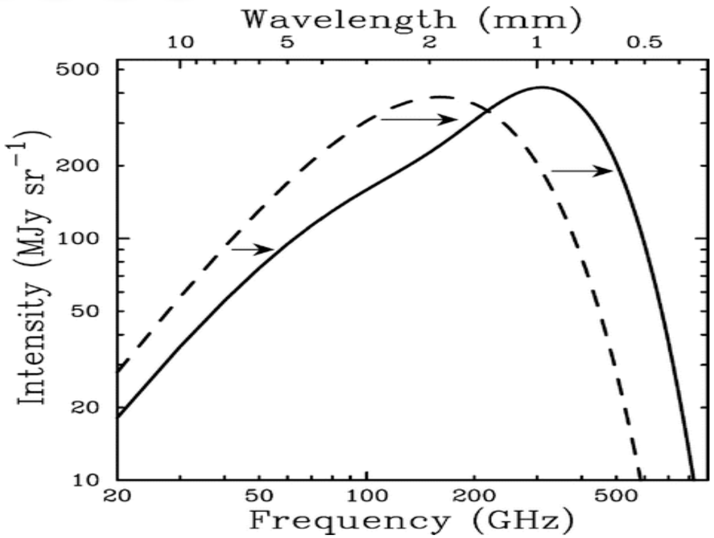
Zel'dovich & Sunyaev 69; Sunyaev & Zel'dovich 80

- Hot electron in a cluster distorts the CMB radiation through the inverse Compton scattering

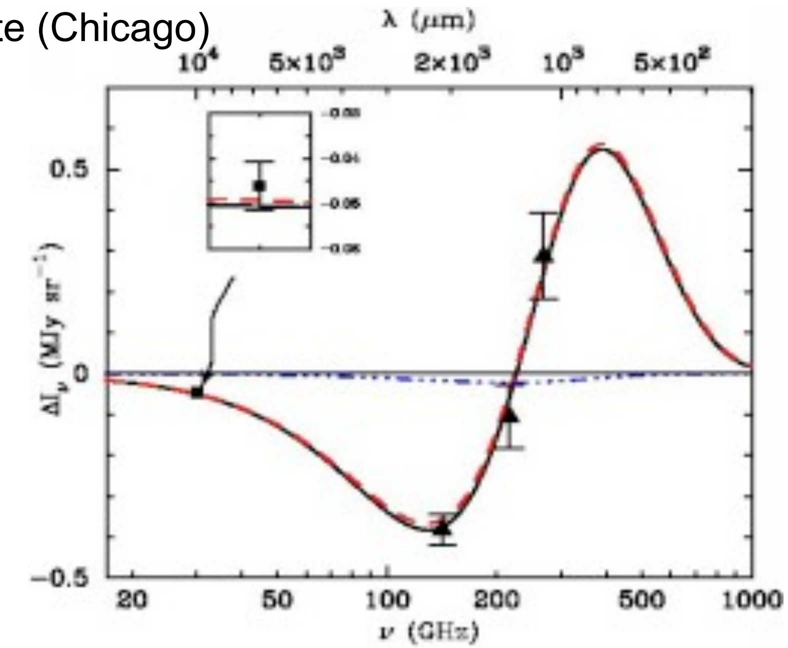


Small distortion
$\tau \sim 0.01$ for 10^{15}Msun

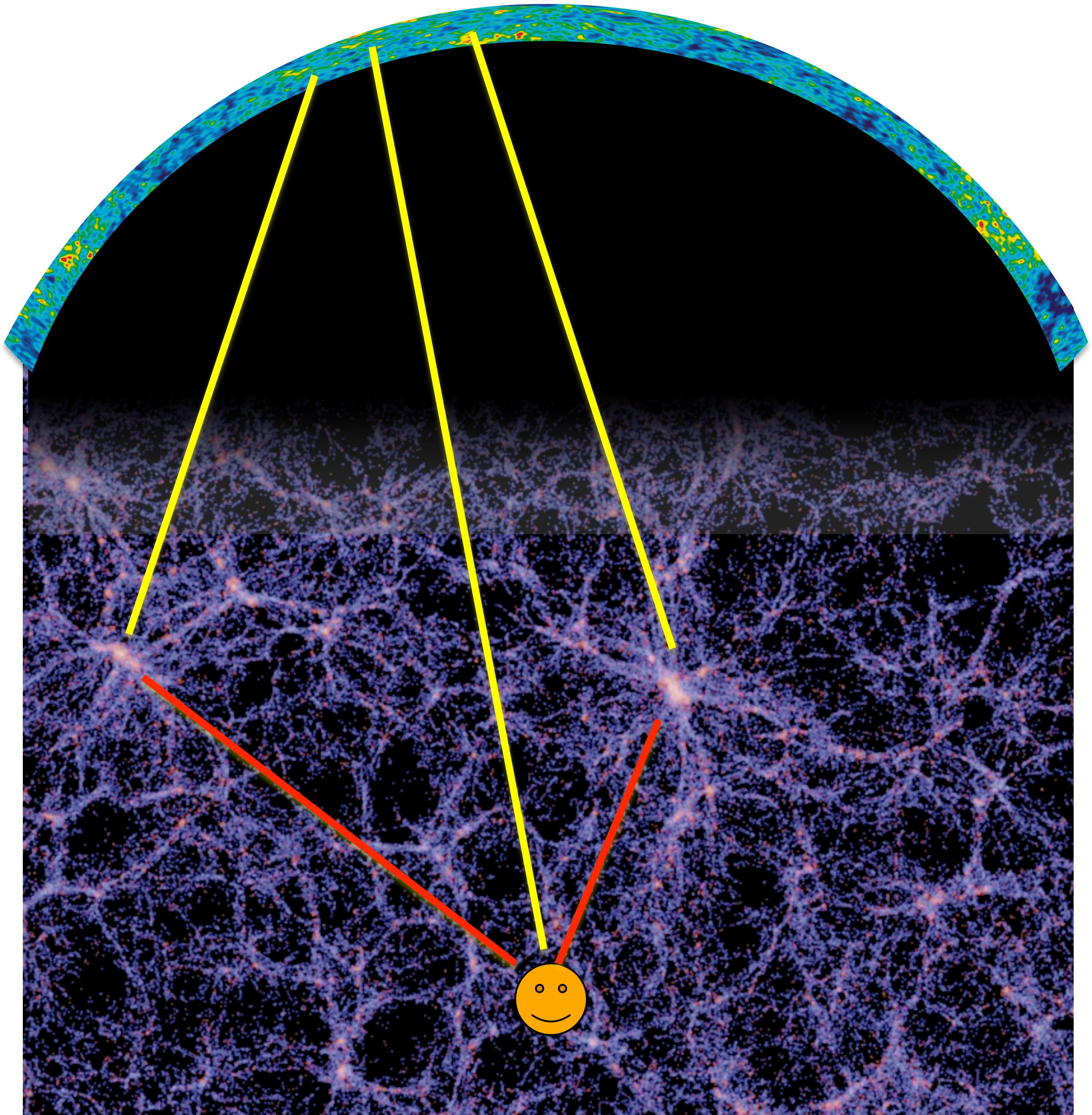
$$\frac{\Delta T_{\rm SZ}}{T_{\rm CMB}} \propto \int n_e T_e dl$$



From SZA website (Chicago)

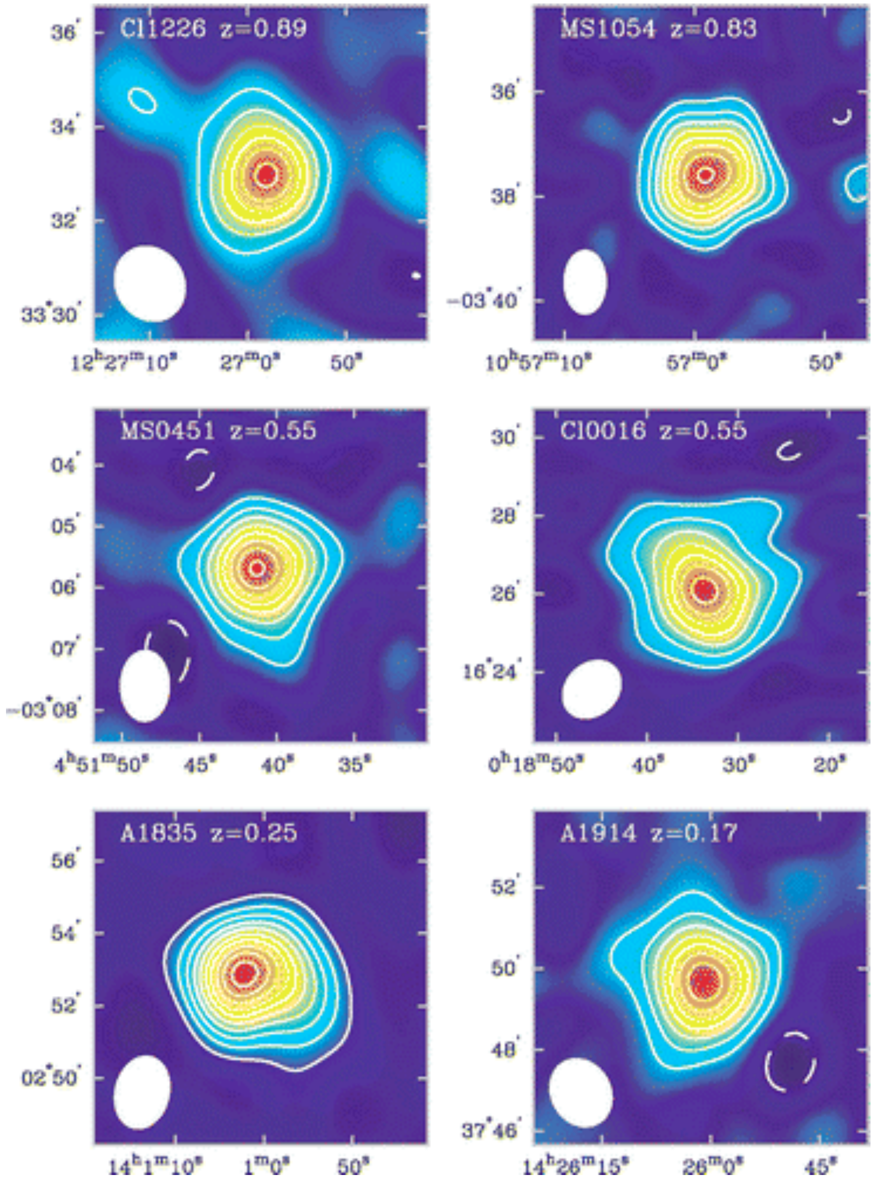


SZ effect (contd.)



- The SZ effect is caused by massive halos (clusters of galaxies)
- Clusters are most massive gravitational objects in the Universe
- Massive clusters are rare (1 per 100^3Mpc^3)

SZ effect (contd.)



Carlstorm et al 02

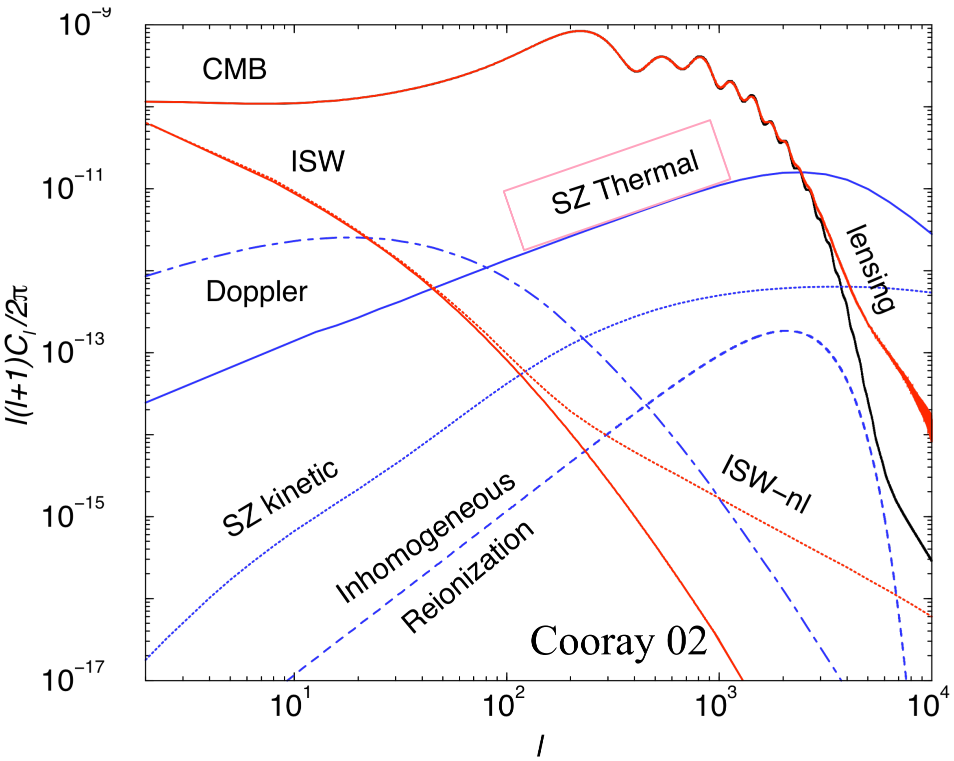
$$\frac{\Delta T_{\mathrm{SZ}}}{T_{\mathrm{CMB}}} \propto \int n_e T_e dl$$

- Unaffected by cosmological dimming effect
- Allow to find clusters up to very high redshifts
- SZE a very powerful probe of structure formation
- SZ flux is proportional to total thermal energy of the cluster and therefore should be good proxy for cluster mass

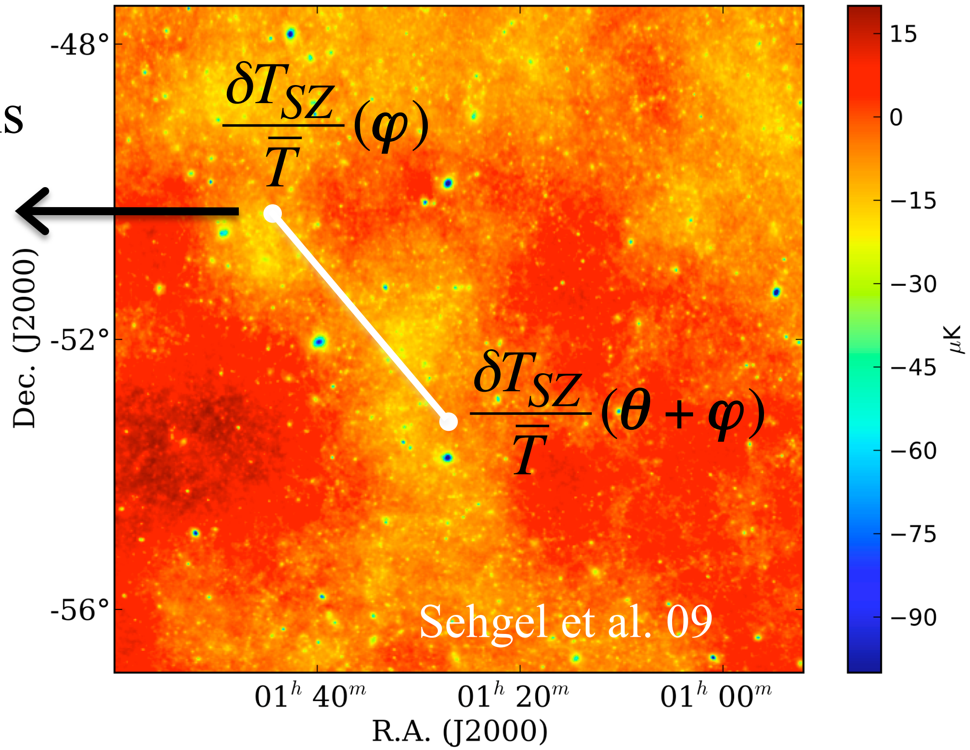
SZ power spectrum

- A two-point correlation of the SZ map is a powerful probe of cosmology: contains the contributions from unresolved small halos

$$\xi^{SZ}(\theta) = \left\langle \frac{\delta T}{T}(\varphi)\frac{\delta T}{T}(\varphi+\theta)\right\rangle \overset{\text{F.T.}}{\Leftrightarrow} C_l^{SZ}$$

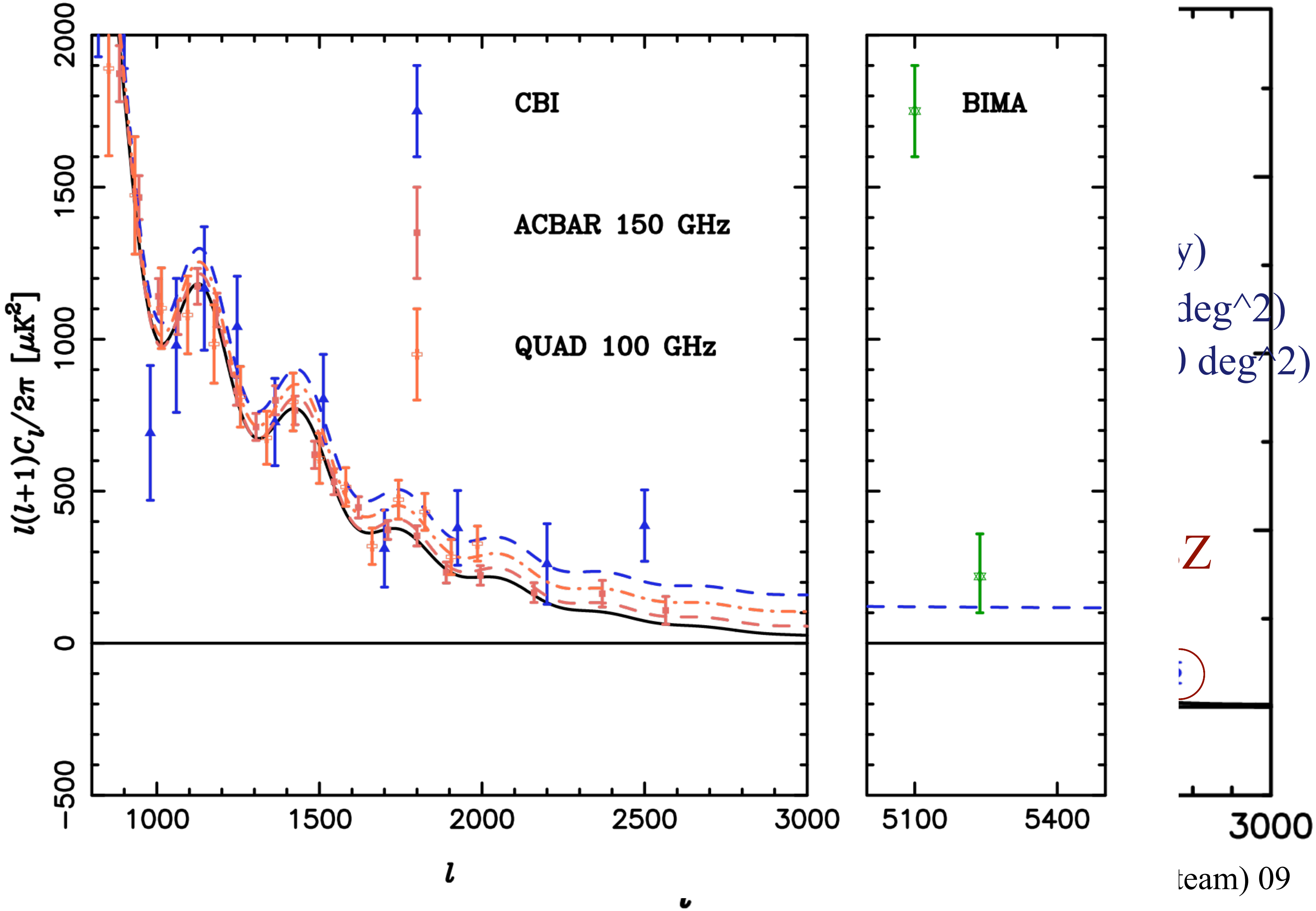


Simulated SZ map (10°×10°: tSZ+kSZ)

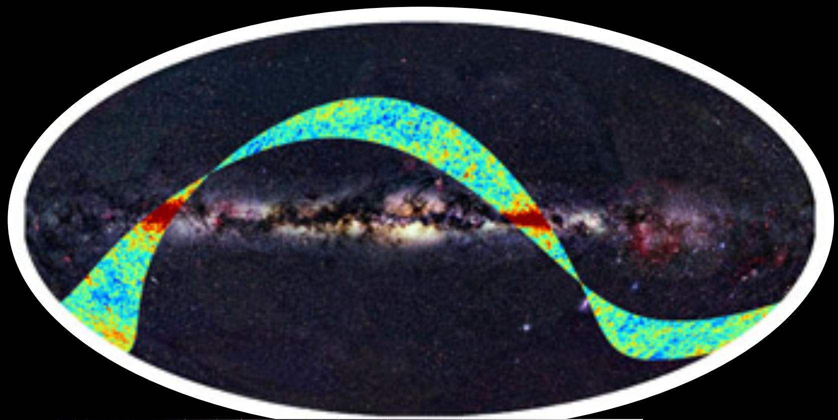


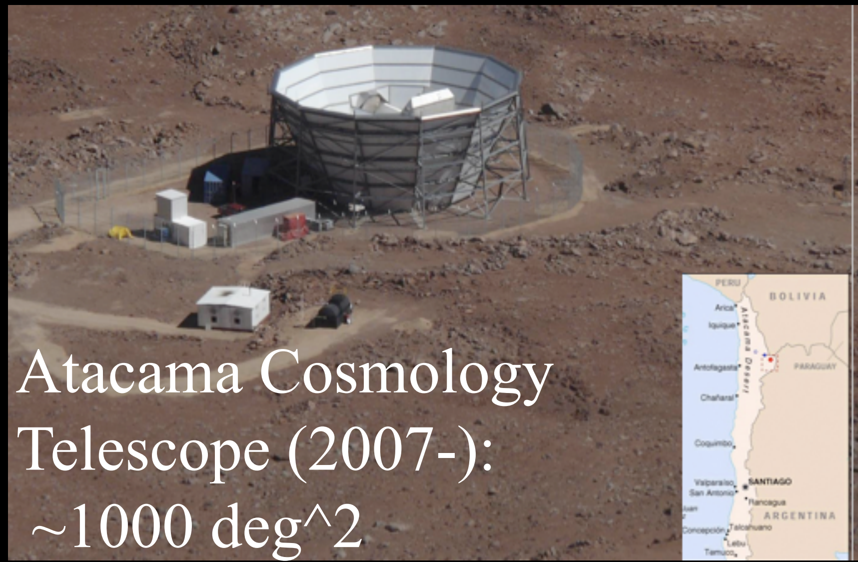
- The SZ power spectrum is dominant over the primary CMB anisotropies and other secondary effects at l>3000

SZ power spectrum present

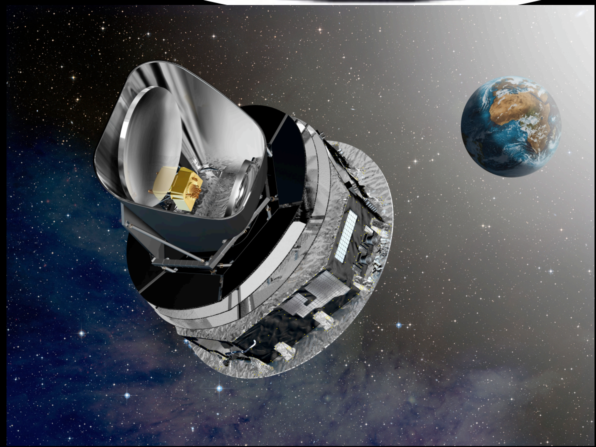


Prospect for SZ science

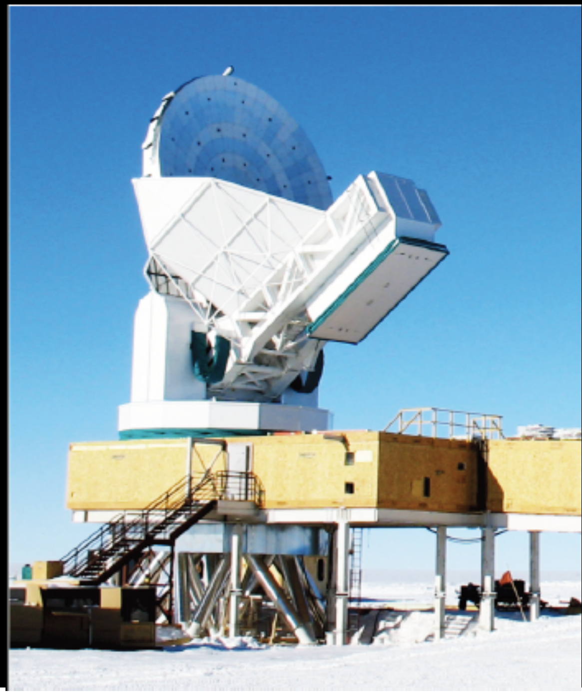




Atacama Cosmology Telescope (2007-): ~1000 deg^2



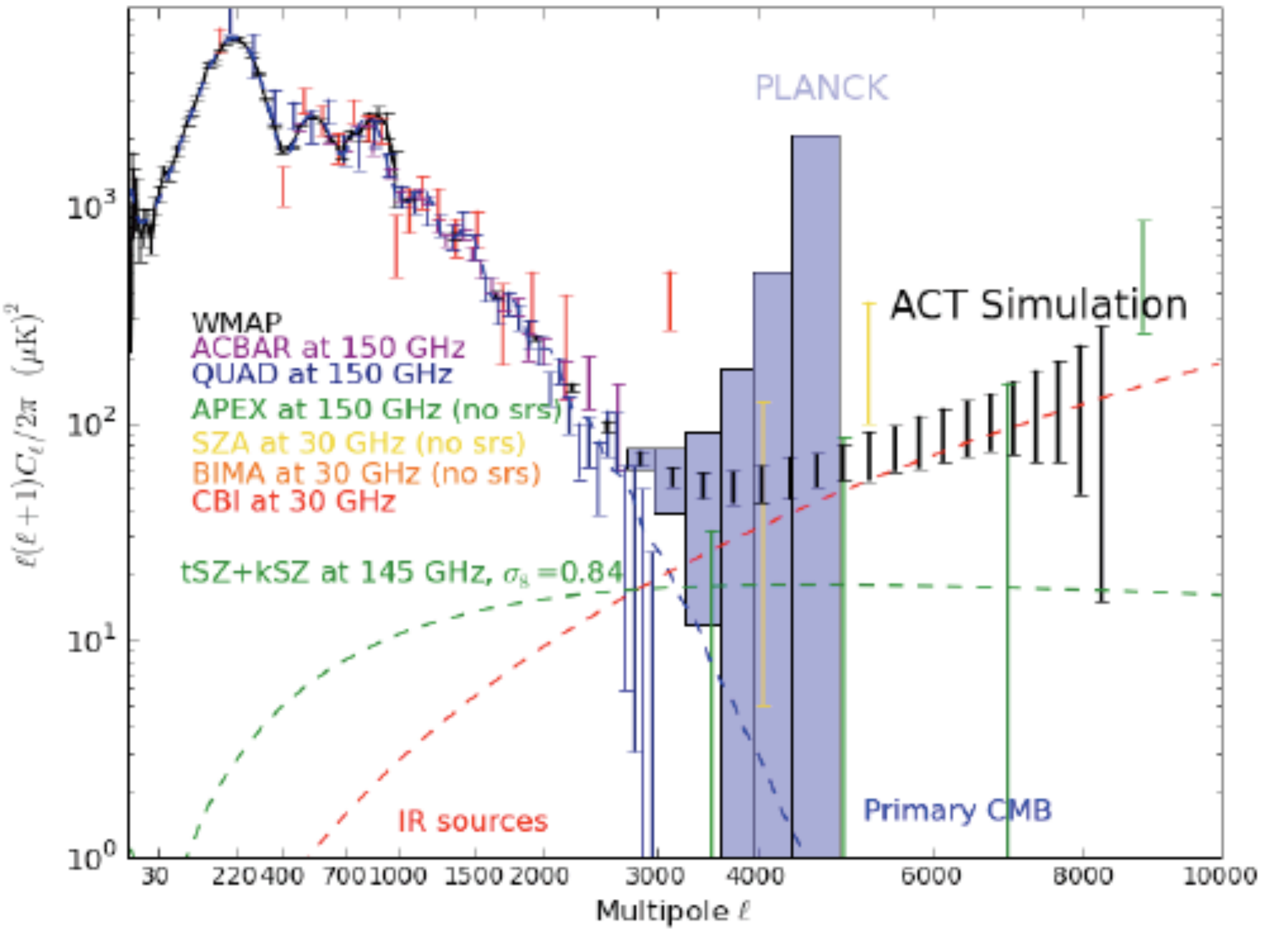
Planck satellite (2009-)



South Pole Telescope (07-): ~1000 deg^2 (so far)

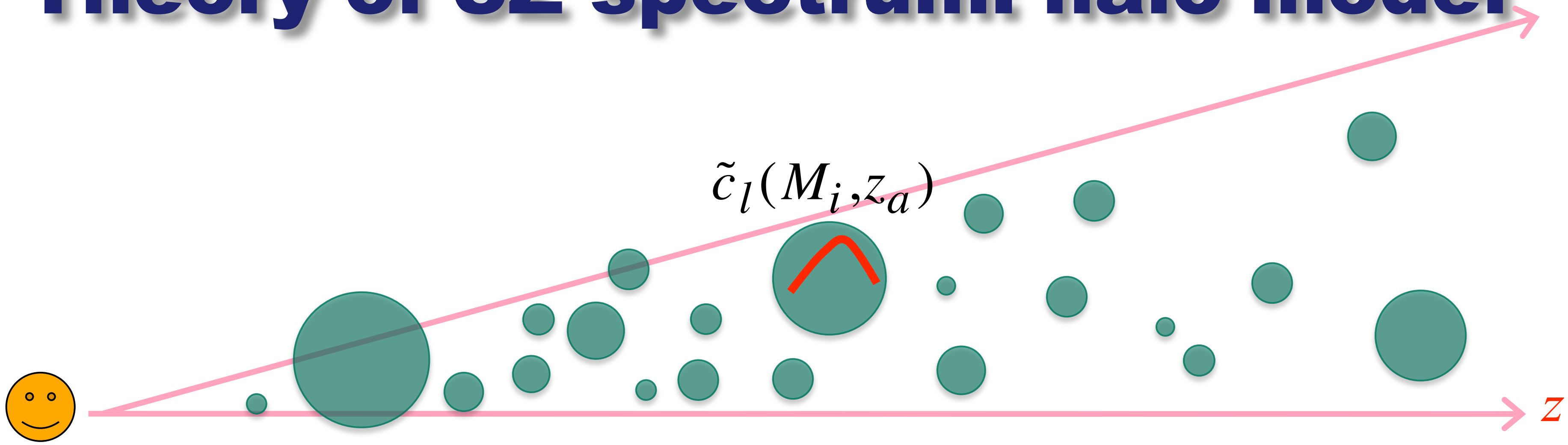
ACT 2008 power spectrum expectations

From Suzanne Staggs's talk @ IPMU conference, June 09



Points are the sums of the **models** shown, not data; error bars from preliminary maps.

Theory of SZ spectrum: halo model



- The SZ PS can be expressed by integrating each halo contributions over the light cone (Komatsu & Kitayama 99; Komatsu & Seljak 02)

$$C_l^{SZ} = \int dz \frac{d^2V}{dz d\Omega} \int dM \frac{dn}{dM} \tilde{c}_l(M,z)$$

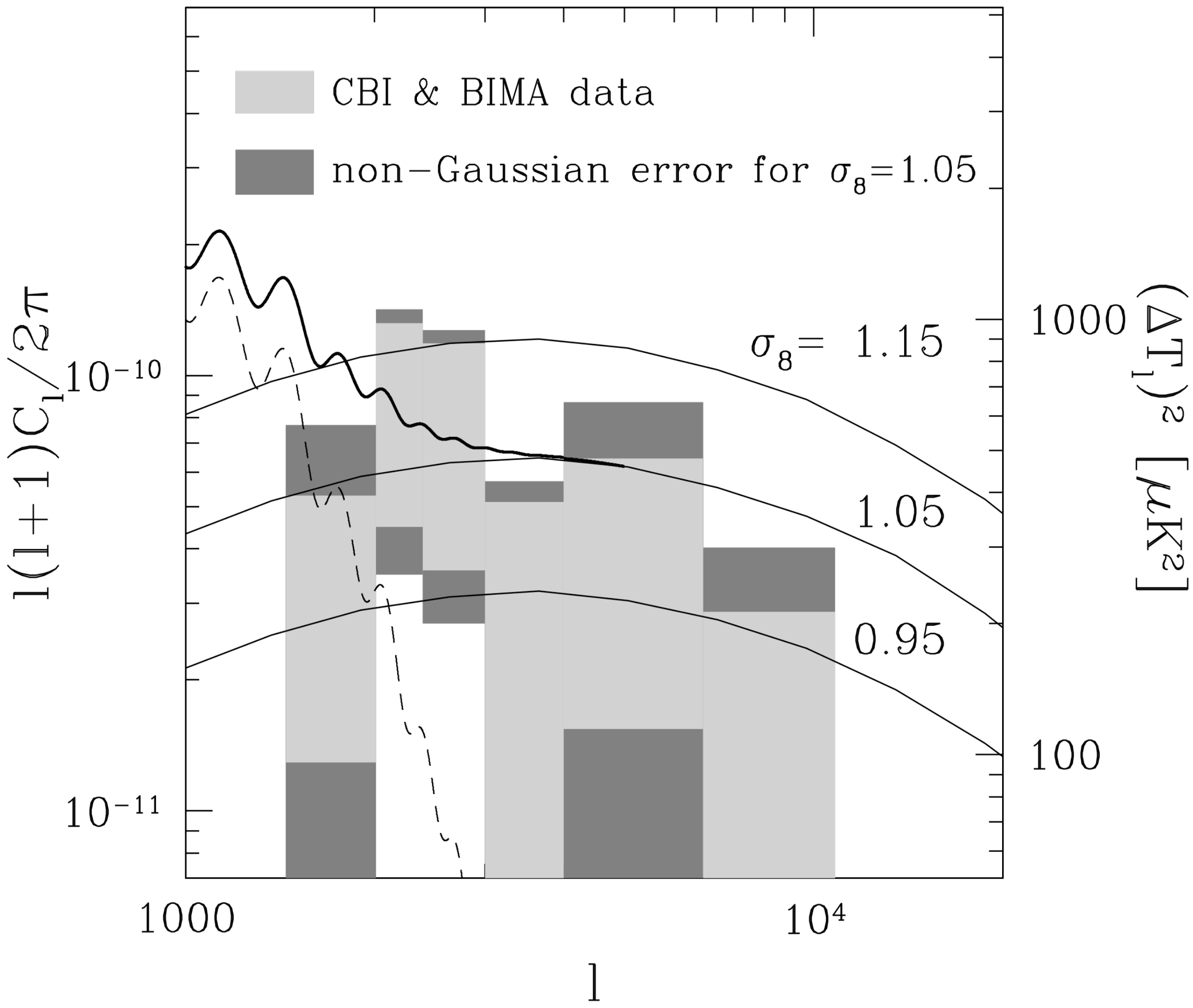
Comoving volume per unit redshift and unit steradian

Number density of halos with masses [M,M+dM] and at redshift [z, z+dz]

Angular ps of a halo with mass M and at redshift z

Cosmology

Cosmological sensitivity of SZ

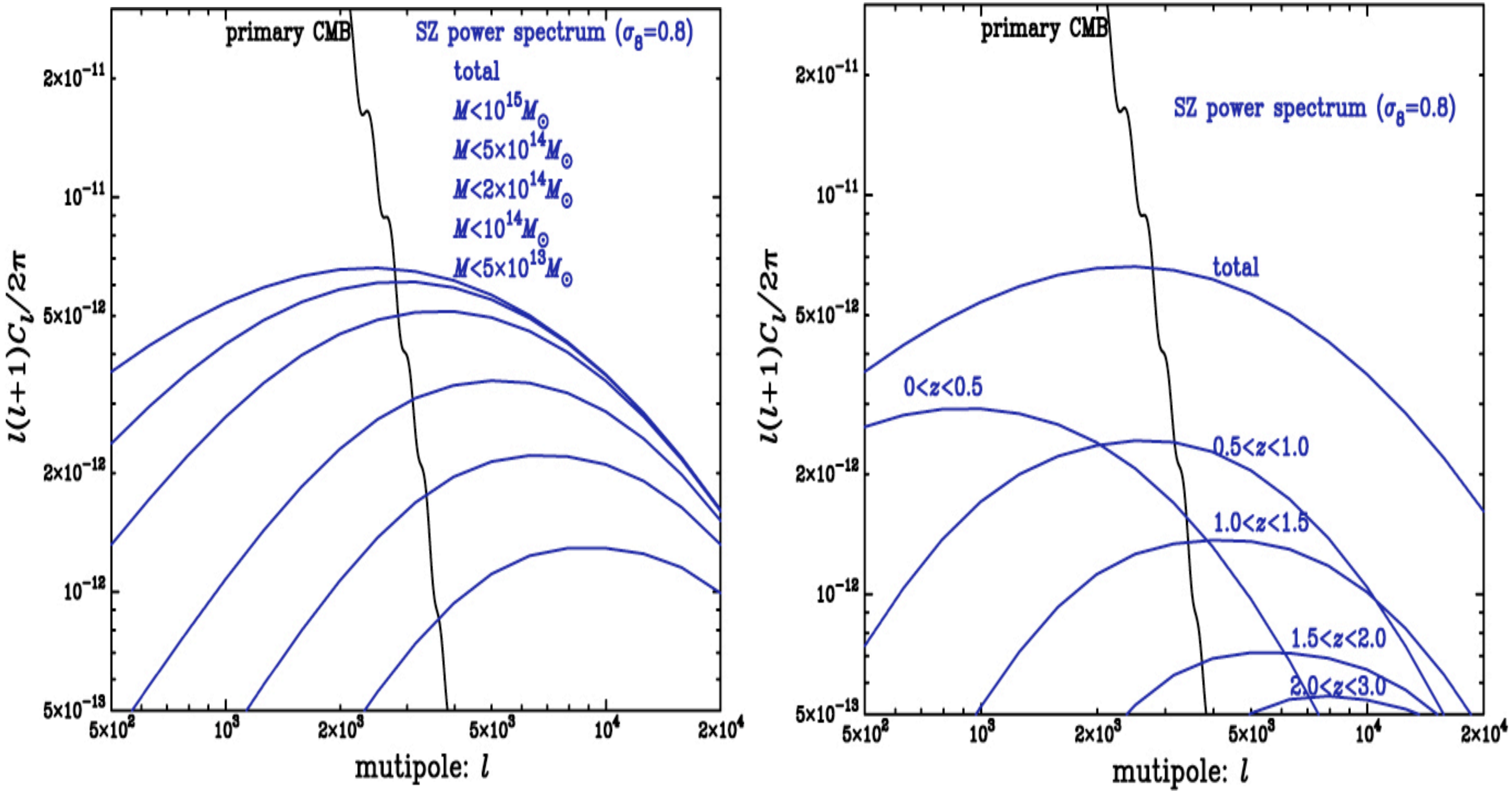


Komatsu & Seljak 02

- The halo model approach fairly well explains the measurements/simulations within ~10% accuracy (simulations needed for a more precise modeling)
- The SZ spectrum is very sensitive to cosmology, especially to one parameter σ_8 (the rms of present-day mass fluctuations within a sphere of 8Mpc/h)

$$C_l \propto \sigma_8^{7-8}$$

SZ vs. halo mass/redshift



Likelihood of SZ power spectrum

(MT & Yoshida in prep.)

- The halo modeled SZ PS can be rewritten as

$$C_l = \int dz \frac{d^2V}{dz d\Omega} \int dM \frac{dn}{dM} \tilde{c}_l(M,z)$$

$$\approx \sum_z \sum_M \frac{d^2V}{dz d\Omega} \Delta z \times \frac{dn}{dM} \Delta M \times \tilde{c}_l(M,z)$$

$$= \frac{1}{\Omega_s} \sum_z \sum_M \left[\Omega_s \times \frac{d^2V}{dz d\Omega} \Delta z \times \frac{dn}{dM} \Delta M \right] \times \tilde{c}_l(M,z) \qquad (\Omega_s : \text{ Survey Area})$$

$$C_l = \frac{1}{\Omega_s} \sum_z \sum_M N(M,z) \times \tilde{c}_l(M,z)$$

The total number of halos with mass M and at redshift z (not number density)

The SZ power spectrum of a halo with mass M and at redshift z (not number density)

SZ likelihood (contd.)

$$C_l = \frac{1}{\Omega_s} \sum_z \sum_M N(M,z) \times \tilde{c}_l(M,z)$$

- The statistical variables, N and c_l, can be safely considered independent

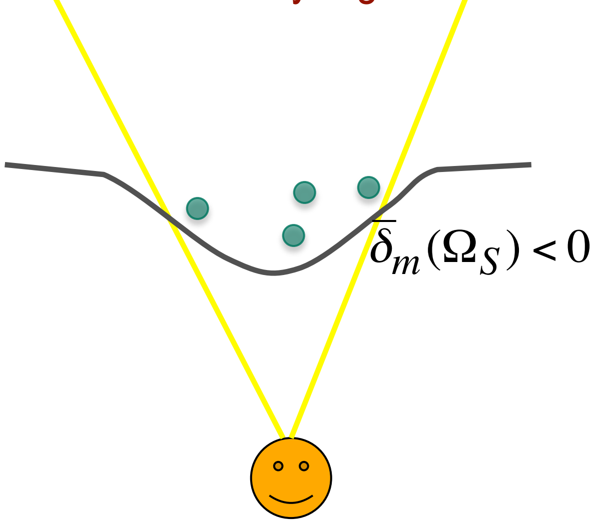
$$P\left[C_l(M,z) \mid \boldsymbol{p}\right] \propto P[N(M,z)] \times P[\tilde{c}_l(M,z)]$$

N: Poisson distribution, what is the mean? (see later)
c_l: assume the Gaussian distribution for high *l* limit (not important)

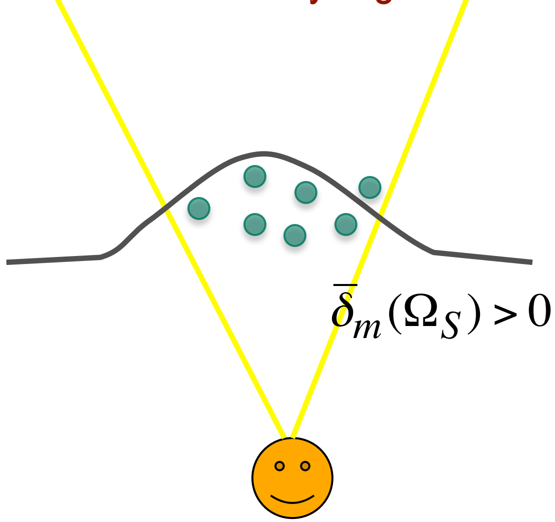
Halo sample variance:
finite survey volume effect

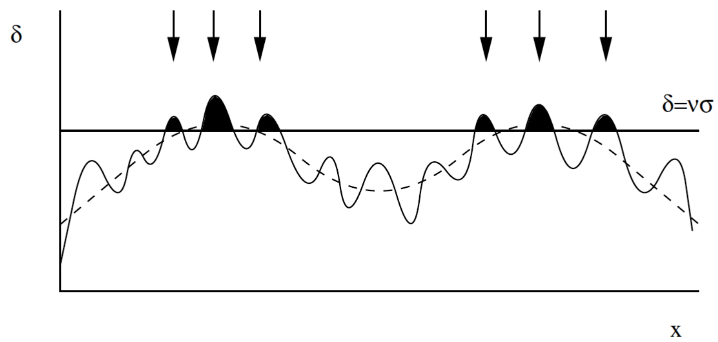
- What is the pdf of halo number for a given survey area?
- The halo number depends on the large-scale density fluctuation of a survey area size via "halo bias"

If survey region is in the large-scale underdensity region:



If survey region in the large-scale overdensity region:





Halo bias concept: a coupling btw large-scale fluctuation and halo distribution/formation (Mo & White 96)

The expected mean of halo number for a given survey area would be

$$\langle N \rangle = \Omega_S \frac{d^2V}{dz\,d\Omega} \Delta z \frac{dn}{dM} \Delta M \left[1 + b(M)\bar{\delta}_m(\Omega_S)\right]$$

PDF of halo number

$$\langle N \rangle = \Omega_S \frac{d^2V}{dz d\Omega} \Delta z \frac{dn}{dM} \Delta M \left[1 + b(M) \bar{\delta}_m (\Omega_S) \right]$$

cosmology

Statistical variables

- Assuming that the large-scale density fluctuations of a survey size obey the Gaussian distribution (or in the linear regime), the PDF of halo number can be given as (also see Hu & Cohn 06)

$$P\left[N(M,z) \mid \boldsymbol{p}\right] = \int d\bar{\delta}_m \frac{1}{\sqrt{2\pi\sigma_m^2(\Omega_S)}} \exp\left[-\frac{\bar{\delta}_m}{2\sigma_m^2(\Omega_S)}\right] \times \frac{\langle N \rangle^N}{N!} e^{-\langle N \rangle}$$

SZ likelihood (contd.)

$$C_l = \frac{1}{\Omega_s} \sum_z \sum_M N(M,z) \times \tilde{c}_l(M,z)$$

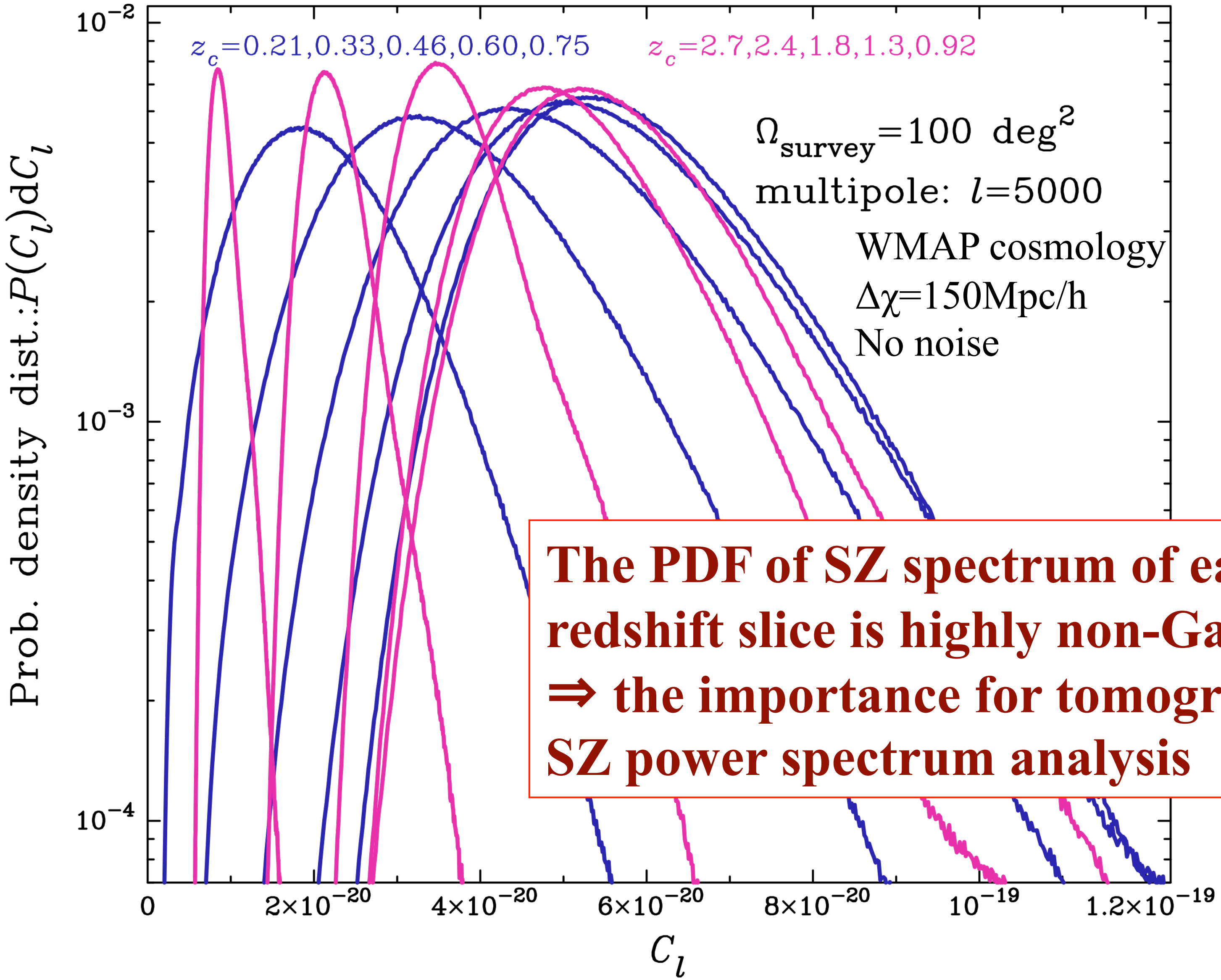
- The statistical variables, N and c_l, can be safely considered independent

$$P\left[C_l(M,z) \mid \boldsymbol{p}\right] \propto P[N(M,z)] \times P[\tilde{c}_l(M,z)]$$

N: Poisson distribution, what is the mean? (see later)
c_l: assume the Gassian distribution for high l limit (not important)

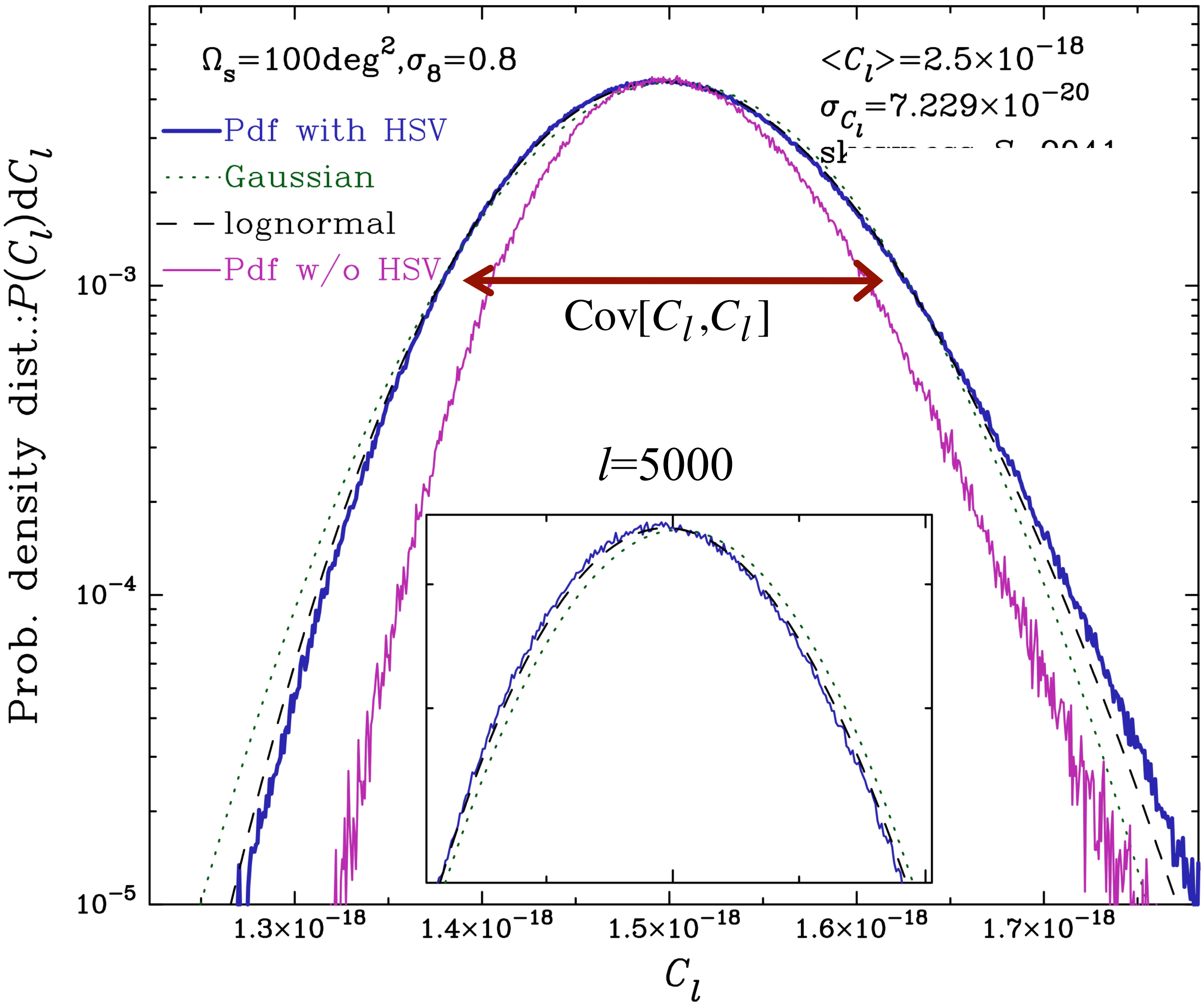
Monte-Carlo method: SZ likelihood

- PDF of SZ power spectra of each redshift slice

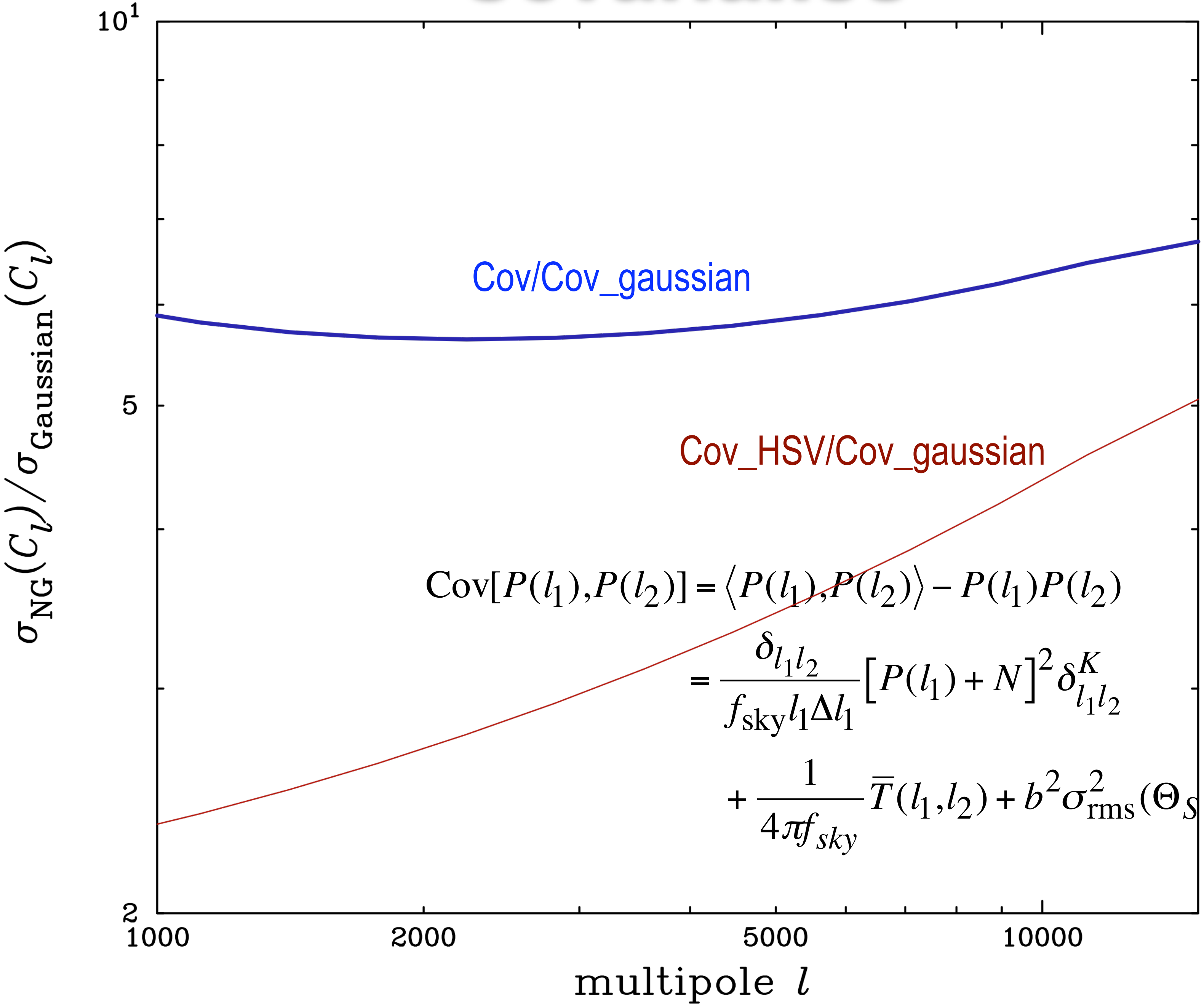


The PDF of SZ spectrum of each redshift slice is highly non-Gaussian ⇒ the importance for tomographic SZ power spectrum analysis

After projection, the PS PDF becomes closer to Gaussian



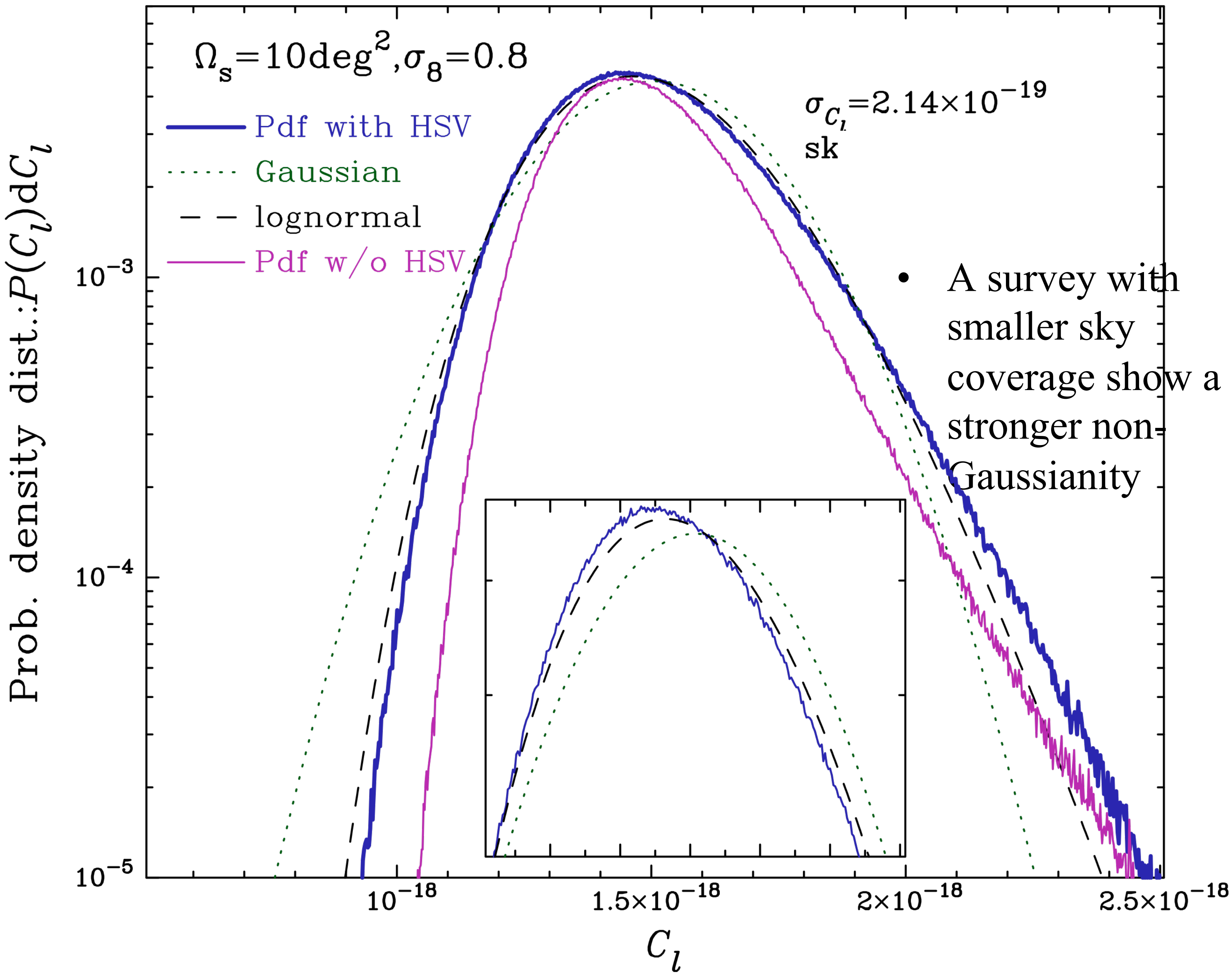
Covariance



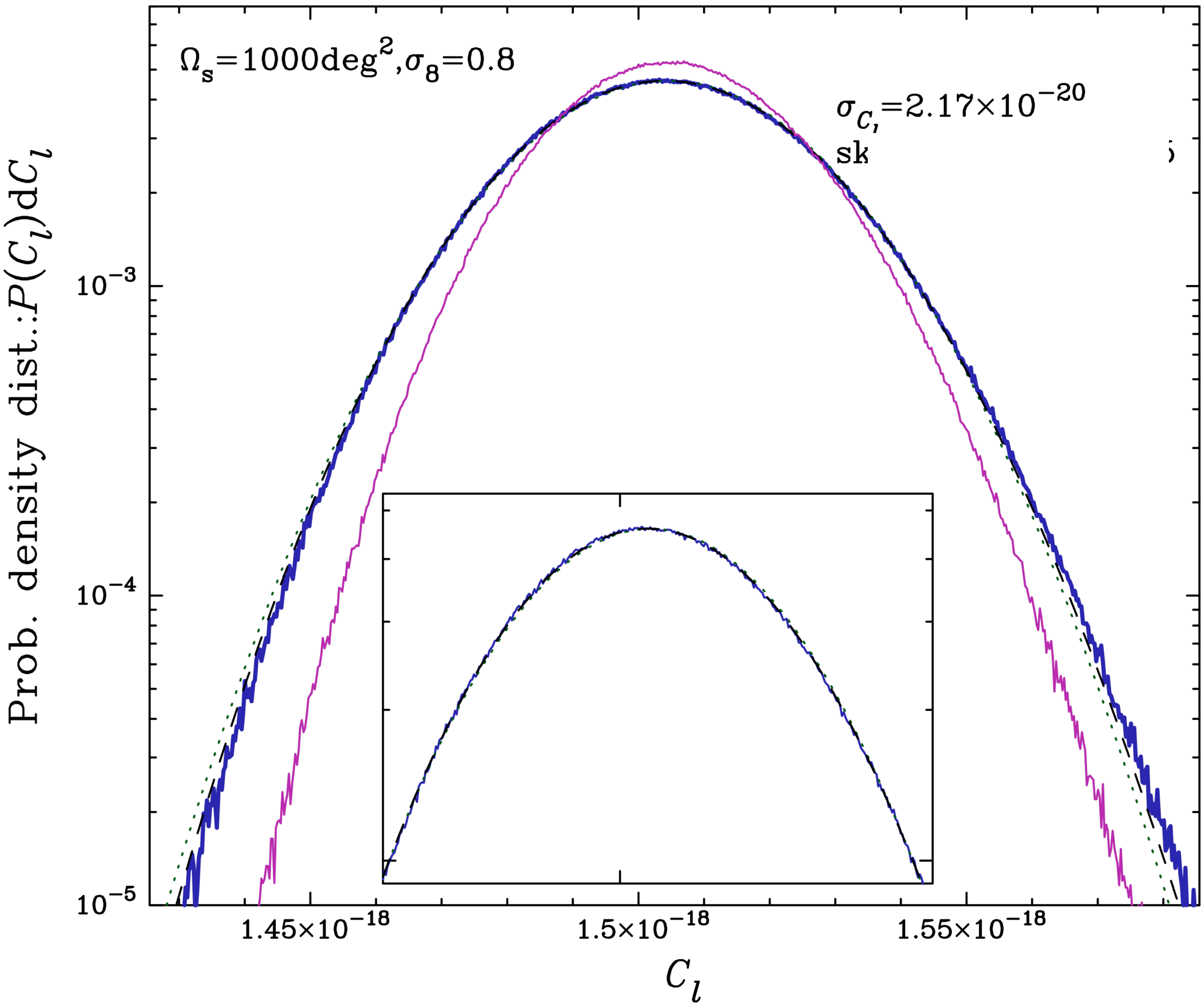
$$\mathrm{Cov}[P(l_1),P(l_2)] = \langle P(l_1),P(l_2)\rangle - P(l_1)P(l_2)$$

$$= \frac{\delta_{l_1 l_2}}{f_{\mathrm{sky}} l_1 \Delta l_1}\left[P(l_1)+N\right]^2 \delta^K_{l_1 l_2}$$

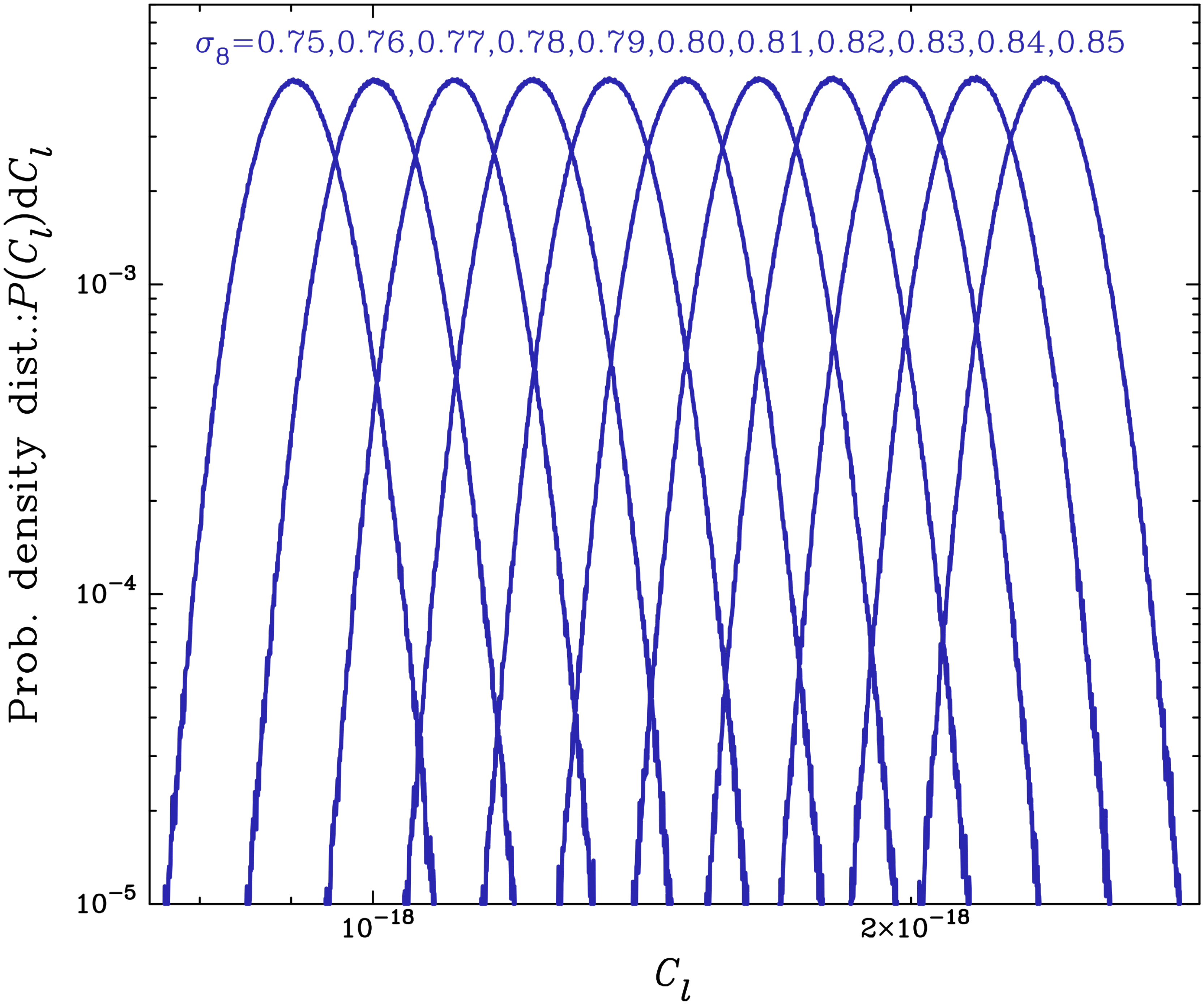
$$+ \frac{1}{4\pi f_{sky}} \bar{T}(l_1,l_2) + b^2 \sigma^2_{\mathrm{rms}}(\Theta_S) P(l_1) P(l_2)$$



- A survey with smaller sky coverage show a stronger non-Gaussianity



Ω_s=1000deg^2,σ_8=0.8
$\sigma_{C,}$=2.17×10^{-20}
sk
Prob. density dist.:$P(C_l)\mathrm{d}C_l$
10^{-3}
10^{-4}
10^{-5}
1.45×10^{-18}
1.5×10^{-18}
1.55×10^{-18}
C_l



σ8=0.75,0.76,0.77,0.78,0.79,0.80,0.81,0.82,0.83,0.84,0.85
Prob. density dist.:P(Cl)dCl
10−3
10−4
10−5
10−18
2×10−18
Cl

Comparing the parameter estimation methods

- Gaussian likelihood analysis

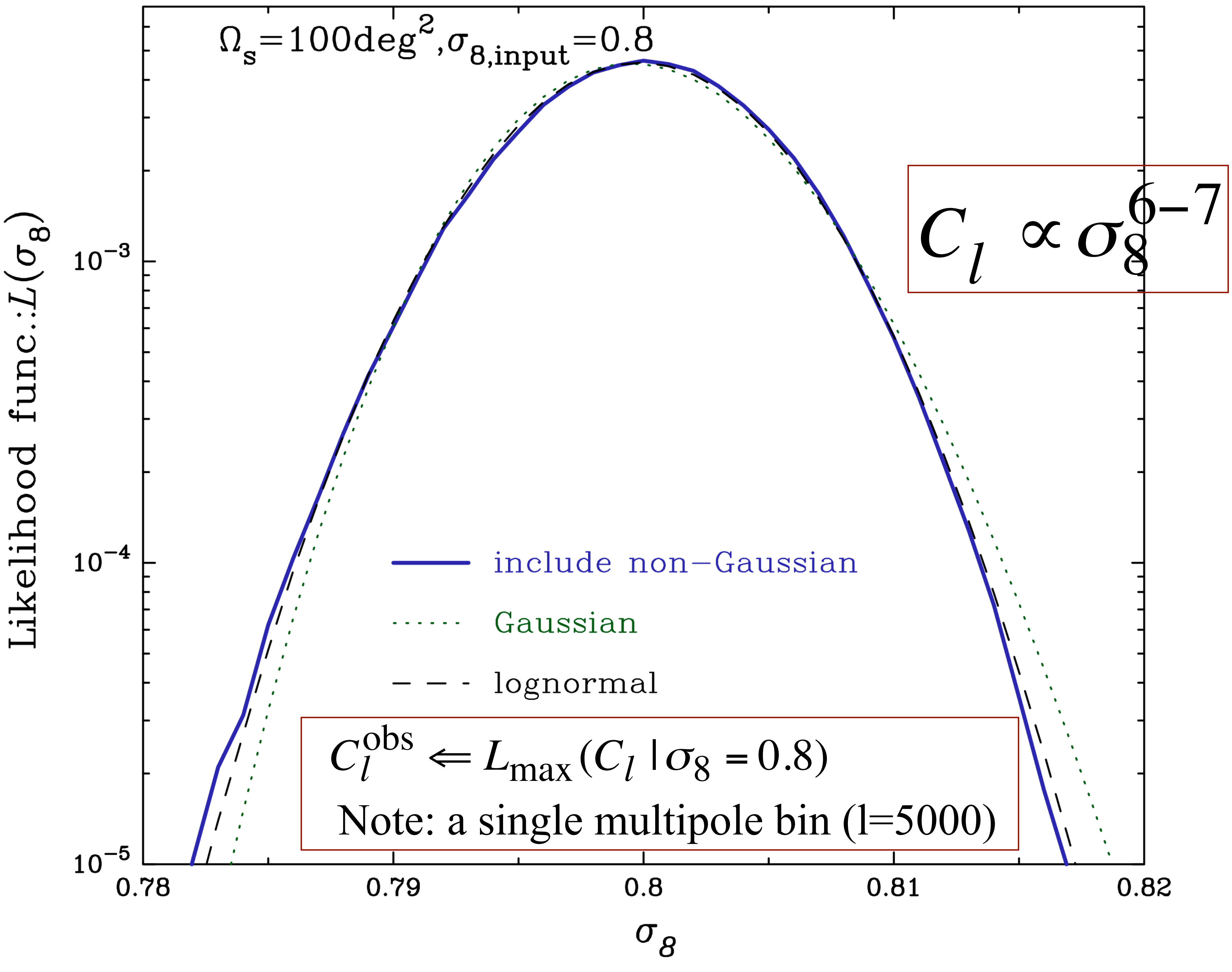
$$-2\ln L_G \propto \left(C_l^{\mathrm{obs}} - C_l^{\mathrm{model}}\right)\left[\mathrm{Cov}(C_l^{\mathrm{m}}, C_{l'}^{\mathrm{m}})\right]^{-1}\left(C_{l'}^{\mathrm{obs}} - C_{l'}^{\mathrm{model}}\right)$$

- Log-normal likelihood analysis

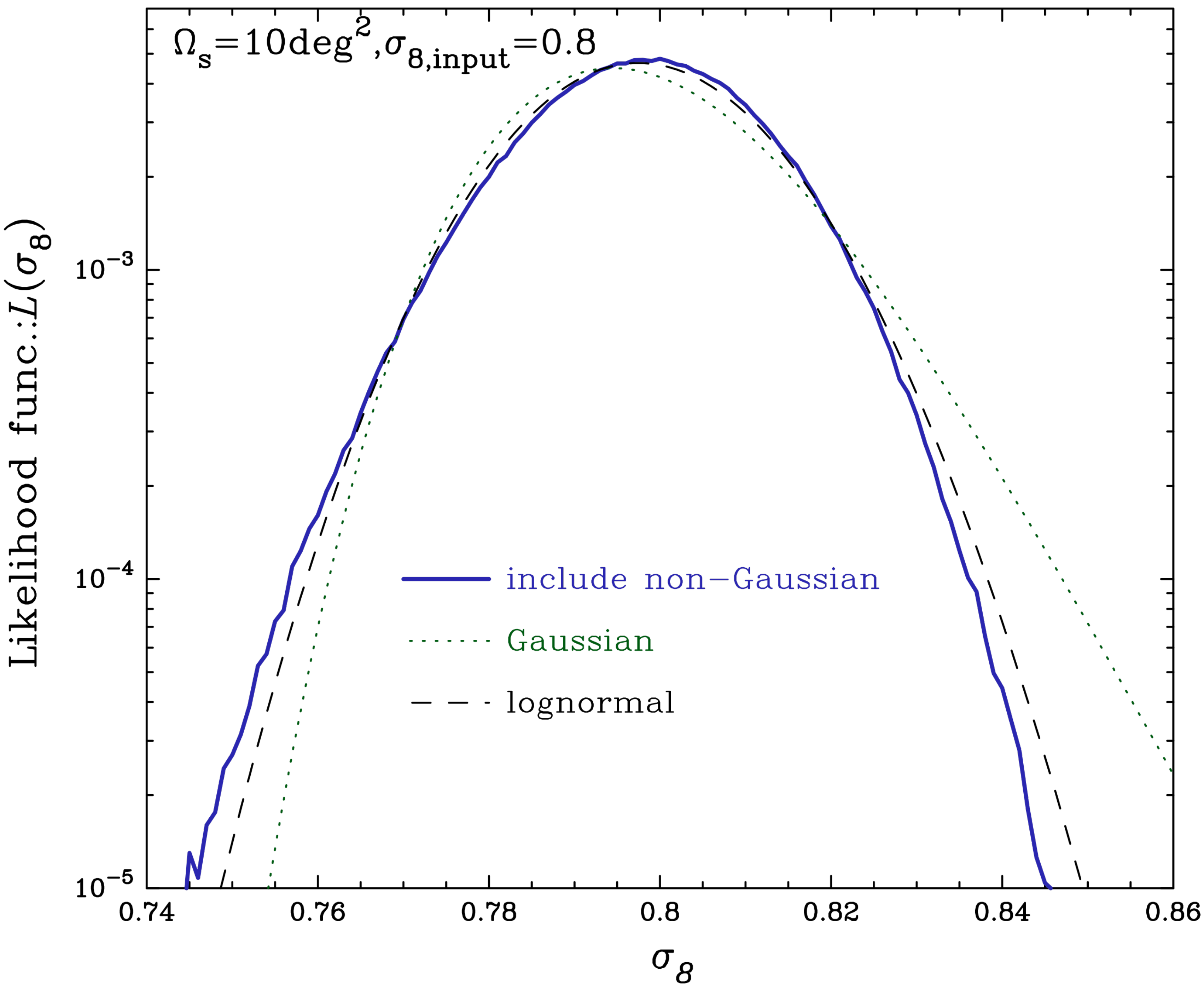
$$-2\ln L_{\mathrm{LG}} \propto \ln\left(C_l^{\mathrm{obs}} / C_l^{\mathrm{model}}\right)\left[\mathrm{Cov}(\ln C_l^{\mathrm{m}}, \ln C_{l'}^{\mathrm{m}})\right]^{-1} \ln\left(C_{l'}^{\mathrm{obs}} / C_{l'}^{\mathrm{model}}\right)$$

- Likelihood analysis

$$P(\boldsymbol{p} \mid C_l^{\mathrm{obs}}) = L[C_l^{\mathrm{obs}} \mid C_l^{\mathrm{model}}(\boldsymbol{p})]P(\boldsymbol{p})$$



$\Omega_s=100\mathrm{deg}^2, \sigma_{8,\mathrm{input}}=0.8$
$C_l \propto \sigma_8^{6-7}$
include non−Gaussian
Gaussian
lognormal
$C_l^{\mathrm{obs}} \Leftarrow L_{\max}(C_l \mid \sigma_8 = 0.8)$
Note: a single multipole bin (l=5000)
Likelihood func.: $L(\sigma_8)$
σ_8
10^{-3}
10^{-4}
10^{-5}
0.78
0.79
0.8
0.81
0.82



Ω_s=10deg^2,$\sigma_{8,\rm input}$=0.8
Likelihood func.:$L(\sigma_8)$
10^{-3}
10^{-4}
10^{-5}
0.74
0.76
0.78
0.8
0.82
0.84
0.86
σ_8
include non−Gaussian
Gaussian
lognormal

Comparison Table

Ω_S	LF	LF w/o HSV	Gaussian	LogG
10deg^2	$0.80^{+0.0049}_{-0.0082}$	$0.80^{+0.0051}_{-0.0058}$	$0.795^{+0.0095}_{-0.0048}$	$0.797^{+0.0076}_{-0.0059}$
100deg^2	$0.80^{+0.0024}_{-0.0024}$	$0.80^{+0.0017}_{-0.0020}$	$0.799^{+0.0032}_{-0.0016}$	$0.80^{+0.0017}_{-0.0020}$

Summary and Conclusion

- Non-Gaussian errors are significant in large-scale structure
  - The marginarized errors on parameter may not be that asymmetric for 2D observables, after projection
  - May be significant for 3D observables at small distance scales
- Hybrid method needed: simulations to have predictions for the mean observable, semi-analytic methods for the scaling of covariance on parameters
- Need more work with statistics + cosmology