XXVI Winter School of Astrophysics
November 3-14th 2014

Bayesian Cosmology

Roberto Trotta
Imperial Centre for Inference and Cosmology
Imperial College London
www.robertotrotta.com

ICIC Imperial Centre for Inference & Cosmology

Imperial College London

Why do we need statistics?

Imperial College London

"If you need statistics, you ought to have done a better experiment"

Attributed to Rutherford

- Increasingly complex models and data: "chi-square by eye" simply not enough
- "If it's real, better data will show it":
  but all the action is in the "discovery zone" around 3-4 sigma significance. This is a moving target.
- Don't waste time explaining effects which are not there
- Plan for the future: which is the best strategy? (survey design & optimization)
- In some cases, there will be no better data! (cosmic variance)

The future of cosmology is in Astrostatistics

Imperial College London

From a data-starved to a data-choked discipline!

Hubble (1929)

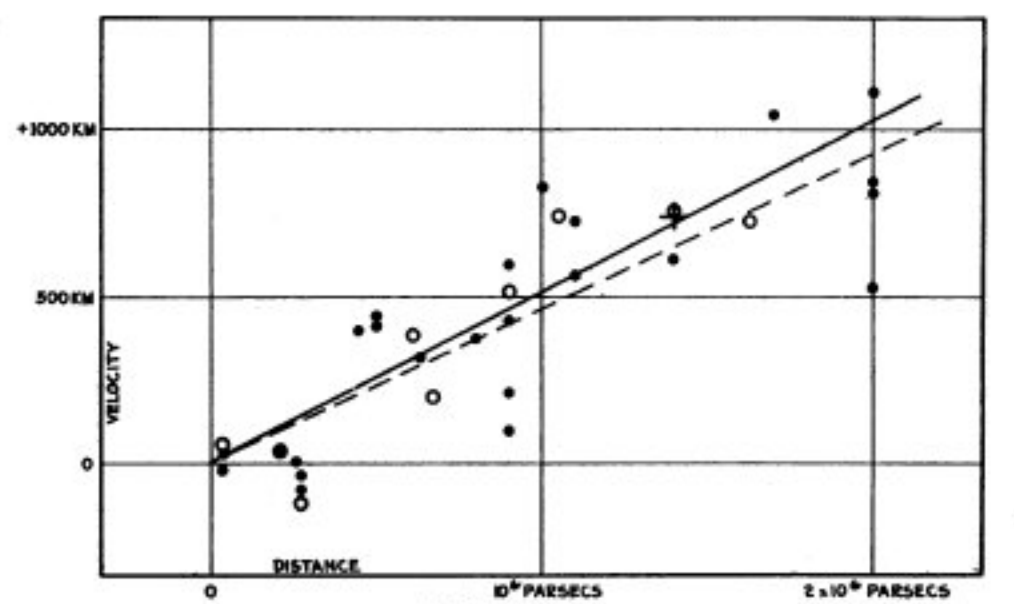


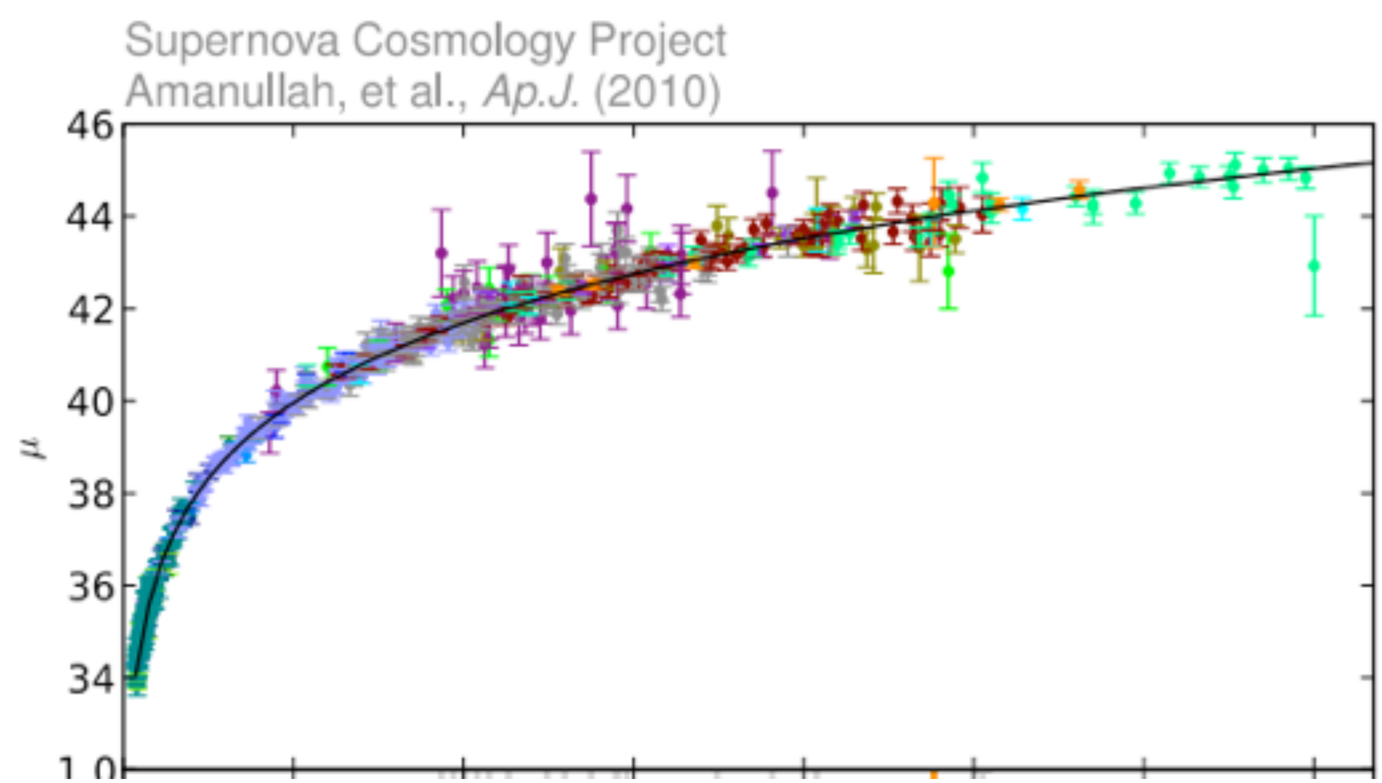
FIGURE 1

Velocity-Distance Relation among Extra-Galactic Nebulae.

Radial velocities, corrected for solar motion, are plotted against distances estimated from involved stars and mean luminosities of nebulae in a cluster. The black discs and full line represent the solution for solar motion using the nebulae individually; the circles and broken line represent the solution combining the nebulae into groups; the cross represents the mean velocity corresponding to the mean distance of 22 nebulae whose distances could not be estimated individually.

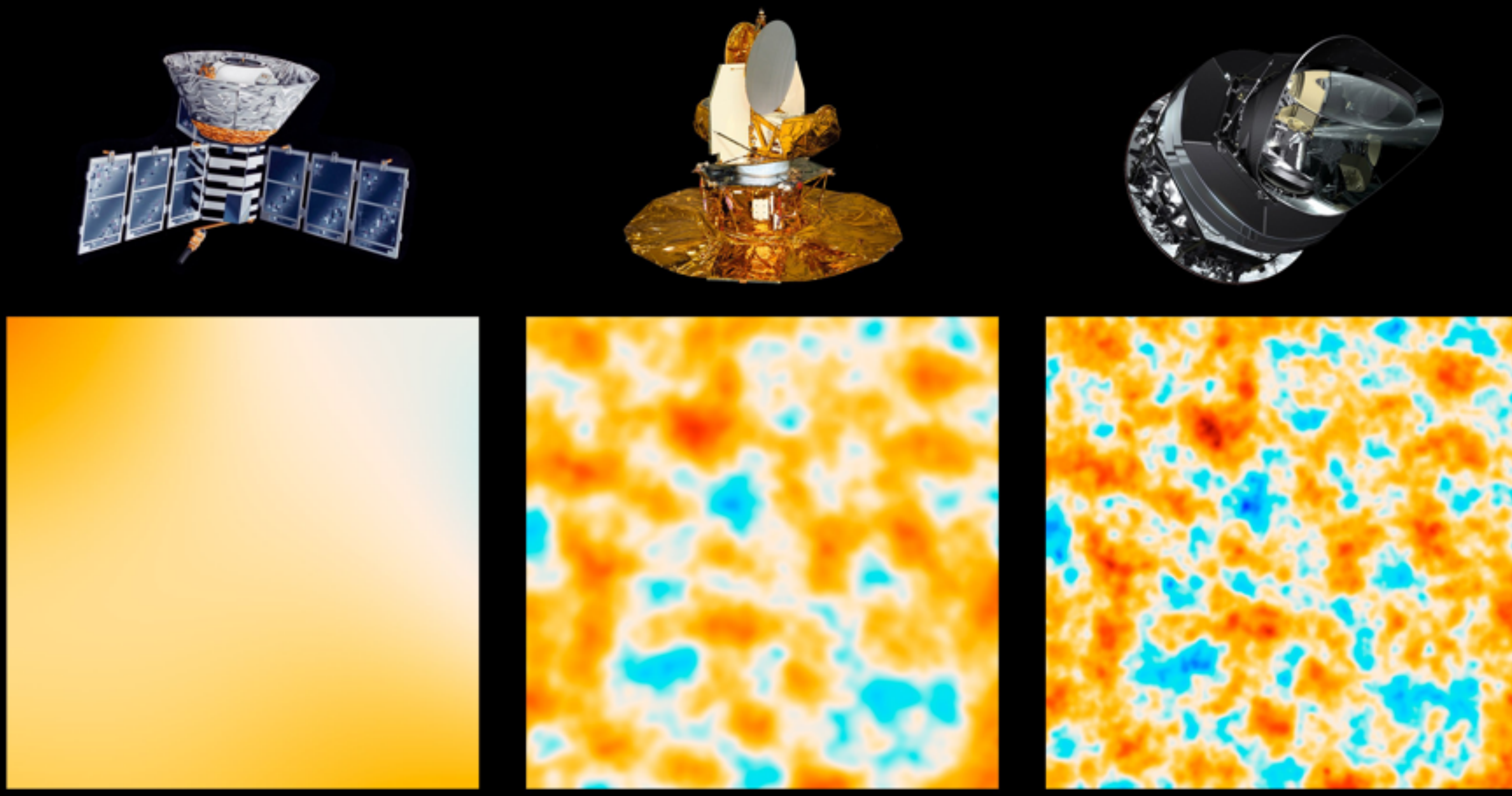
Source: Edwin Hubble, PNAS

"Union 2" compilation (2010)



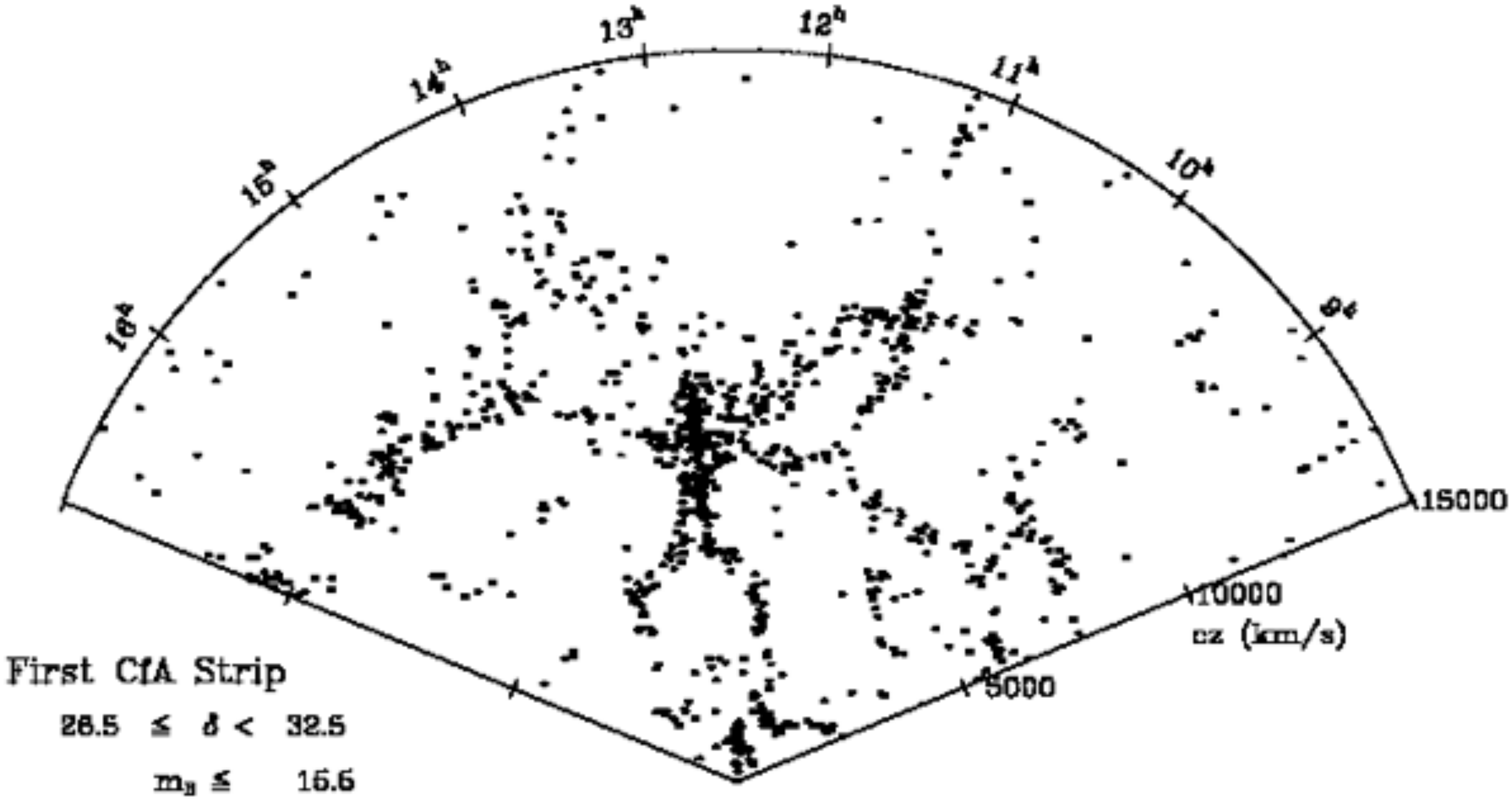
ICIC

1994 2001-2010 2013-2014

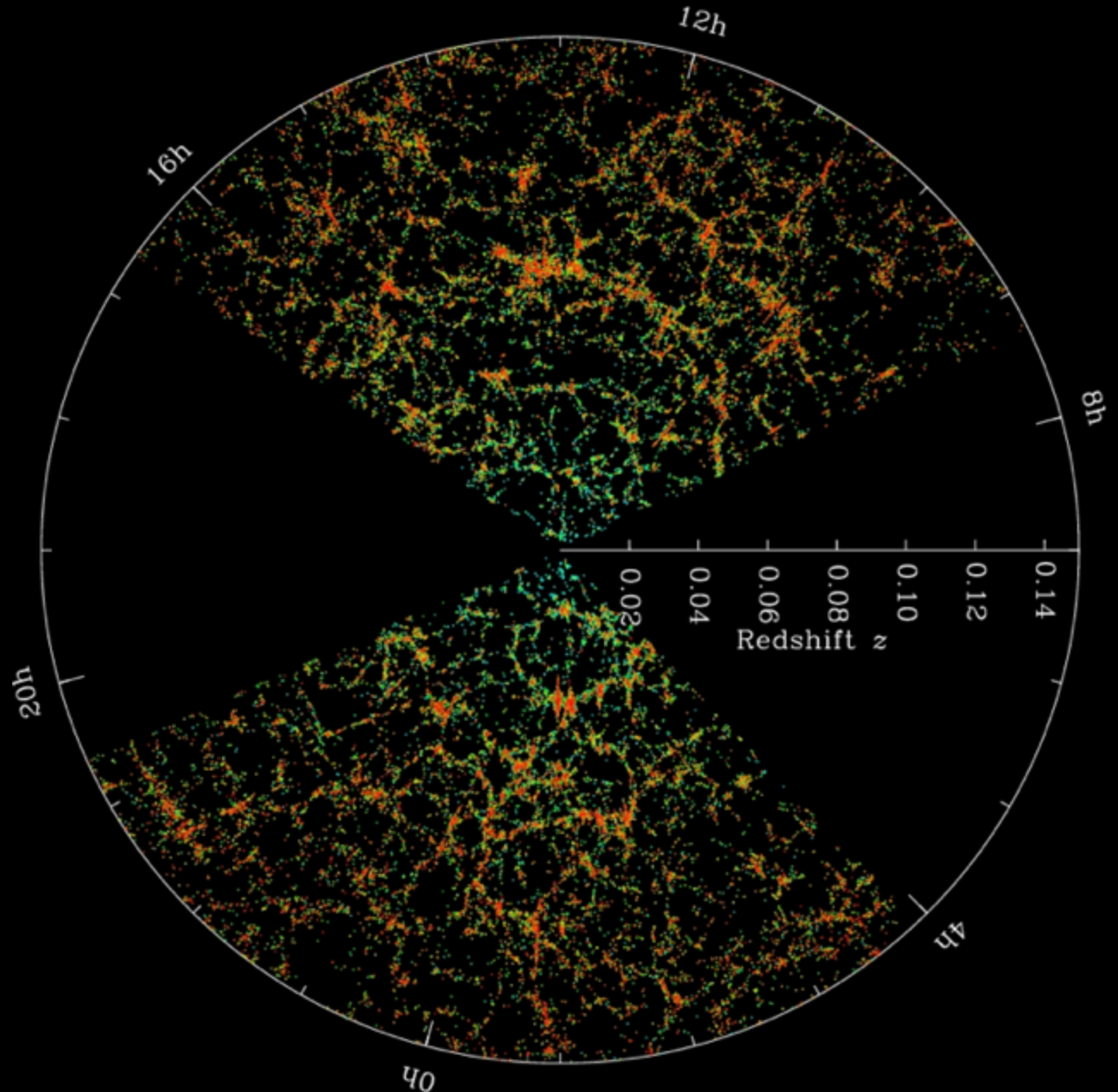


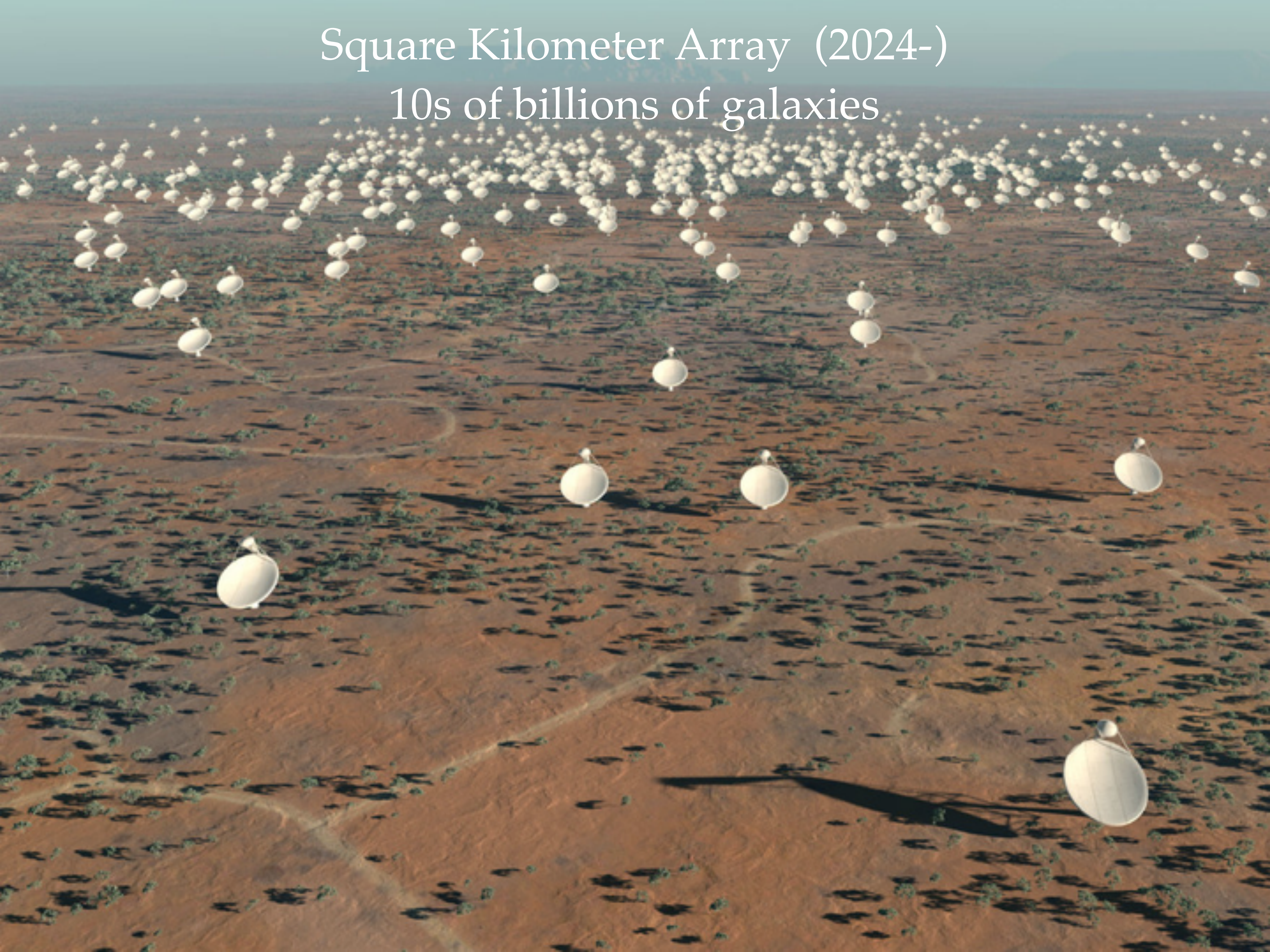
COBE WMAP Planck

CfA redshift survey (1985)
1100 galaxies



Sloan Digital Sky Survey (2000-2008)
1M galaxies




Square Kilometer Array (2024-)
10s of billions of galaxies

The cosmological concordance model

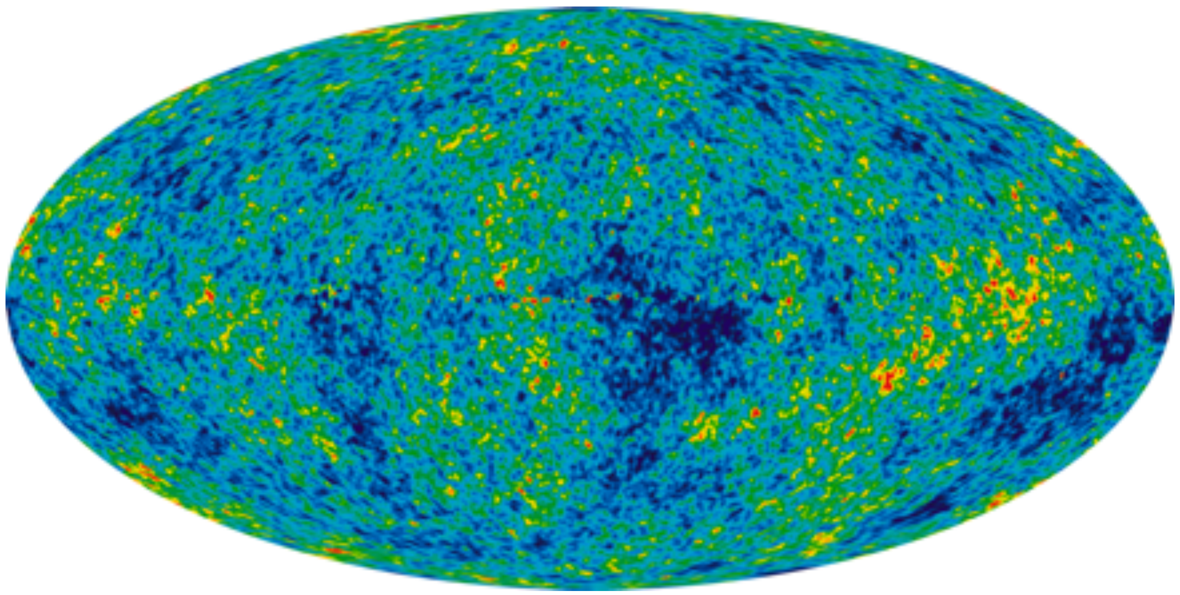
The ΛCDM cosmological concordance model is built on three pillars:

1. **INFLATION:**
   A burst of exponential expansion in the first ~10^{-32} s after the Big Bang, probably powered by a yet unknown scalar field.
2. **DARK MATTER:**
   The growth of structure in the Universe and the observed gravitational effects require a massive, neutral, non-baryonic yet unknown particle making up ~25% of the energy density.
3. **DARK ENERGY:**
   The accelerated cosmic expansion (together with the flat Universe implied by the Cosmic Microwave Background) requires a smooth yet unknown field with negative equation of state, making up ~70% of the energy density.

The next 5 to 10 years are poised to bring major observational breakthroughs in each of those topics!

Cosmic Microwave Background analysis

Imperial College London



WMAP7 internal linear combination map

The observed anisotropies are a superposition of:

1. Initial conditions (inflation/early Universe physics)
2. Temperature/potential fluctuations at decoupling
3. Line-of-sight effects (ISW, SZ, lensing)

Temperature fluctuations:

$$\frac{\delta T}{T}(\vec{n}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\vec{n})$$

2-point correlation function

$$\xi(\theta) = \langle \frac{\delta T}{T}(\vec{n}) \frac{\delta T}{T}(\vec{n}') \rangle$$

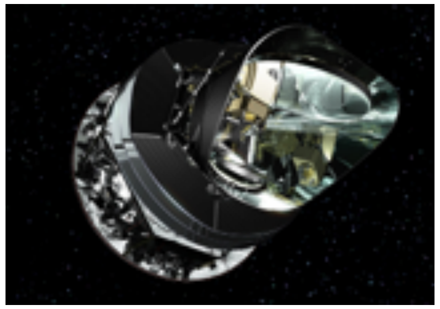
$$= \frac{1}{4\pi} \sum_{\ell} (2\ell + 1) C_\ell P_\ell(\vec{n} \cdot \vec{n}')$$

Angular power spectrum (assumes isotropy)

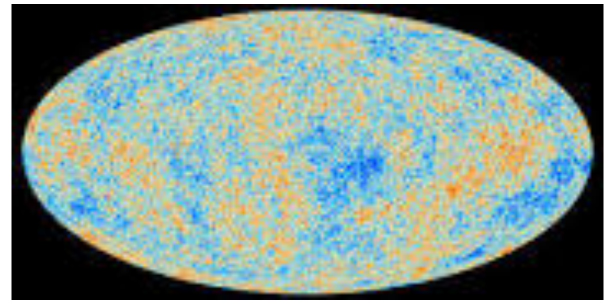
$$C_\ell = \langle |a_{\ell m}|^2 \rangle$$

The power spectrum contains the full statistical information
IF fluctuations are Gaussian

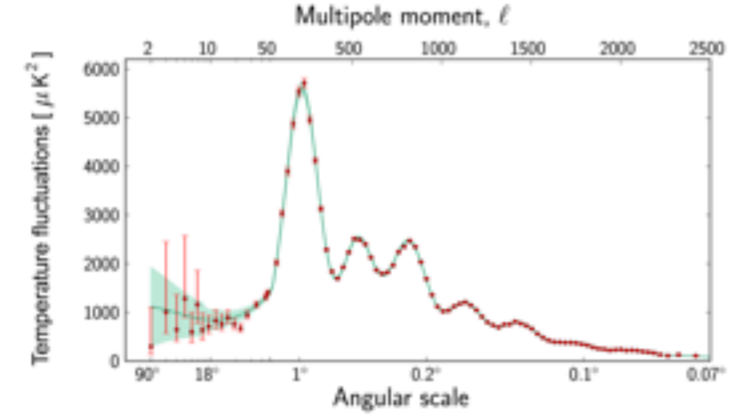
ICIC



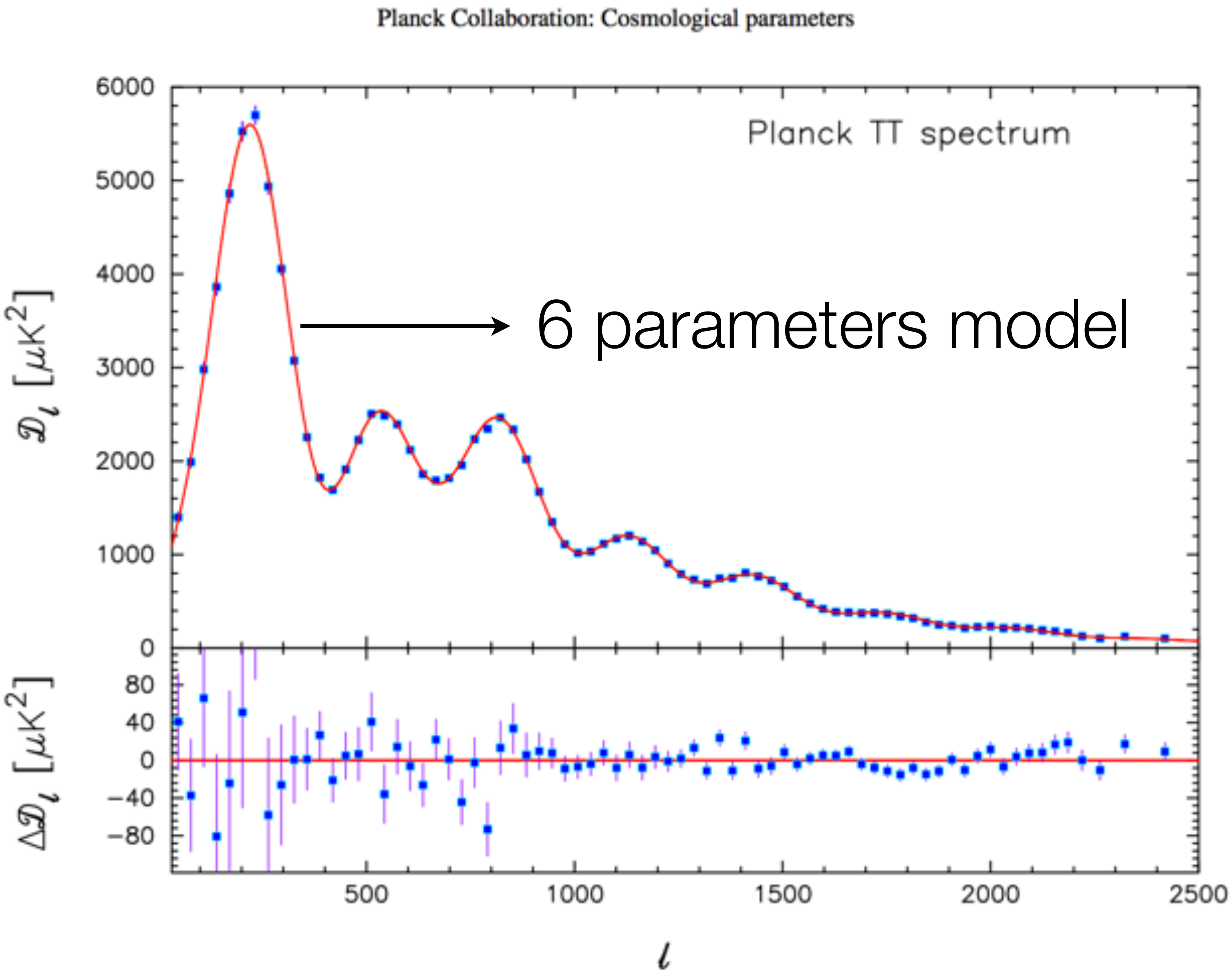
10^{12} bits

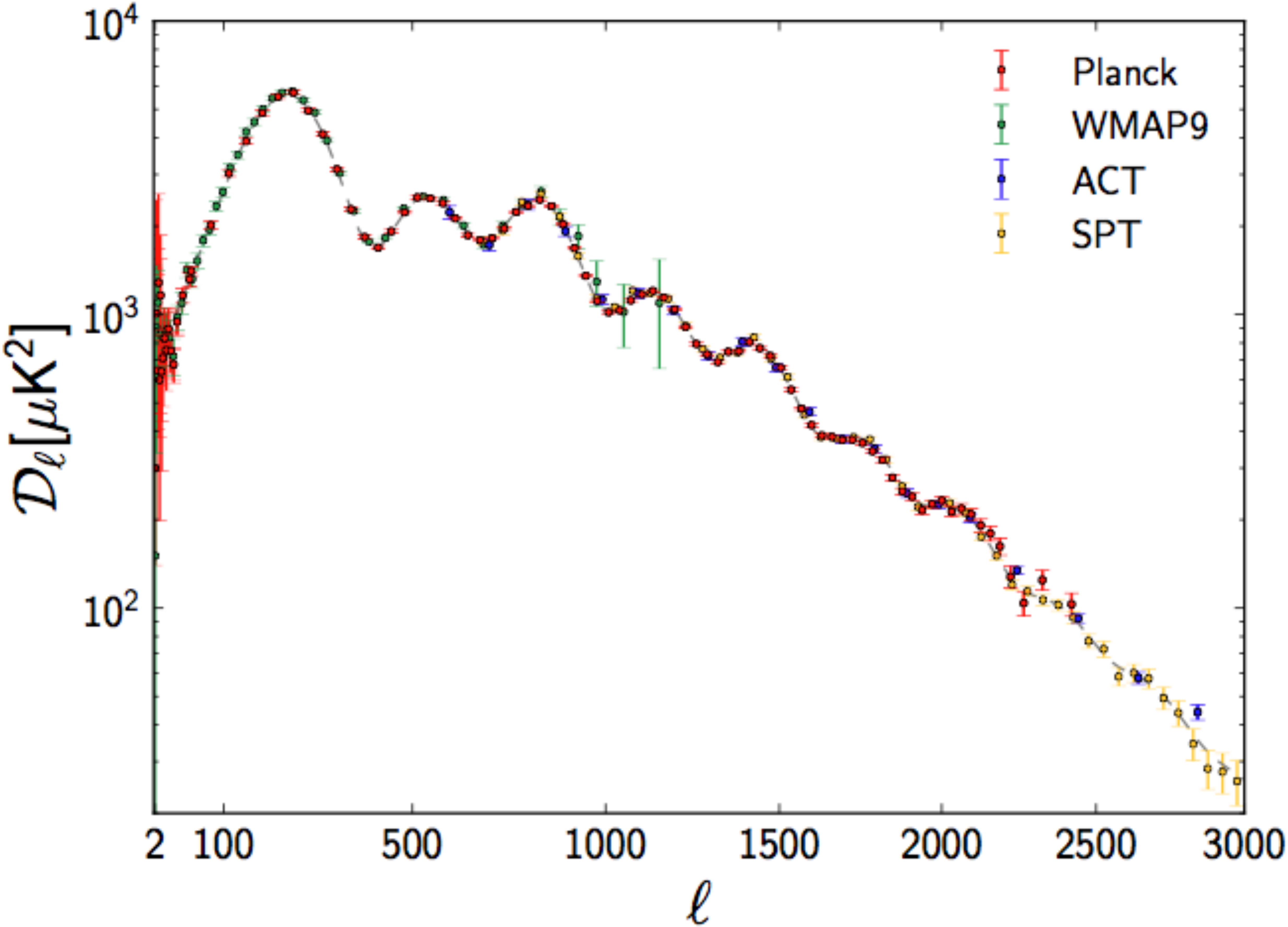


50×10^6 pixels



2500 harmonics

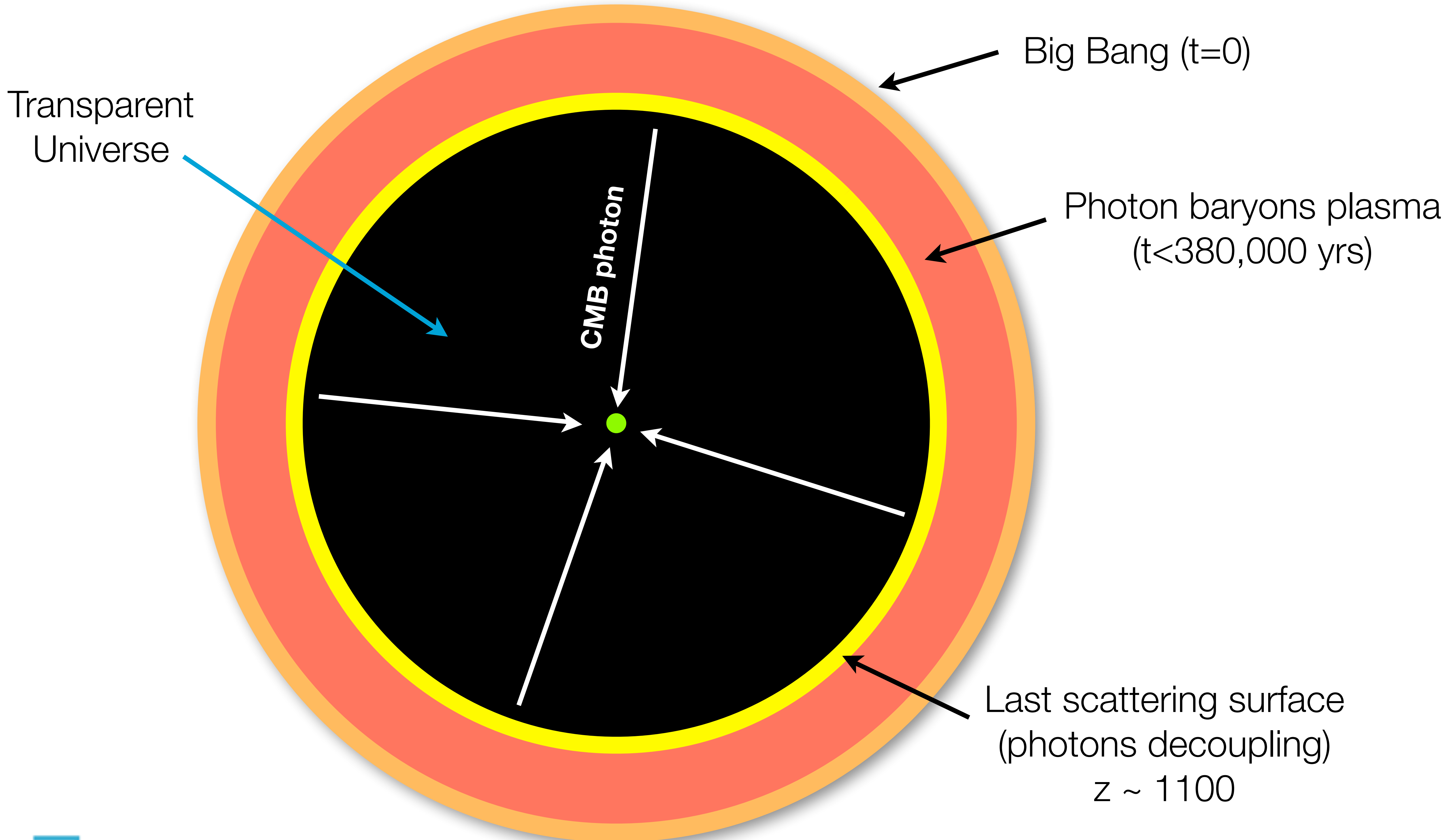




Planck
WMAP9
ACT
SPT
10^4
10^3
10^2
$\mathcal{D}_\ell[\mu K^2]$
2
100
500
1000
1500
2000
2500
3000
ℓ

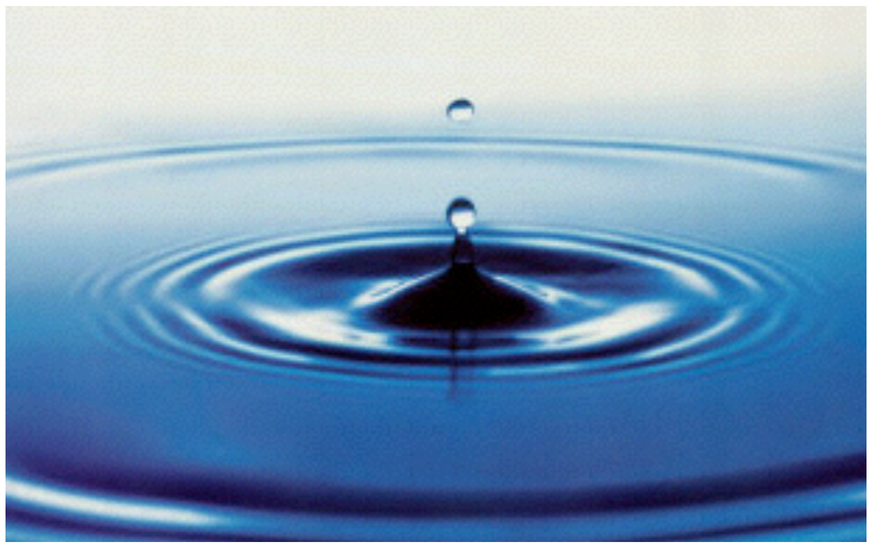
Origin of the CMB

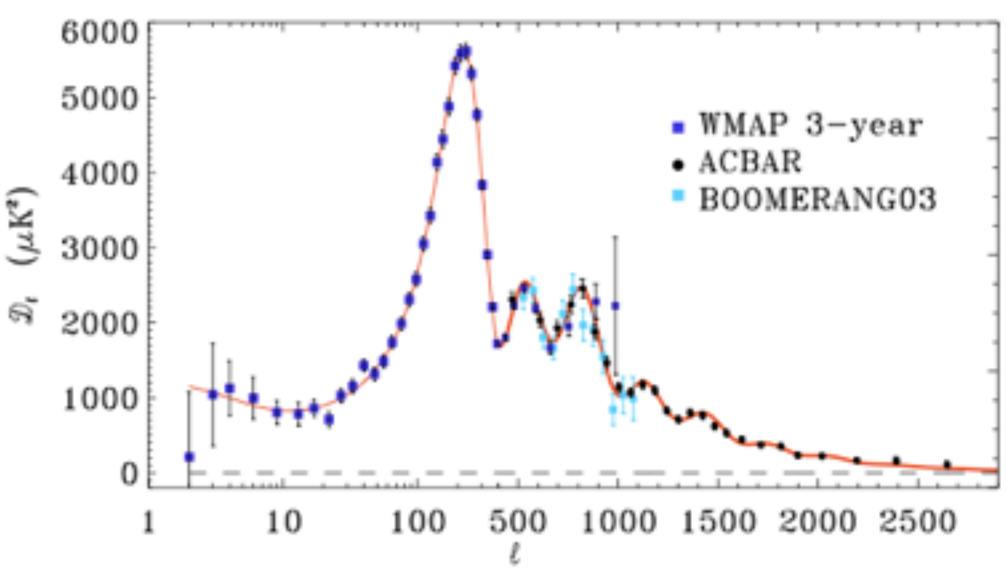
Imperial College London

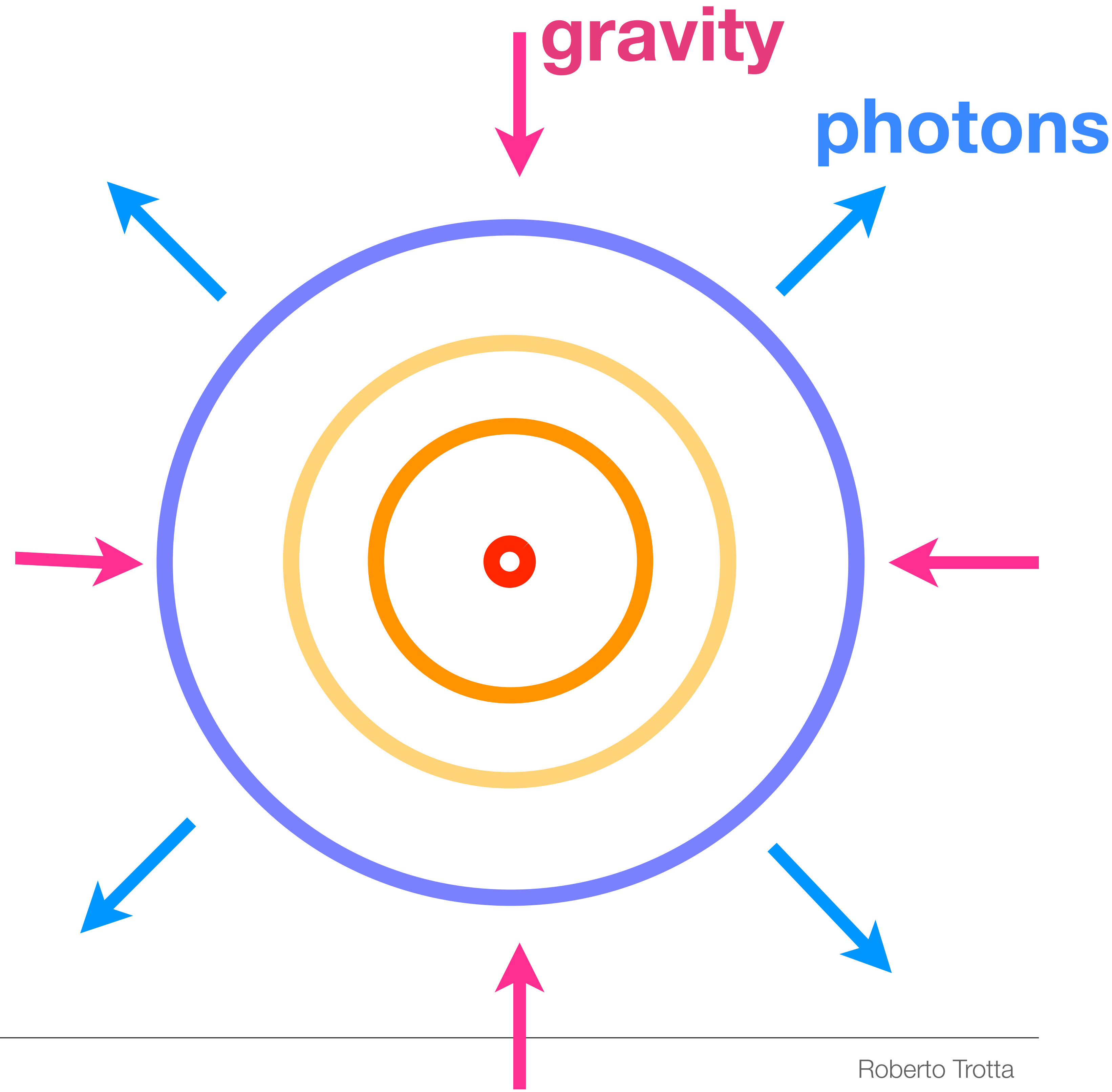


Cosmic sound

Imperial College London

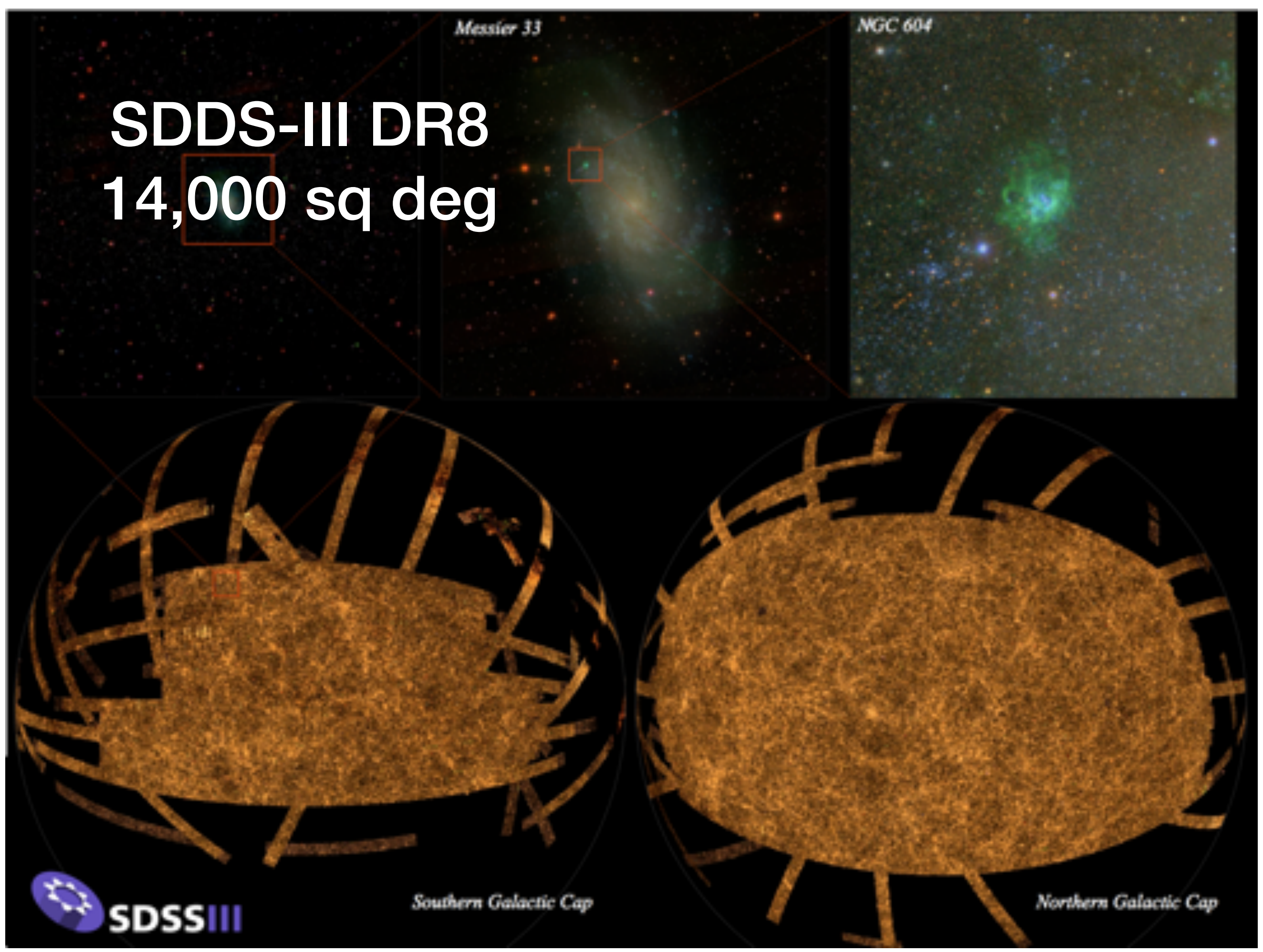




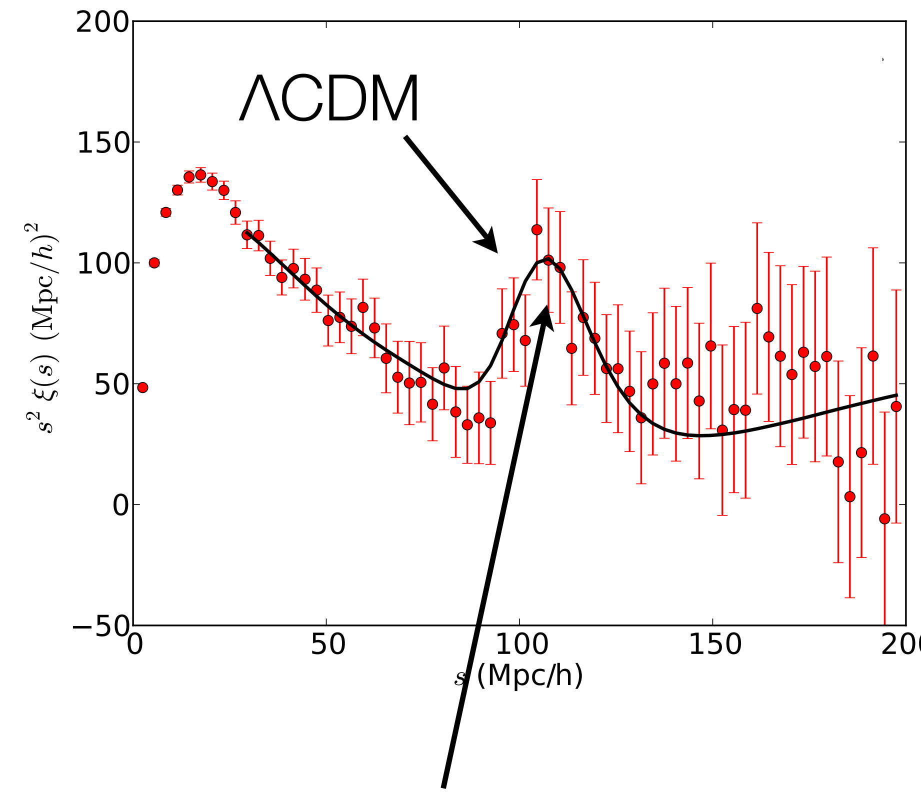


BAO: correlation between galaxies’ position

Primordial sound waves introduce extra correlation between galaxies on scales ~ 150 Mpc: this corresponds to (on average) 1 extra galaxy at this preferential separation



Baryonic acoustic oscillations (z~0.35)



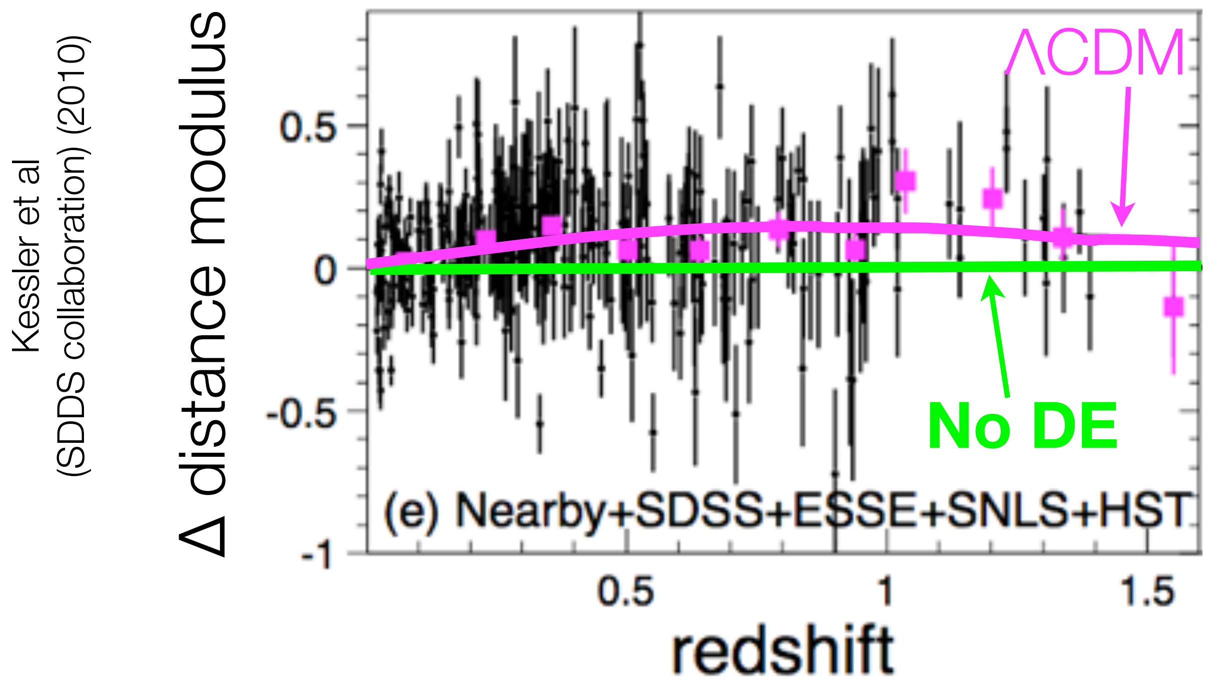
Padmanabhan et al (2012)

Baryonic **A**coustic **O**scillations from ~50,000 LRGs
1.9% distance accuracy to z=0.35
~4 sigma significance after reconstruction

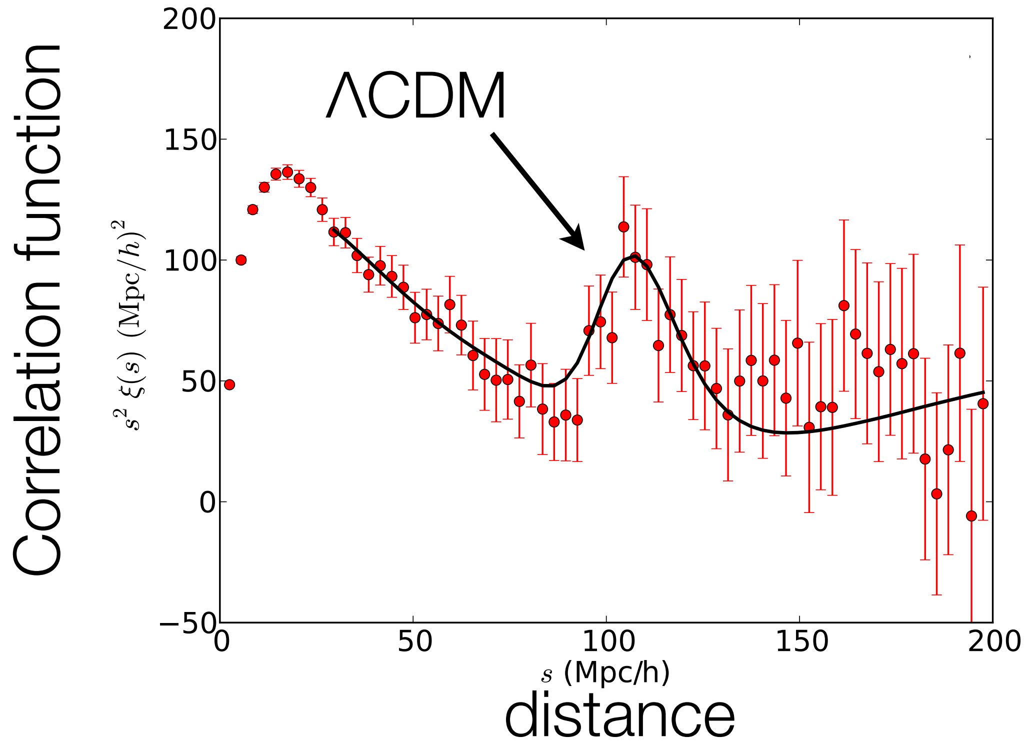
Low redshift cosmological probes

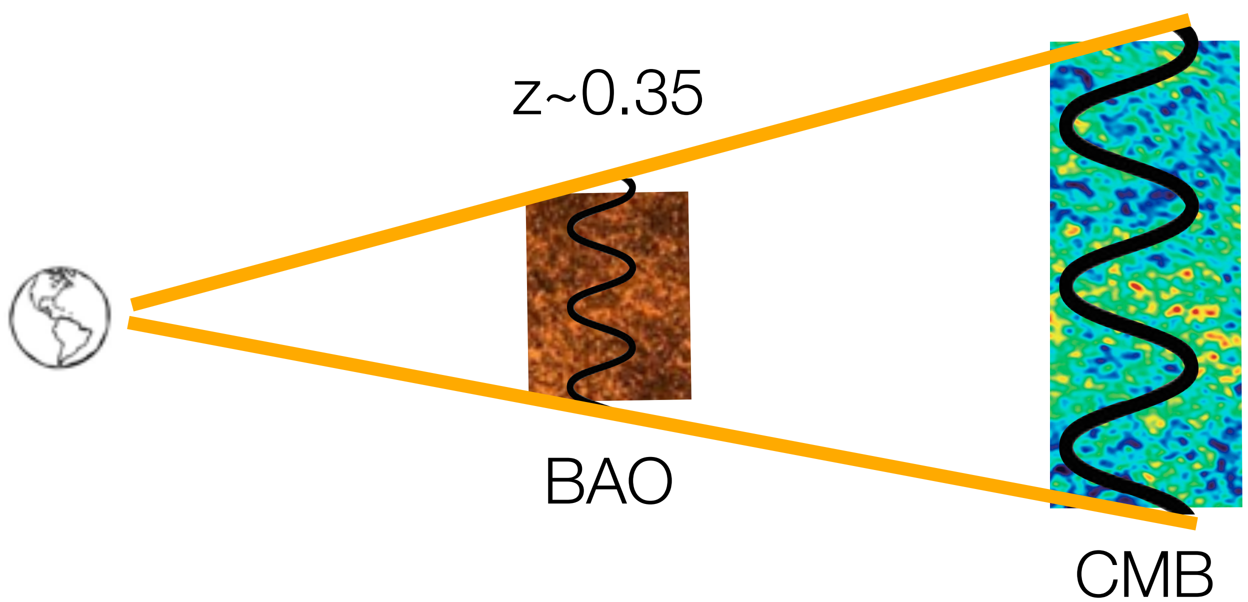
Imperial College London

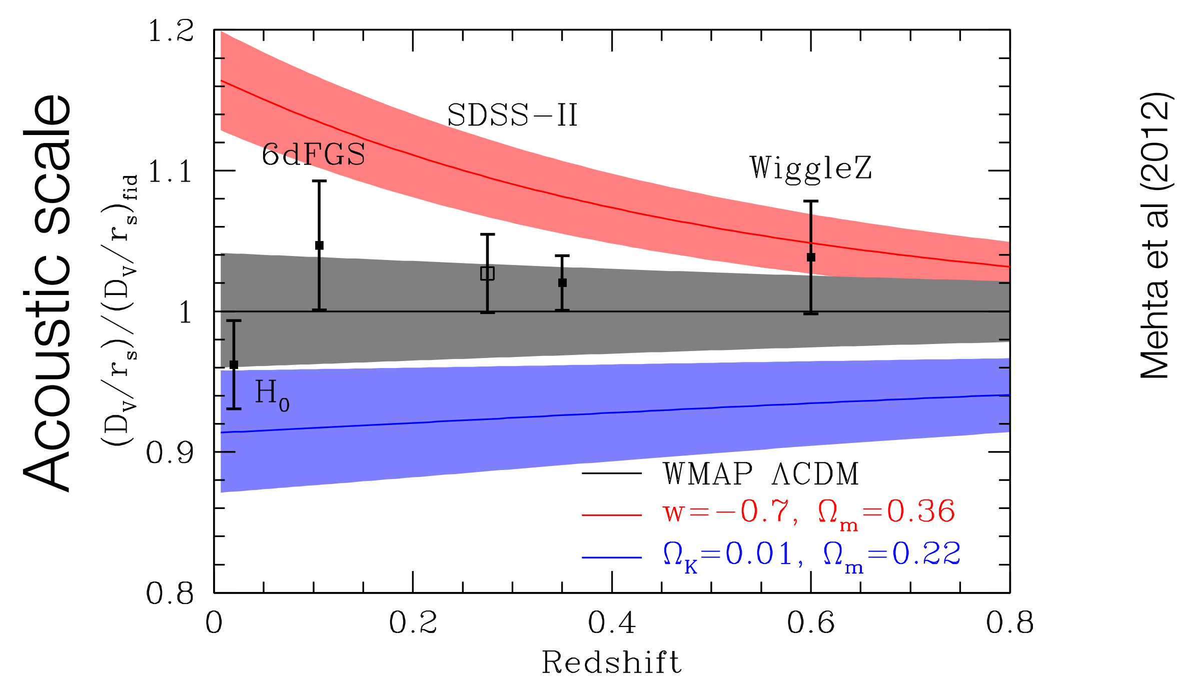
Supernovae type Ia (z < 1.5)



Baryonic acoustic oscillations (z~0.35)







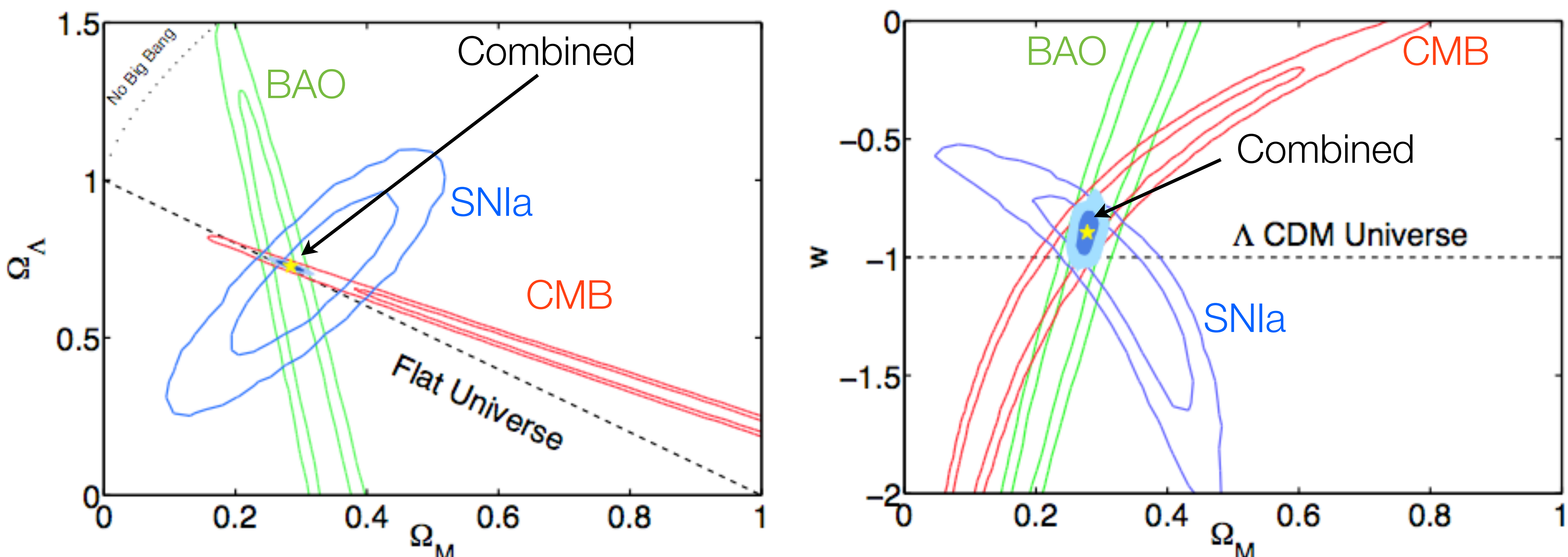
Putting it all together...

Imperial College London

Combined constraints on total matter ($\Omega_M = \Omega_B + \Omega_{CDM}$) and dark energy ($\Omega_\Lambda$) content (dark energy equation of state parameter w = pressure/energy density):

Assuming Λ (w=-1)

Assuming flatness ($\Omega_\Lambda + \Omega_M = 1$)



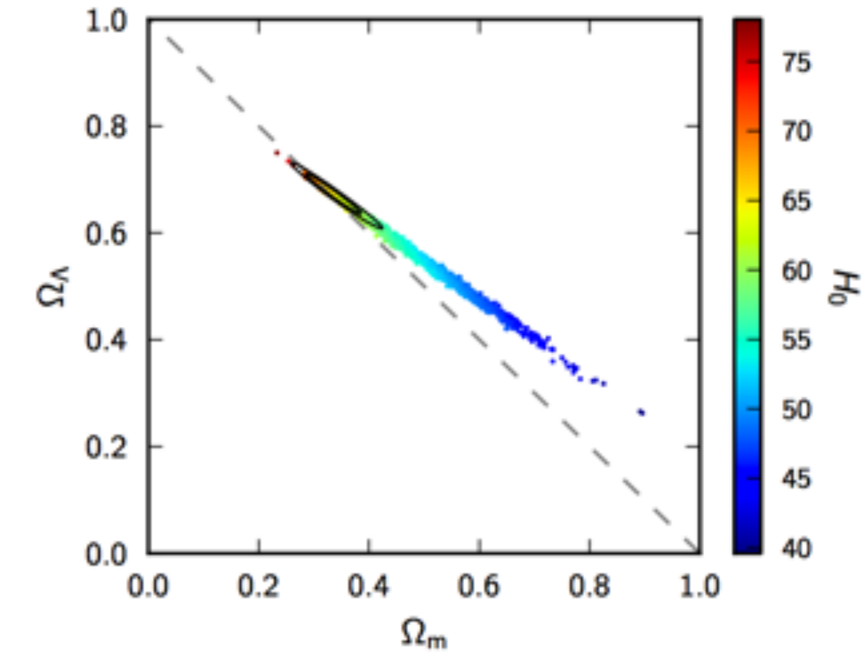
March, RT et al (2012)

Where are we today?

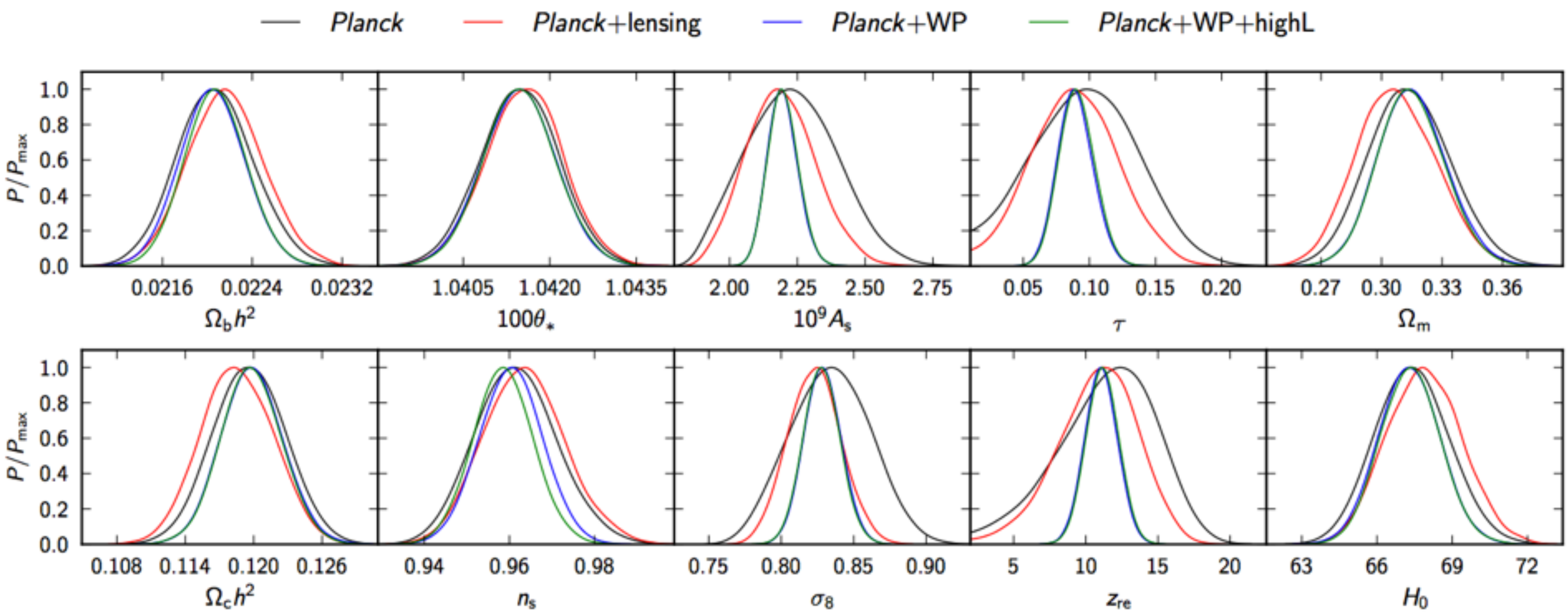
Imperial College London

Parameter	*Planck* Best fit	*Planck* 68% limits	*Planck*+lensing Best fit	*Planck*+lensing 68% limits	*Planck*+WP Best fit	*Planck*+WP 68% limits
$\Omega_b h^2$	0.022068	0.02207 ± 0.00033	0.022242	0.02217 ± 0.00033	0.022032	0.02205 ± 0.00028
$\Omega_c h^2$	0.12029	0.1196 ± 0.0031	0.11805	0.1186 ± 0.0031	0.12038	0.1199 ± 0.0027
$100\theta_{MC}$	1.04122	1.04132 ± 0.00068	1.04150	1.04141 ± 0.00067	1.04119	1.04131 ± 0.00063
τ	0.0925	0.097 ± 0.038	0.0949	0.089 ± 0.032	0.0925	$0.089^{+0.012}_{-0.014}$
n_s	0.9624	0.9616 ± 0.0094	0.9675	0.9635 ± 0.0094	0.9619	0.9603 ± 0.0073
$\ln(10^{10} A_s)$	3.098	3.103 ± 0.072	3.098	3.085 ± 0.057	3.0980	$3.089^{+0.024}_{-0.027}$
Ω_Λ	0.6825	0.686 ± 0.020	0.6964	0.693 ± 0.019	0.6817	$0.685^{+0.018}_{-0.016}$
Ω_m	0.3175	0.314 ± 0.020	0.3036	0.307 ± 0.019	0.3183	$0.315^{+0.016}_{-0.018}$
σ_8	0.8344	0.834 ± 0.027	0.8285	0.823 ± 0.018	0.8347	0.829 ± 0.012
z_{re}	11.35	$11.4^{+4.0}_{-2.8}$	11.45	$10.8^{+3.1}_{-2.5}$	11.37	11.1 ± 1.1
H_0	67.11	67.4 ± 1.4	68.14	67.9 ± 1.5	67.04	67.3 ± 1.2
$10^9 A_s$	2.215	2.23 ± 0.16	2.215	$2.19^{+0.12}_{-0.14}$	2.215	$2.196^{+0.051}_{-0.060}$
$\Omega_m h^2$	0.14300	0.1423 ± 0.0029	0.14094	0.1414 ± 0.0029	0.14305	0.1426 ± 0.0025
$\Omega_m h^3$	0.09597	0.09590 ± 0.00059	0.09603	0.09593 ± 0.00058	0.09591	0.09589 ± 0.00057
Y_P	0.247710	0.24771 ± 0.00014	0.247785	0.24775 ± 0.00014	0.247695	0.24770 ± 0.00012
Age/Gyr	13.819	13.813 ± 0.058	13.784	13.796 ± 0.058	13.8242	13.817 ± 0.048

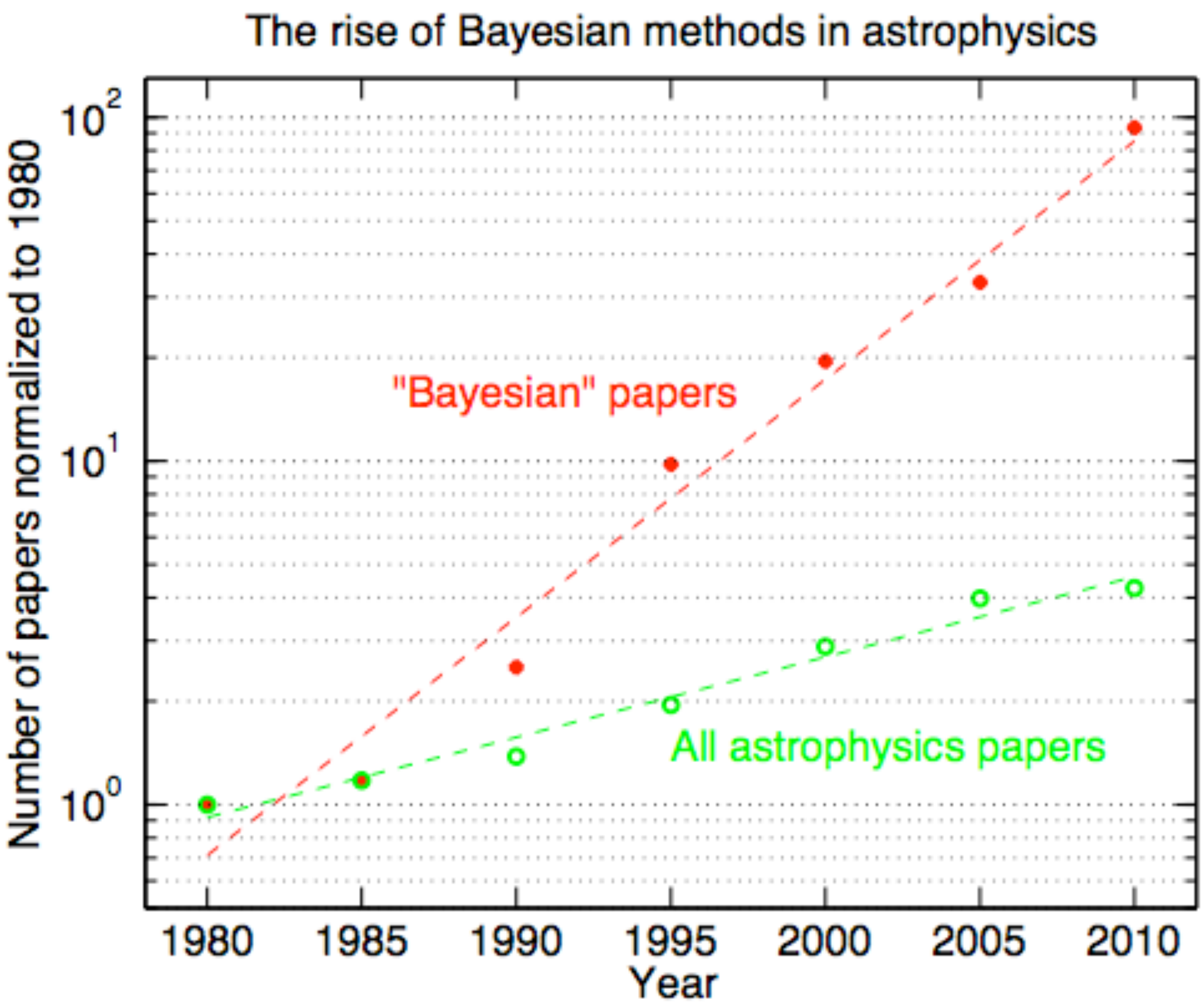
From Planck (2013)



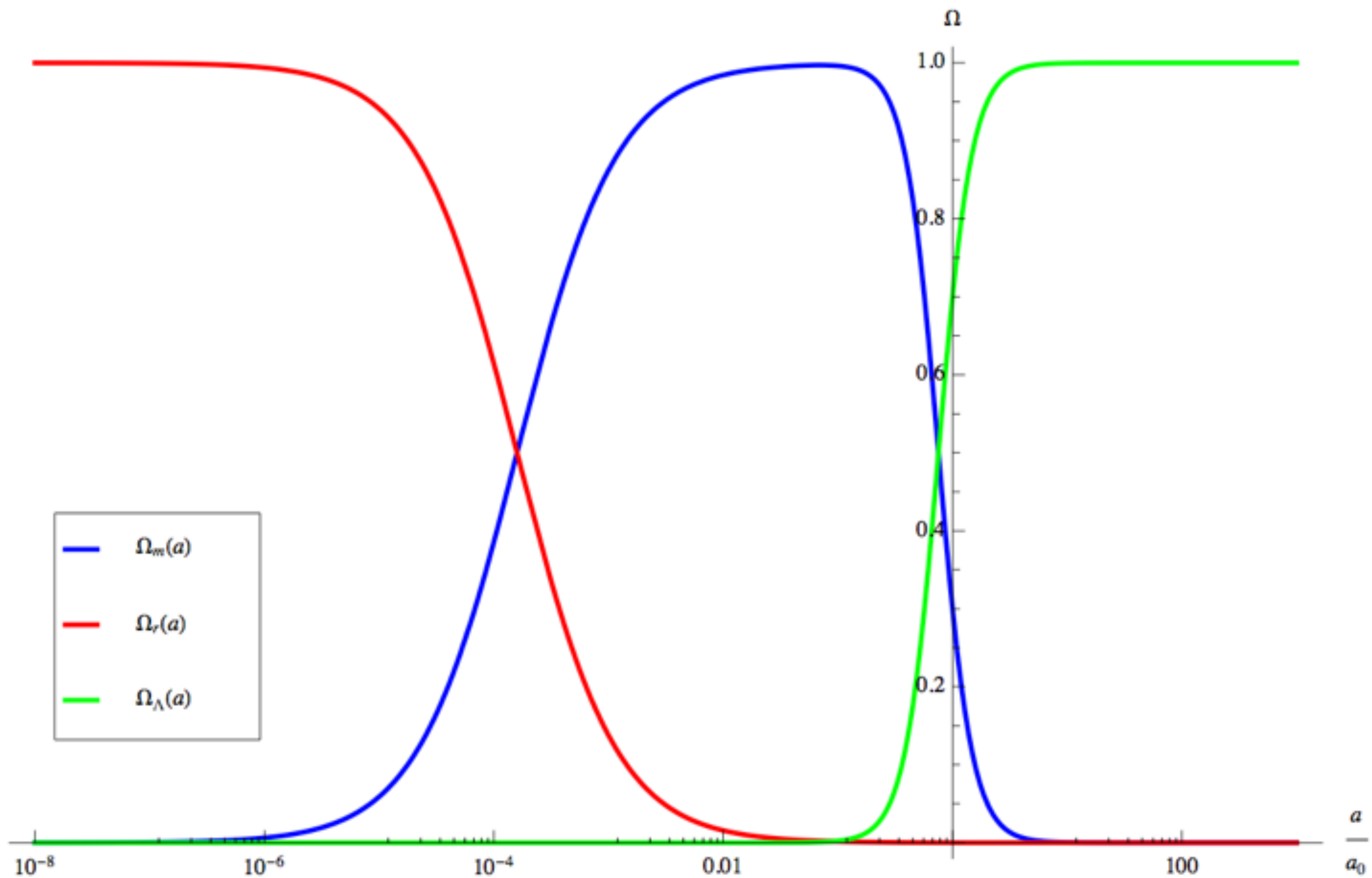
Planck Collaboration: Cosmological parameter



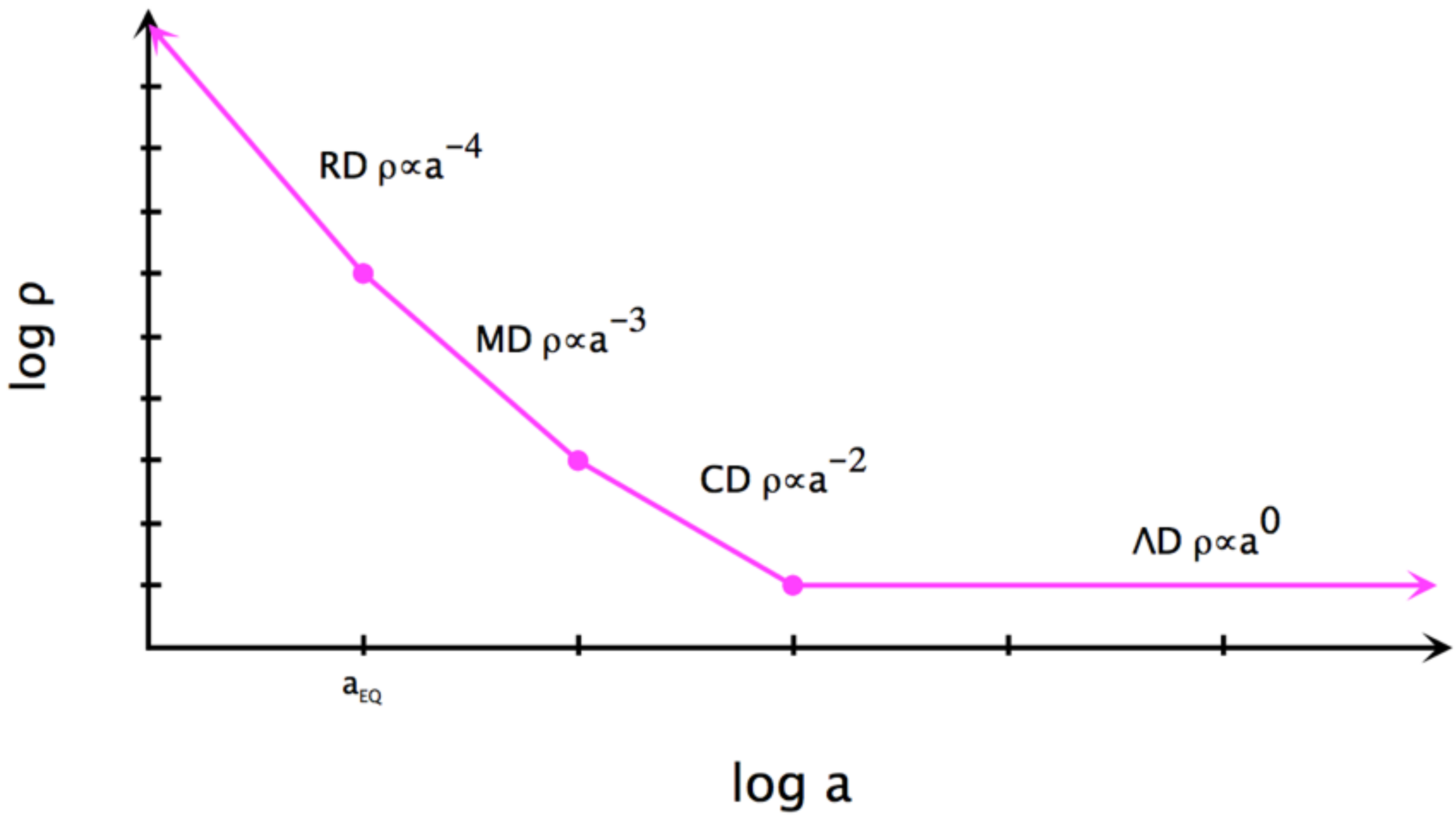
Bayes in the sky



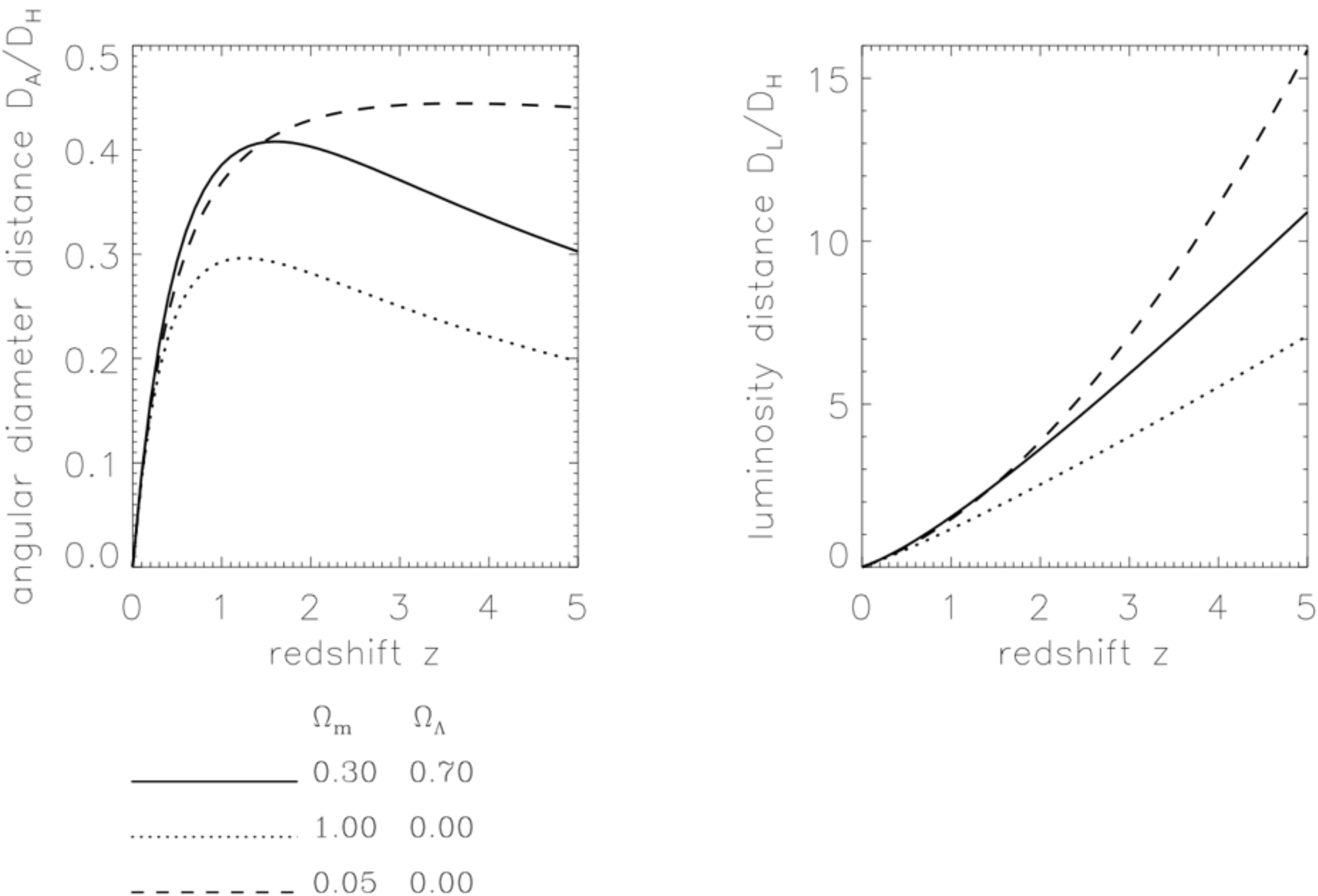
Cosmology



Ω
1.0
0.8
0.6
0.4
0.2
$\Omega_m(a)$
$\Omega_r(a)$
$\Omega_\Lambda(a)$
10^{-8}
10^{-6}
10^{-4}
0.01
100
$\frac{a}{a_0}$

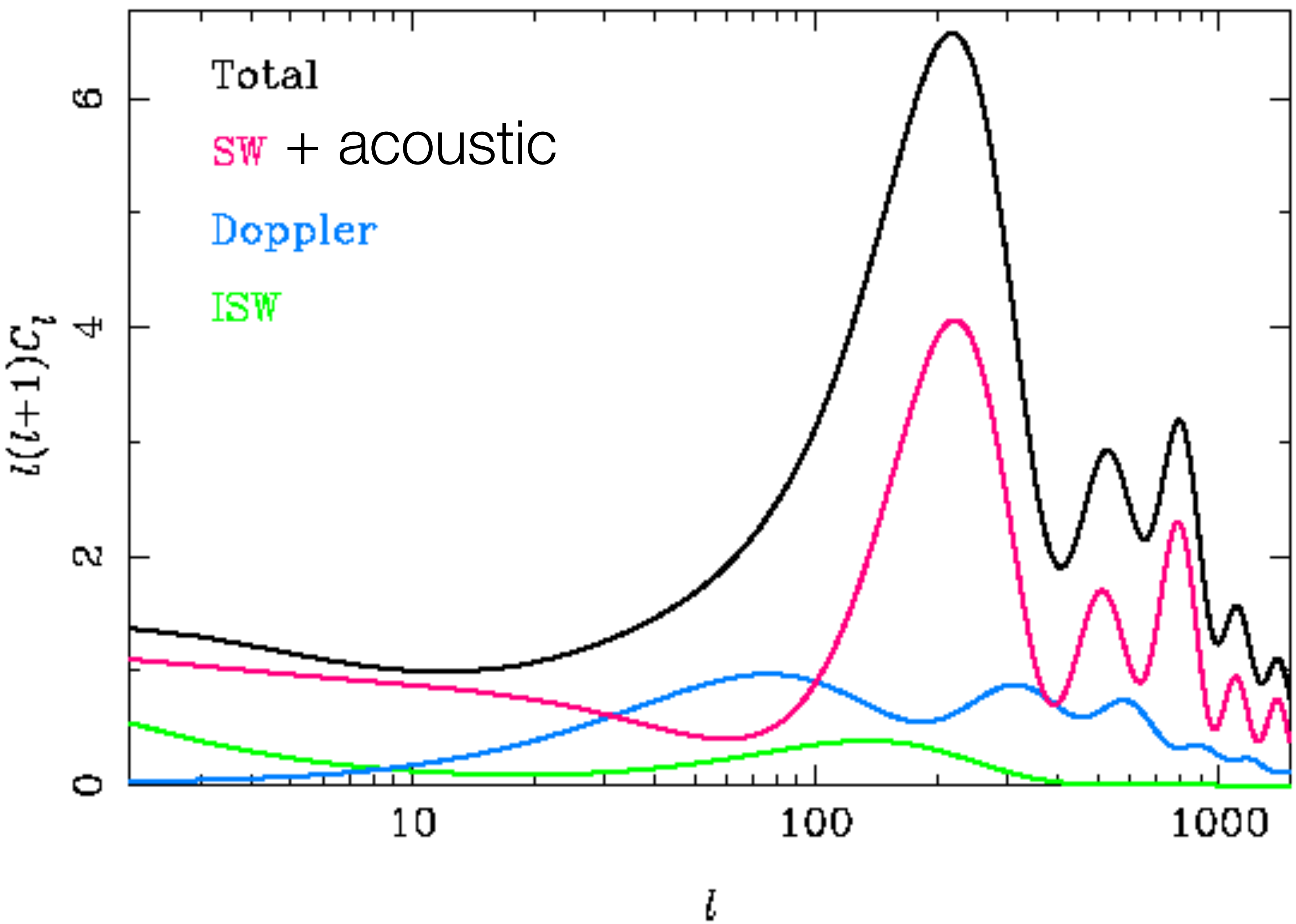


RD $\rho \propto a^{-4}$
MD $\rho \propto a^{-3}$
CD $\rho \propto a^{-2}$
ΛD $\rho \propto a^{0}$
log ρ
a_{EQ}
log a


angular diameter distance D_A/D_H
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3
4
5
redshift z
luminosity distance D_L/D_H
0
5
10
15
0
1
2
3
4
5
redshift z
Ω_m Ω_Λ
0.30 0.70
1.00 0.00
0.05 0.00

CMB decomposition

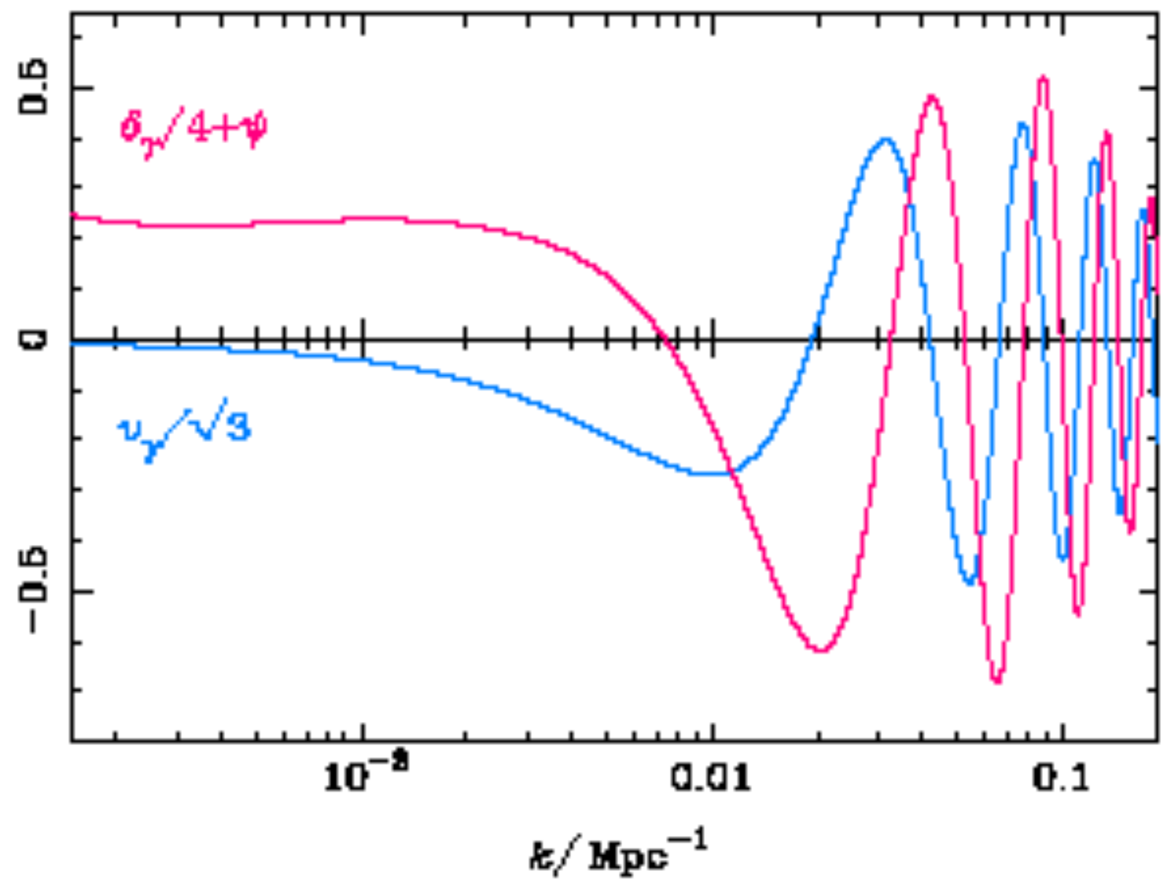
Imperial College London



ICIC

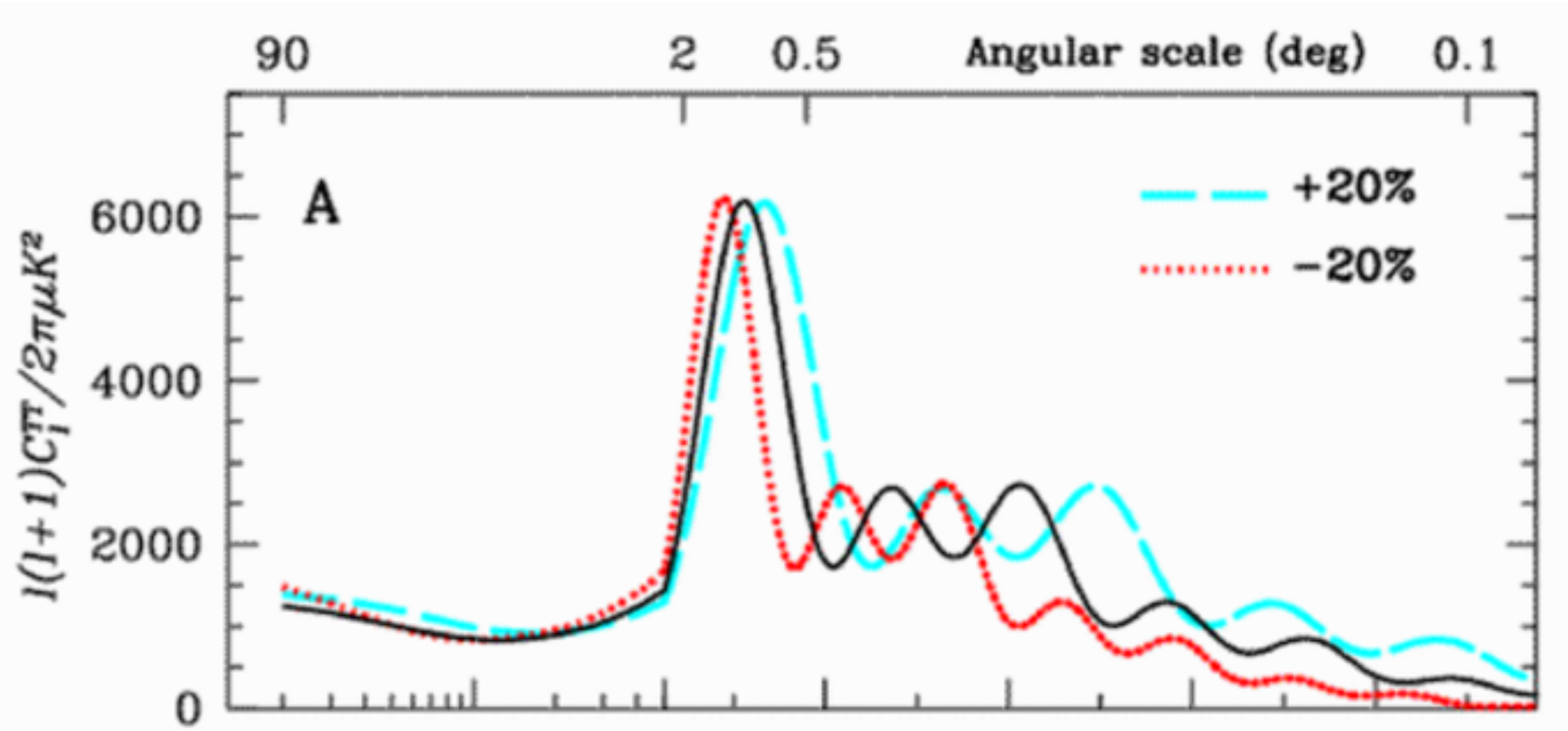
Temperature fluctuations

Imperial College London



Angular projection

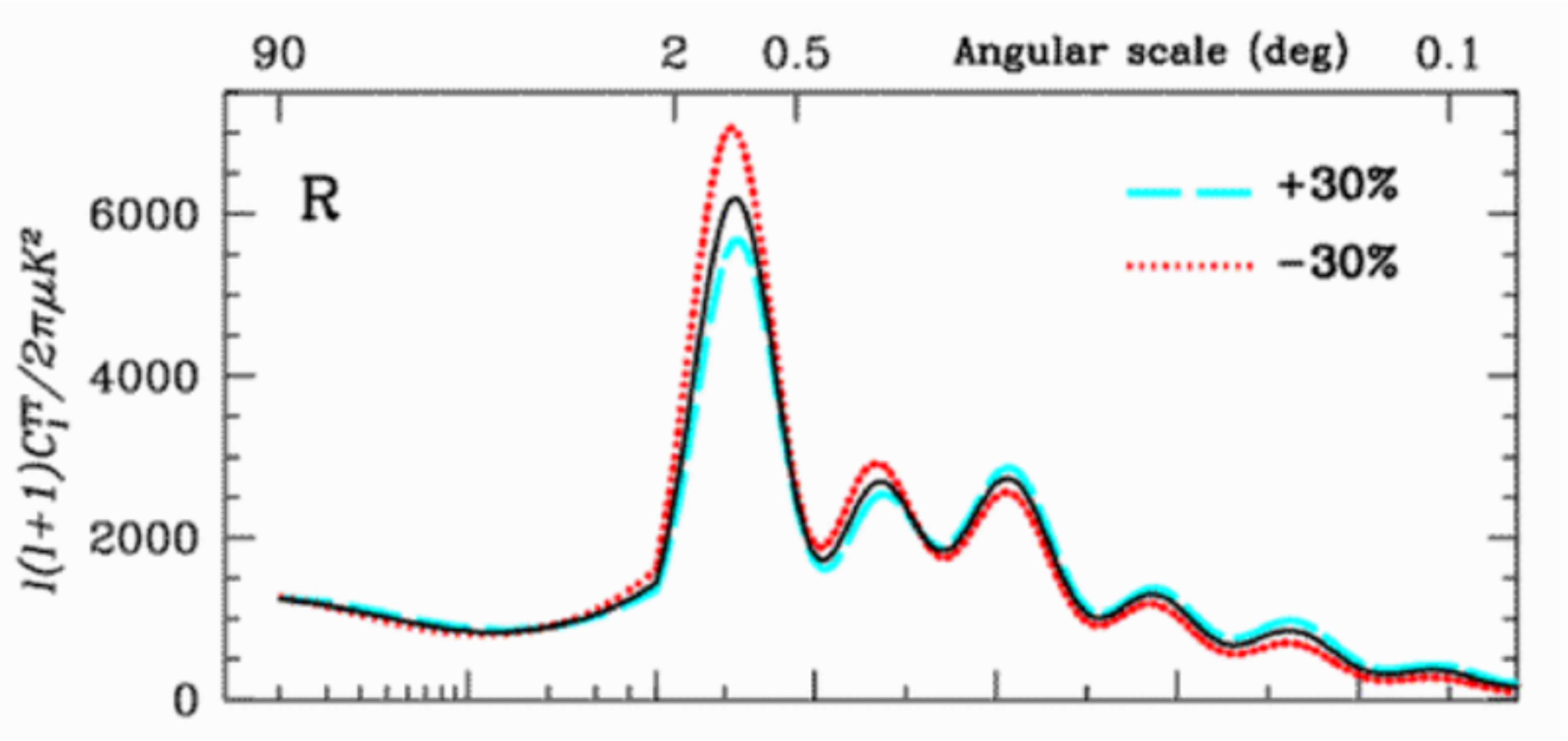
Imperial College London



ICIC

Matter-radiation equality

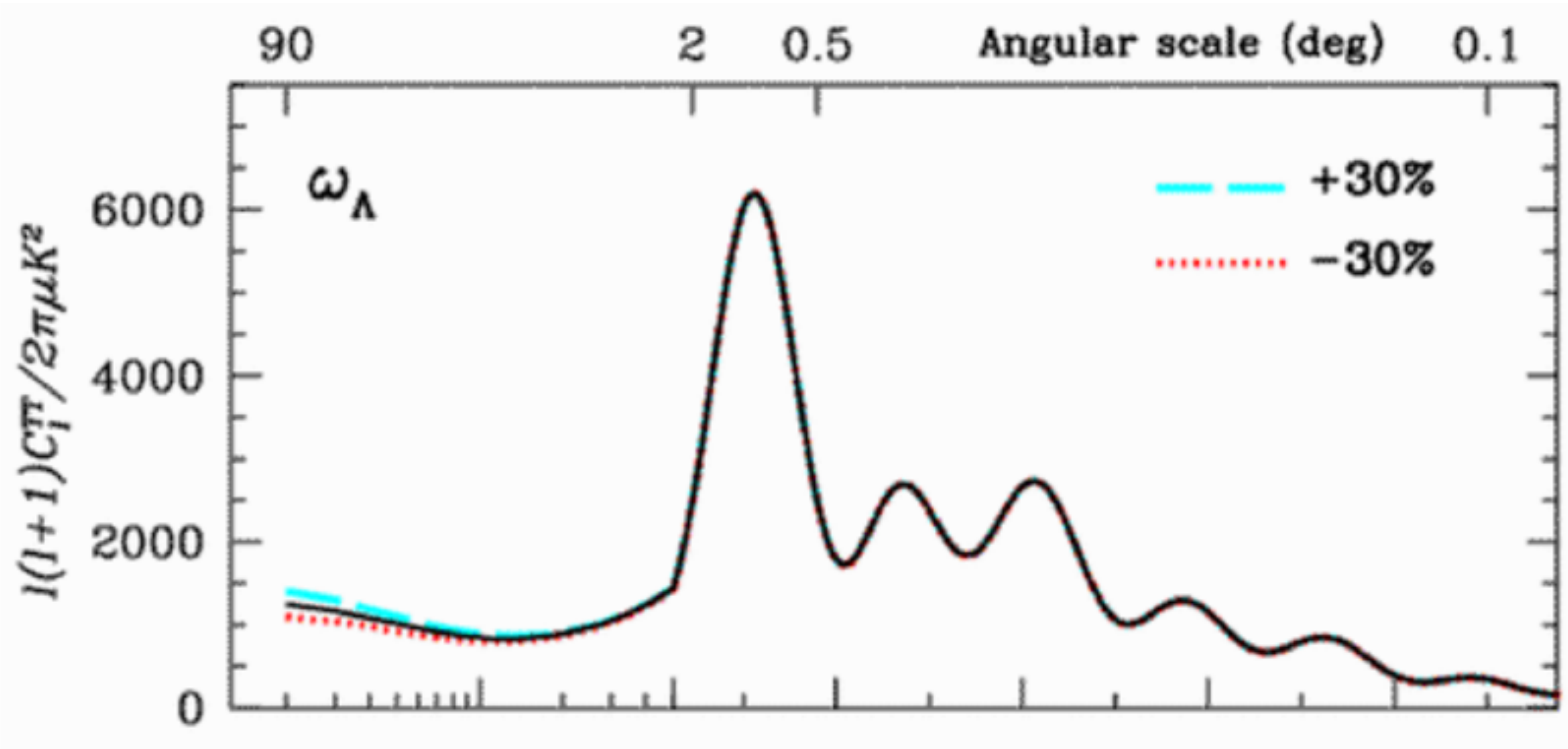
Imperial College London



ICIC

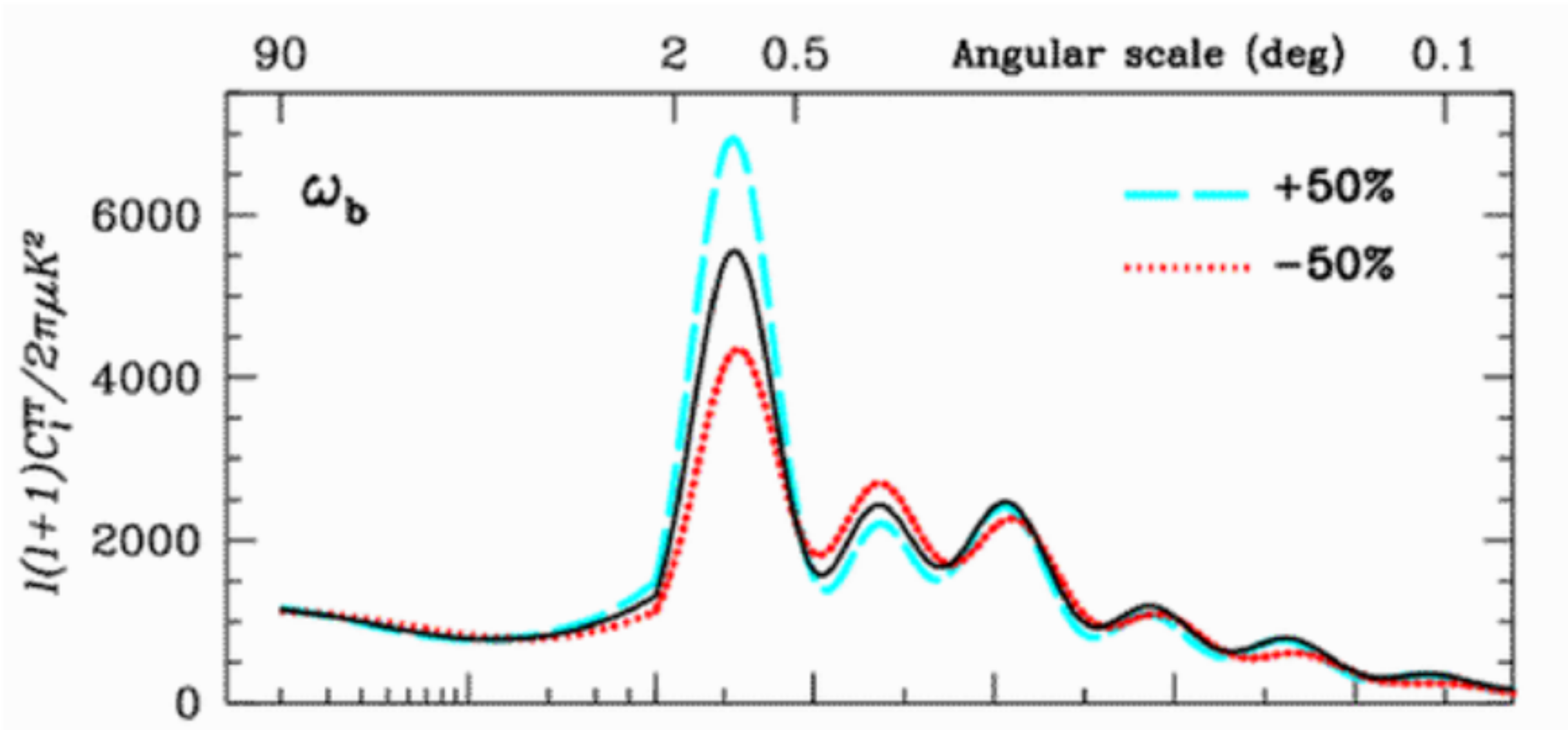
Cosmological constant density

Imperial College London



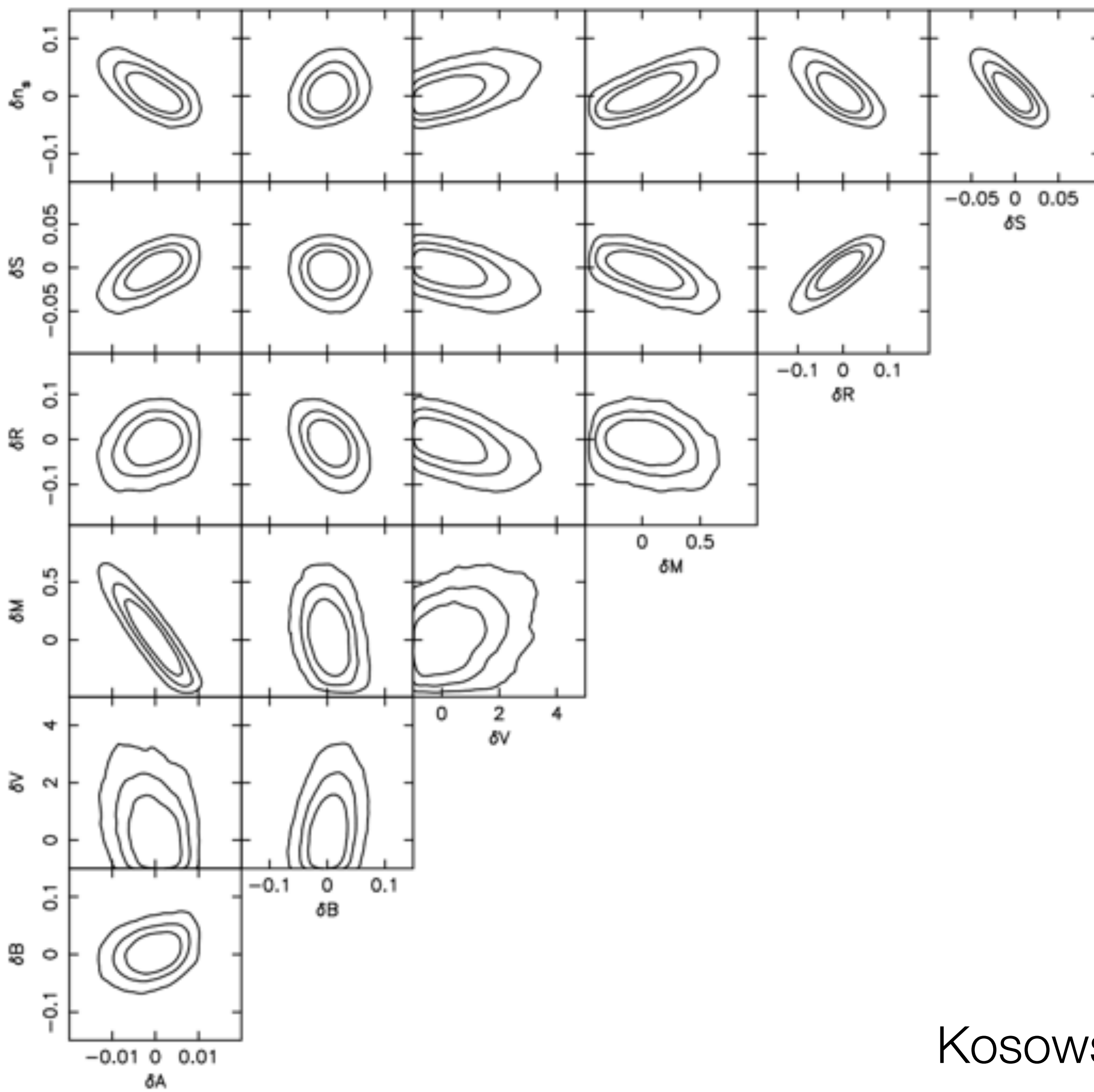
Baryon density

Imperial College London



Normal parameters: good

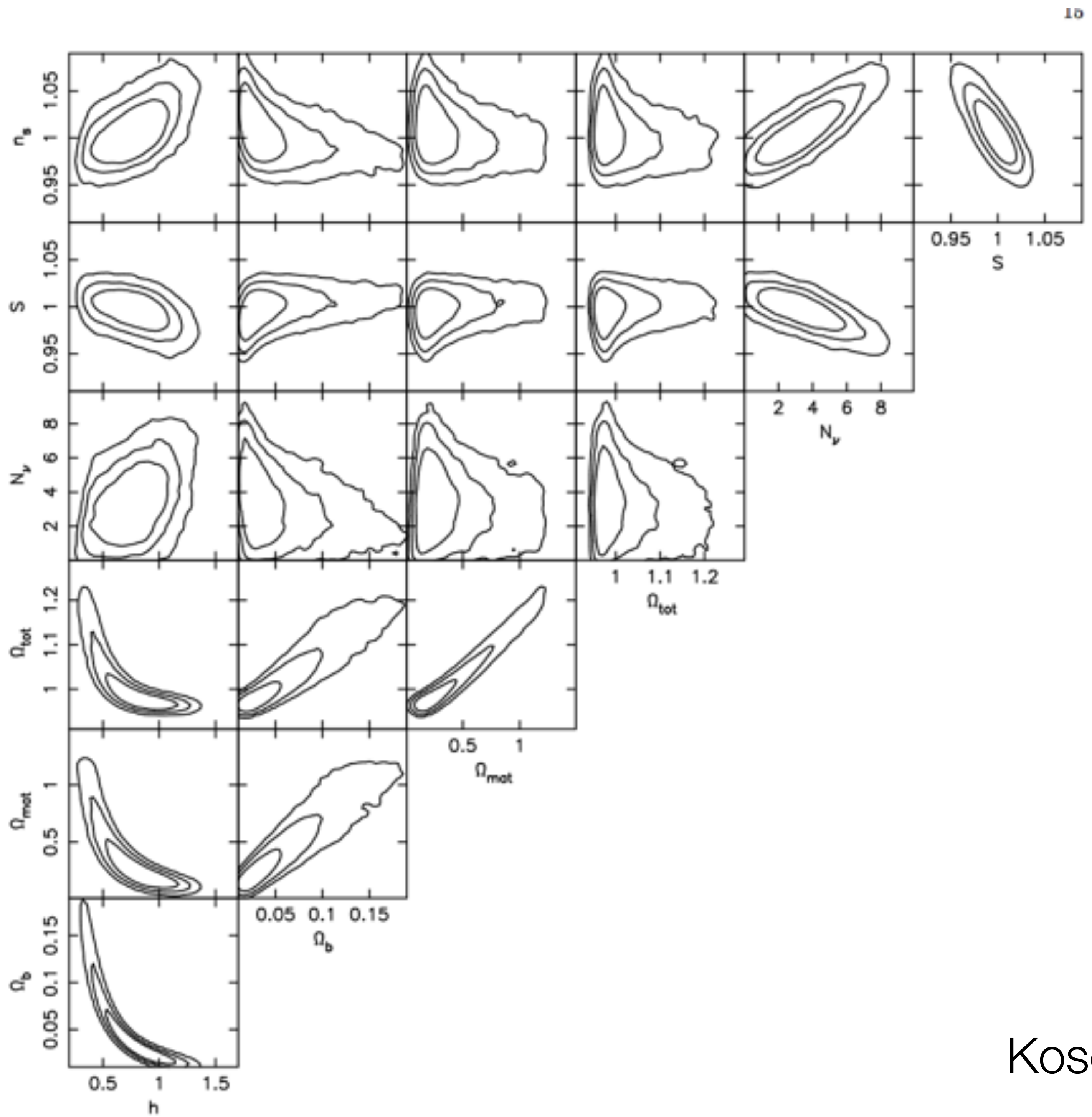
Imperial College London



Kosowsky et al (2002)

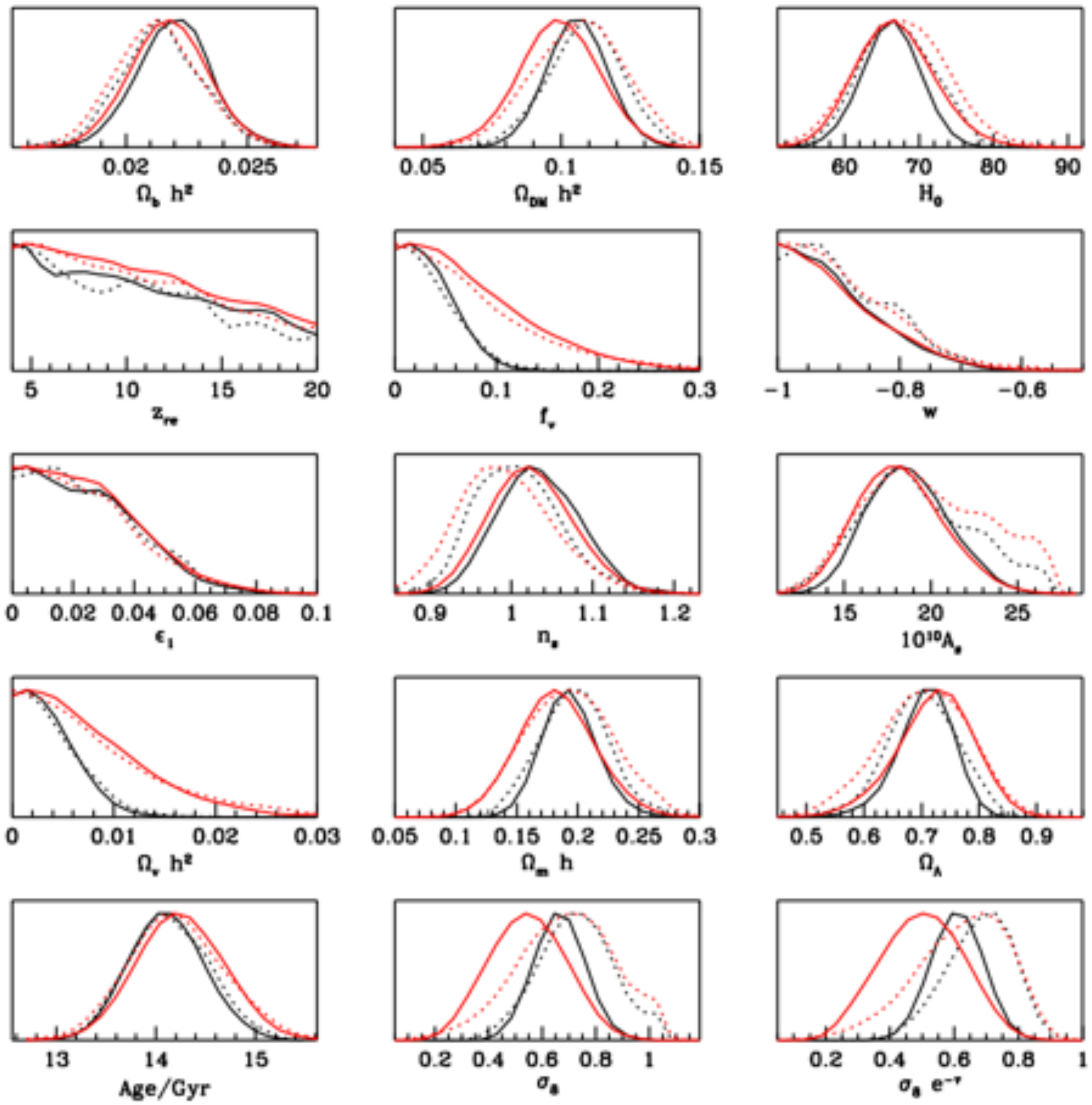
"Physical" parameters: bad

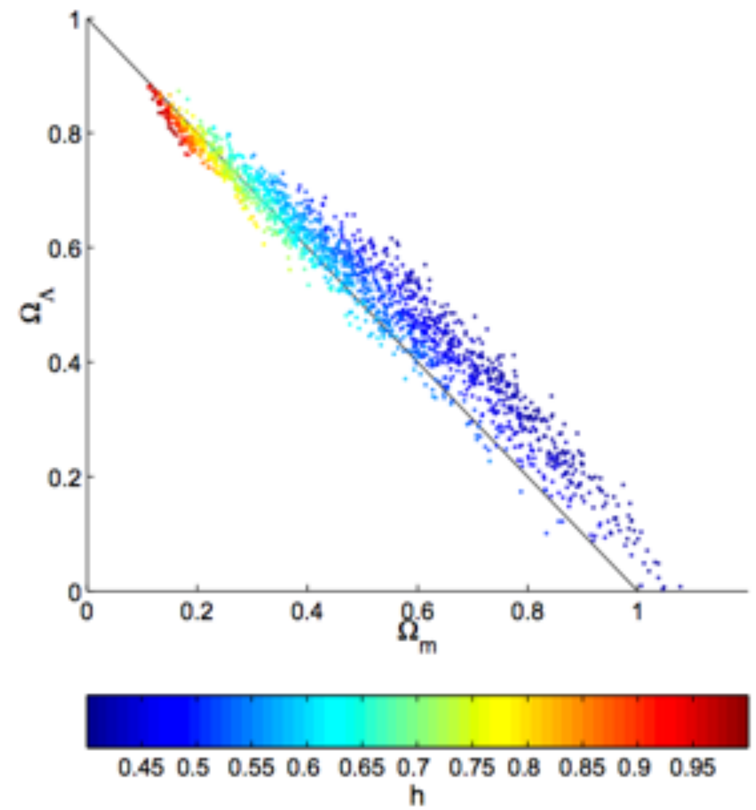
Imperial College London



Kosowsky et al (2002)

Cosmomc: example





Bridle & Lewis (2003)

The general solution

Imperial College London

$$P(\theta|d, I) \propto P(d|\theta, I)P(\theta|I)$$

- Once the RHS is defined, how do we evaluate the LHS?
- Analytical solutions exist only for the simplest cases (e.g. Gaussian linear model)
- Cheap computing power means that numerical solutions are often just a few clicks away!
- Workhorse of Bayesian inference: Markov Chain Monte Carlo (MCMC) methods. A procedure to generate a list of samples from the posterior.

MCMC estimation

Imperial College London

$$P(\theta|d, I) \propto P(d|\theta, I)P(\theta|I)$$

- A Markov Chain is a list of samples θ_1, θ_2, θ_3,... whose density reflects the (unnormalized) value of the posterior
- A MC is a sequence of random variables whose *(n+1)*-th element only depends on the value of the *n*-th element
- **Crucial property:** a Markov Chain converges to a stationary distribution, i.e. one that does not change with time. In our case, the posterior.
- From the chain, expectation values wrt the posterior are obtained very simply:

$$\langle\theta\rangle = \int d\theta P(\theta|d)\theta \approx \frac{1}{N}\sum_i \theta_i$$

$$\langle f(\theta)\rangle = \int d\theta P(\theta|d) f(\theta) \approx \frac{1}{N}\sum_i f(\theta_i)$$

Reporting inferences

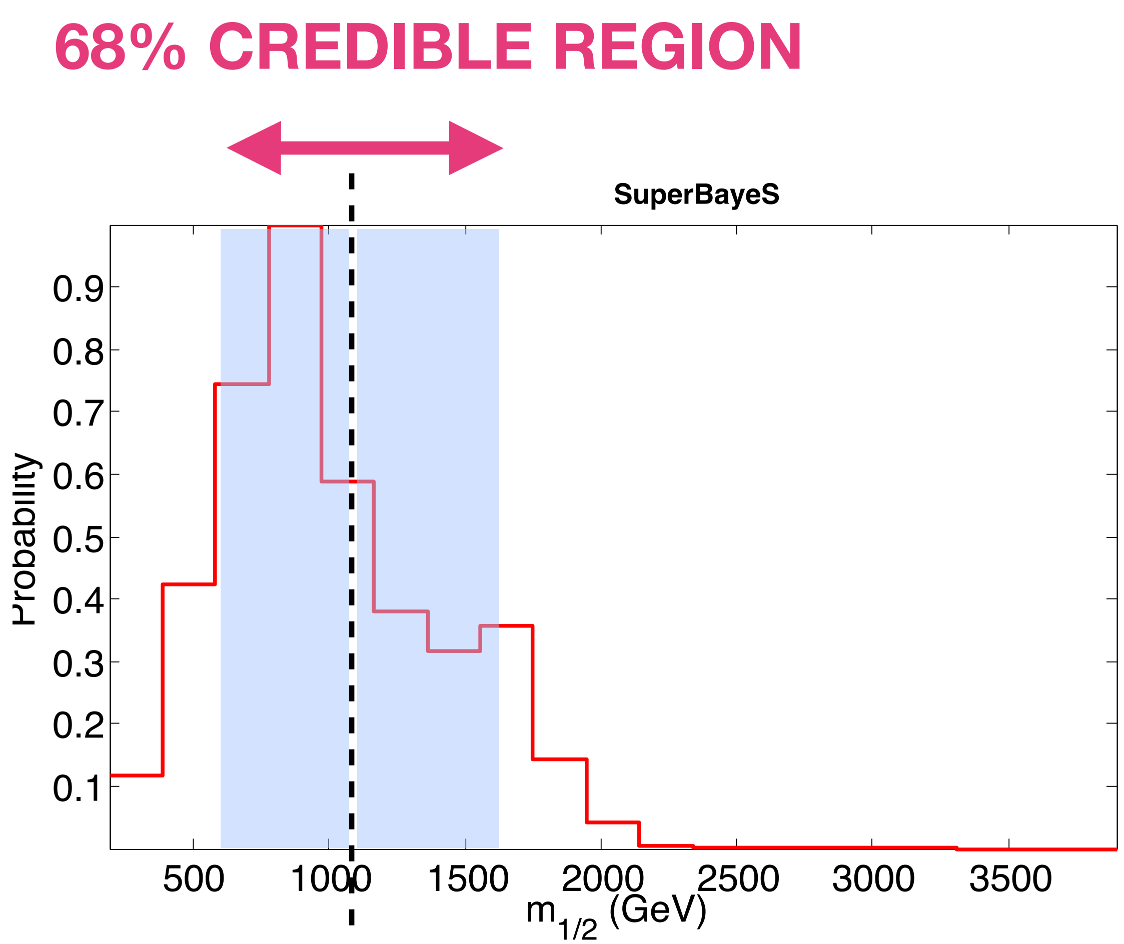
Imperial College London

- **Once P(θ|d, I) found, we can report inference by:**
  - Summary statistics (best fit point, average, mode)
  - Credible regions (e.g. shortest interval containing 68% of the posterior probability for θ). Warning: this has **not** the same meaning as a frequentist confidence interval! (Although the 2 might be formally identical)
  - Plots of the marginalised distribution, integrating out nuisance parameters (i.e. parameters we are not interested in). This generalizes the propagation of errors:

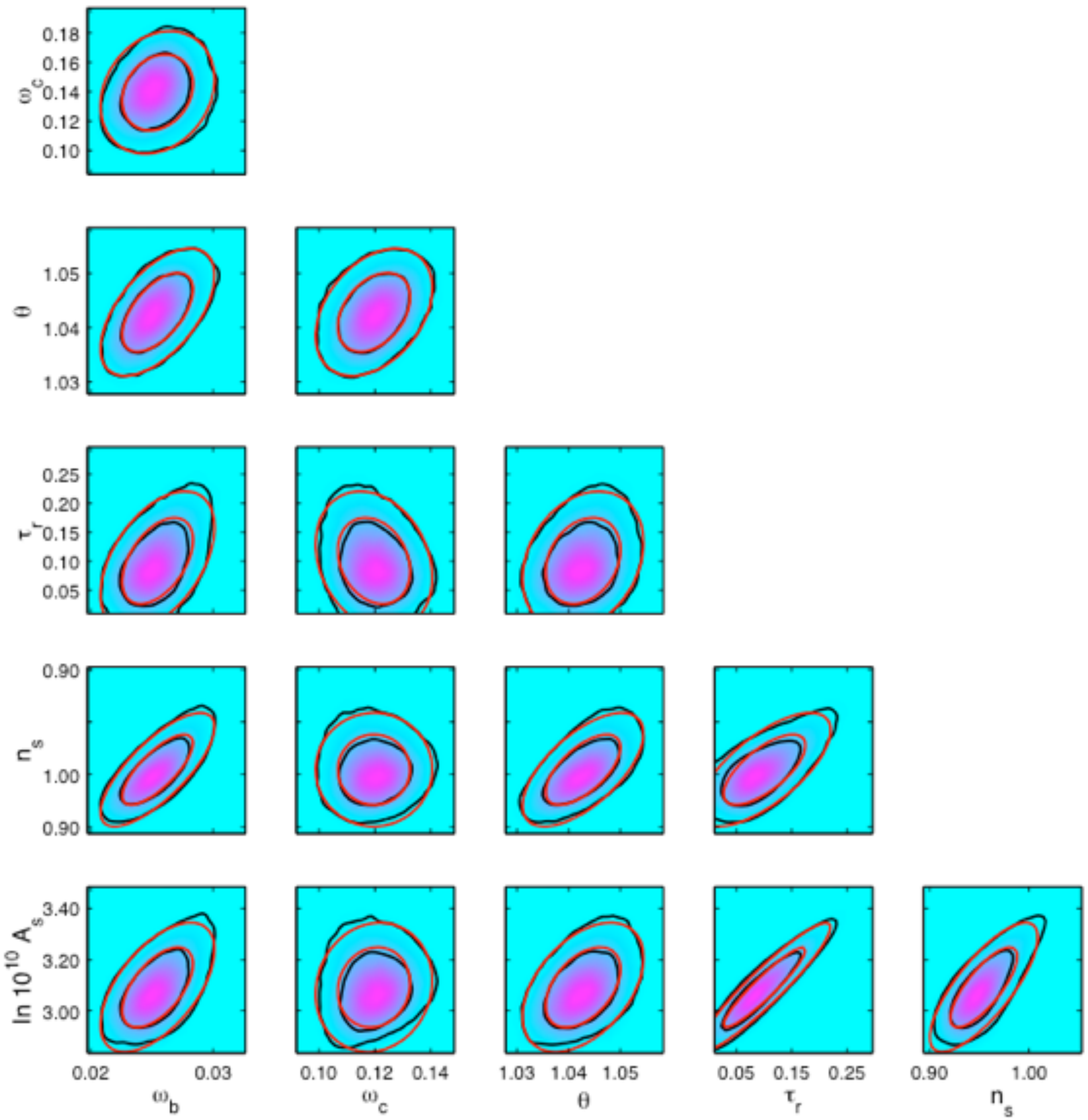
$$P(\theta|d,I) = \int d\phi P(\theta,\phi|d,I)$$

Credible regions: Bayesian approach

- Use the prior to define a metric on parameter space.
- **Bayesian methods:** the best-fit has no special status. Focus on region of large posterior probability mass instead.
    - Markov Chain Monte Carlo (MCMC)
    - Nested sampling
    - Hamiltonian MC
- Determine posterior credible regions: e.g. symmetric interval around the mean containing 68% of samples



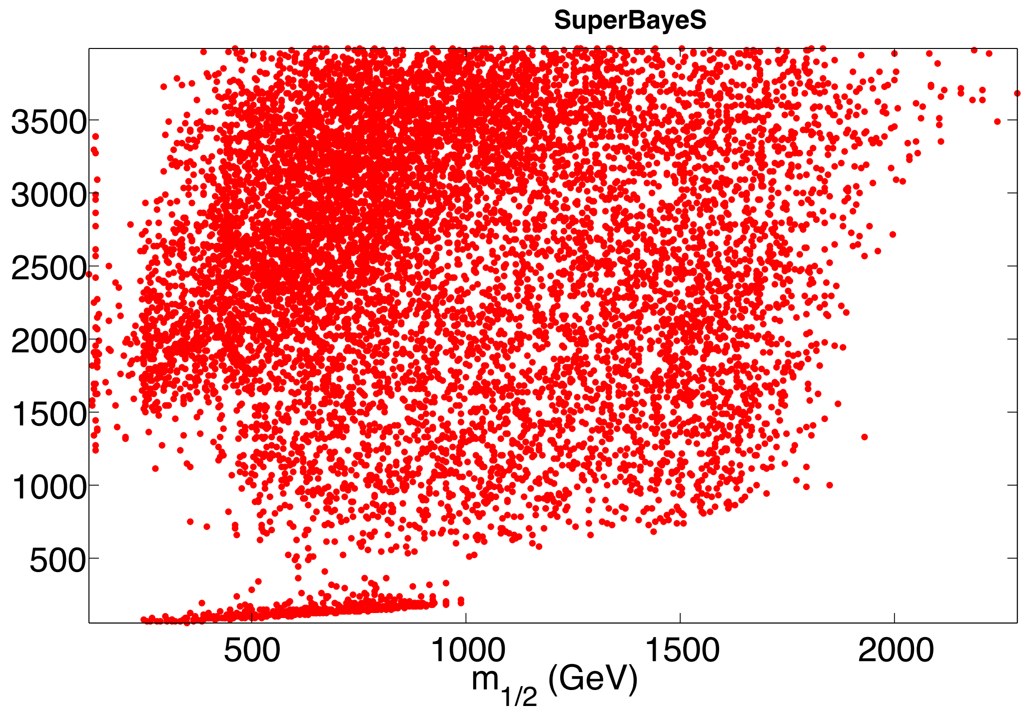
Gaussian case

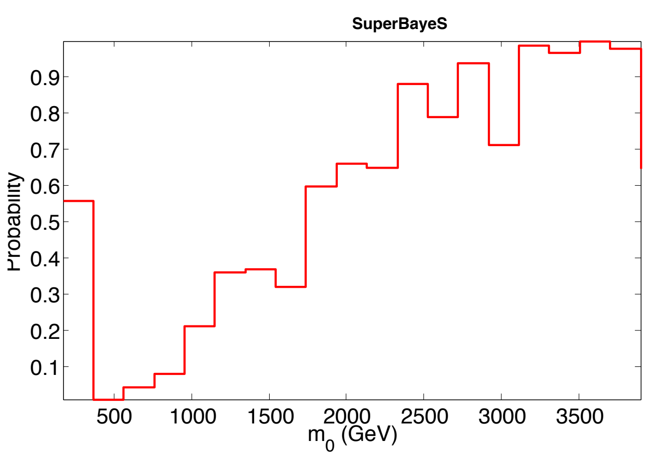


MCMC estimation

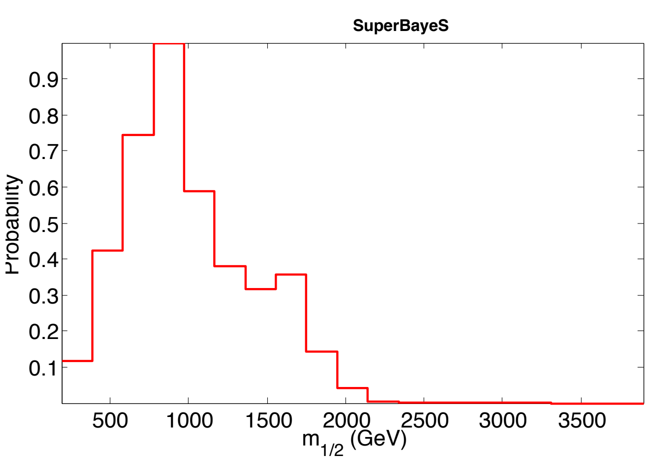
- **Marginalisation becomes trivial:** create bins along the dimension of interest and simply count samples falling within each bins ignoring all other coordinates
- Examples (from **superbayes.org**) :

2D distribution of samples from joint posterior





1D marginalised posterior (along y)



1D marginalised posterior (along x)

The simplest MCMC algorithm

Imperial College London

- Several (sophisticated) algorithms to build a MC are available: e.g. Metropolis-Hastings, Hamiltonian sampling, Gibbs sampling, rejection sampling, mixture sampling, slice sampling and more...

- Arguably the simplest algorithm is the **Metropolis (1954) algorithm:**

  - pick a starting location θ_0 in parameter space, compute $P_0 = p(\theta_0|d)$
  - pick a candidate new location θ_c according to a proposal density $q(\theta_0, \theta_c)$
  - evaluate $P_c = p(\theta_c|d)$ and accept θ_c with probability $\alpha = \min\left(\frac{P_c}{P_0}, 1\right)$
  - if the candidate is accepted, add it to the chain and move there; otherwise stay at θ_0 and count this point once more.

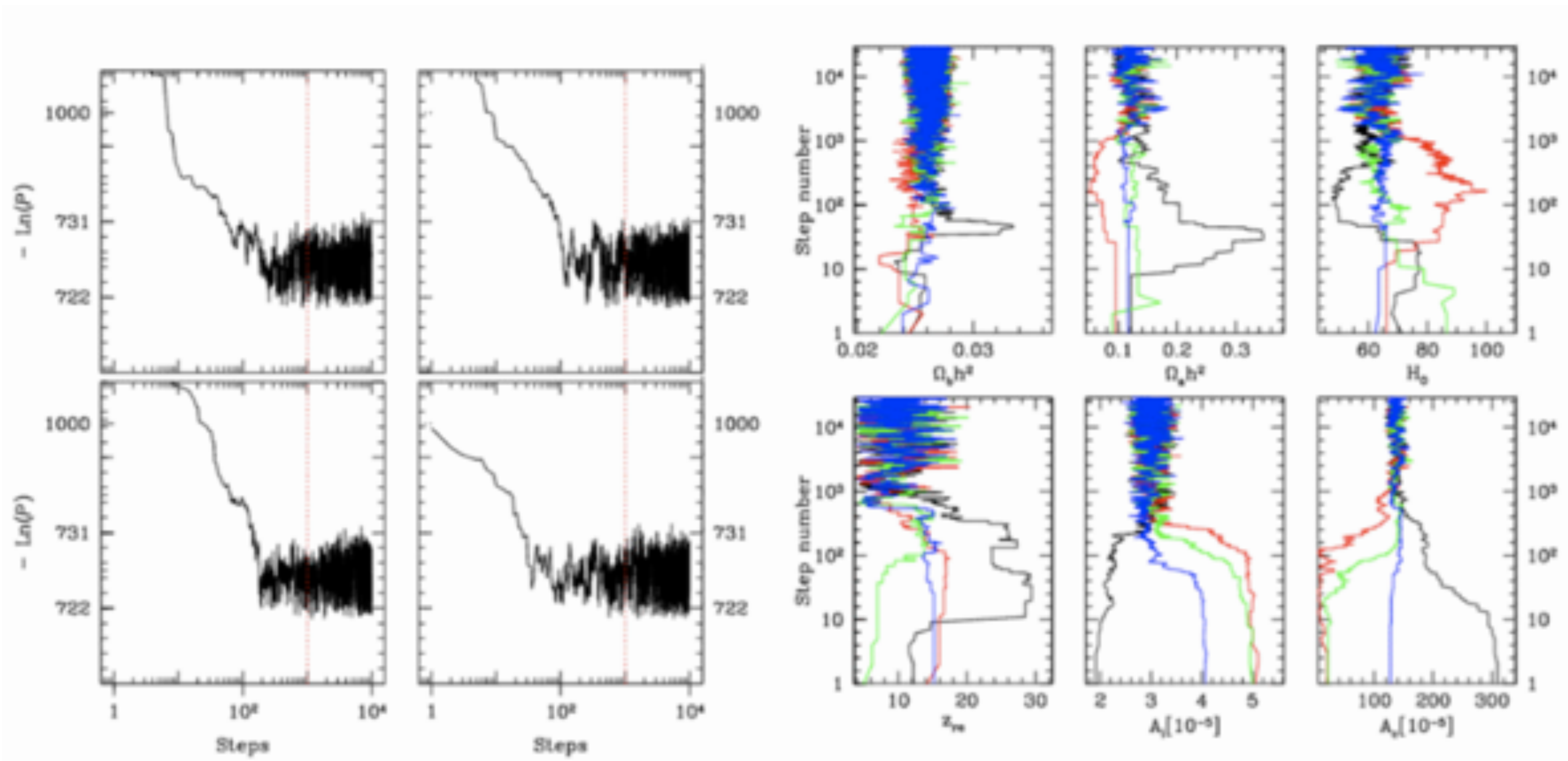
Practicalities

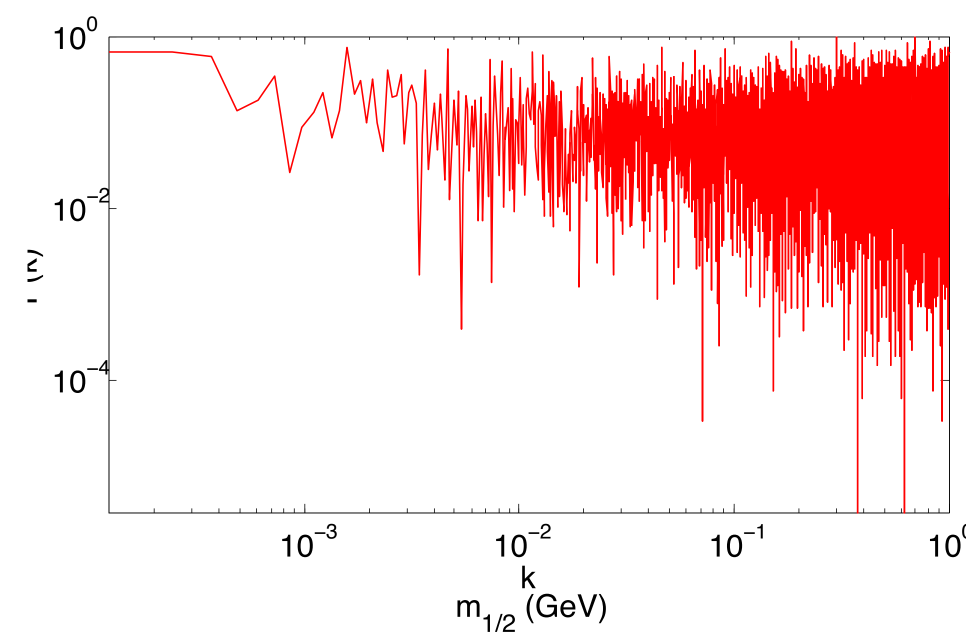
Imperial College London

- Except for simple problems, achieving good MCMC **convergence** (i.e., sampling from the target) and **mixing** (i.e., all chains are seeing the whole of parameter space) can be tricky
- There are several diagnostics criteria around but none is fail-safe. Successful MCMC remains a bit of a black art!
- Things to watch out for:
  - Burn in time
  - Mixing
  - Samples auto-correlation

MCMC diagnostics

Burn in   Mixing   Power spectrum



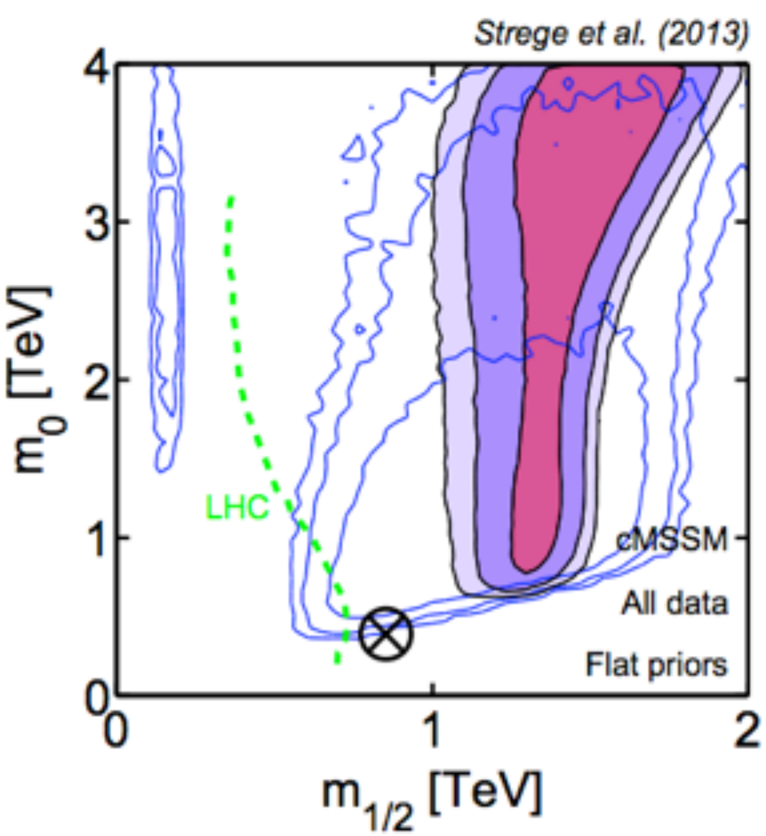


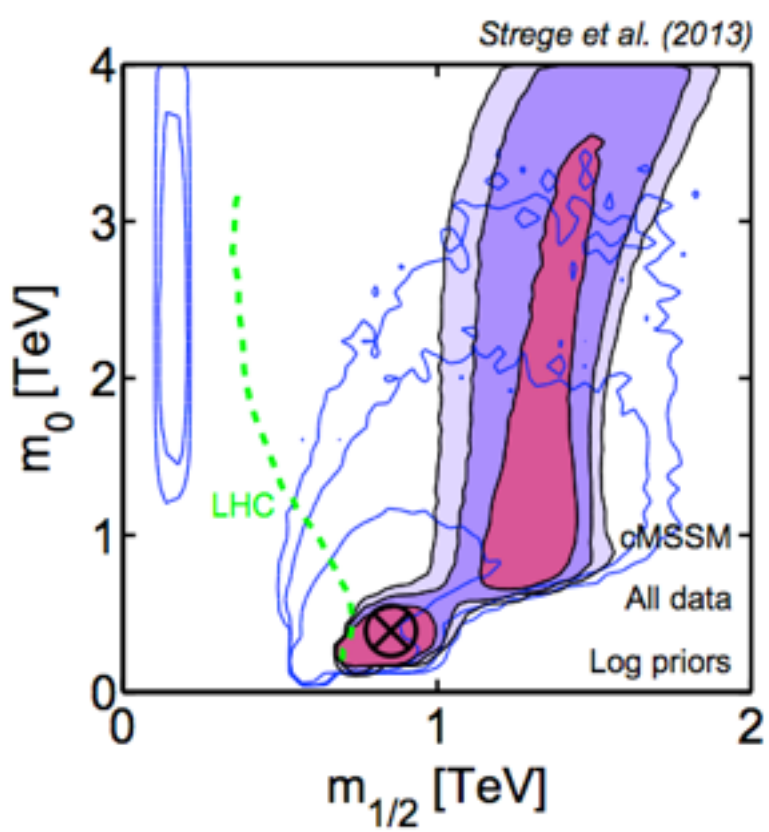
(see astro-ph/0405462 for details)

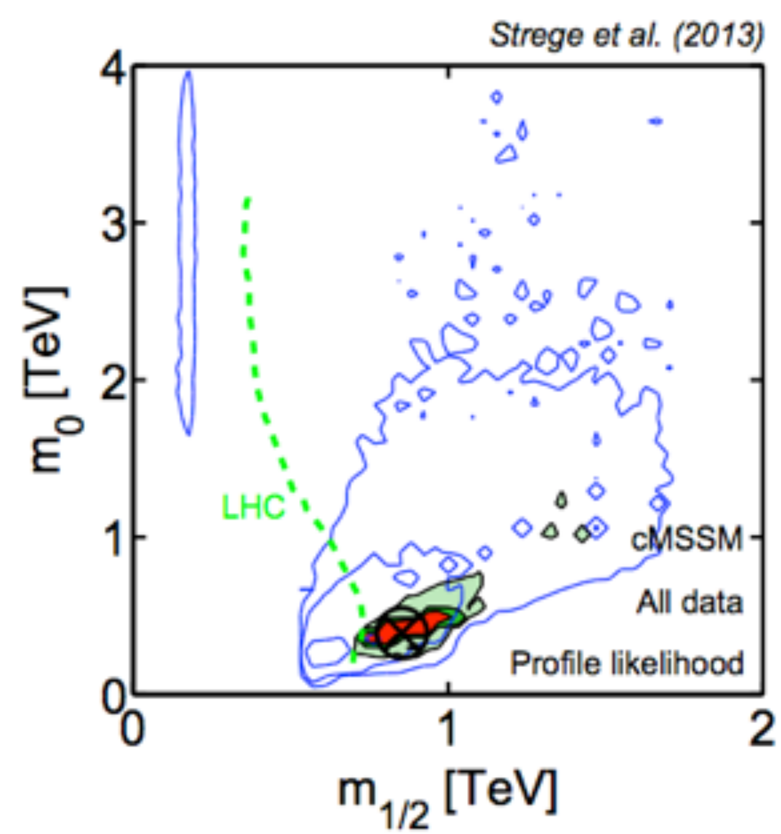
Non-Gaussian example

Imperial College London

Bayesian posterior ("flat priors")	Bayesian posterior ("log priors")	Profile likelihood





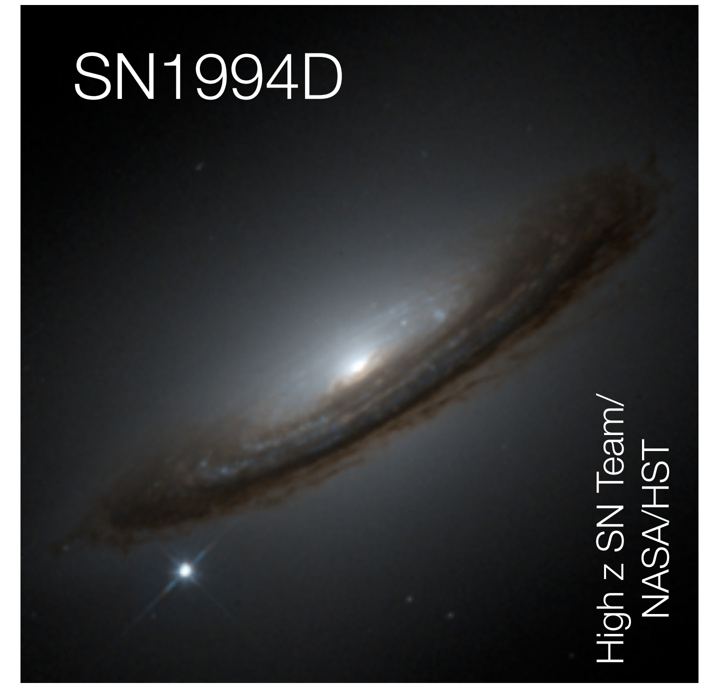


Constrained Minimal Supersymmetric Standard Model (4 parameters)
Strege, RT et al (2013)

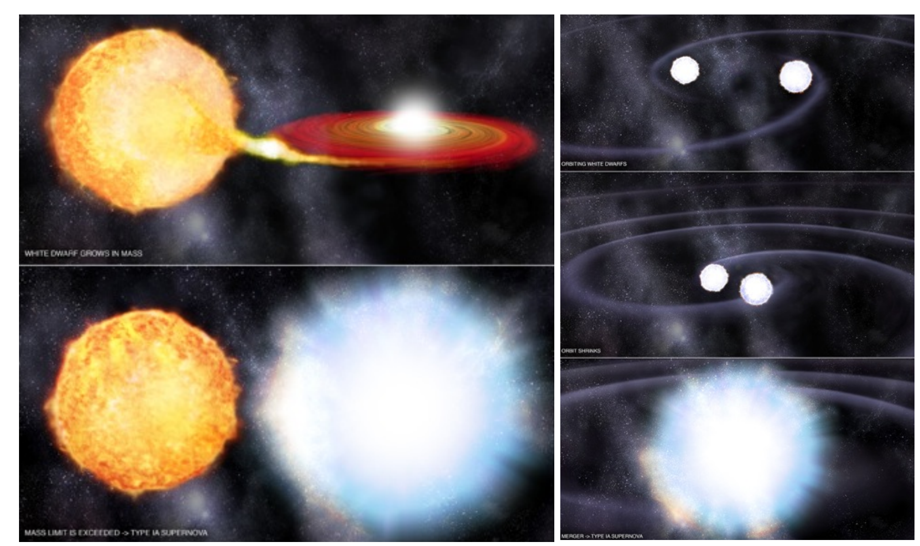
Supernovae Type Ia

Type Ia supernovae

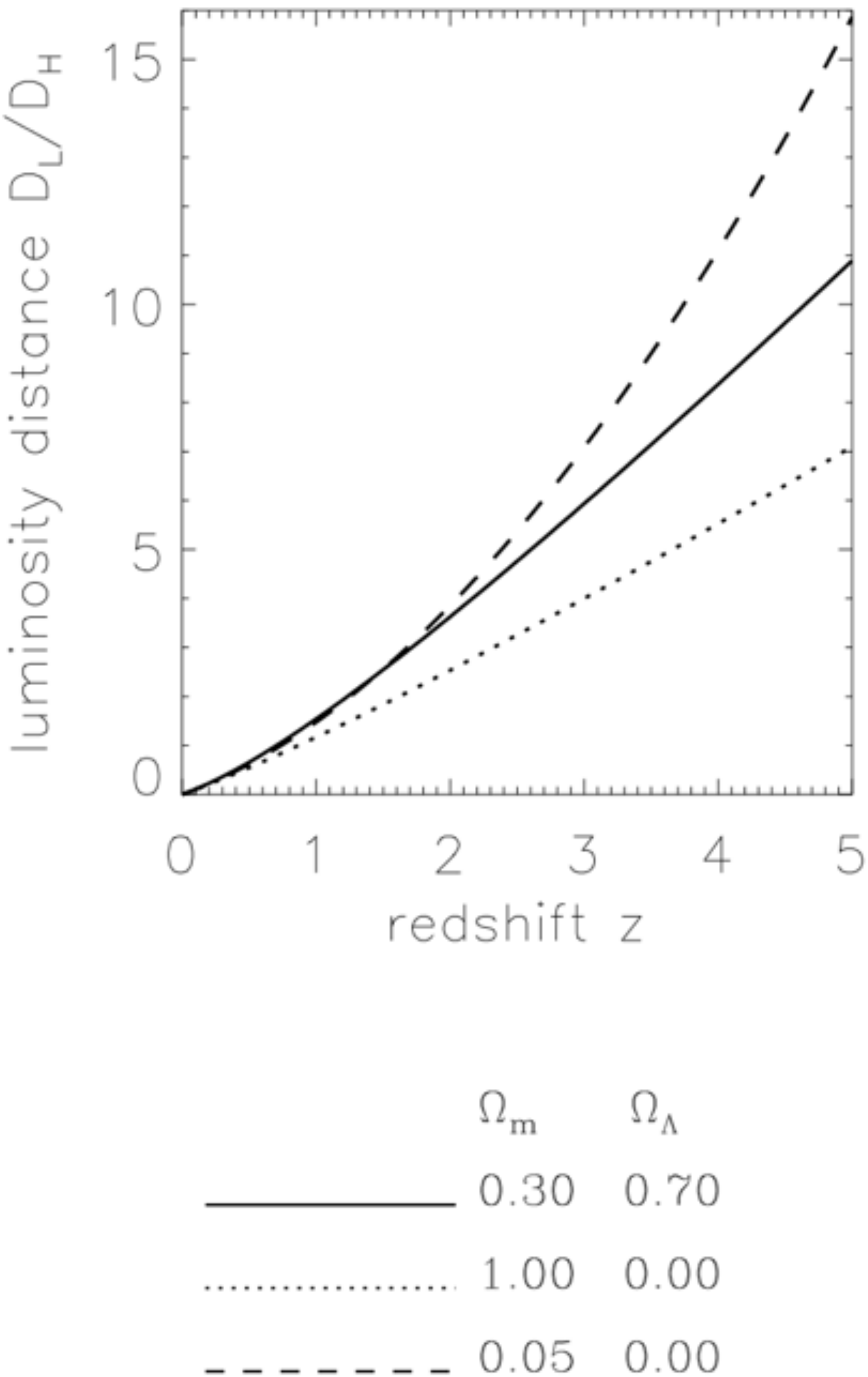
- **Supernovae:** core-collapse thermonuclear explosions of stars, emitting a large (~ 10^{51} erg, cf L_{galaxy} ~ 10^{44} erg/s) amount of energy (photons + neutrinos).
- **Supernovae type Ia (SNIa):** characterized by the lack of H in their spectrum, outcome of a CO white dwarf (WD) in a close binary system accreting mass above the Chandrasekhar limit (1.4 solar masses).
- The nature of the **donor star** is still disputed: *Single Degenerate* (WD + Main sequence or Red giant or a He star companion) vs *Double Degenerate* (WD + WD merger) scenarios (or both)

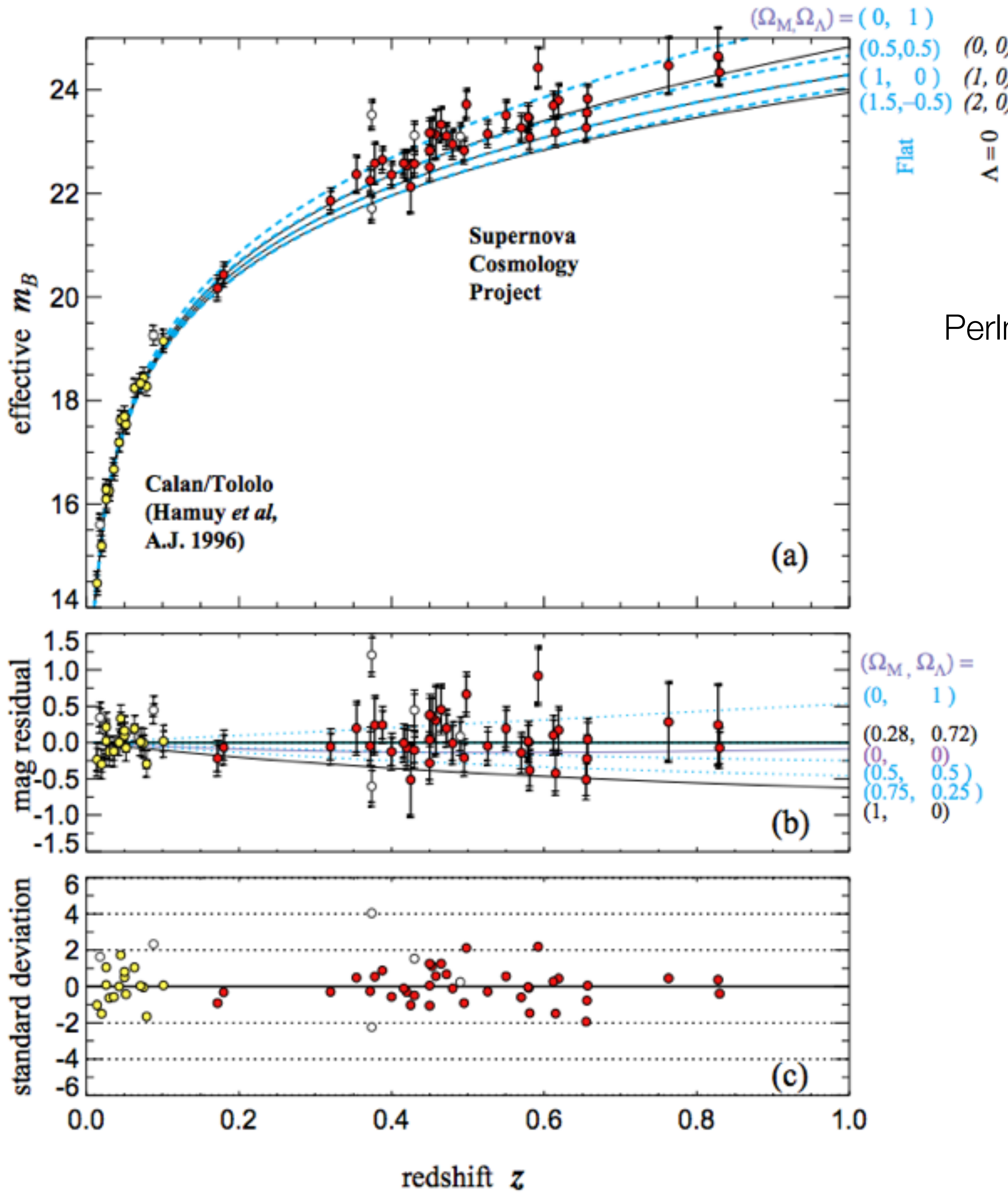


High z SN Team/ NASA/HST

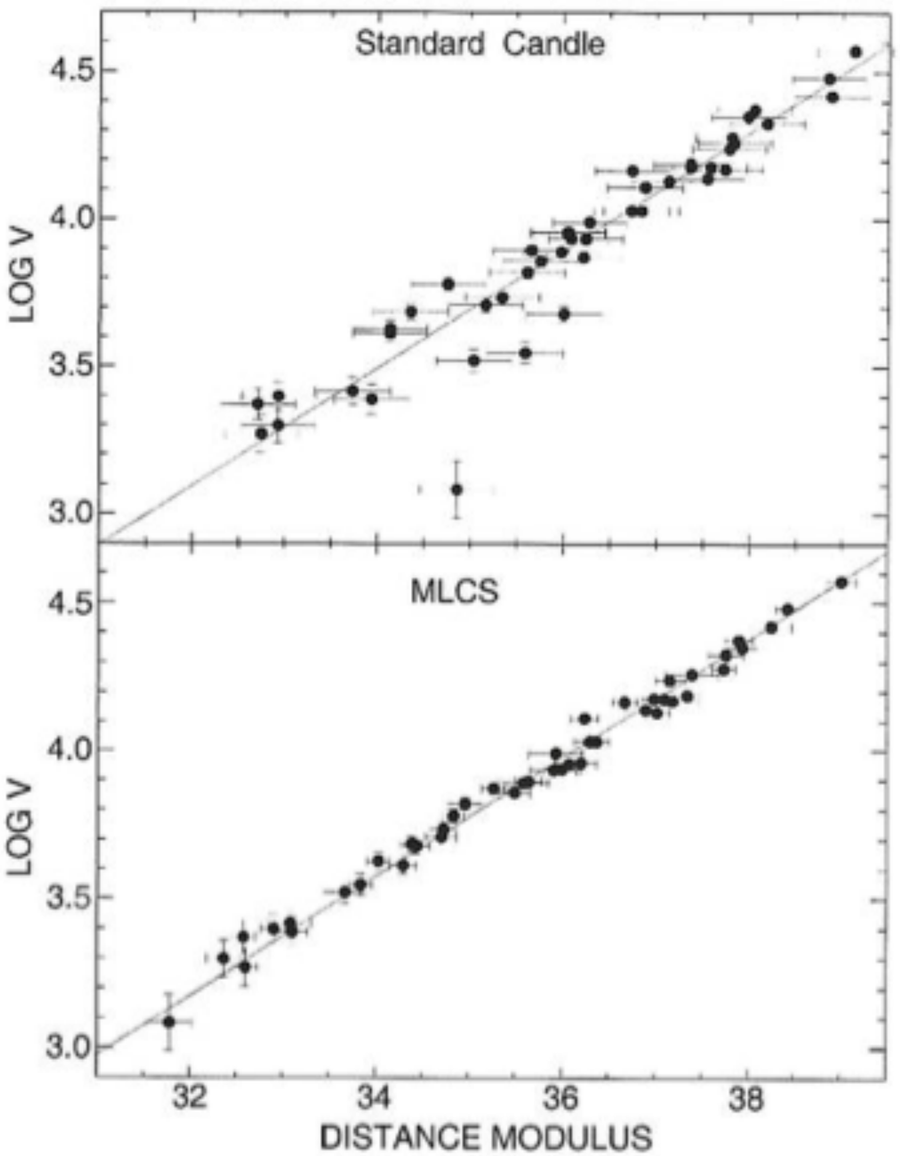


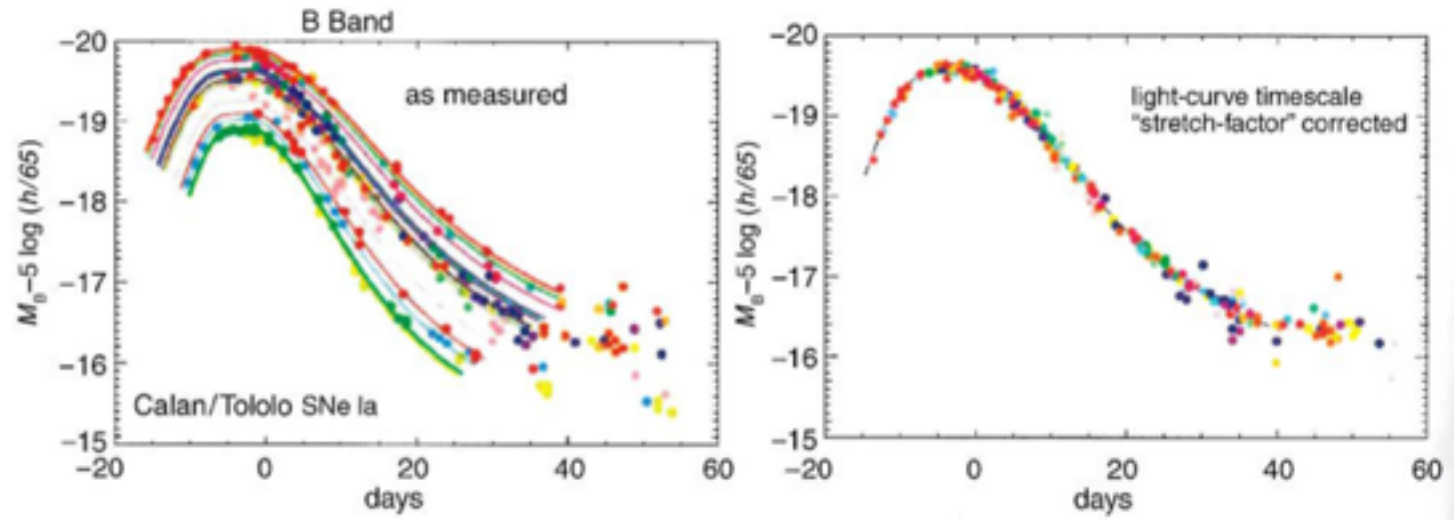
NASA/CXC/M. Weiss

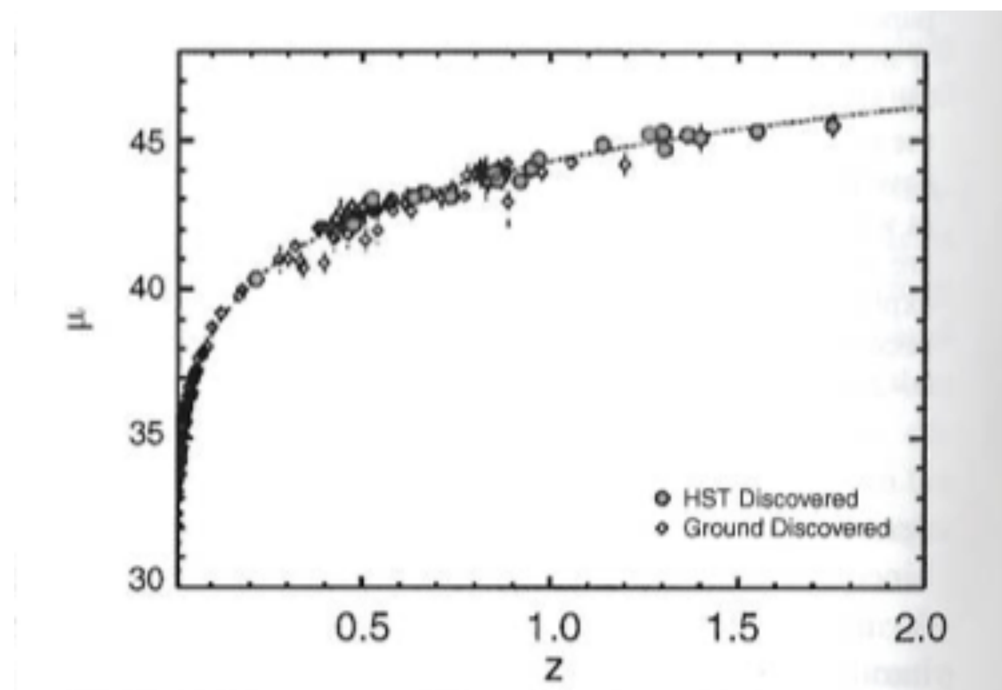

luminosity distance D_L/D_H
0
5
10
15
0
1
2
3
4
5
redshift z
Ω_m Ω_Λ
0.30 0.70
1.00 0.00
0.05 0.00



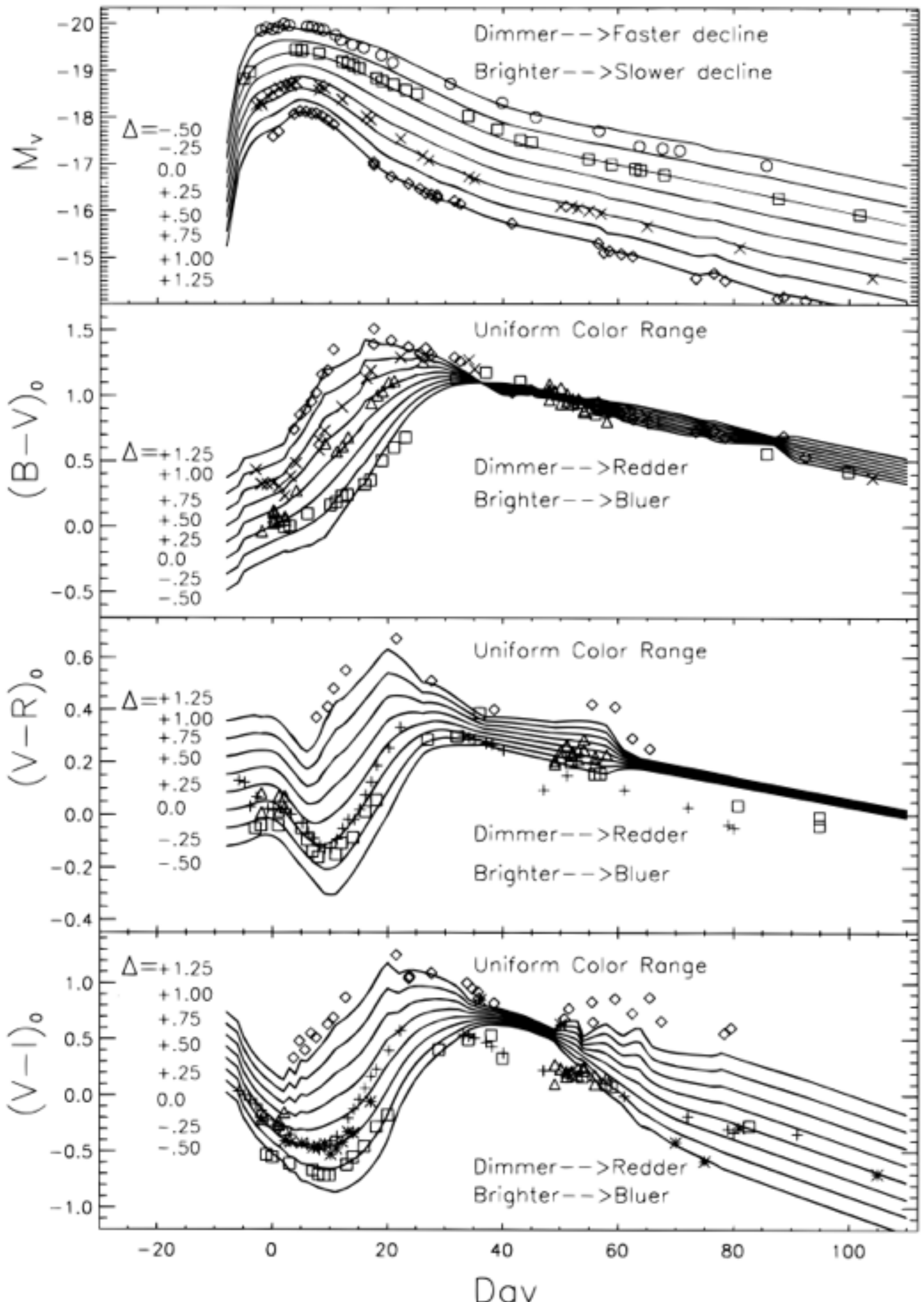
Perlmutter et al (1999)







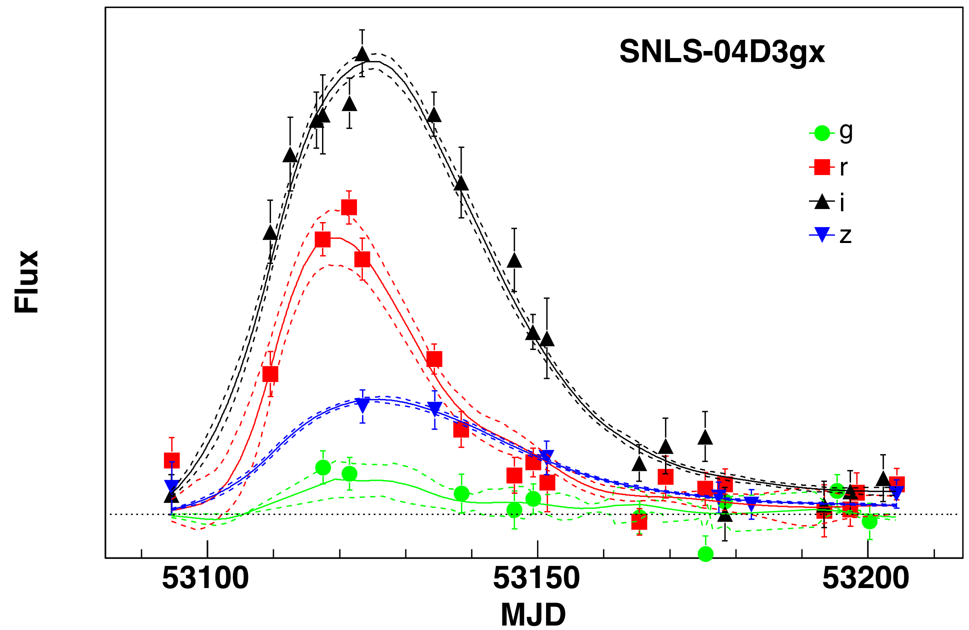
For references, see Chapter 8 in Schneider, Extragalactic Astronomy and Cosmology: An Introduction, Springer (2006).



Riess, Press, Kirshner (1996)

SNIa lightcurves

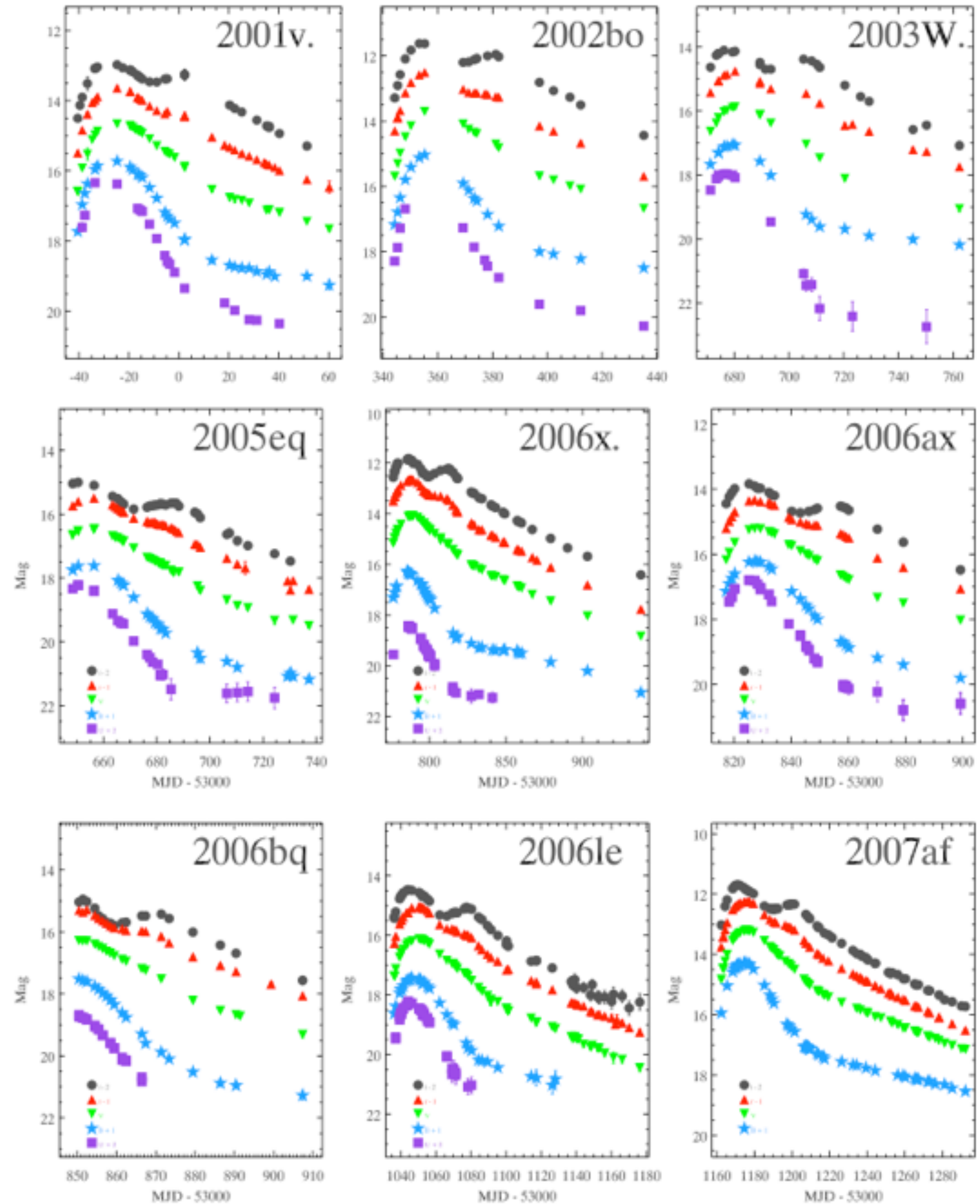
SNLS



Guy et al (2007)

CfA3
185 multi-band optical nearby SNIa

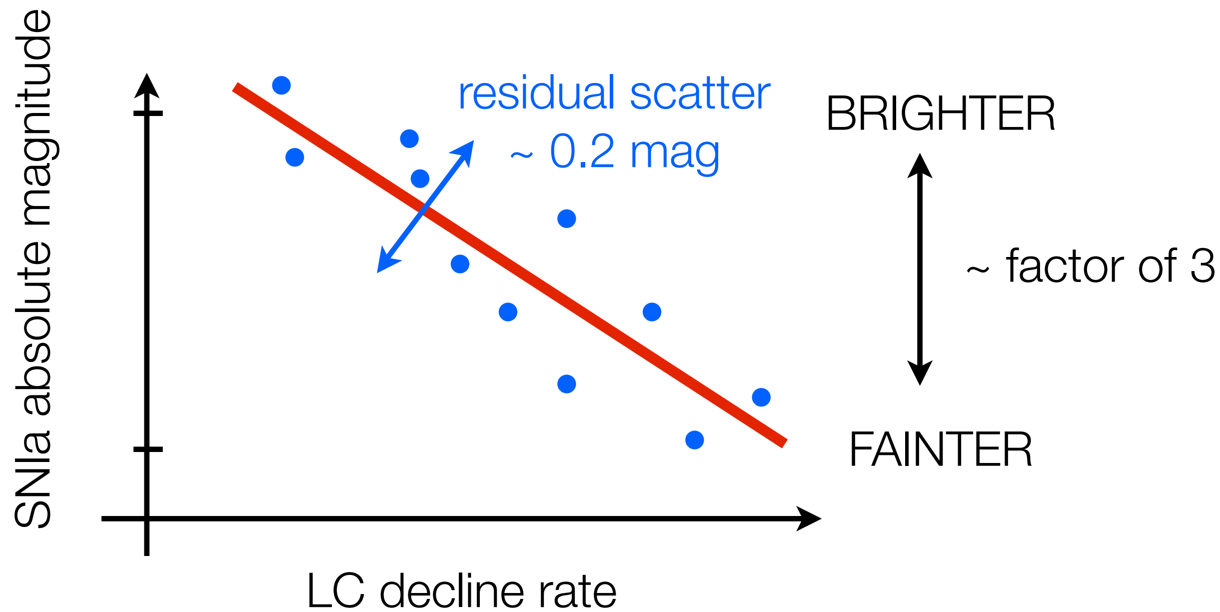
Imperial College

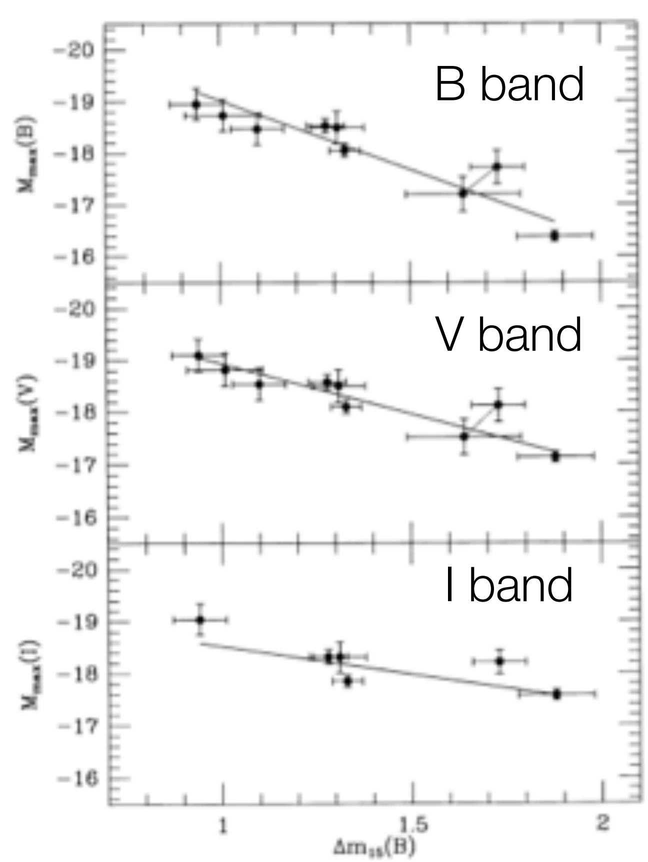


Hicken et al (2009)

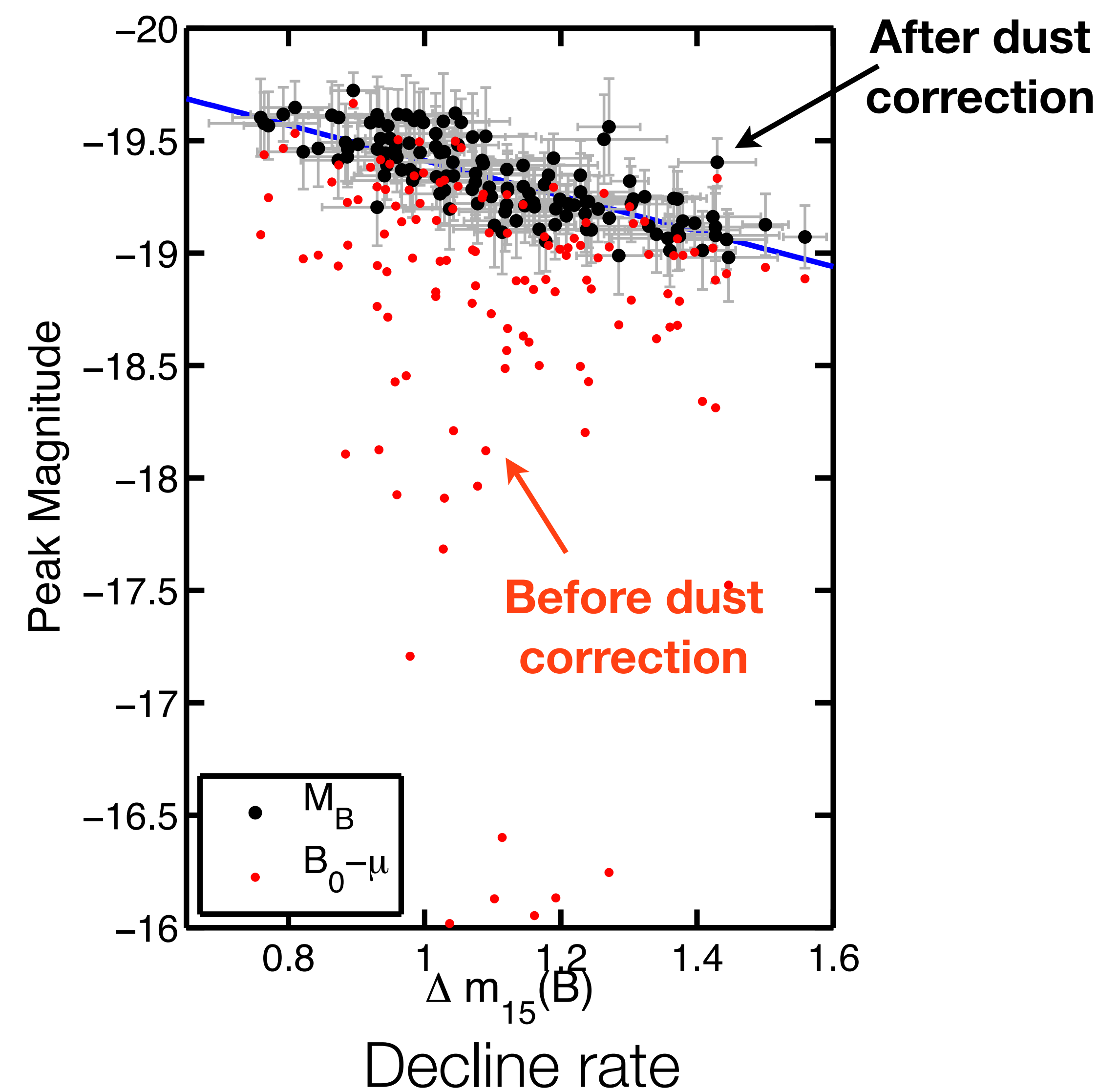
Brightness-width relationship

Imperial College London





Phillips (1993)

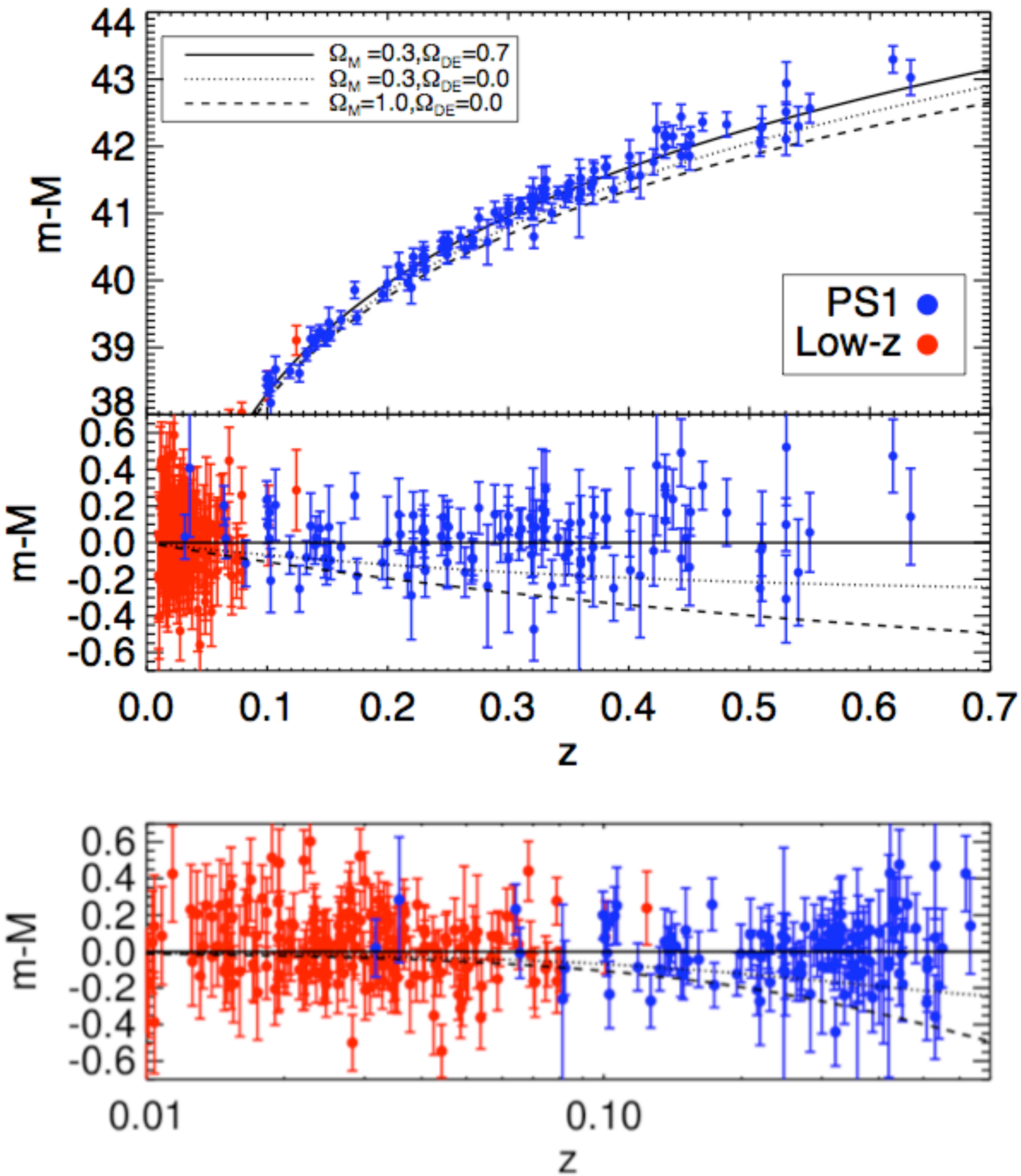


Mandel et al (2011)

Brighter SNIa are slow decliners

PS1 data

- Most recent data set from PAN-STARRS1 survey
- 146 spectroscopically confirmed SNIa
- Cosmological fit: 112 PS1 at high-z (blue) + 201 low-z SNIa (red)



Standard Chi² fits of SALT2 output

Imperial College London

- Standard analysis minimizes the likelihood (typically, C minimized with α, β fixed, then α, β minimized with C fixed), arbitrarily defined as:

observed values (SALT2 fits)

parameters

$$-2\log\mathcal{L} = \chi^2 = \sum_i \frac{\left(\mu(z_i,\mathcal{C}) - [\hat{m}_{B,i} - M + \alpha\hat{x}_{1,i} - \beta\hat{c}_i]\right)^2}{\sigma^2_{\rm int} + \sigma^2_{\rm fit}}$$

$$\sigma^2_{\rm fit} = \sigma^2_{m_B} + \alpha^2\sigma^2_{x_1} + \beta^2\sigma^2_c + \text{correlations}$$

$\sigma^2_{\rm int}$ represents the "intrinsic" (residual) scatter

determined by requiring Chi²/dof ~ 1

ICIC

Problems of the standard analysis

Imperial College London

$$-2\log\mathcal{L} = \chi^2 = \sum_i \frac{\left(\mu(z_i,\mathcal{C}) - [\hat{m}_{B,i} - M + \alpha\hat{x}_{1,i} - \beta\hat{c}_i]\right)^2}{\sigma^2_{\rm int} + \sigma^2_{\rm fit}}$$

- Form of the likelihood function is unjustified
- α, β appear in the variance, too - this is a problem of simultaneous estimation of the mean and of the variance. Chi^2 not the correct distribution.
- Incorrectly normalized - missing $-\frac{1}{2}\log\left(\sigma^2_{\rm int} + \sigma^2_{\rm fit}\right)$ term in front. Adding this in results in a (known) 6-sigma bias of β.
- Chi^2/dof ~ 1 prescription prevents by construction model checking and hypothesis testing
- Marginalization (and use of fast Bayesian MCMC methods) impossible (profile likelihood “fudge” necessary)

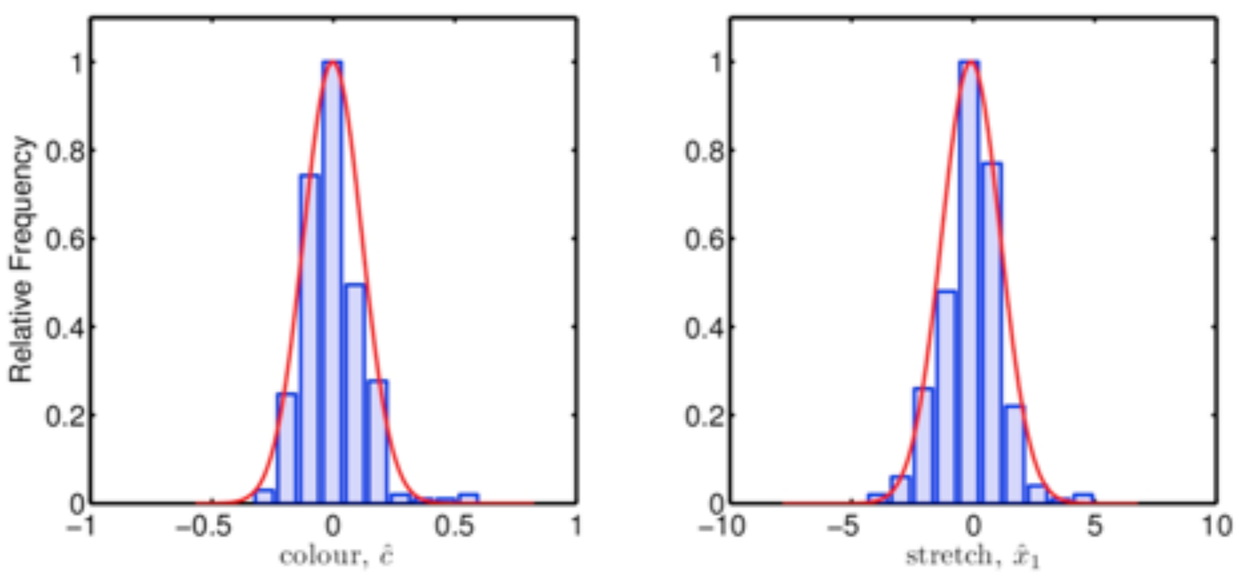
Principled Bayesian solution required!

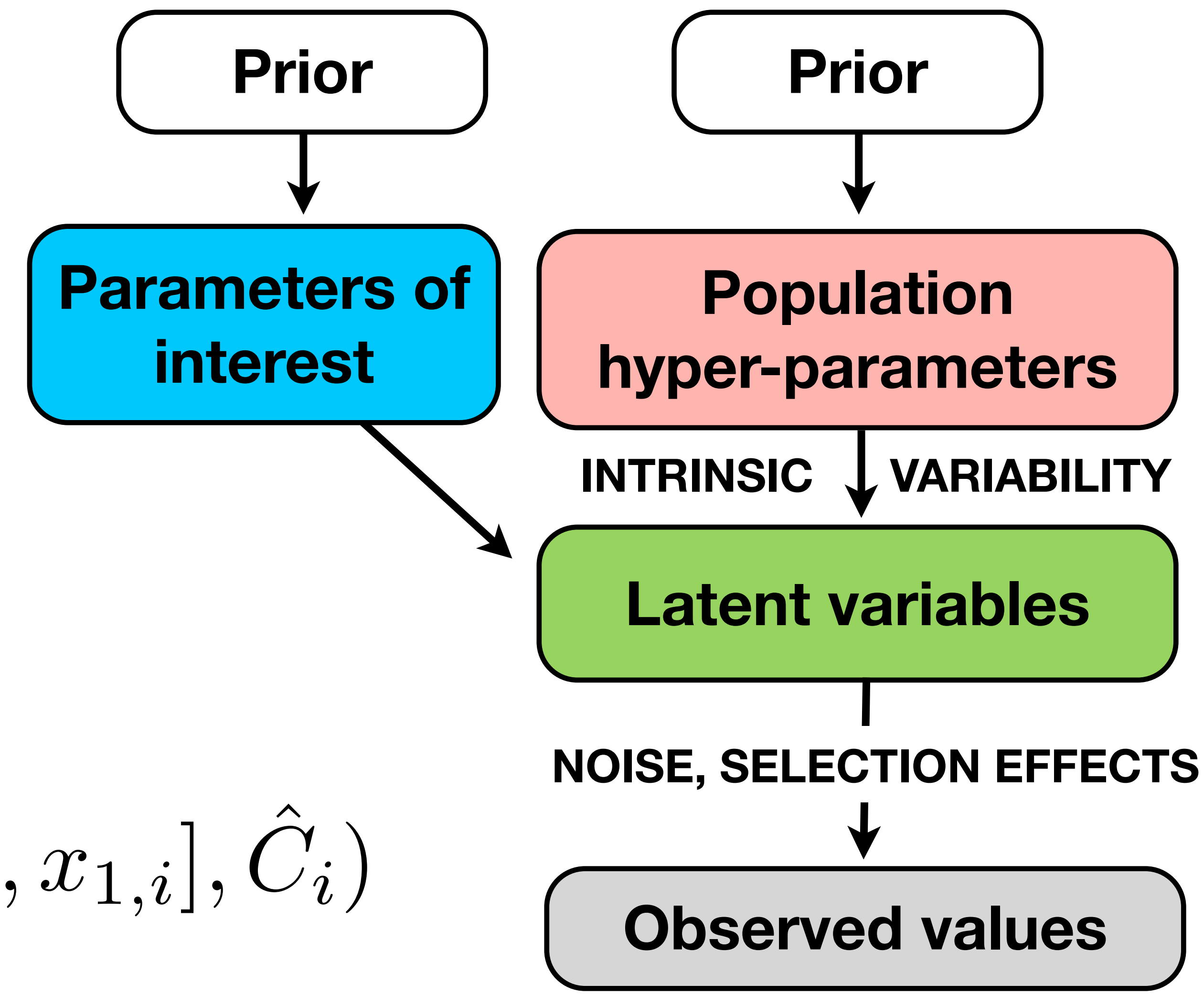
ICIC

Bayesian hierarchical model

For each SNIa, this relation holds **exactly** between **latent** (unobserved) variables:

$$x_{1,i} - \beta c_i$$





$$c_i \sim \mathcal{N}(c_\star, R_c)$$

$$x_{1,i} \sim \mathcal{N}(x_\star, R_x)$$

$$[\hat{m}_{B,i}, \hat{c}_i, \hat{x}_{1,i}] \sim \mathcal{N}([m_{B_i}, c_i, x_{1,i}], \hat{C}_i)$$

Advantages of multi-layer model

Imperial College London

- The Bayesian hierarchical approach allows us to:
  - model explicitly the **population-level** intrinsic variability of SNIa
  - investigate the impact of multiple SNIa populations (e.g., different progenitor models)
  - determine/include correlations with other observables (galaxy mass, metallicity, age, spectral lines, etc) to reduce residual scatter in Hubble diagram
  - obtain a **principled data likelihood** that can be used with Bayesian MCMC/ MultiNest (marginal posteriors, Bayesian evidence for model selection)
  - derive a fully marginalized posterior on the residual (after colour and stretch correction) **intrinsic scatter** in the SNIa intrisic magnitude
  - investigate possible **SNIa evolution** (e.g., $\beta(z)$) and other systematics

At the heart of the method...

Imperial College London

- ... lies the fundamental problem of **linear regression** in the presence of measurement errors on both the dependent and independent variable and intrinsic scatter in the relationship (e.g., Gull 1989, Gelman et al 2004, Kelly 2007):

$$\mu_i = m_{B,i} - M_i + \alpha x_{1,i} - \beta c_i$$

analogous to

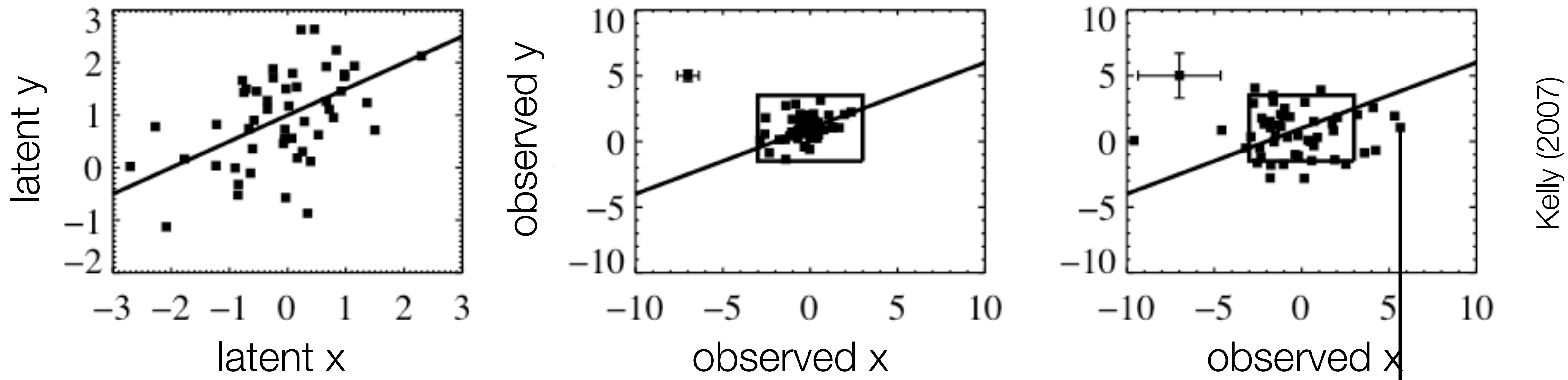
$$y_i = b + a x_i$$

$$x_i \sim p(x|\Psi) = \mathcal{N}_{x_i}(x_\star, R_x)$$ POPULATION DISTRIBUTION

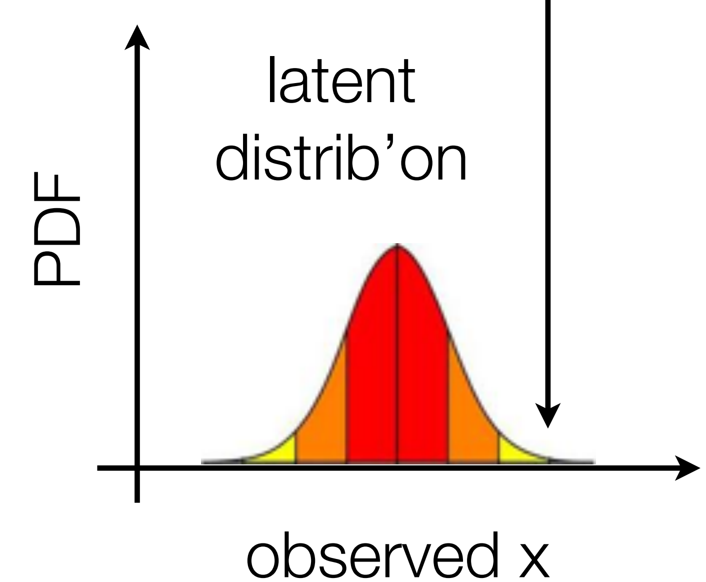
$$y_i|x_i \sim \mathcal{N}_{y_i}(b + a x_i, \sigma^2)$$ INTRINSIC VARIABILITY

$$\hat{x}_i, \hat{y}_i | x_i, y_i \sim \mathcal{N}_{\hat{x}_i,\hat{y}_i}([x_i, y_i], \Sigma^2)$$ MEASUREMENT ERROR

INTRINSIC VARIABILITY **+ MEASUREMENT ERROR**



- Modeling the latent distribution of the independent variable accounts for “Malmquist bias”
- An observed x value far from the origin is more probable to arise from up-scattering (due to noise) of a lower latent x value than down-scattering of a higher (less probable) x value

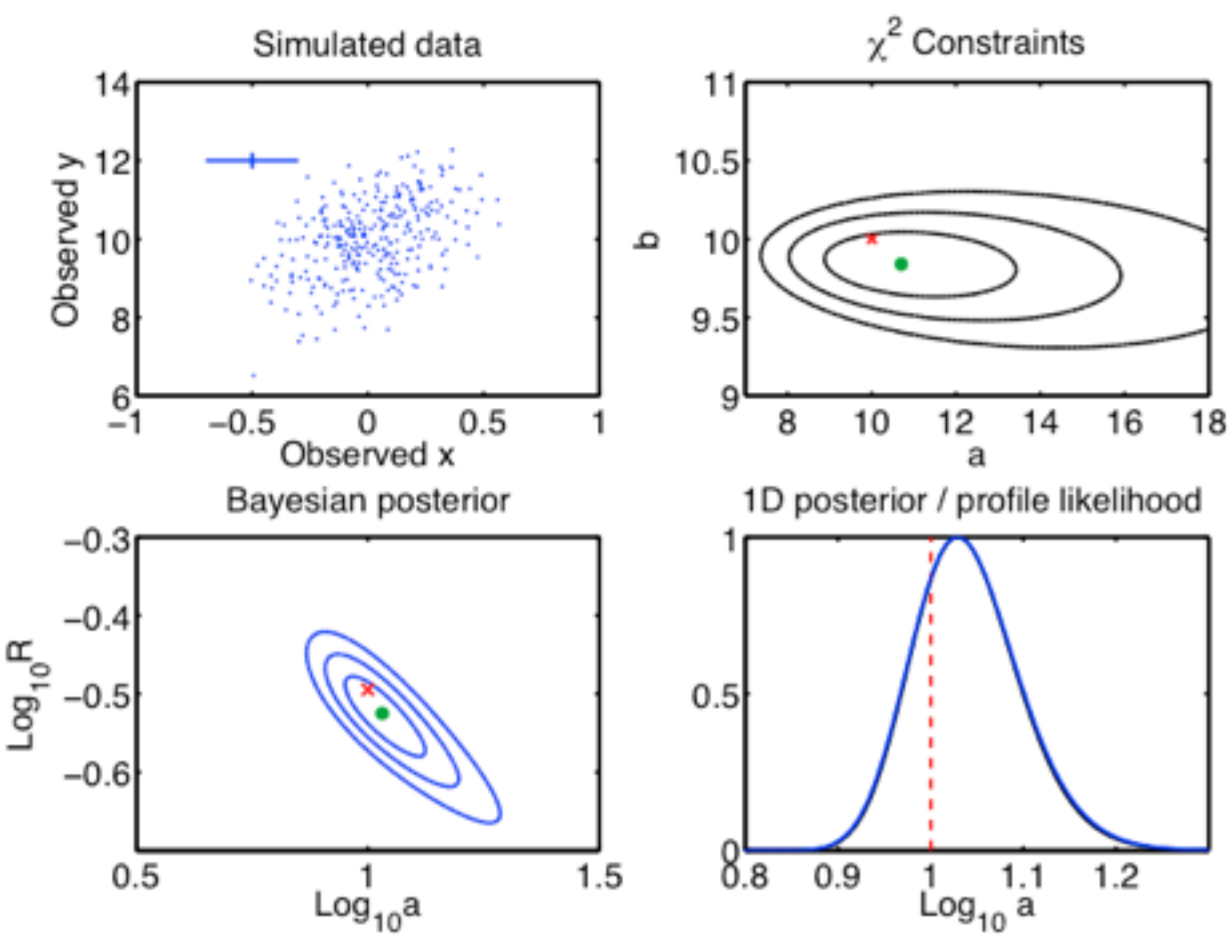


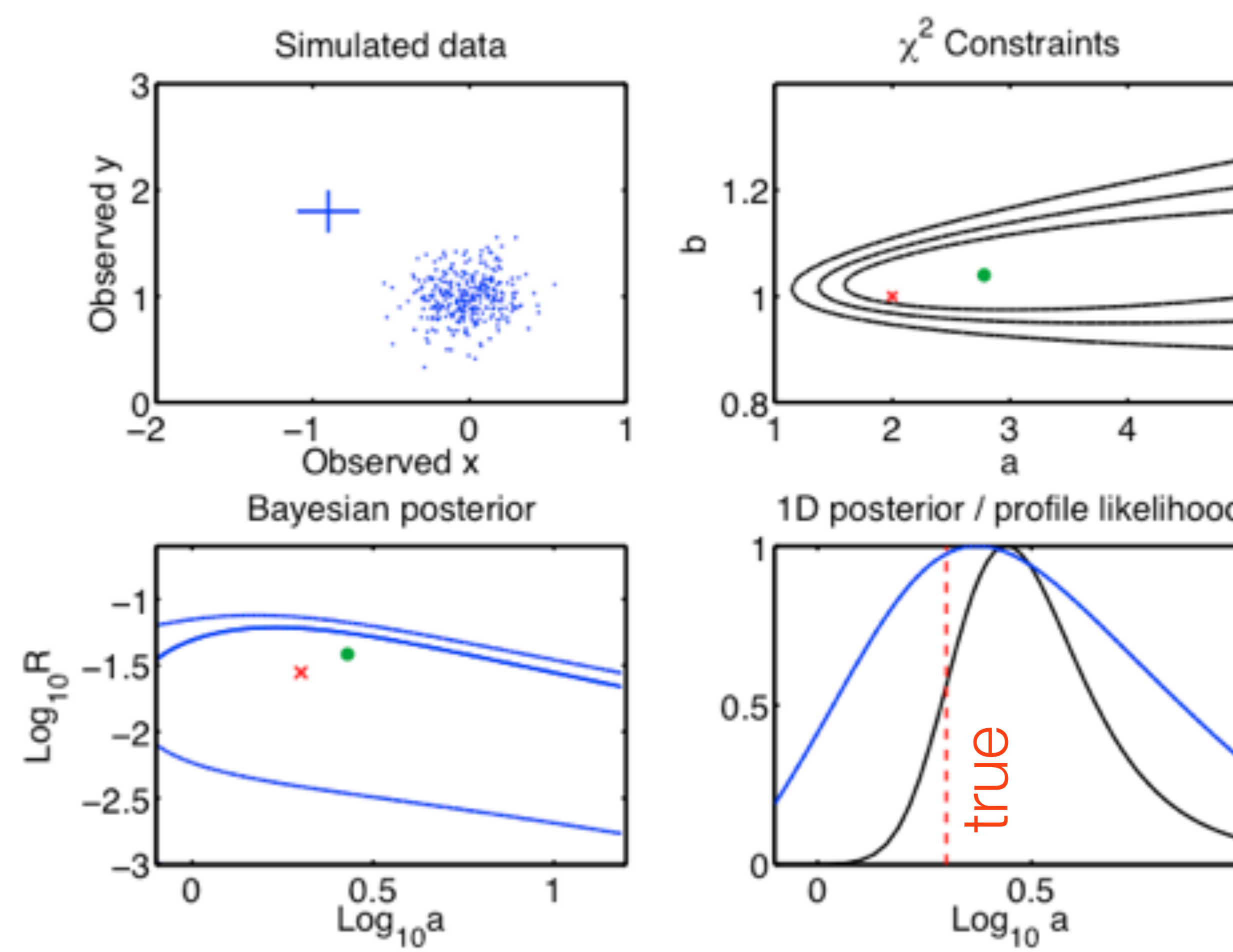
The key parameter is noise/population variance $\sigma_x \sigma_y / R_x$

$\sigma_x \sigma_y / R_x$ small

$$y_i = b + a x_i$$

$\sigma_x \sigma_y / R_x$ large





Bayesian marginal posterior identical to profile likelihood

Bayesian marginal posterior broader but less biased than profile likelihood

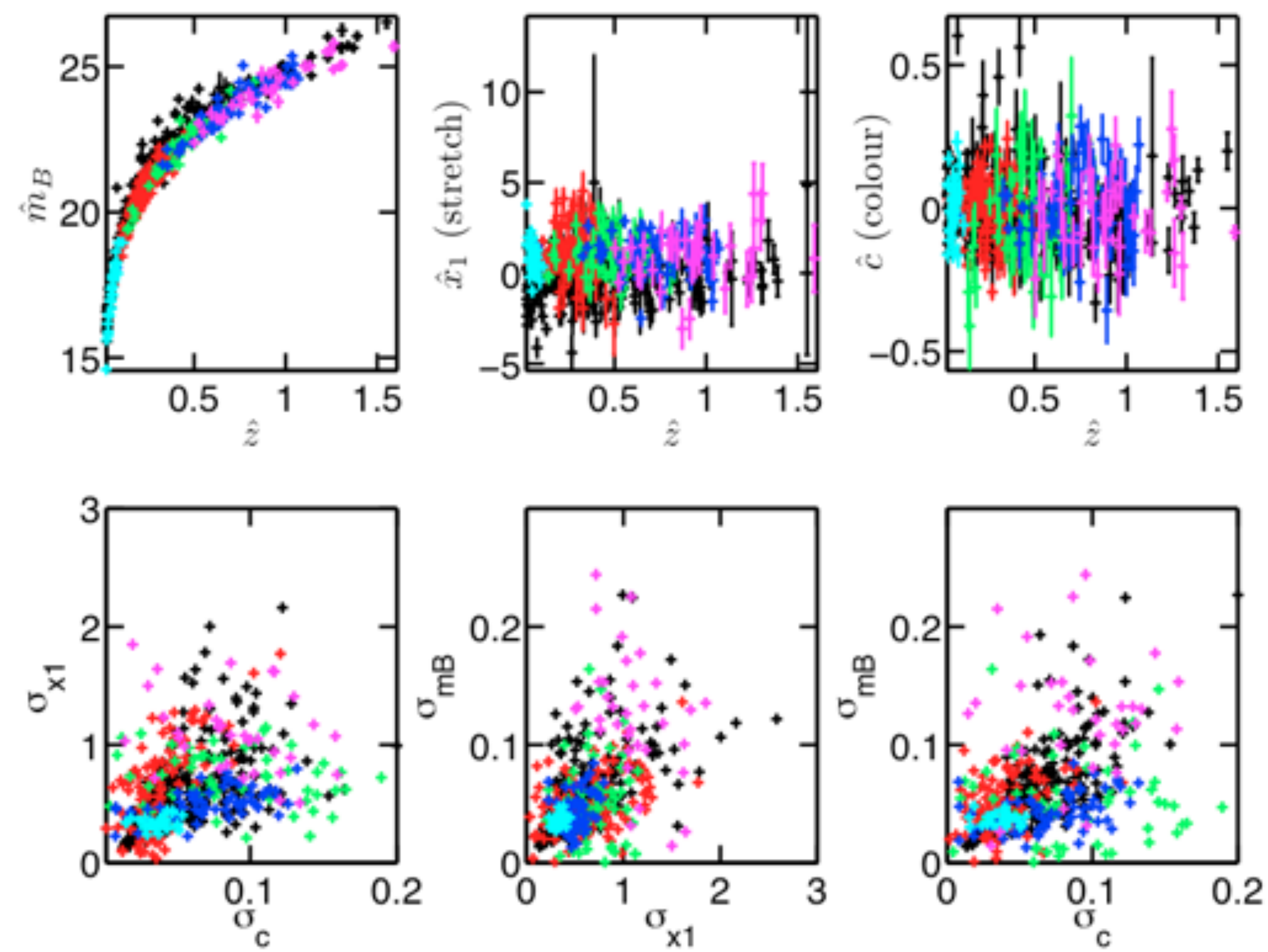
March, RT et al (2011)

Tests on simulated SNIa data

Imperial College London

- Simulated N=288 SNIa with similar characteristics as SDSS +ESSENCE+SNLS+HST +Nearby sample
- Reconstruction of cosmological parameters over 100 realizations, comparing Bayesian hierarchical method with standard Chi^2.

Simulated SNIa realization (colour coded according to "survey")



March et al (2011)

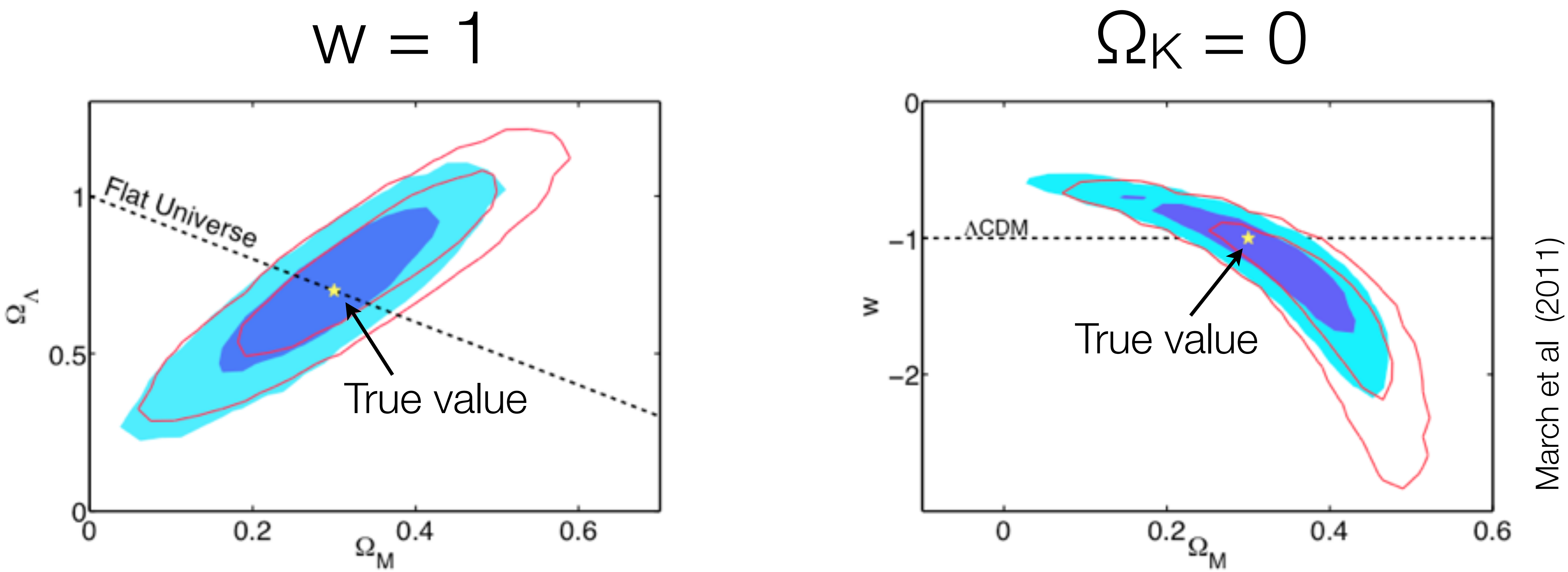
Posterior sampling

Imperial College London

- In the Bayesian hierarchical approach, we have
  - 3 cosmological parameters: H_0, Ω_M, Ω_K (w=1) or H_0, Ω_M, w (Ω_K =0)
  - 2 stretch/colour correction parameters: α, β
  - 6 population-level parameters: M_0, σ^2, x^*, R_x, c^*, R_c
  - 3N (=864) latent variables M_i, x_{1i}, c_i
- Analytical marginalization over all latent variables and linear population-level parameters is possible in Gaussian case (no selection effects). Sampling of the remaining parameters via MultiNest.
- Alternatively, Gibbs sampling can be used to sample over all parameters (conditional distributions are Gaussian in the absence of selection effects. Including them introduces additional accept/reject step).

Marginal posterior (simulated data)

Imperial College London



Red/empty: Chi^2 (68%, 95% CL)

Blue/filled: Bayesian (68%, 95% credible regions)

Bayesian posterior is noticeably different from the Chi^2 CL: which one is “best”?

ICIC

Coverage, bias and mean squared error

Imperial College London

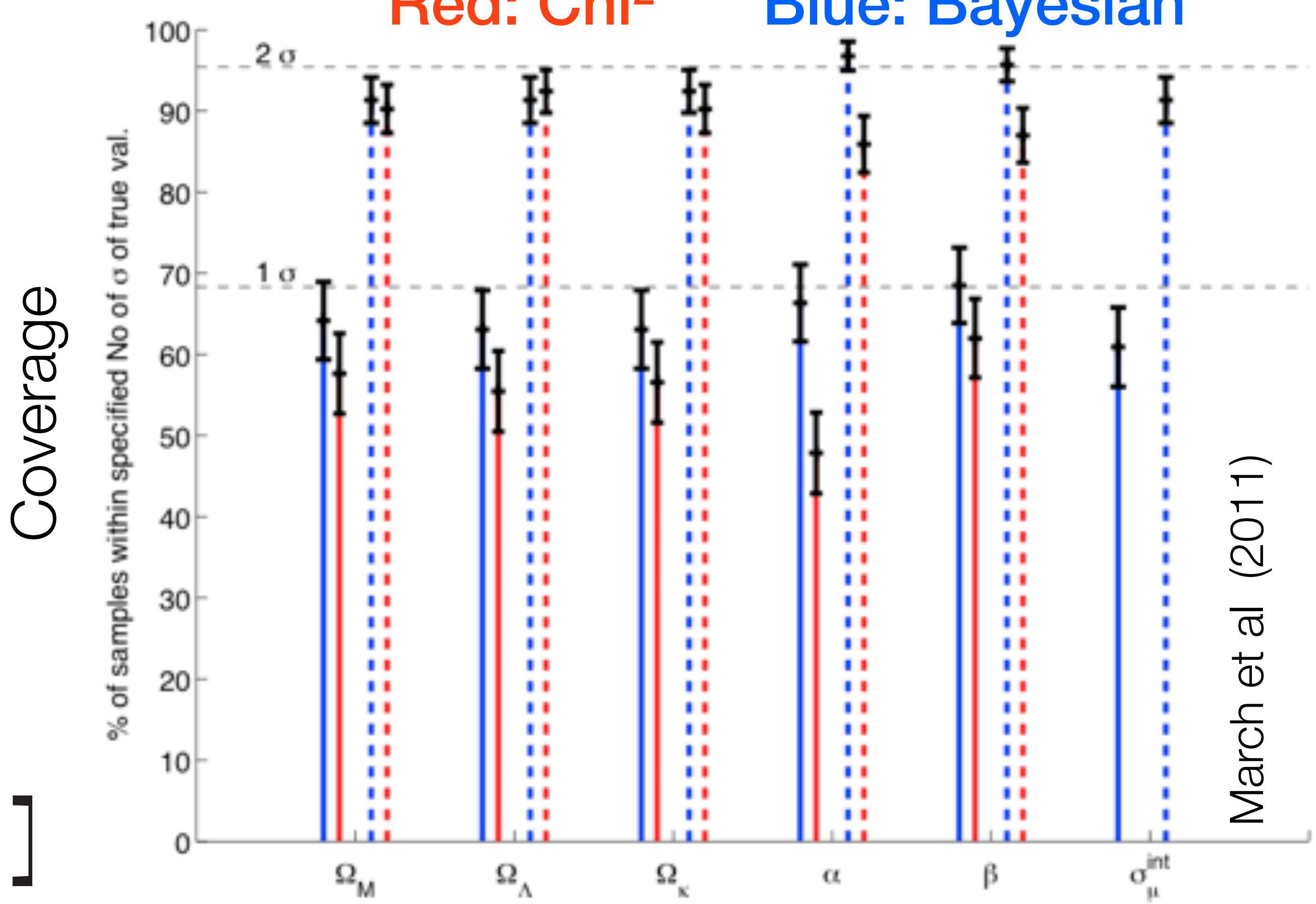
- Coverage of Bayesian 1D marginal posterior CR and of 1D Chi^2 profile likelihood CI computed from 100 realizations
- Bias and mean squared error (MSE) defined as

$$\text{bias} = \langle \hat{\theta} - \theta_{\text{true}} \rangle$$

$$\text{MSE} = \text{bias}^2 + \text{Var}$$

$\hat{\theta}$ is the posterior mean (Bayesian) or the maximum likelihood value (Chi^2).

Red: Chi^2 Blue: Bayesian



March et al (2011)

Results:

Coverage: generally improved (but still some undercoverage observed)

Bias: reduced by a factor ~ 2-3 for most parameters

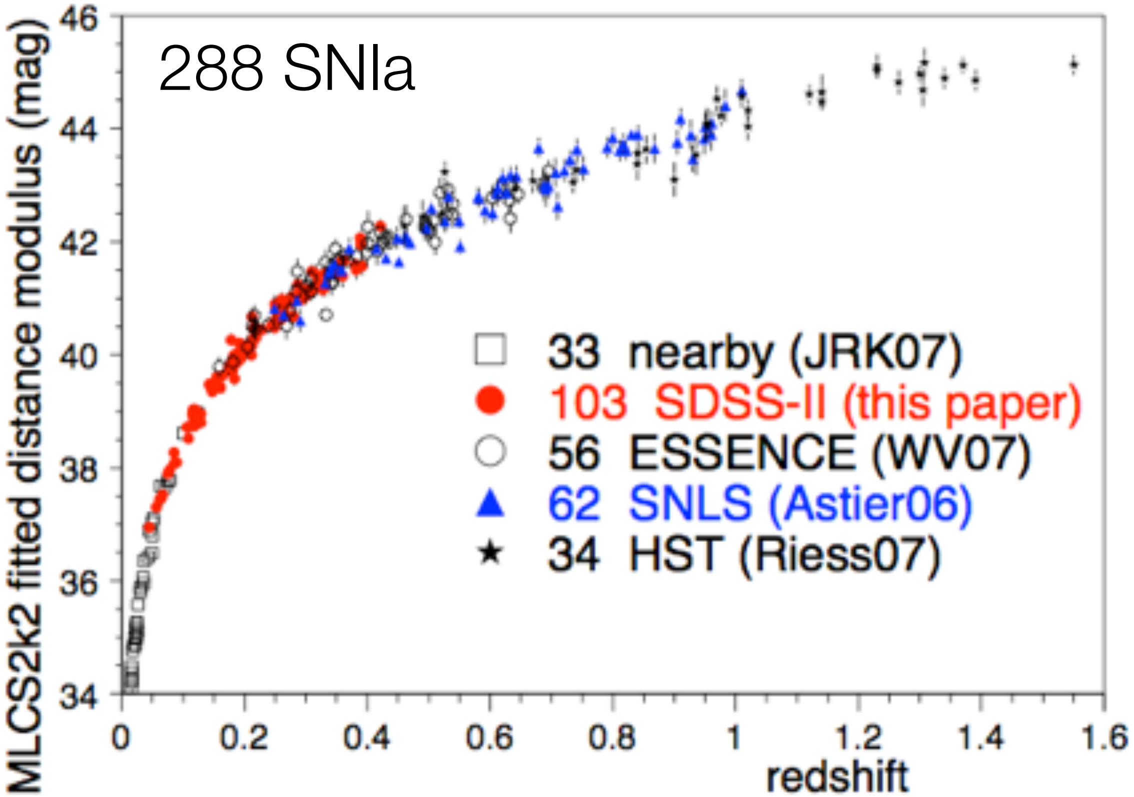
MSE: reduced by a factor 1.5-3.0 for all parameters

Cosmology results

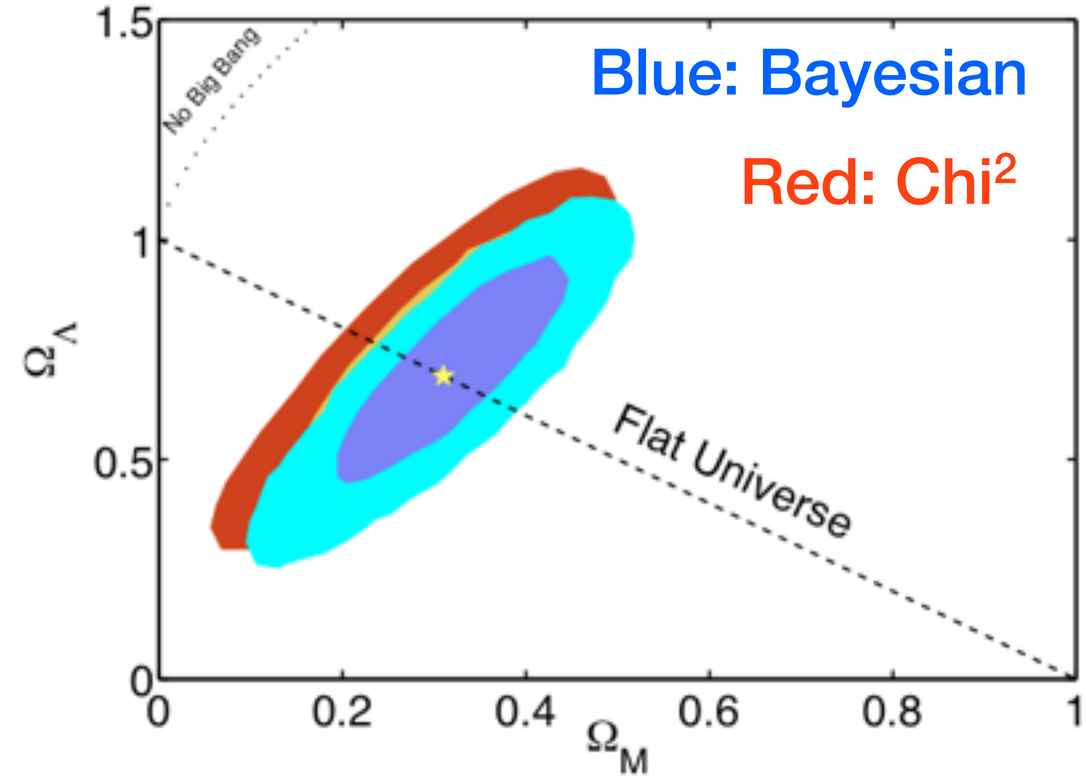
Imperial College London

Combined sample

Kessler et al (SDDS collaboration) (2010)

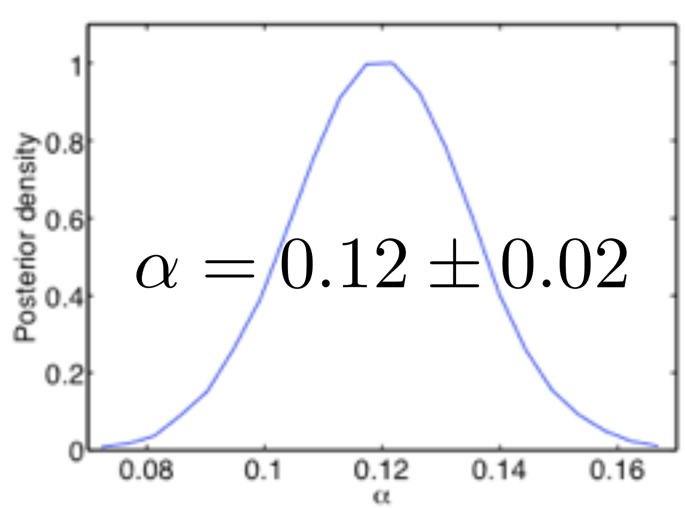


w = 1

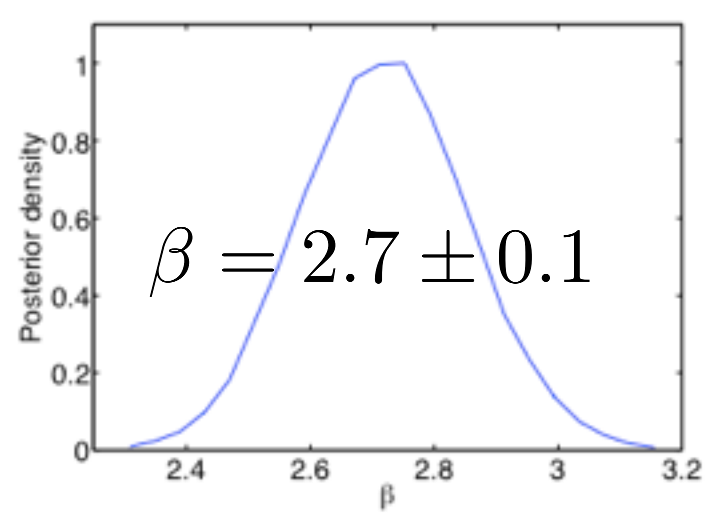


March et al (2011)

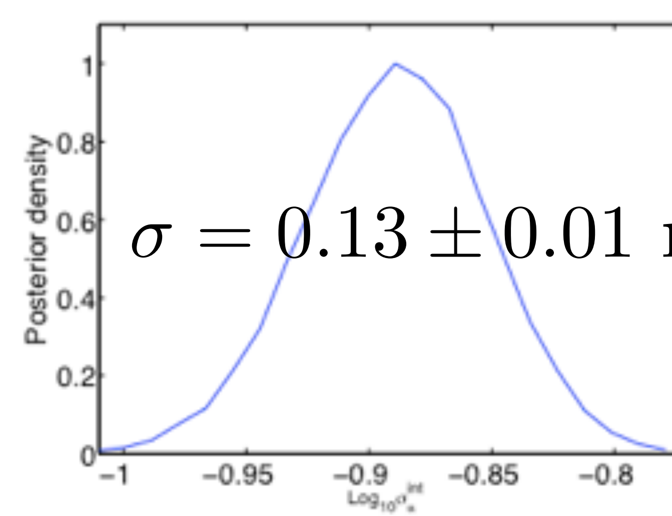
Marginal posteriors



$$\alpha = 0.12 \pm 0.02$$



$$\beta = 2.7 \pm 0.1$$

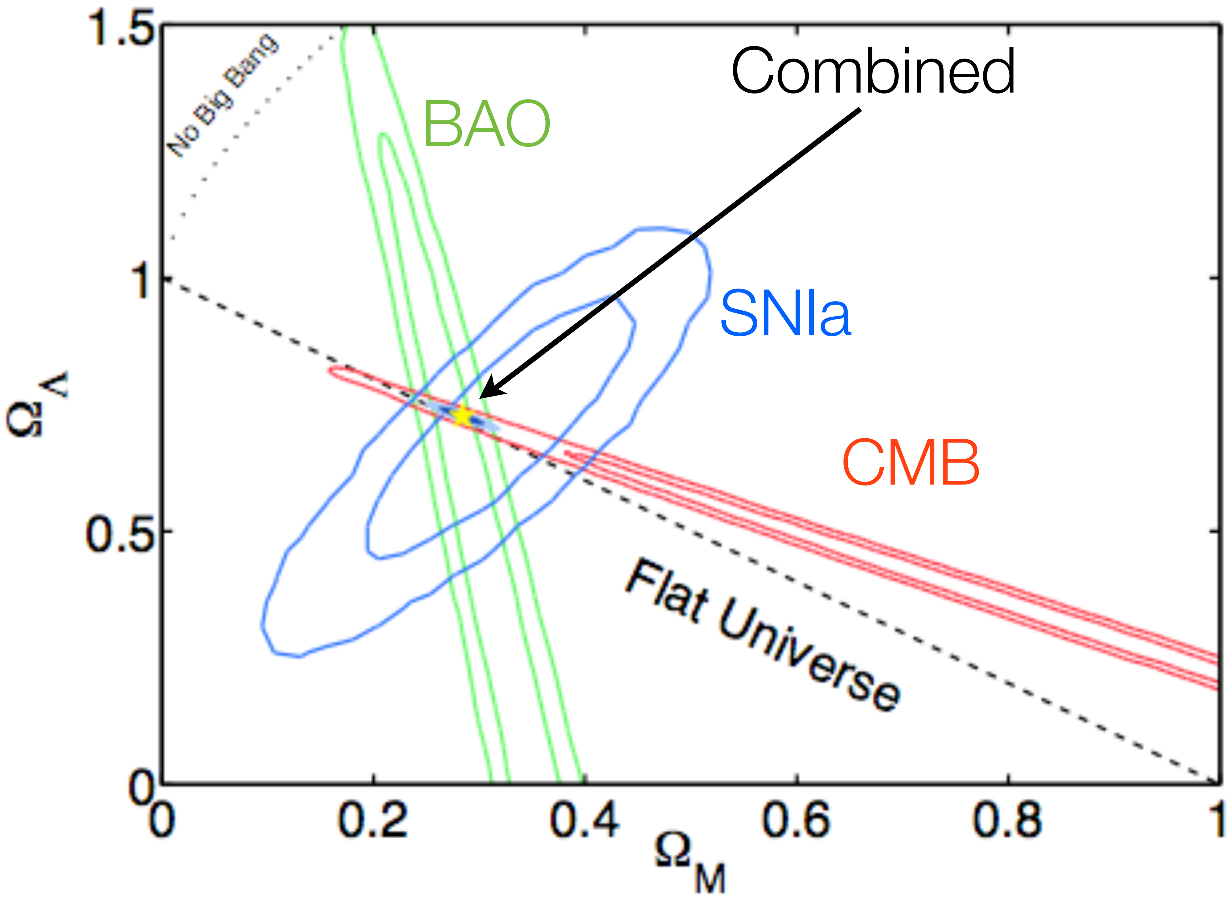


$$\sigma = 0.13 \pm 0.01 \text{ mag}$$

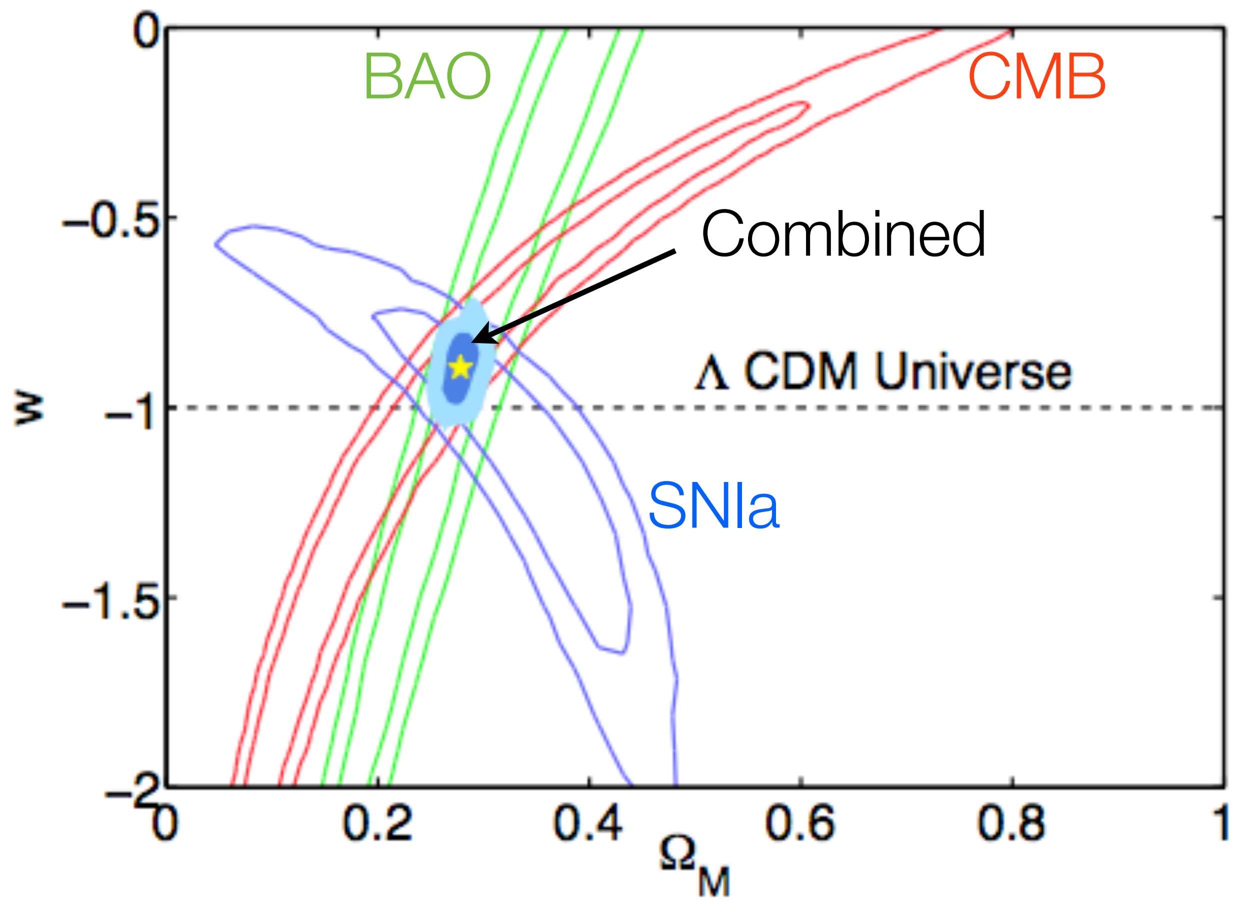
Combined constraints

- Combined cosmological constraints on matter and dark energy content:

w = 1



$\Omega_K = 0$

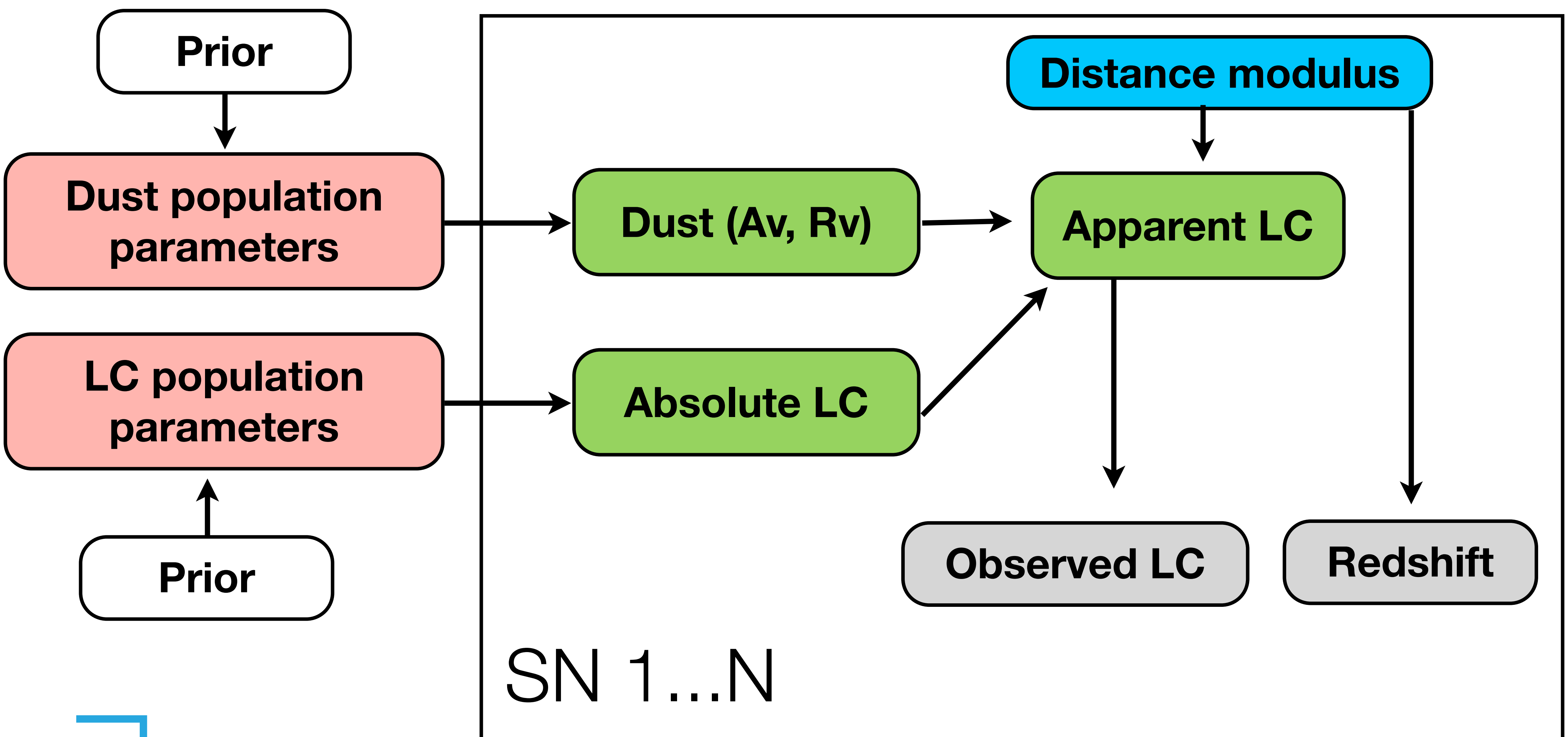


March, RT et al (2011)

The BayeSN approach

Imperial College London

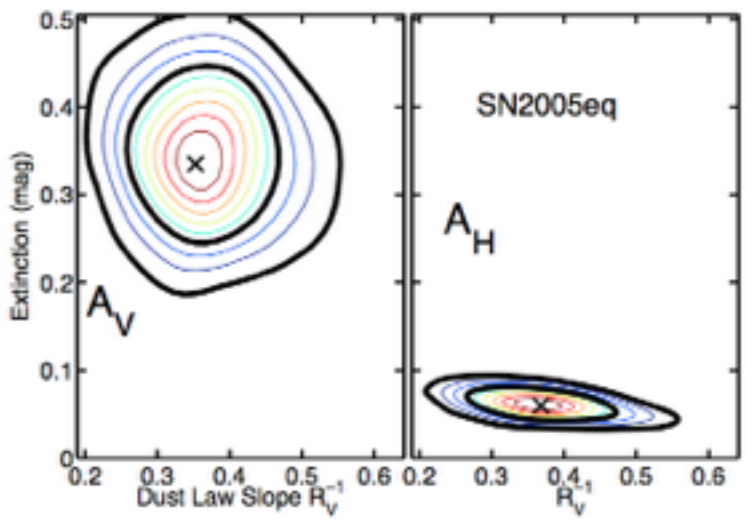
- Developed by K. Mandel (Mandel et al, 2009, 2011) and collaborators: fully Bayesian approach to LC fitting, including random errors, population structure, intrinsic variations/correlations, dust extinction and reddening, incomplete data



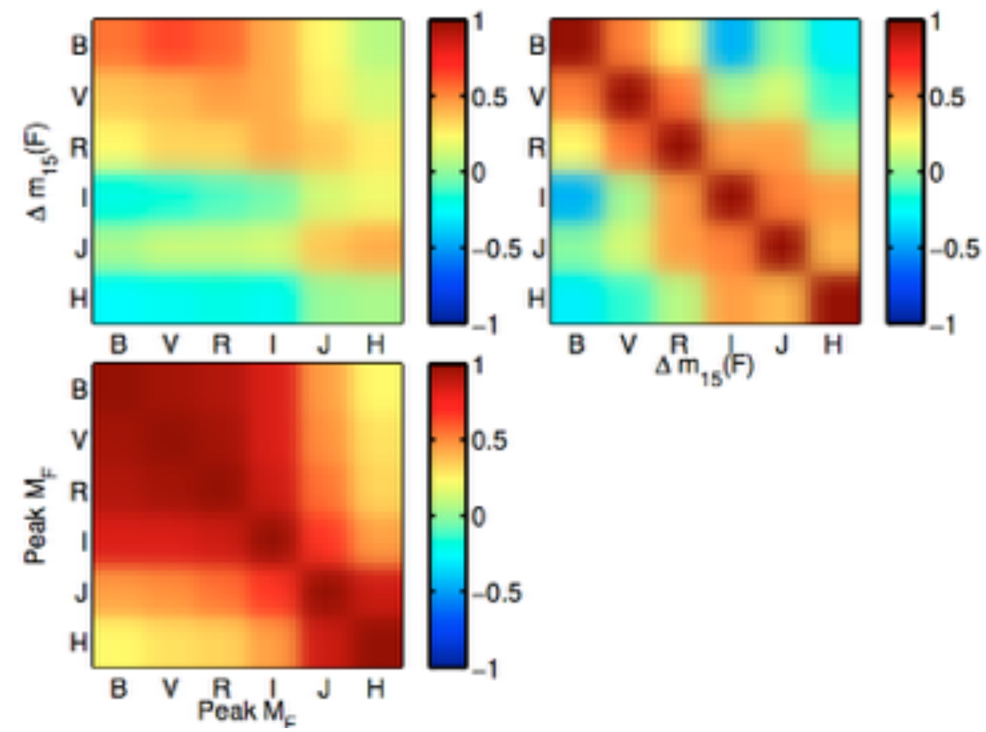
Some results from BayeSN

Mandel et al (2011)

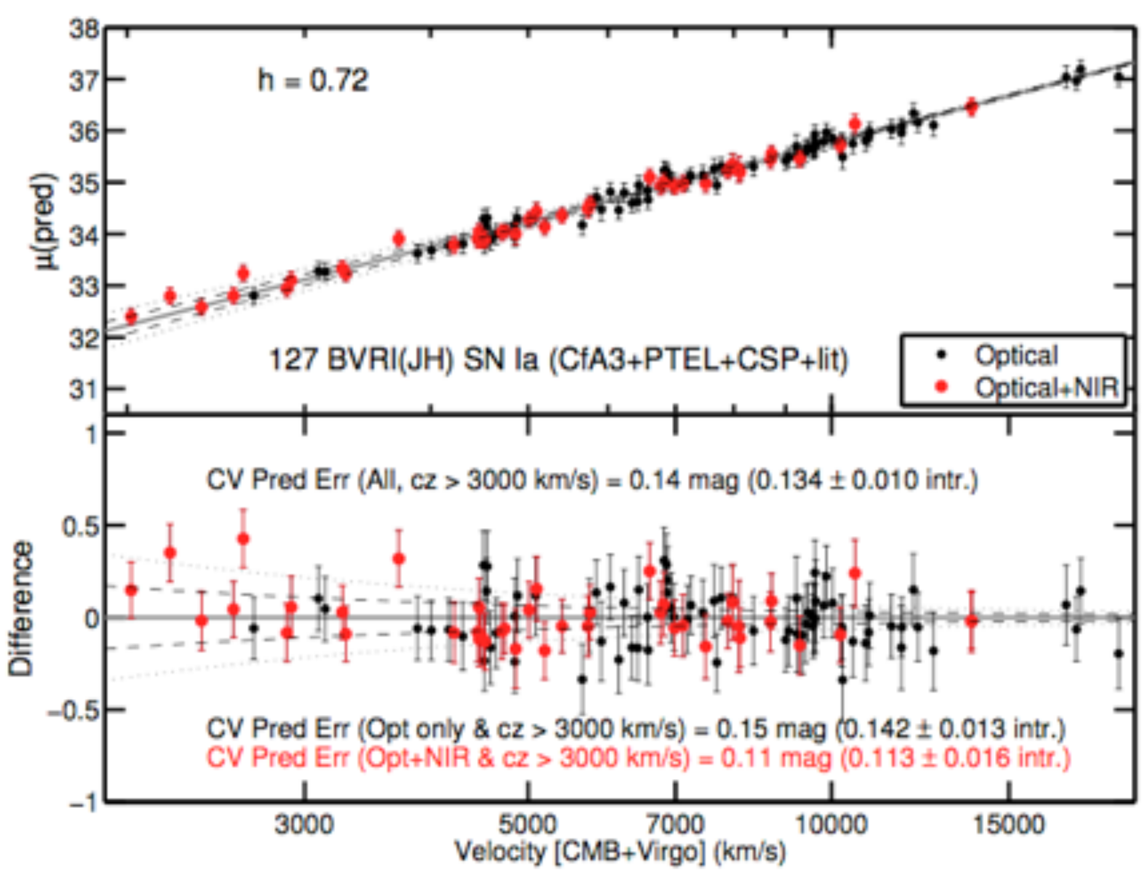
Dust absorption for each SNIa



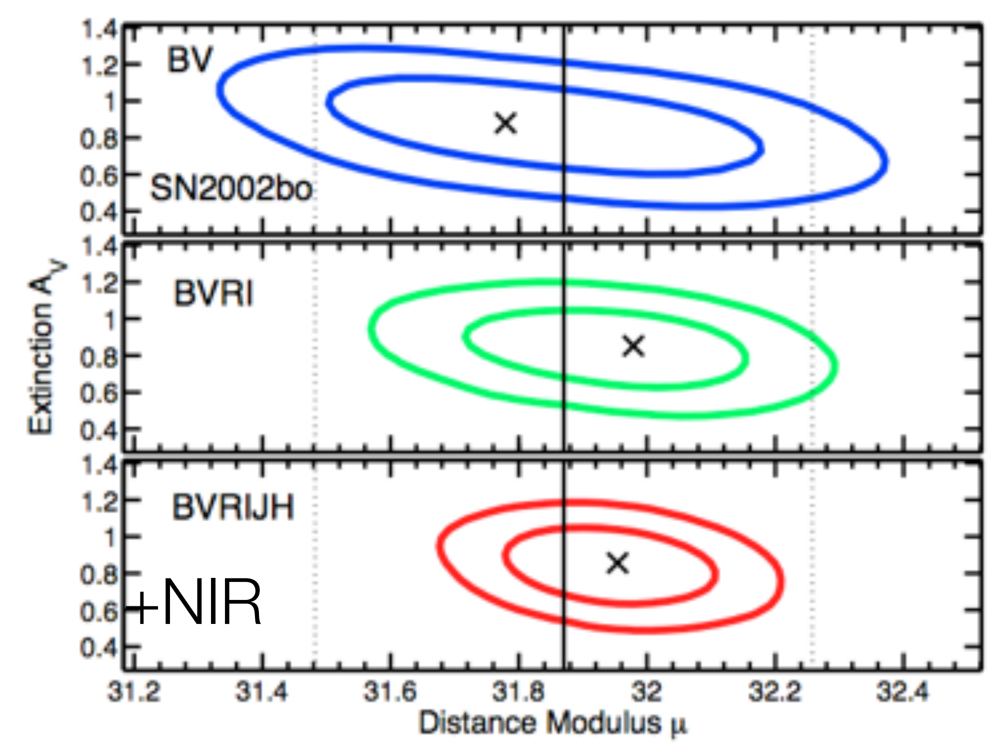
Population level analysis of correlations



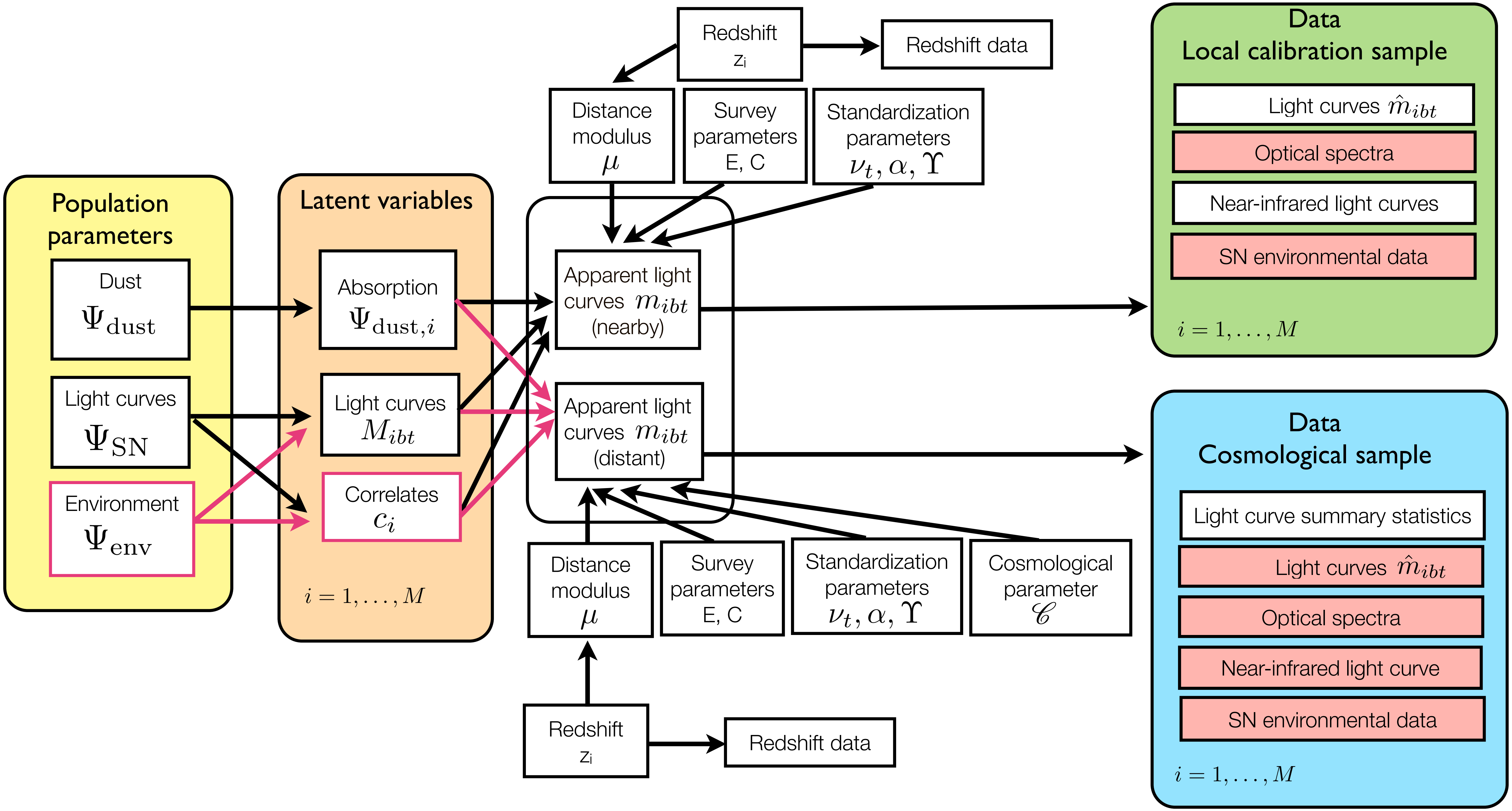
Hubble diagram: residual scatter reduced by ~2 using optical+NIR LC



Inclusion of NIR LC



The complete hierarchical model



Red arrows/boxes indicate elements/data that have never been explored before in such a multi-level setting

Principled Bayesian model selection

The 3 levels of inference

LEVEL 1
I have selected a model M and prior P(θ|M)

Parameter inference
What are the favourite values of the parameters?
(assumes M is true)

$$P(\theta|d,M) = \frac{P(d|\theta,M)P(\theta|M)}{P(d|M)}$$

LEVEL 2
Actually, there are several possible models: M_0, M_1,...

Model comparison
What is the relative plausibility of M_0, M_1,... in light of the data?

$$\text{odds} = \frac{\mathrm{P}(\mathrm{M}_0|\mathrm{d})}{\mathrm{P}(\mathrm{M}_1|\mathrm{d})}$$

LEVEL 3
None of the models is clearly the best

Model averaging
What is the inference on the parameters accounting for model uncertainty?

$$P(\theta|d) = \sum_i P(M_i|d)P(\theta|d,M_i)$$

Examples of model comparison questions

Imperial College London

ASTROPARTICLE

Gravitational waves detection

Do cosmic rays correlate with AGNs?

Which SUSY model is ‘best’?

Is there evidence for DM modulation?

Is there a DM signal in gamma ray/ neutrino data?

COSMOLOGY

Is the Universe flat?

Does dark energy evolve?

Are there anomalies in the CMB?

Which inflationary model is ‘best’?

Is there evidence for modified gravity?

Are the initial conditions adiabatic?

Many scientific questions are of the model comparison type

ASTROPHYSICS

Exoplanets detection

Is there a line in this spectrum?

Is there a source in this image?

Level 2 inference: model comparison

Imperial College London

$$P(\theta|d, M) = \frac{P(d|\theta, M)P(\theta|M)}{P(d|M)}$$

Bayesian evidence or model likelihood

The evidence:

$$P(d|M) = \int_\Omega d\theta P(d|\theta, M)P(\theta|M)$$

Posterior probability for the model M:

$$P(M|d) = \frac{P(d|M)P(M)}{P(d)}$$

When comparing two models:

$$\frac{P(M_0|d)}{P(M_1|d)} = \frac{P(d|M_0)}{P(d|M_1)} \frac{P(M_0)}{P(M_1)}$$

The Bayes factor:

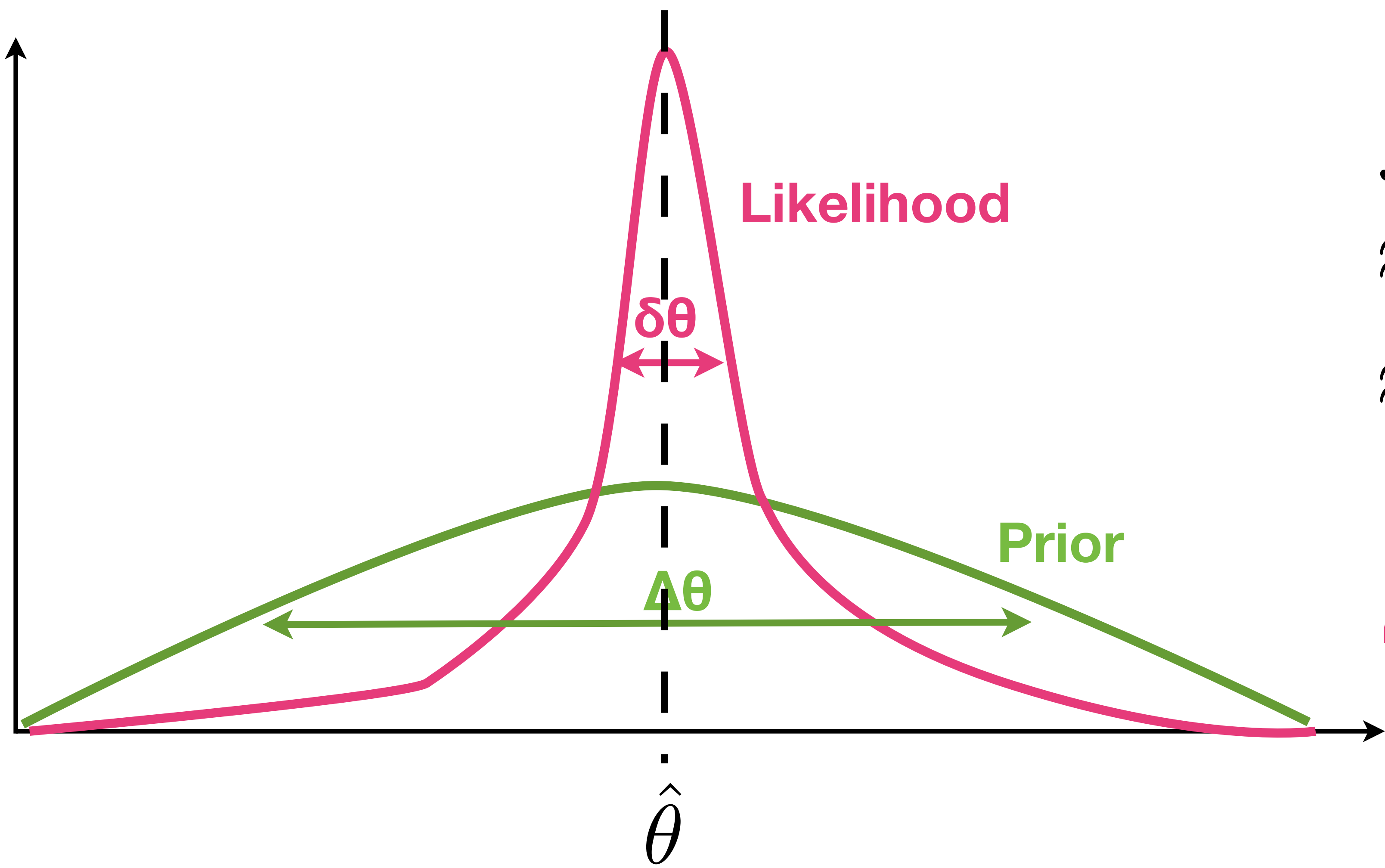
$$B_{01} \equiv \frac{P(d|M_0)}{P(d|M_1)}$$

Posterior odds = Bayes factor × prior odds

An in-built Occam's razor

Imperial College London

- The Bayesian evidence balances *quality of fit* vs *extra model complexity*.
- It rewards highly predictive models, penalizing "wasted" parameter space.
- **The prior here is important:** it quantifies the *predictive power* of the model.



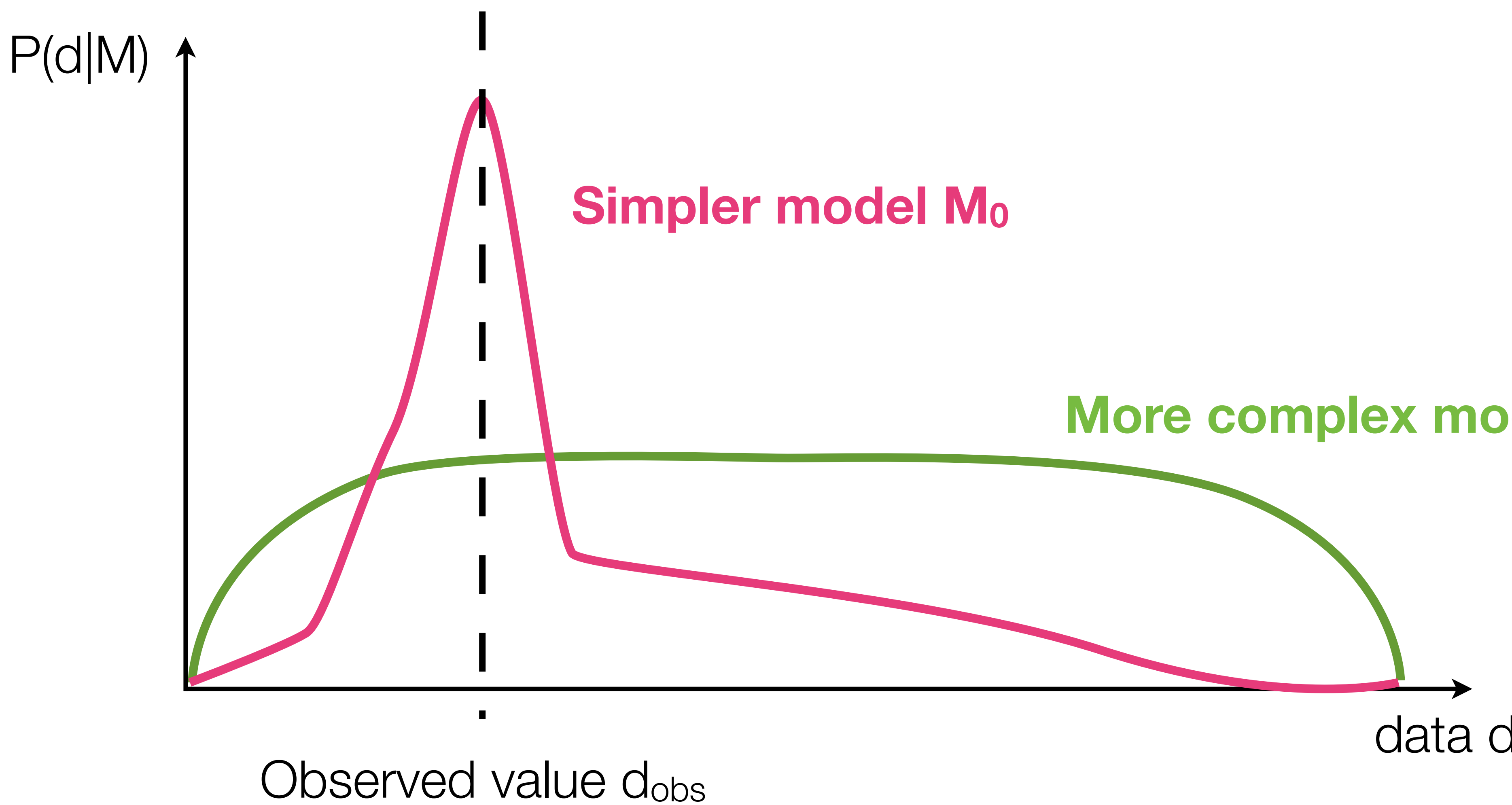
$$P(d|M) = \int d\theta L(\theta) P(\theta|M) \approx L(\hat{\theta}) \delta\theta P(\hat{\theta}) \approx \frac{\delta\theta}{\Delta\theta} L(\hat{\theta})$$

"Occam's factor" ($\frac{\delta\theta}{\Delta\theta}$)

Quality of fit ($L(\hat{\theta})$)

The evidence as predictive probability

- The evidence can be understood as a function of d to give the predictive probability under the model M:



Nested models

M0: θ = 0

M1: θ ≠ 0 with prior p(θ)

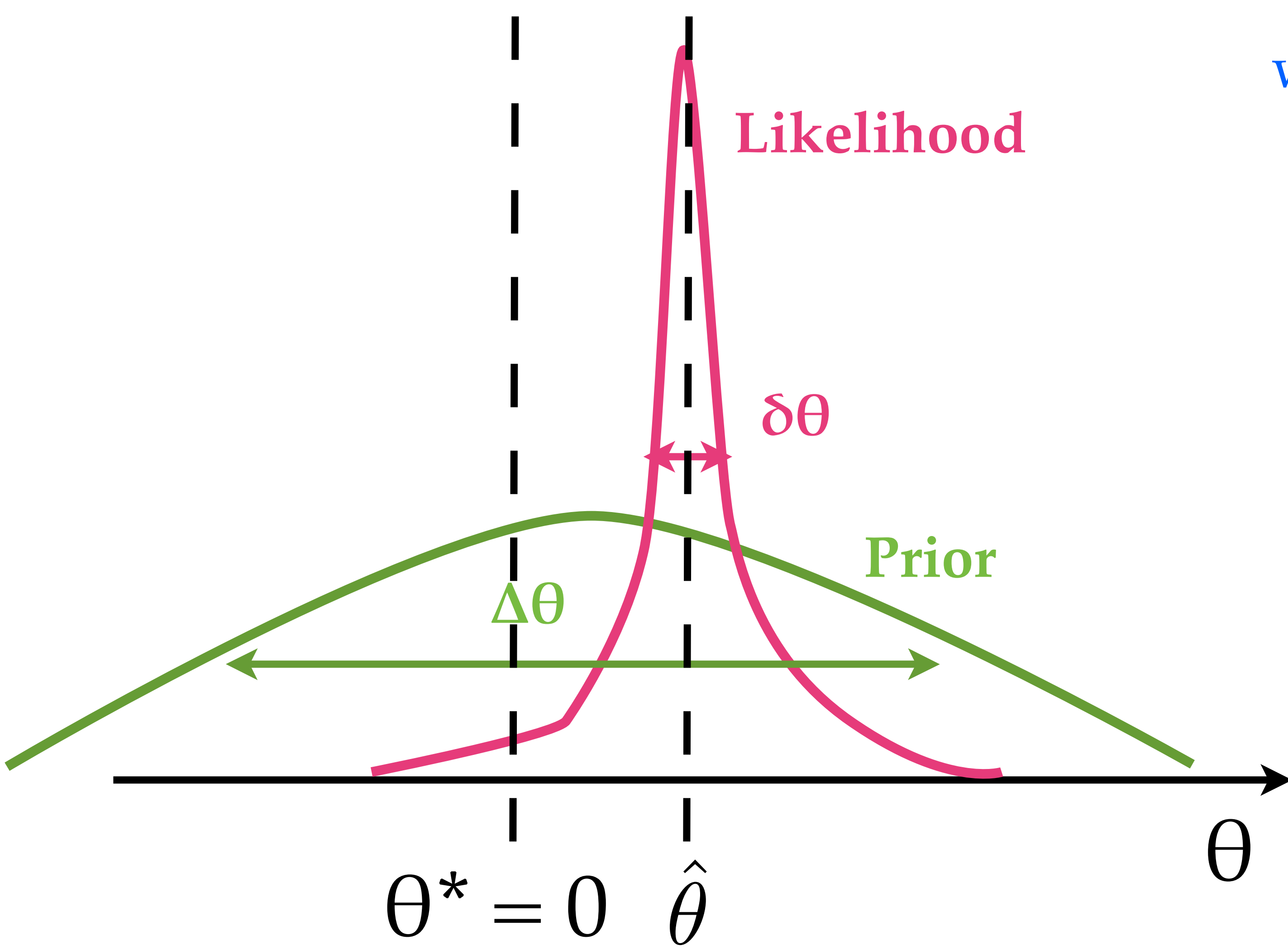
Do we need the extra "complexity"?

$$\lambda \equiv \frac{\hat{\theta}-\theta^\star}{\delta\theta}$$

$$\ln B_{01} \approx \ln \frac{\Delta\theta}{\delta\theta} - \frac{\lambda^2}{2}$$

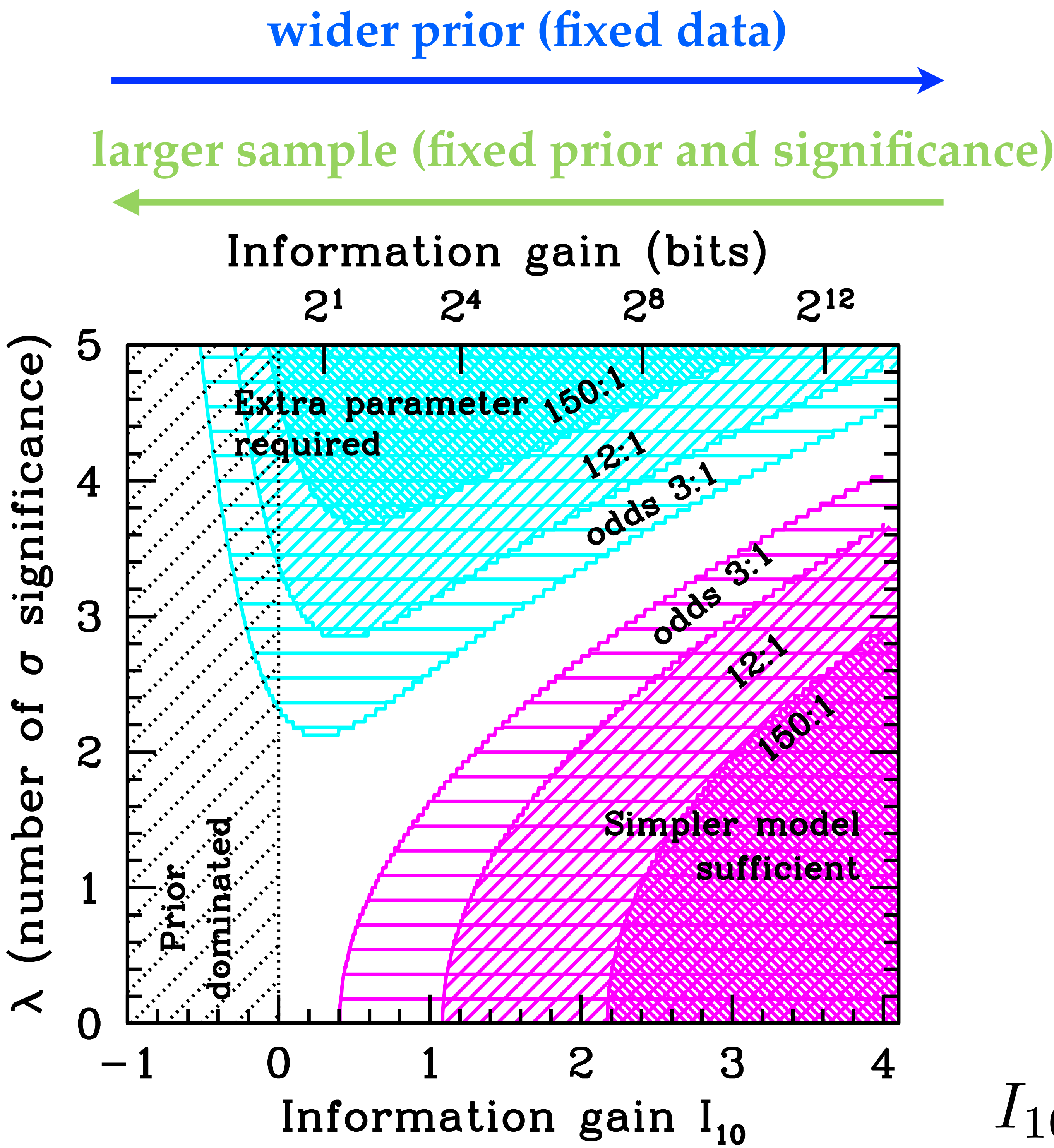
wasted parameter space (favours simpler model)

mismatch of prediction with observed data (favours more complex model)



Model selection for nested models

Imperial College London



In Bayesian model comparison, **the prior scale never goes away.**

Also, the **alternative hypothesis** needs to be formulated from the outset (Jaynes: *"there is no point in rejecting a model unless one has a better alternative"*)

One should look at the scale of the prior and hope that the result is **robust** for "reasonable" prior choices

$$I_{10} \equiv \log_{10} \frac{\Delta\theta}{\delta\theta}$$

Trotta (2008)

Scale for the strength of evidence

Imperial College London

- A (slightly modified) Jeffreys' scale to assess the strength of evidence (**Notice:** this is empirically calibrated!)

$\|\ln B\|$	relative odds	favoured model's probability	Interpretation
< 1.0	< 3:1	< 0.750	not worth mentioning
< 2.5	< 12:1	0.923	weak
< 5.0	< 150:1	0.993	moderate
> 5.0	> 150:1	> 0.993	strong

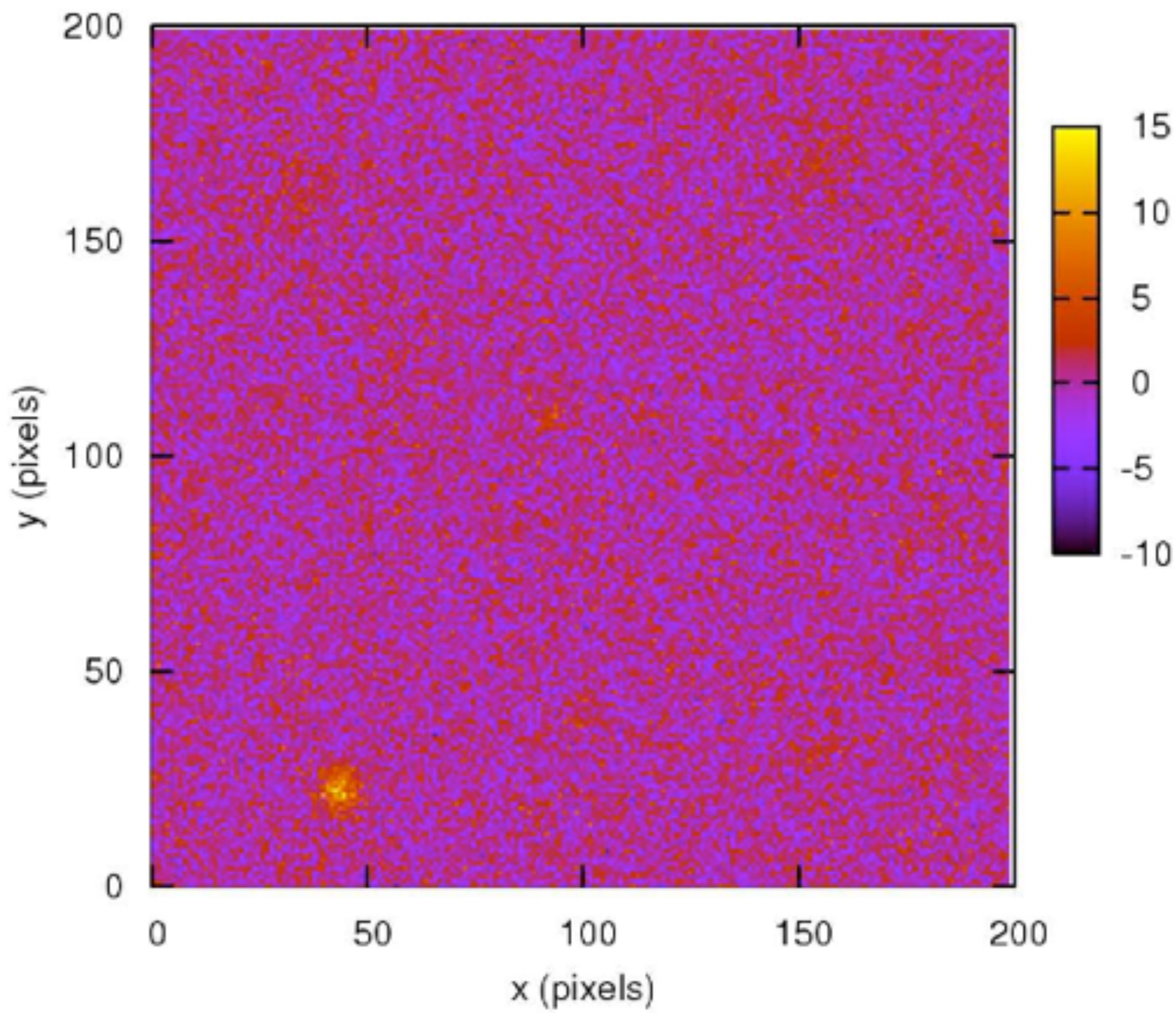
ICIC

Astro example: how many sources?

Imperial College London

Feroz and Hobson (2007)

Signal + Noise

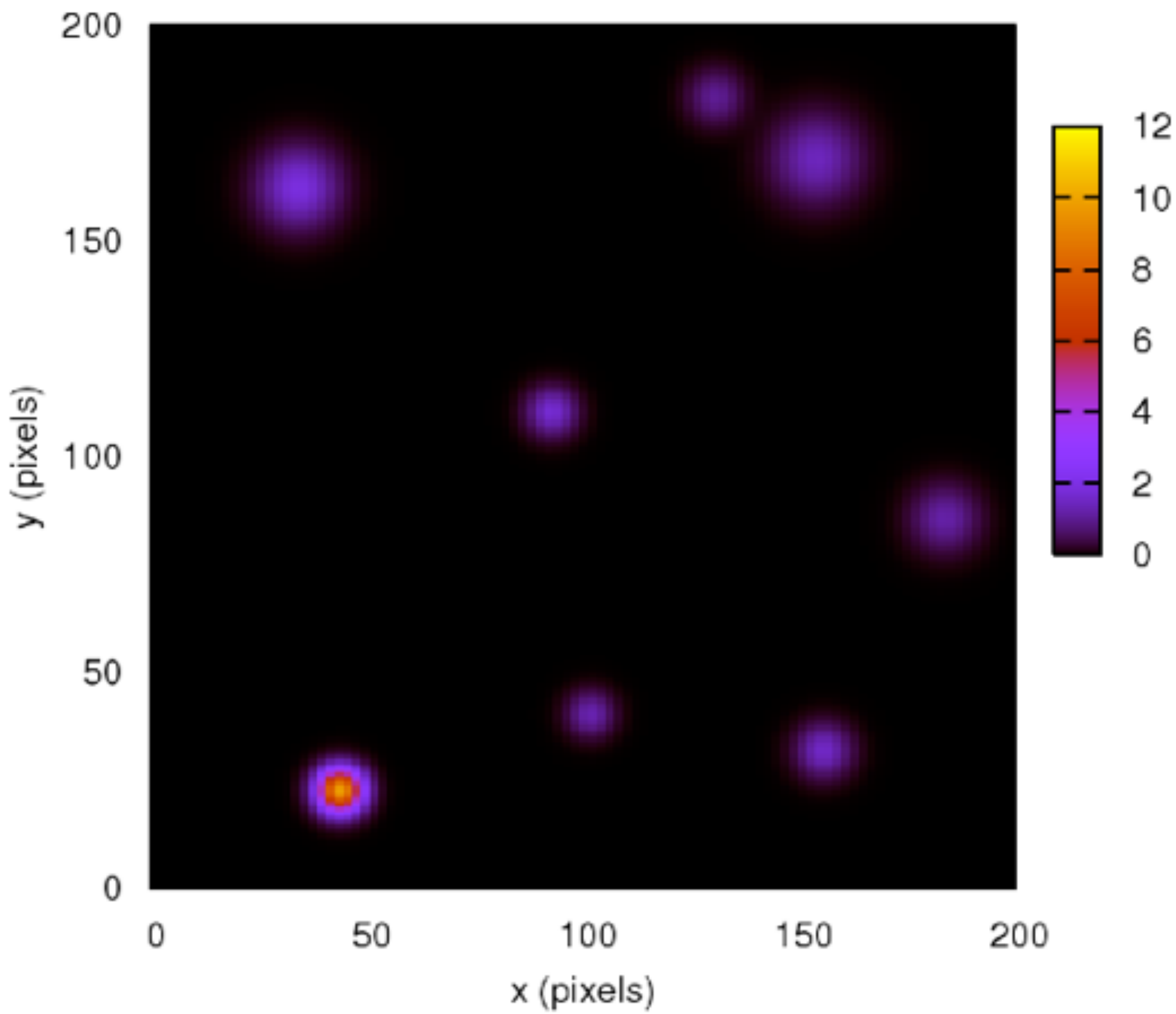


Astro example: how many sources?

Imperial College London

Feroz and Hobson (2007)

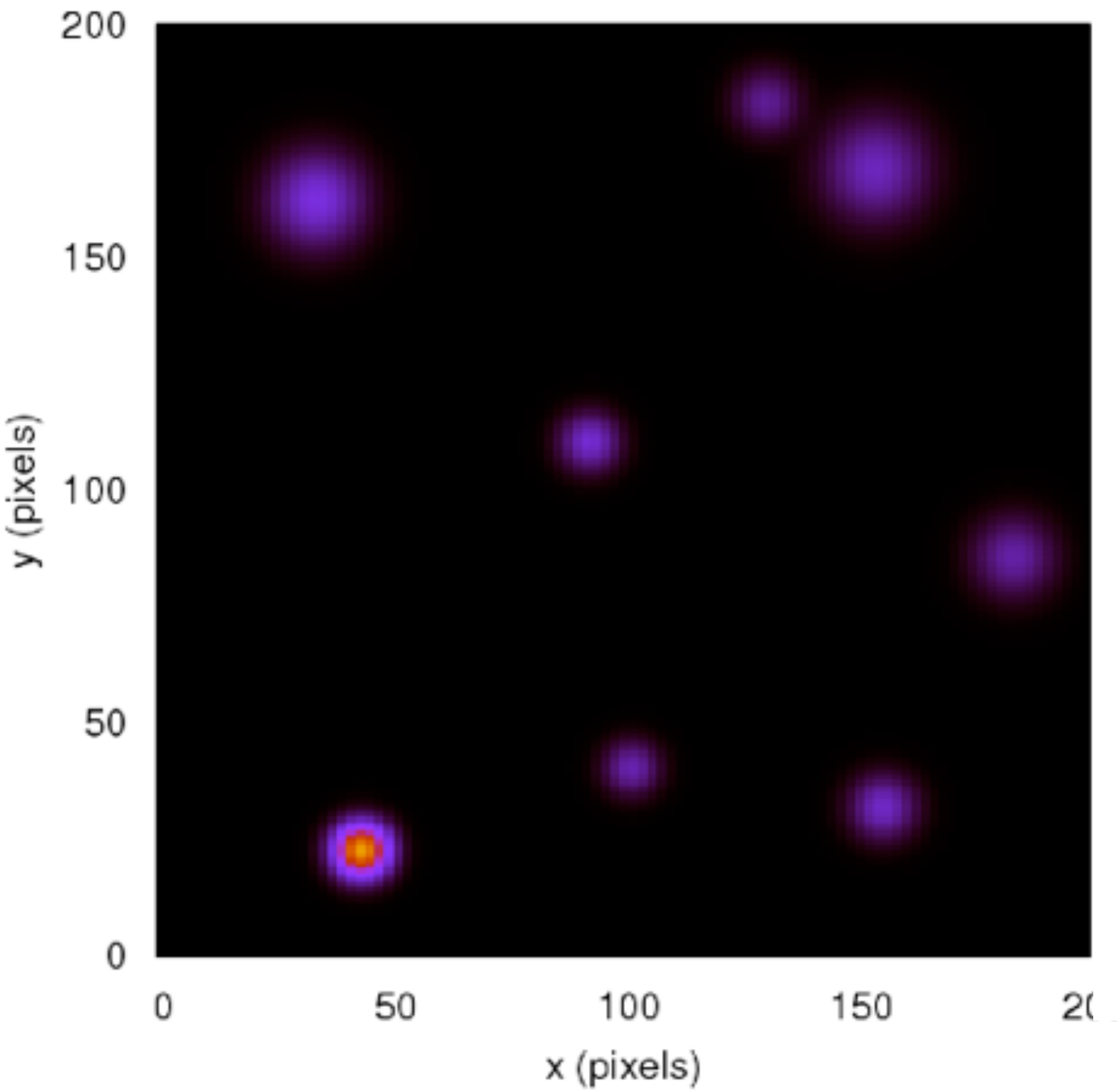
Signal: 8 sources



Astro example: how many sources?

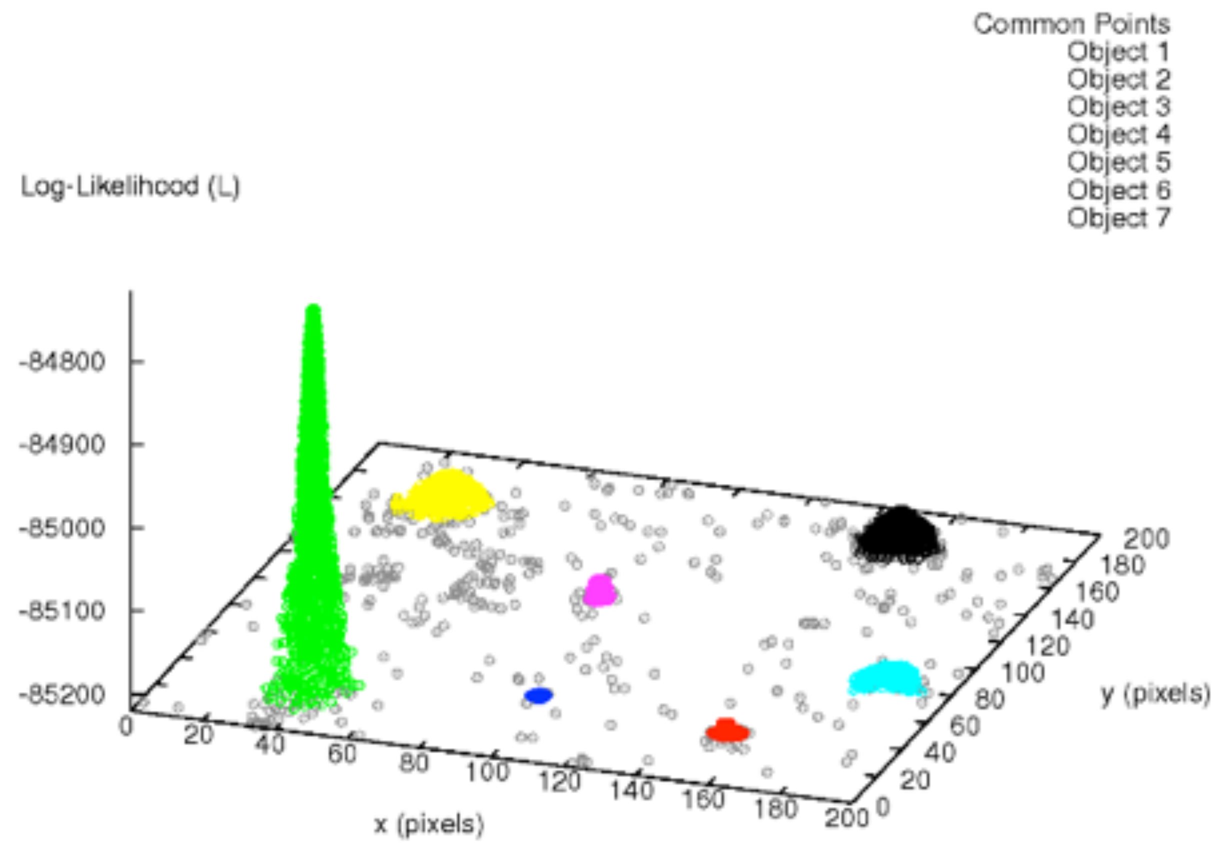
Imperial College London

Feroz and Hobson (2007)

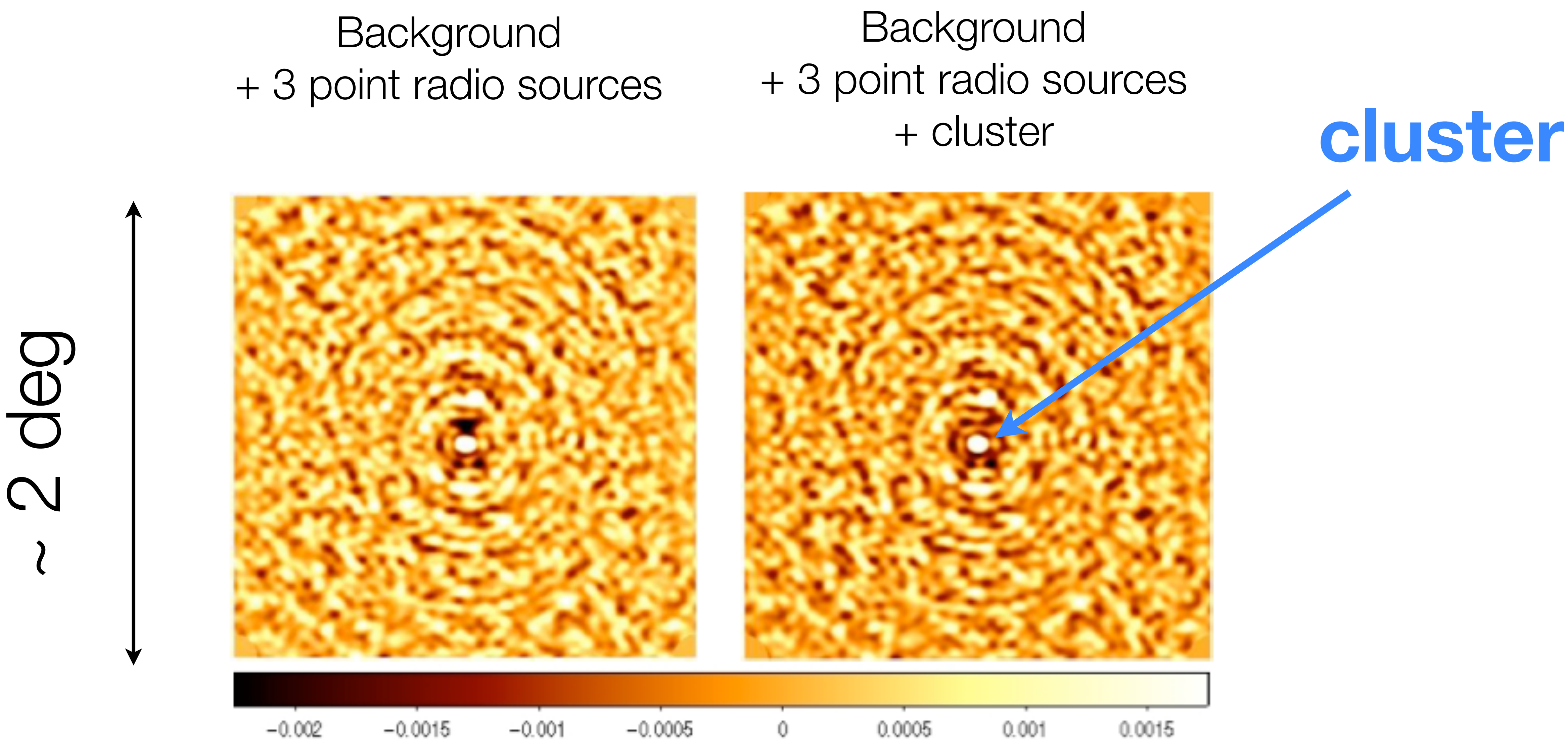


Bayesian reconstruction

7 out of 8 objects correctly identified.
Mistake happens because 2 objects very close.



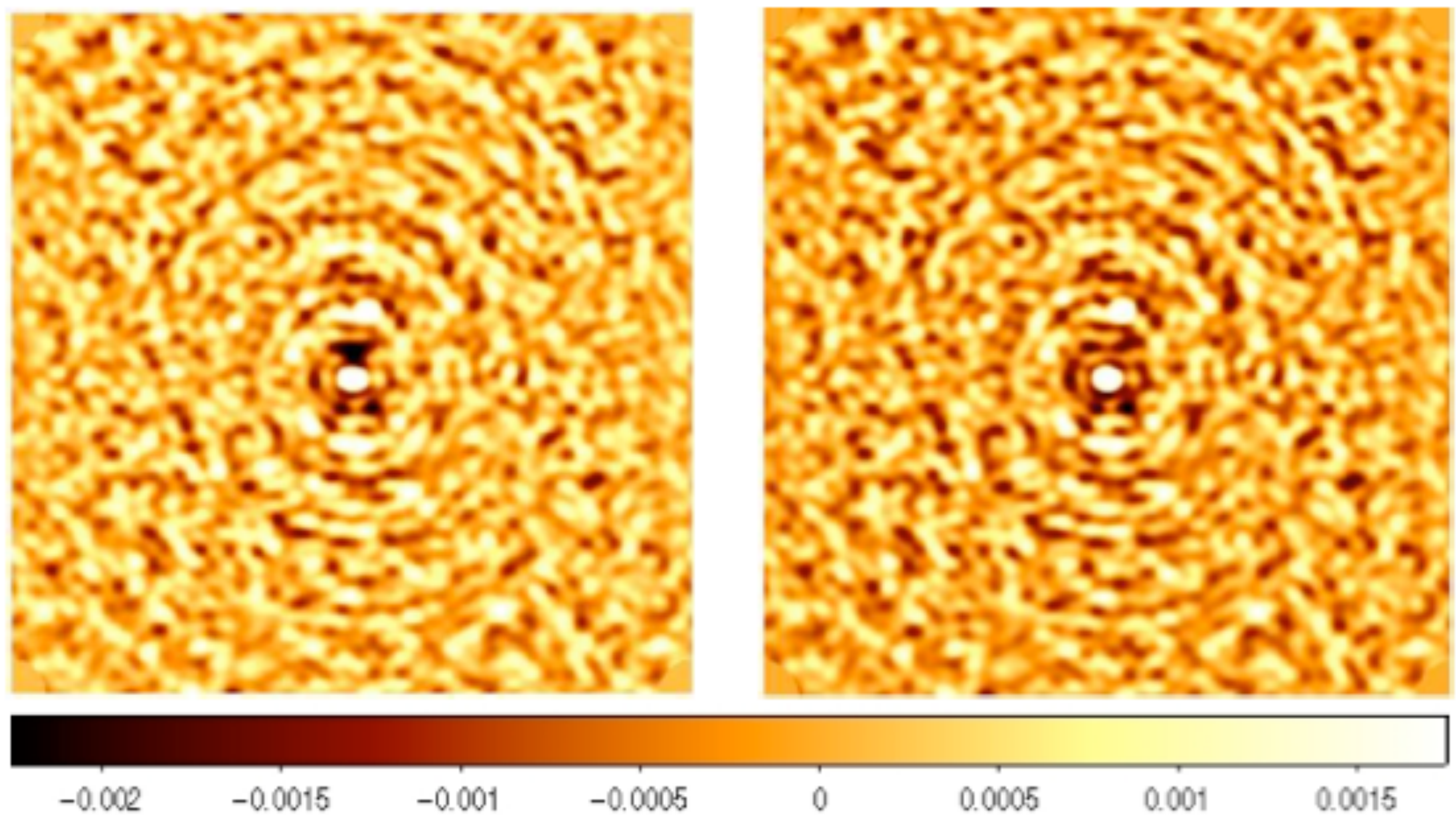
Cluster detection from Sunyaev-Zeldovich effect in cosmic microwave background maps



Feroz et al 2009

Background
+ 3 point radio sources

Background
+ 3 point radio sources
+ cluster



Posterior odds:

R = P(cluster | data)/P(no cluster | data)

$R = 0.35 \pm 0.05$ $R \sim 10^{33}$

Cluster parameters also recovered (position, temperature, profile, etc)

Computing the evidence

Imperial College London

Evidence: $P(d|M) = \int_\Omega d\theta P(d|\theta, M) P(\theta|M)$

Bayes factor: $$B_{01} \equiv \frac{P(d|M_0)}{P(d|M_1)}$$

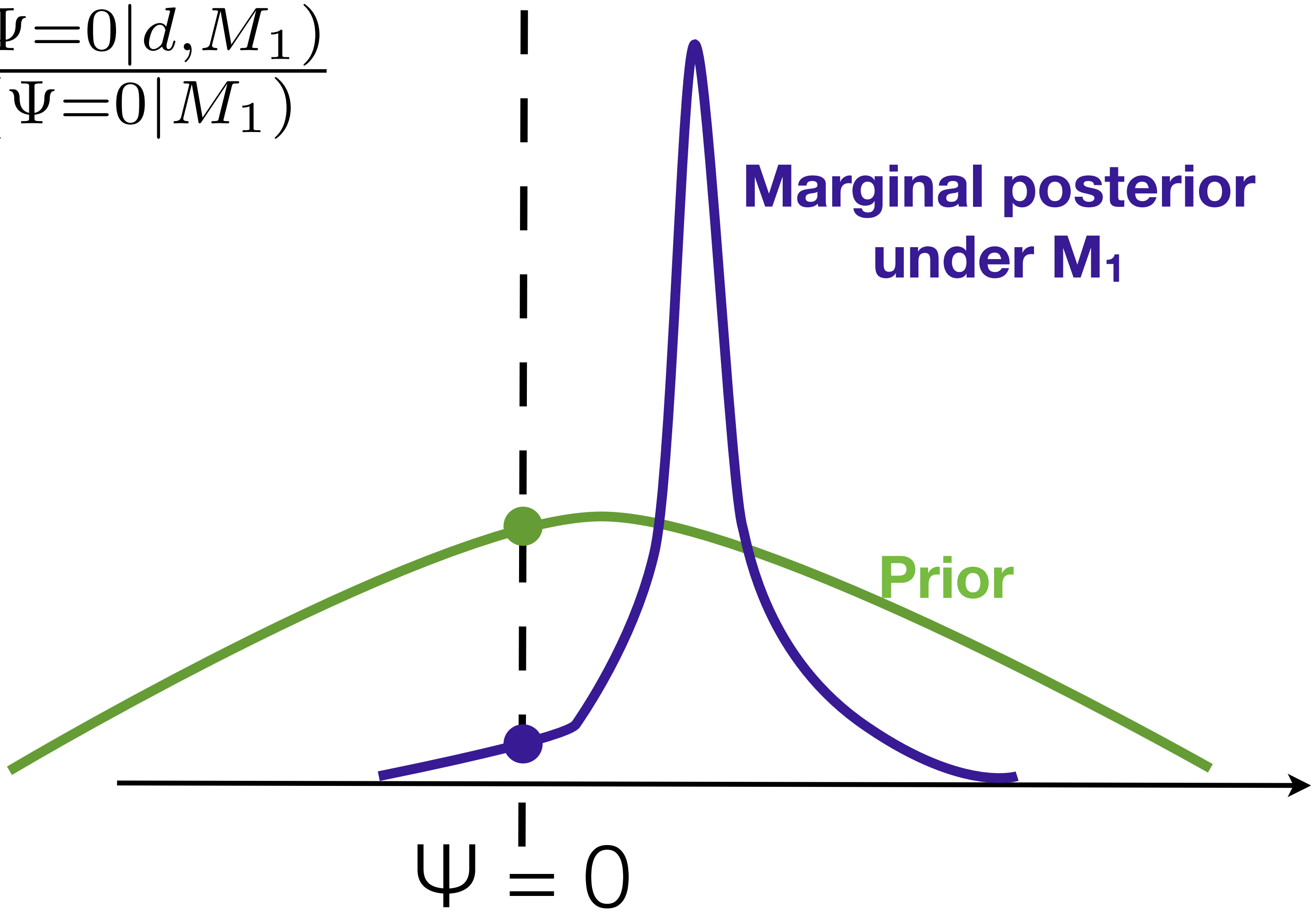
- Usually a computational demanding multi-dimensional integral!
- Several numerical/semi-analytical techniques available:
  - **Thermodynamic integration** or **Population Monte Carlo**
  - **Laplace approximation:** approximate the likelihood to second order around maximum gives Gaussian integrals (for normal prior). Can be inaccurate.
  - **Savage-Dickey density ratio:** good for nested models, gives the Bayes factor
  - **Nested sampling:** clever & efficient, can be used generally

The Savage-Dickey density ratio

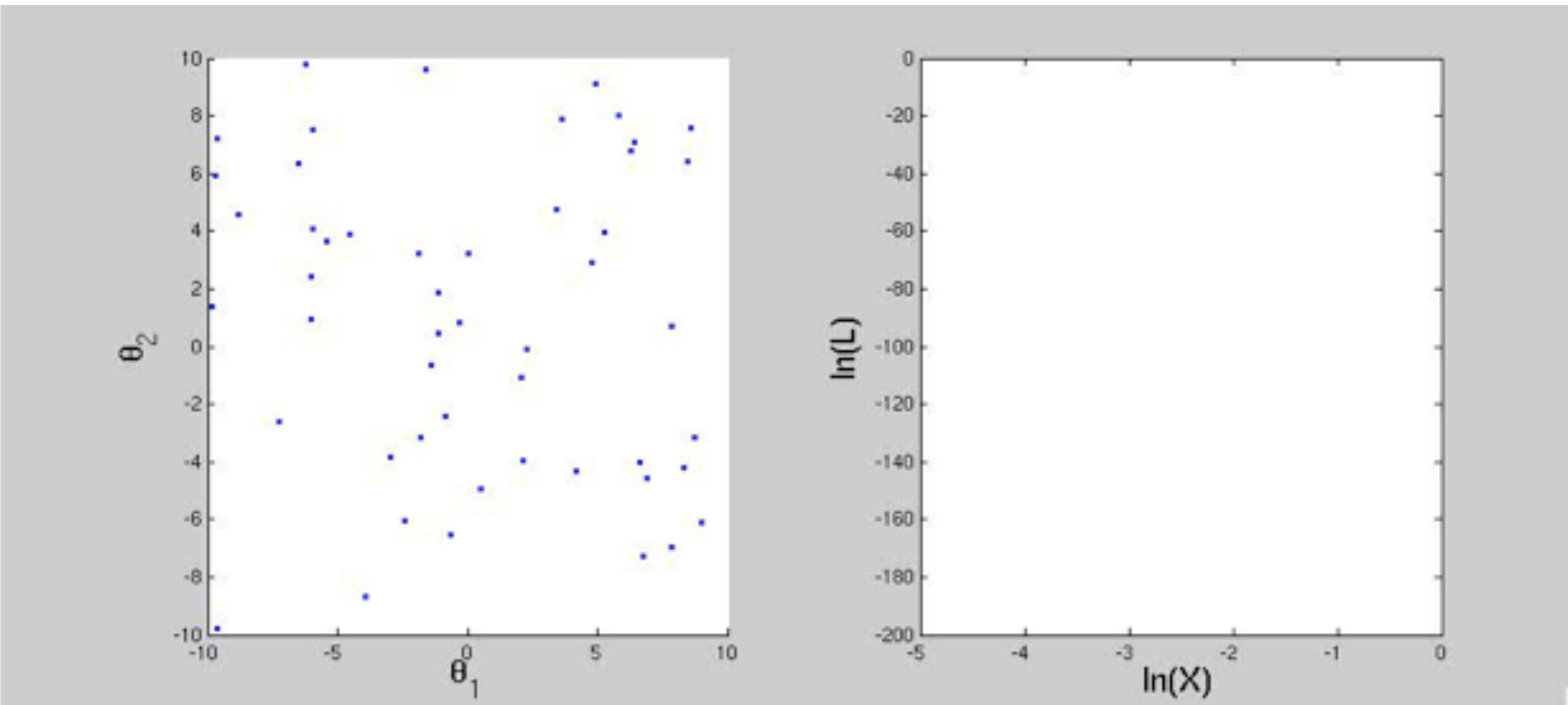
Imperial College London

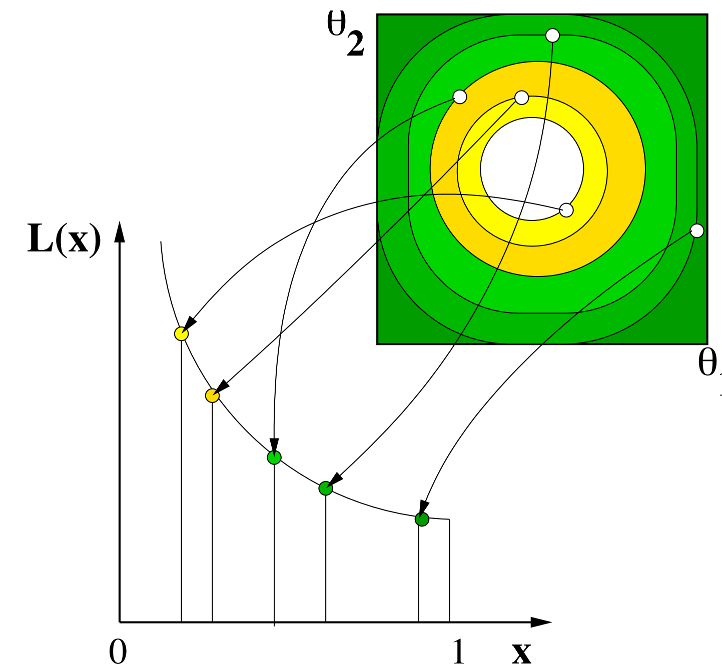
- This methods works for nested models and gives the Bayes factor analytically.
- **Assumptions:** nested models (M_1 with parameters θ,Ψ reduces to M_0 for e.g. Ψ =0) and separable priors (i.e. the prior $P(\theta,\Psi|M_1)$ is uncorrelated with $P(\theta|M_0)$)
- Result:
- **Advantages**:
  - analytical
  - often accurate
  - clarifies the role of prior
  - does not rely on Gaussianity

$$B_{01} = \frac{P(\Psi=0|d,M_1)}{P(\Psi=0|M_1)}$$



Nested sampling





(animation courtesy of David Parkinson)

An algorithm originally aimed primarily at the Bayesian evidence computation (Skilling, 2006):

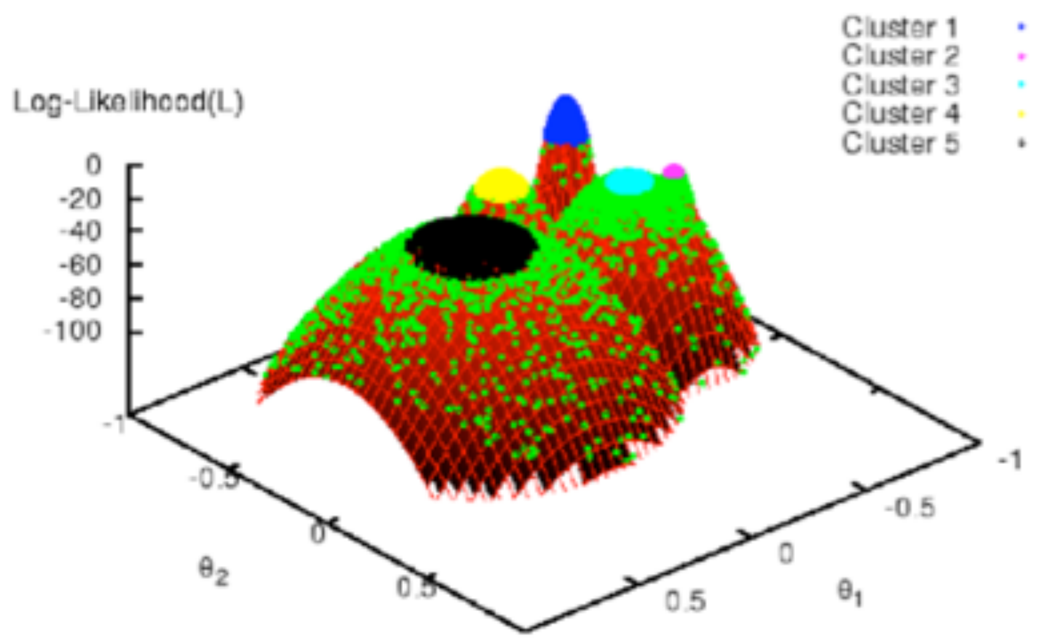
$$X(\lambda) = \int_{\mathcal{L}(\theta)>\lambda} P(\theta) d\theta$$

$$P(d) = \int d\theta L(\theta) P(\theta) = \int_0^1 L(X) dX$$

MultiNest

Feroz and Hobson (2007)

Imperial College London



Gaussian mixture model:

True evidence: log(E) = -5.27

Multinest:

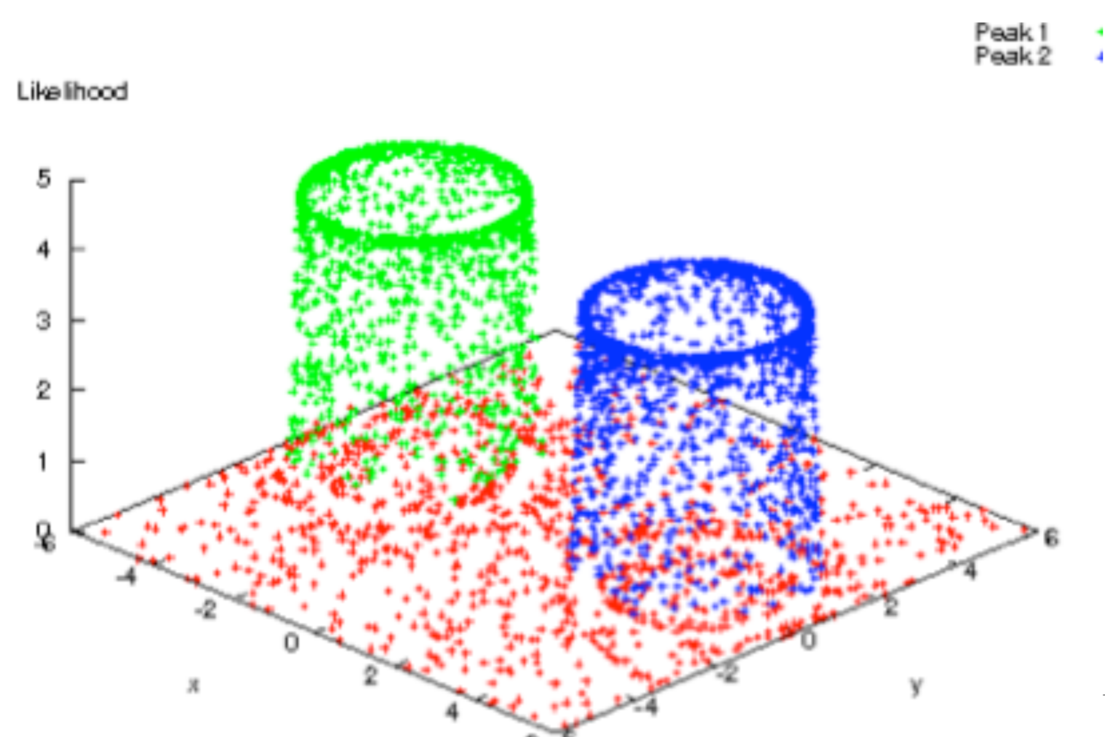
Reconstruction: log(E) = -5.33 ± 0.11

Likelihood evaluations ~ 10^4

Thermodynamic integration:

Reconstruction: log(E) = -5.24 ± 0.12

Likelihood evaluations ~ 10^6



Courtesy Mike Hobson

D	N	efficiency	likes per dimension
2	7000	70%	83
5	18000	51%	7
10	53000	34%	3
20	255000	15%	1.8
30	753000	8%	1.6

The inflationary paradigm

Imperial College London

- Is a high energy phase of accelerated expansion in the early Universe

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\,\mathrm{d}\vec{x}^2 \qquad \ddot{a} > 0$$

- Solves the Hot Big Bang horizon and flatness problem
- Can be implemented with a single scalar field

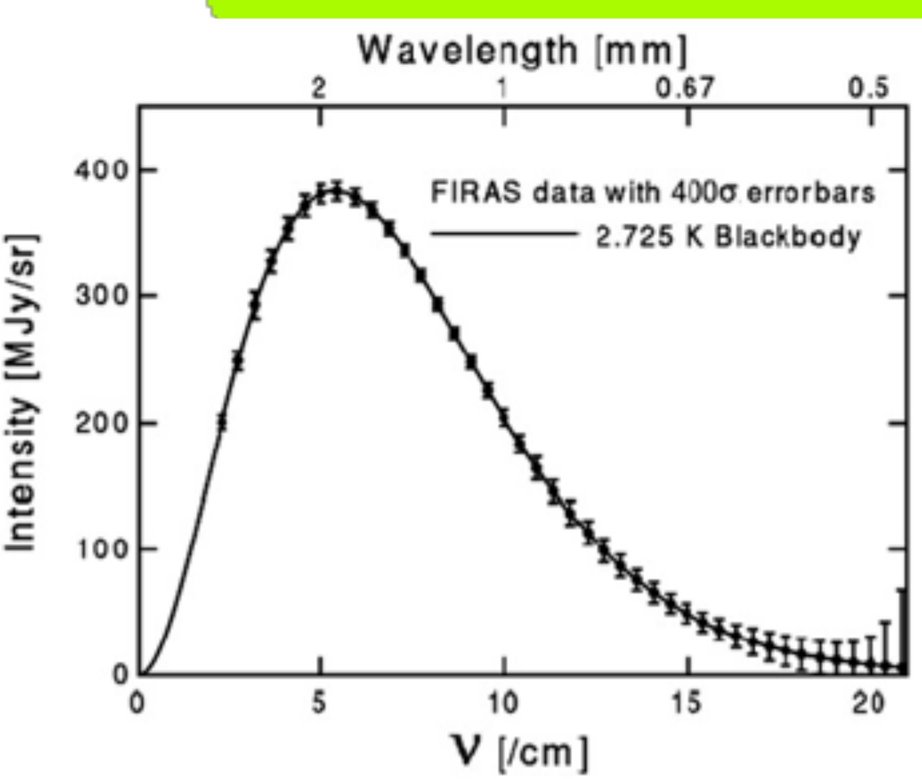
$$S = -\int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + V(\phi)\right]$$

$$\Rightarrow \quad \begin{cases} \rho = \frac{1}{2}\left(\dot{\phi}\right)^2 + V(\phi) \\ p = \frac{1}{2}\left(\dot{\phi}\right)^2 - V(\phi) \end{cases}$$

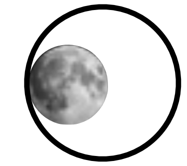
$$\ddot{a}/a = -\frac{1}{6M_{\mathrm{P}}^2}(\rho + 3p) \qquad \Longrightarrow \qquad V(\phi) \gg \dot{\phi}^2$$

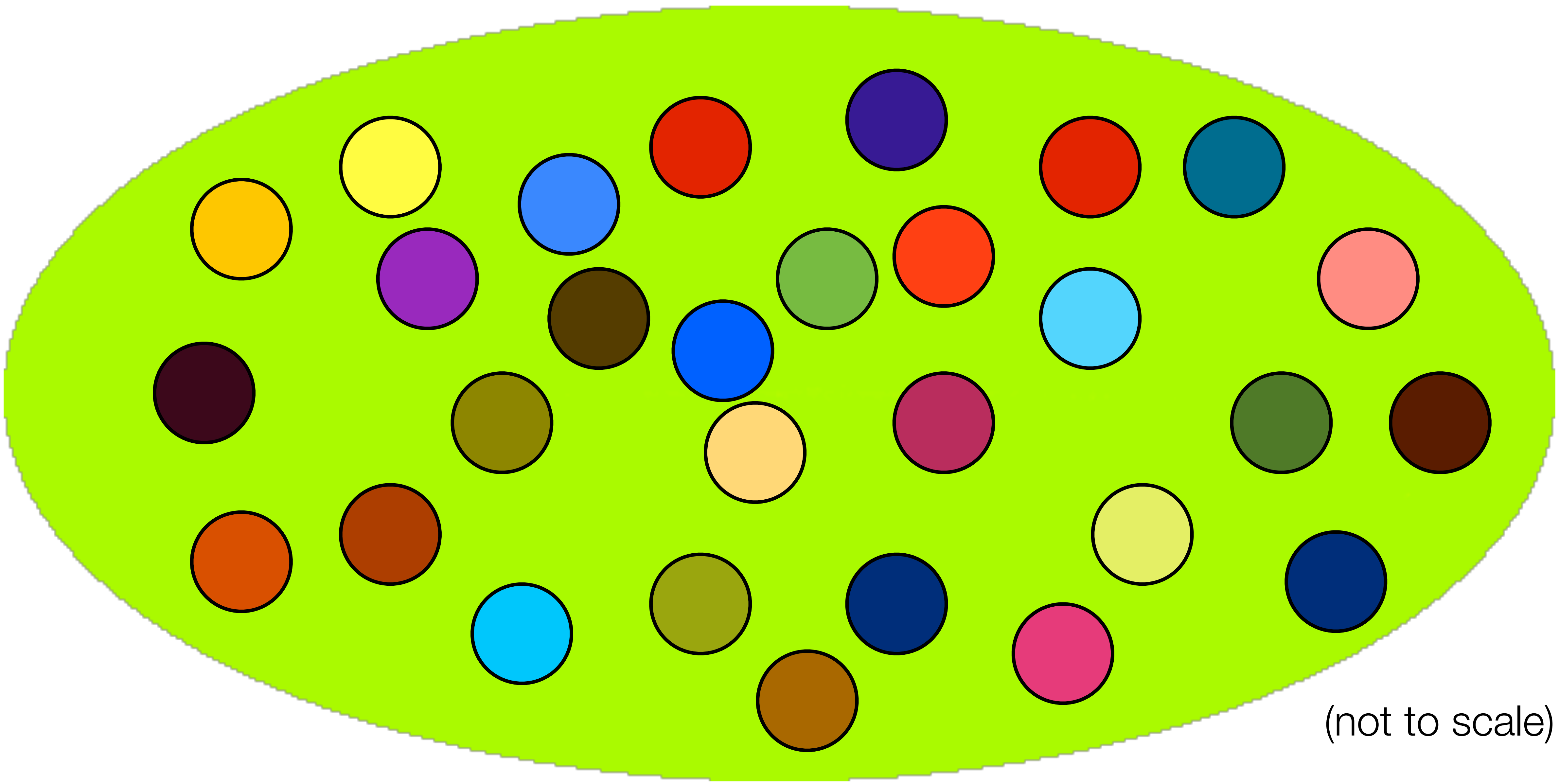
The horizon problem

T = 2.72 K



The horizon problem





Patches separated by more than 1 deg should not have the same temperature!

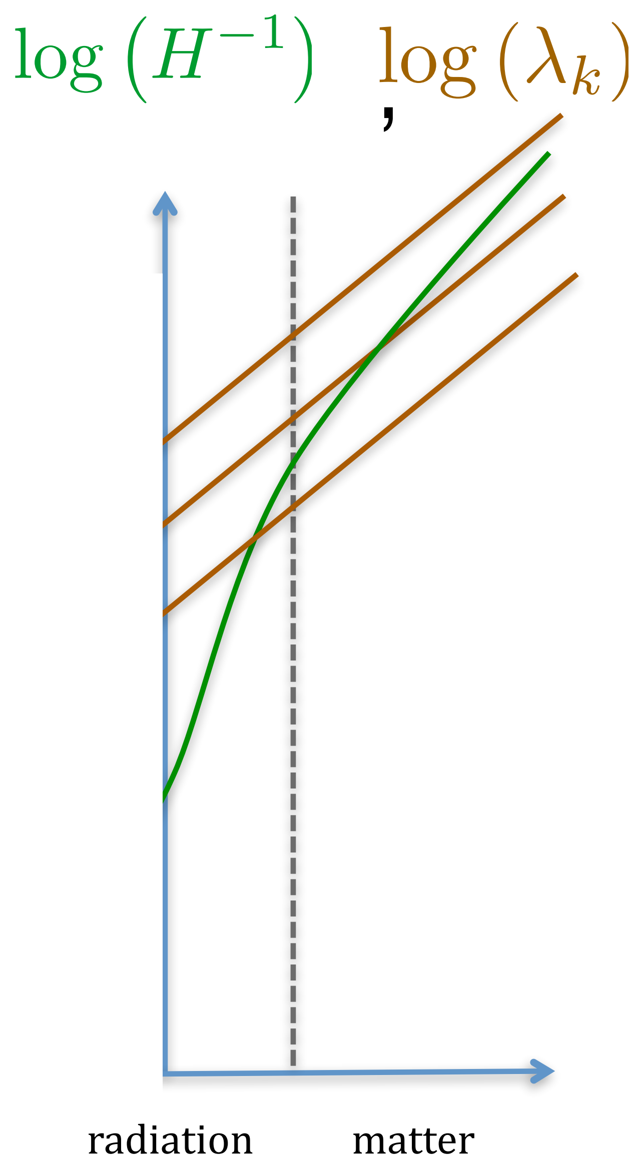
Solution to the "horizon problem"

Imperial College London

$$H^{-1} = \left(\frac{\dot{a}}{a}\right)^{-1}$$ physical horizon length

During inflation: $$\frac{d}{dt}\frac{H^{-1}}{a} < 0$$

$$\lambda_k = \frac{2\pi}{k} a$$ physical scale



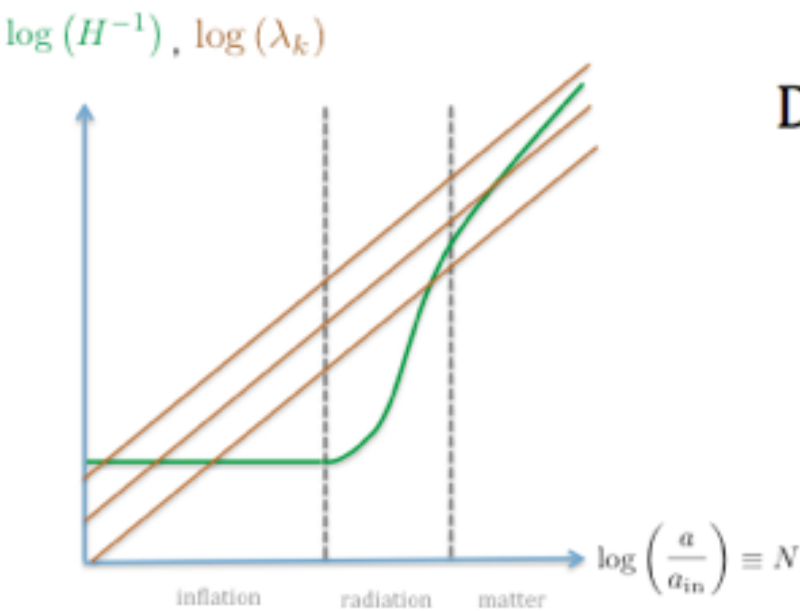
"number of e-folds"

$$\log\left(\frac{a}{a_{\rm in}}\right) \equiv N$$

ICIC

Slow-roll approximation

Imperial College London



During inflation, H is almost constant

$$\epsilon_0 = \frac{H_{\rm in}}{H} \simeq \text{constant}$$

Slow-Roll hierarchy: $\epsilon_{n+1} = \frac{1}{\epsilon_n}\frac{d\epsilon_n}{dN}$

Friedman equation: $3M_{\rm Pl}^2 H^2 = V + \dot{\phi}^2/2$

Klein Gordon equation: $\ddot{\phi} + 3H\dot{\phi} + V_\phi = 0$

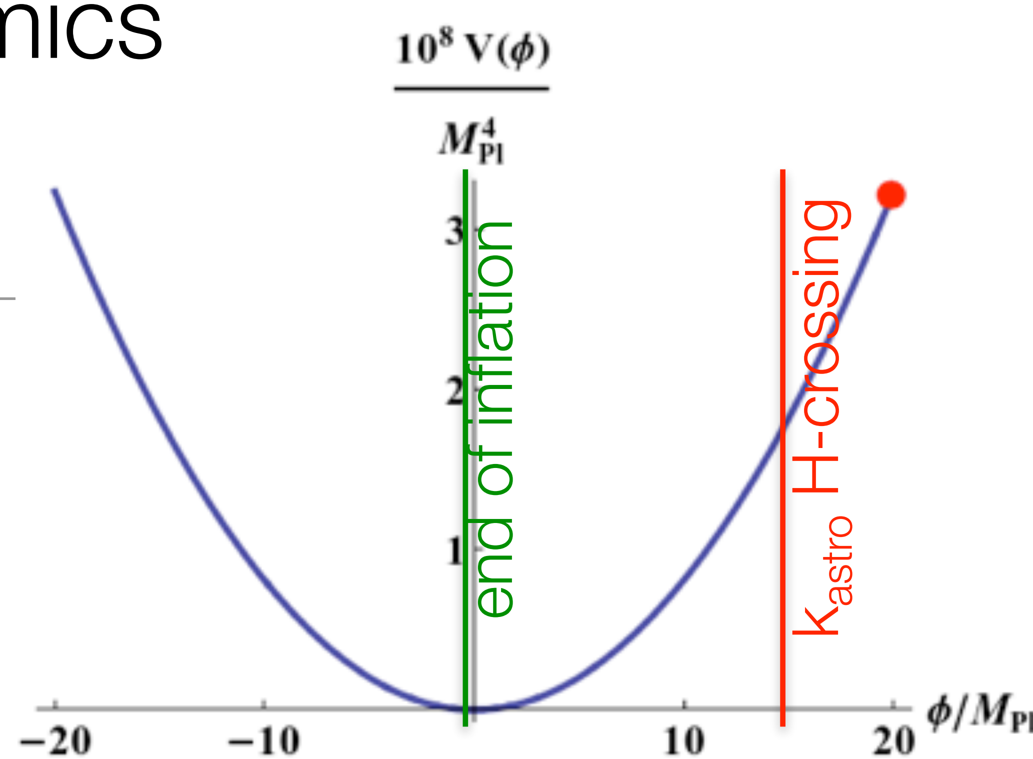
$$\Longrightarrow \epsilon_1 = 3\frac{\dot{\phi}^2/(2V)}{1+\dot{\phi}^2/(2V)} \ll 1$$

$$\frac{\ddot{a}}{a} = \dot{H} + H^2 > 0 \qquad -\frac{\dot{H}}{H^2} = \epsilon_1 < 1 \qquad \Rightarrow \dot{\phi}^2/V \ll 1$$ « slow roll »

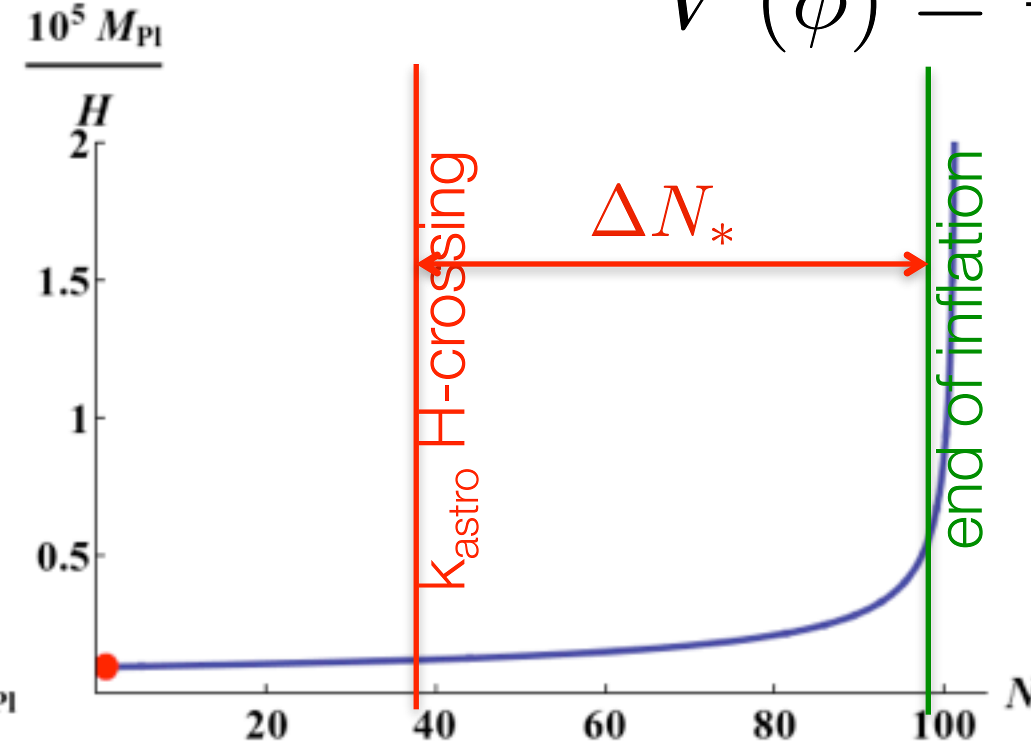
$$\epsilon_1 \simeq \frac{1}{2M_{\rm Pl}^2}\left(\frac{V_\phi}{V}\right)^2 \qquad \epsilon_2 \simeq \frac{2}{M_{\rm Pl}^2}\left[\left(\frac{V_\phi}{V}\right)^2 - \frac{V_{\phi\phi}}{V}\right] \qquad \epsilon_3 \simeq \text{etc...}$$

Dynamics

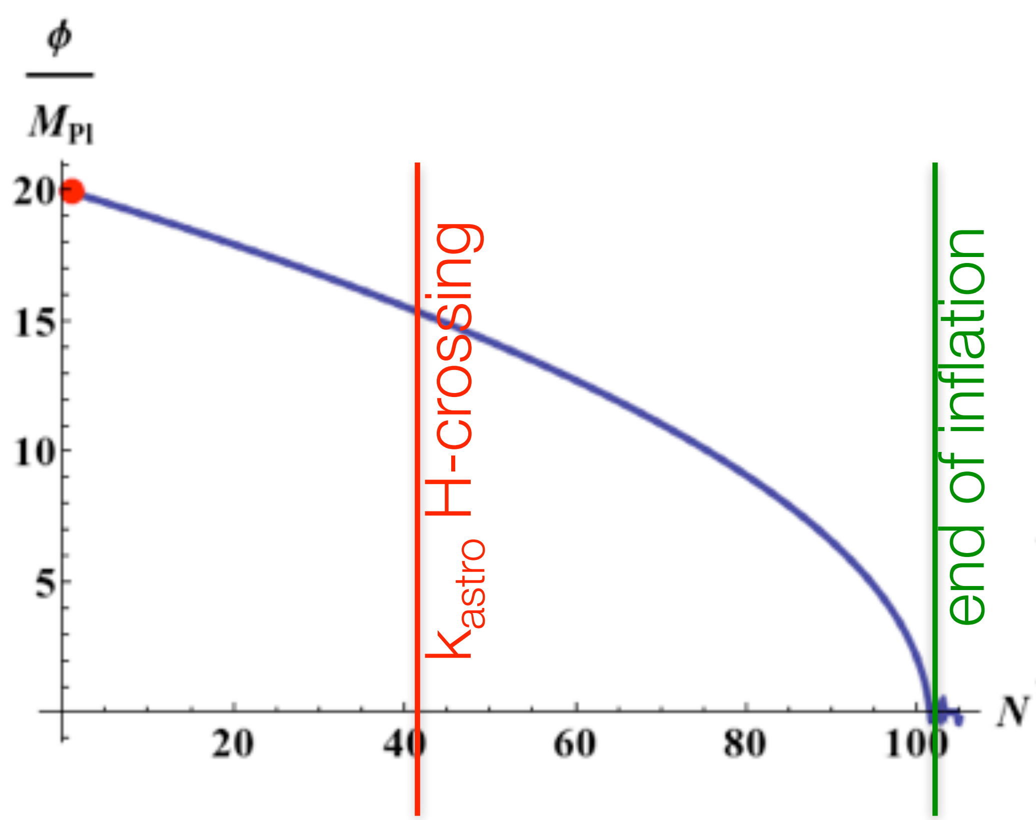
$$V(\phi) = \frac{m^2}{2}\phi^2$$

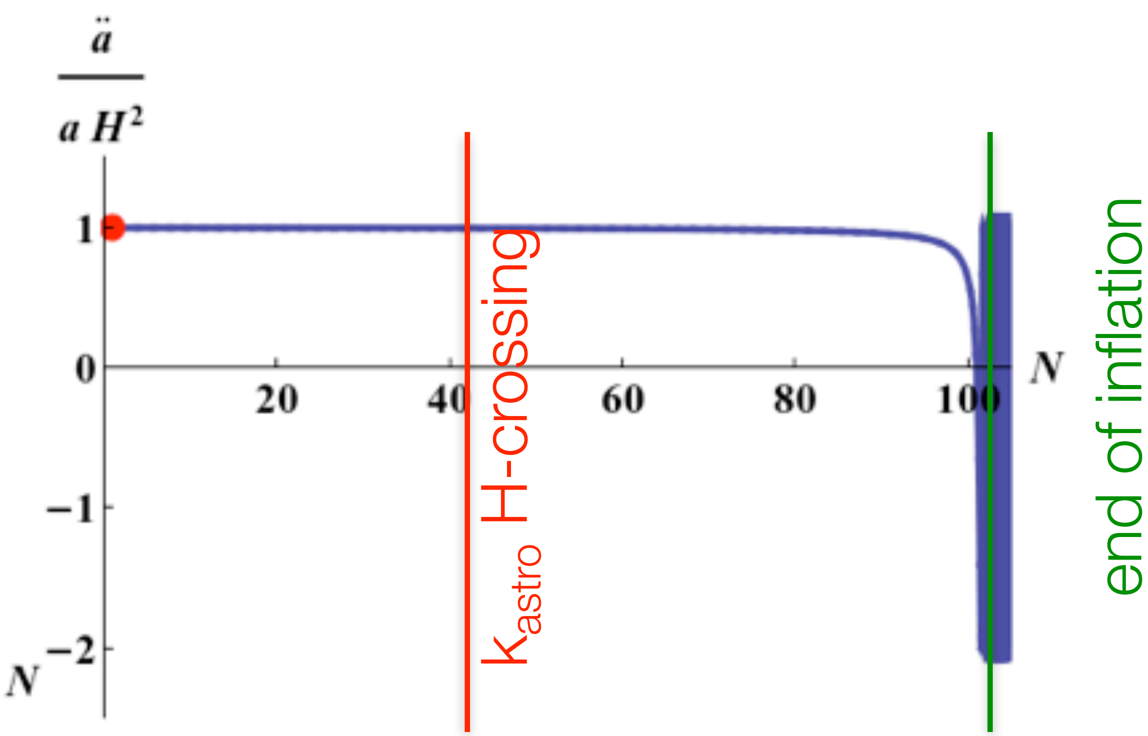


Hubble length



Field evolution

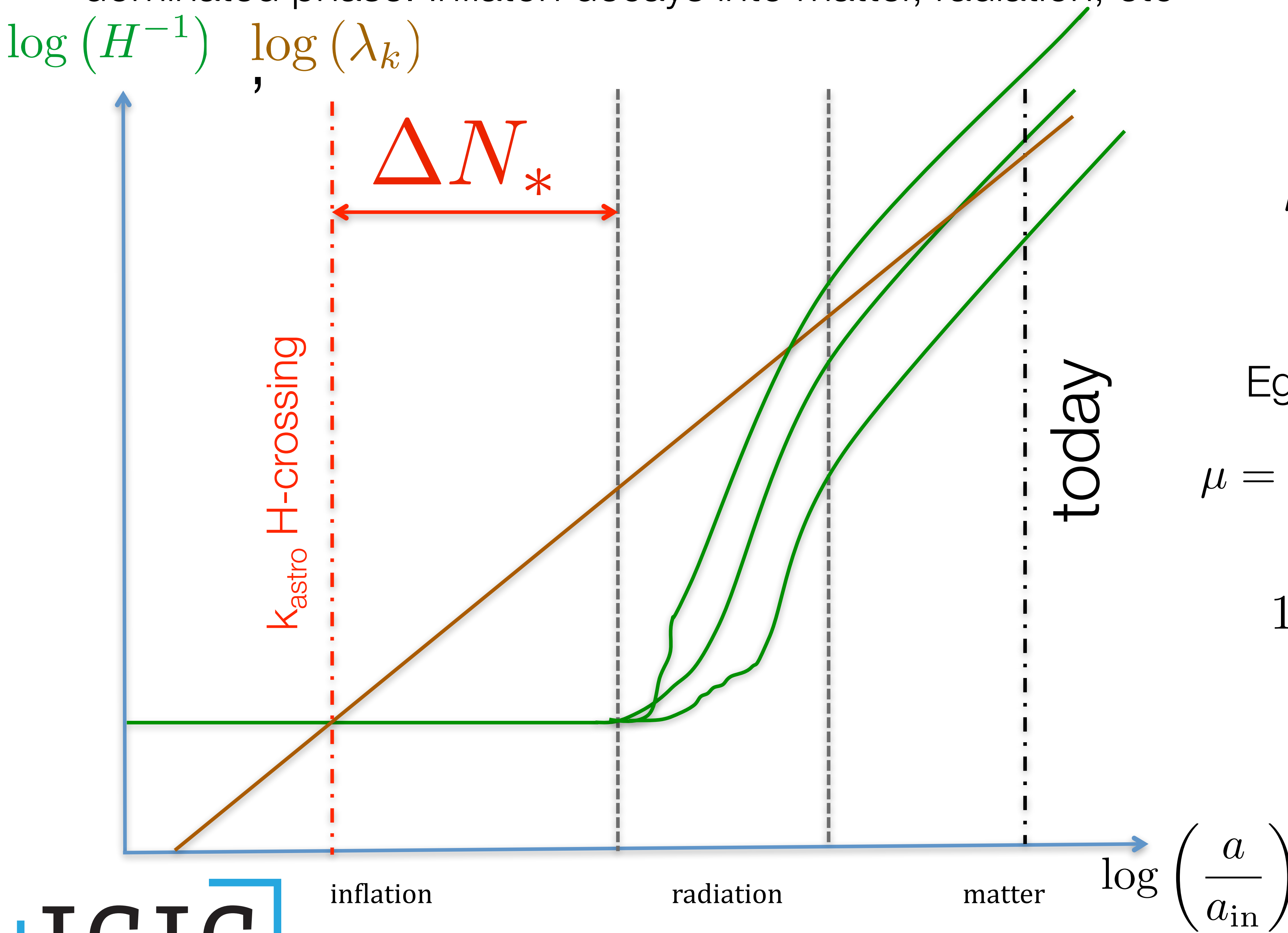




Reheating

Imperial College London

"Reheating" describes the transition from the end of inflation to the usual radiation dominated phase. Inflaton decays into matter, radiation, etc



$$\Delta N_*\left(\bar{\rho}_{\mathrm{reh}}, \bar{w}_{\mathrm{reh}}\right)$$

$$\rho_{\mathrm{nuc}} < \bar{\rho}_{\mathrm{reh}} < \rho_{\mathrm{end}}$$

Constrains ΔN_*

Eg: $V = M^4\left(1 - \phi^3/\mu^3\right)$

$$\mu = 0.001 M_{\mathrm{Pl}}\,,\ \bar{w}_{\mathrm{reh}} = -0.3$$

$$17.2 < \Delta N_* < 46.0$$

$$\Delta N_* = 60\ ?$$

Planck coll. prior:

$$\Delta N_* \in [20, 90]$$

From inflation to CMB fluctuations

Imperial College London

- At the end of inflation, the quantum fluctuations in the inflaton field are transferred via reheating to the matter/radiation content of the Universe
- Single-field inflation = adiabatic fluctuations (ie, curvature perturbations)
- The power spectrum (= Fourier transform of the 2-point correlation function) can be computed as a function of the slow roll parameters:

$$\mathcal{P}_\zeta(k) \propto a_0(\epsilon_n) + a_1(\epsilon_n)\ln\left(\frac{k}{k_*}\right) + \frac{1}{2}a_2(\epsilon_n)\ln^2\left(\frac{k}{k_*}\right) + \dots$$

- ... and is one of the key observables in the Cosmic Microwave Background (CMB) maps.
- Eg. two key quantities ("summary statistics") are

Spectral index $n_\mathrm{S} = \left.\frac{d\ln P}{d\ln k}\right|_{k_*}$

$n_\mathrm{S}^\mathrm{Planck} \sim 0.96$

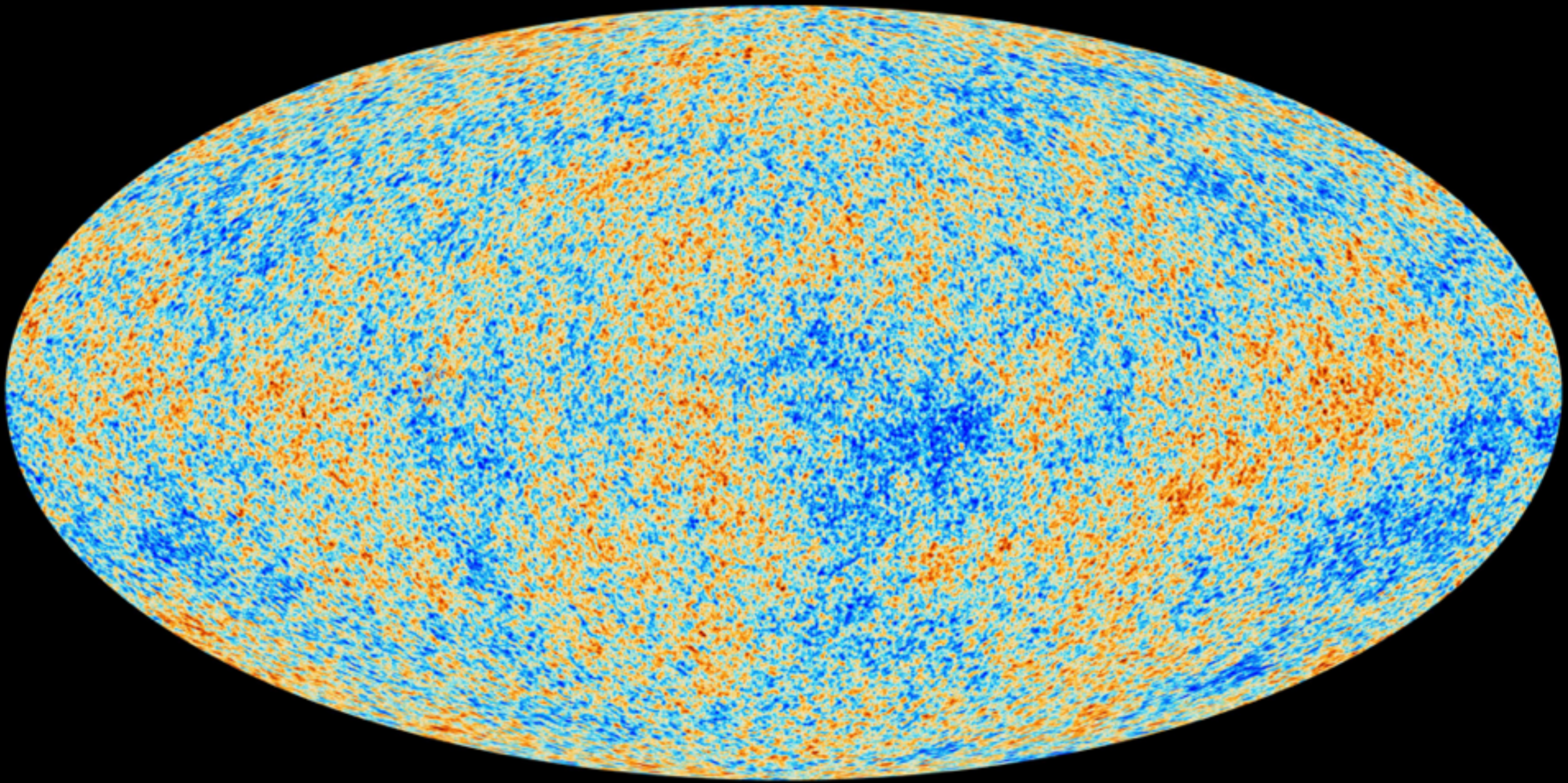
Tensor modes (gravity waves)

$$r = \frac{P_h(k_*)}{P_v(k_*)} = 16\epsilon_{1*} + \cdots$$

r < 0.1 (from Planck)

ICIC

Cosmic Microwave Background



Data from the Planck satellite, 2013

Phenomenological constraints

Imperial College London

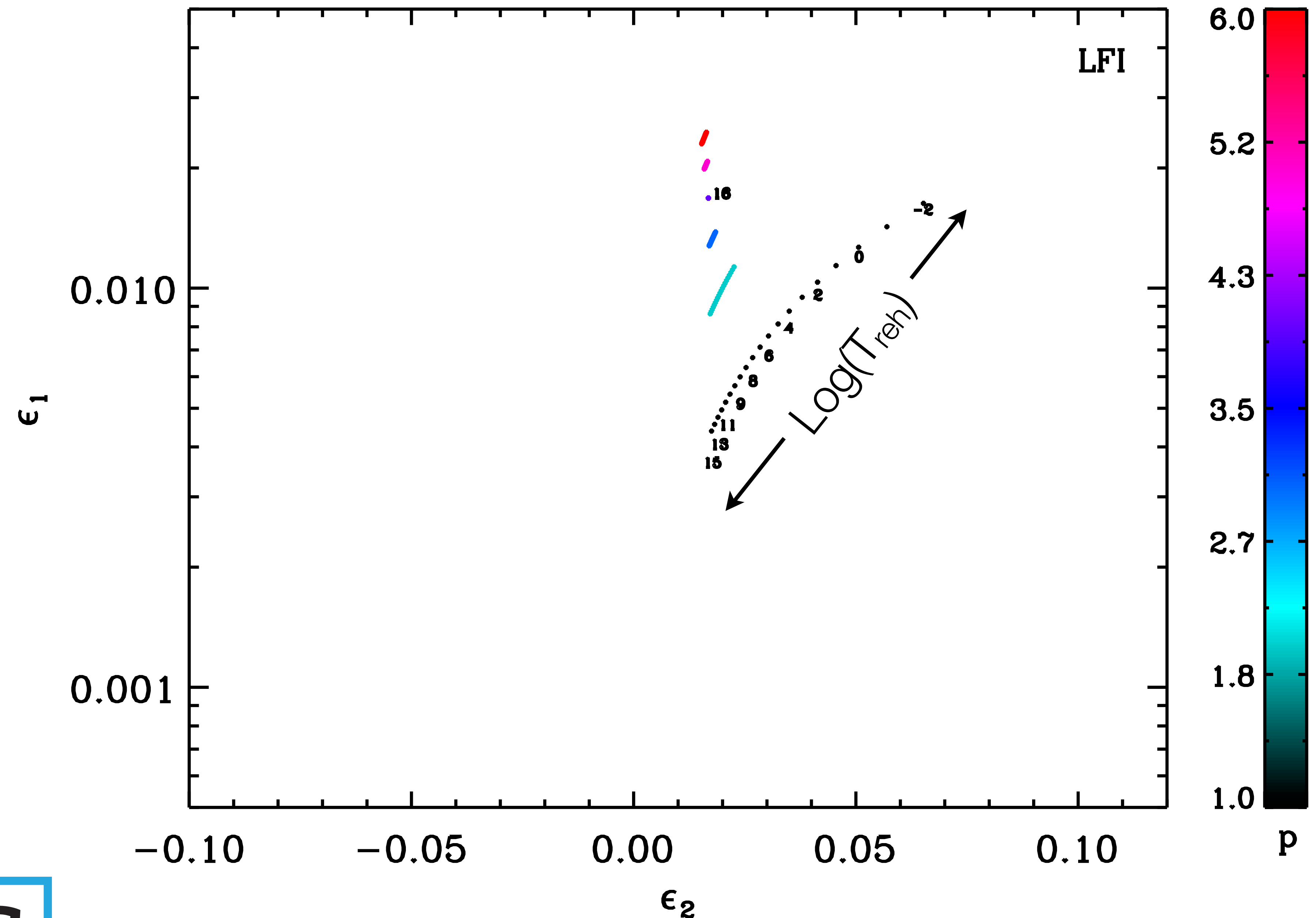
	Tilt	Tensors	Running	Non-Gauss.	Isocurvature
	n	r (95%CL)	α	f_{nl}	$\mathcal{I}/\mathcal{R}$ (95%CL)
COBE 2	1.21±0.57				
COBE 4	1.20±0.3				
WMAP 1	1.20±0.11	<0.81	-0.077±0.05	40±49	<32%
WMAP 3	0.984±0.029	<0.65	-0.055±0.03	30±42	
WMAP 5	0.960±0.013	<0.43	-0.037±0.028	51±30	<16%
WMAP 7	0.968±0.012	<0.36	-0.034±0.026	32±21	<13%
WMAP 9	0.9608±0.008	<0.13	-0.019±0.025	37.2±19.9	<15%
Planck 2013	0.9603±0.007	<0.11	-0.013±0.009	2.7±5.8	<3.6%
$\frac{\sigma_{WMAP1}}{\sigma_{Planck}}$	15.7	7.4	5.5	8.4	8.9

ICIC

Example: Large Field Inflation

Imperial College London

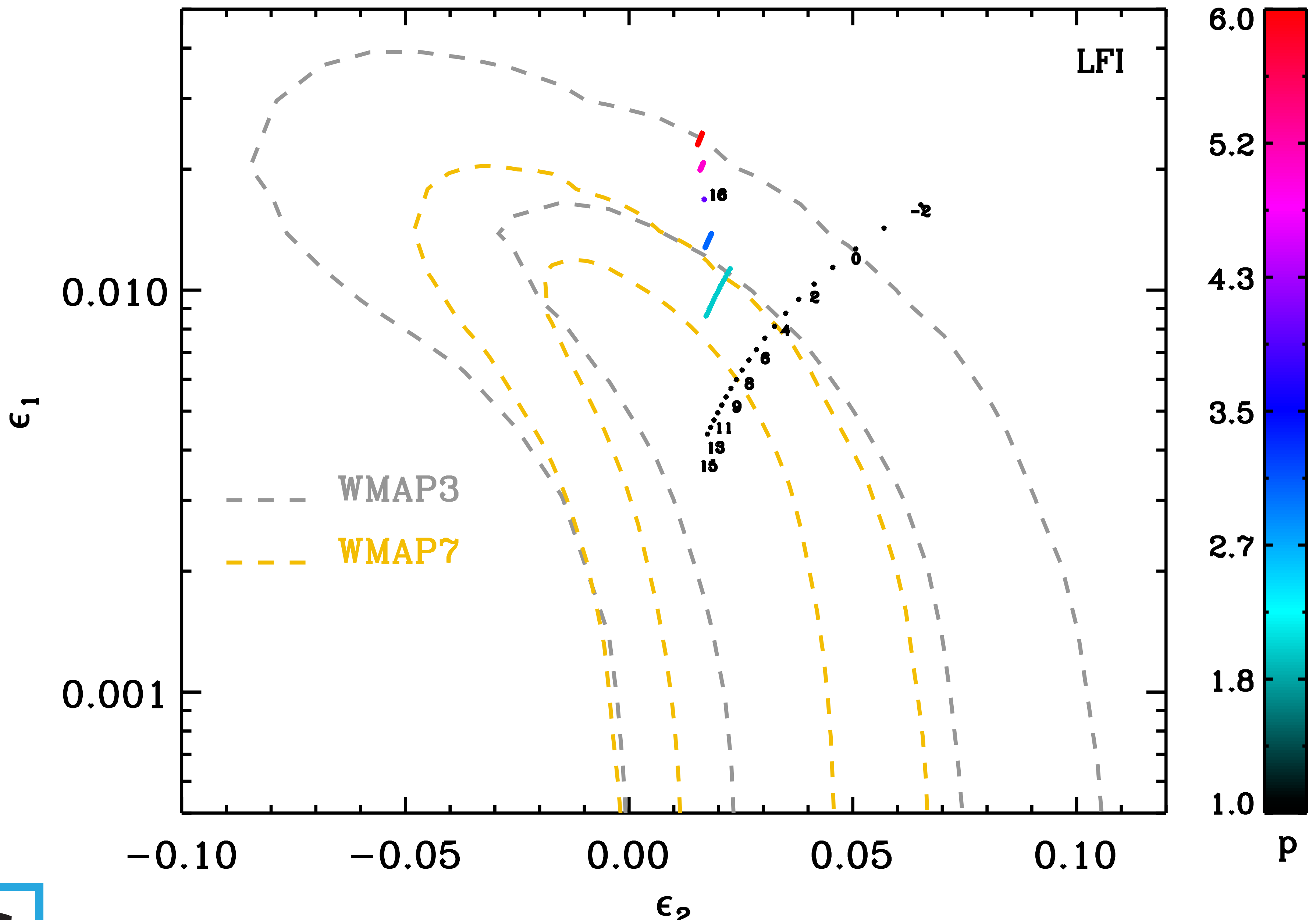
$$V(\phi) = M^4 \left(\frac{\phi}{M_{\rm Pl}} \right)^p$$



ICIC

Example: Large Field Inflation

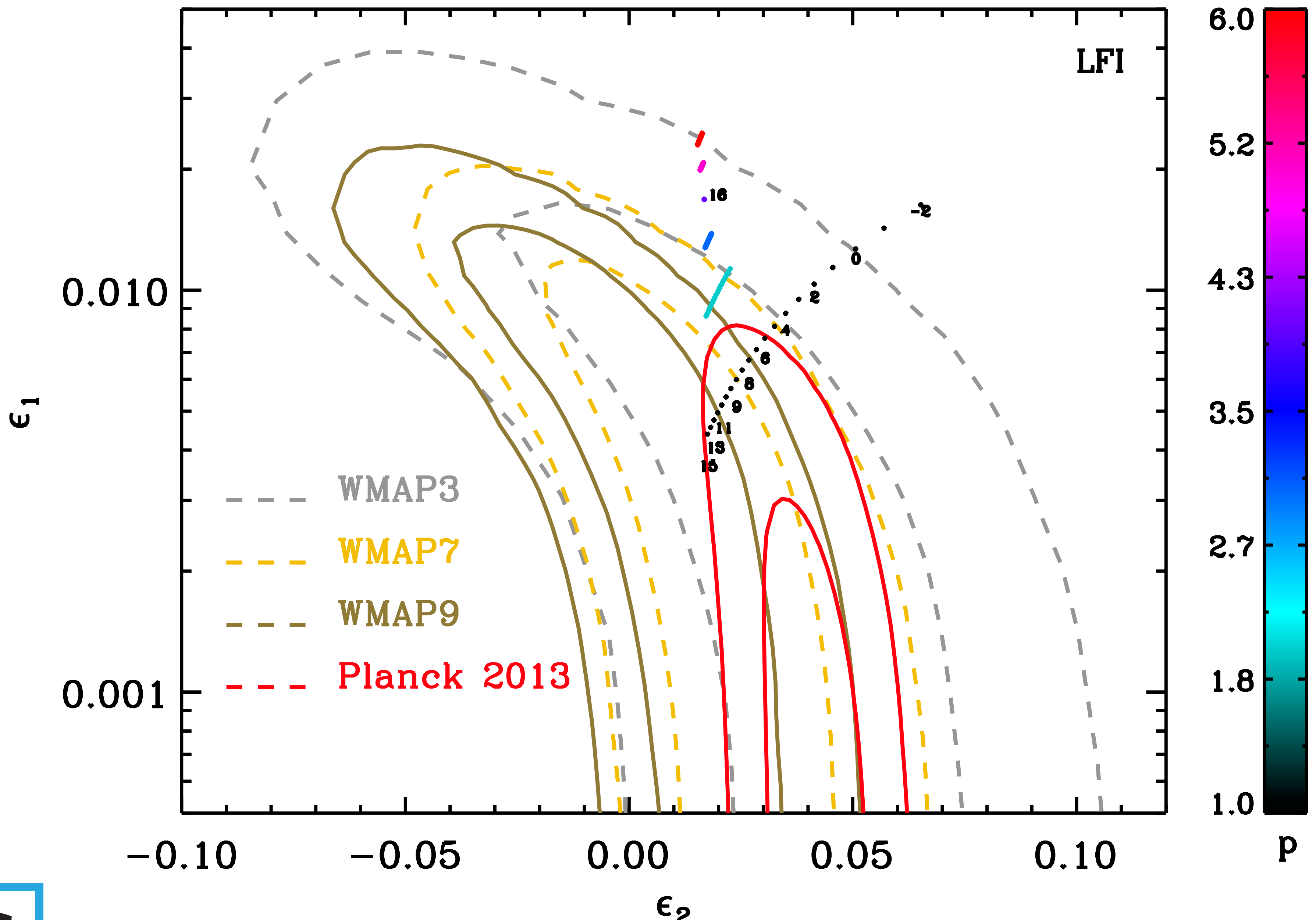
$$V(\phi) = M^4 \left(\frac{\phi}{M_{\rm Pl}} \right)^p$$



Example: Large Field Inflation

Imperial College London

$$V(\phi) = M^4 \left(\frac{\phi}{M_{\rm Pl}}\right)^p$$

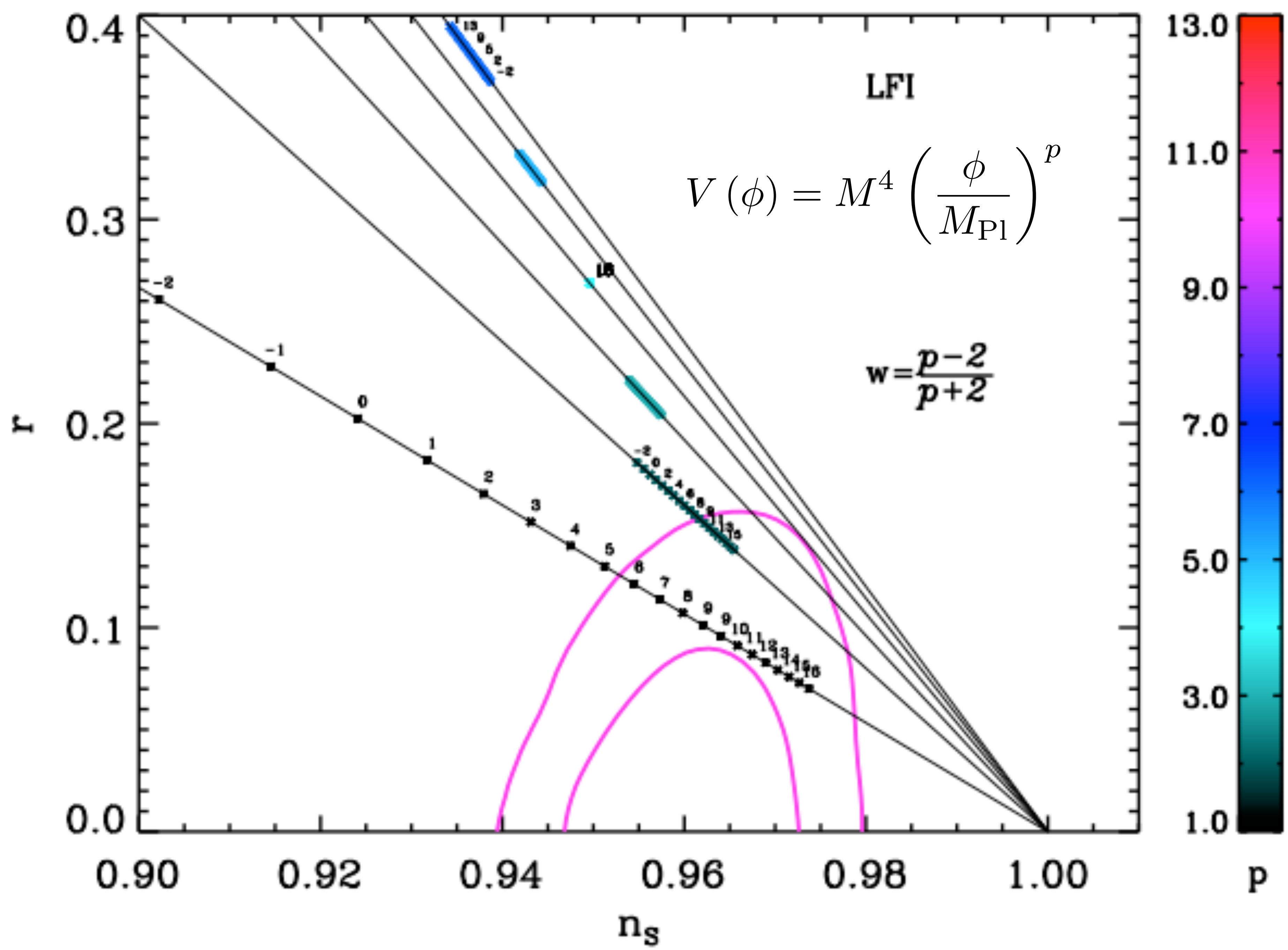


ICIC

Example: Large Field Inflation

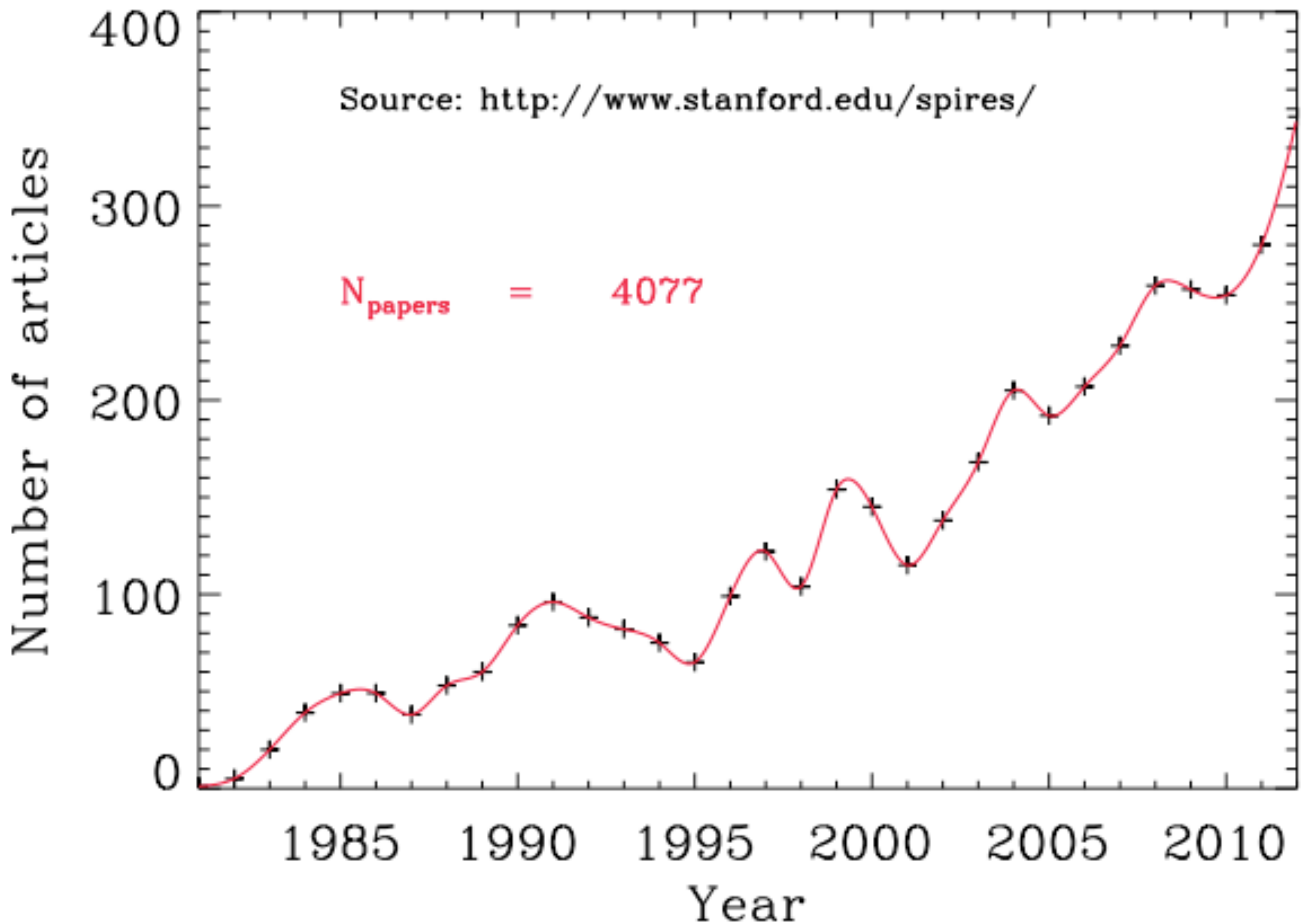
$$n_S = 1 - 2\epsilon_1 - \epsilon_2,$$

$$r \equiv \frac{\mathcal{P}_h}{\mathcal{P}_\zeta} = 16\epsilon_1$$



Inflation of "Inflationary" papers

Imperial College London



[1303.3787]

ASPIC =
Accurate **S**low-roll
Predictions for **I**nflationary **C**osmology

Encyclopædia Inflationaris

Jérôme Martin,[a] Christophe Ringeval[b] and Vincent Vennin[a]

[a]Institut d'Astrophysique de Paris, UMR 7095-CNRS, Université Pierre et Marie Curie, 98bis boulevard Arago, 75014 Paris (France)
[b]Centre for Cosmology, Particle Physics and Phenomenology, Institute of Mathematics and Physics, Louvain University, 2 Chemin du Cyclotron, 1348 Louvain-la-Neuve (Belgium)

E-mail: jmartin@iap.fr, christophe.ringeval@uclouvain.be, vennin@iap.fr

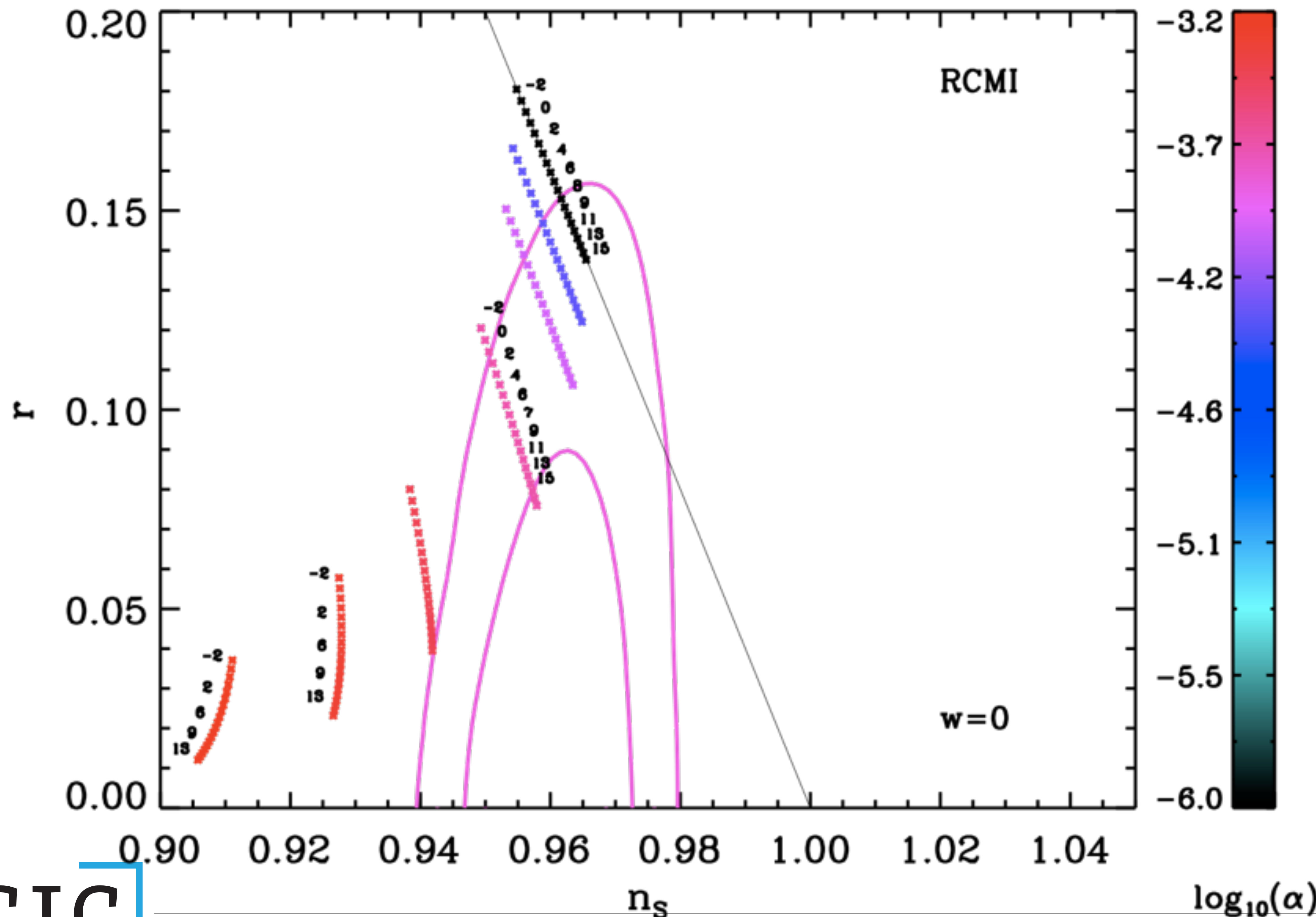
Keywords: Cosmic Inflation, Slow-Roll, Reheating, Cosmic Microwave Background, Aspic

≈ 70 models

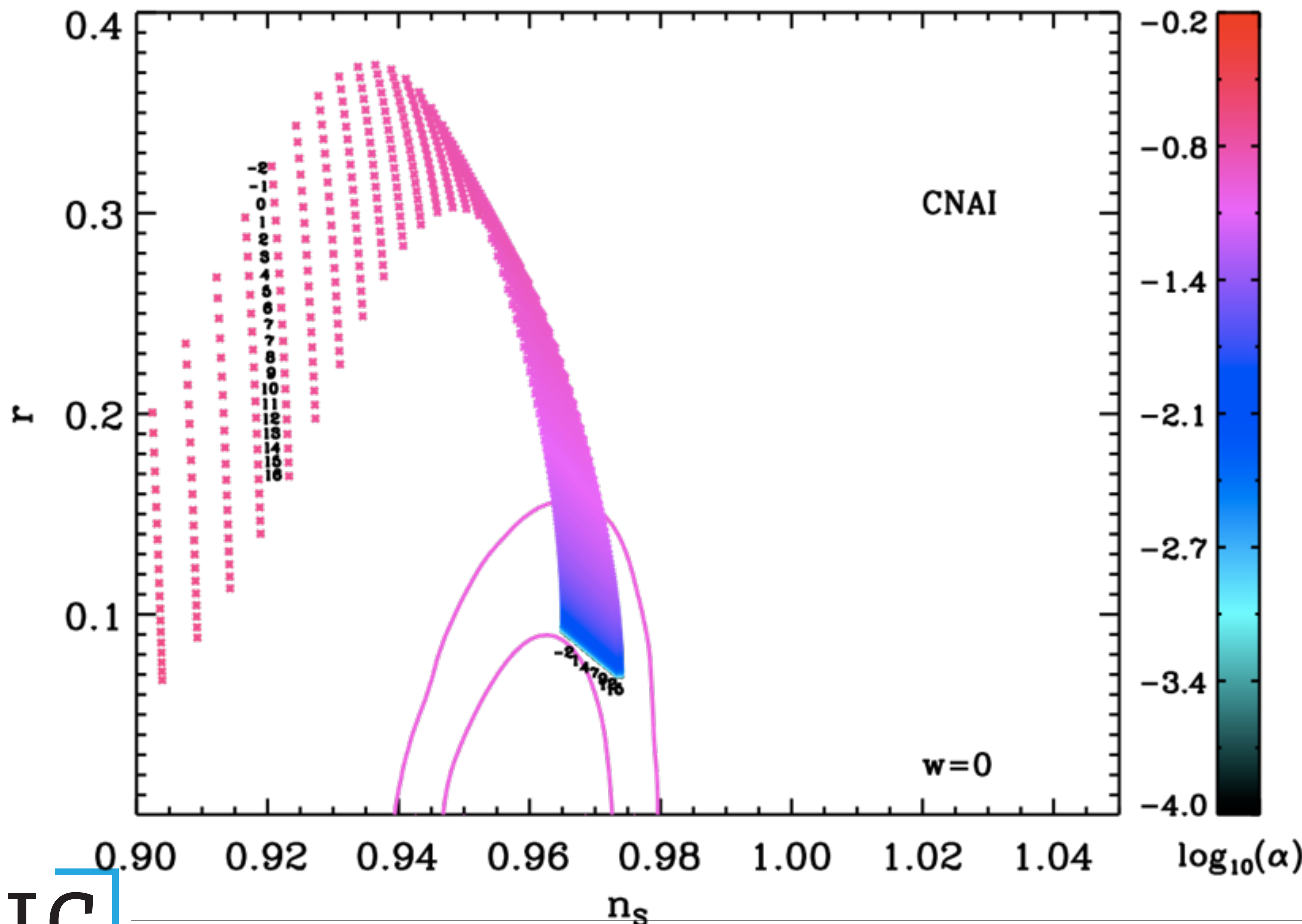
≈ 700 slow roll formulas

≈ 320 pages

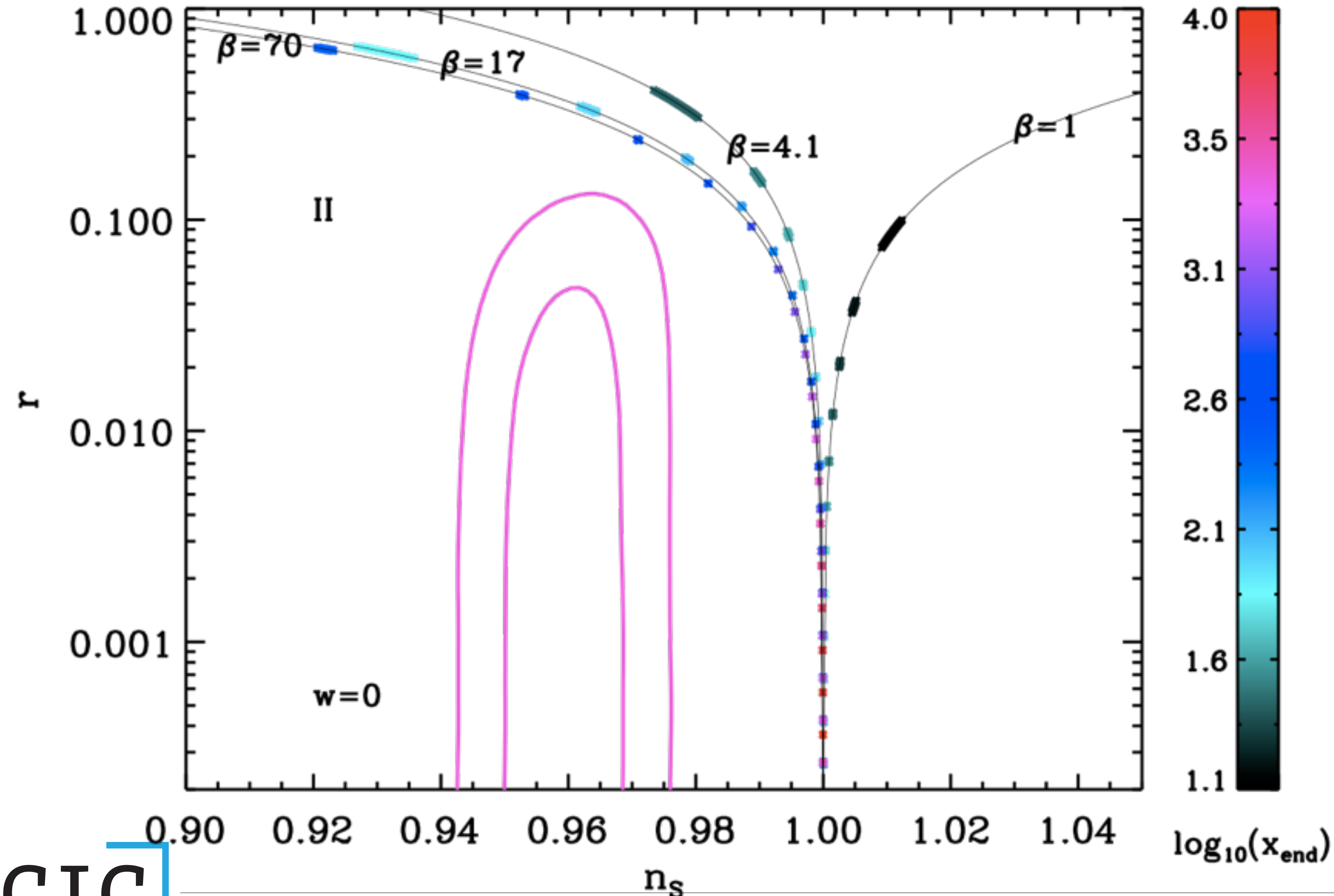
Imperial College London

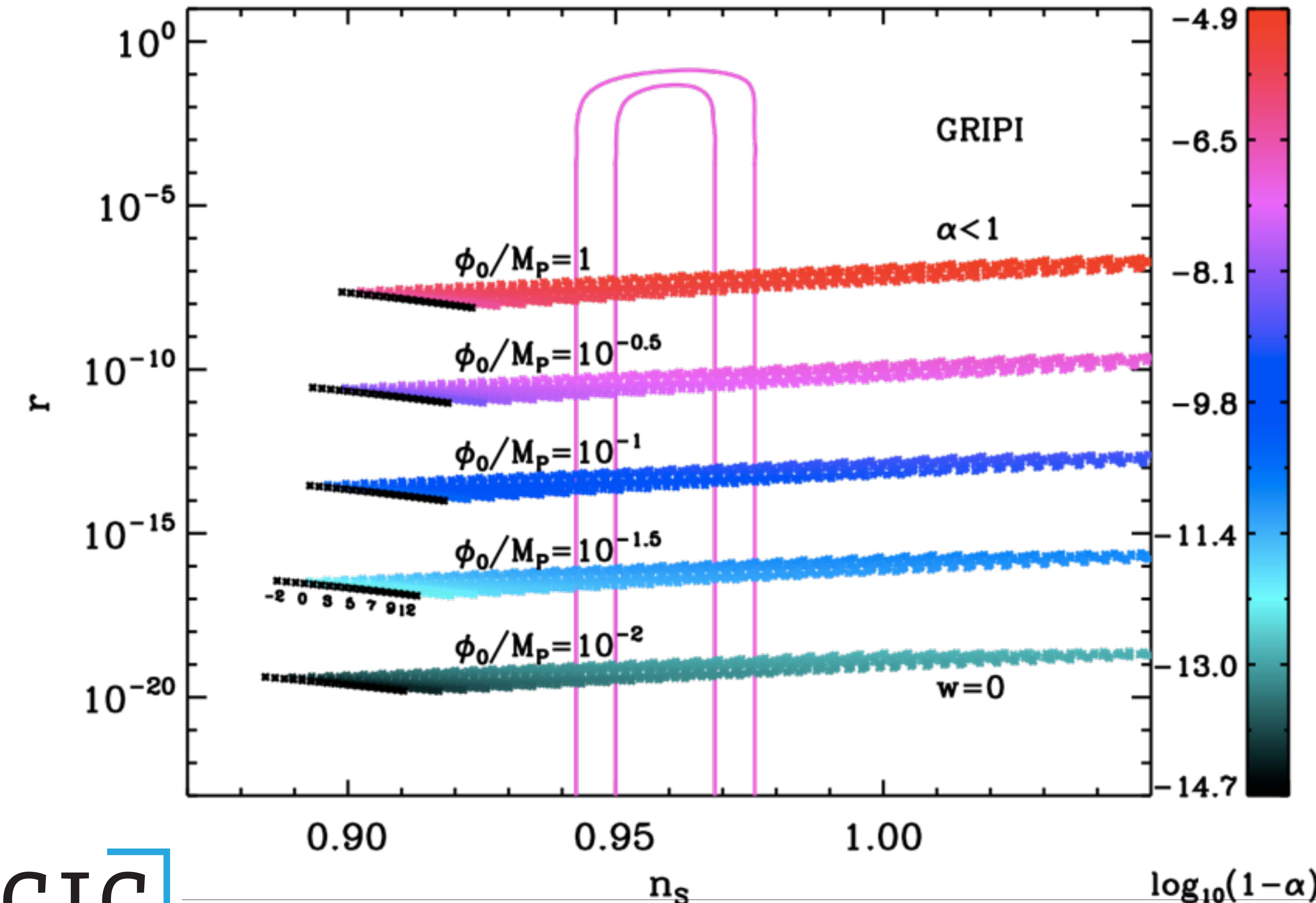


ICIC

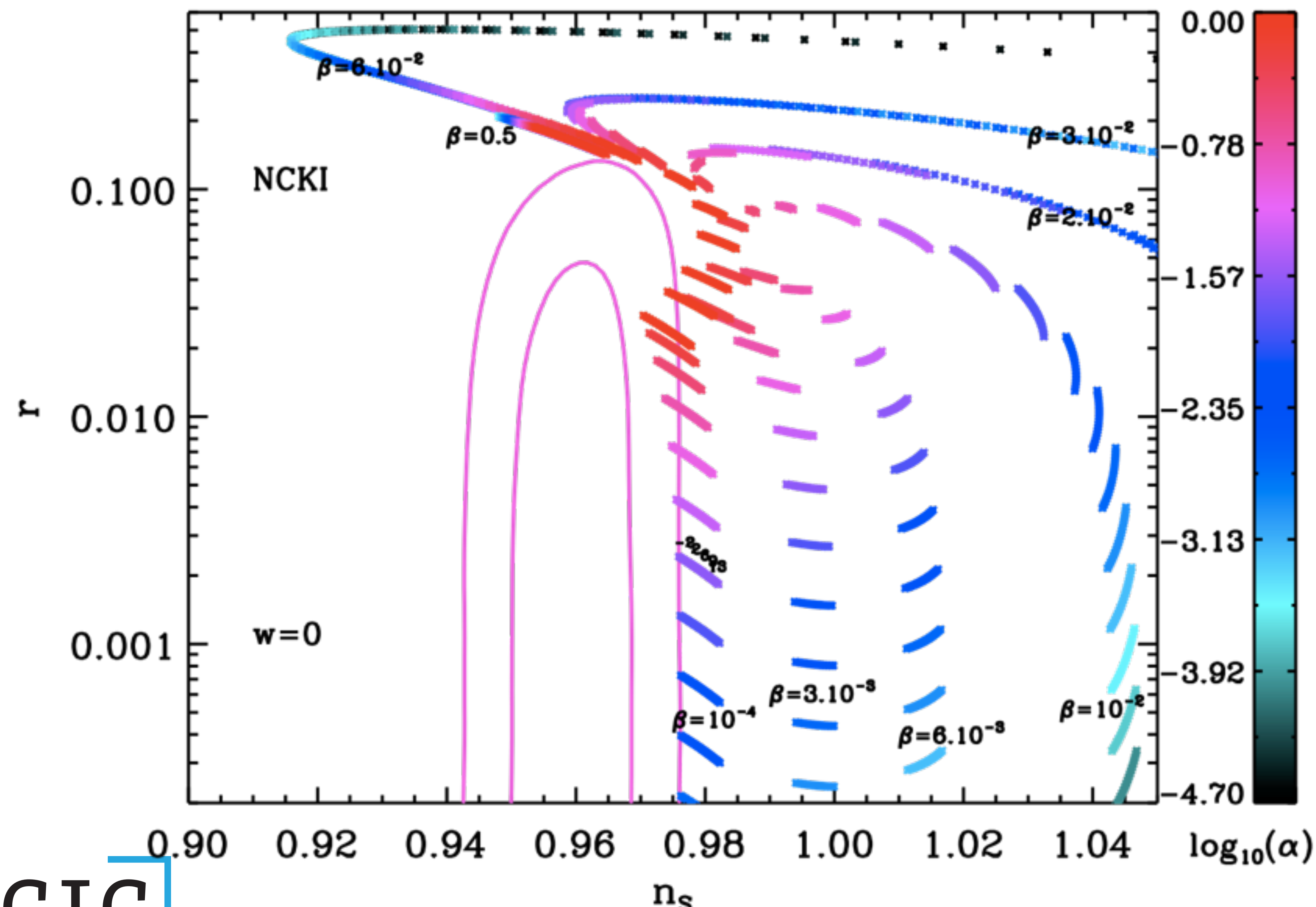


CNAI
w=0
r
n_s
$\log_{10}(\alpha)$
0.1
0.2
0.3
0.4
0.90
0.92
0.94
0.96
0.98
1.00
1.02
1.04
−0.2
−0.8
−1.4
−2.1
−2.7
−3.4
−4.0


1.000
0.100
0.010
0.001
r
0.90
0.92
0.94
0.96
0.98
1.00
1.02
1.04
n_S
β=70
β=17
β=4.1
β=1
II
w=0
4.0
3.5
3.1
2.6
2.1
1.6
1.1
$\log_{10}(x_{end})$



GRIPI
α<1
w=0
$\phi_0/M_P=1$
$\phi_0/M_P=10^{-0.5}$
$\phi_0/M_P=10^{-1}$
$\phi_0/M_P=10^{-1.5}$
$\phi_0/M_P=10^{-2}$
-2 0 3 5 7 9 12
r
10^0
10^{-5}
10^{-10}
10^{-15}
10^{-20}
0.90
0.95
1.00
n_S
-4.9
-6.5
-8.1
-9.8
-11.4
-13.0
-14.7
$\log_{10}(1-\alpha)$


β=6.10^-2
β=0.5
β=3.10^-2
β=2.10^-2
NCKI
w=0
β=10^-4
β=3.10^-3
β=6.10^-3
β=10^-2
r
0.100
0.010
0.001
0.90
0.92
0.94
0.96
0.98
1.00
1.02
1.04
n_S
0.00
0.78
1.57
2.35
3.13
3.92
4.70
log10(α)

Numerical shortcut

Ringeval [1312.2347]

Imperial College London

Theoretical prediction

$$C_\ell^{\rm th}\left(\theta_{\rm s},\theta_{\rm reh},\theta_{\rm inf}\right)=\int_0^{+\infty}\frac{{\rm d}k}{k}j_\ell(kr_{\ell\rm ss})T(k;\theta_{\rm s})\mathcal{P}_\zeta(k;\theta_{\rm reh},\theta_{\rm inf}),$$

θ_s: Standard LCDM parameters + nuisance (18)

θ_{reh}: Reheating parameter (1)

θ_{inf}: Inflationary potential parameters of interest (1-3)

Marginal likelihood

$$\mathcal{E}=\int{\rm d}\theta_{\rm s}{\rm d}\theta_{\rm reh}{\rm d}\theta_{\rm inf}\mathcal{L}\left(\theta_{\rm s},\theta_{\rm reh},\theta_{\rm inf}\right)\pi\left(\theta_{\rm s},\theta_{\rm reh},\theta_{\rm inf}\right)$$

Strategy: numerical marginalization over θ_s and definition of a “Planck effective likelihood” via fast interpolators

ICIC

Effective likelihood

Imperial College London

- The likelihood only depends on inflationary physics via the phenomenological parameters P_* (amplitude) and slow-roll parameters ε_n
- Map the likelihood onto the phenomenological parameters, then numerically marginalize out the standard cosmological parameters

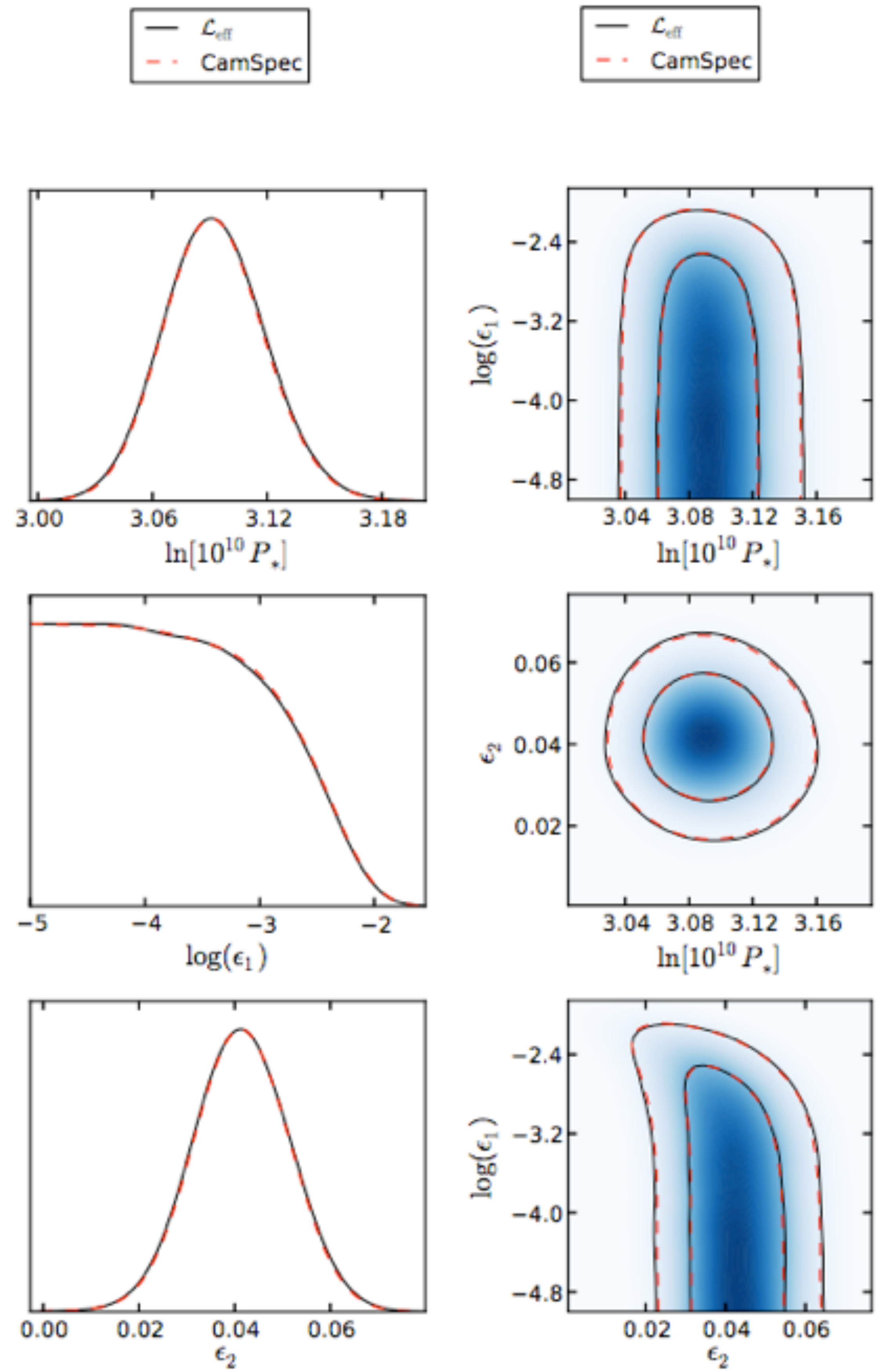
$$\mathcal{E} = \int \mathrm{d}\theta_{\mathrm{s}} \mathrm{d}\theta_{\mathrm{reh}} \mathrm{d}\theta_{\mathrm{inf}} \mathcal{L}\left[\theta_{\mathrm{s}}, P_*(\theta_{\mathrm{reh}}, \theta_{\mathrm{inf}}), \epsilon_n(\theta_{\mathrm{reh}}, \theta_{\mathrm{inf}})\right] \pi(\theta_{\mathrm{s}}) \pi(\theta_{\mathrm{reh}}, \theta_{\mathrm{inf}})$$

$$= \int \mathrm{d}\theta_{\mathrm{reh}} \mathrm{d}\theta_{\mathrm{inf}} \mathcal{L}_{\mathrm{eff}}\left[P_*(\theta_{\mathrm{reh}}, \theta_{\mathrm{inf}}), \epsilon_n(\theta_{\mathrm{reh}}, \theta_{\mathrm{inf}})\right] \pi(\theta_{\mathrm{reh}}) \pi(\theta_{\mathrm{inf}}),$$

- For each inflationary model, map the potential parameters onto the functionals P_* (and ε_n . Now the remaining parameter space is at most 4 dimensional.
- Dramatic speed-up: < 1 μs/likelihood evaluation, ~ 1 CPU hour for the full marginal likelihood.

Comparison

- Comparison with the traditional method shows excellent agreement in the marginal posterior distributions for the slow-roll parameters
- Speed-up is of several orders of magnitude
- Full marginal likelihood can now be obtained with O(100,000) likelihood evaluations in a 4D parameter space



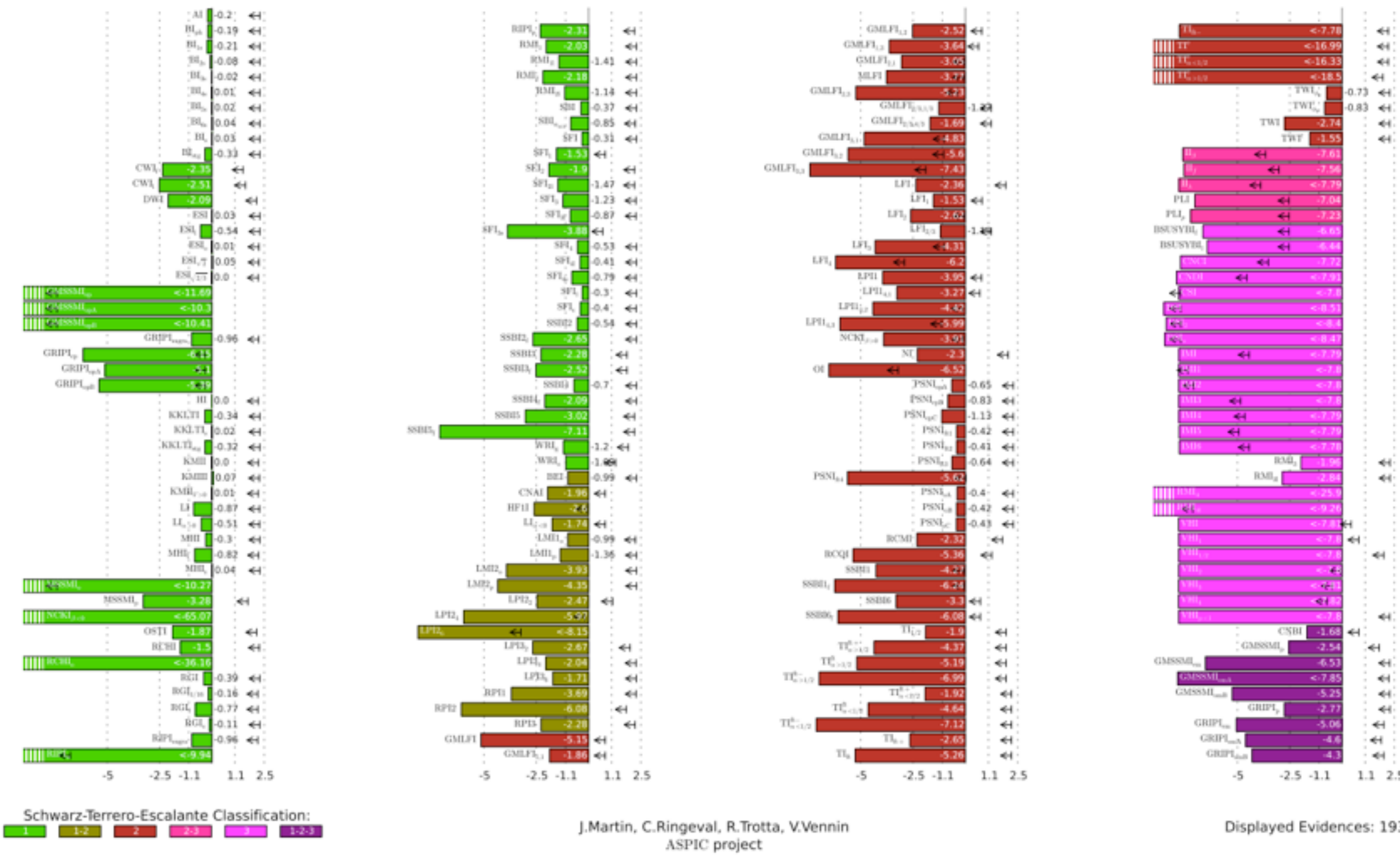
Priors for inflationary parameters

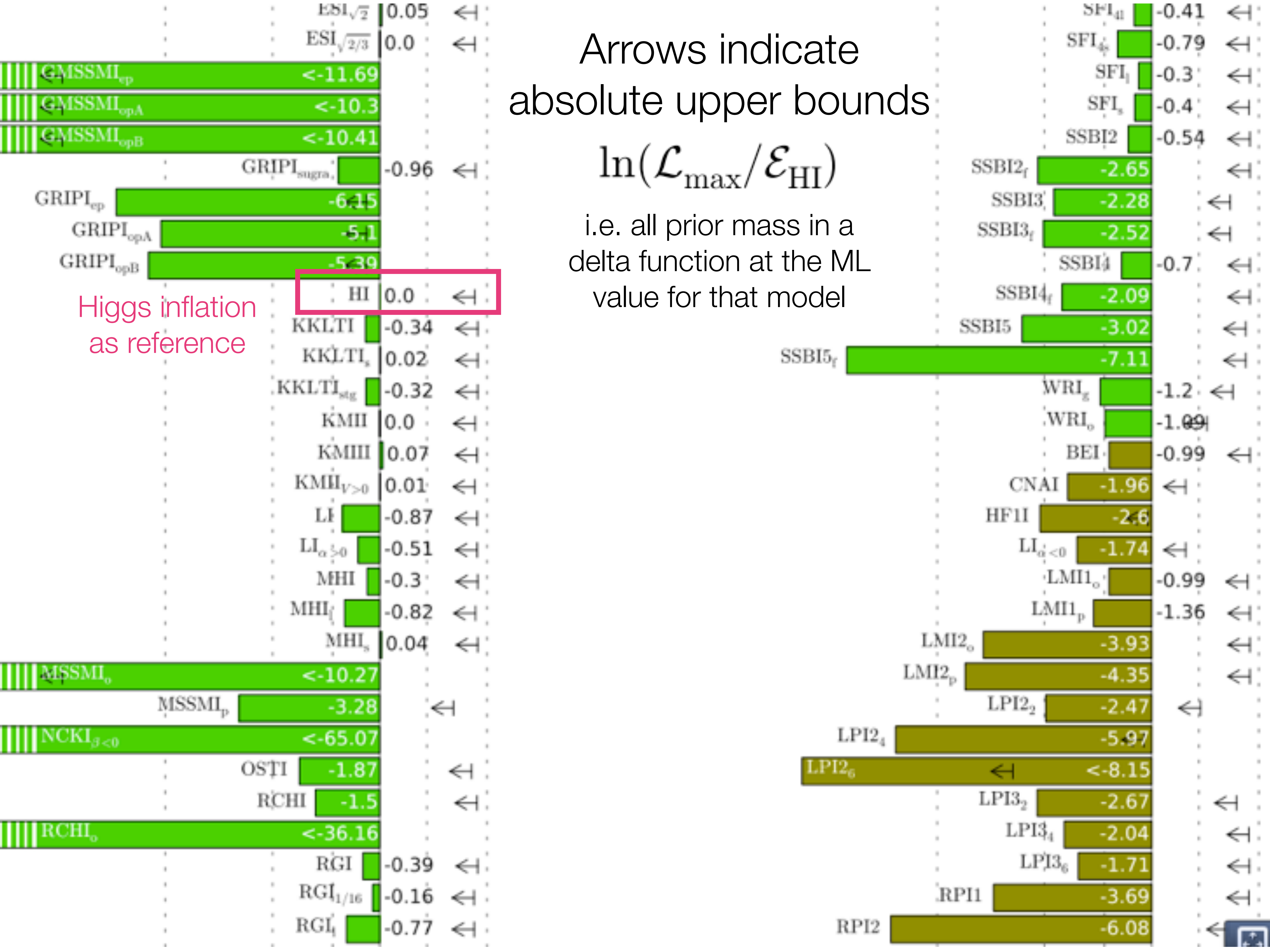
Imperial College London

- The choice of priors for the inflaton potential parameters is crucial for the outcome of the Bayesian model comparison
- Prior shape and width controls the strength of the Occam's razor effect
- Should therefore be motivated by theoretical scenario (i.e. underlying physics)
- 70 potential shapes (and associated parameters): some are split by making different choices of priors, giving a total of 193 "models"
- General rule: for parameters whose order of magnitude is unknown, we use priors uniform in the log of the quantity.
- Priors are proper - boundaries specified by theoretical/physical consideration
- Uniform prior on log of reheating parameter R, ensuring that reheating takes place after inflation and before BBN (and that the mean EOS satisfies $-1/3 < w < 1$).
- Prior on reheating parameter R and normalization are common to all models - their impact does not matter for the outcome of model comparison (SDDR for nested models).

Bayesian model comparison of 193 models
Higgs inflation as reference model

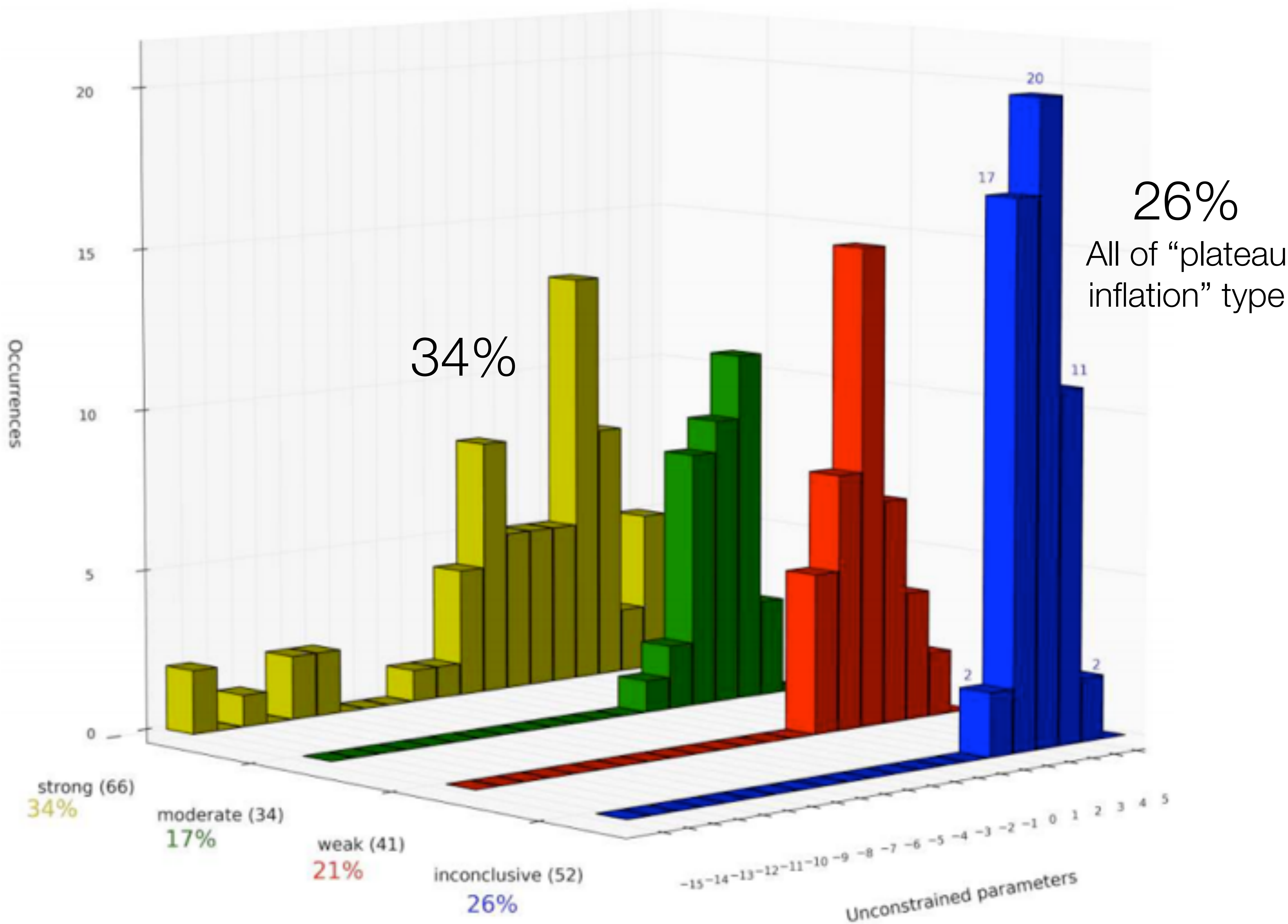
$$\ln(\mathcal{E}/\mathcal{E}_{\mathrm{HI}})$$



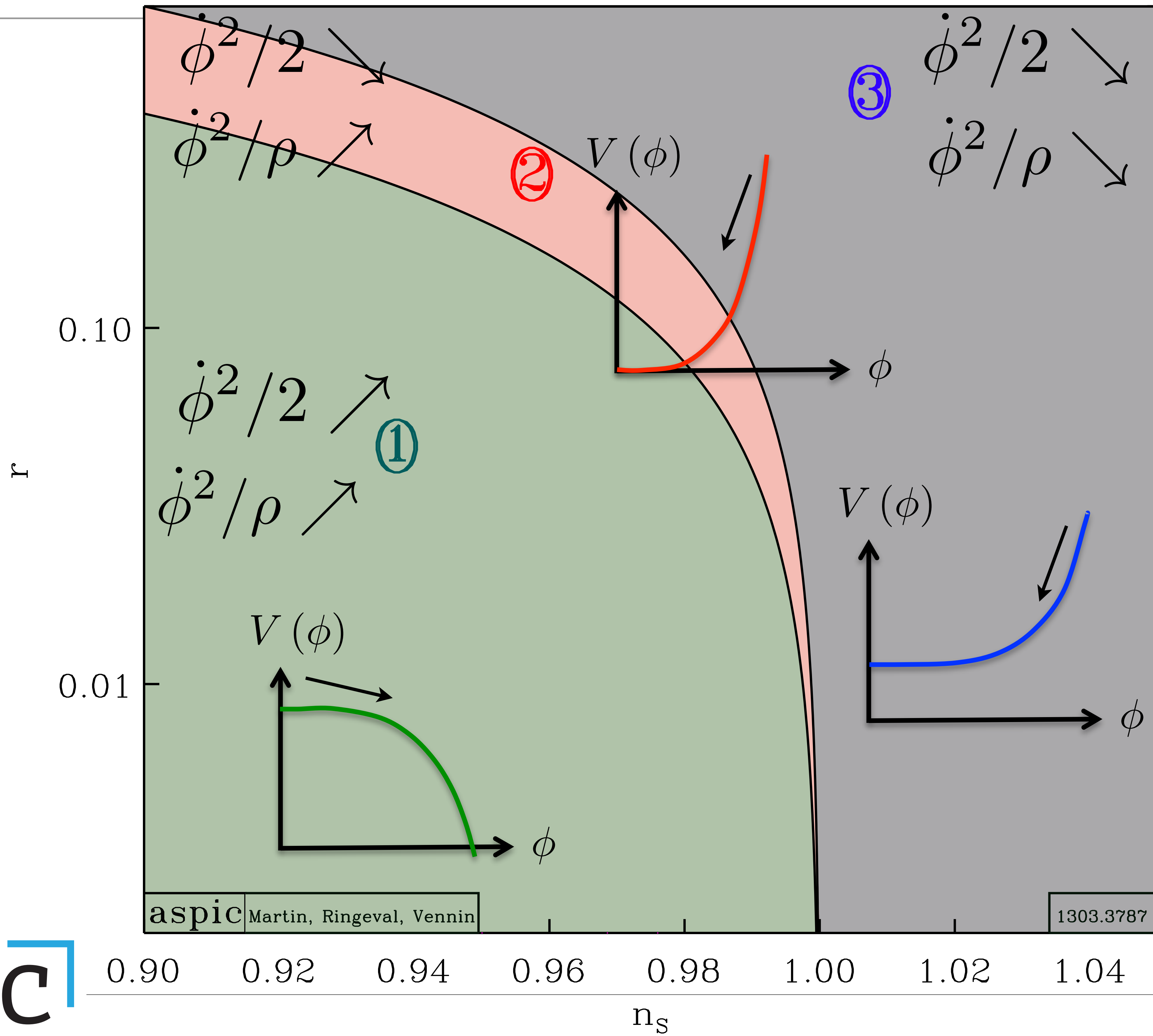


Arrows indicate absolute upper bounds
$\ln(\mathcal{L}_{\max}/\mathcal{E}_{\mathrm{HI}})$
i.e. all prior mass in a delta function at the ML value for that model
Higgs inflation as reference
$\mathrm{ESI}_{\sqrt{2}}$ 0.05
$\mathrm{ESI}_{\sqrt{2/3}}$ 0.0
$\mathrm{GMSSMI_{ep}}$ <-11.69
$\mathrm{GMSSMI_{opA}}$ <-10.3
$\mathrm{GMSSMI_{opB}}$ <-10.41
$\mathrm{GRIPI_{sugra}}$ -0.96
$\mathrm{GRIPI_{ep}}$ -6.15
$\mathrm{GRIPI_{opA}}$ -5.1
$\mathrm{GRIPI_{opB}}$ -5.39
HI 0.0
KKLTI -0.34
$\mathrm{KKLTI_s}$ 0.02
$\mathrm{KKLTI_{stg}}$ -0.32
KMII 0.0
KMIII 0.07
$\mathrm{KMII}_{V>0}$ 0.01
LI -0.87
$\mathrm{LI}_{\alpha>0}$ -0.51
MHI -0.3
$\mathrm{MHI_l}$ -0.82
$\mathrm{MHI_s}$ 0.04
$\mathrm{MSSMI_o}$ <-10.27
$\mathrm{MSSMI_p}$ -3.28
$\mathrm{NCKI}_{\beta<0}$ <-65.07
OSTI -1.87
RCHI -1.5
$\mathrm{RCHI_o}$ <-36.16
RGI -0.39
$\mathrm{RGI}_{1/16}$ -0.16
$\mathrm{RGI_l}$ -0.77
$\mathrm{SFI_{4l}}$ -0.41
$\mathrm{SFI_{4s}}$ -0.79
$\mathrm{SFI_l}$ -0.3
$\mathrm{SFI_s}$ -0.4
SSBI2 -0.54
$\mathrm{SSBI2_f}$ -2.65
SSBI3 -2.28
$\mathrm{SSBI3_f}$ -2.52
SSBI4 -0.7
$\mathrm{SSBI4_f}$ -2.09
SSBI5 -3.02
$\mathrm{SSBI5_f}$ -7.11
$\mathrm{WRI_g}$ -1.2
$\mathrm{WRI_o}$ -1.09
BEI -0.99
CNAI -1.96
HF1I -2.6
$\mathrm{LI}_{\alpha<0}$ -1.74
$\mathrm{LMI1_o}$ -0.99
$\mathrm{LMI1_p}$ -1.36
$\mathrm{LMI2_o}$ -3.93
$\mathrm{LMI2_p}$ -4.35
$\mathrm{LPI2_2}$ -2.47
$\mathrm{LPI2_4}$ -5.97
$\mathrm{LPI2_6}$ <-8.15
$\mathrm{LPI3_2}$ -2.67
$\mathrm{LPI3_4}$ -2.04
$\mathrm{LPI3_6}$ -1.71
RPI1 -3.69
RPI2 -6.08

Strength of evidences

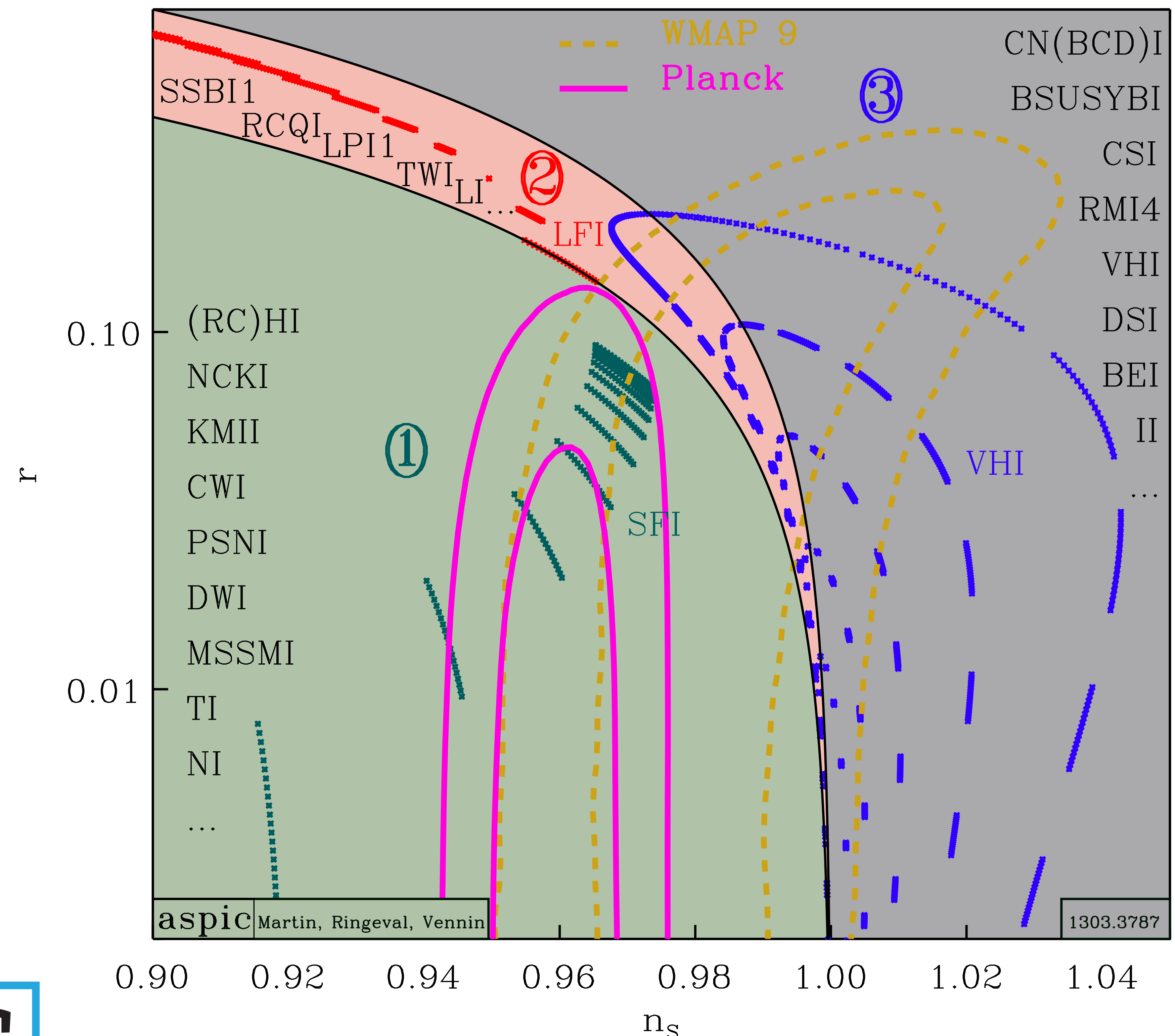


Models classification



Model classes vs constraints

Imperial College London



ICIC

Cosmological model comparison

Competing model	ΔN_{par}	$\ln B$	Ref	Data	Outcome
Initial conditions					
Isocurvature modes					
CDM isocurvature	+1	-7.6	[58]	WMAP3+, LSS	Strong evidence for adiabaticity
+ arbitrary correlations	+4	-1.0	[46]	WMAP1+, LSS, SN Ia	Undecided
Neutrino entropy	+1	$[-2.5, -6.5]^p$	[60]	WMAP3+, LSS	Moderate to strong evidence for adiabaticity
+ arbitrary correlations	+4	-1.0	[46]	WMAP1+, LSS, SN Ia	Undecided
Neutrino velocity	+1	$[-2.5, -6.5]^p$	[60]	WMAP3+, LSS	Moderate to strong evidence for adiabaticity
+ arbitrary correlations	+4	-1.0	[46]	WMAP1+, LSS, SN Ia	Undecided
Primordial power spectrum					
No tilt ($n_s = 1$)	-1	$+0.4$	[47]	WMAP1+, LSS	Undecided
		$[-1.1, -0.6]^p$	[51]	WMAP1+, LSS	Undecided
		-0.7	[58]	WMAP1+, LSS	Undecided
		-0.9	[70]	WMAP1+	Undecided
		$[-0.7, -1.7]^{p,d}$	[186]	WMAP3+	$n_s = 1$ weakly disfavoured
		-2.0	[185]	WMAP3+, LSS	$n_s = 1$ weakly disfavoured
		-2.6	[70]	WMAP3+	$n_s = 1$ moderately disfavoured
		-2.9	[58]	WMAP3+, LSS	$n_s = 1$ moderately disfavoured
		$< -3.9^c$	[65]	WMAP3+, LSS	Moderate evidence at best against $n_s \neq 1$
Running	+1	$[-0.6, 1.0]^{p,d}$	[186]	WMAP3+, LSS	No evidence for running
		$< 0.2^c$	[166]	WMAP3+, LSS	Running not required
Running of running	+2	$< 0.4^c$	[166]	WMAP3+, LSS	Not required
Large scales cut–off	+2	$[1.3, 2.2]^{p,d}$	[186]	WMAP3+, LSS	Weak support for a cut–off
Matter–energy content					
Non–flat Universe	+1	-3.8	[70]	WMAP3+, HST	Flat Universe moderately favoured
		-3.4	[58]	WMAP3+, LSS, HST	Flat Universe moderately favoured
Coupled neutrinos	+1	-0.7	[193]	WMAP3+, LSS	No evidence for non–SM neutrinos
Dark energy sector					
$w(z) = w_{\rm eff} \neq -1$	+1	$[-1.3, -2.7]^p$	[187]	SN Ia	Weak to moderate support for Λ
		-3.0	[50]	SN Ia	Moderate support for Λ
		-1.1	[51]	WMAP1+, LSS, SN Ia	Weak support for Λ
		$[-0.2, -1]^p$	[188]	SN Ia, BAO, WMAP3	Undecided
		$[-1.6, -2.3]^d$	[189]	SN Ia, GRB	Weak support for Λ
$w(z) = w_0 + w_1 z$	+2	$[-1.5, -3.4]^p$	[187]	SN Ia	Weak to moderate support for Λ
		-6.0	[50]	SN Ia	Strong support for Λ
		-1.8	[188]	SN Ia, BAO, WMAP3	Weak support for Λ
$w(z) = w_0 + w_a(1-a)$	+2	-1.1	[188]	SN Ia, BAO, WMAP3	Weak support for Λ
		$[-1.2, -2.6]^d$	[189]	SN Ia, GRB	Weak to moderate support for Λ
Reionization history					
No reionization ($\tau = 0$)	-1	-2.6	[70]	WMAP3+, HST	$\tau \neq 0$ moderately favoured
No reionization and no tilt	-2	-10.3	[70]	WMAP3+, HST	Strongly disfavoured

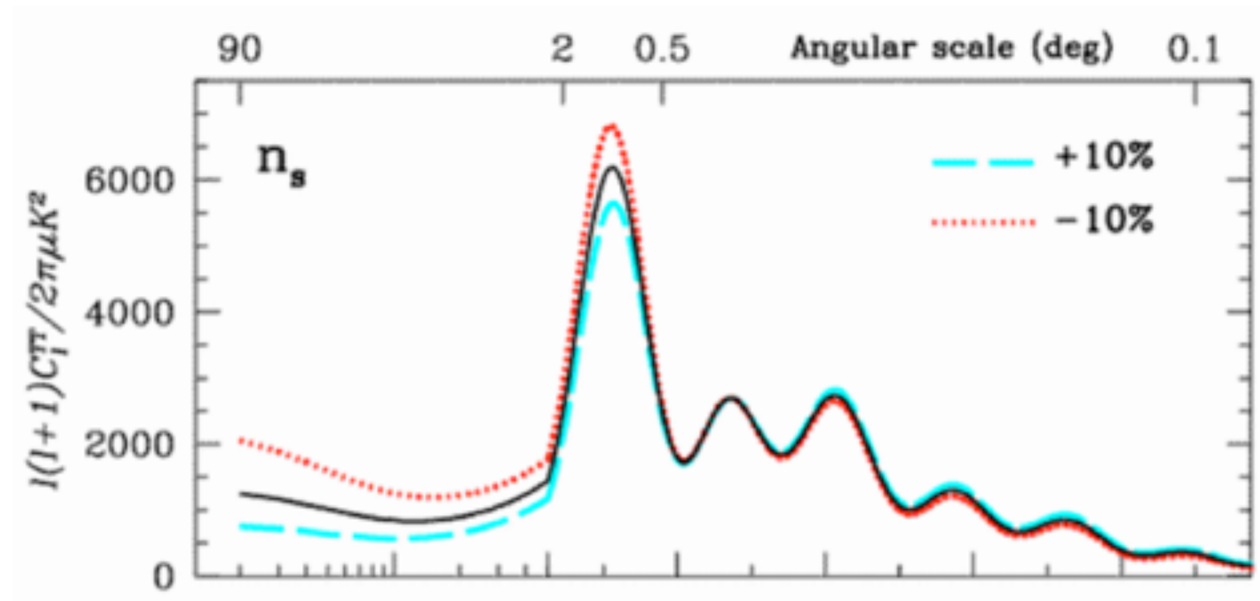
Trotta '08

lnB < 0: ΛCDM remains the "best" model from a Bayesian perspective!

Cosmological model selection

Imperial College London

- Is the spectrum of primordial fluctuations scale-invariant (n = 1)?
- Model comparison:
  n = 1 vs n ≠ 1 (with inflation-motivated prior)
- Results:
  n ≠ 1 favoured with odds of 17:1
  (Trotta 2007)
  n ≠ 1 favoured with odds of 15:1
  (Kunz, Trotta & Parkinson 2007)
  n ≠ 1 favoured with odds of 7:1
  (Parkinson 2007 et al 2006)

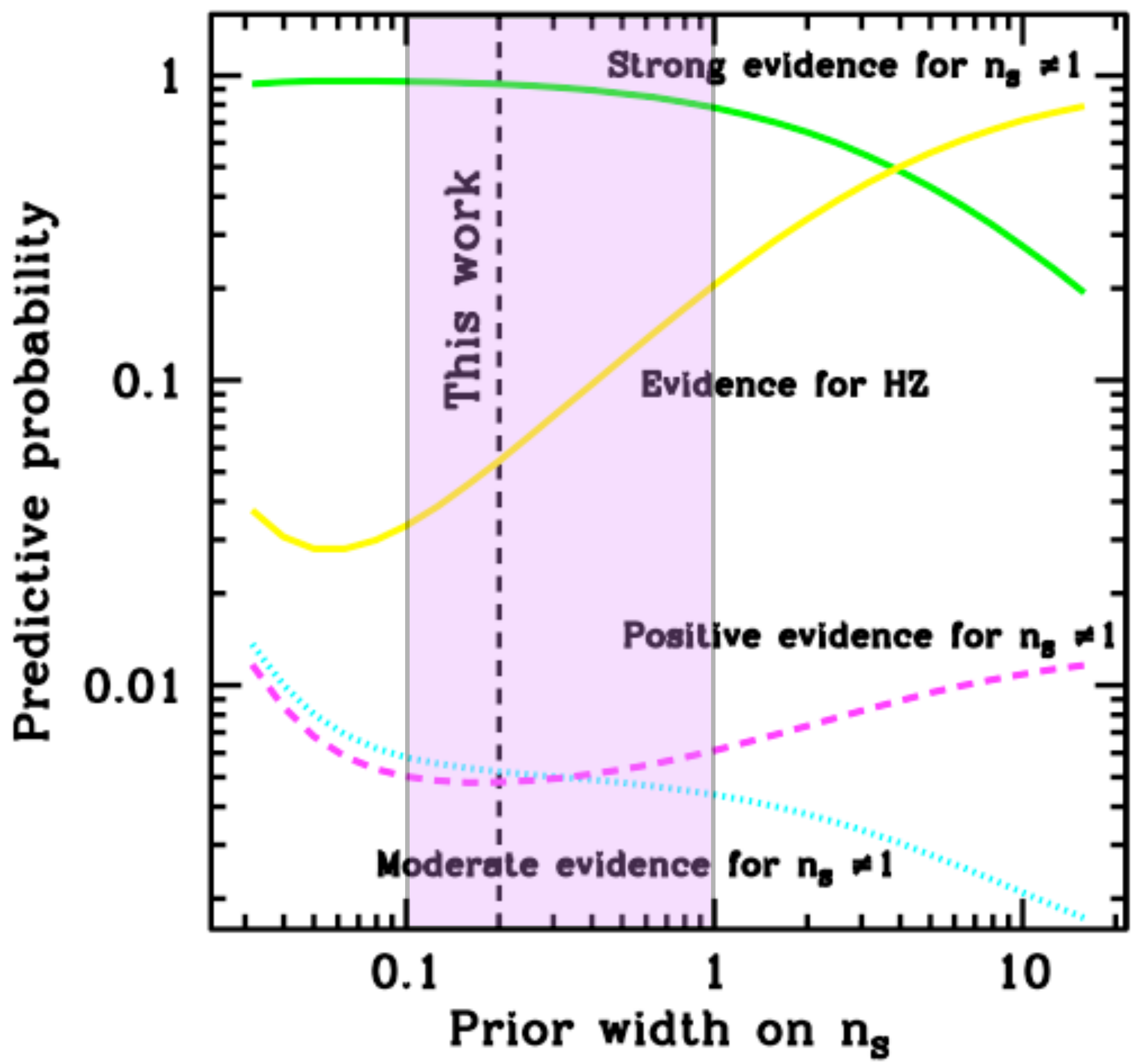


Example of reasonable sensitivity analysis

Imperial College London

- The favoured model (non-scale invariant CMB spectrum) is robust for physically reasonable changes (motivated by inflation) in the prior width

Inflationary range

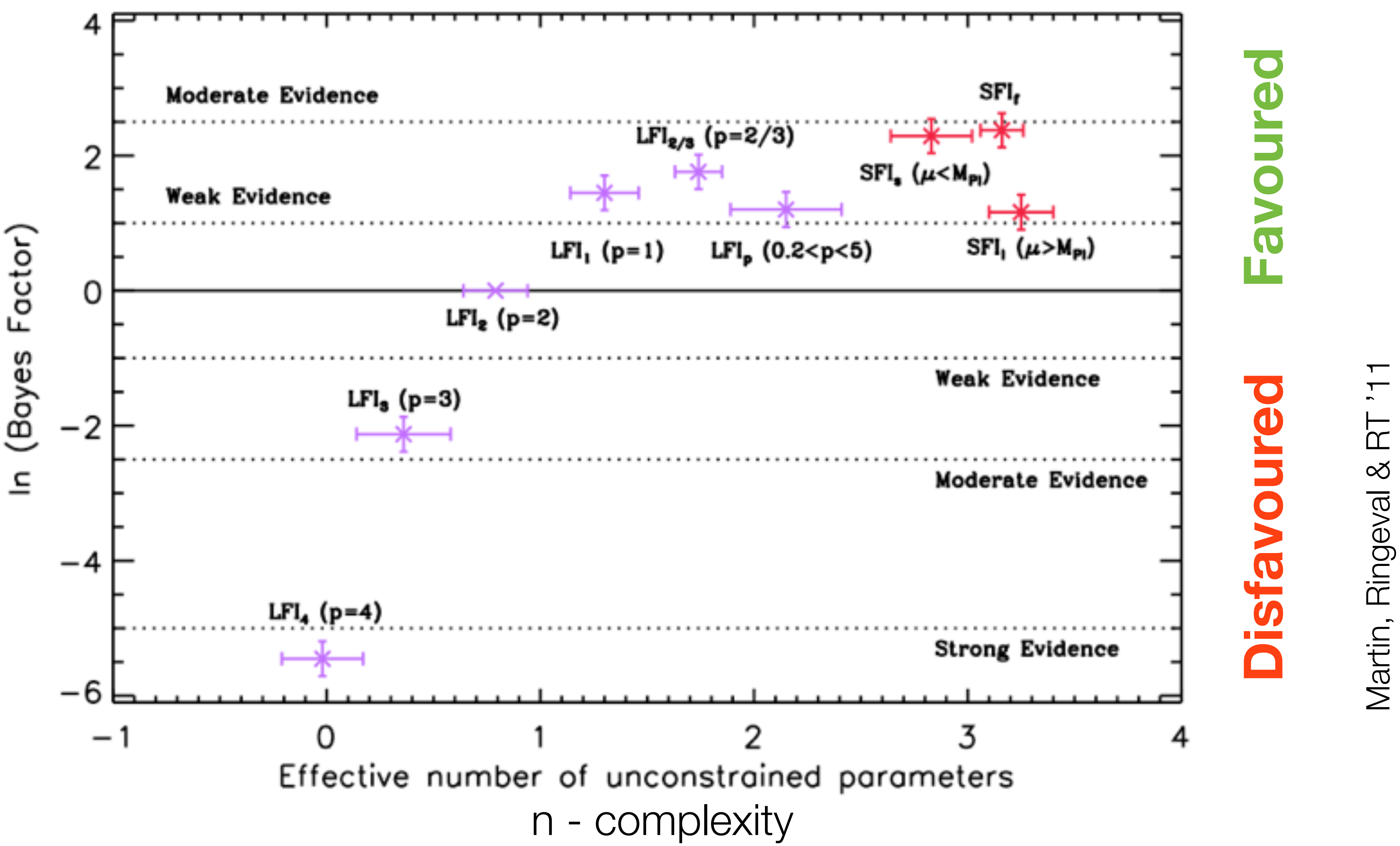


Trotta (2007)

Small field vs large field inflation

Imperial College London

The probability of small field models rises from an initial 50% to
P(small field | all data) = 0.77 ± 0.03



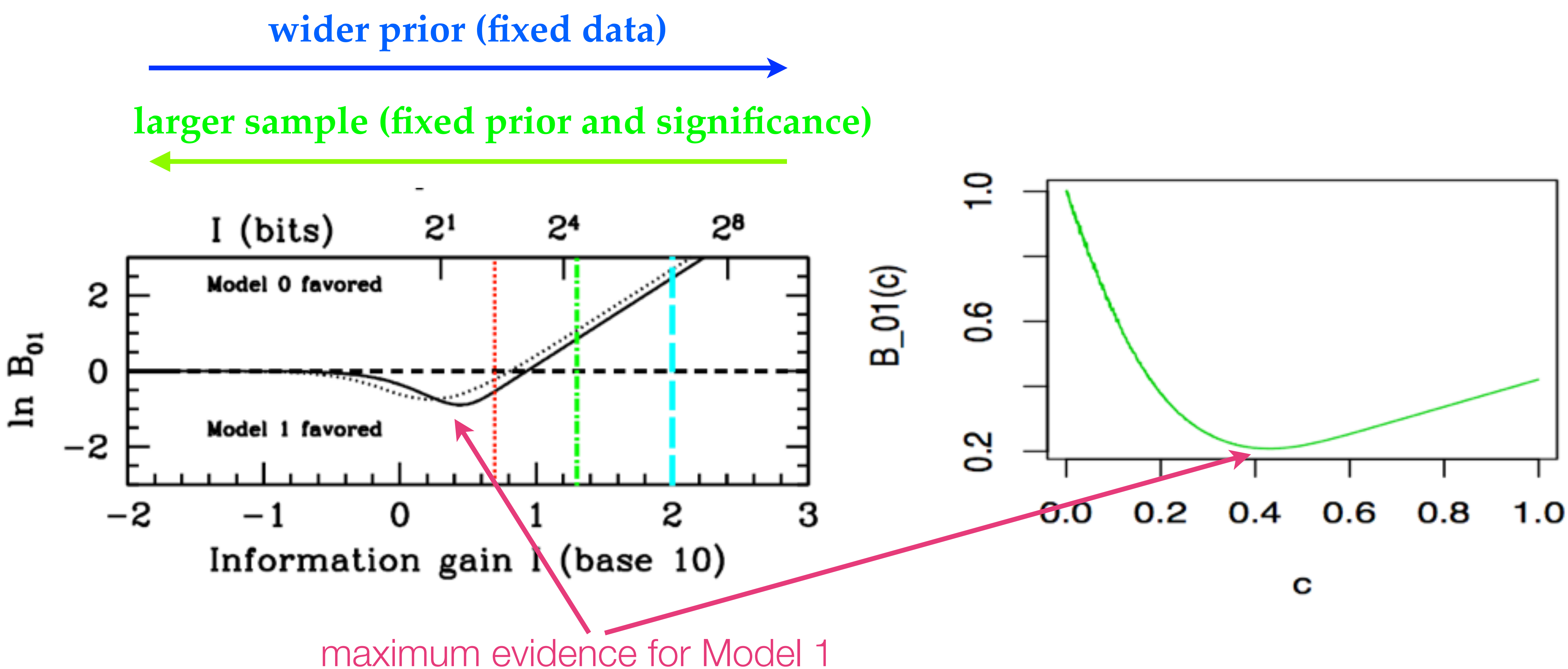
Pitfalls of Bayesian model selection

Imperial College London

- In cosmology/High Energy Physics, there are many situations with nested models with extra unknown parameters for the fundamental theory.
- Little or nothing is known about the metric to be imposed on such a parameter space
- “The concept of total ignorance about θ does not have any precise meaning” (Bob Cousins)
- Often, deviations are looked for using arbitrarily parameterized alternative models (not tied to any specific physics), e.g. Gaussian Processes.
- Occam’s razor factor may be arbitrary. HOWEVER: if the range of your prior is arbitrary (by many orders of magnitude) then arguably the physics behind it is not strongly predictive...
- In some cases, the upper bound formalism might be useful (Jim Berger and collaborators)

“Prior-free” evidence bounds

- What if we do not know how to set the prior? For nested models, we can still choose a prior that will maximise the support for the more complex model:



Maximum evidence for a detection

- **The absolute upper bound:** put all prior mass for the alternative onto the observed maximum likelihood value. Then

$$B < \exp(-\chi^2/2)$$

- **More reasonable class of priors:** symmetric and unimodal around Ψ=0, then (α = significance level)

$$B < \frac{-1}{\exp(1)\alpha \ln \alpha}$$

If the upper bound is small, no other choice of prior will make the extra parameter significant.

Sellke, Bayarri & Berger, *The American Statistician*, 55, 1 (2001)

How to interpret the "number of sigma's"

p	sigma	Absolute bound on lnB (B)	"Reasonable" bound on lnB (B)
0.05	2	2.0 (7:1) weak	0.9 (3:1) undecided
0.003	3	4.5 (90:1) moderate	3.0 (21:1) moderate
0.0003	3.6	6.48 (650:1) strong	5.0 (150:1) strong

A conversion table

Imperial College London

p–value	$\bar{B}$	$\ln \bar{B}$	sigma	category
0.05	2.5	0.9	2.0	
0.04	2.9	1.0	2.1	'weak' at best
0.01	8.0	2.1	2.6	
0.006	12	2.5	2.7	'moderate' at best
0.003	21	3.0	3.0	
0.001	53	4.0	3.3	
0.0003	150	5.0	3.6	'strong' at best
6×10^{-7}	43000	11	5.0	

Rule of thumb:

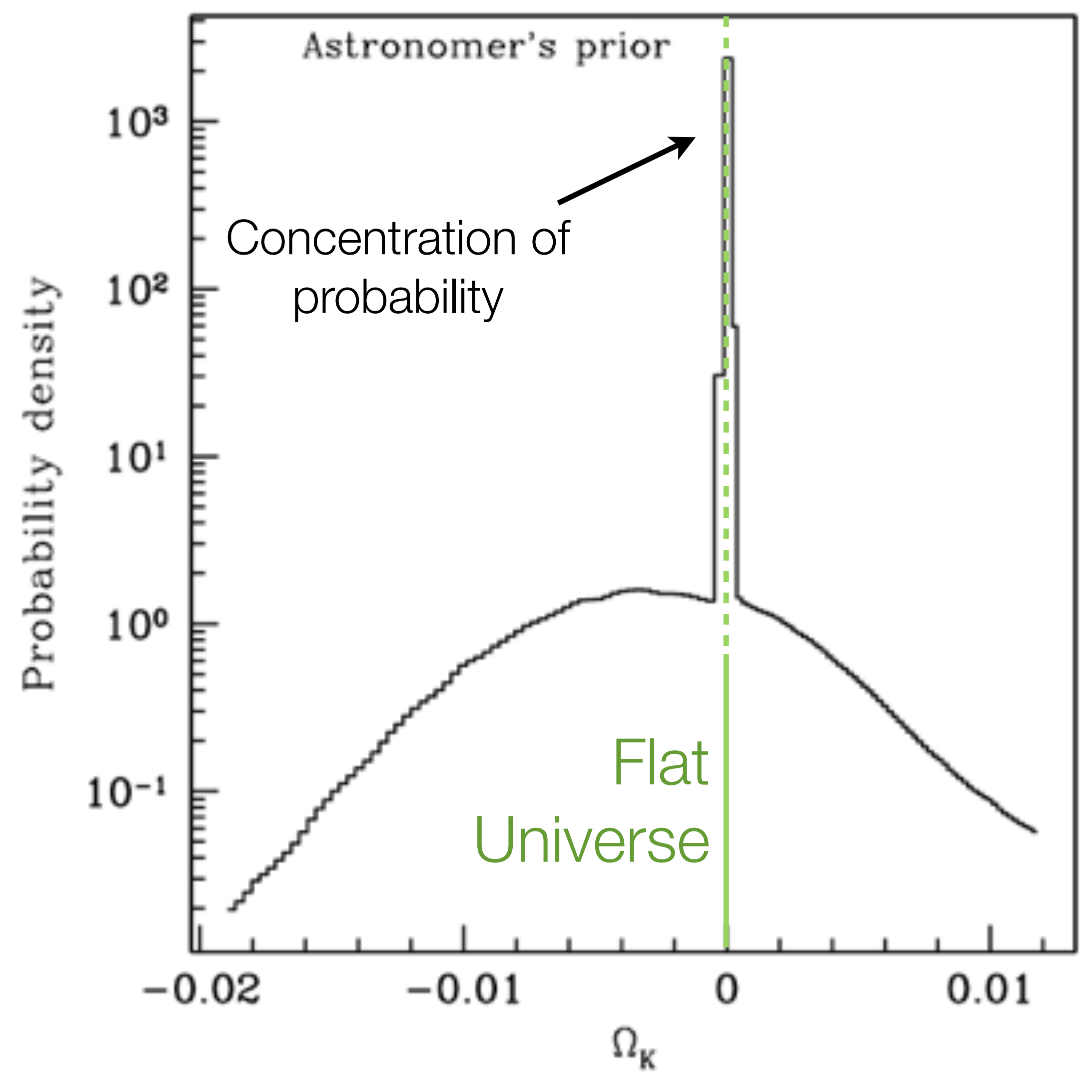
a n-sigma result should be interpreted as a n-1 sigma result

ICIC

Level 3 inference: model averaging

$$P(\theta|d) = \sum_i P(M_i|d)P(\theta|d, M_i)$$

- **Aim:** model-independent constraints that account for model uncertainty
- **Model posterior:** flat models are preferred by Bayesian model selection → probability gets concentrated onto those models
- **Consequence:** constraints on the curvature, number of Hubble spheres and size of the Universe can be **stronger** after Baysian model averaging!
- **Number of Hubble spheres** $N_U > 251$ (99%) ~8 times stronger
  Radius of curvature > 42 Gpc (99%) 1.5 times stronger



Vardanyan, RT & Silk '11

Some references

Imperial College London

- R. Trotta, "Bayes in the sky: Bayesian inference and model selection in cosmology" Contemporary Physics, 49, 2 (2008), 71-104 (arXiv: 0803.4089)
- Bayesian methods in cosmology, Hobson et al (eds), CUP (2010)
- Kelly, Some Aspects of Measurement Error in Linear Regression of Astronomical Data, Astrophys.J. 665 (2007) 1489-1506, arXiv:0705.2774
- Tom Loredo's Bayesian papers: http://www.astro.cornell.edu/staff/loredo/bayes/tjl.html
- G. D'Agostini, Probability and Measurement Uncertainty in Physics - a Bayesian Primer (1995), hep-ph/9512295
- E.T. Jaynes, Probability Theory: The Logic of Science, CUP (2003)
- D. MacKay, Information theory, Inference & Learning Algorithms, CUP (2003) (available for free on the web for onscreen viewing)
- P. Gregory, Bayesian logical data analysis for the physical sciences, CUP (2003)
- Hu & Dodelson, Cosmic Microwave Background Anisotropies, Ann.Rev.Astron.Astrophys.40:171-216,2002
- Schneider, Extragalactic Astronomy and Cosmology: An Introduction, Springer (2006).

Thank you!

www.robertotrotta.com