Astrophysics and Space Science Proceedings 45

Júlio C. Fabris
Oliver F. Piattella
Davi C. Rodrigues
Hermano E.S. Velten
Winfried Zimdahl *Editors*

The Cosmic Microwave Background

Proceedings of the II José Plínio Baptista School of Cosmology

Springer

Astrophysics and Space Science Proceedings

Volume 45

More information about this series at http://www.springer.com/series/7395

Júlio C. Fabris • Oliver F. Piattella •
Davi C. Rodrigues • Hermano E.S. Velten •
Winfried Zimdahl
Editors

The Cosmic Microwave Background

Proceedings of the II José Plínio Baptista School of Cosmology

Springer

Editors
Júlio C. Fabris
Departamento de Física, CCE
Universidade Federal do Espírito Santo
Vitória/ES, Brazil

Oliver F. Piattella
Departamento de Física, CCE
Universidade Federal do Espírito Santo
Vitória/ES, Brazil

Davi C. Rodrigues
Departamento de Física, CCE
Universidade Federal do Espírito Santo
Vitória/ES, Brazil

Hermano E.S. Velten
Departamento de Física, CCE
Universidade Federal do Espírito Santo
Vitória/ES, Brazil

Winfried Zimdahl
Departamento de Física
Universidade Federal do Espírito Santo
Vitória/ES, Brazil

ISSN 1570-6591 ISSN 1570-6605 (electronic)
Astrophysics and Space Science Proceedings
ISBN 978-3-319-44768-1 ISBN 978-3-319-44769-8 (eBook)
DOI 10.1007/978-3-319-44769-8

Library of Congress Control Number: 2016951303

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

The cosmic microwave background (CMB) radiation is one of the most important phenomena in physics and a fundamental probe of our universe when it was only 400,000 years old. It is an extraordinary laboratory where we can learn from particle physics to cosmology; its discovery in 1965 has been a landmark event in the history of physics.

The observations of the anisotropy of the cosmic microwave background radiation through the satellites COBE, WMAP, and Planck provided a huge amount of data which are being analyzed in order to discover important informations regarding the composition of our universe and the process of structure formation.

The series of texts composing this book is based on the lectures presented during the II José Plínio Baptista School of Cosmology, held in Pedra Azul (Espírito Santo, Brazil) between 9 and 14 March 2014. This II JBPCosmo has been entirely devoted to the problem of understanding theoretical and observational aspects of CMB.

We thank the speakers and the participants for their enthusiasm and for having provided a very nice environment to discuss this important topic of modern cosmology. The II JBPCosmo has been supported by CNPq, CAPES, FAPES, and UFES.

Vitória, Brazil

Júlio C. Fabris
Oliver F. Piattella
Davi C. Rodrigues
Hermano E.S. Velten
Winfried Zimdahl

Contents

Part I
Mini Courses

Physics of the Cosmic Microwave Background Radiation

David Wands, Oliver F. Piattella, and Luciano Casarini

Abstract The cosmic microwave background (CMB) radiation provides a remarkable window onto the early universe, revealing its composition and structure. In these lectures we review and discuss the physics underlying the main features of the CMB.

1 Introduction

The cosmic microwave background (CMB) radiation provides a remarkable window onto the early universe, revealing its composition and structure. It is a relic, thermal radiation from a hot dense phase in the early evolution of our Universe which has now been cooled by the cosmic expansion to just 3° above absolute zero. Its existence had been predicted in the 1940s by Alpher and Gamow (Alpher et al. 1948; Alpher 2014) and its discovery by Penzias and Wilson at Bell Labs in New Jersey, announced in 1965 (Penzias and Wilson 1965) was convincing evidence for most astronomers that the cosmos we see today emerged from a Hot Big Bang more than 10 billion years ago.

Since its discovery, many experiments have been performed to observe the CMB radiation at different frequencies, directions and polarisations, mostly with ground- and balloon-based detectors. These have established the remarkable uniformity of the CMB radiation, at a temperature of 2.7 K in all directions, with a small ± 3.3 mK dipole due to the Doppler shift from our local motion (at 1 million km/h) with respect to this cosmic background.

However, the study of the CMB has been transformed over the last 20 years by three pivotal satellite experiments. The first of these was the Cosmic Background

D. Wands (✉)
Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Burnaby Road, Portsmouth PO1 3FX, UK
e-mail: david.wands@port.ac.uk

O.F. Piattella • L. Casarini
Departamento de Física, Universidade Federal do Espírito Santo, Av. Fernando Ferrari, 514, Campus de Goiabeiras, 29075-910 Vitória, Espírito Santo, Brazil
e-mail: oliver.piattella@pq.cnpq.br; casarini.astro@gmail.com

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_1

Explorer (CoBE), launched by NASA in 1990 (Smoot et al. 1992; Mather et al. 1994). It confirmed the black body spectrum with an astonishing precision, with deviations less than 50 parts per million (Fixsen et al. 1996). And in 1992 CoBE reported the detection of statistically significant temperature anisotropies in the CMB, at the level of $\pm 30\,\mu$K on 10° scales (Smoot et al. 1992). COBE was succeeded by the Wilkinson Microwave Anisotropy Probe (WMAP) satellite, launched by NASA in 2001, which produced full sky maps in five frequencies (from 23 to 94 GHz) mapping the temperature anisotropies to sub-degree scales and determining the CMB polarisation on large angular scales for the first time. The Planck satellite, launched by ESA in 2009, sets the current state of the art with nine separate frequency channels, measuring temperature fluctuations to a millionth of a degree at an angular resolution down to 5 arc-min. Planck intermediate data was released in 2013 (Ade et al. 2014a).[1]

These lectures draw upon the excellent reviews of CMB physics by Hu and Dodelson (Hu and Dodelson 2002; Hu 2008, 2016), Komatsu (2016) and Crittenden (2016). We also refer the reader to comprehensive reviews on cosmological perturbations by Mukhanov et al. (1992) and Malik and Wands (2009). Useful textbooks are those of Peebles (1994), Dodelson (2003), Mukhanov (2005) and Weinberg (2008). Throughout this chapter we will use natural units such that $\hbar = k_B = c = 1$.

2 Background Cosmology and the Hot Big Bang Model

We start by recalling the mathematical framework describing the expansion of the universe and the Hot Big Bang. Much of modern cosmology is based on general relativity and the framework of Friedmann, Lemaître, Robertson and Walker in the 1920s and 30s (Friedmann 1924; Lemaitre 1927; Robertson 1935), and Hubble's discovery of the expansion of the universe (Hubble 1929). We can "slice" four-dimensional spacetime into expanding three-dimensional space at each cosmic time, t, with a uniform matter density and spatial curvature. Requiring spatial homogeneity and isotropy at each cosmic time is known as the cosmological principle, which picks out the following space-time metric:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-\kappa r^2} + r^2 d\Omega^2\right] , \tag{1}$$

where $a(t)$ is the scale factor and κ is the curvature of the maximally symmetric spatial slices, and we chose spherical coordinates with infinitesimal solid angle $d\Omega^2$. We will sometimes find it convenient to use conformal time, η, where $dt = a d\eta$ and

[1] After these lectures were given, full-mission data was released in 2015 (Adam et al 2015), with final polarisation data still to come.

the line element takes the form

$$ds^2 = a^2(\eta)\left[-d\eta^2 + \frac{dr^2}{1-\kappa r^2} + r^2 d\Omega^2\right] , \tag{2}$$

The Hubble expansion rate is defined as $H \equiv \dot{a}/a$, where a dot denotes a derivative with respect to cosmic time t. The present value of H is called the Hubble constant and denoted as H_0. The value of H_0 is often given in the form

$$H_0 = 100\,\mathrm{h\,km\,s^{-1}\,Mpc^{-1}} . \tag{3}$$

Using the energy constraint, from Einstein's equations of general relativity, one gets the Friedmann equation for the Hubble expansion

$$H^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{\kappa}{a^2} , \tag{4}$$

where we introduce the cosmological constant, Λ, and ρ, the energy density. The latter includes electrons, baryons (protons, neutrons and atomic nuclei), radiation (photons and neutrinos) and dark matter (non-baryonic massive particles, non-relativistic by the present day).

Dividing through by H^2, Eq. (4) can be cast in the following dimensionless form:

$$1 = \Omega + \Omega_\Lambda + \Omega_\kappa , \tag{5}$$

where we define the relative contributions to the Hubble expansion

$$\Omega \equiv \frac{8\pi G\rho}{3H^3} , \quad \Omega_\Lambda \equiv \frac{\Lambda}{3H^2} , \quad \Omega_\kappa \equiv \frac{-\kappa}{a^2H^2} . \tag{6}$$

In order to get a closed system of equations we must determine the evolution of the density, ρ in Eq. (4), as a function of the scale factor. For this we can use the continuity (energy conservation) equation

$$\dot{\rho} = -3H(\rho + P) , \tag{7}$$

plus an equation of state for the pressure, $P(\rho)$. We will be interested in three important cases:

- $\Omega = \Omega_r = 1$, radiation domination:

$$P_r = \frac{1}{3}\rho_r \quad \Rightarrow \quad \rho_r \propto a^{-4} \quad \Rightarrow \quad a \propto t^{1/2} \propto \eta . \tag{8}$$

- $\Omega = \Omega_m = 1$, matter domination (Einstein-de Sitter):

$$P_m = 0 \quad \Rightarrow \quad \rho_m \propto a^{-3} \quad \Rightarrow \quad a \propto t^{2/3} \propto \eta^2 . \tag{9}$$

- $\Omega_\Lambda = 1$, Λ domination (de Sitter):

$$a \propto e^{Ht} \propto (\eta_\infty - \eta)^{-1} \,. \tag{10}$$

The CMB consists of photons which survive from an early, radiation-dominated, Hot Big Bang and have a small density with respect to non-relativistic matter today. Nonetheless the CMB holds a rich store of information about the history of our Universe, as we shall see. For example, recent observations of the CMB by Planck (Ade et al. 2014b) can be used to infer values for the above cosmological parameters at the present-day:

$$h = 0.674 \pm 0.014 \,, \quad \Omega_0 = 0.314 \pm 0.020 \,, \quad \Omega_{\Lambda 0} = 0.686 \pm 0.020 \,, \quad \Omega_{\kappa 0} = -0.04 \pm 0.05 \,. \tag{11}$$

The data are consistent with a flat universe, $\kappa = 0$, which will be our working hypothesis hereafter. We see that the expansion today is dominated by a cosmological constant (or some form of matter which acts very much like a cosmological constant) but in the recent past it was dominated by non-relativistic matter, and before that by radiation.

2.1 Black-Body Spectrum

The CMB is observed to have a black-body spectrum characteristic of a thermal equilibrium distribution, consistent with the hypothesis that our Universe emerged from a hot, dense Big Bang.

Photons follow a null trajectory in the FLRW metric (2) such that

$$\frac{dx^i}{d\eta} = \hat{n}^i \,, \tag{12}$$

where $\hat{n}^i$ is a unit 3-vector, $g_{ij}\hat{n}^i\hat{n}^j = 1$. The 3-momentum of a photon is $p^i = p\hat{n}^i$, where p is the wavenumber (remembering that we are using units such that $\hbar = 1$ and $c = 1$, so that p also describes the energy of a massless photon).

CMB photons have an isotropic Bose-Einstein distribution function with temperature T

$$f(p) = \frac{1}{\exp(p/T) - 1} \,. \tag{13}$$

Given this isotropic distribution, we can compute the number density of CMB photons

$$n_\gamma = 2 \int \frac{4\pi p^2 dp}{(2\pi)^3} f(p) \approx \frac{2.4}{\pi^2} T^3 \,, \tag{14}$$

where the photons have two independent polarisation states and $4\pi p^2 dp$ is the volume of an infinitesimal shell in three-dimensional momentum-space. Their energy density is

$$\rho_\gamma = 2 \int \frac{4\pi p^2 dp}{(2\pi)^3} p f(p) = \frac{\pi^2}{15} T^4 \,. \tag{15}$$

However, the CMB photons are no longer in equilibrium with the matter we see in the universe today. The photons are free to propagate through the universe after electrons and baryons have recombined into neutral atoms, so the black-body spectrum must be propagated to the present day from the early universe. Freely propagating photons follow the geodesic equation in curved space-time

$$\frac{dP^\mu}{d\lambda} + \Gamma^\mu_{\nu\sigma} P^\nu P^\sigma = 0 \,, \tag{16}$$

where $\Gamma^\mu_{\nu\sigma}$ is the Christoffel symbol. We define the photon 4-momentum as $P^\mu = dx^\mu/d\lambda$, where λ is an affine parameter, and the modulus-squared of the 3-momentum is $p^2 = g_{ij} P^i P^j$ where g_{ij} is the spatial part of FLRW metric (1). From the geodesic equation in the conformal FLRW metric (2) we obtain

$$\frac{1}{p}\frac{dp}{d\eta} = -\frac{1}{a}\frac{da}{d\eta} \,. \tag{17}$$

Integrating this up to the present we obtain the cosmological redshift of the photon momentum, defined as

$$1 + z \equiv \frac{p}{p_0} = \frac{a_0}{a} \,. \tag{18}$$

We can interpret this simply as the expansion of the universe stretching the wavelength of a photon, reducing (redshifting) its energy and momentum.

Note that the form of the Bose-Einstein distribution (13) is preserved

$$f(p) = \frac{1}{\exp(p/T) - 1} = \frac{1}{\exp(p_0/T_0) - 1} \,, \tag{19}$$

where the temperature is also redshifted with the expansion

$$1 + z = \frac{T}{T_0} \,. \tag{20}$$

Thus we see that the energy density (15) of the photons decreases as the universe expands

$$\rho_\gamma \propto a^{-4} \,. \tag{21}$$

Although photon density is small is in the universe today, it dominated the hot, dense, early universe.

2.2 Hot Big Bang

At sufficiently high temperatures we expect all particles to be relativistic. If these particles interact and efficiently redistribute energy they will share the same thermal equilibrium temperature. To be relativistic we require $T \gg m$, i.e., the thermal energy is much larger than the rest mass of a given particle species. At this stage of the primordial universe we can write the energy density using the same form given in Eq. (15) for all the relativistic species:

$$\rho = g_{\rm eff}\frac{\pi^2}{30}T^4 \,, \tag{22}$$

where $g_{\rm eff}$ is the sum of the effective number of degrees of freedom. Each bosonic species in thermal equilibrium contributes one per spin state (e.g., photons contribute $+2$, corresponding to two polarisations), whereas each fermion contributes $7/8$ per spin state, due to the different statistics.[2]

In a radiation-dominated universe (8) the time dependence of the scale factor is given by $a \propto t^{1/2}$ and thus from Eq. (4) we have

$$\rho = \frac{3H^2}{8\pi G} = \frac{3}{32\pi G t^2} \,, \tag{23}$$

so that from (22) time and temperature are related by

$$t = \sqrt{\frac{3}{32\pi G}\frac{30}{g_{\rm eff}\pi^2}}\frac{1}{T^2} \,. \tag{24}$$

Thus we have the simple, approximate temperature-time relation

$$\frac{t}{1\ {\rm s}} \approx \frac{1}{\sqrt{g_{\rm eff}}}\left(\frac{1\ {\rm MeV}}{T}\right)^2 \,. \tag{25}$$

[2]If a species decouples from this thermal bath, but remains relativistic, it can contribute with a different temperature in the above equation. This is what happens for neutrinos. They decouple relativistically from the primordial soup, at $T \approx 1$ MeV and their temperature today is expected to be $(4/11)^{1/3}$ times that of the photons because photons are heated by e^--e^+ annihilation.

2.3 Spectral Distortions

The black-body shape of the CMB spectrum is maintained at early times because of the high interaction rate of photons with the other particles of the primordial plasma. We can identify two principal scattering processes which contribute to maintaining an isotropic, equilibrium distribution:

- *Compton scattering*: scattering of photons and relativistic electrons, redistributing energy and momentum, conserving photon number

$$e^- + \gamma \leftrightarrow e^- + \gamma \,.$$

  At low energies this reduces to *Thomson scattering*, i.e., elastic scattering of photons off non-relativistic electrons, exchanging momentum, but conserving photon energy and number.
- *Double (radiative) Compton scattering*: scattering of photons and relativistic electrons, redistributing energy and momentum, and changing photon number

$$e^- + \gamma \leftrightarrow e^- + \gamma + \gamma \,.$$

Many processes in the early universe before the time of recombination could potentially lead to measurable distortions in the CMB spectrum, which might be measured with future missions. Particle annihilation or decay would heat the primordial plasma, and hence the photons, or even the evaporation of primordial black holes in the relevant mass range. Even the damping of small scale density variations in the primordial plasma due to photon diffusion can lead to deviations from an exact black-body spectrum. For more detail about CMB spectral distortions and what might cause them, see Chluba and Sunyaev (2012).

Efficient Compton and double Compton scattering maintains a full thermal equilibrium spectrum above a redshift (Hu 2008)

$$z_{th} = 2 \times 10^6 \left(\frac{\Omega_b h^2}{0.02} \right)^{-2/5} , \tag{26}$$

where $\Omega_b h^2$ determines the density of baryons and hence (in an electrically neutral universe) electrons.

Below this redshift Compton scattering can still redistribute energy and momentum between photons and electrons, but double Compton scattering becomes inefficient. In the absence of double Compton scattering, interactions cannot create or remove photons from the plasma. Compton scattering still maintains a statistical equilibrium above redshift (Hu 2008)

$$z_\mu = 5 \times 10^4 \left(\frac{\Omega_{b0}}{0.02} \right)^{-1/2} . \tag{27}$$

Thus if additional energy is dumped into the primordial plasma below redshift z_{th} the CMB photons acquire a statistical equilibrium distribution

$$f(p) = \frac{1}{\exp[(p-\mu)/T] - 1} \tag{28}$$

with non-zero chemical potential μ. This is known as a μ-distortion in the CMB spectrum. Limits from the COBE satellite give an upper limit on the size of such a distortion (Fixsen et al. 1996):

$$\frac{|\mu|}{T} < 9 \times 10^{-5} \text{ at } 95\,\% \text{ CL}\,. \tag{29}$$

Below the redshift z_μ Compton scattering off relativistic electrons becomes inefficient. High-energy electrons along the line of sight can still transfer energy to low-frequency photons via inverse Compton scattering, without reaching statistical equilibrium. This leads to a characteristic "y-distortion" where low energy photons are boosted to higher frequencies, leading to a deficit in the CMB intensity at low frequencies in the Rayleigh-Jeans region, equivalent to a temperature deficit

$$\left.\frac{\Delta T}{T}\right|_{p \ll T} = -2y \tag{30}$$

and an enhancement at high frequencies. The Compton y-parameter is defined as the line-of-sight integral of the electron pressure

$$y = \int \frac{T_e}{m_e} n_e \sigma_T dl\,, \tag{31}$$

where n_e is the density of free electrons and σ_T is the Thomson scattering cross-section, see Eq. (33) below. Constraints from COBE/FIRAS give the upper limit (Fixsen et al. 1996)

$$|y| < 1.5 \times 10^{-5} \text{ at } 95\,\% \text{ CL}\,. \tag{32}$$

These constraints still rely on COBE observations, more than 20 years ago.

An important source of y-distortions seen in specific directions in the CMB is the Sunyaev-Zeldovich effect (Sunyaev and Zeldovich 1970), from hot cluster gas along the line of sight after recombination. The Planck satellite has now compiled a catalogue of 439 clusters detected in the Planck data via their SZ signal (Ade et al. 2015) with many more being detected by ground-based experiments such as the Atacama Cosmology Telescope (Hasselfield et al. 2013) and the South Pole Telescope (Bleem et al. 2015).

2.4 Tight-Coupling and Sudden Recombination

At low energies (much smaller than the electron rest mass) electrons and photons interact via Thomson scattering, whose cross-section is[3]

$$\sigma_{\rm T} = \frac{8\pi\alpha^2}{3m_{\rm e}^2} = 6.65 \times 10^{-29}\ {\rm m}^2\ . \tag{33}$$

The corresponding mean-free-path for photons associated with Thomson scattering is given by

$$\lambda_{\rm mfp} = \frac{1}{n_{\rm e}\sigma_{\rm T}}\ . \tag{34}$$

Around $z \approx 1100$ the mean-free-path is approximately 2.5 kpc, corresponding to a comoving scale of order 2.5 Mpc at present (Hu 2008). On scales much larger than the mean-free-path, $\lambda \gg \lambda_{\rm mfp}$, the photons are tightly coupled to the electrons, while electrons are tightly coupled to protons through the Coulomb interaction. In this regime, photons, electrons and protons can be treated as a single fluid with common 3-velocity, and isotropic pressure.

The mean-free-path is time-dependent because the free-electron density, $n_{\rm e}$, is time-dependent. As the Universe cools down the capture of electrons by protons becomes efficient. As the wavelengths of photons are redshifted by the cosmic expansion, fewer photons have sufficient energy (the ionisation energy, 13.6 eV) required to break the binding energy of an electron in a neutral hydrogen atom. Therefore, the density of free electrons, n_e, rapidly drops around $z \approx 1100$, leading to a rapid increase in the Thomson mean-free-path beyond the Hubble radius.

This process is called decoupling, because photons no longer interact with electrons. It is also called recombination because this is the epoch when protons and electrons recombine to form hydrogen atoms. Recombination and decoupling are practically simultaneous because the rapid drop in the density of free electrons due to recombination affects the Thomson scattering rate. By solving the corresponding Boltzmann equation we see that recombination and decoupling occur at redshift (Hu 2008)

$$1 + z_* = 1089 \left(\frac{\Omega_m h^2}{0.14}\right)^{0.0105} \left(\frac{\Omega_b h^2}{0.024}\right)^{-0.028}\ . \tag{35}$$

[3]The full cross-section describing the process $e^- + \gamma \rightarrow e^- + \gamma$ is given by the Klein-Nishina formula (Klein and Nishina 1929), which displays not only the dependence on the photon energy but also on its polarization and the scattering angle. Since the energies involved in the recombination process are much smaller than the electron mass, we can safely use Thomson cross-section.

Note that this is some time after (but not long after) matter-radiation equality,

$$1 + z_{\rm eq} = 3.4 \times 10^3 \left(\frac{\Omega_m h^2}{0.14} \right) . \tag{36}$$

Another way to define when recombination/decoupling takes place is via the Thomson optical depth

$$\tau = \int_{\eta}^{\eta_0} n_e \sigma_T dt \ , \tag{37}$$

which represents the integrated scattering rate from a conformal time η until today η_0, i.e., the average number of scattering events between these two times. The spatial hyper-surface of constant $\eta = \eta_*$, where η_* is the conformal time corresponding to $\tau = 1$, is called the last-scattering surface. Of course, recombination is not an instantaneous phenomenon, but it occurs sufficiently rapidly that a useful approximation on comoving scales greater than about 2.5 Mpc is the so-called *sudden recombination*, as if it really happened at a single instant, η_*.

3 CMB Anisotropies

Anisotropies observed in the CMB radiation are caused by inhomogeneities in the cosmological spacetime and matter distribution. Fortunately, these inhomogeneities are small (about one part in 10^4) with respect to the background homogenous energy density, thereby allowing us to use perturbation theory to model their behaviour. In the following we shall consider a linearly perturbed distribution.

We do not measure the plasma density directly, but rather anisotropies, in the CMB photon distribution function, $f \to \bar{f} + \delta f$. At first order these can be described by a perturbation in the temperature of the Bose-Einstein distribution (13), where

$$T(\eta, \mathbf{x}, \hat{\mathbf{p}}) = \bar{T}(\eta) \left[1 + \Theta(\eta, \mathbf{x}, \hat{\mathbf{p}}) \right] \ , \tag{38}$$

where $\hat{\mathbf{p}}$ denotes the direction of the photon propagation. The temperature fluctuation in the plasma is related to the photon density contrast via Eq. (15) as

$$\Theta \equiv \frac{\delta T}{T} = \frac{1}{4} \frac{\delta \rho_\gamma}{\rho_\gamma} \equiv \frac{1}{4} \delta_\gamma \ . \tag{39}$$

3.1 Spherical Harmonics

Since we observe CMB on the celestial sphere, it is useful to expand Θ in spherical harmonics

$$\Theta(\eta, \mathbf{x}, \hat{\mathbf{p}}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m}(\eta, \mathbf{x}) Y_{\ell m}(\hat{\mathbf{p}}) \ . \tag{40}$$

Since the spherical harmonics form a complete orthonormal basis on the sphere,

$$\int d\Omega_n Y_{\ell m}(\hat{\mathbf{n}}) Y^*_{\ell' m'}(\hat{\mathbf{n}}) = \delta_{\ell\ell'}\delta_{mm'} \ . \tag{41}$$

The coefficients a_{lm} describe the temperature fluctuations at a given angular multipole ℓ. An isotropic distribution has an angular power spectrum C_l:

$$\langle a^*_{\ell m} a_{\ell' m'} \rangle = \delta_{\ell\ell'}\delta_{mm'} C_\ell \ . \tag{42}$$

In this case the correlation between the temperatures in two directions on the CMB sky depends only on the angular distance between the two directions and not on the orientation of the arc which joins them.

For a fixed ℓ, one has $2\ell + 1$ different $a_{\ell m}$'s, i.e., $2\ell + 1$ independent estimates of the true C_ℓ. The "observed" $C_\ell^{\rm obs}$ corresponds to our best estimate of the true angular power spectrum:

$$C_\ell^{\rm obs} \equiv \frac{1}{2\ell + 1} \sum_m (a_{\ell m}^{\rm obs})^* a_{\ell m}^{\rm obs} \ , \tag{43}$$

i.e., it is an average over the observed multipole moments, m, at fixed ℓ. We define the cosmic variance as the expected error in our determination of the true power spectrum

$$\left(\frac{\Delta C_\ell}{C_\ell} \right)^2_{\text{cosmic variance}} \equiv \left\langle \left(\frac{C_\ell - C_\ell^{\rm obs}}{C_\ell} \right)^2 \right\rangle \ . \tag{44}$$

Calculating the expectation in the above equation, with the help of Eq. (42), one obtains

$$\left(\frac{\Delta C_\ell}{C_\ell} \right)_{\text{cosmic variance}} = \sqrt{\frac{2}{2\ell + 1}} \ . \tag{45}$$

Thus at small multipoles, ℓ, corresponding to very large angular scales, the cosmic variance is significant and represents the minimal uncertainty in estimating the true angular power spectrum given that we have only one realisation of the CMB sky.

3.2 Last-Scattering Sphere

Since most photons are last scattered at η_*, we will be mostly interested in their distribution, $\Theta(\eta, \mathbf{x}, \hat{\mathbf{p}})$ in Eq. (38), at evaluated at recombination, i.e., at initial time $\eta = \eta_*$ and comoving displacement with respect to an observer at the origin, $\mathbf{x}_* = -D_*\hat{\mathbf{p}}$, where the comoving distance to last-scattering $D_* = \eta_0 - \eta_* \simeq \eta_0$. Then we propagate this photon distribution until today using the free-streaming equations, i.e., the collision-less Boltzmann equation for photons.

Adopting the sudden-recombination approximation, we assume that the photons are tightly coupled with an isotropic distribution up until last scattering,

$$\Theta_*(\hat{\mathbf{p}}) = \Theta\left(\eta_*, \mathbf{x}_*\right) . \tag{46}$$

The CMB temperature varies across our sky due to the variation in the photon temperature across the last-scattering surface.

We can decompose this 3D CMB temperature field into Fourier modes

$$\Theta(\eta, \mathbf{x}) = \frac{1}{(2\pi)^3} \int d^3\mathbf{k}\, \Theta(\eta, \mathbf{k})\, e^{i\mathbf{k}\cdot\mathbf{x}} . \tag{47}$$

Linear modes with different comoving wavevectors, $\mathbf{k}$, then evolve independently at first order. We assume that these perturbations are stochastic quantities drawn from some distribution, which usually is assumed to be Gaussian.

The expectation value of each mode is zero and its variance is the power spectrum

$$\langle \Theta^*(\eta, \mathbf{k}_1)\Theta(\eta, \mathbf{k}_2)\rangle = (2\pi)^3\delta^3(\mathbf{k}_1 + \mathbf{k}_2)P_\Theta(k_1, \eta) . \tag{48}$$

Note that P_Θ is function of the modulus of $\mathbf{k}_1$ only, i.e., we assume statistical isotropy. The correlation function in real space is given by the Fourier transform of the power spectrum

$$\xi_\Theta(\mathbf{r}) \equiv \langle \Theta(\eta, \mathbf{x})\Theta(\eta, \mathbf{x} + \mathbf{r})\rangle = \frac{1}{(2\pi)^3} \int d^3\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{r}} P_\Theta(k) . \tag{49}$$

Angular brackets denote the ensemble average. That is, one imagines different possible realizations of our universe. In theories such as inflation, where primordial fluctuations are quantum in their origin and then become effectively classical through an exponential phase of expansion, it is possible to predict the primordial form of the power spectrum. After that, it is evolved up until today using the classical

equations of cosmological perturbation theory. Thanks to the ergodic theorem, we can swap the ensemble average into a position average, see Appendix D of Weinberg (2008).

Since P_Θ depends only on the modulus k, we can perform the angular integration in (49) and find

$$\xi_\Theta(r) = \frac{1}{2\pi^2}\int_0^\infty \frac{dk}{k} k^3 P_\Theta(k) \frac{\sin kr}{kr} \,. \tag{50}$$

From the above result, we can identify the dimensionless power spectrum

$$\mathcal{P}_\Theta(k) \equiv \frac{k^3 P_\Theta(k)}{2\pi^2} \,. \tag{51}$$

We can decompose the temperature field on the last-scattering surface into spherical harmonics using the plane-wave expansion

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} i^\ell j_\ell(kr) Y^*_{\ell m}(\hat{\mathbf{k}}) Y_{\ell m}(\hat{\mathbf{r}}) \,. \tag{52}$$

where the spherical Bessel function $j_\ell(x)$ is defined in terms of the regular Bessel function $J_{\ell+1/2}(x)$ as $j_\ell(x) = (\pi/2x)^{1/2} J_{\ell+1/2}(x)$. Substituting this expansion into (47) and comparing with (40) evaluated at $\mathbf{x}_* = -D_* \hat{\mathbf{p}}$ we obtain the spherical harmonic coefficients

$$a_{\ell m} = \frac{i^\ell}{2\pi^2} \int d^3\mathbf{k}\, \Theta(\eta_*, \mathbf{k})\, j_l(kD_*)\, Y^*_{\ell m}(\hat{\mathbf{k}}) \,, \tag{53}$$

and hence the angular power spectrum (42), by using Eqs. (41) and (48), becomes:

$$C_\ell = 4\pi \int_0^\infty \frac{dk}{k} \mathcal{P}_\Theta(k) j_\ell^2(kD_*) \,. \tag{54}$$

The window function

$$W_\ell(k) \equiv 4\pi j_\ell(kD_*)^2 \,, \tag{55}$$

peaks about $k = \ell/D_*$, so one obtains approximately that

$$\frac{\ell(\ell+1)}{2\pi} C_\ell \approx \mathcal{P}_\Theta(\ell/\eta_0) \,, \tag{56}$$

by using $D_* \approx \eta_0$ and the result

$$\int_0^\infty \frac{dk}{k} j_\ell^2(k\eta_0) = \frac{1}{2l(l+1)} \,. \tag{57}$$

This is the origin of the ubiquitous prefactor $l(l+1)$ in CMB spectrum plots. In order to obtain the full result one should include contributions from the metric perturbations and the dipole at recombination and the ISW effect, which we present in the following section.

4 Sachs-Wolfe Formula

In the previous section we discussed the basic quantities which describe the CMB temperature anisotropies at last-scattering, and in particular the angular power spectrum, C_ℓ. In this section we link these to the observed temperature fluctuations including the effect of inhomogeneities in the metric and the density distribution of the matter content in the universe. We will derive the Sachs-Wolfe formula. In order to do this, we present the essential elements of relativistic cosmological perturbation theory, focusing on first-order fluctuations. The pioneering work in this field is due to Lifshitz (1946) but we also refer the reader to more recent reviews, such as Malik and Wands (2009).

4.1 Metric Perturbations

The starting point for discussing cosmological perturbations is the perturbed FRLW metric (Malik and Wands 2009)

$$ds^2 = a^2\left\{-(1+2A)d\eta^2 + 2\nabla_i B dx^i d\eta + \left[(1+2C)\delta_{ij} + 2\nabla_i\nabla_j E\right]dx^i dx^j\right\}\ , \tag{58}$$

where A, B, C, and E are scalar functions of the coordinates. In the above metric, we are considering only scalar perturbations, neglecting for now vector and tensor (gravitational wave) perturbations. Because of the tensorial nature of the metric, the above scalar functions change when changing the reference frame. It could happen that a reference frame exists in which $A = B = C = D = 0$. In this case then there are no metric perturbations, since we recover the original unperturbed FLRW metric. So, the fact of having four scalar functions of the coordinates in the above metric does not guarantee that we are actually dealing with cosmological perturbations, because the latter may be coordinate artifacts. This is the well-known *gauge problem*.

In order to know if we are really dealing with cosmological perturbations, a useful tool is to construct combinations of the above scalars which remain invariant under first order coordinate changes. There are three combinations independent of the spatial threading: A, C and $\sigma \equiv E' - B$, where the prime denotes differentiation with respect to the conformal time η. There are then two combinations independent of time slicing, for example, the Bardeen potentials (Bardeen 1980; Mukhanov et al.

1992; Malik and Wands 2009)

$$\Psi \equiv A - \mathcal{H}\sigma - \sigma' \,, \qquad \Phi \equiv C - \mathcal{H}\sigma \,. \tag{59}$$

In the above definition $\mathcal{H} \equiv a'/a$, is the conformal Hubble parameter, i.e. defined with respect to the conformal time.

A particularly useful gauge is the conformal Newtonian gauge, where the metric becomes diagonal since the choice is $B = E = 0$. The Bardeen potentials (59) can be identified with the metric perturbations A and C in this conformal Newtonian gauge (where $\sigma = 0$). The perturbed metric thus takes the form (Hu 2008)

$$ds^2 = a^2 \left\{-(1+2\Psi)d\eta^2 + (1+2\Phi)\delta_{ij}dx^i dx^j\right\} \,. \tag{60}$$

It can be shown by writing down explicitly the Einstein equations that their spatial traceless part depends on $\Phi + \Psi$. For example, the quadruple moment of the matter distribution acts as source of the spatial traceless part of the Einstein equation. In the tight coupling limit, there is no anisotropic stress because the high interaction rate of photons due to Thomson scattering establishes an isotropic distribution, which implies that $\Phi + \Psi = 0$.

One can construct other gauge-invariant variables, e.g., involving matter quantities, such as the density contrast and velocity potential in the conformal Newtonian gauge

$$\delta \equiv \frac{\delta\rho - \rho'\sigma}{\rho} \,, \qquad V \equiv v + E' \,, \tag{61}$$

or the curvature perturbation

$$\zeta \equiv C - \frac{\mathcal{H}}{\rho'}\delta\rho \,, \tag{62}$$

which can be identified with the metric perturbations C in the uniform-density gauge. This is a particularly useful variable on large scales since ζ is conserved for adiabatic perturbations on super-Hubble scales ($k \ll aH$) (Wands et al. 2000). For example, simple slow-roll inflation models typically produce an approximately scale-invariant dimensionless power spectrum, $\mathcal{P}_\zeta(k)$, on large scales at the start of the radiation dominated era. Thus we will typically set initial conditions in terms of ζ and/or isocurvature perturbations.

Note that these different perturbation variables are not necessarily independent. For example we can express ζ in terms of the conformal Newtonian gauge quantities:

$$\zeta = \Phi - \frac{\mathcal{H}\rho}{\rho'}\delta \,. \tag{63}$$

4.2 Perturbed Geodesics

What is the form of a perturbed geodesics in the conformal Newtonian gauge (60)? By setting $ds^2 = 0$ for a null trajectory, we find the coordinate velocity of a photon

$$\frac{dx^i}{d\eta} = (1 + \Psi - \Phi)\hat{p}^i \,, \tag{64}$$

where $\hat{p}^i$ is a unit vector, $\delta_{ij}\hat{p}^i\hat{p}^j = 1$. Defining the 4-momentum as $P^\mu = dx^\mu/d\lambda$ and the modulus of the 3-momentum $p^2 = g_{ij}P^iP^j$, the perturbed geodesic equation (16) can be written as follows:

$$\frac{1}{p}\frac{dp}{d\eta} = -\left(\frac{1}{a}\frac{da}{d\eta} + \frac{\partial\Phi}{\partial\eta}\right) - \hat{p}^i\frac{\partial\Psi}{\partial x^i} \,. \tag{65}$$

The term in parenthesis is the usual Hubble redshift corrected by the metric perturbation, which makes the expansion not homogeneous and isotropic, as it was in the background. The last term represents the gravitational blueshift or redshift experienced by a photon falling into or climbing out of a potential well. Introducing the total time derivative along the photon path, i.e.

$$\frac{d\Psi}{d\eta} = \frac{\partial\Psi}{\partial\eta} + \hat{p}^i\frac{\partial\Psi}{\partial x^i} \,, \tag{66}$$

the geodesic equation (65) becomes

$$\frac{1}{p}\frac{dp}{d\eta} = -\frac{1}{a}\frac{da}{d\eta} - \frac{d\Psi}{d\eta} + \frac{\partial}{\partial\eta}(\Psi - \Phi) \,. \tag{67}$$

This can be formally integrated along the photon trajectory from recombination, η_*, until today, η_0,

$$\ln\left(\frac{p_0}{p_*}\right) = -\ln\left(\frac{a_0}{a_*}\right) - \Psi_0 + \Psi_* + \int_{\eta_*}^{\eta_0}(\Psi' - \Phi')d\eta \,. \tag{68}$$

Splitting the momentum in a background part plus perturbation, i.e. $p \to p + \delta p$, one obtains

$$\left(\frac{\delta p}{p}\right)_0 = \left(\frac{\delta p}{p}\right)_* + \Psi_* - \Psi_0 + \int_{\eta_*}^{\eta_0}(\Psi' - \Phi')d\eta \,. \tag{69}$$

This relative perturbation in the photon momentum causes a relative temperature fluctuation in the CMB, $\Theta = \delta p/p$. So, one sees that at recombination photons get a redshift escaping from over-densities on the last-scattering surface with a negative gravitational potential Ψ_*. This is part of the Sachs-Wolfe effect (Sachs and Wolfe

1967). The integral term in (69), named the integrated Sachs-Wolfe effect, accounts for the time-dependence of the potentials along the line of sight from recombination until today.

The observed temperature fluctuation includes a Doppler shift due to the relative peculiar velocity (in addition to the expansion) between the last-scattering surface and the observer

$$\Theta = \frac{\delta p}{p} + \hat{\mathbf{n}} \cdot \mathbf{V} \, . \tag{70}$$

The full Sachs-Wolfe formula is thus

$$\Theta_{\rm obs} = \frac{1}{4}\delta_{\gamma *} + \Psi_* - \hat{\mathbf{n}} \cdot \mathbf{V}_* + \int_{\eta_*}^{\eta_0} (\Psi' - \Phi') d\eta - \Psi_0 + \hat{\mathbf{n}} \cdot \mathbf{V}_{\rm obs} \, , \tag{71}$$

where we have identified the relative momentum perturbation for photons on the last scattering surface with the radiation density contrast $\delta_\gamma = 4\delta p/p$. The first three terms represent the intrinsic Sachs-Wolfe effect (on the last-scattering surface) and the fourth the integrated one we already mentioned. Ψ_0 is the gravitational potential at the observer today and gives an undetectable correction to the monopole (that is, the solid-angle-averaged temperature). The last term is a dipole anisotropy induced by the observer's velocity.

4.3 Adiabatic and Isocurvature Perturbations

In order to evaluate the relative contribution of different terms in the Sachs-Wolfe formula (71) we need to determine the evolution of linear perturbations, in particular at the time of last scattering.

The behaviour of the scalar perturbations previously introduced is given by the Einstein evolution equations (coming from the spatial part of the Einstein equations written in an arbitrary gauge):

$$C'' + 2\mathcal{H}C' - \mathcal{H}A' - (2\mathcal{H}' + \mathcal{H}^2)A = -4\pi G a^2 \left(\delta P + \frac{2}{3}\nabla^2 \Pi \right) \, , \tag{72}$$

$$\sigma' + 2\mathcal{H}\sigma - C - A = 8\pi G a^2 \Pi \, , \tag{73}$$

subject to the Einstein energy-momentum constraints (the time-time and time-space components of the Einstein equations):

$$3\mathcal{H}(-C + \mathcal{H}A) + \nabla^2 (C - \mathcal{H}\sigma) = -4\pi G a^2 \delta\rho \, , \tag{74}$$

$$-C' + \mathcal{H}A = -4\pi G a^2 (\rho + P)(v + B) \, . \tag{75}$$

Finally, not independent from the above equations, we also have the energy and momentum conservation (continuity and Euler) equations:

$$\delta\rho' + 3\mathcal{H}(\delta\rho + \delta P) + 3(\rho + P)C' + (\rho + P)\nabla^2(v + E') = 0 \,, \quad (76)$$

$$(v + B)' + (1 - 3c_s^2)\mathcal{H}(v + B) + \phi + \frac{1}{\rho + P}\left(\delta P + \frac{2}{3}\nabla^2\Pi\right) = 0 \,. \quad (77)$$

Note that the fluid quantities above introduced ($\delta\rho$, δP, etc.) refer to the total matter content, but if the components do not interact among themselves, these equations can also be considered individually for each one of the components that make up the balance of the cosmic energy budget.

Exploiting the gauge freedom, we may consider the continuity equation written in the uniform density gauge, i.e. for $\delta\rho = 0$ where $\zeta = C$ in Eq. (62):

$$3\mathcal{H}\delta P_{\rm nad} + 3(\rho + P)\zeta' + (\rho + P)\nabla^2 V = 0 \,, \quad (78)$$

where we identify the pressure perturbation with the non-adiabatic pressure in this uniform-density gauge, $\delta P_{\rm and} \equiv \delta P - (p'/\rho')\delta\rho$, and the conformal Newtonian velocity V was defined in Eq. (61). Rearranging (78) we have

$$\zeta' = -\mathcal{H}\frac{\delta P_{\rm nad}}{\rho + P} - \frac{1}{3}\nabla^2 V \,. \quad (79)$$

For fluids with a barotropic equation of state, $P = P(\rho)$, we automatically have zero non-adiabatic pressure, $\delta P_{\rm nad} = 0$. Thus, on large scales, where the contribution from the divergence of the conformal Newtonian velocity, $\nabla^2 V$, can be neglected, we have ζ being conserved.

The same argument can be applied to any non-interacting barotropic fluids (Lyth and Wands 2003). Thus, generalising the definition of the curvature perturbation (62), we get conserved perturbations on large scales for radiation and matter, ζ_γ and ζ_m. These can be written in terms of conformal Newtonian gauge quantities as

$$\zeta_\gamma = \frac{\delta_\gamma}{4} + \Phi \,, \quad (80)$$

$$\zeta_m = \frac{\delta_m}{3} + \Phi \,. \quad (81)$$

Initial conditions are set up at sufficiently early times and on very large scales. Considering only radiation and matter, we have the total curvature and entropy perturbations

$$\zeta = \frac{4\rho_\gamma\zeta_\gamma + 4\rho_m\zeta_m}{4\rho_\gamma + 3\rho_m} \,, \qquad S_m = 3(\zeta_m - \zeta_\gamma) \,. \quad (82)$$

Adiabatic initial conditions are defined as:

$$\zeta = \zeta_m = \zeta_\gamma = \text{constant} , \qquad S_m = 0 . \tag{83}$$

And the isocurvature initial conditions are defined as

$$\zeta_\gamma = 0 , \qquad S_m = 3\zeta_m = \text{constant} . \tag{84}$$

A full treatment of the initial conditions requires neutrinos and baryons to be dealt with separately, giving rise to two more isocurvature density modes (Bucher et al. 2000).

The above constants are in general dependent on the scale. This dependence is set during an inflationary era and is thought to come from primordial quantum fluctuations. A slow time-dependence of the evolution during inflation leads to a weak scale-dependence of the dimensionless power spectrum.

$$n - 1 \equiv \frac{d \ln \mathcal{P}_\zeta}{d \ln k} \approx 0 , \tag{85}$$

The intrinsic Sachs-Wolfe effect (71) on large scales (neglecting the velocity $\mathbf{V}_*$) can then be written as

$$\frac{\delta T}{T} = \frac{1}{4}\delta_{\gamma *} + \Psi_* = \zeta_\gamma + 2\Psi_* . \tag{86}$$

Let's now see how the evolution of the gravitational potential looks like, since we have seen its role in determining the Sachs-Wolfe effect. Consider a barotropic fluid with equation of state $p = w\rho$ and constant w. Neglecting anisotropic stresses, so that $\Phi = -\Psi$, the evolution equation is computed from the spatial trace of the Einstein equation (72) and reads:

$$\Psi'' + 3(1 + w)\mathcal{H}\Psi' + w\nabla^2\Psi = 0 . \tag{87}$$

On large scales or during matter domination ($w = 0$), where one can neglect the spatial gradient, we obtain a constant solution

$$\Psi_0 = -\frac{3(1 + w)}{5 + 3w}\zeta , \tag{88}$$

where we related this constant to ζ which we have already shown to be conserved on large scales and for adiabatic perturbations. This sets the initial conditions for the scales which subsequently enter the horizon.

The evolution of the gravitational potential behaves in a two very different ways from the radiation dominated phase compared to the matter dominated era. In

radiation, $w = 1/3$, for a comoving wavenumber k, we find

$$\Psi_k(\eta) = \zeta_\gamma \left[\frac{6}{(k\eta)^2} \cos\left(\frac{k\eta}{\sqrt{3}}\right) - \frac{6\sqrt{3}}{(k\eta)^3} \sin\left(\frac{k\eta}{\sqrt{3}}\right) \right] . \tag{89}$$

For super horizon scales ($k\eta \ll 1$) the above solution tends to a constant: $\Psi_k \to -2\zeta_\gamma/3$. But for sub horizon scales ($k\eta \to \infty$) the potential oscillates and decays, $\Psi_k \to 0$. Thus the growth of matter inhomogeneities will also be suppressed in this regime.

In the matter dominated era ($w = 0$) the gravitational potential is constant at all scales

$$\Psi_k(\eta) = -3\zeta_m/5 . \tag{90}$$

With this result, the intrinsic Sachs-Wolfe effect on large scales can be written as

$$\frac{\delta T}{T} = \zeta_\gamma + 2\Psi_* = \zeta_\gamma - \frac{6}{5}\zeta_m , \tag{91}$$

where we used the matter-dominated solution for the gravitational potential because, as previously noted, recombination takes place in this regime, $z_* < z_{\rm eq}$. For adiabatic perturbations ($S_m = 0$ and $\zeta_m = \zeta_\gamma$) the contribution is:

$$\left.\frac{\delta T}{T}\right|_{\rm ad} = -\frac{1}{5}\zeta_\gamma = \frac{1}{3}\Psi_* , \tag{92}$$

whereas for isocurvature perturbations ($\zeta_\gamma = 0$, $S_m = 3\zeta_m$):

$$\left.\frac{\delta T}{T}\right|_{\rm iso} = -\frac{2}{5}S_m = 2\Psi_* . \tag{93}$$

With these two formulas, the large-scale approximation (56) can be written as follows:

$$\frac{l(l+1)}{2\pi} C_l \approx \frac{1}{25}\mathcal{P}_\zeta(l/\eta_0) + \frac{4}{25}\mathcal{P}_S(l/\eta_0) + \frac{4}{25}\mathcal{C}_{\zeta S}(l/\eta_0) , \tag{94}$$

i.e. a contribution coming from the adiabatic perturbations, $\mathcal{P}_\zeta$, another from the isocurvature perturbations, $\mathcal{P}_S$, and their cross-correlation, $\mathcal{C}_{\zeta S}$ (recall that the power spectrum is a quadratic function of the perturbation variables).

5 Acoustic Oscillations

One of the striking features of CMB power spectrum is the presence of acoustic oscillations, which originate from sound waves in the baryon-photon fluid at the time of last scattering. We present here the physics which lies behind this phenomenon, following the approach of Hu (2008).

The coupled first-order energy and momentum conservation equations for radiation perturbations in conformal Newtonian gauge (neglecting for the moment the effect of baryons) are

$$\frac{1}{4}\delta'_\gamma = -\frac{1}{3}\nabla\cdot\mathbf{V}_\gamma - \Phi' \,, \tag{95}$$

$$\mathbf{V}'_\gamma = -\frac{1}{4}\nabla\delta_\gamma - \nabla\Psi \,. \tag{96}$$

Eliminating $\mathbf{V}_\gamma$ yields an oscillator equation:

$$\left(\frac{1}{4}\delta_\gamma + \Psi\right)'' - \frac{1}{3}\nabla^2\left(\frac{1}{4}\delta_\gamma + \Psi\right) = (\Psi - \Phi)'' \,. \tag{97}$$

5.1 Matter Era

It is easy to solve this equation in the matter era, since $\Psi = -\Phi =$ constant, therefore the right-hand-side vanishes and we get

$$\left(\frac{1}{4}\delta_\gamma + \Psi\right) = \left(\frac{1}{4}\delta_\gamma + \Psi\right)_0 \cos(ks) \,, \tag{98}$$

for $k \gg k_{\rm eq}$, where $k_{\rm eq}$ is the comoving wave-number of a scale which crosses the Hubble horizon at matter-radiation equivalence, i.e. $k_{\rm eq} = H_{\rm eq}a_{\rm eq}$. The sound horizon for the relativistic fluid is given by

$$s = \int_0^\eta c_s d\eta' = \frac{1}{\sqrt{3}}\eta \,. \tag{99}$$

We shall see later that the presence of baryons affects the sound horizon since it modifies the speed of sound.

At recombination, for adiabatic perturbations, the above solution can be written as

$$\left(\frac{1}{4}\delta_\gamma + \Psi\right)_* = -\frac{1}{5}\zeta_\gamma \cos(ks_*) \,. \tag{100}$$

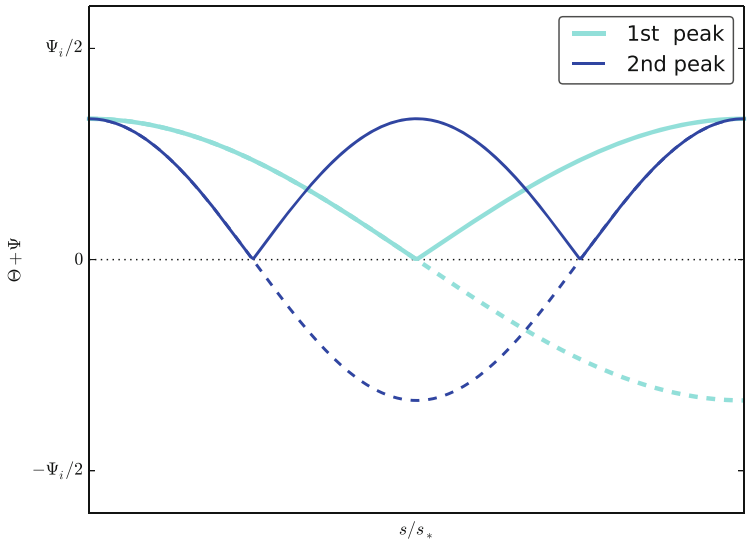

Fig. 1 Evolution of the effective temperature, Eq. (98), and its absolute value (corresponding to the power peaks, *solid lines*) (Hu 2008; Hu and Dodelson 2002)

We see in Fig. 1 the characteristic oscillating behaviour of the CMB temperature fluctuations, known as acoustic oscillations. Note that the power spectrum corresponds to the square of the amplitude so that peaks in the power spectrum occur at maxima or minima of the amplitude.

The first peak takes place when the argument of the cosine is equal to π, i.e. the comoving acoustic scale, λ_A, is defined as

$$k_A s_* = \pi \quad \Rightarrow \quad \lambda_A = \frac{2\eta_*}{\sqrt{3}} \,. \tag{101}$$

Dividing by the comoving distance to recombination, we find the angular scale

$$\theta_A = \frac{\lambda_A}{\eta_0 - \eta_*} \,, \tag{102}$$

which in the matter-dominated universe can be approximated as

$$\theta_A \approx \frac{\eta_*}{\eta_0} \approx z_*^{-1/2} \approx 2^\circ \,. \tag{103}$$

This is the angular scale of the particle horizon at recombination. It spans only 2° in the sky, and yet we see a high degree of isotropy in the CMB sky on all angular scales. No causal process could lead to isotropy on scales separated by more than 2° in the classical hot big bang because of the lack of causal connection on any large

scales. This is the well-known horizon problem in the big bang model which is most clearly seen in the CMB, and to which inflation offers a solution.

5.2 Radiation Driving

In the radiation dominated era, the gravitational potentials cannot be assumed to be constant and they rapidly oscillate and decay as each scale enters the horizon, i.e. $\Phi = -\Psi \to 0$ as $k\eta \to \infty$. Thus we can neglect the source term in the oscillator equation (97) for $k\eta \gg 1$ to obtain

$$\left(\frac{1}{4}\delta_\gamma + \Psi\right) \approx \frac{1}{4}\delta_\gamma \propto \cos(ks) \ . \tag{104}$$

Extrapolating forward to the matter era, we obtain the same oscillations for the radiation density as before, Eq. (100), but with larger amplitude the potentials are negligible. For scales that enter the horizon well before matter-radiation equality, $k \gg k_{\rm eq}$, we find

$$\left(\frac{1}{4}\delta_\gamma + \Psi\right)_* \approx -\zeta_\gamma \cos(ks_*) \ . \tag{105}$$

Comparing with Eq. (100), we find that the amplitude of oscillations for $k > k_{\rm eq}$ can be five times larger for low matter density, as this delays matter-domination, which causes the gravitational potential to decay on sub-horizon scales during the radiation era, see Fig. 2.

5.3 Baryon Loading

What happens if we include the effect of baryons which are strongly coupled to the electrons, and hence photons, before recombination? Intuitively, since baryons are massive particles, they would slow down the oscillations.

We define the coupled baryon-photon velocity, $V_{b\gamma}$, using

$$(\rho_\gamma + P_\gamma)\mathbf{V}_\gamma + (\rho_b + P_b)\mathbf{V}_b = (1+R)(\rho_\gamma + P_\gamma)\mathbf{V}_{b\gamma} \ , \tag{106}$$

where the baryon-to-photon ratio is given by

$$R \equiv \frac{\rho_b + P_b}{\rho_\gamma + P_\gamma} = \frac{3}{4}\frac{\rho_b}{\rho_\gamma} \ . \tag{107}$$

In the tight-coupling limit, $\mathbf{V}_{b\gamma} = \mathbf{V}_\gamma = \mathbf{V}_b$.

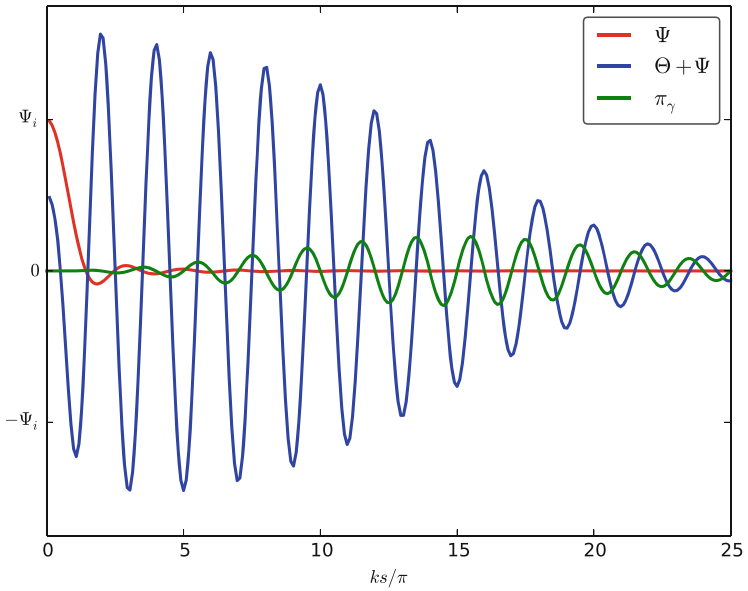

Fig. 2 Evolution, starting in the radiation era, of the gravitational potential, Ψ (*red*), the effective temperature, $(1/4)\delta_\gamma + \Psi$ (*blue*), and the anisotropic pressure due to photon diffusion, π_γ (*green*) (Hu 2008)

Using the coupled energy and momentum conservation equations we can write

$$\frac{1}{4}\delta'_\gamma = -\frac{1}{3}\nabla\cdot\mathbf{V}_{b\gamma} - \Phi' \,, \tag{108}$$

$$\left[(1+R)\mathbf{V}_{b\gamma}\right]' = -\frac{1}{4}\nabla\delta_\gamma - (1+R)\nabla\Psi \,. \tag{109}$$

Neglecting the time-variation of Ψ and R (i.e., $\Psi' \approx R' \approx 0$, which is a reasonable approximation) one gets the following oscillator equation:

$$\left[\frac{1}{4}\delta_\gamma + (1+R)\Psi\right]'' - \frac{1}{3(1+R)}\nabla^2\left[\frac{1}{4}\delta_\gamma + (1+R)\Psi\right] \approx 0 \,. \tag{110}$$

Comparing with the earlier oscillator equation neglecting baryons, (97), we see that the adiabatic speed of sound (the term multiplying the Laplacian is the square of the sound speed), and now is reduced by the presence of baryons by a factor $(1+R)$. The solution for the above equation is:

$$\left[\frac{1}{4}\delta_\gamma + (1+R)\Psi\right] = \left[\frac{1}{4}\delta_\gamma + (1+R)\Psi\right]_0 \cos(ks) \,, \tag{111}$$

where now the sound horizon is

$$s \equiv \int c_s d\eta = \int \frac{d\eta}{3(1+R)} \,. \tag{112}$$

So the matter-era solution for $k \gg k_{\rm eq}$, (100), corrected by the presence of baryons is now:

$$\left(\frac{1}{4}\delta_\gamma + \Psi\right)_* \approx [3R - (1+3R)\cos(ks_*)]\,\frac{1}{5}\zeta_\gamma \,. \tag{113}$$

The oscillations are now asymmetric: compression and rarefaction are no longer symmetric because baryons tend to collapse under their own gravity, more than they become more rarefied.

In Fig. 3 we present how the baryon loading modifies the shape of the CMB temperature power spectrum. We vary the baryon density $\Omega_b h^2$ but keep all the other parameters fixed. Because of Friedmann constraint (5)

$$1 = \Omega_b + \Omega_c + \Omega_\Lambda + \Omega_k \,, \tag{114}$$

we must therefore let h vary while keeping Ω_b fixed. Increasing the baryonic density increases the height of the first peak and lowers the second peak. This is due to the

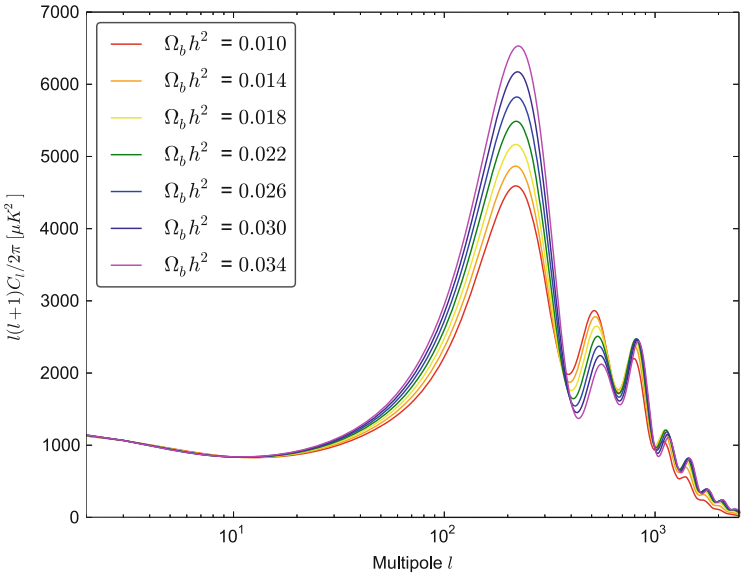

Fig. 3 CMB temperature power spectra, for different values of $\Omega_b h^2$, showing the effect of baryon loading. We used the CAMB code (Lewis et al. 2000) for the fiducial ΛCDM model given with the following parameters fixed $\Omega_b h^2 = 0.022$, $\Omega_c h^2 = 0.12$, $\Omega_\Lambda = 0.7$, $\tau = 0.1$ and $\Omega_k = 0$

fact that, the first peak is a maximum of compression, it is enhanced by a heavier baryon load, while the second peak is a rarefaction peak.

In Fig. 4 we present how the dark matter content modifies the shape of the CMB power spectrum. We use the same strategy as in Fig. 3, but now we let $\Omega_c h^2$ vary. As one can see, increasing this quantity causes the whole peak structure to decrease in amplitude. Since $\Omega_b h^2$ is fixed, the matter-radiation equivalence epoch takes place at earlier times when increasing $\Omega_c h^2$. This implies weaker radiation driving and a smaller amplitude of oscillations.

There is also another very important feature characterising CMB anisotropies on small scales. On scales comparable to the photon diffusion scale the tight-coupling approximation breaks down. The diffusion comoving scale has the following expression in terms of cosmological parameters (Hu 2008)

$$\lambda_{\rm D} \approx 64.5\ {\rm Mpc} \left(\frac{\Omega_m h^2}{0.14}\right)^{-0.278} \left(\frac{\Omega_b h^2}{0.024}\right)^{-0.18} . \tag{115}$$

Photon diffusion translates into a damping of the oscillations, see Fig. 2, and a decay in the correlation, i.e., the angular power spectrum, C_l. This is easily understood. On very small scales, below $\lambda_{\rm D}$, cold and hot photons mix thereby averaging to zero the correlation.

Thus far we have assumed that all photons last scattered at the time of recombination, but this is an approximation. The absence of a Gunn-Petersen trough (no

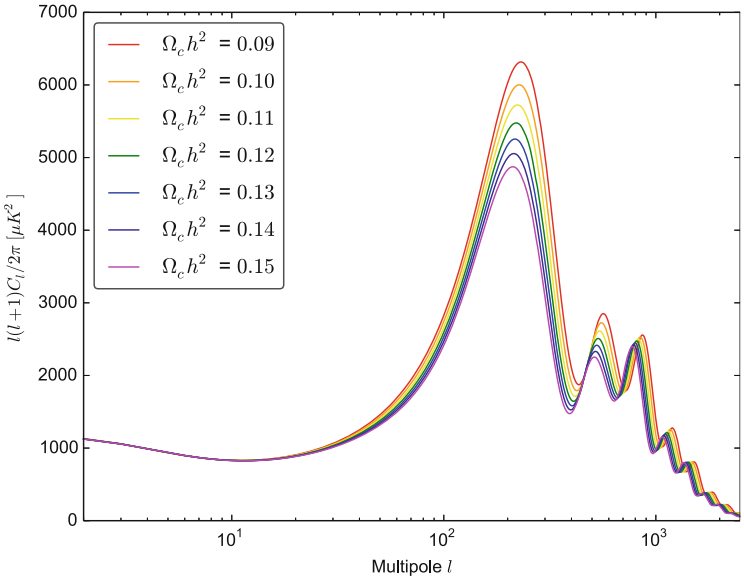

Fig. 4 CMB temperature power spectra, for different values of $\Omega_c h^2$, showing the effect of radiation driving

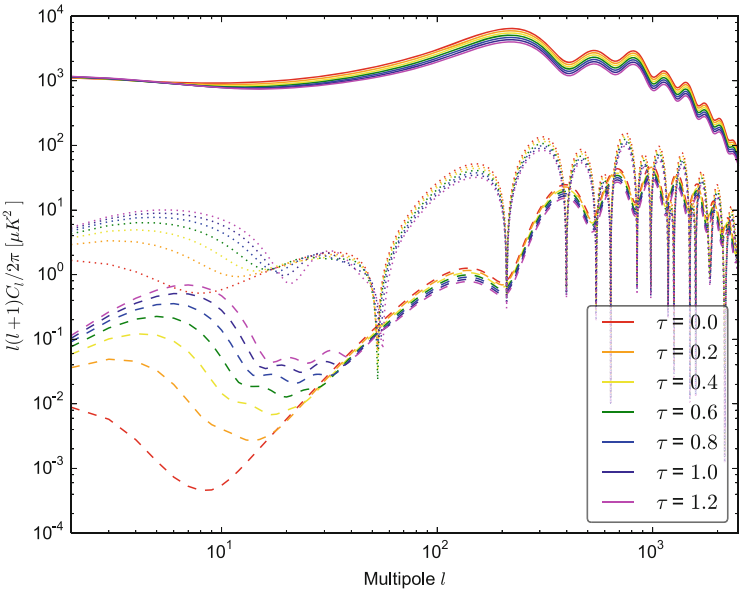

Fig. 5 Effects of reionisation on the temperature (*solid lines*), EE (*dotted lines*) and BB (*dashed lines*) power spectra

absorption by neutral hydrogen in quasar spectra) suggests that neutral gas has been reionised by a redshift $z > 6$. Reionisation leads to an optically thin "smog" between us and recombination with an optical depth $\tau \approx 0.1$, i.e., 10 % of the photons are scattered again. This reionisation suppresses small-scale anisotropies by rescattering the photons which also tends to average out the temperature anisotropies. The effect of varying the reionisation optical depth is shown in Fig. 5.

5.4 *Parameter Constraints from Peak Structure*

The dependence of the acoustic peak structure on a variety of different physical parameters enables cosmologists to determine these cosmological parameters with unprecedented accuracy from detailed measurements of the CMB sky. This is often referred to as the era of "precision cosmology".

From Planck, we have a precise constraint on the angular scale of the first acoustic peak (Ade et al. 2014b):

$$\theta_* = (1.04148 \pm 0.00066) \times 10^{-2} \,. \tag{116}$$

In flat space this angle is given by $\theta_* = s_*/D_*$, i.e., the ratio between the sound horizon at recombination

$$s_* = \int_0^{\eta_*} c_s d\eta \propto \Omega_m h^2 \,, \tag{117}$$

and the present horizon distance to recombination

$$D_* = \eta_0 - \eta_* = \int_0^{z_*} \frac{dz}{H(z)} \propto h^{-1} \,. \tag{118}$$

Thus in flat space we find

$$\theta_* = \frac{s_*}{D_*} \propto \Omega_m h^3 = 0.0959 \pm 0.0006 \,. \tag{119}$$

More generally, in a curved space with curvature radius R, we have $\theta_* = s_*/D_A$, see Fig. 6. The angular diameter distance, D_A, can be written for $D_* \ll R$ as

$$D_A = R \sin\left(\frac{D_*}{R}\right) \approx D_* \left(1 + \frac{\Omega_k H_0^2 D_*^2}{6}\right) \,, \tag{120}$$

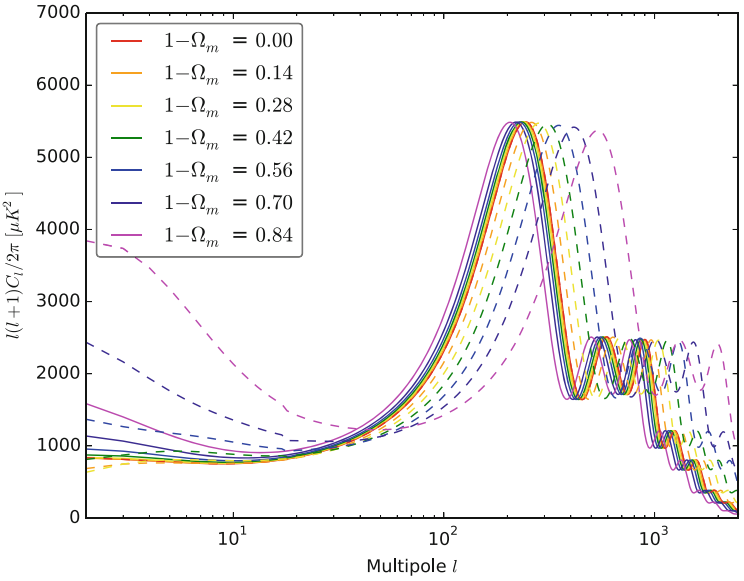

Fig. 6 Effects of varying the cosmological constant (*solid lines* showing $\Omega_m + \Omega_\Lambda = 1$) and spatial curvature (*dashed lines* showing $\Omega_m + \Omega_k = 1$)

from which we obtain the bound (Ade et al. 2014b)

$$\Omega_k = -0.0042^{+0.043}_{-0.048} \,. \tag{121}$$

As yet there is no evidence for spatial curvature.

We can derive a lot more information from the peak structure. For example, as discussed earlier, the second peak height relative to the first peak is related to the baryon loading. As shown in Fig. 3, this suppresses the even (rarefaction peaks) peaks with respect to the odd numbered peaks (compression peaks). From the relative height of the second peak, the following constraint on the baryon content is found (Ade et al. 2014b):

$$\Omega_b h^2 = 0.02207 \pm 0.00033 \,. \tag{122}$$

The decay of the gravitational potential in the radiation era enhances the third and higher peaks. The less matter there is, the higher the peaks are enhanced, as shown in Fig. 4. This enables us to put a constraint on the matter density today (Ade et al. 2014b):

$$\Omega_m h^2 = 0.1423 \pm 0.0029 \,. \tag{123}$$

Note that combining the two bounds (119) and (123) yields a bound on the Hubble constant (3) directly from the CMB (Ade et al. 2014b)

$$H_0 = 67.3 \pm 1.2 \text{ km s}^{-1} \text{ Mpc}^{-1} \,. \tag{124}$$

The diffusion length (115) and hence the damping tail is then fixed once $\Omega_b h^2$ and $\Omega_m h^2$ are specified in the basic ΛCDM cosmology. Reionisation also suppresses anisotropies at all small angles, above $l = 20$. This is approximately degenerate with the primordial spectral tilt (85), but this degeneracy can be broken by polarisation (see next section). Planck data combined with WMAP polarisation data requires a spectral tilt (Ade et al. 2014b)

$$n = 0.9603 \pm 0.0073 \,. \tag{125}$$

6 Polarisation

Thomson scattering is due to the motion of charged particles responding to an incident electromagnetic wave. The outgoing radiation emitted by an electron responding to a single incident wave is polarised in the direction of motion of the electron. However for the CMB photons emitted in a given direction at last scattering to have a net polarisation requires the electrons on the last scattering surface to see an anisotropic distribution of incoming photons. We have seen that tight-coupling

between electrons and photons leads to an approximately isotropic distribution of photons before recombination, therefore CMB radiation from last-scattering is only weakly polarised, due to a small quadrupole moment caused by small, but finite photon diffusion before last scattering. However about 10 % of CMB photons are re-scattered long after recombination due to reionisation, when the photon field is anisotropic, and hence reionisation provides a source of polarisation on large angular scales.

In general we define the polarization tensor in terms of the electric field, **E** (Hu 2008),

$$\mathbf{P}_{ij} \propto \langle \mathbf{E}\mathbf{E}^{\dagger} \rangle \propto \begin{pmatrix} \Theta + Q & U \\ -U & \Theta - Q \end{pmatrix} , \tag{126}$$

where Q and U are the two possible states of linear polarization and the angular brackets here denote the time average. Circular polarization is neglected here.

Since the polarisation state is invariant under a 90° rotation, it corresponds to a spin-2 field. Like the temperature anisotropy, we can decompose the polarisation in any direction into harmonic functions, E and B modes, across the whole sky,

$$Q(\hat{\mathbf{n}}) \pm iU(\hat{\mathbf{n}}) = \sum_{\ell,m} (E_{\ell m} \pm iB_{\ell m}) \, {}_{\pm 2}Y_{\ell m}(\hat{\mathbf{n}}) , \tag{127}$$

where ${}_{\pm 2}Y_{\ell m}(\hat{\mathbf{n}})$ are spin-2 spherical harmonics.

Scalar perturbations are longitudinal wave-modes where the inhomogeneities which give rise to polarised radiation are in the same direction as the wave propagates. This symmetry ensures that only E-mode polarisation is generated by scalar perturbations at linear order, corresponding to polarisation parallel or perpendicular to density gradients. Figure 7 shows the angular power spectrum for E-modes alongside the temperature power spectrum, and their cross-correlation. It shows the E-mode polarisation "bump" at large angular scales ($\ell < 20$) which can be used as a sensitive measure of reionisation and optical depth τ.

B-mode polarisation can only be generated at first order by transverse waves, i.e., vector or tensor perturbations. Initial vector perturbations decay at linear order in an expanding cosmology, and are completely absent in inflation driven by scalar fields. However tensor perturbations correspond to free perturbations of the metric, i.e., gravitational waves.

Consider a spatial metric perturbation:

$$h_{ij}(\mathbf{x}, \eta) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} e^{i\mathbf{k}\cdot\mathbf{x}} \left[h_{\mathbf{k}} e_{ij}(\mathbf{k}) + \bar{h}_{\mathbf{k}} \bar{e}_{ij}(\mathbf{k}) \right] . \tag{128}$$

Tensor perturbations must be transverse and traceless, i.e.,

$$k^i e_{ij} = 0 , \qquad k^i \bar{e}_{ij} = 0 , \qquad g^{ij} e_{ij} = 0 , \qquad g^{ij} \bar{e}_{ij} = 0 . \tag{129}$$

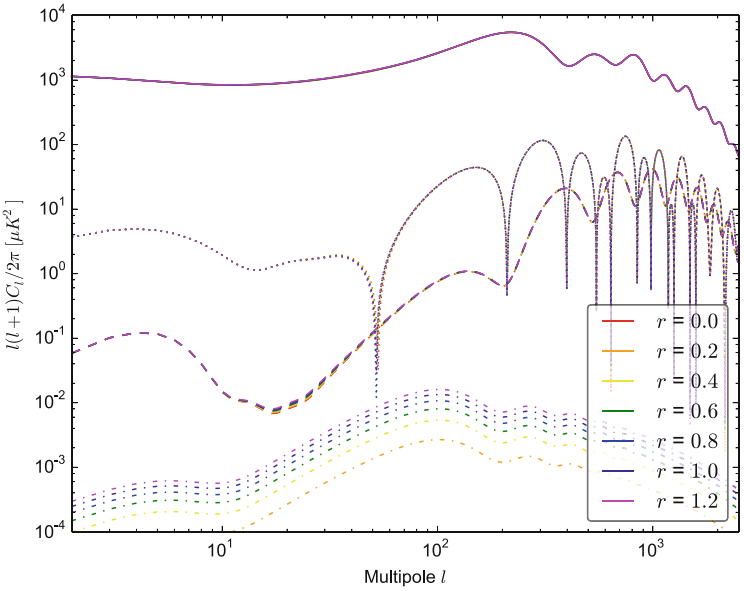

Fig. 7 Effects of varying tensor-to-scalar ratio, r. The *solid lines* (*top*) correspond to the total intensity (temperature) angular power spectrum. The *dashed lines* correspond to the E-mode power spectrum and the *dotted lines* correspond to the temperature-E-mode cross-correlation. These are dominated by scalar modes. The *dot-dashed lines* show the B-mode angular power spectrum for different values of the primordial tensor-to-scalar ratio, r

This leaves only two independent degrees of freedom, which are the two polarisation modes characterising a gravitational wave.

The Einstein equations give the following evolution equation for the gravitational wave amplitude

$$\ddot{h}_k + 3H\dot{h}_k + \frac{k^2}{a^2}h_k = 0\ , \tag{130}$$

i.e., a wave equation for a massless field, including a damping term due to the expansion of the Universe. Quantum vacuum fluctuations in massless fields during slow-roll inflation generate an almost scale-invariant spectrum of primordial perturbations on super-Hubble scales $k < aH$. Therefore, we can predict a primordial power spectrum for gravitational waves from inflation:

$$\mathcal{P}_T(k) = 2\frac{4\pi k^3}{(2\pi)^3}\langle hh^\dagger\rangle \approx 2\frac{32\pi}{M_{Pl}^2}\left.\frac{H^2}{2\pi}\right|_{k=aH}\ , \tag{131}$$

where the average is the vacuum expectation value, classically promoted to a variance. The power spectrum is thus directly determined by the Hubble parameter, H, during inflation.

It is customary to introduce the tensor-to-scalar ratio:

$$r \equiv \frac{\mathcal{P}_T}{\mathcal{P}_\zeta} \,. \tag{132}$$

From Planck constraints on the temperature power spectrum we obtain (Ade et al. 2014b), $r < 0.11$, and the energy scale at the end of inflation is constrained to be $V < 2 \times 10^{16}$ GeV. Nonetheless there is considerable interest in searching for B-mode polarisation in the CMB, either from future satellite experiments or dedicated ground-based experiments, as a direct signal of primordial gravitational waves. If an almost scale invariant, Gaussian distribution of primordial gravitational waves were discovered then these would surely be strong evidence for inflation, and our first evidence for the quantum nature of gravity.

7 The Next Frontier in CMB Theory

The observed CMB sky is remarkably uniform with temperature variations less than one part in 10^4. These lectures have reviewed the analysis of the homogeneous "background" CMB sky and anisotropies in the CMB temperature and polarisation, modelled using first-order perturbation theory. As the precision of CMB experiments improves the next challenge in CMB theory may be to accurately model non-linearity in the CMB anisotropies, both at last scattering and along the subsequent line-of-sight.

Non-linear interactions lead to departures from Gaussianity, evident in the CMB bispectrum, i.e., a non-zero correlation between different spherical harmonics. Primordial non-Gaussianity is often described in terms of the dimensionless non-linearity parameter (Bartolo et al. 2004)

$$f_{NL} \equiv \frac{B_\zeta(k_1, k_2, k_3)}{P_\zeta(k_1)P_\zeta(k_2) + P_\zeta(k_2)P_\zeta(k_3) + P_\zeta(k_1)P_\zeta(k_3)} \,, \tag{133}$$

i.e., the primordial bispectrum, $B_\zeta(k_1, k_2, k_3)$, relative to the square of the power spectrum. This $f_{NL} \sim 1$ corresponds to a primordial bispectrum $B_\zeta(k_1, k_2, k_3) \sim 10^{-18}$. It is related to the three-point correlation function, which is identically vanishing in the Gaussian case, along with all the odd-order correlation functions. In general, f_{NL} defined in this way is a scale- and shape-dependent function of the three wave numbers, but for local-type non-Gaussianity f_{NL} is a constant parameter (Wands 2010).

Non-linear interactions also lead to important conceptual differences from the simple assumptions valid in linear theory. Small but non-zero vector and tensor perturbations, hence B-mode polarisation, are generated at second-order from first-order scalar perturbations (Mollerach et al. 2004; Fidler et al. 2014). Also, second-order anisotropies in the photon distribution can no longer be described simply by

a black-body spectrum with anisotropic temperature. Second-order effects lead to anisotropic spectral distortions. For example, the angular power spectrum of the Compton y-distortion could provide a powerful tracer of reionisation (Pitrou et al. 2010).

Weak lensing along the line of sight is already an important nonlinear effect seen in current data, and needs to be taken into account in parameter estimates using Planck data. It is caused by many small-angle deflections by non-linear structures along the line of sight. This redistributes power in the small-scale angular power spectrum, smoothing out peaks at high ℓ in the power spectrum. This anisotropic lensing of the small scale power in the CMB has been used by the Planck team to create a map of the lensing potential and hence a measure of the integrated mass distribution along the line of sight (Ade et al. 2014c).

CMB weak-lensing also provides a non-zero contribution the angular bispectrum. Gravitational lensing leads to a second order anisotropy which is correlated with the integrated Sachs-Wolfe effect along the line of sight (Goldberg and Spergel 1999; Seljak and Zaldarriaga 1999; Lewis et al. 2011). This contributes a significant bias to estimates of the non-linearity parameter, and $f_{\rm NL} \approx 7$ has been seen in the Planck analysis. After subtracting this effect, Planck measurements remain consistent with vanishing primordial non-Gaussianity, $f_{NL} = 2.7 \pm 5.8$ (Ade et al. 2014d).

Most current bounds on primordial non-Gaussianity are based upon theoretical templates based on non-linear modelling of inflationary (or alternative) models, as non-Gaussian initial conditions for standard, linear Boltzmann codes such as CAMB (Lewis et al. 2000) or CLASS (Lesgourgues 2011). However the process of decoupling and the Sachs-Wolfe effect on temperature (and spectrum) anisotropies is in reality a non-linear process. As bounds on primordial non-Gaussianity become tighter we also need templates for the intrinsic non-Gaussianity expected from non-linear physics at recombination. There are now second-order general relativistic Boltzmann codes which have been developed (Huang and Vernizzi 2013; Pettinari et al. 2013; Su et al. 2014) building on pioneering early work (Pitrou 2011). Intrinsic non-Gaussianity at last-scattering provides a small bias, $f_{NL} \simeq 1$, which remains below the observational uncertainty of current experiments. However we are now in the position to be able to build a template for the intrinsic non-linearity at recombination which could be a target for future all-sky (hence space-based) missions, as shown in Fig. 8. This would make novel tests of physical process at last-scattering, e.g., gravitational wave production from density waves at second order.

8 Outlook

The standard ΛCDM cosmology provides a remarkably successful base model, able to explain many detailed features of the CMB revealed over the past 50 years by a series of ground-, balloon- and space-based experiments. In particular the Sachs-Wolfe plateau at large angular scales, the series of acoustic peaks in the angular

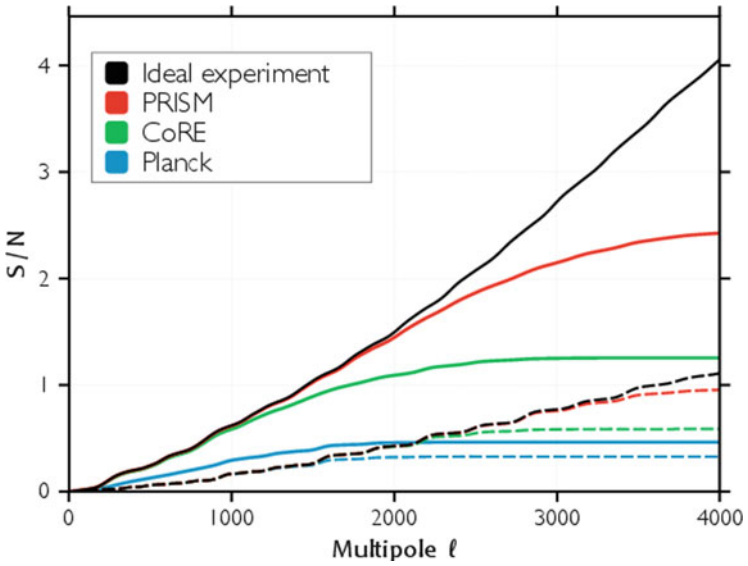

Fig. 8 Signal-to-noise ratio of the intrinsic bispectrum signal as a function of maximum multipole for Planck and proposed future satellite missions using temperature only (*dashed lines*) or temperature and polarisation (*solid lines*). Figure reproduced from Pettinari et al. (2014)

power spectrum, and the damping tail at small angular scales can be described by this model with just six cosmological parameters: the Hubble constant, the baryon and matter densities, the reionisation optical depth, and the amplitude and tilt of primordial perturbations. These six parameters are increasingly tightly constrained in the new era of precision cosmology, and bounds are set to become ever tighter, especially through new combinations with other data sets, such as high redshift galaxy surveys and HI (neutral hydrogen) survey data. The framework already successfully accommodates new observational discoveries such as the effect of weak lensing now seen in the CMB power spectrum and bispectrum.

Nonetheless there is no reason to believe this is the final theory of cosmology. Even the simplest, single-field models of inflation in the very early universe make predictions for additional features in the primordial perturbations, including a spectrum of tensor (gravitational wave) perturbations and small but finite running of the scalar spectral index. Many inflation models make further predictions including primordial isocurvature perturbations and/or non-Gaussianity. Any of these additional parameters would radically change our views about the likely mechanisms generating primordial structure. There are many additional cosmological parameters possible, including additional particle species and/or interactions, but there is no clear evidence yet requiring any more than the six basic parameters.

The present theoretical framework now being constrained by data was established in the 1970s and 80s well before the golden age of CMB experiments was begun by the COBE satellite results. Work now in progress will set new theo-

retical challenges for future experiments. Ground-based experiments are currently targeting CMB polarisation from weak-lensing and the elusive B-mode signature of primordial gravitational waves. Future space-based experiments are likely to focus on polarisation, spectral distortions and/or non-Gaussianity. The CMB will remain a key testing ground for cosmological theory for many years to come.

Acknowledgements DW is grateful to the organisers of the second José Plínio Baptista school for their warm hospitality. The authors are grateful to Rob Crittenden for helpful comments. This work is supported by STFC grants ST/K00090/1 and ST/L005573/1.

References

Adam, R., et al.: Planck 2015 results. I. Overview of products and scientific results (2015). arXiv:1502.01582

Ade, P.A.R., et al.: Planck 2013 results. I. Overview of products and scientific results. Astron. Astrophys. **571**, A1 (2014a). arXiv:1303.5062. doi:10.1051/0004-6361/201321529.

Ade, P.A.R., et al.: Planck 2013 results. XVI. Cosmological parameters. Astron. Astrophys. **571**, A16 (2014b). arXiv:1303.5076. doi:10.1051/0004-6361/201321591

Ade, P.A.R., et al.: Planck 2013 results. XVII. Gravitational lensing by large-scale structure. Astron. Astrophys. **571**, A17 (2014c). arXiv:1303.5077. doi:10.1051/0004-6361/201321543

Ade, P.A.R., et al.: Planck 2013 Results. XXIV. Constraints on primordial non-Gaussianity. Astron. Astrophys. **571**, A24 (2014d). arXiv:1303.5084. doi:10.1051/0004-6361/201321554

Ade, P.A.R., et al.: Planck 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts (2015). arXiv:1502.01597

Alpher, V.S.: Ralph A. Alpher, George Antonovich Gamow, and the prediction of the cosmic microwave background radiation. Asian J. Phys. **2**, 17–26 (2014). arXiv:1411.0172

Alpher, R.A., Bethe, H., Gamow, G.: The origin of chemical elements. Phys. Rev. **73**, 803–804 (1948). doi:10.1103/PhysRev.73.803

Bardeen, J.M.: Gauge invariant cosmological perturbations. Phys. Rev. **D22**, 1882–1905 (1980). doi:10.1103/PhysRevD.22.1882

Bartolo, N., Komatsu, E., Matarrese, S., Riotto, A.: Non-Gaussianity from inflation: theory and observations. Phys. Rep. **402**, 103–266 (2004). arXiv:astro-ph/0406398. doi:10.1016/j.physrep.2004.08.022

Bleem, L.E., et al.: Galaxy clusters discovered via the Sunyaev-Zel'dovich effect in the 2500-square-degree SPT-SZ survey. Astrophys. J. Suppl. **216**(2), 27 (2015). arXiv:1409.0850. doi:10.1088/0067-0049/216/2/27

Bucher, M., Moodley, K., Turok, N.: The general primordial cosmic perturbation. Phys. Rev. **D62**, 083508 (2000). arXiv:astro-ph/9904231. doi:10.1103/PhysRevD.62.083508

Chluba, J., Sunyaev, R.A.: The evolution of CMB spectral distortions in the early Universe. Mon. Not. R. Astron. Soc. **419**, 1294–1314 (2012). arXiv:1109.6552. doi:10.1111/j.1365-2966.2011.19786.x

Crittenden, R.: http://icg.port.ac.uk/~crittenr/teaching.html (2016). Online. Accessed 7 July 2016

Dodelson, S.: Modern cosmology. Academic (2003)

Fidler, C., Pettinari, G.W., Beneke, M., Crittenden, R., Koyama, K., et al.: The intrinsic B-mode polarisation of the cosmic microwave background. JCAP **1407**, 011 (2014). arXiv:1401.3296. doi:10.1088/1475-7516/2014/07/011

Fixsen, D.J., Cheng, E.S., Gales, J.M., Mather, J.C., Shafer, R.A., et al.: The cosmic microwave background spectrum from the full COBE FIRAS data set. Astrophys. J. **473**, 576 (1996). arXiv:astro-ph/9605054. doi:10.1086/178173.

Friedmann, A.: On the possibility of a world with constant negative curvature of space. Z. Phys. **21**, 326–332 (1924). doi:10.1007/BF01328280

Goldberg, D.M., Spergel, D.N.: Microwave background bispectrum. 2. A probe of the low redshift universe. Phys. Rev. **D59**, 103002 (1999). arXiv:astro-ph/9811251. doi:10.1103/PhysRevD.59.103002

Hasselfield, M., Hilton, M., Marriage, T.A., Addison, G.E., Barrientos, L.F., et al.: The Atacama Cosmology Telescope: Sunyaev-Zel'dovich selected galaxy clusters at 148 GHz from three seasons of data. JCAP **1307**, 008 (2013). arXiv:1301.0816. doi:10.1088/1475-7516/2013/07/008

Hu, W.: Lecture Notes on CMB Theory: From Nucleosynthesis to Recombination (2008). arXiv:0802.3688

Hu, W.: http://background.uchicago.edu/~whu/ (2016). Online. Accessed 29 August 2016

Hu, W., Dodelson, S.: Cosmic microwave background anisotropies. Ann. Rev. Astron. Astrophys. **40**, 171–216 (2002). arXiv:astro-ph/0110414. doi:10.1146/annurev.astro.40.060401.093926

Huang, Z., Vernizzi, F.: Cosmic microwave background bispectrum from recombination. Phys. Rev. Lett. **110**(10), 101303 (2013). arXiv:1212.3573. doi:10.1103/PhysRevLett.110.101303

Hubble, E.: A relation between distance and radial velocity among extra-galactic nebulae. Proc. Natl. Acad. Sci. **15**, 168–173 (1929). doi:10.1073/pnas.15.3.168

Klein, O., Nishina, T.: Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac. Zeitschrift fur Physik **52**, 853–868 (1929). doi:10.1007/BF01366453

Komatsu, E.: http://www.mpa-garching.mpg.de/~komatsu/ (2016). Online. Accessed 7 July 2016

Lemaitre, G.: A homogeneous Universe of constant mass and growing radius accounting for the radial velocity of extragalactic nebulae. Ann. Soc. Sci. Brux. Ser. I Sci. Math. Astron. Phys. **A47**, 49–59 (1927). doi:10.1007/s10714-013-1548-3

Lesgourgues, J.: The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview (2011). arXiv:1104.2932

Lewis, A., Challinor, A., Lasenby, A.: Efficient computation of CMB anisotropies in closed FRW models. Astrophys. J. **538**, 473–476 (2000). arXiv:astro-ph/9911177

Lewis, A., Challinor, A., Hanson, D.: The shape of the CMB lensing bispectrum. JCAP **1103**, 018 (2011). arXiv:1101.2234. doi:10.1088/1475-7516/2011/03/018

Lifshitz, E.: On the Gravitational stability of the expanding universe. J. Phys. (USSR) **10**, 116 (1946)

Lyth, D.H., Wands, D.: Conserved cosmological perturbations. Phys. Rev. **D68**, 103515 (2003). arXiv:astro-ph/0306498. doi:10.1103/PhysRevD.68.103515

Malik, K.A., Wands, D.: Cosmological perturbations. Phys. Rep. **475**, 1–51 (2009). arXiv:0809.4944. doi:10.1016/j.physrep.2009.03.001

Mather, J.C., Cheng, E.S., Cottingham, D.A., Eplee, R.E., Fixsen, D.J., et al.: Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument. Astrophys.J. **420**, 439–444 (1994). doi:10.1086/173574

Mollerach, S., Harari, D., Matarrese, S.: CMB polarization from secondary vector and tensor modes. Phys. Rev. **D69**, 063002 (2004). arXiv:astro-ph/0310711. doi:10.1103/PhysRevD.69.063002

Mukhanov, V.: Physical foundations of cosmology. Cambridge university press, Cambridge (2005)

Mukhanov, V.F., Feldman, H.A., Brandenberger, R.H.: Theory of cosmological perturbations. Part 1. Classical perturbations. Part 2. Quantum theory of perturbations. Part 3. Extensions. Phys. Rep. **215**, 203–333 (1992). doi:10.1016/0370-1573(92)90044-Z

Peebles, P.J.E.: Principles of physical cosmology. Princeton University Press, Princeton (1993)

Penzias, A.A., Wilson, R.W.: A measurement of excess antenna temperature at 4080-Mc/s. Astrophys.J. **142**, 419–421 (1965). doi:10.1086/148307

Pettinari, G.W., Fidler, C., Crittenden, R., Koyama, K., Wands, D.: The intrinsic bispectrum of the cosmic microwave background. JCAP **1304**, 003 (2013). arXiv:1302.0832. doi:10.1088/1475-7516/2013/04/003

Pettinari, G.W., Fidler, C., Crittenden, R., Koyama, K., Lewis, A., et al.: Impact of polarization on the intrinsic cosmic microwave background bispectrum. Phys. Rev. **D90**(10), 103010 (2014). arXiv:1406.2981. doi:10.1103/PhysRevD.90.103010

Pitrou, C.: The tight-coupling approximation for baryon acoustic oscillations. Phys. Lett. **B698**, 1–5 (2011). arXiv:1012.0546. doi:10.1016/j.physletb.2011.02.058.

Pitrou, C., Bernardeau, F., Uzan, J.-P.: The y-sky: diffuse spectral distortions of the cosmic microwave background. JCAP **1007**, 019 (2010). arXiv:0912.3655. doi:10.1088/1475-7516/2010/07/019

Robertson, H.P.: Kinematics and world-structure. Astrophys. J. **82**, 284–301 (1935). doi:10.1086/143681

Sachs, R.K., Wolfe, A.M.: Perturbations of a cosmological model and angular variations of the microwave background. Astrophys. J. **147**, 73–90 (1967). doi:10.1007/s10714-007-0448-9

Seljak, U., Zaldarriaga, M.: Direct signature of evolving gravitational potential from cosmic microwave background. Phys. Rev. **D60**, 043504 (1999). arXiv:astro-ph/9811123. doi:10.1103/PhysRevD.60.043504

Smoot, G.F., Bennett, C.L., Kogut, A., Wright, E.L., Aymon, J., et al.: Structure in the COBE differential microwave radiometer first year maps. Astrophys.J. **396**, L1–L5 (1992). doi:10.1086/186504

Su, S.C., Lim, E.A., Shellard, E.P.S.: Cosmic microwave background bispectrum from nonlinear effects during recombination. Phys. Rev. **D90**(2), 023004 (2014). doi:10.1103/PhysRevD.90.023004

Sunyaev, R.A., Zeldovich, Ya.B.: Small scale fluctuations of relic radiation. Astrophys. Space Sci. **7**, 3–19 (1970)

Wands, D.: Local non-Gaussianity from inflation. Class. Quant. Grav. **27**, 124002 (2010). arXiv:1004.0818. doi:10.1088/0264-9381/27/12/124002

Wands, D., Malik, K.A., Lyth, D.H., Liddle, A.R.: A new approach to the evolution of cosmological perturbations on large scales. Phys. Rev. **D62**, 043527 (2000). arXiv:astro-ph/0003278. doi:10.1103/PhysRevD.62.043527

Weinberg, S.: Cosmology. Oxford University Press, Oxford (2008)

The Observational Status of Cosmic Inflation After Planck

Jérôme Martin

Abstract The observational status of inflation after the Planck 2013 and 2015 results and the BICEP2/Keck Array and Planck joint analysis is discussed. These pedagogical lecture notes are intended to serve as a technical guide filling the gap between the theoretical articles on inflation and the experimental works on astrophysical and cosmological data. After a short discussion of the central tenets at the basis of inflation (negative self-gravitating pressure) and its experimental verifications, it reviews how the most recent Cosmic Microwave Background (CMB) anisotropy measurements constrain cosmic inflation. The fact that vanilla inflationary models are, so far, preferred by the observations is discussed and the reason why plateau-like potential versions of inflation are favored within this subclass of scenarios is explained. Finally, how well the future measurements, in particular of *B*-Mode CMB polarization or primordial gravity waves, will help to improve our knowledge about inflation is also investigated.

1 Introduction

With the release of the Planck data 2013 (Ade et al. 2014a,b, 2013) and 2015 (Planck 2015; Ade et al. 2015a,b), and the recent BICEP2/Keck Array and Planck joint analysis (BICEP2/Keck et al. 2015), the theory of cosmic inflation (Starobinsky 1979, 1980; Guth 1981; Linde 1982; Mukhanov and Chibisov 1981, 1982; Starobinsky 1982) has acquired a new status. Several of its predictions such as spatial flatness of our Universe, the presence of Doppler peaks in the Cosmic Microwave Background (CMB) multipole moments, almost scale invariant power spectrum for density perturbations have been definitively confirmed by the recent CMB anisotropy measurements. That makes inflation a predictive and verified theory of the early Universe.

In fact, another remarkable outcome of the Planck data is that they also allow us to identify which version of inflation is most likely to have been realized in

J. Martin (✉)
Institut d'Astrophysique de Paris, UMR 7095-CNRS, 98bis boulevard Arago, 75014 Paris, France
e-mail: jmartin@iap.fr

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_2

Nature (Martin et al. 2014a,b; Martin 2013). As is well-known, inflation comes in different flavors but these different scenarios make different predictions and, thus, one can, at least in principle, distinguish among them. The fact that the primordial fluctuations are adiabatic and Gaussian to a relatively high degree of accuracy (Ade et al. 2014b, 2013) is an important indication that we probably deal with single-field slow-roll inflation (with standard kinetic term), the simplest but non-trivial model of inflation. Of course, the final word has not yet been spoken since many non-vanilla inflationary scenarios are still compatible with the data. But, presently, they are just not needed in order to explain CMB measurements even if this situation could change in the future.

The fact that we now have high accuracy CMB data at our disposal also allows us to detect the "fine structure" of inflation and to constrain the shape of the inflaton potential. Here again, the Planck data have provided precious information. We now know that the potential is of the plateau type and that simple monomials are disfavored (Ijjas et al. 2013; Martin et al. 2014a,b; Martin 2013). Moreover, we now start probing the reheating epoch (Martin and Ringeval 2010; Martin et al. 2015). Reheating is the epoch, after inflation and before the radiation dominated era of the standard hot Big bang phase, where the inflaton field decays and where all matter we see around us was produced (Turner 1983; Traschen and Brandenberger 1990; Kofman et al. 1997; Amin et al. 2015). It is therefore of major conceptual importance. And Planck 2013 and 2015 data put non trivial constraints on the physical processes that took place at that time (Martin and Ringeval 2006, 2010; Lorenz et al. 2008a; Martin et al. 2015; Munoz and Kamionkowski 2015; Dai et al. 2014; Gong et al. 2015).

The goal of these lectures, given at the second Jose Plinio Baptista school on Cosmology held in Pedra Azul (Brazil) in March 2014, is to review how the above conclusions can be established. Many reviews on inflation can be found in the literature (Martin 2004, 2005, 2008; Sriramkumar 2009) and there are technical papers reporting the astrophysical and cosmological observations, such as the Planck papers (Ade et al. 2014a,b, 2013). But, in between, few things can be found and the present article aims at filling this gap. In some sense, it can be viewed as a technical guide which, from a reasonable prior knowledge of inflation, permits a detailed understanding of the implications for inflation of the recent high accuracy CMB data.

These lecture notes are also written at a special time: the Planck 2013 and 2015 data (Ade et al. 2013, 2014a,b, 2015a,b; Planck 2015) have been released and their consequences (in fact, mainly the consequences of Planck 2013) already analyzed in several works. Moreover, very recently, a joint analysis made by the BICEP2/Keck Array team and the Planck collaboration (BICEP2/Keck et al. 2015) has been published showing that the BICEP2 detection of B-mode CMB polarization announced in Ade et al. (2014c) is mainly due to dust and cannot be attributed to primordial gravity waves produced during inflation. At the time of writing, the Planck 2015 scientific products (in particular, the likelihood) are expected to be delivered in June 2015 only. This means that reproducing or extending the Planck 2015 analysis is not yet possible. However, from what is already known, the Planck

2015 results are in good agreement with Planck 2013. Therefore, the conclusions discussed in the present article (model comparison, constraints on reheating etc ...) will most likely remain valid for the second release of the Planck data. Whenever available, we quote the values obtained by Planck 2015.

These lectures are also related to the lectures given by C. Byrnes on Non-Gaussianities (Byrnes 2014) and by D. Wands on CMB physics. Hopefully, these three reviews should provide the reader with a rather complete overview of modern primordial Cosmology and its observational implications. In particular, Byrnes (2014) reviews how Non-Gaussianities are produced in non-vanilla inflationary models while, here, we restrict ourselves to simple scenarios for which Non-Gaussianities are very small. The two lectures are therefore complementary. The lecture notes by D. Wands, O.F. Piattella and L. Casarini explain in details how CMB anisotropies are generated while, here, we just take it as a known fact (see also the recent review (Bucher 2015)). Therefore, the present article and the one on CMB physics are also complementary.

These lecture notes are organized as follows. In the next section, Sect. 2, we present general considerations on inflation. Rather than discussing inflation in detail, which can be found in many review articles, we just give the basics and choose to focus on the fundamental principles at the basis of the inflationary mechanism and its experimental justifications. In Sect. 3, we discuss how inflation can be realized in practice. In particular, in Sect. 3.1, we review how inflation can be embedded in high energy physics. Recently, alternative parameterizations have been considered and we discuss them in Sect. 3.2. In Sect. 3.3, we also review how the reheating phase can be described. Then, in Sect. 4, we discuss the theory of inflationary cosmological perturbations of quantum-mechanical origin. This part of the inflationary scenario is especially important because this is how one can relate theoretical predictions to astrophysical observations. In Sect. 4.1, we present the calculation of the two-point correlation functions, or power spectra, for scalar and tensor perturbations in the slow-roll approximation. In Sect. 4.2, we review the calculation of the three-point correlation function, or bispectrum, and in Sect. 4.3, the calculation of the four-point correlation function, or tri-spectrum. All these considerations are made in the slow-roll approximation and for single-field models with minimal kinetic terms. In Sect. 4.4, we discuss the isocurvature perturbations and how they can be produced in the framework of inflation. In Sect. 5, we use the tools introduced before and compare the inflation predictions to the high accuracy CMB Planck data. In Sect. 5.1, we consider the measurements of spatial curvature, in Sect. 5.2 the measurements of isocurvature perturbations and, in Sect. 5.3, those of Non-Gaussianities. Since these data indicate that single field models are preferred, we then focus on this class of scenarios. In Sect. 5.4, we give the constraints on the slow-roll parameters and on the derived power-law parameters, such as the spectral index, the running or the tensor-to-scalar ratio. We also discuss the implications of the recent joint analysis made by the BICEP2/Keck Array team and the Planck collaboration. In Sect. 5.5, we carry out a Bayesian analysis to do model comparison and determine what are the best models of inflation. In Sect. 5.6, we present the constraints on reheating that can be inferred from the Planck data. Finally, in

Sect. 6, we recap our main results and discuss which lesson can be drawn for our understanding of inflation and primordial cosmology.

2 General Considerations on Inflation

The motivations for introducing a phase of inflation, i.e. a phase of accelerated expansion, are well-known: postulating $\ddot{a} > 0$ (a is the Friedmann-Lemaître-Roberston-Walker -FLRW- scale factor) allows us to avoid the puzzles of the standard hot Big Bang theory (for a detailed discussion of these issues, see Martin 2004; Mukhanov 2005; Peter and Uzan 2009). If gravity is described by General Relativity (GR), then, in a homogeneous and isotropic Universe, the equations of motion are given by

$$H^2 + \frac{\mathcal{K}}{a^2} = \left(\frac{\dot{a}}{a}\right)^2 + \frac{\mathcal{K}}{a^2} = \frac{1}{3M_{\rm Pl}^2}\sum_i \rho_i \equiv \frac{1}{3M_{\rm Pl}^2}\rho, \tag{1}$$

$$-\left(2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{\mathcal{K}}{a^2}\right) = \frac{1}{M_{\rm Pl}^2}\sum_i p_i \equiv \frac{1}{M_{\rm Pl}^2}p, \tag{2}$$

where ρ_i and p_i are respectively the energy density and the pressure of the fluid "i". In the standard model of Cosmology, we have indeed a collection of different fluids, pressure-less matter (made of baryons and cold dark matter), radiation (made of photons and neutrinos) and dark energy. These different types of matter source the Einstein equations and control the dynamics of the expansion. Notice that the expansion rate of the Universe is given by the Hubble parameter which, according to the above equations, is defined by $H \equiv \dot{a}/a$ where a dot means a derivative with respect to cosmic time. The quantity $M_{\rm Pl}$ is the reduced Planck mass and, in the following, we will also use the quantity $\kappa \equiv 1/M_{\rm Pl}^2 = 8\pi G_{\rm N}$, $G_{\rm N}$ being the Newton constant. Finally, the quantity $\mathcal{K}$, that can always be normalized to 0 or ± 1, represents the curvature of the spatial sections. Notice that one can also define an effective curvature energy density by $\rho_{\rm curv} \equiv -3\mathcal{K}/(\kappa a^2)$ such that the Friedmann equation takes the form $H^2 = (\kappa/3)\sum_i \rho_i + (\kappa/3)\rho_{\rm curv}$. Defining $\Omega_i \equiv \rho/\rho_{\rm cri}$ and $\Omega_{\mathcal{K}} = \rho_{\rm curv}/\rho_{\rm cri}$, where the critical energy density is $\rho_{\rm cri} = 3H^2/\kappa$, the Friedmann equation can be rewritten as $\sum_i \Omega_i + \Omega_{\mathcal{K}} = 1$.

Let us now discuss under which physical conditions inflation can be obtained. The above equations can be combined and lead to the following formula which relates the acceleration of the expansion to the matter content of the Universe

$$\frac{\ddot{a}}{a} = -\frac{1}{6M_{\rm Pl}^2}\sum_i (\rho_i + 3p_i) \ . \tag{3}$$

This immediately implies that, in order to have inflation, the pressure must be negative, i.e. $p < -\rho/3$ where ρ and p are defined in Eqs. (1) and (2). Having realized that we need a negative pressure, the next question is of course which kind of matter can possess this property and this will be the subject of the two next sections. Of course, as is well-known, we will see that scalar fields are ideal candidates.

But before starting this discussion, it is interesting to notice that inflation is a genuine relativistic effect since it involves the term $3p$ in the above Eq. (3), which is absent in Newtonian physics. Indeed, let us consider a sphere of radius $R(t)$ and of uniform density ρ. A galaxy of mass m, located at the edge of the sphere, feels a gravitational field $\boldsymbol{G}$ that can be simply evaluated by means of the Gauss's law, $\int \boldsymbol{G} \cdot \mathrm{d}\boldsymbol{S} = 4\pi G_{\mathrm{N}} M$, where M is the mass of the sphere. This gives $G = G_{\mathrm{N}} M/R^2$. As a consequence, the acceleration of the galaxy can be written as

$$m\ddot{R} = -m\frac{G_{\mathrm{N}} M}{R^2}, \tag{4}$$

or

$$\frac{\ddot{R}}{R} = -\frac{4\pi G_{\mathrm{N}}}{3}\rho = -\frac{\rho}{6M_{\mathrm{Pl}}^2}, \tag{5}$$

where we have used $M = 4\pi\rho R^3/3$. This equation is similar to Eq. (3) except that the term $3p$ is not present. The physical reason behind the presence of this term is deeply rooted in the fundamental principles of GR: in GR, every form of energy weighs, including pressure.

The term $3p$ is so important for inflation that it is interesting to ask whether it plays a role in other physical situations and if its appearance has been tested experimentally and/or observationally. This is a difficult question since, in ordinary cases, the contribution of pressure is usually negligible, $p \ll \rho$. In fact, four situations where a gravitating pressure is important can be identified: inflation, dark energy but in some sense this is the same as inflation, neutron stars and Big Bang Nucleosynthesis (BBN). In particular, it is interesting to see what can be said about the $3p$ terms in the last two examples.

Let us start with the internal structure of a neutron star (Schwab et al. 2008). As is well-known, it is controlled by the Tolman-Oppenheimer-Volkoff equations that can be obtained in the following way. The metric for a static and spherically symmetric solution can be written as

$$\mathrm{d}s^2 = -\mathrm{e}^{2\Phi}\mathrm{d}t^2 + \mathrm{e}^{2\lambda}\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2\right), \tag{6}$$

where t is time, r a radial coordinate and θ and φ angular coordinates. The quantities Φ and λ are functions of r only. Matter is assumed to be described by a perfect fluid, the stress energy tensor of which can be expressed as

$$T_{\mu\nu} = (\rho + p)\, u_\mu u_\nu + p g_{\mu\nu}, \tag{7}$$

where $g_{\mu\nu}$ is the metric tensor and the normalized 4-velocity reads $u_\mu = \left(-\mathrm{e}^{\Phi}, 0, 0, 0\right)$. Then, the time-time and $r-r$ component of the Einstein equations read

$$\frac{1}{r^2}\mathrm{e}^{-2\lambda}\left(2r\frac{\mathrm{d}\Phi}{\mathrm{d}r}+1-\mathrm{e}^{2\lambda}\right)=\frac{1}{M_{\mathrm{Pl}}^2}p, \tag{8}$$

$$\frac{1}{r^2}\mathrm{e}^{-2\lambda}\left(-1+2r\frac{\mathrm{d}\lambda}{\mathrm{d}r}+\mathrm{e}^{2\lambda}\right)=\frac{1}{M_{\mathrm{Pl}}^2}\rho. \tag{9}$$

On the other hand, energy conservation, $\nabla_\mu T^{\mu\nu} = 0$, more precisely its radial component, implies that

$$\frac{\mathrm{d}p}{\mathrm{d}r}=-\left(\rho+p\right)\frac{\mathrm{d}\Phi}{\mathrm{d}r}. \tag{10}$$

The other components lead to the fact that ρ does not depend on time, θ or φ, that is to say $\rho = \rho(r)$. If we now define the gravitational mass $m(r)$ by

$$G_{\mathrm{N}}m(r)=\frac{r}{2}\left(1-\mathrm{e}^{-2\lambda}\right), \tag{11}$$

then Eq. (9) implies that

$$\frac{\mathrm{d}m}{\mathrm{d}r}=4\pi\rho(r)r^2. \tag{12}$$

Introducing the expression of the mass (11) in Eq. (8) in order to express $\mathrm{d}\Phi/\mathrm{d}r$ and, then, inserting the corresponding expression in the conservation Eq. (10) leads to

$$\frac{\mathrm{d}p}{\mathrm{d}r}=-(\rho+p)\frac{G_{\mathrm{N}}}{r^2(1-2mG_{\mathrm{N}}/r)}\left[m(r)+3p(r)\left(\frac{4}{3}\pi r^3\right)\right]. \tag{13}$$

The important point in this formula is that the term $3p$ participates to this expression. This means that self-gravity of pressure affects the internal structure of the neutron stars. In practice, in order to calculate this internal structure, one has first to choose an equation of state $\rho = \rho(p)$. Once this is done, one can integrate the two following equations

$$\frac{\mathrm{d}\rho}{\mathrm{d}r}=\frac{\mathrm{d}\rho}{\mathrm{d}p}\frac{\mathrm{d}p}{\mathrm{d}r},\quad \frac{\mathrm{d}m}{\mathrm{d}r}=4\pi\rho(r)r^2, \tag{14}$$

the last equation being nothing but Eq. (12). This leads to the functions $\rho(r)$ and $m(r)$. The radius of the star, R_{star}, is defined by $\rho\left(R_{\mathrm{star}}\right) = 0$ and its mass is given by $M_{\mathrm{star}} \equiv m\left(R_{\mathrm{star}}\right)$. One can then plot the mass-radius relation $M_{\mathrm{star}}(R_{\mathrm{star}})$. Of course, one obtains different mass-radius relations for different equations of state. Let us also notice that, at fixed equation of state, one obtains a curve, and not a

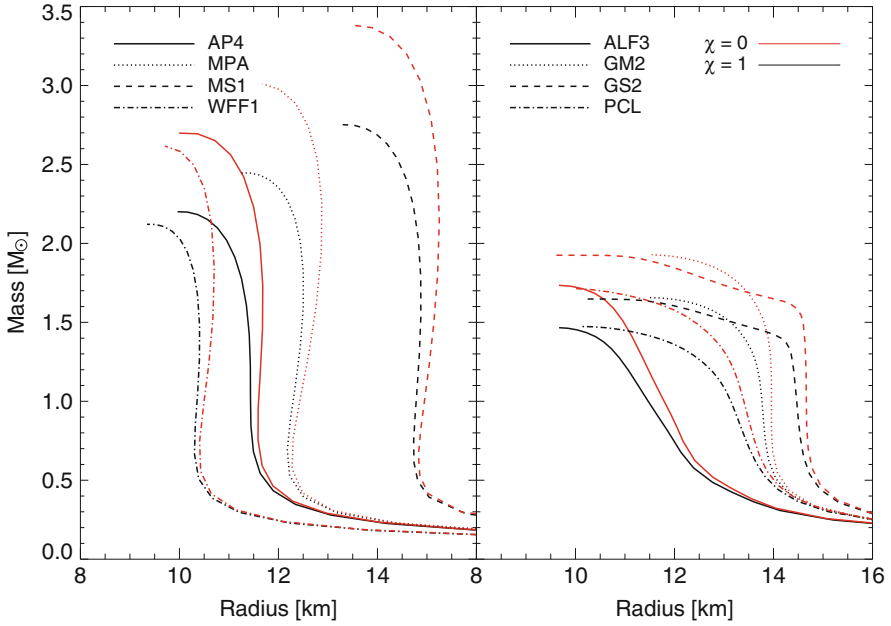

Fig. 1 Mass-radius relations of neutron stars for different equations of state ("standard" in the *left panel*, more "exotic" in the *right panel*). *Black curves* correspond to the standard GR calculation while *red curves* represent the case where self-gravity of pressure is absent. Figure taken from Schwab et al. (2008)

unique prediction, because one needs to specify $\rho\,(r=0)$ to be able to integrate the above equations. One thus has a family of points parametrized by $\rho\,(r=0)$. Several examples are displayed in Fig. 1 (black lines).

The fact that the structure of a neutron star depends on the general relativistic $3p$ term opens the possibility to experimentally test it. In order to do so, the idea of Schwab et al. (2008) is to study an ad-hoc modification of the Tolman-Openheimer-Volkoff equation such that

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -(\rho+p)\frac{G_{\mathrm{N}}}{r^2(1-2mG_{\mathrm{N}}/r)}\left[m(r)+3\chi p(r)\left(\frac{4}{3}\pi r^3\right)\right], \tag{15}$$

where χ is a new, phenomenological, parameter introduced by hand. The term $3p$ weighs normally when $\chi=1$ and does not weigh at all when $\chi=0$. Notice that $\chi=0$ is not the Newtonian limit because there are other relativistic terms in Eq. (15) (for instance $1-2mG_{\mathrm{N}}/r$ at the denominator). So the idea is now to re-derive the mass-radius relation for neutron stars and to see the influence of a parameter $\chi\neq 1$, the hope being to be able to put constraints on χ from astronomical observations. The results are shown in Fig. 1. The fact that red curves (namely those obtained with

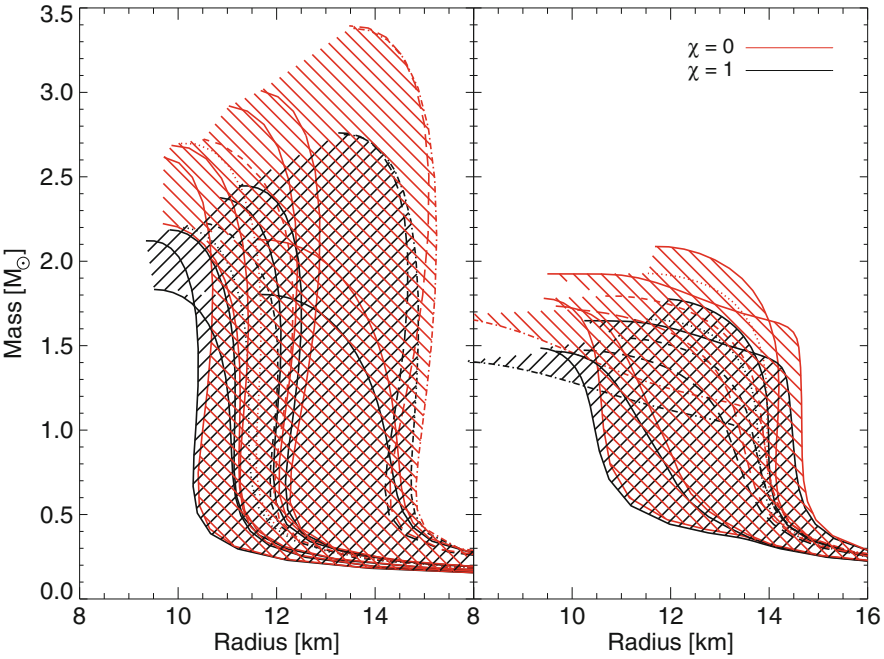

Fig. 2 Mass radius relations for different equations of state and associated theoretical uncertainties. In *black* are represented the mass radius relations obtained when $\chi = 1$ (standard GR calculation) while, in *red*, are represented the mass radius relations obtained without self-gravity pressure (namely $\chi = 0$). The *hatched regions* show the theoretical uncertainty associated with the fact that the equation of state is in fact unknown. It is clear from the plot that this completely dominates the differences between the $\chi = 1$ and $\chi = 0$ situations. Figure taken from Schwab et al. (2008)

$\chi = 0$) are different from the black ones (those obtained in the standard GR case) confirms that the $3p$ term has a significant influence of the mass-radius relation.

However, as shown in Fig. 2, the fact that the equation of state is not known accurately completely blurs the effect. Indeed, one sees that the corresponding uncertainty is typically of the same order of the effect we try to detect. Therefore, the conclusion is that, although it is true that self-gravity is crucial in order to predict correctly their mass-radius relation, at least for the moment, neutron stars cannot be used to experimentally test the $3p$ term.

Let us now turn to the other possibility, namely BBN (Rappaport et al. 2008). Since BBN takes place during the radiation dominated era for which $p = \rho/3$, it is clear that the $3p$ term should have an important impact on BBN. In order test the influence of the $3p$ term, we follow the same strategy as for neutron stars and introduce an ad-hoc modification of GR characterized by the χ parameter, namely

$$\frac{\ddot{a}}{a} = -\frac{1}{6M_{\rm Pl}^2}(\rho + 3\chi p). \tag{16}$$

This equation should be compared to Eq. (3). In order to derive the Friedmann equation, we need another equation and we can use the first law of thermodynamics for an adiabatic expansion, namely $\mathrm{d}(a^3\rho) = -p\mathrm{d}(a^3)$, written for a co-moving volume or, equivalently, $\dot{\rho} + 3\dot{a}(\rho + p)/a = 0$. Then, noticing that $\ddot{a}/a = 1/(2\dot{a}a)\mathrm{d}(\dot{a}^2)/\mathrm{d}t$ and using the conservation equation, it is straightforward to derive the following relation

$$\mathrm{d}\left(\dot{a}^2\right) = -\frac{1}{3M_{\mathrm{Pl}}^2}\left[(1-3\chi)\,\rho a\mathrm{d}a - \chi a^2\mathrm{d}\rho\right]. \tag{17}$$

If $\chi = 1$, it is easy to check that

$$\mathrm{d}\left(\dot{a}^2\right) = \frac{1}{3M_{\mathrm{Pl}}^2}\mathrm{d}\left(\rho a^2\right), \tag{18}$$

which gives

$$H^2 = \frac{\rho}{3M_{\mathrm{Pl}}^2} + \frac{C}{a^2}, \tag{19}$$

where C is an integration constant leading to a curvature term. Now, if $\chi \neq 1$ and $p = w\rho$, where w is a constant equation of state parameter, then one obtains

$$H^2 = \frac{1+3\chi w}{1+3w}\frac{\rho}{3M_{\mathrm{Pl}}^2} + \frac{C}{a^2}. \tag{20}$$

Using this modified Friedmann equation with $w = 1/3$ and ignoring the curvature term (which is sub-dominant in presence of radiation as shown by the cosmological data), one obtains

$$H^2 = \frac{1+\chi}{2}\frac{\rho_{\mathrm{rad}}}{3M_{\mathrm{Pl}}^2}. \tag{21}$$

Therefore, the effect of the term proportional to χ is to modify the expansion rate of the Universe in the radiation dominated era. Or, if one uses the fact that the energy density of radiation is $\pi^2 g_* T^4/30$, we see that this is also equivalent to changing the effective number of relativistic degrees of freedom, namely $g'_* = g_*(1+\chi)/2$.

Rappaport et al. (2008) has performed BBN calculations, assuming Eq. (21), and computed the abundance of deuterium, helium-4 and lithium-7. The isocontours are represented in Fig. 3 in the plane (η_{10}, χ). The parameter η_{10} is defined by $\eta_{10} \equiv 10^{10}\eta$ where $\eta \equiv n_{\mathrm{B}}/n_\gamma = \Omega_{\mathrm{B}}h^2\pi^2\rho_{\mathrm{cri}}/[m_{\mathrm{B}}h^2 2T^3\zeta(3)] \simeq 2.73\times 10^{-8}\Omega_{\mathrm{B}}h^2$

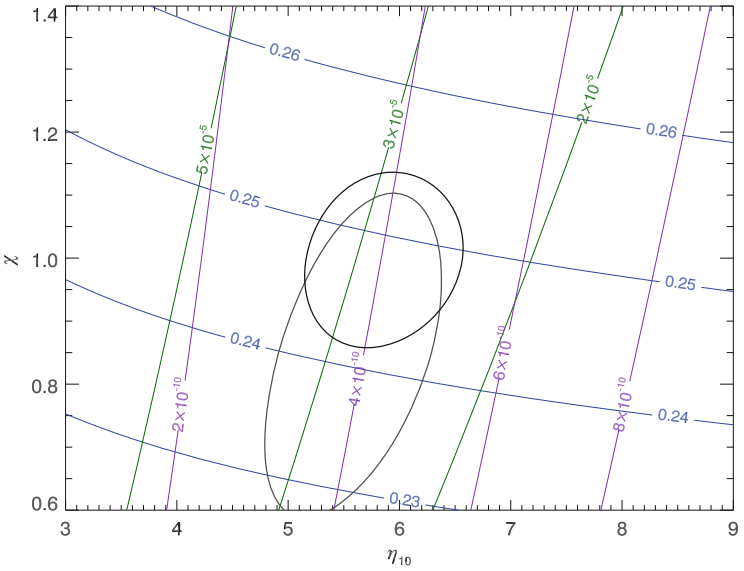

Fig. 3 Light elements abundances calculated when the Friedmann equation is modified according to Eq. (21). *Green* contours are for deuterium abundance, *blue* ones for helium-4 and *purple* ones for lithium-7. The two *gray ellipses* indicate the region in parameter space allowed by observations. Figure taken from Rappaport et al. (2008)

[in the last expression, $m_{\mathrm{B}} \simeq 939.6\,\mathrm{MeV}$ is the baryon (neutron) mass, $\rho_{\mathrm{cri}} \simeq 8.099 \times 10^{-47}\mathrm{GeV}^4$ is the critical energy density today, $T \simeq 2.7255\mathrm{K}$ is the CMB temperature and $\zeta(3) \simeq 1.20206$]. Green contours represent the deuterium abundance $(D/H)_{\mathrm{P}}$, blue contours are helium-4 abundance Y_{P} and purple contours are lithium-7 abundance. We see that deuterium abundance mainly determines η_{10} while helium-4 abundance gives good constraints on the new parameter χ. Observations indicate that $\log(D/H)_{\mathrm{P}} = -4.55 \pm 0.04$ (O'Meara et al. 2006) and $Y_{\mathrm{P}} = 0.24 \pm 0.006$ (Steigman 2007). Then, one can identify the region in the space (η_{10}, χ) which is consistent with those observations. This is indicated in Fig. 3 by the two gray ellipses (corresponding to two slightly different assumptions about the abundances inferred from the observations). Without entering a detailed discussion, the conclusion is that $\chi \simeq 1$ is compatible with observations and that the value $\chi = 0$ is strongly ruled out. Therefore, self-gravity of pressure is, in some sense, confirmed by cosmological observations.

The previous considerations "validate" the mechanism on which inflation is based. Inflation thus appears as a well-justified theory. In the next section, we therefore describe this theory in more detail and discuss the micro-physics of inflation.

3 The Micro-Physics of Inflation or How to Parametrize Inflation

3.1 Inflation and High Energy Physics

We have seen in the last section that, in order to have a phase of inflation, we need a situation where the fluid dominating the matter content of the Universe has a negative pressure. The next question is of course which type of matter can have this property. In order to answer this question, let us first remark that inflation is a high energy phenomenon by particle physics standards since it is supposed to occur in the early Universe. In this situation, the relevant framework to describe matter is not fluid mechanics but field theory. And the simplest field, compatible with isotropy and homogeneity, is a time dependent scalar field $\phi(t)$ since it has no preferred direction and is space-independent. Moreover, in a FLRW Universe, the energy density and pressure of a scalar field are given by

$$\rho = \frac{\dot{\phi}^2}{2} + V(\phi), \quad p = \frac{\dot{\phi}^2}{2} - V(\phi). \tag{22}$$

As a consequence, in a situation where the potential energy dominates over the kinetic energy, namely when the field moves slowly or, equivalently, when the potential is flat, one obtains a negative pressure and, hence, inflation. The field which drives inflation is called the "inflaton".

Let also notice that, when $V(\phi) \gg \dot{\phi}^2$, the equation of state is $p \simeq -\rho$ which, using the conservation equation, immediately implies that the energy density, and therefore the Hubble parameter H, is almost a constant. The Friedmann equation then leads to a scale factor $a(t) \propto \mathrm{e}^{Ht}$. In other words inflation is also a phase of quasi-exponential expansion. Moreover, using the expressions established above, one also has

$$|\Omega_{\mathcal{K}}| \equiv \left|\frac{\rho_{\mathrm{curv}}}{\rho_{\mathrm{cri}}}\right| = \frac{|\mathcal{K}|}{a^2H^2}, \tag{23}$$

and we see that $\Omega_{\mathcal{K}}$ goes exponentially to zero during inflation. We therefore expect to measure a vanishing spatial curvature: this is a first generic prediction of inflation and we will see in Sect. 5 that it is good agreement with the most recent cosmological observations.

As mentioned before, inflation is a high energy phenomenon and, therefore, a concrete implementation necessarily rests on high energy physics. In the modern view, the micro-physics of inflation should therefore be described by an effective field theory characterized by a cutoff Λ. If the gravitational sector is described by GR, which itself is viewed as an effective theory with a cutoff at the Planck scale, then $\Lambda < M_{\mathrm{Pl}}$. On the other hand, we know that the Hubble parameter during the part of the inflationary phase we have observationally access to can be as large as

10^{15}GeV and this suggests $\Lambda > 10^{15}$GeV. Clearly, at those energy scales, particle physics remains speculative and this is the reason why there is currently a plethora of different inflationary scenarios. A priori, without any further theoretical guidance, the effective action can therefore be written as

$$S = \int \mathrm{d}^4x\sqrt{-g}\Bigg[M_{\rm Pl}^2\Lambda_{\rm B} + \frac{M_{\rm Pl}^2}{2}R + aR^2 + bR_{\mu\nu}R^{\mu\nu} + \frac{c}{M_{\rm Pl}^2}R^3 + \cdots$$
$$-\frac{1}{2}\sum_i g^{\mu\nu}\partial_\mu\phi_i\partial_\nu\phi_i - V(\phi_1,\cdots,\phi_n) + \sum_i d_i\frac{\mathcal{O}_i}{\Lambda^{n_i-4}}\Bigg]$$
$$+S_{\rm int}(\phi_1,\cdots,\phi_n,A_\mu,\Psi) + \cdots \tag{24}$$

In the above equation, the first line represents the effective Lagrangian for gravity ($\Lambda_{\rm B}$ is the cosmological constant). In practice, we will mainly work with the Einstein-Hilbert term only. The second line represents the contribution of matter. We assumed that several scalar field are present (a priori, there is no reason to assume that only one field plays a role). The two first terms are the canonical Lagrangian while $\mathcal{O}_i$ represents a higher order operator of dimension $n_i > 4$, the amplitude of which is determined by the coefficient d_i. Those corrections can modify the potential but also the (standard) kinetic term (Chialva and Mazumdar 2014). The last term encodes the interaction between the inflaton fields and the rest of the world, i.e. the gauge fields A_μ and the fermions Ψ. The dots stand for the rest of the terms such as the Lagrangians of A_μ, of Ψ, the corresponding higher order operators etc .... Notice that the above description is not completely general. For instance, suppose that the action of the inflation field is of the Dirac-Born-Infeld (DBI) type (Alishahiha et al. 2004), namely

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{M_{\rm Pl}^2}{2}R - T(\phi)\sqrt{1-2\frac{X}{T(\phi)}} + T(\phi) - V(\phi)\right], \tag{25}$$

where $X \equiv -1/2g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$. An expansion in X gives

$$S = \int \mathrm{d}^4x\sqrt{-g}\left(\frac{M_{\rm Pl}^2}{2}R - X + V + \frac{X^2}{2T(\phi)} + \cdots\right), \tag{26}$$

and we see that the higher order terms are not suppressed by a fixed cutoff Λ but by $T(\phi)$. In this case, in some sense, the cutoff has become field dependent. As a consequence, the canonical Lagrangian $X - V$ is not necessarily always the first term of the series and it makes sense to also consider more complicated cases, even at "leading order".

Another, but related, question is whether the higher order operators can be neglected during inflation. Firstly, it is necessary that the field excursion $\Delta\phi$ be small in comparison with the cutoff scale, i.e. $\Delta\phi < \Lambda$. Whether this is the case or not depends on the model. Second, the tree level potential V can receive corrections

that can be difficult to control. For instance, if there is a mass term, then typically the mass m becomes

$$m^2 \to m^2 + gM^2 \ln\left(\frac{\Lambda}{\mu}\right), \tag{27}$$

where μ is a renormalization scale, M the mass of a heavy field and g the coupling between ϕ and the heavy field. If $M > \Lambda$ then one has $m > H$ since we have $\Lambda > H$. This means that the potential is no longer flat enough to support inflation, an embarrassing problem indeed! Ways out consist in assuming that the coupling g is small or, more convincingly, that symmetries forbid this type of corrections.

Finally, let us say a few words about the interaction term. Usually, it is considered to be negligible during the slow-roll phase. If this is not the case, it leads to warm inflation (Berera 1995; Yokoyama and Linde 1999; Berera et al. 2009). Even if it does not play a role during the accelerated phase, the interaction term is of fundamental importance for inflation since it is responsible for the reheating stage, that is to say it explains how inflation is smoothly connected to the standard hot Big Bang epoch.

We see that, using theoretical considerations only, it is difficult to restrict the Lagrangian of inflation to a simple form. But, in fact, the point is that the CMB Planck data can do the job and can constrain the Lagrangian (24). For instance, we will see in the following that the perturbations are adiabatic (at least for the moment; this could of course very well change when more accurate data are collected) and this supports the idea that only one scalar field is at play during inflation. Moreover, we will also show that Non-Gaussianities have been measured to be compatible with zero and this supports the fact that the kinetic term must be standard. We are therefore led to consider that inflation is described by the simplest scenario, namely single-field slow-roll with a standard kinetic term. It is important to emphasize that we are pushed to this class of models, which is clearly easier to analyze than Eq. (24), not because we want to simplify the scenario but because this is what the CMB data suggest. In this framework, the inflationary Lagrangian can be written as

$$\mathcal{L} = -\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi) + \mathcal{L}_{\rm int}(\phi, A_\mu, \Psi). \tag{28}$$

In the following, we will ignore the interaction term during the accelerated phase and will consider its effect only at the end of inflation (the "reheating" phase). We see that we are left with a model that contains only one arbitrary function, the potential $V(\phi)$. Therefore, what remains to be done in order to completely characterize inflation is to constrain this a priori arbitrary function with cosmological data. This line of research has played a dominant role in the recent years.

Let us now describe the slow-roll formalism which is used in practice to derive the inflationary predictions of the models mentioned above. As already remarked previously, one can distinguish two different phases of evolution: the slow-roll phase and the reheating phase. In principle, once $V(\phi)$ and $\mathcal{L}_{\rm int}(\phi, A_\mu, \Psi)$ are known,

the model is completely specified. In practice, however, one proceeds in a slightly different way. The function $V(\phi)$ is considered to be relevant for a limited range of field values only, corresponding to our observable window. Then, the evolution of the system is controlled by the Friedmann and Klein-Gordon equations, namely

$$H^2 = \frac{1}{3M_{\rm Pl}^2}\left[\frac{\dot{\phi}^2}{2} + V(\phi)\right], \tag{29}$$

$$\ddot{\phi} + 3H\dot{\phi} + V_\phi = 0, \tag{30}$$

where we remind that $H \equiv \dot{a}/a$ denotes the Hubble parameter and where a subscript ϕ means a derivative with respect to the inflaton field. It is also interesting to introduce the Hubble flow functions ϵ_n defined by Schwarz et al. (2001) and Leach et al. (2002)

$$\epsilon_{n+1} \equiv \frac{{\rm d}\ln|\epsilon_n|}{{\rm d}N}, \quad n \geq 0, \tag{31}$$

where $\epsilon_0 \equiv H_{\rm ini}/H$ starts the hierarchy and $N \equiv \ln(a/a_{\rm ini})$ is the number of e-folds. These parameters provide useful information about the inflationary dynamics. For instance, the first slow-roll parameter can be expressed as

$$\epsilon_1 = -\frac{\dot{H}}{H^2} = 1 - \frac{\ddot{a}}{aH^2}, \tag{32}$$

and, therefore, inflation ($\ddot{a} > 0$) occurs if $\epsilon_1 < 1$. In fact, since the parameters ϵ_n are defined in terms of H and since H is determined once $V(\phi)$ is known, see Eqs. (29) and (30), it follows that one can also express them in terms of the potential. For instance, ϵ_1 is given by

$$\epsilon_1 = \frac{3\dot{\phi}^2}{2}\frac{1}{\dot{\phi}^2/2 + V(\phi)}. \tag{33}$$

In fact, it is not sufficient to have $\epsilon_1 < 1$ but one also needs $\epsilon_1 \ll 1$. Indeed, from the above expression, we see that this corresponds to a situation where $\dot{\phi}^2/2 \ll V(\phi)$ or, in other words, to a situation where the potential is very flat since the field must roll very slowly. We just recover the case considered in the previous section. In this situation, referred to as the slow-roll approximation, one has in fact $\epsilon_n \ll 1$ for any n. If this is the case, then the Hubble flow functions can be expressed as (Liddle et al. 1994)

$$\epsilon_1 \simeq \frac{M_{\rm Pl}^2}{2}\left(\frac{V_\phi}{V}\right)^2, \tag{34}$$

$$\epsilon_2 \simeq 2M_{\rm Pl}^2\left[\left(\frac{V_\phi}{V}\right)^2-\frac{V_{\phi\phi}}{V}\right], \tag{35}$$

$$\epsilon_2\epsilon_3 \simeq 2M_{\rm Pl}^4\left[\frac{V_{\phi\phi\phi}V_\phi}{V^2}-3\frac{V_{\phi\phi}}{V}\left(\frac{V_\phi}{V}\right)^2+2\left(\frac{V_\phi}{V}\right)^4\right]. \tag{36}$$

The slow-roll approximation allows us to simplify the equations of motion and to analytically integrate the inflaton trajectory. Indeed, Eqs. (29) and (30), which control the evolution of ϕ, can be rewritten as

$$H^2 = \frac{V}{M_{\rm Pl}^2(3-\epsilon_1)}, \tag{37}$$

$$\left(1+\frac{\epsilon_2}{6-2\epsilon_1}\right)\frac{{\rm d}\phi}{{\rm d}N} = -M_{\rm Pl}^2\frac{{\rm d}\ln V}{{\rm d}\phi}. \tag{38}$$

As a consequence, in the slow-roll approximation, one has $H^2 \simeq V/(3M_{\rm Pl}^2)$ and ${\rm d}\phi/{\rm d}N \simeq -M_{\rm Pl}^2{\rm d}\ln V/{\rm d}\phi$, from which one obtains

$$N-N_{\rm ini} = -\frac{1}{M_{\rm Pl}^2}\int_{\phi_{\rm ini}}^{\phi}\frac{V(\chi)}{V_\chi(\chi)}\,{\rm d}\chi\,, \tag{39}$$

$\phi_{\rm ini}$ being the initial vacuum expectation value of the field. It is clear from the above considerations that the inflaton dynamics is entirely determined once the potential $V(\phi)$ has been specified. Since, in addition, the function $V(\phi)$ allows us to make the connection with high energy physics, it appears as a natural tool to parameterize inflation.

3.2 *Other Parameterizations?*

Recently, other parameterizations of inflation have been considered. The motivation of these works was to establish a general framework in order to characterize what the generic or typical predictions of cosmic inflation are. In this section, we discuss them and show that, in fact, they all boil down to choosing a specific potential.

The first alternative parametrization that we discuss is the so-called "horizon-flow approach" (Hoffman and Turner 2001; Kinney 2002; Liddle 2003; Ramirez and Liddle 2005; Chongchitnan and Efstathiou 2005). It has been recently discussed in detail in Vennin (2014). Let us define a new set of parameters ${}^\ell\lambda$ given by

$${}^\ell\lambda = \left(2M_{\rm Pl}^2\right)^\ell\frac{(H')^{\ell-1}}{H^\ell}\frac{{\rm d}^{\ell+1}H}{{\rm d}\phi^{\ell+1}}. \tag{40}$$

Of course, this new definition does not bring any new information. The new parameters can be expressed in terms of the previous ones, for instance ${}^1\lambda = \epsilon_1 - \epsilon_2/2$, ${}^2\lambda = \epsilon_1^2 - 3\epsilon_1\epsilon_2/2 + \epsilon_2\epsilon_3/2$, etc .... It only shows that, if the ϵ_n's are all of the same order in slow-roll, the ${}^\ell\lambda$ are of increasing order. Then, the simple equation, see Eq. (31)

$$\frac{\mathrm{d}\epsilon_n}{\mathrm{d}N} = \epsilon_n\epsilon_{n+1} \tag{41}$$

is replaced with

$$\frac{\mathrm{d}\epsilon_1}{\mathrm{d}N} = \epsilon_1\epsilon_2, \tag{42}$$

$$\frac{\mathrm{d}\epsilon_2}{\mathrm{d}N} = 2^2\lambda - 2\epsilon_1^2 - 3\epsilon_1\epsilon_2, \tag{43}$$

$$\frac{\mathrm{d}^\ell\lambda}{\mathrm{d}N} = -{}^{\ell+1}\lambda - {}^\ell\lambda\left(\frac{\ell-1}{2}\epsilon_2 - \epsilon_1\right). \tag{44}$$

The idea is now to truncate this hierarchy at some order M, i.e. to assume that ${}^\ell\lambda = 0$ for $\ell > M$, maybe motivated by the fact that higher order equations deal with higher order slow-roll parameters and are thus, in some sense, negligible. Then, this finite set of equations (in practice, the case $M = 5$ has been considered) is numerically integrated many times with different initial conditions (Hoffman and Turner 2001). In this way, one obtains different values of the slow-roll parameters at Hubble radius crossing and, since the observables such as the spectral index n_S or the tensor-to-scalar ratio r can be expressed in terms of these parameter (see below), different inflationary predictions. The next step consists in searching systematic patterns in these predictions which, as a consequence, would be considered as "typical" of inflation. In particular, it has been claimed that the different predictions for n_S and r obtained in this way cluster around the relation (Kinney 2002; Ramirez and Liddle 2005)

$$r_{16} \simeq \frac{1}{3}(1 - n_\mathrm{S}), \tag{45}$$

where $r_{16} \equiv r/16$. The above equation is then viewed as a generic prediction of inflation, obtained without the need to specify a particular potential $V(\phi)$.

However, the above claim is not correct (Vennin 2014). Indeed, truncating the hierarchy at order M clearly means that one assumes, see Eq. (40),

$$\frac{\mathrm{d}^{M+2}H}{\mathrm{d}\phi^{M+2}} = 0, \tag{46}$$

an equation which can be easily integrated (!) and leads to Liddle (2003)

$$H(\phi) = H_0 \left[1 + \sum_{i=1}^{M+1} A_i \left(\frac{\phi}{M_{\rm Pl}} \right)^i \right]. \tag{47}$$

Then, from this expression of the Hubble parameter, one can easily calculate the corresponding inflationary potential and one obtains

$$V(\phi) = 3M_{\rm Pl}^2 H^2(\phi) - 2M_{\rm Pl}^4 H'(\phi) \tag{48}$$

$$= 3M_{\rm Pl}^2 H_0^2 \left[1 + \sum_{i=1}^{M+1} A_i \left(\frac{\phi}{M_{\rm Pl}} \right)^i \right]^2 - 2M_{\rm Pl}^3 H_0 \sum_{i=1}^{M+1} i A_i \left(\frac{\phi}{M_{\rm Pl}} \right)^{i-1}. \tag{49}$$

The whole procedure is therefore nothing but a particular choice of a potential $V(\phi)$ depending on $M + 1$ parameters, A_i. Moreover, Vennin (2014) has shown that the "mysterious" coefficient $1/3$ in Eq. (45) can be easily recovered if one carries out a standard slow-roll analysis of the potential (48). We conclude that this approach is not generic at all and only consists in studying a very particular potential.

More recently, it has also been argued that, rather than choosing a potential $V(\phi)$, it is more generic to choose the equation of state during inflation, see Mukhanov (2013, 2015) and Binetruy et al. (2015). So, in practice, what is done is an educated guess for $w(N) = p/\rho$. Notice that, since

$$1 + w(N) = \frac{2}{3}\epsilon_1(N), \tag{50}$$

this is also equivalent to choosing a particular function $\epsilon_1(N)$, which is the strategy followed in Roest (2014), Garcia-Bellido and Roest (2014) and Binetruy et al. (2015). Concretely, one takes

$$1 + w(N) = \frac{\beta}{(N_{\rm end} - N)^\alpha}, \tag{51}$$

where α and β are two free and positive parameters and $N_{\rm end}$ is the number of e-folds at the end of inflation. However, again, this choice is in fact a choice of $V(\phi)$. Indeed, the slow-roll trajectory (39), $\mathrm{d}N = -V\mathrm{d}\phi/(M_{\rm Pl}^2 V')$ can be re-written as

$$M_{\rm Pl}^2 \frac{\mathrm{d}}{\mathrm{d}N} (\ln V) \simeq -\left(\frac{\mathrm{d}\phi}{\mathrm{d}N} \right)^2, \tag{52}$$

and, from the exact formula

$$\epsilon_1 = \frac{1}{2M_{\rm Pl}^2} \left(\frac{\mathrm{d}\phi}{\mathrm{d}N} \right)^2, \tag{53}$$

one obtains the following system of equations

$$\left(\frac{\mathrm{d}\phi}{\mathrm{d}N}\right)^2 = 3M_{\mathrm{Pl}}^2\left[1+w(N)\right], \tag{54}$$

$$\frac{\mathrm{d}}{\mathrm{d}N}\left(\ln V\right) = -3\left[1+w(N)\right]. \tag{55}$$

When the above set of equations is solved one obtains $\phi(N)$ and $V(N)$ and, eventually eliminating N, the function $V(\phi)$. We conclude that giving $w(N)$ and/or $\epsilon_1(N)$ is not a new generic parametrization but just a particular choice of a potential. In order to illustrate this point, let us see how it works in practice for the case of Eq. (51). The trajectory, given by Eq. (54), can be written as

$$\frac{\phi}{M_{\mathrm{Pl}}} = C_1 \pm \sqrt{3\beta}\frac{2}{\alpha-2}\left(N_{\mathrm{end}}-N\right)^{(2-\alpha)/2}, \tag{56}$$

where C_1 is an integration constant. For the potential, the integration of Eq. (55) is also straightforward and one finds

$$\ln V = C_2 + \frac{3\beta}{1-\alpha}\left(N_{\mathrm{end}}-N\right)^{1-\alpha}, \tag{57}$$

where C_2 is another integration constant. Then, from Eq. (56), one arrives at

$$N_{\mathrm{end}}-N = \left[\pm\frac{\alpha-2}{2\sqrt{3\beta}}\left(\frac{\phi}{M_{\mathrm{Pl}}}-C_1\right)\right]^{2/(2-\alpha)}, \tag{58}$$

and, inserting this result in Eq. (57), one obtains

$$\ln V = C_2 + \frac{3\beta}{1-\alpha}\left[\pm\frac{\alpha-2}{2\sqrt{3\beta}}\left(\frac{\phi}{M_{\mathrm{Pl}}}-C_1\right)\right]^{2(1-\alpha)/(2-\alpha)}. \tag{59}$$

This shows that Eq. (51) is, in the slow-roll approximation, completely equivalent to the choice

$$V(\phi) = M^4\mathrm{e}^{\delta\phi^\gamma}, \tag{60}$$

where δ and γ are constants (Mukhanov 2015). This potential is almost identical to Logamediate inflation, LMI in the terminology of Martin et al. (2014a,b), $V(\phi) = M^4x^\alpha\mathrm{e}^{\delta x^\gamma}$, with $x = (\phi-\phi_0)/M_{\mathrm{Pl}}$, in the case where $\alpha = 0$. This model was studied in detail in Martin et al. (2014a,b). The only difference is that, for LMI, one has $\alpha = 4(1-\gamma)$ implying $\gamma = 1$ when $\alpha = 0$, which is not the case here (i.e. $\alpha = 0$ but γ is still free).

We conclude that all the so-called "alternative" parameterizations of inflation considered so far are in fact strictly equivalent to specifying a potential. Claiming that it is either new or different or better seems definitively far-fetched. In addition, discussing inflation in terms of $V(\phi)$ has the advantage to make the link with high energy physics explicit. For these reasons, we conclude that working in terms of $V(\phi)$ and scanning the inflationary landscape by considering all possible models seems to be the most efficient method to learn about inflation.

3.3 Parametrization of Reheating

Let us now consider the end of inflation, namely the reheating phase, and how one can describe it. When $\epsilon_1 = 1$, the potential is no longer flat enough to support an accelerated phase and inflation stops. Usually, this happens in the vicinity of the ground state (concretely, the minimum of the potential). At this time, the inflaton field starts oscillating and decaying. Then, these decay products thermalize (Podolsky et al. 2006) and the radiation dominated epoch of the hot Big Bang phase commences. The micro-physics of reheating is described by the term $\mathcal{L}_{\rm int}(\phi, A_\mu, \Psi)$ in Eq. (28). But, in fact, in order to parameterize reheating, we do not need to have such a detailed description. Indeed, as we will see in the following, the inflationary observational predictions are expressed in terms of $\epsilon_{n*} \equiv \epsilon_n(\phi_*)$, where ϕ_* is the value of ϕ when the pivot scale $k_{\rm P}$ leaves the Hubble radius during inflation (the pivot scale is conveniently chosen in the middle of the observable window). Since, in the slow-roll approximation, we know the trajectory $\phi = \phi(N)$, we just need to determine N_* such that $\phi_* = \phi(N_*)$. This can be done as follows. The physical pivot scale during inflation is given by

$$\frac{k_{\rm P}}{a(N)} = \frac{k_{\rm P}}{a_{\rm now}}\frac{a_{\rm now}}{a_{\rm reh}}\frac{a_{\rm reh}}{a_{\rm end}}\frac{a_{\rm end}}{a(N)} = \frac{k_{\rm P}}{a_{\rm now}}\frac{a_{\rm now}}{a_{\rm reh}}\frac{a_{\rm reh}}{a_{\rm end}}{\rm e}^{N_{\rm end}-N}, \tag{61}$$

where $a_{\rm end}$ denotes the scale factor at the end of inflation and $a_{\rm reh}$ the scale factor at the end of reheating. In the above expression, $k_{\rm P}/a_{\rm now}$ is known and, concretely, we take $k_{\rm P}/a_{\rm now} = 0.05{\rm Mpc}^{-1}$. The quantity $a_{\rm now}/a_{\rm reh}$ is also known since it only involves the standard thermal history of the Universe. On the other hand, the ratio $a_{\rm reh}/a_{\rm end}$ depends on what happens during reheating and this is precisely the reason why the inflationary predictions are sensitive to this phase of evolution. To go further, we write the above equation at the time $N = N_*$. Since, by definition, $k_{\rm P}/a(N_*) = H(N_*)$, Eq. (61) becomes

$$H(N_*) = \frac{1}{M_{\rm Pl}}\sqrt{\frac{V(N_*)}{3-\epsilon_1(N_*)}} = \frac{k_{\rm P}}{a_{\rm now}}\frac{a_{\rm now}}{a_{\rm reh}}\frac{a_{\rm reh}}{a_{\rm end}}{\rm e}^{N_{\rm end}-N_*}, \tag{62}$$

the first expression being just the Friedmann equation, see Eq. (37). We see that this is a transcendental equation for N_* which, therefore, needs to be solved numerically. We also see that it depends on the potential $V(\phi)$ and, hence, on the model under consideration. Finally, in order to solve this equation, one needs to estimate the quantity $a_{\rm reh}/a_{\rm end}$. Let ρ and p be the total energy density and pressure during reheating. Notice that one can have several fluids, possibly interacting which each other. The treatment presented here is therefore completely general. Conservation of total energy density (we emphasize again that it is not necessary to assume that the energy density of each fluid is separately conserved) implies that

$$\rho\left(N\right) = \rho_{\rm end}\exp\left\{-3\int_{N_{\rm end}}^{N}\left[1+w_{\rm reh}\left(n\right)\right]{\rm d}n\right\}, \tag{63}$$

where $w_{\rm reh} \equiv p/\rho$ is the "instantaneous" equation of state during reheating. Then, let us define the mean equation of state parameter, $\overline{w}_{\rm reh}$, by

$$\overline{w}_{\rm reh} \equiv \frac{1}{\Delta N}\int_{N_{\rm end}}^{N_{\rm reh}} w_{\rm reh}(n){\rm d}n, \tag{64}$$

where $\Delta N \equiv N_{\rm reh} - N_{\rm end}$ is the total number of e-folds during reheating. It follows that

$$\rho_{\rm reh} = \rho_{\rm end}{\rm e}^{-3(1+\overline{w}_{\rm reh})\Delta N}, \tag{65}$$

and, therefore,

$$ {\rm e}^{\Delta N} = \frac{a_{\rm reh}}{a_{\rm end}} = \left(\frac{\rho_{\rm reh}}{\rho_{\rm end}}\right)^{-1/(3+3\overline{w}_{\rm reh})}. \tag{66}$$

As a consequence, the ratio $a_{\rm reh}/a_{\rm end}$ depends on two quantities only: the energy density at the end of reheating, $\rho_{\rm reh}$, and the mean equation of state during reheating, $\overline{w}_{\rm reh}$. Once a model of inflation is known, $\rho_{\rm end}$ can be calculated so this is not a new quantity (but, again, it introduces an additional dependence on the inflationary potential). Inserting Eq. (66) into the above expression (62) leads to

$$H(N_*) = \frac{1}{M_{\rm Pl}}\sqrt{\frac{V(N_*)}{3-\epsilon_1(N_*)}} = \frac{k_{\rm P}}{a_{\rm now}}\frac{a_{\rm now}}{a_{\rm reh}}\left(\frac{\rho_{\rm reh}}{\rho_{\rm end}}\right)^{-1/(3+3\overline{w}_{\rm reh})}{\rm e}^{N_{\rm end}-N_*}. \tag{67}$$

The above formula still contains $a_{\rm reh}$, a quantity that we would like to eliminate from the final expression. For this purpose, we write $a_{\rm now}/a_{\rm reh}$ as $a_{\rm now}/a_{\rm eq}\times a_{\rm eq}/a_{\rm reh}$, where $a_{\rm eq}$ is the scale factor at matter-radiation equality. Then, we use the fact that,

during the radiation dominated era, $a \propto \rho^{-1/4}$, to write

$$H(N_*) = \frac{1}{M_{\rm Pl}}\sqrt{\frac{V(N_*)}{3-\epsilon_1(N_*)}} = \frac{k_{\rm P}}{a_{\rm now}}\frac{a_{\rm now}}{a_{\rm eq}}\left(\frac{\rho_{\rm reh}}{\rho_{\rm eq}}\right)^{1/4}\left(\frac{\rho_{\rm reh}}{\rho_{\rm end}}\right)^{-1/(3+3\overline{w}_{\rm reh})} {\rm e}^{N_{\rm end}-N_*} \tag{68}$$

$$= \frac{k_{\rm P}}{a_{\rm now}}\frac{a_{\rm now}}{a_{\rm eq}}\frac{M_{\rm Pl}}{\rho_{\rm eq}^{1/4}}\frac{\rho_{\rm end}^{1/2}}{M_{\rm Pl}^2}\frac{M_{\rm Pl}}{\rho_{\rm end}^{1/4}}\left(\frac{\rho_{\rm reh}}{\rho_{\rm end}}\right)^{1/4-1/(3+3\overline{w}_{\rm reh})} {\rm e}^{N_{\rm end}-N_*}. \tag{69}$$

Except the quantities that are known from standard cosmology (since they only depend on post-inflationary physics), such as $a_{\rm now}/a_{\rm eq} \times M_{\rm Pl}/\rho_{\rm eq}^{1/4}$, we see that this equation singles out the following combination (by definition, the "reheating" parameter) (Martin and Ringeval 2006, 2010; Lorenz et al. 2008a; Martin et al. 2015)

$$R \equiv \frac{\rho_{\rm end}^{1/4}}{M_{\rm Pl}} R_{\rm rad}, \tag{70}$$

with

$$R_{\rm rad} \equiv \left(\frac{\rho_{\rm reh}}{\rho_{\rm end}}\right)^{-1/4+1/(3+3\overline{w}_{\rm reh})} = \left(\frac{\rho_{\rm reh}}{\rho_{\rm end}}\right)^{(1-3\overline{w}_{\rm reh})/(12+12\overline{w}_{\rm reh})}. \tag{71}$$

Notice that we have a term $\rho_{\rm end}^{1/2}/M_{\rm Pl}^2$ left in Eq. (69). It is introduced because it produces a term proportional to the square root of the potential at the end of inflation and combines nicely with the $\sqrt{V_*}$ on the left hand side of Eq. (69). The arguments presented above can be easily generalized to take into account a change of relativistic degrees of freedom between the reheating epoch and today, see Martin et al. (2014a).

The reheating parameter encodes what can be learned about reheating from the CMB. In Sect. 5, we will see that the Planck data already put constraints on its value.

4 Inflationary Perturbations

In this section, we review the theory of inflationary perturbations (Mukhanov et al. 1992; Bardeen 1980; Peter 2013). This part of the inflationary scenario is very important because it allows us to use astrophysical data to put constraints on cosmic inflation. In the following, we pay special attention to the question of how one can calculate the correlation functions of the perturbations and to the concept of adiabatic and isocurvature perturbations. As will be seen in Sect. 5, these quantities carry useful information about the type of inflationary model that is realized in Nature. This section can therefore be viewed as a preparation to Sect. 5 in the sense

that we discuss in some detail the meaning of the quantities that have been measured recently by the Planck experiment.

To describe CMB anisotropies and large scale structures, one must go beyond the cosmological principle. This is a priori a technically difficult task but since the inhomogeneities are small in the early Universe, one can use a perturbative approximation which, obviously, greatly simplifies the problem. Then, the idea is to write the metric tensor as $g_{\mu\nu}(\eta,\boldsymbol{x}) = g^{\rm FLRW}_{\mu\nu}(\eta) + \delta g_{\mu\nu}(\eta,\boldsymbol{x}) + \cdots$, where $g^{\rm FLRW}_{\mu\nu}(\eta)$ represents the metric tensor of the FLRW Universe and where $\delta g_{\mu\nu}(\eta,\boldsymbol{x}) \ll g^{\rm FLRW}_{\mu\nu}(\eta)$. In fact, $\delta g_{\mu\nu}(\eta,\boldsymbol{x})$ can be expressed in terms of three types of perturbations, scalar, vector and tensor. In the context of inflation, only scalar and tensor are important. Scalar perturbations are directly coupled to the perturbed stress-energy tensor while tensor fluctuations are independent of $\delta T_{\mu\nu}$ and, in fact, are nothing but gravity waves. The equations of motion of each type of fluctuations are given by the perturbed Einstein equations, namely $\delta G_{\mu\nu} = \kappa \delta T_{\mu\nu}$.

In order to calculate the behavior of the fluctuations, we also need to specify the initial conditions. This is done by postulating that the perturbations are of quantum-mechanical origin and that, initially, their quantum state is the vacuum. This is possible because, at the beginning of inflation, the physical wavelengths of the Fourier modes of the perturbations are smaller than the Hubble radius. This means that, initially, space-time curvature is not felt and that, as a consequence, a well-motivated vacuum state can be defined.

4.1 Inflationary Two-Point Correlation Functions

Once the equations of motion have been derived and the initial conditions specified, one can determine all the statistical properties of the fluctuations, in particular their two-point correlation functions or, in Fourier space, power spectra. The scalar perturbations are curvature perturbations defined by $\zeta(\eta,\boldsymbol{x}) \equiv \Phi + 2(\mathcal{H}^{-1}\Phi' + \Phi)/(3+3w)$, with $w = p/\rho$ the equation of state during inflation and Φ being the Bardeen potential (Bardeen 1980) (not to be confused with the scalar field ϕ). As usual in a linear theory, it is convenient to work in Fourier space and, therefore, we write

$$\zeta(\eta,\boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}} \int {\rm d}\boldsymbol{k}\, \zeta_{\boldsymbol{k}}(\eta)\, {\rm e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}. \tag{72}$$

As explained before, in the framework of the theory of cosmological perturbations of quantum-mechanical origin, the source of the perturbations is the unavoidable zero-point vacuum fluctuations. As a consequence, $\zeta(\eta,\boldsymbol{x})$ must in fact be viewed as a quantum operator and can be expressed as

$$\hat{\zeta}(\eta,\boldsymbol{x}) = \int \frac{{\rm d}^3\boldsymbol{k}}{(2\pi)^{3/2}} \left[a_{\boldsymbol{k}} g_{\boldsymbol{k}}(\eta) {\rm e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} + a^{\dagger}_{\boldsymbol{k}} g^{*}_{\boldsymbol{k}}(\eta) {\rm e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\right], \tag{73}$$

where $a_{\boldsymbol{k}}$ and $a_{\boldsymbol{k}}^{\dagger}$ are respectively the annihilation and creation operators satisfying $[a_{\boldsymbol{k}}, a_{\boldsymbol{p}}^{\dagger}] = \delta^{(3)}(\boldsymbol{k}-\boldsymbol{p})$. The quantum state of the perturbations is the vacuum $|0\rangle$ which is, by definition, annihilated by the operator $a_{\boldsymbol{k}}$, namely $a_{\boldsymbol{k}}|0\rangle = 0$. The function $g_{\boldsymbol{k}}(\eta)$ is the mode function and the Fourier transform of $\zeta(\eta, \boldsymbol{x})$ is given by $\zeta_{\boldsymbol{k}}(\eta) = a_{\boldsymbol{k}} g_{\boldsymbol{k}}(\eta) + a_{-\boldsymbol{k}}^{\dagger} g_{\boldsymbol{k}}^{*}(\eta)$. This last equation leads to $\langle 0|\zeta_{\boldsymbol{k}_1}\zeta_{\boldsymbol{k}_2}|0\rangle = |g_{\boldsymbol{k}_1}|^2 \delta^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2)$. From the previous considerations, it follows that the two-point correlation function is given by

$$\langle \zeta^2(\eta, \boldsymbol{x})\rangle = \int \frac{\mathrm{d}k}{k}\mathcal{P}_{\zeta}(k) = \int \frac{\mathrm{d}k}{k}\frac{k^3}{2\pi^2}|g_{\boldsymbol{k}}|^2, \tag{74}$$

where $\mathcal{P}_{\zeta}(k)$ is, by definition, the power spectrum of scalar perturbations. An exact calculation of this power spectrum is rarely available but a perturbative expansion into the slow-roll parameters (since they are small parameters) can be done and results in

$$\frac{\mathcal{P}_{\zeta}(k)}{\mathcal{P}_{\zeta 0}(k_{\mathrm{P}})} = a_0^{(\mathrm{S})} + a_1^{(\mathrm{S})} \ln\left(\frac{k}{k_{\mathrm{P}}}\right) + \frac{a_2^{(\mathrm{S})}}{2}\ln^2\left(\frac{k}{k_{\mathrm{P}}}\right) + \ldots, \tag{75}$$

where, as already mentioned, k_{P} is the pivot scale and the overall amplitude can be written as

$$\mathcal{P}_{\zeta_0} = \frac{H_*^2}{8\pi^2\epsilon_{1*}M_{\mathrm{Pl}}^2}, \tag{76}$$

a star meaning that a quantity is evaluated at the time at which the pivot scale crossed out the Hubble radius during inflation. We see that the amplitude of the power spectrum depends on H_* but also on the first slow-roll parameter, ϵ_{1*}. The coefficients $a_i^{(\mathrm{S})}$ can be expressed in terms of the Hubble flow functions. For scalar perturbations, at second order in the slow-roll approximation, one gets (Schwarz et al. 2001; Martin and Schwarz 2003; Casadio et al. 2005a,b,c; Gong and Stewart 2001; Choe et al. 2004; Leach et al. 2002; Lorenz et al. 2008b; Martin et al. 2013; Jimenez et al. 2013)

$$\begin{aligned} a_0^{(\mathrm{S})} &= 1 - 2(C+1)\epsilon_{1*} - C\epsilon_{2*} + \left(2C^2 + 2C + \frac{\pi^2}{2} - 5\right)\epsilon_{1*}^2 \\ &+ \left(C^2 - C + \frac{7\pi^2}{12} - 7\right)\epsilon_{1*}\epsilon_{2*} + \left(\frac{1}{2}C^2 + \frac{\pi^2}{8} - 1\right)\epsilon_{2*}^2 \\ &+ \left(-\frac{1}{2}C^2 + \frac{\pi^2}{24}\right)\epsilon_{2*}\epsilon_{3*} + \cdots, \end{aligned} \tag{77}$$

$$a_1^{(\mathrm{S})} = -2\epsilon_{1*} - \epsilon_{2*} + 2(2C+1)\epsilon_{1*}^2 + (2C-1)\epsilon_{1*}\epsilon_{2*} + C\epsilon_{2*}^2 - C\epsilon_{2*}\epsilon_{3*} + \cdots, \tag{78}$$

$$a_2^{(\mathrm{S})} = 4\epsilon_{1*}^2 + 2\epsilon_{1*}\epsilon_{2*} + \epsilon_{2*}^2 - \epsilon_{2*}\epsilon_{3*} + \cdots, \tag{79}$$

$$a_3^{(\mathrm{S})} = \mathcal{O}(\epsilon_{n*}^3), \tag{80}$$

where $C \equiv \gamma_{\mathrm{E}} + \ln 2 - 2 \approx -0.7296$, γ_{E} being the Euler constant.

For tensor fluctuations, the approach is exactly similar to what we have just described. In particular, the tensor power spectrum $\mathcal{P}_h$ can be written in the same way as Eq. (75) but with a global amplitude now given by

$$\mathcal{P}_{h_0} = \frac{2H_*^2}{\pi^2 M_{\mathrm{Pl}}^2}. \tag{81}$$

This time, the amplitude only depends on the Hubble parameter during inflation. Moreover, the coefficients $a_i^{(\mathrm{T})}$ have a similar structure and can be written as

$$a_0^{(\mathrm{T})} = 1 - 2(C+1)\epsilon_{1*} + \left(2C^2 + 2C + \frac{\pi^2}{2} - 5\right)\epsilon_{1*}^2 + \left(-C^2 - 2C + \frac{\pi^2}{12} - 2\right)\epsilon_{1*}\epsilon_{2*} + \cdots, \tag{82}$$

$$a_1^{(\mathrm{T})} = -2\epsilon_{1*} + 2(2C+1)\epsilon_{1*}^2 - 2(C+1)\epsilon_{1*}\epsilon_{2*} + \cdots, \tag{83}$$

$$a_2^{(\mathrm{T})} = 4\epsilon_{1*}^2 - 2\epsilon_{1*}\epsilon_{2*} + \cdots, \tag{84}$$

$$a_3^{(\mathrm{T})} = \mathcal{O}(\epsilon_{n*}^3). \tag{85}$$

The coefficients in front of the $\ln k$ term are related to the spectral indices and, at first order in the slow-roll parameters (we will discuss them in more detail in Sect. 5), they can be expressed as

$$n_{\mathrm{S}} = 1 - 2\epsilon_1 - \epsilon_2, \quad n_{\mathrm{T}} = -2\epsilon_1, \tag{86}$$

where the first expression refers to scalar perturbations while the second is for tensor perturbations. Notice that, sometimes, the power spectrum is written as $k^{n_{\mathrm{S}}-1}$. In the context of slow-roll inflation, this is clearly not justified as it would amount to keep an infinite number of higher order terms while n_{S} has been evaluated at first order only. It is worth stressing that power-law power spectra are predictions of power-law inflation only, that is to say the inflationary model for which $V(\phi) \propto \exp(-C\phi)$ (Lucchin and Matarrese 1985). From Eqs. (76) and (81), one can also estimate the relative contribution of tensor and scalar amplitudes

$$r \equiv \frac{\mathcal{P}_h}{\mathcal{P}_\zeta} = 16\epsilon_{1*}, \tag{87}$$

which means that, since $\epsilon_{1*} \ll 1$, tensor are sub-dominant. This is of course rather unfortunate since a direct measurement of gravity wave would directly lead to the energy scale during inflation, H_*.

4.2 Inflationary Three-Point Correlation Functions

We have just derived the slow-roll inflationary two-point correlation functions but, of course, higher-order correlation functions are also interesting and the field of Non-Gaussianity has played an important role in the recent years, see Gangui et al. (1994), Gangui (1994), Wang and Kamionkowski (2000), Gangui and Martin (2000a,b) and Gangui et al. (2002) for original works on this question and Maldacena (2003), Seery and Lidsey (2005), Chen et al. (2007), Hotchkiss and Sarkar (2010), Chen (2010), Martin and Sriramkumar (2012), Hazra et al. (2012, 2013), Sreenath et al. (2013, 2015), Martin et al. (2014c), and Sreenath and Sriramkumar (2014) for later works. For a complete overview of the subject, we refer to the lecture notes by Byrnes (2014). Here, in order to be able to fully appreciate the relevance of the Planck data on Non-Gaussianities, we discuss how the three-point inflationary correlation functions can be calculated in the case of single-field slow-roll inflation with a minimal kinetic term.

For the two-point correlation, we have seen that it is convenient to work in Fourier space and to define the power spectrum. In the same way, for the three-point correlation function, we can define the bispectrum as a correlator in Fourier space, namely $\langle \zeta_{\boldsymbol{k}_1}(\eta)\,\zeta_{\boldsymbol{k}_2}(\eta)\,\zeta_{\boldsymbol{k}_3}(\eta)\rangle$. In fact, we will rather calculate the quantity $\langle \mathcal{R}_{\boldsymbol{k}_1}(\eta)\,\mathcal{R}_{\boldsymbol{k}_2}(\eta)\,\mathcal{R}_{\boldsymbol{k}_3}(\eta)\rangle$ where $\mathcal{R} \equiv -\Psi - \mathcal{H}\delta\phi^{(\mathrm{gi})}/\phi'$, $\Psi = \Phi$ (valid if a scalar field dominates the matter content of the Universe) being another Bardeen potential and $\delta\phi^{(\mathrm{gi})}$ being the gauge invariant scalar field fluctuation (Mukhanov et al. 1992). This amounts to a simple change of sign of the three-point function (and no change at the power spectrum level, namely $\mathcal{P}_\zeta = \mathcal{P}_\mathcal{R}$, because the power spectrum is quadratic in the Fourier amplitudes) since $\mathcal{R} = -\zeta$.[1] Concretely one has

$$\langle \mathcal{R}(\eta,\boldsymbol{x})\,\mathcal{R}(\eta,\boldsymbol{x})\,\mathcal{R}(\eta,\boldsymbol{x})\rangle = \int \frac{\mathrm{d}^3\boldsymbol{k}_1}{(2\pi)^{3/2}} \int \frac{\mathrm{d}^3\boldsymbol{k}_2}{(2\pi)^{3/2}} \int \frac{\mathrm{d}^3\boldsymbol{k}_3}{(2\pi)^{3/2}}\, \langle \mathcal{R}_{\boldsymbol{k}_1}(\eta)\,\mathcal{R}_{\boldsymbol{k}_2}(\eta)\,\mathcal{R}_{\boldsymbol{k}_3}(\eta)\rangle \times \mathrm{e}^{i\,(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)\cdot\boldsymbol{x}}. \tag{91}$$

[1] Indeed, the space time component of the perturbed Einstein equation reads

$$-\frac{2}{a^2}\partial_i\left(\mathcal{H}\Phi + \Phi'\right) = \kappa(\rho + p)\partial_i v^{(\mathrm{gi})}, \tag{88}$$

where, for a scalar field, $v^{(\mathrm{gi})} = -\delta\phi^{(\mathrm{gi})}/\phi'$. As a consequence

$$\Phi + \mathcal{H}^{-1}\Phi' = \frac{\kappa a^2}{2\mathcal{H}}(\rho + p)\frac{\delta\phi^{(\mathrm{gi})}}{\phi'}. \tag{89}$$

Using this last expression in the definition of ζ and the Friedmann equation $\mathcal{H}^2 = \kappa a^2\rho/3$, one obtains

$$\zeta = \Phi + \frac{2}{3}\frac{\mathcal{H}^{-1}\Phi' + \Phi}{1+w} = \Phi + \mathcal{H}\frac{\delta\phi^{(\mathrm{gi})}}{\phi'} = -\mathcal{R}, \tag{90}$$

namely the equation mentioned in the text.

In the above expression $\mathcal{R}_{\boldsymbol{k}}(\eta)$ obviously represents the Fourier transform of the curvature (scalar) perturbation $\mathcal{R}(\eta,\boldsymbol{x})$, namely

$$\mathcal{R}(\eta,\boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}}\int \mathrm{d}\boldsymbol{k}\,\mathcal{R}_{\boldsymbol{k}}(\eta)\,\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}. \tag{92}$$

As explained before, in the framework of the theory of cosmological perturbations of quantum-mechanical origin, it is an operator and it can be expressed as

$$\hat{\mathcal{R}}(\eta,\boldsymbol{x}) = \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^{3/2}}\left[a_{\boldsymbol{k}}f_{\boldsymbol{k}}(\eta)\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} + a^{\dagger}_{\boldsymbol{k}}f^{*}_{\boldsymbol{k}}(\eta)\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\right], \tag{93}$$

leading to $\langle\mathcal{R}_{\boldsymbol{k}_1}\mathcal{R}_{\boldsymbol{k}_2}\rangle = |f_{k_1}|^2\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2)$ since $\mathcal{R}_{\boldsymbol{k}} = a_{\boldsymbol{k}}f_{\boldsymbol{k}} + a^{\dagger}_{-\boldsymbol{k}}f^{*}_{\boldsymbol{k}}$. Here, the creation and annihilation operators are the same as those appearing in Eq. (73). Of course working in terms of $\mathcal{R}_{\boldsymbol{k}}(\eta)$ instead of $\zeta_{\boldsymbol{k}}(\eta)$ is both harmless and trivial since $\mathcal{R}_{\boldsymbol{k}} = -\zeta_{\boldsymbol{k}}$ and $f_{\boldsymbol{k}}(\eta) = -g_{\boldsymbol{k}}(\eta)$! We do it since many papers on Non-Gaussianities use this variable.

At this stage, it may be useful to say a few words about conventions. In this article, we are using Fourier transforms as defined in Eq. (92). Another convention, often used in the literature on Non-Gaussianities, is

$$\mathcal{R}(\eta,\boldsymbol{x}) = \frac{1}{(2\pi)^{3}}\int \mathrm{d}\boldsymbol{k}\,\bar{\mathcal{R}}_{\boldsymbol{k}}\,\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{94}$$

so that $\bar{\mathcal{R}}_{\boldsymbol{k}} = (2\pi)^{3/2}\mathcal{R}_{\boldsymbol{k}}$. This implies that $\langle\bar{\mathcal{R}}_{\boldsymbol{k}_1}\bar{\mathcal{R}}_{\boldsymbol{k}_2}\rangle = (2\pi)^3|f_{k_1}|^2\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2)$. Notice that the two-point correlation function is sometimes defined as $\langle\bar{\mathcal{R}}_{\boldsymbol{k}_1}\bar{\mathcal{R}}_{\boldsymbol{k}_2}\rangle \equiv (2\pi)^3P_{\mathcal{R}}(k_1)\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2)$ which leads to the identification $P_{\mathcal{R}}(k_1) = |f_{k_1}|^2$ [the quantity $P_{\mathcal{R}}(k_1)$ should not be confused with $\mathcal{P}_{\mathcal{R}}(k_1) = k_1^3|f_{k_1}|^2/(2\pi^2)$]. These definitions imply that $\bar{\mathcal{R}}_{\boldsymbol{k}} = (2\pi)^{3/2}\left(a_{\boldsymbol{k}}f_{\boldsymbol{k}} + a^{\dagger}_{-\boldsymbol{k}}f^{*}_{\boldsymbol{k}}\right)$ which can be rewritten as $\bar{\mathcal{R}}_{\boldsymbol{k}} = \bar{a}_{\boldsymbol{k}}f_{\boldsymbol{k}} + \bar{a}^{\dagger}_{-\boldsymbol{k}}f^{*}_{\boldsymbol{k}}$ with $\bar{a}_{\boldsymbol{k}} = (2\pi)^{3/2}a_{\boldsymbol{k}}$. In particular, since $[a_{\boldsymbol{k}},a^{\dagger}_{\boldsymbol{p}}] = \delta^{(3)}(\boldsymbol{k}-\boldsymbol{p})$, we now have $[\bar{a}_{\boldsymbol{k}},\bar{a}^{\dagger}_{\boldsymbol{p}}] = (2\pi)^3\delta^{(3)}(\boldsymbol{k}-\boldsymbol{p})$. Different conventions basically correspond to different choices for where the factors 2π appear in the equations. In principle straightforward, it can sometimes be confusing when one tries to check a result in the existing literature.

The bispectrum can be evaluated using the standard rules of quantum field theory. It is given by Maldacena (2003) and Seery and Lidsey (2005)

$$\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta)\,\mathcal{R}_{\boldsymbol{k}_2}(\eta)\,\mathcal{R}_{\boldsymbol{k}_3}(\eta)\rangle = -i\int_{\eta_{\mathrm{ini}}}^{\eta_{\mathrm{e}}}\mathrm{d}\tau\,a(\tau)\,\langle[\mathcal{R}_{\boldsymbol{k}_1}(\eta)\,\mathcal{R}_{\boldsymbol{k}_2}(\eta)\,\mathcal{R}_{\boldsymbol{k}_3}(\eta),H_{\mathrm{int}}(\tau)]\rangle\,, \tag{95}$$

where η_{ini} represents an initial time at the beginning of inflation (in practice we take $\eta_{\mathrm{ini}}\to-\infty$) and η_{e} a final time at the end of inflation when all the scales relevant

to the problem are outside the Hubble radius (in practice we take $\eta_{\rm e} \to 0$). The quantity $H_{\rm int}$ is the interaction Hamiltonian. It can be obtained from the action of the system expanded up to third order in $\mathcal{R}$, the action of the system being the Einstein-Hilbert action plus that of a scalar field (the inflaton). A now standard calculation gives (Maldacena 2003; Seery and Lidsey 2005; Chen et al. 2007; Chen 2010)

$$\mathcal{S}_3[\mathcal{R}] = M_{\rm Pl}^2 \int {\rm d}t\, {\rm d}^3\boldsymbol{x} \left[a^3\, \epsilon_1^2\, \mathcal{R}\, \dot{\mathcal{R}}^2 + a\, \epsilon_1^2\, \mathcal{R}\, (\partial\mathcal{R})^2 - 2\, a\, \epsilon_1\, \dot{\mathcal{R}}\, (\partial^i\mathcal{R})\, (\partial_i\chi) \right.$$
$$\left. + \frac{a^3}{2}\, \epsilon_1\, \dot{\epsilon}_2\, \mathcal{R}^2\, \dot{\mathcal{R}} + \frac{\epsilon_1}{2a}\, (\partial^i\mathcal{R})\, (\partial_i\chi)\, (\partial^2\chi) + \frac{\epsilon_1}{4\, a}\, (\partial^2\mathcal{R})\, (\partial\chi)^2 + \mathcal{F}\left(\frac{\delta\mathcal{L}_2}{\delta\mathcal{R}}\right) \right], \tag{96}$$

where $\delta\mathcal{L}_2/\delta\mathcal{R}$ denotes the variation of the second order action with respect to $\mathcal{R}$, and is given by

$$\frac{\delta\mathcal{L}_2}{\delta\mathcal{R}} = \dot{\Lambda} + H\,\Lambda - \epsilon_1\, \partial^2\mathcal{R}, \tag{97}$$

and the quantities Λ and χ are defined by

$$\Lambda \equiv \frac{a^2\dot{\phi}^2}{2M_{\rm Pl}^2H^2}\dot{\mathcal{R}} = a^2\epsilon_1\dot{\mathcal{R}}, \quad \chi \equiv \partial^{-2}\Lambda. \tag{98}$$

The term $\mathcal{F}(\delta\mathcal{L}_2/\delta\mathcal{R})$ introduced in Eq. (96) stands for the following complicated expression

$$\mathcal{F}\left(\frac{\delta\mathcal{L}_2}{\delta\mathcal{R}}\right) = \frac{a}{2}\epsilon_2\left(\frac{\delta\mathcal{L}_2}{\delta\mathcal{R}}\right)\mathcal{R}^2 + \frac{2a}{H}\left(\frac{\delta\mathcal{L}_2}{\delta\mathcal{R}}\right)\dot{\mathcal{R}}\mathcal{R}$$
$$+ \frac{1}{2aH}\left\{(\partial^i\mathcal{R})\,(\partial_i\chi)\left(\frac{\delta\mathcal{L}_2}{\delta\mathcal{R}}\right) + \delta^{ij}\left[\Lambda\,(\partial_i\mathcal{R}) + (\partial^2\mathcal{R})\,(\partial_i\chi)\right]\right.$$
$$\left.\times\,\partial_j\left[\partial^{-2}\left(\frac{\delta\mathcal{L}_2}{\delta\mathcal{R}}\right)\right] + \frac{\delta^{im}\delta^{jn}}{H}(\partial_i\mathcal{R})\,(\partial_j\mathcal{R})\,\partial_m\partial_n\left[\partial^{-2}\left(\frac{\delta\mathcal{L}_2}{\delta\mathcal{R}}\right)\right]\right\}. \tag{99}$$

The terms which involves $\delta\mathcal{L}_2/\delta\mathcal{R}$ can be removed by a suitable field redefinition of $\mathcal{R}$ of the following form (Maldacena 2003; Seery and Lidsey 2005; Chen et al. 2007; Chen 2010):

$$\mathcal{R} \to \mathcal{R}_n + \epsilon_2\frac{\mathcal{R}_n^2}{4}. \tag{100}$$

After this redefinition, the perturbed action (96) becomes a functional of $\mathcal{R}_n$. In the following, in order to avoid too complicated notations, we will still use $\mathcal{R}$ in place of $\mathcal{R}_n$. Then, with the redefinition (100), the interaction Hamiltonian can be expressed as

$$H_{\rm int}(\eta) = -M_{\rm Pl}^2 \int {\rm d}^3\boldsymbol{x} \left[a\,\epsilon_1^2\,\mathcal{R}\,\mathcal{R}'^2 + a\,\epsilon_1^2\,\mathcal{R}\,(\partial\mathcal{R})^2 - 2\,\epsilon_1\,\mathcal{R}'\,(\partial^i\mathcal{R})\,(\partial_i\chi) \right.$$
$$\left. + \frac{a}{2}\,\epsilon_1\,\epsilon_2'\,\mathcal{R}^2\,\mathcal{R}' + \frac{\epsilon_1}{2\,a}\,(\partial^i\mathcal{R})\,(\partial_i\chi)\,\left(\partial^2\chi\right) + \frac{\epsilon_1}{4\,a}\,\left(\partial^2\mathcal{R}\right)\,(\partial\chi)^2 \right]. \quad (101)$$

where we remind that a prime means a derivative with respect to conformal time. The first three terms are second order in the slow-roll parameters while the three last ones are third order. As a consequence, already at this stage, we see that the bispectrum will be a small quantity. Since we now know the interaction Hamiltonian we can insert its expression in Eq. (95) in order to derive the bispectrum explicitly. One finds that

$$\langle \mathcal{R}_{\boldsymbol{k}_1}(\eta_{\rm e})\,\mathcal{R}_{\boldsymbol{k}_2}(\eta_{\rm e})\,\mathcal{R}_{\boldsymbol{k}_3}(\eta_{\rm e})\rangle = \frac{(2\,\pi)^3}{(2\pi)^{9/2}} M_{\rm Pl}^2 \sum_{C=1}^{6} \left[f_{\boldsymbol{k}_1}(\eta_{\rm e}) f_{\boldsymbol{k}_2}(\eta_{\rm e}) f_{\boldsymbol{k}_3}(\eta_{\rm e}) \mathcal{G}_C(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) \right.$$
$$\left. + f_{\boldsymbol{k}_1}^*(\eta_{\rm e}) f_{\boldsymbol{k}_2}^*(\eta_{\rm e}) f_{\boldsymbol{k}_3}^*(\eta_{\rm e}) \mathcal{G}_C^*(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) \right] \delta^{(3)}\left(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3\right), \quad (102)$$

where the delta function ensures momentum conservation. Written in this way, the correlator is obviously real. In the above expression, the term $\mathcal{G}_C(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3)$ with $C = (1,6)$ correspond to the six terms in the interaction Hamiltonian (101) (the six "vertices"), and are explicitly given by Maldacena (2003)

$$\mathcal{G}_1(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = 2i \int_{\eta_{\rm ini}}^{\eta_{\rm e}} {\rm d}\tau\, a^2\,\epsilon_1^2\,\left(f_{\boldsymbol{k}_1}^* f_{\boldsymbol{k}_2}'^* f_{\boldsymbol{k}_3}'^* + \text{two permutations}\right), \quad (103)$$

$$\mathcal{G}_2(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = -2i \int_{\eta_{\rm ini}}^{\eta_{\rm e}} {\rm d}\tau a^2\,\epsilon_1^2 f_{\boldsymbol{k}_1}^* f_{\boldsymbol{k}_2}^* f_{\boldsymbol{k}_3}^*\,\left(\boldsymbol{k}_1\cdot\boldsymbol{k}_2 + \text{two perms}\right), \quad (104)$$

$$\mathcal{G}_3(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = -2i \int_{\eta_{\rm ini}}^{\eta_{\rm e}} {\rm d}\tau\; a^2\,\epsilon_1^2 \left[f_{\boldsymbol{k}_1}^* f_{\boldsymbol{k}_2}'^* f_{\boldsymbol{k}_3}'^* \left(\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_2^2}\right) \right. \quad (105)$$
$$\left. + \text{five permutations} \right], \quad (106)$$

$$\mathcal{G}_4(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = i \int_{\eta_{\rm ini}}^{\eta_{\rm e}} {\rm d}\tau\; a^2\,\epsilon_1\,\epsilon_2'\,\left(f_{\boldsymbol{k}_1}^* f_{\boldsymbol{k}_2}^* f_{\boldsymbol{k}_3}'^* + \text{two permutations}\right), \quad (107)$$

$$\mathcal{G}_5(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = \frac{i}{2} \int_{\eta_{\rm ini}}^{\eta_{\rm e}} {\rm d}\tau\; a^2\,\epsilon_1^3 \left[f_{\boldsymbol{k}_1}^* f_{\boldsymbol{k}_2}'^* f_{\boldsymbol{k}_3}'^* \left(\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_2^2}\right) \right. \quad (108)$$
$$\left. + \text{five permutations} \right], \quad (109)$$

$$\mathcal{G}_6(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = \frac{i}{2}\int_{\eta_{\rm ini}}^{\eta_{\rm e}} {\rm d}\tau\; a^2\,\epsilon_1^3\left[f_{\boldsymbol{k}_1}^{*} f_{\boldsymbol{k}_2}'^{*} f_{\boldsymbol{k}_3}'^{*}\left(\frac{k_1^2}{k_2^2\,k_3^2}\right)(\boldsymbol{k}_2\cdot\boldsymbol{k}_3)\right.$$
$$\left. + \text{two permutations}\right]. \tag{110}$$

Actually, an additional seventh term arises due to the field redefinition (100), and its contribution to the three point correlation function is found to be

$$\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta_{\rm e})\,\mathcal{R}_{\boldsymbol{k}_2}(\eta_{\rm e})\,\mathcal{R}_{\boldsymbol{k}_3}(\eta_{\rm e})\rangle^{(7)} = \frac{(2\,\pi)^3}{(2\pi)^{9/2}}\frac{\epsilon_2}{2}\left(|f_{k_2}|^2\,|f_{k_3}|^2 + \text{two permutations}\right)$$
$$\times\delta^{(3)}\left(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3\right). \tag{111}$$

The other terms in Eq. (99) do not contribute because they all contain a derivative (time derivative and/or space derivative) and, at the end of inflation, on super Hubble scales, $\zeta = -\mathcal{R}$ is constant.

In order to calculate each of the above terms, one obviously needs to know the mode function $f_{\boldsymbol{k}}$. Since we evaluate the bispectrum at leading order in slow roll, it is in fact sufficient to use the de Sitter mode function, namely $f_{\boldsymbol{k}} = iH(1+ik\eta){\rm e}^{-ik\eta}/(2M_{\rm Pl}\sqrt{k^3\epsilon_1})$ (which is properly normalized). Moreover, we only need to calculate the first three terms and $\mathcal{G}_7(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3)$, the other contributions being of higher orders in slow-roll. In order to illustrate how the calculation proceeds, let us explain in detail how $\mathcal{G}_2(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3)$ can be calculated (this term is easier than the others since we do not have to use the derivative of the mode function). Inserting the de Sitter mode function into Eq. (104), one obtains

$$\mathcal{G}_2(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = -2i\frac{(-iH)^3}{8M_{\rm Pl}^3\sqrt{\epsilon_{1*}^3k_1^3k_2^3k_3^3}}\frac{1}{H^2}\epsilon_{1*}^2\left(\boldsymbol{k}_1\cdot\boldsymbol{k}_2 + \text{two permutations}\right)$$
$$\times\int_{\eta_{\rm ini}}^{\eta_{\rm e}}\frac{{\rm d}\tau}{\tau^2}{\rm e}^{ik_{\rm T}\tau}(1-ik_1\tau)(1-ik_2\tau)(1-ik_3\tau), \tag{112}$$
$$= -2i\frac{(-iH)^3}{8M_{\rm Pl}^3\sqrt{\epsilon_{1*}^3k_1^3k_2^3k_3^3}}\frac{1}{H^2}\epsilon_{1*}^2\left(\boldsymbol{k}_1\cdot\boldsymbol{k}_2 + \text{two permutations}\right)$$
$$\times\int_{\eta_{\rm ini}}^{\eta_{\rm e}}\frac{{\rm d}\tau}{\tau^2}\left[1-ik_{\rm T}\tau-(k_1k_2+k_2k_3+k_1k_3)\,\tau^2+ik_1k_2k_3\tau^3\right]{\rm e}^{ik_{\rm T}\tau}, \tag{113}$$

where $k_{\rm T}\equiv k_1+k_2+k_3$ is the "total" wave-number. This expression is made of four integrals that we need to calculate. The first and the fourth ones can be integrated

by parts and the third one can be directly performed. This leads to

$$\mathcal{G}_2(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = -2i\frac{(-iH)^3}{8M_{\rm Pl}^3\sqrt{\epsilon_{1*}^3k_1^3k_2^3k_3^3}}\frac{1}{H^2}\epsilon_{1*}^2\,(\boldsymbol{k}_1\cdot\boldsymbol{k}_2 + \text{two permutations})$$
$$\times\left[\frac{-1}{\tau}\mathrm{e}^{ik_{\rm T}\tau}\Big|_{\eta_{\rm ini}}^{\eta_{\rm e}} + ik_{\rm T}\int_{\eta_{\rm ini}}^{\eta_{\rm e}}\frac{\mathrm{e}^{ik_{\rm T}\tau}}{\tau}\mathrm{d}\tau - ik_{\rm T}\int_{\eta_{\rm ini}}^{\eta_{\rm e}}\frac{\mathrm{e}^{ik_{\rm T}\tau}}{\tau}\mathrm{d}\tau\right.$$
$$\left. - (k_1k_2+k_2k_3+k_1k_3)\,\frac{\mathrm{e}^{ik_{\rm T}\tau}}{ik_{\rm T}}\Big|_{\eta_{\rm ini}}^{\eta_{\rm e}} + ik_1k_2k_3\left(\frac{\tau\mathrm{e}^{ik_{\rm T}\tau}}{ik_{\rm T}} + \frac{\mathrm{e}^{ik_{\rm T}\tau}}{k_{\rm T}^2}\Big|_{\eta_{\rm ini}}^{\eta_{\rm e}}\right]\right., \tag{114}$$

and we see that the second integral exactly cancels the term arising from the integration by parts of the first integral. In principle, at this stage, it is sufficient to take $\eta_{\rm ini} = -\infty$ in the above expression in order to get the final result. But, obviously, the result would be ill-defined. So what is done is to slightly rotate the integration path in the complex plane and replace $\eta_{\rm ini}$ with $-\infty(1-i\delta)$ where δ is a small parameter. This produces a term $\mathrm{e}^{-ik_{\rm T}\infty-k_{\rm T}\delta\infty}$ which, in fact, kills all terms proportional to $\mathrm{e}^{ik_{\rm T}\eta_{\rm ini}}$. It is worth noticing that this should not be viewed as an arbitrary technical trick but as the standard method to properly identify the correct vacuum state (Peskin and Schroeder 1995). As a result, one obtains the following expression

$$\mathcal{G}_2(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = -2i\frac{(-iH)^3}{8M_{\rm Pl}^3\sqrt{\epsilon_{1*}^3k_1^3k_2^3k_3^3}}\frac{1}{H^2}\epsilon_{1*}^2\,(\boldsymbol{k}_1\cdot\boldsymbol{k}_2 + \text{two permutations})$$
$$\times\left[\frac{-1}{\eta_{\rm e}}\mathrm{e}^{ik_{\rm T}\eta_{\rm e}} - (k_1k_2+k_2k_3+k_1k_3)\,\frac{\mathrm{e}^{ik_{\rm T}\eta_{\rm e}}}{ik_{\rm T}}\right.$$
$$\left. + ik_1k_2k_3\left(\frac{\eta_{\rm e}\mathrm{e}^{ik_{\rm T}\eta_{\rm e}}}{ik_{\rm T}} + \frac{\mathrm{e}^{ik_{\rm T}\eta_{\rm e}}}{k_{\rm T}^2}\right)\right]. \tag{115}$$

Then, the final step is to take $\eta_{\rm e}\to 0$. Clearly, there is a problem with the first term and, therefore, in the following expressions, we will keep $\eta_{\rm e}$ unspecified. For the other terms, the above expression simplifies and one is led to

$$\mathcal{G}_2(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = -2i\frac{(-iH)^3}{8M_{\rm Pl}^3\sqrt{\epsilon_{1*}^3k_1^3k_2^3k_3^3}}\frac{1}{H^2}\epsilon_{1*}^2\,(\boldsymbol{k}_1\cdot\boldsymbol{k}_2 + \text{two permutations})$$
$$\times\left[\frac{-1}{\eta_{\rm e}}\mathrm{e}^{ik_{\rm T}\eta_{\rm e}} + \frac{i}{k_{\rm T}}(k_1k_2+k_2k_3+k_1k_3) + \frac{i}{k_{\rm T}^2}k_1k_2k_3\right]. \tag{116}$$

This completes the calculation of $\mathcal{G}_2(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3)$. Now, we insert the above result into Eq. (102) in order to determine the contribution of $\mathcal{G}_2(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3)$ to $\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_2}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_3}(\eta_\mathrm{e})\rangle$. To perform this calculation, we need $f_{\boldsymbol{k}}(\eta_\mathrm{e})$, which we take to be $iH/[8M_\mathrm{Pl}^3\sqrt{\epsilon_1^3(\eta_\mathrm{e})k_1^3k_2^3k_3^3}]$ since the limit $\eta_\mathrm{e}\to 0$ does not cause any problem in that case. As a result, one finds that

$$\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_2}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_3}(\eta_\mathrm{e})\rangle^{(2)}=\frac{(2\pi)^3}{(2\pi)^{9/2}}M_\mathrm{Pl}^2\delta^{(3)}\left(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3\right)$$
$$\times\left[\frac{(iH)^3}{8M_\mathrm{Pl}^3\sqrt{\epsilon_1^3(\eta_\mathrm{e})k_1^3k_2^3k_3^3}}\mathcal{G}_2(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3)+\frac{(-iH)^3}{8M_\mathrm{Pl}^3\sqrt{\epsilon_1^3(\eta_\mathrm{e})k_1^3k_2^3k_3^3}}\mathcal{G}_2^*(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3)\right], \tag{117}$$

which, combined with Eq. (116), leads to

$$\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_2}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_3}(\eta_\mathrm{e})\rangle^{(2)}=\frac{(2\pi)^3}{(2\pi)^{9/2}}M_\mathrm{Pl}^2\delta^{(3)}\left(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3\right)\frac{H^3}{8M_\mathrm{Pl}^3\sqrt{\epsilon_1^3(\eta_\mathrm{e})k_1^3k_2^3k_3^3}}$$
$$\times\frac{2H^3}{8M_\mathrm{Pl}^3\sqrt{\epsilon_{1*}^3k_1^3k_2^3k_3^3}}\frac{1}{H^2}\epsilon_{1*}^2\left(\boldsymbol{k}_1\cdot\boldsymbol{k}_2+\text{two permutations}\right)\left\{-i(-i)^3i^3\left[\frac{-1}{\eta_\mathrm{e}}\mathrm{e}^{ik_\mathrm{T}\eta_\mathrm{e}}\right.\right.$$
$$\left.+\frac{i}{k_\mathrm{T}}\left(k_1k_2+k_2k_3+k_1k_3\right)+\frac{i}{k_\mathrm{T}^2}k_1k_2k_3\right]+i(i)^3(-i)^3\left[\frac{-1}{\eta_\mathrm{e}}\mathrm{e}^{-ik_\mathrm{T}\eta_\mathrm{e}}\right.$$
$$\left.\left.-\frac{i}{k_\mathrm{T}}(k_1k_2+k_2k_3+k_1k_3)-\frac{i}{k_\mathrm{T}^2}k_1k_2k_3\right]\right\}. \tag{118}$$

This expression can be simplified further and one obtains the following formula

$$\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_2}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_3}(\eta_\mathrm{e})\rangle^{(2)}=\frac{(2\pi)^3}{(2\pi)^{9/2}}M_\mathrm{Pl}^2\delta^{(3)}\left(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3\right)$$
$$\times\frac{2H^6\epsilon_{1*}^2}{64H^2M_\mathrm{Pl}^6\epsilon_{1*}^{3/2}\epsilon_1^{3/2}(\eta_\mathrm{e})k_1^3k_2^3k_3^3}\left(\boldsymbol{k}_1\cdot\boldsymbol{k}_2+\text{two permutations}\right)$$
$$\times\left[-2k_\mathrm{T}\frac{\sin(k_\mathrm{T}\eta_\mathrm{e})}{k_\mathrm{T}\eta_\mathrm{e}}+\frac{2}{k_\mathrm{T}}\left(k_1k_2+k_2k_3+k_1k_3\right)+\frac{2}{k_\mathrm{T}^2}k_1k_2k_3\right]. \tag{119}$$

We see that the limit $\eta_\mathrm{e}\to 0$ is now well defined and can be taken. The term in $\mathcal{G}_2(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3)$ was singular but, combined with its complex conjugate in the correlator, the limit has become regular. Therefore, the appearance of a singular limit was just a temporary technical problem and, in the expression of the physical

quantity, the problematic term has disappeared. The final expression reads

$$\langle \mathcal{R}_{\boldsymbol{k}_1}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_2}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_3}(\eta_\mathrm{e})\rangle^{(2)} = \frac{(2\pi)^3}{(2\pi)^{9/2}}\frac{H^4}{16M_\mathrm{Pl}^4\epsilon_1}\frac{1}{(k_1k_2k_3)^3}\,(\boldsymbol{k}_1\cdot\boldsymbol{k}_2 + \text{two permutations})$$
$$\times\left[-k_\mathrm{T} + \frac{1}{k_\mathrm{T}}\,(k_1k_2 + k_2k_3 + k_1k_3) + \frac{1}{k_\mathrm{T}^2}k_1k_2k_3\right]\delta^{(3)}\,(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3)\,. \qquad (120)$$

As expected the amplitude is controlled by the Hubble parameter (to the power four while the amplitude of the power spectrum was quadratic in H) and the (first) slow-roll parameter. We also see that the scale dependence is quite complicated.

The calculation proceeds exactly the same way for the first and third terms. Explicitly, one obtains

$$\langle \mathcal{R}_{\boldsymbol{k}_1}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_2}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_3}(\eta_\mathrm{e})\rangle^{(1)} = \frac{(2\pi)^3}{(2\pi)^{9/2}}\frac{H^4}{16M_\mathrm{Pl}^4\epsilon_1}\frac{1}{(k_1k_2k_3)^3}$$
$$\times\left[\left(1+\frac{k_1}{k_\mathrm{T}}\right)\frac{k_2^2k_3^2}{k_\mathrm{T}} + \left(1+\frac{k_2}{k_\mathrm{T}}\right)\frac{k_1^2k_3^2}{k_\mathrm{T}} + \left(1+\frac{k_3}{k_\mathrm{T}}\right)\frac{k_1^2k_2^2}{k_\mathrm{T}}\right]\delta^{(3)}\,(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3)\,, \qquad (121)$$

$$\langle \mathcal{R}_{\boldsymbol{k}_1}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_2}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_3}(\eta_\mathrm{e})\rangle^{(3)} = -\frac{(2\pi)^3}{(2\pi)^{9/2}}\frac{H^4}{16M_\mathrm{Pl}^4\epsilon_1}\frac{1}{(k_1k_2k_3)^3}$$
$$\times\left[(\boldsymbol{k}_1\cdot\boldsymbol{k}_2)\frac{k_3^2}{k_\mathrm{T}}\left(2+\frac{k_1+k_2}{k_\mathrm{T}}\right) + (\boldsymbol{k}_1\cdot\boldsymbol{k}_3)\frac{k_2^2}{k_\mathrm{T}}\left(2+\frac{k_1+k_3}{k_\mathrm{T}}\right)\right.$$
$$\left. + (\boldsymbol{k}_2\cdot\boldsymbol{k}_3)\frac{k_1^2}{k_\mathrm{T}}\left(2+\frac{k_2+k_3}{k_\mathrm{T}}\right)\right]\delta^{(3)}\,(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3)\,. \qquad (122)$$

Finally, the seventh term given by Eq. (111) can be re-written in terms of the two-point correlation function

$$\langle \mathcal{R}_{\boldsymbol{k}_1}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_2}(\eta_\mathrm{e})\,\mathcal{R}_{\boldsymbol{k}_3}(\eta_\mathrm{e})\rangle^{(7)} = \frac{(2\pi)^3}{(2\pi)^{9/2}}2\pi^4\epsilon_2\frac{1}{(k_1k_2k_3)^3}$$
$$\times\left[k_1^3\mathcal{P}_\mathcal{R}(k_2)\mathcal{P}_\mathcal{R}(k_3) + k_2^3\mathcal{P}_\mathcal{R}(k_1)\mathcal{P}_\mathcal{R}(k_3) + k_3^3\mathcal{P}_\mathcal{R}(k_1)\mathcal{P}_\mathcal{R}(k_2)\right]\delta^{(3)}\,(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3)\,, \qquad (123)$$

where, in order to evaluate the last term, we have made use of the definition introduced before: $\mathcal{P}_\mathcal{R}(k) = k^3|f_k|^2/(2\pi^2)$, see Eq. (74).

We have now completed the calculation of the three-point correlation function in Fourier space. We notice that, as already mentioned above, the dependence in k_1, k_2, k_3 is rather non trivial. In order to emphasize this point, it is interesting to recalculate the three-point correlation function in the following simple setup. Suppose that we

write the curvature perturbation as

$$\mathcal{R}(\eta,\boldsymbol{x}) = \mathcal{R}_{\rm G}(\eta,\boldsymbol{x}) - \frac{3f_{\rm NL}^{\rm loc}}{5}\mathcal{R}_{\rm G}^2(\eta,\boldsymbol{x}) + \cdots, \tag{124}$$

where $\mathcal{R}_{\rm G}$ denotes a Gaussian quantity, and the factor of $3/5$ arises due to the relation between the Bardeen potential and the curvature perturbation during the matter dominated epoch. The amplitude of the quadratic term is constant and conventionally called $f_{\rm NL}^{\rm loc}$. Let us notice that this assumption is highly non trivial and that, a priori, the coefficient in front of the quadratic term is expected to be a function of space. Postulating that it is a constant enforces a particular scale dependence of the three-point correlation function as we are going to see. In Fourier space, the Gaussian part is written $\mathcal{R}_{\rm G} = (2\pi)^{-3/2}\int {\rm d}\boldsymbol{k}\mathcal{R}_{\boldsymbol{k}}^{\rm G}{\rm e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}$ and it follows that

$$\mathcal{R}^2(\eta,\boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}}\int {\rm d}\boldsymbol{k}\,(2\pi)^{-3/2}\int {\rm d}\boldsymbol{p}\,\mathcal{R}_{\boldsymbol{p}}^{\rm G}\,\mathcal{R}_{\boldsymbol{k}-\boldsymbol{p}}^{\rm G}\,{\rm e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{125}$$

from which we can read the Fourier coefficient of the non-linear curvature perturbation, namely

$$\mathcal{R}_{\boldsymbol{k}} = \mathcal{R}_{\boldsymbol{k}}^{\rm G} - \frac{3f_{\rm NL}^{\rm loc}}{5}(2\pi)^{-3/2}\int {\rm d}\boldsymbol{p}\,\mathcal{R}_{\boldsymbol{p}}^{\rm G}\,\mathcal{R}_{\boldsymbol{k}-\boldsymbol{p}}^{\rm G}. \tag{126}$$

Using this expression, one can now evaluate the bispectrum. One obtains

$$\begin{aligned}\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta)\mathcal{R}_{\boldsymbol{k}_2}(\eta)\mathcal{R}_{\boldsymbol{k}_3}(\eta)\rangle = \Bigg\langle&\left[\mathcal{R}_{\boldsymbol{k}_1}^{\rm G} - \frac{3f_{\rm NL}^{\rm loc}}{5}(2\pi)^{-3/2}\int {\rm d}\boldsymbol{p}_1\,\mathcal{R}_{\boldsymbol{p}_1}^{\rm G}\,\mathcal{R}_{\boldsymbol{k}_1-\boldsymbol{p}_1}^{\rm G}\right]\\ &\times\left[\mathcal{R}_{\boldsymbol{k}_2}^{\rm G} - \frac{3f_{\rm NL}^{\rm loc}}{5}(2\pi)^{-3/2}\int {\rm d}\boldsymbol{p}_2\,\mathcal{R}_{\boldsymbol{p}_2}^{\rm G}\,\mathcal{R}_{\boldsymbol{k}_2-\boldsymbol{p}_2}^{\rm G}\right]\\ &\times\left[\mathcal{R}_{\boldsymbol{k}_3}^{\rm G} - \frac{3f_{\rm NL}^{\rm loc}}{5}(2\pi)^{-3/2}\int {\rm d}\boldsymbol{p}_3\,\mathcal{R}_{\boldsymbol{p}_3}^{\rm G}\,\mathcal{R}_{\boldsymbol{k}_3-\boldsymbol{p}_3}^{\rm G}\right]\Bigg\rangle,\end{aligned} \tag{127}$$

and, therefore,

$$\begin{aligned}&\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta)\mathcal{R}_{\boldsymbol{k}_2}(\eta)\mathcal{R}_{\boldsymbol{k}_3}(\eta)\rangle = \left\langle\mathcal{R}_{\boldsymbol{k}_1}^{\rm G}(\eta)\mathcal{R}_{\boldsymbol{k}_2}^{\rm G}(\eta)\mathcal{R}_{\boldsymbol{k}_3}^{\rm G}(\eta)\right\rangle\\ &-\frac{3f_{\rm NL}^{\rm loc}}{5}(2\pi)^{-3/2}\int {\rm d}\boldsymbol{p}_3\,\left\langle\mathcal{R}_{\boldsymbol{k}_1}^{\rm G}(\eta)\mathcal{R}_{\boldsymbol{k}_2}^{\rm G}(\eta)\mathcal{R}_{\boldsymbol{p}_3}^{\rm G}(\eta)\mathcal{R}_{\boldsymbol{k}_3-\boldsymbol{p}_3}^{\rm G}(\eta)\right\rangle + \text{two permutations} + \cdots,\end{aligned} \tag{128}$$

where the dots denote the higher order terms. Since the three point correlation function vanishes for Gaussian statistics, the previous expression reduces to

$$\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta)\mathcal{R}_{\boldsymbol{k}_2}(\eta)\mathcal{R}_{\boldsymbol{k}_3}(\eta)\rangle = -\frac{3f_{\rm NL}^{\rm loc}}{5}(2\pi)^{-3/2}\int {\rm d}\boldsymbol{p}_3\,\left\langle\mathcal{R}^{\rm G}_{\boldsymbol{k}_1}(\eta)\mathcal{R}^{\rm G}_{\boldsymbol{k}_2}(\eta)\mathcal{R}^{\rm G}_{\boldsymbol{p}_3}(\eta)\mathcal{R}^{\rm G}_{\boldsymbol{k}_3-\boldsymbol{p}_3}(\eta)\right\rangle + \text{two permutations} + \cdots . \tag{129}$$

As expected, the three-point correlation function is proportional to the coefficient $f_{\rm NL}^{\rm loc}$. To proceed, one can evaluate this expression by means of the Wick's theorem. Then, one obtains

$$\begin{aligned}
&\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta)\mathcal{R}_{\boldsymbol{k}_2}(\eta)\mathcal{R}_{\boldsymbol{k}_3}(\eta)\rangle = \\
&-\frac{3f_{\rm NL}^{\rm loc}}{5}(2\pi)^{-3/2}\int {\rm d}\boldsymbol{p}_3\Bigg[\left\langle\mathcal{R}^{\rm G}_{\boldsymbol{k}_1}(\eta)\mathcal{R}^{\rm G}_{\boldsymbol{k}_2}(\eta)\right\rangle\left\langle\mathcal{R}^{\rm G}_{\boldsymbol{p}_3}(\eta)\mathcal{R}^{\rm G}_{\boldsymbol{k}_3-\boldsymbol{p}_3}(\eta)\right\rangle \\
&+\left\langle\mathcal{R}^{\rm G}_{\boldsymbol{k}_1}(\eta)\mathcal{R}^{\rm G}_{\boldsymbol{p}_3}(\eta)\right\rangle\left\langle\mathcal{R}^{\rm G}_{\boldsymbol{k}_2}(\eta)\mathcal{R}^{\rm G}_{\boldsymbol{k}_3-\boldsymbol{p}_3}(\eta)\right\rangle + \left\langle\mathcal{R}^{\rm G}_{\boldsymbol{k}_1}(\eta)\mathcal{R}^{\rm G}_{\boldsymbol{k}_3-\boldsymbol{p}_3}(\eta)\right\rangle\left\langle\mathcal{R}^{\rm G}_{\boldsymbol{k}_2}(\eta)\mathcal{R}^{\rm G}_{\boldsymbol{p}_3}(\eta)\right\rangle \\
&+\text{two permutations} + \cdots\Bigg].
\end{aligned} \tag{130}$$

Since the two-point correlation functions are nothing but the power spectrum, the above expression takes the following form

$$\begin{aligned}
\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta)\mathcal{R}_{\boldsymbol{k}_2}(\eta)\mathcal{R}_{\boldsymbol{k}_3}(\eta)\rangle = &-\frac{3f_{\rm NL}^{\rm loc}}{5}(2\pi)^{-3/2} \\
&\times\int {\rm d}\boldsymbol{p}_3\Bigg[\frac{(2\pi)^2}{2}\frac{\mathcal{P}_{\mathcal{R}}(k_1)}{k_1^3}\delta^{(3)}\left(\boldsymbol{k}_1+\boldsymbol{k}_2\right)\frac{(2\pi)^2}{2}\frac{\mathcal{P}_{\mathcal{R}}(p_3)}{k_3^3}\delta^{(3)}\left(\boldsymbol{p}_3+\boldsymbol{k}_3-\boldsymbol{p}_3\right) \\
&+\frac{(2\pi)^2}{2}\frac{\mathcal{P}_{\mathcal{R}}(k_1)}{k_1^3}\delta^{(3)}\left(\boldsymbol{k}_1+\boldsymbol{p}_3\right)\frac{(2\pi)^2}{2}\frac{\mathcal{P}_{\mathcal{R}}(k_2)}{k_2^3}\delta^{(3)}\left(\boldsymbol{k}_2+\boldsymbol{k}_3-\boldsymbol{p}_3\right) \\
&+\frac{(2\pi)^2}{2}\frac{\mathcal{P}_{\mathcal{R}}(k_1)}{k_1^3}\delta^{(3)}\left(\boldsymbol{k}_1+\boldsymbol{k}_3-\boldsymbol{p}_3\right)\frac{(2\pi)^2}{2}\frac{\mathcal{P}_{\mathcal{R}}(k_2)}{k_2^3}\delta^{(3)}\left(\boldsymbol{k}_2+\boldsymbol{p}_3\right) \\
&+\text{two permutations} + \cdots\Bigg].
\end{aligned} \tag{131}$$

Then, the integral over $\boldsymbol{p}_3$ can be easily performed, thanks to the presence of the Dirac delta functions. We see that the first term in the above expression is different from the two next ones. Indeed, it leads to a term $\delta^{(3)}(\boldsymbol{k}_3)$ which can be ignored since, in some sense, it is homogeneous and only participates to the background. The two other terms yield a $\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)$ which ensures momentum conservation.

The final expression reads

$$\langle \mathcal{R}_{\boldsymbol{k}_1} \mathcal{R}_{\boldsymbol{k}_2} \mathcal{R}_{\boldsymbol{k}_3} \rangle = -\frac{3 f_{\rm NL}^{\rm loc}}{10} \, (2\,\pi)^4 \, (2\,\pi)^{-3/2} \, \frac{1}{k_1^3 \, k_2^3 \, k_3^3} \, \delta^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) \times \left[k_1^3 \, \mathcal{P}_{\mathcal{R}}(k_2) \, \mathcal{P}_{\mathcal{R}}(k_3) + \text{two permutations} \right] . \tag{132}$$

We see that the scale dependence of the bispectrum for this simple model does not reproduce what we obtained in the case of inflation, see Eqs. (120)–(123). The inflationary case is clearly much more complicated. In fact, Eq. (132) has a similar structure as $\langle \mathcal{R}_{\boldsymbol{k}_1}(\eta_{\rm e}) \, \mathcal{R}_{\boldsymbol{k}_2}(\eta_{\rm e}) \, \mathcal{R}_{\boldsymbol{k}_3}(\eta_{\rm e}) \rangle^{(7)}$, see Eq. (123). But the three extra terms $\langle \mathcal{R}_{\boldsymbol{k}_1}(\eta_{\rm e}) \, \mathcal{R}_{\boldsymbol{k}_2}(\eta_{\rm e}) \, \mathcal{R}_{\boldsymbol{k}_3}(\eta_{\rm e}) \rangle^{(1,2,3)}$ are such that the full slow-roll bispectrum differs from Eq. (132).

At this stage, it is worth discussing again our conventions. We have seen below Eq. (94) that, often in the literature, the two-point correlation function is defined as $\langle \bar{\mathcal{R}}_{\boldsymbol{k}_1} \bar{\mathcal{R}}_{\boldsymbol{k}_2} \rangle \equiv (2\pi)^3 P_{\mathcal{R}}(k_1) \delta^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2)$, where $\bar{\mathcal{R}}_{\boldsymbol{k}_1} \equiv (2\pi)^{3/2} \mathcal{R}_{\boldsymbol{k}_1}$ and $P_{\mathcal{R}}(k_1) \equiv |f_{\boldsymbol{k}_1}|^2 \neq \mathcal{P}_{\mathcal{R}}(k_1)$. Then, in order to mimic and/or generalize the definition of the two-point correlation function, the following definition of the bispectrum $\mathcal{B}_{\mathcal{R}}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$ is introduced

$$\langle \bar{\mathcal{R}}_{\boldsymbol{k}_1} \bar{\mathcal{R}}_{\boldsymbol{k}_2} \bar{\mathcal{R}}_{\boldsymbol{k}_3} \rangle = (2\pi)^3 \mathcal{B}_{\mathcal{R}}(k_1, k_2, k_3) \, \delta^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) . \tag{133}$$

Notice that we could have also used another definition $\langle \mathcal{R}_{\boldsymbol{k}_1} \mathcal{R}_{\boldsymbol{k}_2} \mathcal{R}_{\boldsymbol{k}_3} \rangle = (2\pi)^3 \mathcal{B}_{\mathcal{R}}(k_1, k_2, k_3) \, \delta^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3)$, which would have resulted in a difference by a factor of $(2\pi)^{9/2}$ [and, by the way, explains the appearance of such a factor in Eq. (102)]. Here, we do not follow this route and use the convention (133). Then, Eq. (132) implies that

$$\mathcal{B}_{\mathcal{R}}(k_1, k_2, k_3) = -\frac{6}{5} f_{\rm NL}^{\rm loc} \left(|f_{k_2}|^2 |f_{k_3}|^2 + \text{two permutations} \right) \tag{134}$$

$$= -\frac{6}{5} f_{\rm NL}^{\rm loc} \left[P_{\mathcal{R}}(k_2) P_{\mathcal{R}}(k_3) + \text{two permutations} \right] . \tag{135}$$

It is also frequent to define the bispectrum of the Bardeen potential Φ rather than the conserved quantity $\mathcal{R}$. Concretely, the definition reads[2]

$$\langle \bar{\Phi}_{\boldsymbol{k}_1} \bar{\Phi}_{\boldsymbol{k}_2} \bar{\Phi}_{\boldsymbol{k}_3} \rangle = (2\pi)^3 \mathcal{B}_{\Phi}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) \, \delta^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) . \tag{138}$$

[2] As already mentioned, our convention for the Fourier transform is such that

$$\Phi(\eta, \boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}} \int \mathrm{d}\boldsymbol{k} \, \Phi_k(\eta) \, \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \tag{136}$$

Since $\zeta = 5\Phi/3 = -\mathcal{R}$, we have $\mathcal{B}_\Phi = -27\mathcal{B}_\mathcal{R}/125$. However, since $f_{\boldsymbol{k}} = -(5/3)u_{\boldsymbol{k}}$, we also have $P_\mathcal{R} = (25/9)P_\Phi$. As a consequence, from Eq. (135), one obtains that

$$\mathcal{B}_\Phi(k_1,k_2,k_3) = 2f_{\rm NL}^{\rm loc}\left[P_\Phi(k_2)P_\Phi(k_3) + \text{two permutations}\right], \tag{139}$$

which is a formula that often appears in the literature.

Of course, we can also put the slow-roll bispectrum calculated before under the form given by Eq. (133). For this purpose, let us write Eqs. (121), (120), (122) and (123) as

$$\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta_{\rm e})\,\mathcal{R}_{\boldsymbol{k}_2}(\eta_{\rm e})\,\mathcal{R}_{\boldsymbol{k}_3}(\eta_{\rm e})\rangle^{(i)} \equiv \frac{(2\pi)^3}{(2\pi)^{9/2}}\mathcal{F}^{(i)}\frac{1}{k_1^3\,k_2^3\,k_3^3}\,\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3), \tag{140}$$

where the concrete expression of the $\mathcal{F}^{(i)}$ can be read off from those equations. Then, the bispectrum for single-field slow-roll models can be written as

$$\mathcal{B}_\mathcal{R}^{\rm sr}(k_1,k_2,k_3) = \frac{1}{k_1^3\,k_2^3\,k_3^3}\sum_{i=1,2,3,7}\mathcal{F}^{(i)}. \tag{141}$$

The previous result can also be used to define an effective, scale dependent, $f_{\rm NL}$ parameter. If we equate the full bispectrum $\sum_{i=1,2,3,7}\langle\mathcal{R}_{\boldsymbol{k}_1}(\eta_{\rm e})\,\mathcal{R}_{\boldsymbol{k}_2}(\eta_{\rm e})\,\mathcal{R}_{\boldsymbol{k}_3}(\eta_{\rm e})\rangle^{(i)}$ to the expression of Eq. (132), one obtains

$$f_{\rm NL}^{\rm sr}(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = -\frac{10}{3}\,(2\,\pi)^{-4}\sum_{i=1,2,3,7}\mathcal{F}^{(i)} \times\left[k_1^3\,\mathcal{P}_\mathcal{R}(k_2)\,\mathcal{P}_\mathcal{R}(k_3) + \text{two permutations}\right]^{-1}. \tag{142}$$

If, for instance, we evaluate this quantity for $\boldsymbol{k}_1 = -\boldsymbol{k}_2$ and a vanishing $\boldsymbol{k}_3$ (so that $\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3$ is also zero which is mandatory given the presence of the Dirac function in the above expressions), then the expressions of $\mathcal{F}^{(i)}$ simplify such that one obtains

$$\sum_{i=1,2,3,7}\mathcal{F}^{(i)} = \frac{H^4k^3}{16M_{\rm Pl}^4\epsilon_1}\left(\frac{1}{2}+\frac{3}{2}+0+\frac{\epsilon_2}{\epsilon_1}\right) = \frac{H^4k^3}{16M_{\rm Pl}^4\epsilon_1^2}(2\epsilon_1+\epsilon_2), \tag{143}$$

---

and, following the notation that we have already introduced, $\bar{\Phi}_{\boldsymbol{k}} = (2\pi)^{3/2}\Phi_{\boldsymbol{k}}$. Moreover, if the Bardeen potential quantum operator is written as

$$\hat{\Phi}(\eta,\boldsymbol{x}) = \int\frac{{\rm d}^3\boldsymbol{k}}{(2\pi)^{3/2}}\left[a_{\boldsymbol{k}}u_{\boldsymbol{k}}(\eta){\rm e}^{{\rm i}\boldsymbol{k}\cdot\boldsymbol{x}} + a_{\boldsymbol{k}}^\dagger u_{\boldsymbol{k}}^*(\eta){\rm e}^{-{\rm i}\boldsymbol{k}\cdot\boldsymbol{x}}\right], \tag{137}$$

then one has $P_\Phi \equiv |u_{\boldsymbol{k}}|^2$ and $\mathcal{P}_\phi \equiv k^3|u_{\boldsymbol{k}}|^2/(2\pi^2)$.

which, using Eq. (86), can be written as (Maldacena 2003; Renaux-Petel 2010)

$$f_{\rm NL}^{\rm sr,sq} = \frac{5}{12}(n_{\rm S} - 1), \tag{144}$$

where "sq" means "squeezed" and refers to the fact that we have taken the particular configuration $\boldsymbol{k}_1 = -\boldsymbol{k}_2$ and a vanishing $\boldsymbol{k}_3$. Notice that, since we calculate a three-point correlation function, the sign of $f_{\rm NL}$ is non trivial. The sign that we have obtained results from the choice made in Eq. (124) and from the fact that we evaluate the correlator of $\mathcal{R}$. Finally, very roughly speaking (see Sect. 5 for a more complete discussion) the present status of the art is such that one can detect Non-Gaussianities if $|f_{\rm NL}| > 5$. For slow-roll models, since $n_{\rm S} \simeq 0.96$, one obtains $f_{\rm NL}^{\rm sr,sq} \simeq -1.6\times10^{-2}$, a number that is therefore undetectable. This conclusion is in fact valid for any configuration one may choose. Let us also mention that other consistency relations for Non-Gaussianities have recently been studied in Sreenath et al. (2013, 2015) and Sreenath and Sriramkumar (2014).

Clearly, a detection of a non-vanishing three-point correlation function, given present day technology, would immediately rule out single field slow-roll models with a standard kinetic term. It is therefore quite remarkable that Non-Gaussianity has not been detected so far. Let us also stress that the opposite statement is not true. The fact that we do not see Non-Gaussianities does not imply that the more complicated models of inflation are necessarily ruled out even if some of them do predict large Non-Gaussianities. For the calculation of the three-point correlation functions of these more complicated models, we again refer to Byrnes (2014).

4.3 Inflationary Four-Point Correlation Functions

Obviously, the next step is to calculate the four-point correlation function or trispectrum (Seery et al. 2007; Byrnes et al. 2006; Arroja and Koyama 2008). Of course, when we consider higher order correlation functions, the calculations become more and more complicated. In the previous sub-section, we calculated the action at third order in the perturbations in order to derive the inflationary three-point correlation function. In order to calculate the four-point correlation function, one therefore needs to evaluate the perturbed action at fourth order. In order to get an idea of how involved it can be, let us consider again Eq. (124) but expanded up to third order

$$\mathcal{R}(\eta,\boldsymbol{x}) = \mathcal{R}_{\rm G}(\eta,\boldsymbol{x}) - \frac{3f_{\rm NL}}{5}\mathcal{R}_{\rm G}^2(\eta,\boldsymbol{x}) + \frac{9}{25}g_{\rm NL}\mathcal{R}_{\rm G}^3(\eta,\boldsymbol{x})\cdots, \tag{145}$$

thus introducing the parameter $g_{\rm NL}$. Here, we write $f_{\rm NL}$ in order to avoid cumbersome notations but it should be clear that $f_{\rm NL} = f_{\rm NL}^{\rm loc}$ (and this will be the case in the rest of

this section). The cube of the curvature perturbation can be expressed as

$$\mathcal{R}_{\mathrm{G}}^{3}(\eta,\boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}}\int \mathrm{d}\boldsymbol{k}\,(2\pi)^{-3}\int \mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}\,\mathcal{R}_{\boldsymbol{k}-\boldsymbol{p}-\boldsymbol{q}}\,\mathcal{R}_{\boldsymbol{p}}\,\mathcal{R}_{\boldsymbol{q}}\,\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{146}$$

which allows us to identify the Fourier transform of the cube of the curvature perturbation [as we identified the Fourier transform of the square of the curvature perturbation in Eq. (126)]. Then, the four-point correlator takes the form

$$\begin{aligned}
&\left\langle \mathcal{R}_{\boldsymbol{k}_1}\mathcal{R}_{\boldsymbol{k}_2}\mathcal{R}_{\boldsymbol{k}_3}\mathcal{R}_{\boldsymbol{k}_4}\right\rangle = \\
&\left\langle\left[\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_1} - \frac{3f_{\mathrm{NL}}}{5}(2\pi)^{-3/2}\int \mathrm{d}\boldsymbol{p}_1\,\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_1}\,\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_1-\boldsymbol{p}_1} + \frac{9g_{\mathrm{NL}}}{25}(2\pi)^{-3}\int \mathrm{d}\boldsymbol{p}_1\mathrm{d}\boldsymbol{q}_1\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_1-\boldsymbol{p}_1-\boldsymbol{q}_1}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_1}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{q}_1}\right]\right. \\
&\left[\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_2} - \frac{3f_{\mathrm{NL}}}{5}(2\pi)^{-3/2}\int \mathrm{d}\boldsymbol{p}_2\,\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_2}\,\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_2-\boldsymbol{p}_2} + \frac{9g_{\mathrm{NL}}}{25}(2\pi)^{-3}\int \mathrm{d}\boldsymbol{p}_2\mathrm{d}\boldsymbol{q}_2\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_2-\boldsymbol{p}_2-\boldsymbol{q}_2}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_2}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{q}_2}\right] \\
&\left[\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_3} - \frac{3f_{\mathrm{NL}}}{5}(2\pi)^{-3/2}\int \mathrm{d}\boldsymbol{p}_3\,\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_3}\,\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_3-\boldsymbol{p}_3} + \frac{9g_{\mathrm{NL}}}{25}(2\pi)^{-3}\int \mathrm{d}\boldsymbol{p}_3\mathrm{d}\boldsymbol{q}_3\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_3-\boldsymbol{p}_3-\boldsymbol{q}_3}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_3}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{q}_3}\right] \\
&\left.\left[\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_4} - \frac{3f_{\mathrm{NL}}}{5}(2\pi)^{-3/2}\int \mathrm{d}\boldsymbol{p}_4\,\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_4}\,\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_4-\boldsymbol{p}_4} + \frac{9g_{\mathrm{NL}}}{25}(2\pi)^{-3}\int \mathrm{d}\boldsymbol{p}_4\mathrm{d}\boldsymbol{q}_4\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_4-\boldsymbol{p}_4-\boldsymbol{q}_4}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_4}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{q}_4}\right]\right\rangle.
\end{aligned} \tag{147}$$

Expanding this expression, one arrives at the following formula

$$\begin{aligned}
&\left\langle \mathcal{R}_{\boldsymbol{k}_1}\mathcal{R}_{\boldsymbol{k}_2}\mathcal{R}_{\boldsymbol{k}_3}\mathcal{R}_{\boldsymbol{k}_4}\right\rangle = \left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_1}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_2}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_3}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_4}\right\rangle \\
&- \frac{3f_{\mathrm{NL}}}{5}(2\pi)^{-3/2}\left(\int \mathrm{d}\boldsymbol{p}_4\left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_1}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_2}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_3}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_4}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_4-\boldsymbol{p}_4}\right\rangle + \text{three permutations}\right) \\
&+ \frac{9}{25}f_{\mathrm{NL}}^{2}(2\pi)^{-3}\left(\int \mathrm{d}\boldsymbol{p}_1\int \mathrm{d}\boldsymbol{p}_2\left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_1}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_1-\boldsymbol{p}_1}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_2}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_2-\boldsymbol{p}_2}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_3}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_4}\right\rangle + \text{five permutations}\right) \\
&+ \frac{9}{25}g_{\mathrm{NL}}(2\pi)^{-3}\left(\int \mathrm{d}\boldsymbol{p}_4\int \mathrm{d}\boldsymbol{q}_4\left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_1}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_2}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_3}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_4-\boldsymbol{p}_4-\boldsymbol{q}_4}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_4}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{q}_4}\right\rangle\right. \\
&\left.+ \text{three permutations}\right) + \cdots,
\end{aligned} \tag{148}$$

where the dots denote higher order terms. The first term in the above expansion is non-vanishing but can be expressed as the square of two-point correlation functions and will be ignored in the following. The second term is zero since it involves five-point correlation functions of Gaussian quantities. The two last terms are the terms of interest. We see that they are given in terms of a six-point correlation function, a quantity which is not zero for Gaussian quantities. These terms can be evaluated by means of the Wick's theorem and lead to the sum of fifteen terms, each of them being made of the product of three two-point correlation functions. For the term proportional to f_{NL}^{2}, among the fifteen only eight of them actually contribute. An

example of a term contributing is given by

$$\begin{aligned}
&\int \mathrm{d}\boldsymbol{p}_1\mathrm{d}\boldsymbol{p}_2 \left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_1}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_2}\right\rangle \left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_1-\boldsymbol{p}_1}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_3}\right\rangle \left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_2-\boldsymbol{p}_2}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_4}\right\rangle \\
&= \int \mathrm{d}\boldsymbol{p}_1\mathrm{d}\boldsymbol{p}_2 \frac{(2\pi)^2}{2p_1^3}\mathcal{P}_{\mathcal{R}}(p_1)\delta^{(3)}(\boldsymbol{p}_1+\boldsymbol{p}_2)\frac{(2\pi)^2}{2k_3^3}\mathcal{P}_{\mathcal{R}}(|\boldsymbol{k}_1-\boldsymbol{p}_1|)\delta^{(3)}(\boldsymbol{k}_1-\boldsymbol{p}_1+\boldsymbol{k}_3) \\
&\times \frac{(2\pi)^2}{2k_4^3}\mathcal{P}_{\mathcal{R}}(|\boldsymbol{k}_2-\boldsymbol{p}_2|)\delta^{(3)}(\boldsymbol{k}_2-\boldsymbol{p}_2+\boldsymbol{k}_4) \\
&= \frac{(2\pi)^6}{8}\int \mathrm{d}\boldsymbol{p}_1 \frac{\mathcal{P}_{\mathcal{R}}(p_1)}{p_1^3}\frac{\mathcal{P}_{\mathcal{R}}(|\boldsymbol{k}_1-\boldsymbol{p}_1|)}{k_3^3}\frac{\mathcal{P}_{\mathcal{R}}(|\boldsymbol{k}_2+\boldsymbol{p}_1|)}{k_4^3}\delta^{(3)}(\boldsymbol{k}_1-\boldsymbol{p}_1+\boldsymbol{k}_3) \\
&\times \delta^{(3)}(\boldsymbol{k}_2+\boldsymbol{p}_1+\boldsymbol{k}_4) \\
&= \frac{(2\pi)^6}{8}\frac{1}{|\boldsymbol{k}_1+\boldsymbol{k}_3|^3k_3^3k_4^3}\mathcal{P}_{\mathcal{R}}(|\boldsymbol{k}_1+\boldsymbol{k}_3|)\mathcal{P}_{\mathcal{R}}(k_3)\mathcal{P}_{\mathcal{R}}(k_4)\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3+\boldsymbol{k}_4).
\end{aligned} \tag{149}$$

In fact among the eight terms mentioned above, four are identical to the one we have just calculated and the remaining four are all given by Eq. (149), but with $|\boldsymbol{k}_1+\boldsymbol{k}_3|$ replaced with $|\boldsymbol{k}_1+\boldsymbol{k}_4|$. On the other hand, an example of a non-contributing term is

$$\int \mathrm{d}\boldsymbol{p}_1\mathrm{d}\boldsymbol{p}_2 \left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_1}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_1-\boldsymbol{p}_1}\right\rangle \left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_2}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_2-\boldsymbol{p}_2}\right\rangle \left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_3}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_4}\right\rangle. \tag{150}$$

We see that the first two-point correlation function appearing in the above integral will lead to a term proportional to $\delta^{(3)}(\boldsymbol{p}_1+\boldsymbol{k}_1-\boldsymbol{p}_1)=\delta^{(3)}(\boldsymbol{k}_1)$, which, in some sense, is homogeneous. This explains why the term in Eq. (150) can be ignored.

Let us now come back to Eq. (148) and consider the term proportional to g_{NL}. Using again Wick's theorem, this term can be expressed as the sum of fifteen terms made of the product of three two-point correlation functions. Among these fifteen terms, only six participate to the final expression (and they all give the same contribution). One example is

$$\begin{aligned}
&\int \mathrm{d}\boldsymbol{p}_4\mathrm{d}\boldsymbol{q}_4 \left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_1}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_4-\boldsymbol{p}_4-\boldsymbol{q}_4}\right\rangle \left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_2}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{p}_4}\right\rangle \left\langle \mathcal{R}^{\mathrm{G}}_{\boldsymbol{k}_3}\mathcal{R}^{\mathrm{G}}_{\boldsymbol{q}_4}\right\rangle \\
&= \int \mathrm{d}\boldsymbol{p}_4\mathrm{d}\boldsymbol{q}_4 \frac{(2\pi)^2}{2k_1^3}\mathcal{P}_{\mathcal{R}}(k_1)\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_4-\boldsymbol{p}_4-\boldsymbol{q}_4)\frac{(2\pi)^2}{2k_2^3}\mathcal{P}_{\mathcal{R}}(k_2)\delta^{(3)}(\boldsymbol{k}_2+\boldsymbol{p}_4) \\
&\times \frac{(2\pi)^2}{2k_3^3}\mathcal{P}_{\mathcal{R}}(k_3)\delta^{(3)}(\boldsymbol{k}_3+\boldsymbol{q}_4) \\
&= \frac{(2\pi)^6}{8}\frac{1}{k_1^3k_2^3k_3^3}\mathcal{P}_{\mathcal{R}}(k_1)\mathcal{P}_{\mathcal{R}}(k_2)\mathcal{P}_{\mathcal{R}}(k_3)\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3+\boldsymbol{k}_4).
\end{aligned} \tag{151}$$

Putting everything together, one obtains the following expression

$$\left\langle \mathcal{R}_{\boldsymbol{k}_1}\mathcal{R}_{\boldsymbol{k}_2}\mathcal{R}_{\boldsymbol{k}_3}\mathcal{R}_{\boldsymbol{k}_4}\right\rangle = \frac{9f_{\rm NL}^2}{25}(2\pi)^{-3}\Bigg[4\times\frac{(2\pi)^6}{8}\frac{1}{|\boldsymbol{k}_1+\boldsymbol{k}_3|^3k_3^3k_4^3}\mathcal{P}_{\mathcal{R}}(|\boldsymbol{k}_1+\boldsymbol{k}_3|)\mathcal{P}_{\mathcal{R}}(k_3)\mathcal{P}_{\mathcal{R}}(k_4)$$
$$+\text{eleven permutations}\Bigg]\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3+\boldsymbol{k}_4)+\frac{9g_{\rm NL}}{25}(2\pi)^{-3}\Bigg[6\times\frac{(2\pi)^6}{8}\frac{1}{k_1^3k_2^3k_3^3}$$
$$\times\mathcal{P}_{\mathcal{R}}(k_1)\mathcal{P}_{\mathcal{R}}(k_2)\mathcal{P}_{\mathcal{R}}(k_3)+\text{three permutations}\Bigg]\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3+\boldsymbol{k}_4) \tag{152}$$
$$=\frac{36f_{\rm NL}^2}{25}(2\pi)^{-3}\Big(|f_{\boldsymbol{k}_1+\boldsymbol{k}_3}|^2|f_{\boldsymbol{k}_3}|^2|f_{\boldsymbol{k}_4}|^2+\text{eleven permutations}\Big)\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3+\boldsymbol{k}_4)$$
$$+\frac{54g_{\rm NL}}{25}(2\pi)^{-3}\Big(|f_{\boldsymbol{k}_1}|^2|f_{\boldsymbol{k}_3}|^2|f_{\boldsymbol{k}_3}|^2+\text{three permutations}\Big)\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3+\boldsymbol{k}_4). \tag{153}$$

The fact that we have eleven permutation in the first term comes from the fact that we had six terms and that each of these terms separates in two groups. At the end, this gives twelve terms. Usually, the definition of the trispectrum is given in terms of $\bar{\mathcal{R}}_{\boldsymbol{k}}$ (see the above discussions about conventions) and reads

$$\left\langle \bar{\mathcal{R}}_{\boldsymbol{k}_1}\bar{\mathcal{R}}_{\boldsymbol{k}_2}\bar{\mathcal{R}}_{\boldsymbol{k}_3}\bar{\mathcal{R}}_{\boldsymbol{k}_4}\right\rangle = (2\pi)^3\mathcal{T}_{\mathcal{R}}(k_1,k_2,k_3,k_4)\delta^{(3)}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3+\boldsymbol{k}_4), \tag{154}$$

with

$$\mathcal{T}_{\mathcal{R}}(k_1,k_2,k_3,k_4) = \tau_{\rm NL}\Big(|f_{\boldsymbol{k}_1+\boldsymbol{k}_3}|^2|f_{\boldsymbol{k}_3}|^2|f_{\boldsymbol{k}_4}|^2+\text{eleven permutations}\Big)$$
$$+\frac{54g_{\rm NL}}{25}\Big(|f_{\boldsymbol{k}_1}|^2|f_{\boldsymbol{k}_3}|^2|f_{\boldsymbol{k}_3}|^2+\text{three permutations}\Big). \tag{155}$$

One can check that our result (153) matches exactly this form provided that

$$\tau_{\rm NL} = \frac{36f_{\rm NL}^2}{25}. \tag{156}$$

This equation is called the Suyama-Yamaguchi consistency relation (Suyama et al. 2010) (more precisely, it is in fact a particular case of $\tau_{\rm NL}\geq 36f_{\rm NL}^2/25$). The above equation indicates that the tri-spectrum is expected to be quadratic in the slow-roll parameters and, hence, even harder to detect than the three-point correlation function. Of course, it should be stressed again that the scale dependence of Eq. (155) is not what would emerge from an exact calculation starting from the perturbed action at fourth order. In Sect. 5, we will discuss the constraints put by the Planck experiment on the tri-spectrum.

4.4 Adiabatic and Isocurvature Perturbations

Another important consequence that follows from the Planck data is that the perturbations are adiabatic. Before discussing in more detail in Sect. 5 how this conclusion is reached, we now explain what it means and what it implies for inflation.

The post inflationary Universe is made of four fluids: photons, neutrinos, baryons and cold dark matter (we are ignoring dark energy). In order to calculate the CMB anisotropies, one needs to integrate the equations governing the behavior of these four fluids. But we also need to specify initial conditions, just after inflation, at the onset of the radiation dominated era. Different initial conditions will lead to different subsequent evolutions and, therefore, to different CMB patterns. Adiabaticity refers to a situation where one has (Bucher et al. 2000)

$$\delta_{\rm cdm} = \delta_{\rm b} = \frac{3}{4}\delta_\gamma = \frac{3}{4}\delta_\nu, \tag{157}$$

where $\delta_X \equiv \delta\rho_X/\rho_X$ is the density contrast ("cdm" stands for cold dark matter, "b" for baryons, γ for photons and ν for neutrinos). It may be surprising that CMB data single out particular initial conditions and it is interesting to discuss why the conditions (157) play an important role. Equally important is the question of what they can teach us about inflation: after all, these initial conditions are the results of what happened during inflation. As a consequence, they certainly tell us something about the type of inflationary expansion that took place in the early Universe.

Let us start by giving the equations controlling the evolution of the four fluids mentioned before. Each fluid is characterized by its density contrast δ_X and by its velocity v_X. From energy conservation, one can derive the following equations

$$(\delta_{\rm c} - 3\Psi)' - k^2 v_{\rm c} = \Delta_{\rm c}' - k^2 v_{\rm c} = 0, \tag{158}$$

$$(\delta_{\rm b} - 3\Psi)' - k^2 v_{\rm b} = \Delta_{\rm b}' - k^2 v_{\rm b} = 0, \tag{159}$$

$$\left(\delta_\gamma - 4\Psi\right)' - \frac{4}{3}k^2 v_\gamma = \Delta_\gamma' - \frac{4}{3}k^2 v_\gamma = 0, \tag{160}$$

$$(\delta_\nu - 4\Psi)' - \frac{4}{3}k^2 v_\nu = \Delta_\nu' - \frac{4}{3}k^2 v_\nu = 0, \tag{161}$$

where Ψ is the second Bardeen potential already considered before (but, in the present context, we no longer necessarily have $\Psi = \Phi$) and where the quantities Δ_X are defined by the above equations. The space component of the conservation equation gives an equation for the velocities. For cold dark matter, one obtains

$$v_{\rm c}' + \mathcal{H} v_{\rm c} + \Phi = 0, \tag{162}$$

where Φ is the other Bardeen potential. In the early Universe, baryons and photons are tightly coupled. This means that $v_{\rm b} = v_\gamma \equiv v_{{\rm b}\gamma}$. The corresponding equation of motion reads

$$v'_{{\rm b}\gamma} + \frac{R}{1+R}\mathcal{H}v_{{\rm b}\gamma} + \Phi + \frac{1}{4}\frac{\delta_\gamma}{1+R} + \frac{4\eta_{{\rm b}\gamma}}{3a\rho_\gamma}\frac{R}{1+R}k^2 v_{{\rm b}\gamma} = 0, \tag{163}$$

where $\eta_{{\rm b}\gamma}$ is the viscosity (or anisotropic stress) of the fluid made of baryons and photons and R is three quarters of the baryon to photon energy density ratio, namely $R \equiv 3\rho_{\rm b}/(4\rho_\gamma)$. Finally, the conservation equation for the neutrinos can be written as

$$v'_\nu + \Phi + \frac{1}{4}\delta_\nu + \frac{\eta_\nu}{a\rho_\nu}k^2 v_\nu = 0, \tag{164}$$

where η_ν is the neutrinos viscosity (notice that the viscosity does not appear in the time component of the conservation equations). Since the above formulas contain the two Bardeen potentials, they must be supplemented by additional equations governing the behavior of Φ and Ψ. These are of course the perturbed Einstein equations. By combining the time-time and time-space Einstein equations, one arrives at

$$-\frac{k^2}{\mathcal{H}^2}\Psi - \frac{9}{2}\Psi\sum_X \Omega_X(1+w_X) = \frac{3}{2}\sum_X \Omega_X\Delta_X - \frac{9}{2}\mathcal{H}\sum_X \Omega_X(1+w_X)v_X, \tag{165}$$

where the sum runs over the four species mentioned above, where w_X is the equation of state parameter of the fluid X and $\mathcal{H} \equiv a'/a$. Finally the space component of the Einstein equations (with $i \neq j$) leads to

$$\frac{k^2}{\mathcal{H}^2}(\Phi - \Psi) = \frac{6k^2}{a\rho_{\rm cri}}\left(\eta_{{\rm b}\gamma}k^2 v_{{\rm b}\gamma} + \eta_\nu k^2 v_\nu\right), \tag{166}$$

where we remind that $\rho_{\rm cri}$ is the critical energy density. At this stage we have all the equations necessary to understand the behavior of the four fluids: we have ten quantities (namely four δ_X, four v_X, Ψ and Φ) and ten equations, namely Eqs. (158)–(166) (the tenth equation is simply $v_{\rm b} = v_\gamma$). The only thing which remains to be done is to specify the initial conditions. Integrating this system of ten equations analytically is not possible (even if linear). This has to be done numerically. However, since we are mainly interested in the behavior of the system on large scales, the problem gets simplified. Indeed, let us introduce the quantity, introduced by Bardeen, Steinhardt and Turner, $\zeta_{\rm BST}$ defined by Bardeen et al. (1983) and Martin and Schwarz (1998)

$$\zeta_{\rm BST} = -\Psi - \frac{\mathcal{H}}{\rho'}\delta\rho = \sum_X \frac{\rho'_X}{\rho'}\zeta_X\,, \tag{167}$$

where $\rho = \sum_X \rho_X$ is the total energy density and ζ_X can be expressed as

$$\zeta_X = -\Psi + \frac{\delta_X}{3\,(1+w_X)}. \tag{168}$$

From Eqs. (158)–(161), we see that, on large scales (where the terms $\propto k^2 v_X$ go to zero), each ζ_X is conserved, namely $\zeta_X' = 0$. Now, we understand the particular role of the conditions (157). Indeed, they amount to simply choose

$$\zeta_{\rm cdm} = \zeta_{\rm b} = \zeta_\gamma = \zeta_\nu \equiv \zeta_{\rm adia}. \tag{169}$$

and, in this case, we have

$$\zeta_{\rm BST} = \zeta_{\rm adia} \sum_X \frac{\rho_X'}{\rho'} = \zeta_{\rm adia}, \tag{170}$$

which is a constant. Therefore, for adiabatic initial conditions, the quantity $\zeta_{\rm BST}$ is conserved on large scales. Another way to see the same thing is to differentiate $\zeta_{\rm BST}$ (using the expression of $\delta\rho'$ obtained from energy conservation). Then, one arrives at the following equation

$$\zeta_{\rm BST}' = -\frac{\mathcal{H}}{\rho+p}\delta p_{\rm nad} - \frac{1}{3}\partial_i\partial^i v^{\rm (gi)}\,, \tag{171}$$

which shows that, on large scales, the conservation of $\zeta_{\rm BST}$ is controlled by the non-adiabatic pressure [here, $v^{\rm (gi)}$ is the scalar component of the gauge-invariant velocity]. This quantity is defined by the following expression

$$\delta p_{\rm nad} = \delta p - c_{\rm s}^2 \delta\rho\,, \tag{172}$$

where $\delta\rho$, δp are the total perturbed energy density and pressure, respectively. The quantity $c_{\rm s}^2 \equiv p'/\rho'$ is the (total) sound velocity. In the case where one has two fluids (in order to keep things simple), expressing the perturbed energy density and the perturbed pressure explicitly, one arrives at

$$\begin{aligned}\delta p_{\rm nad} = {} & \left(\delta p_1 - c_{\rm s1}^2\,\delta\rho_1\right) + \left(\delta p_2 - c_{\rm s2}^2\,\delta\rho_2\right) \\ & + \left(c_{\rm s1}^2 - c_{\rm s2}^2\right)\frac{(\rho_1+p_1)\,(\rho_2+p_2)}{\rho+p}S_{12}\,,\end{aligned} \tag{173}$$

where S_{12} is given by

$$S_{12} = \frac{\delta\rho_1}{\rho_1+p_1} - \frac{\delta\rho_2}{\rho_2+p_2} = 3\,(\zeta_1-\zeta_2)\,. \tag{174}$$

and where $c_{\mathrm{Si}} \equiv p_i'/\rho_i'$. The non-adiabatic pressure contains two contributions. The terms $\delta p_i - c_{\mathrm{Si}}^2 \delta\rho_i$ originate from intrinsic entropy perturbations (if any) of the fluids under consideration while the term proportional to S_{12} represents the entropy of mixing. Let us summarize: for adiabatic perturbations, ζ_{BST} is a conserved quantity. For non adiabatic perturbations, this quantity can evolve even on large scales and this evolution is given by Eq. (167).

Let us also remark that one can work in terms of the quantity ζ defined by Martin and Schwarz (1998) and already introduced before

$$\zeta = \Phi + \frac{2}{3}\frac{\mathcal{H}^{-1}\Phi' + \Phi}{1+\omega}. \tag{175}$$

If one has $\Psi = \Phi$, then

$$\zeta_{\mathrm{BST}} = -\zeta - \frac{k^2}{3\epsilon_1\mathcal{H}^2}\Phi, \tag{176}$$

and, in the standard situation, when there is no entropy perturbations, the quantities ζ and ζ_{BST} are both conserved on super-Hubble scales. Notice that, strictly speaking, ζ stays constant only in absence of shear viscosity.

Let us now try to understand how the presence or the absence of adiabatic perturbations can affect CMB anisotropies. On large scales, the temperature fluctuations can be expressed as

$$\frac{\delta T}{T} \simeq \frac{1}{4}\delta_\gamma|_{\mathrm{lss}} + \Phi|_{\mathrm{lss}}, \tag{177}$$

where "lss" means "last scattering surface" and indicates when the radiation density contrast and the Bardeen potential must be evaluated. Since last scatterings occur during the matter dominated era, using the time-time component of the perturbed Einstein equation, one obtains $-2\Phi|_{\mathrm{lss}} \simeq R_{\mathrm{cdm}}\delta_{\mathrm{cdm}}|_{\mathrm{lss}} + R_{\mathrm{b}}\delta_{\mathrm{b}}|_{\mathrm{lss}}$ where $R_{\mathrm{cdm}} \equiv \rho_{\mathrm{cdm}}/(\rho_{\mathrm{cdm}} + \rho_{\mathrm{b}})$ and $R_{\mathrm{b}} \equiv \rho_{\mathrm{b}}/(\rho_{\mathrm{cdm}} + \rho_{\mathrm{b}})$. It is conventional to measure the non-adiabatic perturbation with respect to photons. Therefore, one introduces the notation $S_X \equiv S_{X\gamma} \equiv 3(\zeta_X - \zeta_\gamma)$. Then, one obtains

$$\Phi|_{\mathrm{lss}} = -\frac{3}{5}\zeta_\gamma - \frac{1}{5}R_{\mathrm{cdm}}S_{\mathrm{cdm}} - \frac{1}{5}R_{\mathrm{b}}S_{\mathrm{b}}, \tag{178}$$

where we recall that R_{b} and R_{cdm} are evaluated at last scattering. Notice also that, in principle, we do not need a subscript "lss" for ζ_X or S_X because they are constant (in time) quantities since ζ_X is conserved. In particular, they should be viewed as the value of ζ_X at the onset of the radiation dominated era, just after inflation and, therefore, S_X could also be written as S_X^{ini} in order to emphasize this point. To calculate the temperature anisotropies, we use Eq. (177) and write

$\delta T/T = \delta_\gamma/4|_{\rm lss} + \Phi_{\rm lss} = \zeta_\gamma + 2\Phi|_{\rm lss}$. As a result, one obtains

$$\frac{\delta T}{T} = -\frac{1}{5}\zeta_\gamma - \frac{2}{5}R_{\rm cdm}S_{\rm cdm} - \frac{2}{5}R_{\rm b}S_{\rm b}. \tag{179}$$

Finally, during the Radiation Dominated (RD) era, one can write

$$\zeta_{\rm RD} = R_\gamma\zeta_\gamma + R_\nu\zeta_\nu = \zeta_\gamma + R_\nu\frac{S_\nu}{3}, \tag{180}$$

with $R_\gamma \equiv \rho_\gamma/(\rho_\gamma + \rho_\nu)$ and $R_\nu \equiv \rho_\nu/(\rho_\gamma + \rho_\nu)$, these quantities being evaluated during the RD era. Using Eq. (180) to obtain an expression of ζ_γ and using this expression in Eq. (179), it follows that

$$\frac{\delta T}{T} = -\frac{1}{5}\zeta_{\rm RD} - \frac{2}{5}R_{\rm cdm}S_{\rm cdm} - \frac{2}{5}R_{\rm b}S_{\rm b} + \frac{1}{15}R_\nu S_\nu, \tag{181}$$

which coincides with Eq. (7) of Gordon and Lewis (2003). The term $\zeta_{\rm RD}/5$ represents the adiabatic contribution. In fact, one can also define an effective isocurvature mode taking into account both cold dark matter and baryons entropy fluctuations by defining

$$S_{\rm cdm}^{\rm eff} \equiv S_{\rm cdm} + \frac{R_{\rm b}}{R_{\rm cdm}}S_{\rm b}, \tag{182}$$

such that Eq. (181) now reads

$$\frac{\delta T}{T} = -\frac{1}{5}\zeta_{\rm RD} - \frac{2}{5}R_{\rm cdm}S_{\rm cdm}^{\rm eff} + \frac{1}{15}R_\nu S_\nu. \tag{183}$$

We therefore have two adiabatic modes that are, as will be seen in Sect. 5, denoted by the Planck collaboration CDI (for the effective cold dark matter) and NDI (for neutrinos). In fact, there is a third mode, NVI, related to neutrinos velocity. Since the expression of the temperature is modified by the presence of isocurvature modes, the temperature multipole moments will also be affected, for concrete and quantitative results see for instance Langlois and Riazuelo (2000). As a consequence, when compared to the CMB data, one can put constraints on their amplitude.

As will be discussed in Sect. 5, so far, CMB measurements are consistent with adiabaticity. This gives non trivial information about inflation. Indeed, if non adiabatic perturbations were observed it would mean that inflation can not be driven by a single scalar field. As for Non-Gaussianities, this would have implied that single-field slow-roll inflation with a standard kinetic term were ruled out. This class of models has therefore passed another non-trivial test. Of course, this is the situation now and this could very well change in the future. In that case, what would be the implications for inflation? A natural explanation would be to have multiple field inflation and we now explain in detail why using a simple

example (Polarski and Starobinsky 1992, 1994; Peter et al. 1994; Langlois 1999; Langlois and Riazuelo 2000; Gordon et al. 2001; Amendola et al. 2002; Bartolo et al. 2001; Wands et al. 2002; Gordon and Lewis 2003; Byrnes and Wands 2006; Wands 2008; Choi et al. 2009).

Assume that, instead of having one field, we now have a collection of fields that all play a role during inflation. For simplicity, and because we want to be explicit, let us consider the case where we have two fields, $\phi_{\rm h}$ and ϕ_ℓ, and where the potential is quadratic for each field, without interaction term, namely $V = V_{\rm h} + V_\ell \equiv m_{\rm h}^2\phi_{\rm h}^2/2 + m_\ell^2\phi_\ell/2$. Then, the equations of motion for the background are given by

$$H^2 = \frac{\kappa}{3}\left(\frac{1}{2}\dot{\phi}_{\rm h}^2 + \frac{1}{2}\dot{\phi}_\ell^2 + \frac{1}{2}m_{\rm h}^2\phi_{\rm h}^2 + \frac{1}{2}m_\ell^2\phi_\ell^2\right), \tag{184}$$

$$\ddot{\phi}_{\rm h} + 3H\dot{\phi}_{\rm h} + m_{\rm h}^2\phi_{\rm h} = 0, \tag{185}$$

$$\ddot{\phi}_\ell + 3H\dot{\phi}_\ell + m_\ell^2\phi_\ell = 0, \tag{186}$$

where, as is standard in the literature, we have used notations that make obvious the fact that one field is heavy and the other light, meaning that $R \equiv m_{\rm h}/m_\ell > 1$ [not to be confused with the R introduced in Eq. (163)]. These equations cannot be solved exactly but one can use the slow-roll approximation. The first Hubble flow parameter is given by, see Eq. (34)

$$\epsilon_1 = \frac{\kappa}{2H^2}\left(\dot{\phi}_{\rm h}^2 + \dot{\phi}_\ell^2\right), \tag{187}$$

and $\epsilon_1 \ll 1$ implies that $\kappa\dot{\phi}_{\rm h}^2/(2H^2) \ll 1$ and $\kappa\dot{\phi}_\ell^2/(2H^2) \ll 1$. These two conditions are similar to what would be obtained in the single-field case. This means that, as usual, the kinetic term can be neglected in the Friedmann equation. On the other hand, the second Hubble flow parameter can be written as

$$\epsilon_2 = 2\epsilon_1 + \frac{2}{H}\frac{\ddot{\phi}_{\rm h}\dot{\phi}_{\rm h} + \ddot{\phi}_\ell\dot{\phi}_\ell}{\dot{\phi}_{\rm h}^2 + \dot{\phi}_\ell^2}. \tag{188}$$

In the single-field case, this relation reduces to $\epsilon_2 = 2\epsilon_1 + 2\ddot{\phi}/(H\dot{\phi})$ and the acceleration in the Klein-Gordon equation can also be neglected since $\epsilon_2 \ll 1$ implies that $\ddot{\phi}/(H\dot{\phi}) \ll 1$. However, in the two-field case, the properties $\ddot{\phi}_{\rm h}/(H\dot{\phi}_{\rm h}) \ll 1$ and $\ddot{\phi}_\ell/(H\dot{\phi}_\ell) \ll 1$ cannot be deduced from $\epsilon_2 \ll 1$. As a consequence, neglecting the acceleration in the Klein-Gordon equations for the heavy and light fields is in fact an additional assumption that we will make in the following. Then, using that $\dot{H} = -\kappa\left(\dot{\phi}_{\rm h}^2 + \dot{\phi}_\ell^2\right)/2$ (which, by the way, shows that the Hubble parameter always decreases) and the slow-roll Klein-Gordon equation to relate the first time derivative of the fields to the derivative of the potential, one obtains

$$\epsilon_1 \simeq 2M_{\rm Pl}^2\frac{R^4\phi_{\rm h}^2 + \phi_\ell^2}{\left(R^2\phi_{\rm h}^2 + \phi_\ell^2\right)^2}. \tag{189}$$

If the heavy field dominates, $R\phi_{\rm h} \gg \phi_\ell$ or, equivalently, $m_{\rm h}\phi_{\rm h} \gg m_\ell\phi_\ell$, then $\epsilon_1 \simeq 2M_{\rm Pl}^2/\phi_{\rm h}^2$. As expected, this expression is similar to that one would obtain in single Large Field Inflation (LFI). And if the light field dominates, i.e. if $\phi_\ell \gg R^2\phi_{\rm h}$, then $\epsilon_1 \simeq 2M_{\rm Pl}^2/\phi_\ell^2$. Therefore, we will assume the following initial conditions which guarantee that slow-roll is valid

$$\phi_{\rm h} \gg \sqrt{2}M_{\rm Pl}, \quad \phi_\ell \gg \sqrt{2}M_{\rm Pl}, \quad R\phi_{\rm h} \gg \phi_\ell. \tag{190}$$

The second condition is a priori less obvious so let us discuss it a little bit more. The domination of the heavy field comes to an end when $R\phi_{\rm h} = \phi_\ell$. At this transition, the first Hubble flow parameter is given by $\epsilon_{1\rm t} = 2M_{\rm Pl}^2(1+R^2)/\phi_{\ell\rm t}^2$. So inflation does not stop provided $\phi_{\ell\rm t} \gg \sqrt{2}\sqrt{1+R^2}M_{\rm Pl} \sim \sqrt{2}RM_{\rm Pl}$. Since the light field is almost constant during the phase dominated by the heavy field (see below), this justifies our initial condition. But it is also possible to consider a situation where inflation stops at the transition, namely $\phi_{\ell\rm t} < \sqrt{2}RM_{\rm Pl}$ (Polarski and Starobinsky 1992, 1994; Peter et al. 1994). After the transition, $\epsilon_1 \simeq 2M_{\rm Pl}^2/\phi_\ell^2$ and if one wants inflation to start again, one needs the condition (190) for the light field.

As long as the slow-roll approximation is valid, the equations of motion can be integrated and the solution for the field vacuum expectation values reads (Polarski and Starobinsky 1992, 1994; Peter et al. 1994)

$$\phi_{\rm h} = \sqrt{\frac{4s}{\kappa}}\sin\left[\theta(s)\right], \quad \phi_\ell = \sqrt{\frac{4s}{\kappa}}\cos\left[\theta(s)\right], \tag{191}$$

while the Hubble parameter is given by

$$H^2(s) = \frac{2s}{3}m_\ell^2\left[1+\left(R^2-1\right)\sin^2\theta\right]. \tag{192}$$

This is a parametric representation of the solution in terms of the variable s defined by $s = -\ln(a/a_{\rm end})$, with

$$s = s_0\frac{(\sin\theta)^{2m_\ell^2/\left(m_{\rm h}^2-m_\ell^2\right)}}{(\cos\theta)^{2m_{\rm h}^2/\left(m_{\rm h}^2-m_\ell^2\right)}}. \tag{193}$$

The initial phase, dominated by the heavy field, corresponds to $\theta \to \pi/2$, $s \to \infty$, $\epsilon_1 \to 0$ and $\phi_{\rm h}/\phi_\ell = \tan\theta \to \infty$. As already mentioned, this happens when $m_{\rm h}\phi_{\rm h} > m_\ell\phi_\ell$ or $\theta < \theta_{\rm t}$ where the "transition angle" $\theta_{\rm t}$ is given by $\tan\theta_{\rm t} \equiv R^{-1}$. The fact that $R > 1$ implies that $\theta_{\rm t} < \pi/4$. If $R \gg 1$ then $\theta_{\rm t} \simeq R^{-1}$ is a small angle. In that case (i.e. $R \gg 1$), in the regime where $\theta \gg \theta_{\rm t}$ (i.e. θ is not a small angle), we have the following behavior for s: $s \simeq s_0\cos^{-2}\theta$. This implies that the heavy and light fields are given by

$$\phi_{\rm h} \simeq \sqrt{\frac{4s_0}{\kappa}}\tan\theta \gg M_{\rm Pl}, \quad \phi_\ell \simeq \sqrt{\frac{4s_0}{\kappa}}. \tag{194}$$

In this regime, the heavy field is super-Planckian and the model is effectively equivalent to large field inflation (LFI). This is confirmed by writing the Friedmann Eq. (192) using Eq. (194)

$$H^2 \simeq \frac{2s}{3} m_\ell^2 R^2 \sin^2\theta = \frac{2s_0}{3} m_\mathrm{h}^2 \tan^2\theta = \frac{\kappa}{6} m_\mathrm{h}^2 \phi_\mathrm{h}^2, \tag{195}$$

which is exactly the Friedmann equation for LFI. On the other hand, as announced above, the light field is frozen and its back-reaction is negligible. This provides an interpretation for the parameter s_0: it is nothing but the vacuum expectation value of the frozen light field. Let us also notice that the condition $\phi_\ell \gg \sqrt{2} M_\mathrm{Pl}$ translates into $s_0 \gg 1/2$. Moreover the condition for avoiding an interruption of inflation reads $s_0 \gg R^2/2$.

Then, the next question is to calculate the behavior of the two scalar fields after the transition. The light field now drives the expansion of space-time. The situation is a little subtle because one can still have $\dot{H} \ll H^2$ but $3H\dot{\phi}_\mathrm{h} \neq -m_\mathrm{h}^2\phi_\mathrm{h}$. In other words, the background still inflates but the heavy field, that has become a test field, is not necessarily in slow-roll. In that case, this, in principle, invalidates Eqs. (191)–(193) since they all assume $\dot{H} \ll H^2$ and the two fields in slow-roll: in other words, having the kinetic terms negligible in the Friedmann equation and only one field in slow-roll is not sufficient to derive Eqs. (191)–(193). In that case, we need to return to the exact Klein-Gordon equation for the heavy field. If we write $\phi_\mathrm{h} = a^{-3/2} f_\mathrm{h}$, then it takes the form

$$\ddot{f}_\mathrm{h} - \left[\frac{3}{2}\left(\dot{H} + \frac{3}{2}H^2\right) - m_\mathrm{h}^2\right] f_\mathrm{h} = 0. \tag{196}$$

Since the background is still in slow-roll, one can neglect the term $\dot{H}$ in the above equation. Then, we see that the behavior of the field depends on the ratio H/m_h. Since H is decreasing, the term proportional to the mass necessarily becomes dominant at some time and then the field oscillates, namely

$$\phi_\mathrm{h} \simeq a^{-3/2} \cos\left(m_\mathrm{h} t\right). \tag{197}$$

The frequency of the oscillations is given by the mass of the field. The amplitude of the oscillations decreases as $\propto a^{-3/2}$ and, therefore, the heavy field becomes negligible very rapidly. During the oscillations of the heavy field, inflation continues driven by the light field. It comes to an end when the vacuum expectation of the light field becomes sub-Planckian.

Having described the behavior of the background, we can now turn to the perturbations. They are described by the Bardeen potential already introduced before, Φ, and the two perturbed scalar fields $\delta\phi_\mathrm{h}$ and $\delta\phi_\ell$. The corresponding

equations of motion read

$$\dot{\Phi} + H\Phi = \frac{\kappa}{2}\left(\dot{\phi}_{\rm h}\delta\phi_{\rm h} + \dot{\phi}_{\ell}\delta\phi_{\ell}\right), \tag{198}$$

$$\delta\ddot{\phi}_{\rm h} + 3H\delta\dot{\phi}_{\rm h} + \left(\frac{k^2}{a^2} + m_{\rm h}^2\right)\delta\phi_{\rm h} = 4\dot{\phi}_{\ell}\dot{\Phi} - 2m_{\ell}^2\Phi\phi_{\ell}, \tag{199}$$

$$\delta\ddot{\phi}_{\ell} + 3H\delta\dot{\phi}_{\ell} + \left(\frac{k^2}{a^2} + m_{\ell}^2\right)\delta\phi_{\ell} = 4\dot{\phi}_{\ell}\dot{\Phi} - 2m_{\ell}^2\Phi\phi_{\ell}. \tag{200}$$

Unfortunately, this system of equations cannot be solved analytically. However, on large scales, namely for wavelengths larger than the Hubble radius, the expression of the growing mode of the Bardeen potential and of the two perturbed scalar fields can be established. They read (Polarski and Starobinsky 1992, 1994; Peter et al. 1994)

$$\Phi = -C_1\frac{\dot{H}}{H^2} - H\frac{\rm d}{{\rm d}t}\left(\frac{d_{\rm h}V_{\rm h} + d_{\ell}V_{\ell}}{V_{\rm h} + V_{\ell}}\right), \tag{201}$$

$$\frac{\delta\phi_{\rm h}}{\dot{\phi}_{\rm h}} = \frac{C_1}{H} - 2H\left(\frac{d_{\rm h}V_{\rm h} + d_{\ell}V_{\ell}}{V_{\rm h} + V_{\ell}} - d_{\rm h}\right), \tag{202}$$

$$\frac{\delta\phi_{\ell}}{\dot{\phi}_{\ell}} = \frac{C_1}{H} - 2H\left(\frac{d_{\rm h}V_{\rm h} + d_{\ell}V_{\ell}}{V_{\rm h} + V_{\ell}} - d_{\ell}\right), \tag{203}$$

where $C_1(k)$, $d_{\rm h}(k)$ and $d_{\ell}(k)$ are integration constants. At this point, the following remark is in order. We have seen that, in the theory of cosmic inflation, the source of the perturbations are the quantum vacuum fluctuations. This of course remains true in a model where we have several scalar fields. This means that the quantities Φ, $\delta\phi_{\rm h}$ and $\delta\phi_{\ell}$ are in fact quantum operators. A convenient way to describe this situation without introducing all the machinery of quantum field theory is simply to write that the amplitude of the perturbed fields at Hubble radius crossing are given by $\delta\phi_{\rm h} = H/\sqrt{2k^3}e_{\rm h}(\boldsymbol{k})$ and $\delta\phi_{\ell} = H/\sqrt{2k^3}e_{\ell}(\boldsymbol{k})$, where $e_{\rm h}(\boldsymbol{k})$ and $e_{\ell}(\boldsymbol{k})$ are two independent Gaussian stochastic processes satisfying $\langle e_{\rm h}(\boldsymbol{k})\rangle = \langle e_{\ell}(\boldsymbol{k})\rangle = 0$ and $\langle e_{\rm h}(\boldsymbol{k})e_{\rm h}(\boldsymbol{k}')\rangle = \delta^{(3)}(\boldsymbol{k} - \boldsymbol{k}')$, $\langle e_{\ell}(\boldsymbol{k})e_{\ell}(\boldsymbol{k}')\rangle = \delta^{(3)}(\boldsymbol{k} - \boldsymbol{k}')$, $\langle e_{\rm h}(\boldsymbol{k})e_{\ell}(\boldsymbol{k}')\rangle = 0$. This parametrization raises in fact non trivial questions such as the quantum-to-classical transition of quantum cosmological perturbations but, in this review, we will not discuss these issues (Grishchuk and Martin 1997; Polarski and Starobinsky 1996; Martin et al. 2012).

Then, let us simplify the expression of the perturbed heavy scalar field, see Eq. (202), by using the explicit form of the potential. One arrives at

$$\frac{\delta\phi_{\rm h}}{\dot{\phi}_{\rm h}} = = \frac{C_1}{H} - 2H\left(d_\ell - d_{\rm h}\right)\frac{V_\ell}{V_{\rm h}+V_\ell} \tag{204}$$

$$= \frac{C_1}{H} - 2H\left(d_\ell - d_{\rm h}\right)\frac{m_\ell^2\phi_\ell^2}{m_{\rm h}^2\phi_{\rm h}^2 + m_\ell^2\phi_\ell^2} \tag{205}$$

$$\simeq \frac{C_1}{H} + 2HC_3, \tag{206}$$

where $C_3 \equiv d_{\rm h} - d_\ell$ and where, in the last equality, we have assumed that the light field was dominant (namely the second phase of inflation). Then, one can use the slow-roll relation $3H\dot{\phi}_{\rm h} \simeq -m_{\rm h}^2\phi_{\rm h}$ and obtains

$$\delta\phi_{\rm h} \simeq -\frac{C_1}{3}\frac{m_{\rm h}^2}{H^2}\phi_{\rm h} - \frac{2}{3}C_3 m_{\rm h}^2\phi_{\rm h} \simeq -\frac{2}{3}C_3 m_{\rm h}^2\phi_{\rm h}, \tag{207}$$

where, in the last equality, we have used the fact that, before the onset of oscillations, $H \gg m_{\rm h}$. Then, as already mentioned above, the field starts oscillating and the slow-roll approximation is no longer valid. As a consequence, the above equations can no longer be used. During the oscillations, one has equipartition between the kinetic and potential energy. As a consequence, $\langle\rho_{\rm h}\rangle \simeq m^2\langle\phi_{\rm h}^2\rangle$. This implies that $\delta\rho_{\rm h} \simeq m_{\rm h}^2 2\phi_{\rm h}\delta\phi_{\rm h}$ and, therefore,

$$\frac{\delta\rho_{\rm h}}{\rho_{\rm h}} \simeq 2\frac{\delta\phi_{\rm h}}{\phi_{\rm h}}. \tag{208}$$

But, in fact, the perturbed Klein-Gordon equation for large scales modes, if one neglects its right hand side, is the same as the background Klein-Gordon equation provided the potential is quadratic in the field (which is precisely the case in the present situation). As a consequence, $\delta\phi_{\rm h}$ is in fact always proportional to $\phi_{\rm h}$, the slow-roll approximation being satisfied or not. In other words, $\delta\phi_{\rm h}/\phi_{\rm h}$ and hence $\delta\rho_{\rm h}/\rho_{\rm h}$ are constant. So if we assume that the heavy field decays into cold dark matter after its oscillations, one has (Polarski and Starobinsky 1992, 1994; Peter et al. 1994)

$$\delta_{\rm cdm} = \left.\frac{\delta\rho_{\rm h}}{\rho_{\rm h}}\right|_{\rm end\ osci} = \left.\frac{\delta\rho_{\rm h}}{\rho_{\rm h}}\right|_{\rm start\ osci} = 2\left.\frac{\delta\phi_{\rm h}}{\phi_{\rm h}}\right|_{\rm start\ osci} = -\frac{4}{3}C_3(k)m_{\rm h}^2. \tag{209}$$

We conclude that, if one is able to express the constant $C_3(k)$, then one can establish the expression of the cold dark matter density contrast. But this is in fact an easy task. Indeed, reproducing the same calculation as the one which led to Eq. (205),

one obtains

$$\frac{\delta\phi_\ell}{\dot{\phi}_\ell} = \frac{C_1}{H} - 2HC_3\frac{m_{\rm h}^2\phi_{\rm h}^2}{m_{\rm h}^2\phi_{\rm h}^2 + m_\ell^2\phi_\ell^2}. \tag{210}$$

Then, using this formula and Eq. (205), one can eliminate $C_1(k)$ and find an expression for $C_3(k)$. Straightforward manipulations lead to

$$C_3(k) = -\frac{1}{2H}\left(\frac{\delta\phi_\ell}{\dot{\phi}_\ell} - \frac{\delta\phi_{\rm h}}{\dot{\phi}_{\rm h}}\right). \tag{211}$$

The next step is to replace the derivatives of the fields by their slow-roll expressions and $\delta\phi_{\rm h}$ by $He_{\rm h}/\sqrt{2k^3}$ (and a similar expression for $\delta\phi_\ell$) as was discussed after Eq. (203). One arrives at

$$C_3(k) = \frac{3H}{2m_{\rm h}^2}\frac{1}{\sqrt{2k^3}}\left(\frac{m_{\rm h}^2}{m_\ell^2}\phi_\ell^{-1}e_\ell - \phi_{\rm h}^{-1}e_{\rm h}\right). \tag{212}$$

Finally, one can estimate the entropy perturbation. According to the definitions introduced before, see Eq. (174), one has $S_{\rm cdm} \equiv S_{{\rm cdm}\gamma} \equiv 3\left(\zeta_{\rm cdm} - \zeta_\gamma\right) = \delta_{\rm cdm} - 4\delta_\gamma/3 \simeq \delta_{\rm cdm}$ because $\delta\rho_{\rm cdm} \sim \delta\rho_\gamma$ and, during the radiation dominated era, $\rho_\gamma \gg \rho_{\rm cdm}$. As a consequence, one has

$$S_{{\rm cdm},\boldsymbol{k}} \simeq -\sqrt{\frac{2}{k^3}}H\left[R^2\phi_\ell^{-1}e_\ell(\boldsymbol{k}) - \phi_{\rm h}^{-1}e_{\rm h}(\boldsymbol{k})\right], \tag{213}$$

where the quantities H, $\phi_{\rm h}$ and ϕ_ℓ should be viewed here as scale dependent quantities since they are expressed at Hubble radius crossing. In fact, as will be discussed in Sect. 5, their scale dependence permits the calculation of the isocurvature perturbations power spectrum. With these equations, one can now predict the CMB temperature anisotropies by using Eq. (181). But, in fact, the most important conclusion is of course that $S_{{\rm cdm},\boldsymbol{k}} \neq 0$. This means that, in a model of inflation with more than one field, isocurvature perturbations can be produced. This justifies our claim that, if non-adiabatic perturbations are observed in the future, a natural explanation will be to consider that several scalar fields play a role during inflation.

In fact, there is even more. Indeed, during the radiation dominated era, the adiabatic perturbations can be written as $\zeta_{\rm RD} = 3\Phi/2 = C_1(k)$. The constant $C_1(k)$ can also be evaluated easily using the solutions (202) and (203). The corresponding expression reads

$$\zeta_{\rm RD} = -\frac{\kappa}{2}\frac{H}{\sqrt{2k^3}}\left[\phi_{\rm h}e_{\rm h}(\boldsymbol{k}) + \phi_\ell e_\ell(\boldsymbol{k})\right]. \tag{214}$$

In particular, one has introduced before the power spectrum of the conserved quantity $\zeta_{\boldsymbol{k}}$ according to $\langle \zeta_{\boldsymbol{k}_1} \zeta^*_{\boldsymbol{k}_2} \rangle = 2\pi^2/k_1^3 \mathcal{P}(k_1)\delta^{(3)}(\boldsymbol{k}_1 - \boldsymbol{k}_2)$.[3] In the same manner, one can define the power spectrum of the non-adiabatic perturbations by

$$\langle S_{\mathrm{cdm},\boldsymbol{k}_1} S^*_{\mathrm{cdm},\boldsymbol{k}_2} \rangle \equiv \frac{2\pi^2}{k_1^3} \mathcal{P}_{S_{\mathrm{cdm}}}(k_1)\delta^{(3)}\left(\boldsymbol{k}_1 - \boldsymbol{k}_2\right). \tag{215}$$

But the most important aspect of the above calculations is that adiabatic and isocurvature perturbations turn out to be correlated (Langlois 1999; Langlois and Riazuelo 2000). This means that the correlator $\langle \zeta_{\boldsymbol{k}_1} S^*_{\mathrm{cdm},\boldsymbol{k}_2} \rangle$ is non-vanishing. This correlator can be expressed as

$$\Re \langle \zeta_{\boldsymbol{k}_1} S^*_{\mathrm{cdm},\boldsymbol{k}_2} \rangle \equiv \frac{2\pi^2}{k_1^3} \mathcal{C}_{\zeta,S_{\mathrm{cdm}}}\delta^{(3)}\left(\boldsymbol{k}_1 - \boldsymbol{k}_2\right). \tag{216}$$

This is because the expressions of $\zeta_{\boldsymbol{k}}$, see Eq. (214), and $S_{\mathrm{cdm},\boldsymbol{k}}$, see Eq. (213), both depend on e_{h} and e_ℓ. From the above definition, let us also notice that one can define a correlation spectrum by

$$\mathcal{P}_{\zeta,S_{\mathrm{cdm}}} \equiv \frac{\mathcal{C}_{\zeta,S_{\mathrm{cdm}}}}{\sqrt{\mathcal{P}_\zeta}\sqrt{\mathcal{P}_{S_{\mathrm{cdm}}}}}. \tag{217}$$

Let us stress that, when one constrains the amplitude of isocurvature modes using the CMB data, it is of course important to take into account the fact that adiabatic and isocurvature perturbations can be correlated. As we will see in the next section, this was done in the analysis of the Planck data.

5 Inflation After Planck

We have previously studied the predictions of inflation for different cosmological observables. In this section, we review what is experimentally known about these observables and discuss the corresponding implications for cosmic inflation.

The Planck CMB data have been released for the first time in 2013 (Ade et al. 2014a) and, more recently, in 2015, new measurements have been made public (Planck 2015). Planck 2013 has measured the CMB temperature anisotropies and the corresponding multipole moments C_ℓ^{TT} are represented in Fig. 4. Let us remind that these quantities are defined as follows. After foregrounds subtraction, the Planck measurements can be used to construct a map of the CMB temperature

[3] Notice that this expression is consistent with the definition given above, in the text between Eqs. (73) and (74), namely $\langle \zeta_{\boldsymbol{k}_1} \zeta_{\boldsymbol{k}_2} \rangle = 2\pi^2/k_1^3 \mathcal{P}(k_1)\delta^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2)$ because $\zeta_{-\boldsymbol{k}_2} = \zeta^*_{\boldsymbol{k}_2}$.

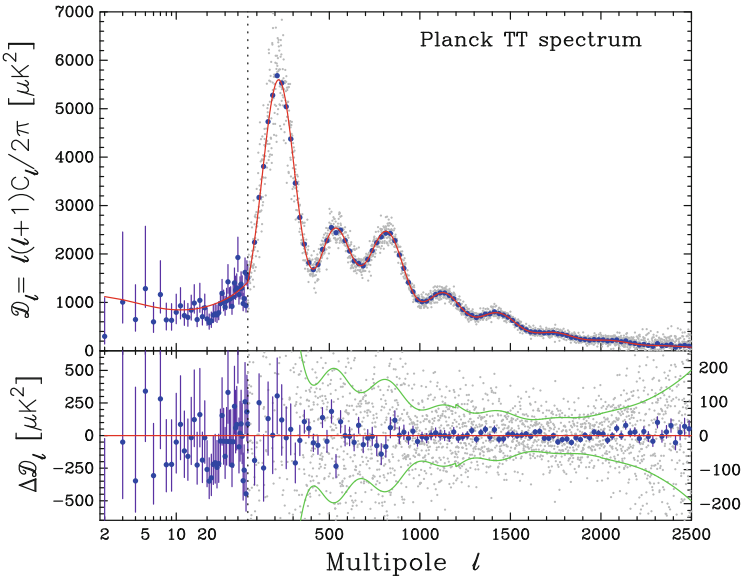

Fig. 4 Temperature anisotropy multipole moments obtained from the Planck 2013 data versus the angular scale ℓ (notice that, for $\ell \leq 49$, the scale is logarithmic). The *gray points* denote the value of the multipole C_ℓ for each ℓ while the *blue points* represent the value of C_ℓ averaged in bands of width $\Delta\ell \simeq 31$. The *red solid line* shows the prediction of the best fit six-parameters ΛCDM model. The error bars correspond to $\pm 1\sigma$ uncertainties. The *lower panel* shows the residual signal once the best fit model has been subtracted. Figure taken from Ade et al. (2014a)

anisotropy, namely

$$\frac{\delta T}{T}(\boldsymbol{e}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\boldsymbol{e}), \tag{218}$$

where $Y_{\ell m}$ are the spherical harmonics and where the vector $\boldsymbol{e}$ specifies a direction in the sky. In practice, $\delta T/T$ can be expressed as

$$\frac{\delta T}{T}(\boldsymbol{e}) = \int \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^{3/2}} \left[F(\boldsymbol{k}) + G(\boldsymbol{k}) \frac{\partial}{\partial \eta_0} \right] \mathrm{e}^{i d_\mathrm{A} \boldsymbol{k}\cdot\boldsymbol{e}/a(\eta_\mathrm{lss})}, \tag{219}$$

where $d_\mathrm{A} = a(\eta_\mathrm{lss}) r_0 + a(\eta_\mathrm{lss})\,(\eta_0 - \eta_\mathrm{lss})$ (r_0 being Earth's radial coordinate and η_0 denoting the present time) is the angular distance to the surface of last scattering and the quantity $\boldsymbol{k}/a(\eta_\mathrm{lss})$ represents the physical wavenumber of the Fourier mode under consideration at the time of recombination. The quantities $F(\boldsymbol{k})$ and $G(\boldsymbol{k})$ encode the behavior of cosmological perturbations and are called "form factors" in Weinberg (2002). Already at this stage, we see that the configuration where the wavelengths of the perturbations become equal to the angular distance of the last scattering surface plays a preferred role. Then, the two-point correlation function in

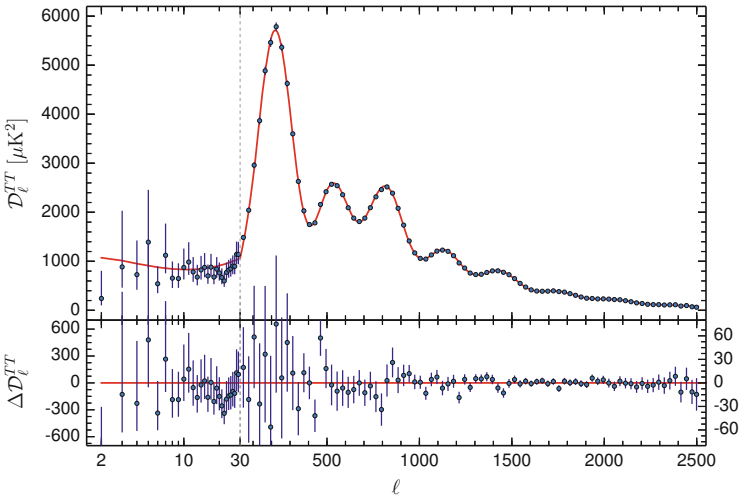

Fig. 5 Same as Fig. 4 but with the Planck 2015 data. Notice that the quantity $\mathcal{D}_\ell$ is defined by $\mathcal{D}_\ell = \ell(\ell+1)C_\ell/(2\pi)$. This plot should be compared to Fig. 4. Figure taken from Planck (2015)

real space can be written as

$$\left\langle \frac{\delta T}{T}(\boldsymbol{e}_1)\frac{\delta T}{T}(\boldsymbol{e}_2)\right\rangle = \frac{1}{4\pi}\sum_{\ell=0}^{+\infty}(2\ell+1)C_\ell^{\mathrm{TT}}P_\ell(\cos\theta), \tag{220}$$

where θ is the angle between the two vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. This expression defines the multipole moments C_ℓ^{TT}.

The big novelty of the Planck 2015 data (Planck 2015; Ade et al. 2015a) is that they not only lead to a more accurate measurements of the C_ℓ^{TT}, see Fig. 5, but they also provide measurements of the E-mode CMB polarization. One can then define quantities similar to the C_ℓ^{TT} for the correlation between temperature and E-mode polarization fluctuations and for the E-mode power spectrum. The corresponding multipole moments C_ℓ^{TE} and C_ℓ^{EE} are represented in Figs. 6 and 7.

Before focusing on the consequences of these data for inflation, let us briefly discuss their implications for the standard model of Cosmology. It is important to understand that the constraints on the cosmological parameters can depend on the model analyzed and on the data used to perform the analysis. In 2013, Planck used the temperature anisotropy measurement plus the WMAP polarization measurement on large scales ($\ell \leq 23$), the corresponding likelihood function being denoted PlanckTT+WP. In 2015, at least five different likelihoods have been used: PlanckTT utilizes temperature data only and is an hybrid, meaning that the temperature likelihood is not the same for low multipoles ($\ell \leq 30$) and high multipoles; PlanckTT+lowP makes use of PlanckTT and low-ℓ polarization data; PlanckTE+lowP corresponds to the TE likelihood at $\ell \geq 30$ plus low-ℓ polarization

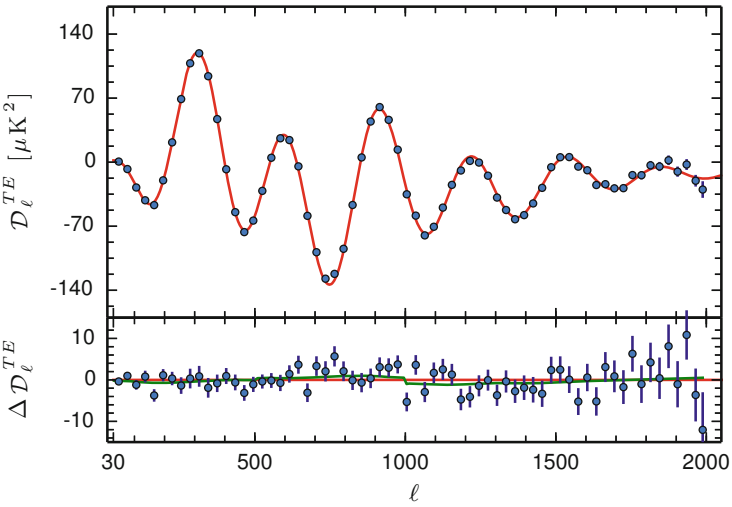

Fig. 6 Multipole moments corresponding to the correlation between temperature and *E*-mode polarization anisotropies obtained from Planck 2015. The *red solid line* corresponds to the best fit obtained with temperature measurements only. The *lower panel* shows the residual with respect to this best fit. Figure taken from Planck (2015)

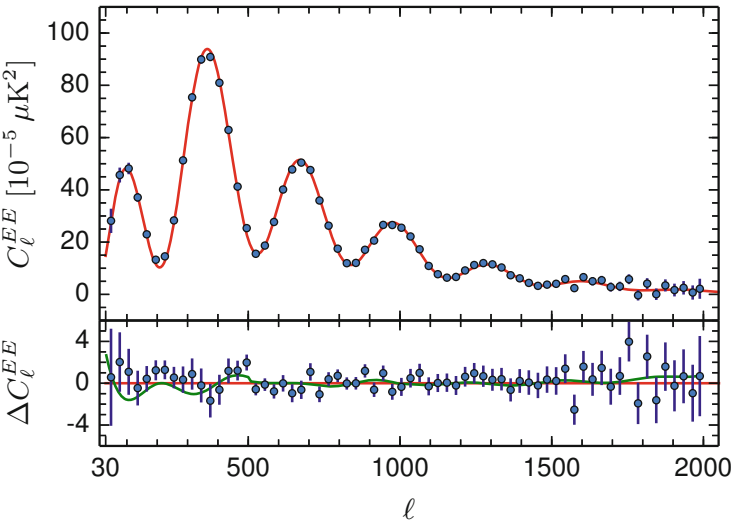

Fig. 7 Same as in Fig. 6 but for the *E*-mode power spectrum obtained from Planck 2015. Figure taken from Planck (2015)

data only; PlanckTT,TE,EE +lowP makes use of the TT, TE and EE likelihoods at $\ell \geq 30$ and of the temperature and polarization data at small scales. Depending on which likelihood is used, the constraints on cosmological parameters can slightly change.

The theoretical framework used to analyze the data is the flat [i.e. $\mathcal{K} = 0$ in Eq. (1)] ΛCDM model. In order to specify it, we need to know the energy budget of the Universe, i.e. the photon energy density ρ_γ, the neutrino energy density ρ_ν, the baryons energy density ρ_b, the cold dark matter energy density ρ_c and the dark energy density ρ_Λ (here assumed to be a cosmological constant). Then, in principle, on can calculate the behavior of the scale factor $a(t)$ since we know $\rho \equiv \sum_i \rho_i$ in the right hand side of the Friedmann Eq. (1). Of course we also need the Hubble rate today, H_0 or $h \equiv H_0/(100\,\mathrm{km} \times \mathrm{s}^{-1} \times \mathrm{Mpc}^{-1})$ which is, therefore, another free parameter. We also need the power spectrum of scalar fluctuations assumed to be of the power-law form $P(k) \propto A_\mathrm{s} k^{n_\mathrm{s}-1}$ where A_s is the amplitude of the fluctuations and n_s the spectral index. Here, gravitational waves are supposed to be absent, $r = 0$. So, in this simple framework, the perturbations are characterized by two numbers. In the case of inflation, we need three parameters, the amplitude of scalar fluctuations and the two first slow-roll parameters. We notice (again!) that the parametrization used here is different from what we generically obtain from inflation where the power spectrum is not of the power-law form and where r is necessarily non-vanishing (but can be very small). Finally, we need a parameter describing reionization and we take the optical depth τ. The interpretation of this parameter is as follows. After recombination, the photons are supposed to propagate freely from the surface of last scattering to us. However, at the epoch of the formation of the first stars, estimated to be $z_\mathrm{re} \sim 10$, the Universe is ionized again. As a consequence, some of the CMB photons scatter off free electrons again. The probability to "avoid" this additional scattering is $\mathrm{e}^{-\tau}$ where

$$\tau \equiv \sigma_\mathrm{T} \int_{t_\mathrm{re}}^{t_\mathrm{now}} n_\mathrm{e} \mathrm{d}t, \tag{221}$$

is the optical depth. In this expression, σ_T is the Thomson cross-section and n_e is the number density of free electrons. When additional scattering occur, the direction of the photon change randomly and this washes out the CMB anisotropy on small angular scales, $\ell > \ell_\mathrm{re}$, namely $a_{\ell m} \to a_{\ell m}\mathrm{e}^{-\tau}$. For $\ell < \ell_\mathrm{re}$, the CMB anisotropies are not changed. The value of ℓ_re clearly depends on z_re. The previous considerations imply that, on small scales, the amplitude of the fluctuations becomes $A_\mathrm{s}\mathrm{e}^{-2\tau}$ and there is therefore a partial degeneracy between A_s and τ.

In the Planck papers, one of the free parameters is in fact taken to be θ_MC, where the subscript "MC" reminds that this quantity is used in COSMOMC. By definition, it is equal to be $\theta_\mathrm{MC} \equiv 100(r_\mathrm{s}/d_\mathrm{A})|_\mathrm{approx}$. Here, r_s is the sound horizon at last scattering, namely

$$r_\mathrm{s} = \int_0^{\eta_\mathrm{lss}} c_\mathrm{s} \mathrm{d}\eta, \tag{222}$$

where $c_{\rm s}$ is the sound speed of the baryon-photon fluid, i.e.

$$c_{\rm s}^2 = \frac{\delta p_{{\rm b}-\gamma}}{\delta\rho_{{\rm b}-\gamma}} = \frac{\delta p_\gamma}{\delta\rho_{\rm b}+\delta\rho_\gamma} = \frac{1}{3}\frac{4\rho_\gamma}{4\rho_\gamma+3\rho_{\rm b}} = \frac{1}{3}\frac{1}{1+R}, \tag{223}$$

where $R = 3\rho_{\rm b}/(4\rho_\gamma)$. The quantity $d_{\rm A}$ is, as already mentioned, the angular distance to the last scattering surface and naturally appears in the expression of the multipole moments, see Eq. (219). Therefore, $r_{\rm s}/d_{\rm A}$ is in fact the angular size of the sound horizon. $\theta_{\rm MC}$ is defined approximately because its value is calculated at a redshift which is given by a fitting formula (Hu and Sugiyama 1996)

$$z_{\rm lss} = 1048\left[1+0.00124\left(\Omega_{\rm b}h^2\right)^{-0.738}\right]\left[1+g_1\left(\Omega_{\rm m}h^2\right)^{g_2}\right], \tag{224}$$

where the function g_1 and g_2 can be expressed as

$$g_1 = 0.0783\left(\Omega_{\rm b}h^2\right)^{-0.238}\left[1+39.5\left(\Omega_{\rm m}h^2\right)^{0.763}\right]^{-1}, \tag{225}$$

$$g_2 = 0.560\left[1+21.1\left(\Omega_{\rm b}h^2\right)^{1.81}\right]^{-1}. \tag{226}$$

with $\Omega_{\rm m} = \Omega_{\rm b}+\Omega_{\rm c}$. In practice, instead of including h in the list of free parameters, we consider $\theta_{\rm MC}$.

We conclude that, a priori, we have a 9 parameters: h or $\theta_{\rm MC}$, ρ_γ, ρ_ν, $\rho_{\rm b}$, $\rho_{\rm c}$, ρ_Λ, $A_{\rm s}$, $n_{\rm s}$ and τ. However, the photon energy density is not a free parameter because it is given by $\pi^2T_0^4/15$ where $T_0 = 2.7255 \pm 0.00006$ K is the CMB temperature. In the same way, the neutrino energy density is fixed since $\rho_\nu = N_{\rm eff}(7/8)(4/11)^{4/3}\rho_\gamma$ with $N_{\rm eff} = 3$. Moreover, the fact that the spatial sections are assumed to be flat means that, say ρ_Λ, can be deduced from the knowledge of the other parameters. Therefore, the "base" model used in the Planck articles is in fact a six-parameter scenario and it is sufficient to fit the CMB data.

Planck 2013 (i.e. PlanckTT+WP using the terminology introduced before) found the following results (68 % confidence limits) (Ade et al. 2014a)

$$\Omega_{\rm b}h^2 = 0.022032 \pm 0.00028, \quad \Omega_{\rm c}h^2 = 0.1199 \pm 0.0027, \tag{227}$$

$$100\theta_{\rm MC} = 1.04131 \pm 0.00063, \quad \tau = 0.089^{+0.012}_{-0.014}, \tag{228}$$

$$n_{\rm s} = 0.9603 \pm 0.0073, \quad \ln\left(10^{10}A_{\rm s}\right) = 3.089^{+0.024}_{-0.027}. \tag{229}$$

On the other hand, Planck 2015 with PlanckTT, TE, EE+lowP (as already mentioned, using the other likelihoods described before would lead to slightly different numbers) gives (Planck 2015)

$$\Omega_{\rm b}h^2 = 0.02225 \pm 0.00016, \quad \Omega_{\rm c}h^2 = 0.1198 \pm 0.0015, \tag{230}$$

$$100\theta_{\rm MC} = 1.04077 \pm 0.00032, \quad \tau = 0.079 \pm 0.017, \tag{231}$$

$$n_{\rm s} = 0.9645 \pm 0.0049, \quad \ln\left(10^{10}A_{\rm s}\right) = 3.094 \pm 0.0049. \tag{232}$$

The consistency between Planck 2013 and Planck 2015 is evidently very good.

More involved data analysis can be carried out by opening the parameter space (for instance by considering gravitational waves, a running for the scalar power spectrum, a time-dependent dark energy equation of state etc ...) and/or adding more data sets. In the following, we will describe the corresponding results for the observables that are especially relevant for inflation.

5.1 Spatial Curvature

As discussed in Sect. 3.1, see Eq. (23), maybe the most important prediction of inflation is that our Universe should be spatially flat (although there are contrived inflationary models for which this is not true (Linde and Mezhlumian 1995)). Therefore, one can follow the strategy described above and relax the assumption that the curvature of spacelike sections is flat. Then, the Planck 2013 data plus the WMAP data on large scale polarization imply that (Ade et al. 2014a,b)

$$\Omega_{\mathcal{K}} = -0.058^{+0.046}_{-0.026}. \tag{233}$$

If, in addition, Baryonic Acoustic Oscillations (BAO) data are included, one obtains $\Omega_{\mathcal{K}} = -0.004 \pm 0.0036$.

Planck (2015) results have confirmed and tightened this conclusion. Indeed, at 95 % confidence level, PlanckTT,TE,EE+lowP leads to $\Omega_{\mathcal{K}} = -0.040^{+0.038}_{-0.041}$. If lensing data and BAO are taken into account, one arrives at the impressive following result

$$\Omega_{\mathcal{K}} = 0.000 \pm 0.005. \tag{234}$$

Therefore, we live in a spatially flat Universe in agreement with one of the most basic prediction of inflation.

As already mentioned, when one relaxes the assumption that the Universe is spatially flat, this introduces a new parameter and, therefore, we are no longer in the framework of the six-parameters ΛCDM base model considered before. As a consequence, a priori, the constraints on the other parameters may change. This is in particular the case of the spectral index n_{s} and its significant deviation from the scale-invariant case which is very important for inflation. However, Ade et al. (2015a) has shown that in the framework where $\Omega_{\mathcal{K}} \neq 0$ and where tensor modes are present, the constraint on n_{s} becomes

$$n_{\mathrm{s}} = 0.969 \pm 0.005, \tag{235}$$

using PlanckTT,TE,EE+lowP. The conclusion that the scale invariant case is ruled out seems therefore robust. In fact, Ade et al. (2015a) has shown that this conclusion is valid for other type of extensions such as different relativistic degrees of freedom (the parameter N_{eff} defined above, running, dark energy equation of state etc ...). This is of course crucial for inflation.

5.2 Isocurvature Perturbations

Let us now investigate the Planck constraints on isocurvature perturbations (Ade et al. 2014a,b, 2015a). We have discussed before, in Sect. 4.4, two types of isocurvature perturbations. Firstly, there is the effective mode taking into account cold dark matter and baryons entropy fluctuations, see Eq. (182) denoted, as already mentioned, CDI in the Planck papers. Secondly, there is also the Neutrino Density Isocurvature (NDI) mode and the Neutrino Velocity Isocurvature (NVI) mode. Each mode is characterized by its power spectrum as in Eq. (215) and each cross term can also be described by the correlation spectrum as in Eq. (217). Therefore, the most general situation can be parametrized by the 4×4 matrix $\mathcal{P}_{ab}(k)$ where $a = \zeta, S_{\rm CDI}, S_{\rm NDI}, S_{\rm NVI}$ with the convention that $\mathcal{P}_{\zeta\zeta} \equiv \mathcal{P}_{\zeta}$ and similar expressions for the diagonal terms. Of course, this matrix is symmetrical.

Usually, only a 2×2 matrix is analyzed and a power law is assumed for each of the power spectra with independent spectral index. But this is not the route followed by the Planck team. Instead, they have assumed the following phenomenological form for $\mathcal{P}_{ab}(k)$

$$\mathcal{P}_{ab}(k) = \exp\left[\left(\frac{\ln k - \ln k_2}{\ln k_1 - \ln k_2}\right)\ln \mathcal{P}_{ab}(k_1) + \left(\frac{\ln k - \ln k_1}{\ln k_2 - \ln k_2}\right)\ln \mathcal{P}_{ab}(k_2)\right], \tag{236}$$

where the two scales k_1 and k_2 are chosen to be $k_1 = 2 \times 10^{-3}\,{\rm Mpc}^{-1}$ and $k_2 = 0.1\,{\rm Mpc}^{-1}$ so that the entire Planck window is spanned. The positive definiteness of the matrix requires $(\mathcal{P}_{ab})^2 \leq \mathcal{P}_{aa}\mathcal{P}_{bb}$.

Then, the following quantities are defined

$$\alpha_{ab}\,(\ell_{\rm min}, \ell_{\rm max}) = \frac{(\Delta T)^2_{ab}\,(\ell_{\rm min}, \ell_{\rm max})}{(\Delta T)^2_{\rm tot}\,(\ell_{\rm min}, \ell_{\rm max})}, \tag{237}$$

where

$$(\Delta T)^2_{ab}\,(\ell_{\rm min}, \ell_{\rm max}) = \sum_{\ell=\ell_{\rm min}}^{\ell=\ell_{\rm max}} (2\ell + 1)\, C^{\rm TT}_{ab,\ell}. \tag{238}$$

In this expression, the quantity $C^{\rm TT}_{ab,\ell}$ represents the multipole moments calculated with the primordial spectrum taken to be $\mathcal{P}_{ab}$. $(\Delta T)^2_{\rm tot}$ is just the sum of all contributions. So, in the standard situation, there is just one contribution and the multipole moments are computed with $\mathcal{P}_{\zeta\zeta} = \mathcal{P}_{\zeta}$. If one has $\alpha_{\zeta\zeta} = 1$, this means that the perturbations are fully adiabatic.

In Figs. 8 and 9, we have respectively represented the one-dimensional posterior distribution of $\alpha_{ab}\,(\ell_{\rm min}, \ell_{\rm max})$ and the two-dimensional distribution for the power spectra $\mathcal{P}_{ab}(k_{1,2})$ for the three modes, CDI, NDI and NVI obtained from Planck

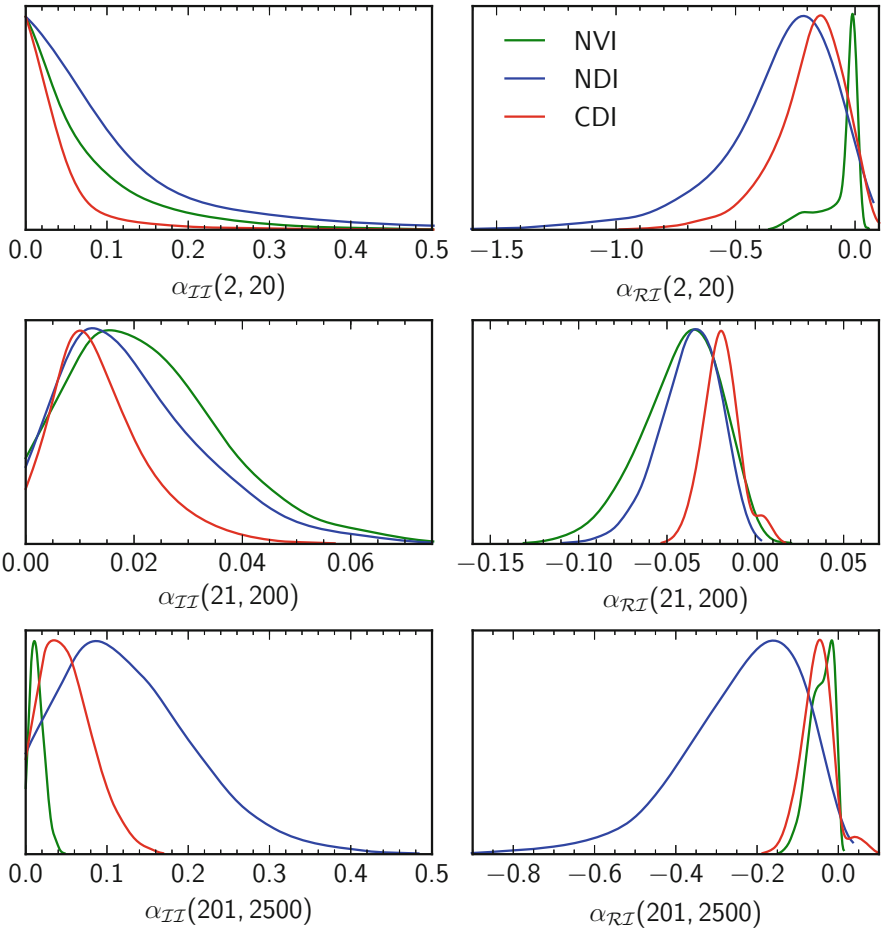

Fig. 8 Posterior distributions for the quantities $\alpha_{ab}\,(\ell_{\rm min}, \ell_{\rm max})$ introduced in Eq. (237), inferred from the Planck 2013 data. No statistically significant deviation from adiabaticity is found. Notice that $\mathcal{R} = -\zeta$ is used in this plot. Figure taken from Ade et al. (2014b)

2013. The conclusion is clear: there is no statistically significant deviation from pure adiabaticity.

In Ade et al. (2015a), the constraints on isocurvature modes implied by the Planck 2015 data have been derived. This work is particularly interesting since one expects the polarization data to have a good constraining power on the amplitude of the isocurvature modes. In this analysis, uniform priors for $\mathcal{P}_{\zeta\zeta}(k_1)$ and $\mathcal{P}_{\zeta\zeta}(k_2)$ are assumed in the range $\left[10^{-9}, 10^{-8}\right]$. For the power spectrum of the isocurvature power spectra, the same choice is made in the range $\left[0, 10^{-8}\right]$. Finally, the adiabatic-isocurvature correlation function at $k = k_1$ is taken in the range $\left[-10^{-8}, 10^{8}\right]$. The same quantity, but at $k = k_2$, is fixed through an assumption about the correlation

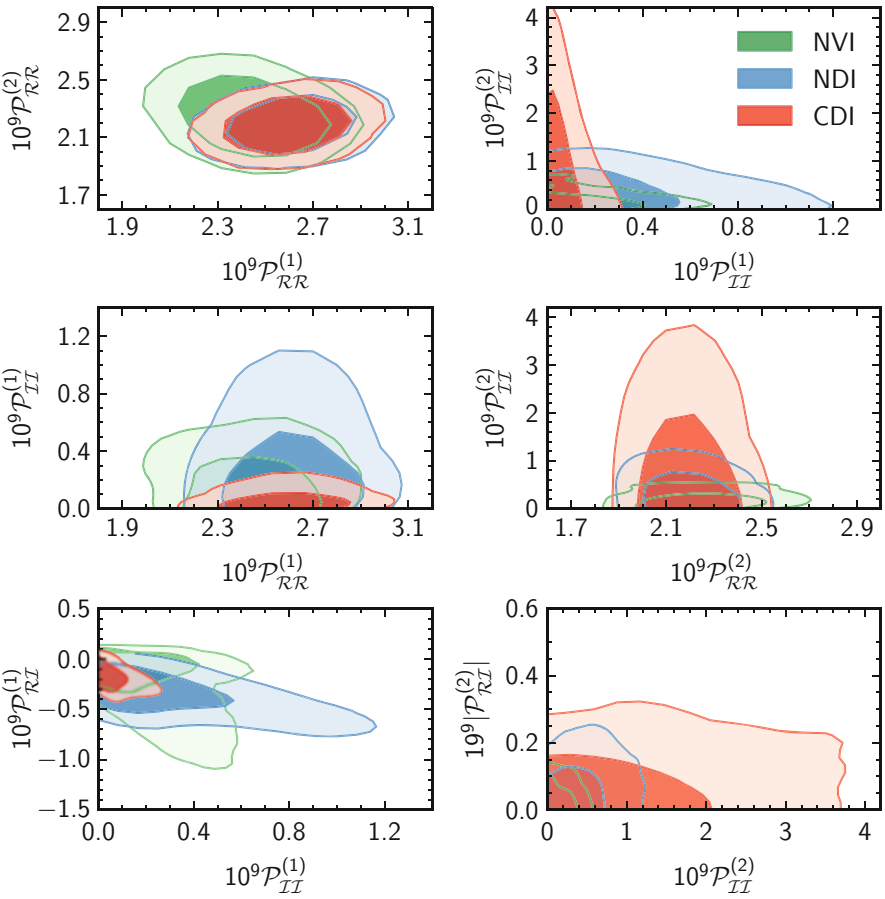

Fig. 9 Two-dimensional distributions for the quantities $\mathcal{P}_{ab}(k_{1,2}) \equiv \mathcal{P}^{(1,2)}_{ab}$ inferred from the Planck 2013 data for $a = \mathcal{R}$, CDI (*red*), NDI (*blue*) and NVI (*green*). Again, amplitude of isocurvature spectra and correlation spectra are all consistent with adiabaticity. In these plots, notice that $\mathcal{P}_{ab}$ has indices $a = \mathcal{R}, \mathcal{I}$ with $\mathcal{I} \equiv$ CDI, NDI, NVI. Figure taken from Ade et al. (2014b)

spectrum, see Eq. (217). Ade et al. (2015a) restricts itself to scale independent correlation spectrum,

$$\cos \Delta_{ab} = \frac{\mathcal{P}_{ab}}{\sqrt{\mathcal{P}_{aa}\mathcal{P}_{bb}}} \tag{239}$$

in the range $[-1, 1]$. Writing the above equation at $k = k_1$ and $k = k_2$ and requiring that the value be the same (since the correlation spectrum is scale-independent) allows us to derive the parameter $\mathcal{P}_{ab}(k_2)$.

The constraints on $\mathcal{P}_{ab}(k_1)$ and $\mathcal{P}_{ab}(k_2)$ obtained from Planck 2015 are represented in Fig. 10. The constraints on the quantities $\alpha_{ab}(\ell_{\rm min}, \ell_{\rm max})$ are displayed

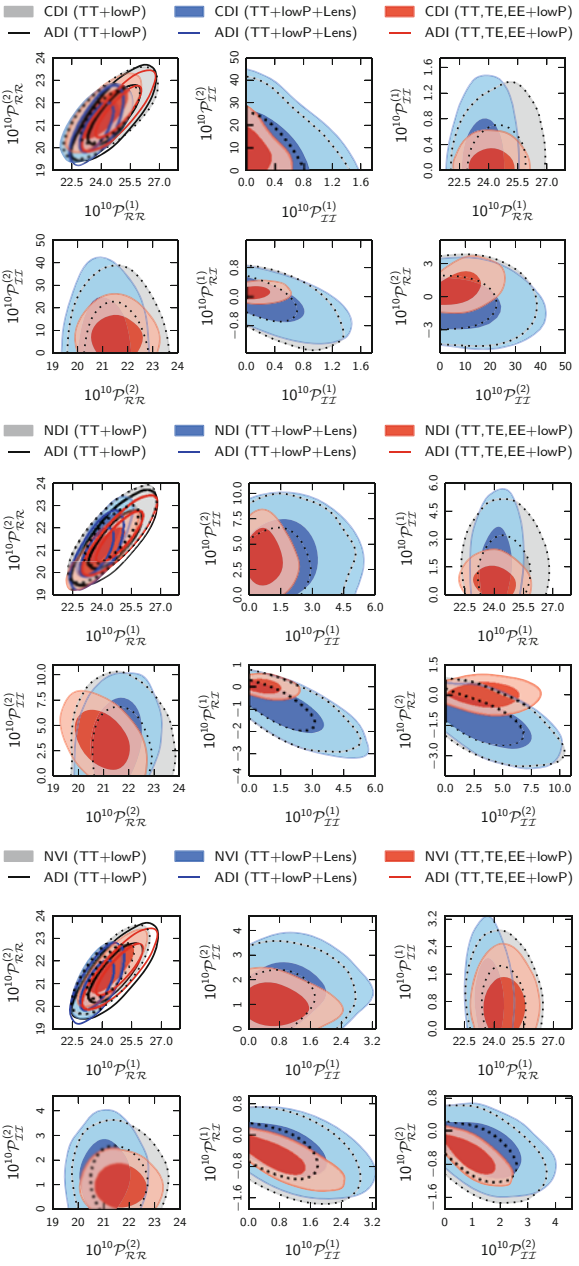

Fig. 10 Two-dimensional distributions for the quantities $\mathcal{P}_{ab}(k_{1,2}) \equiv \mathcal{P}_{ab}^{(1,2)}$ inferred from the Planck 2015 for different choices of likelihoods indicated by different colors (*gray, blue* and *red*). This plot should be compared to Fig. 9. The *six upper plots* correspond to a situation where we have a mixture of adiabatic (denoted ADI) and CDI modes, the *six middle plots* to a situation where we have ADI and NDI and the *bottom six plots* to a case where one has ADI and NVI. Figure taken from Ade et al. (2015a)

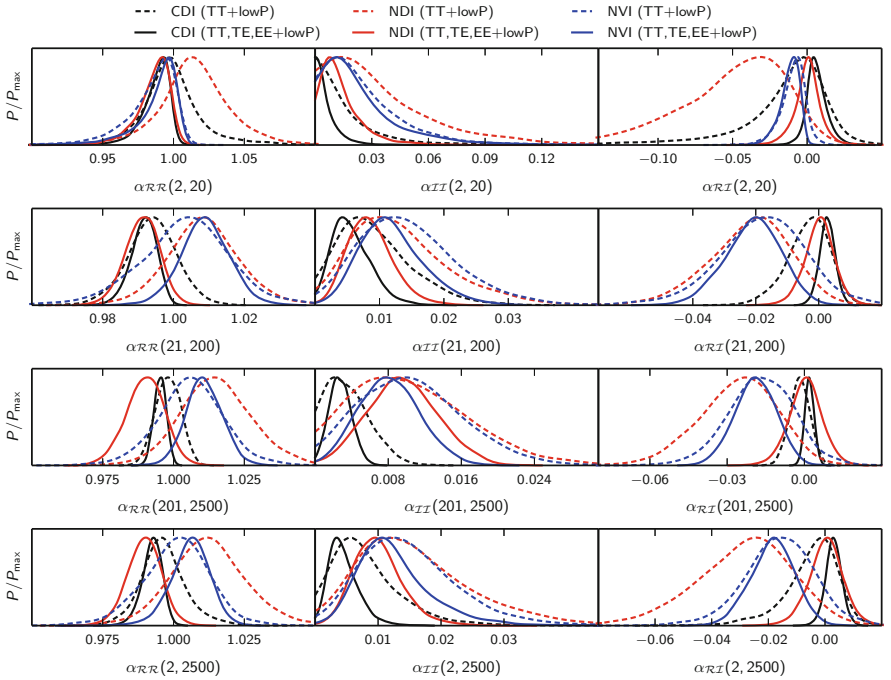

Fig. 11 Posterior distributions for the quantities $\alpha_{ab}\,(\ell_{\rm min}, \ell_{\rm max})$ introduced in Eq. (237), inferred from the Planck 2015 data. This plot should be compared to Fig. 8. Figure taken from Ade et al. (2015a)

in Fig. 11. The conclusions obtained from Planck 2013 are confirmed and even tightened. No isocurvature mode is detected and the primordial fluctuations are fully compatible with exact adiabaticity. This has of course very important implications for inflation. As explained before, this is a non trivial test for single-field slow-roll models. The Planck 2013 data were compatible with this simple class of models and did not require to introduce additional fields. The results of Planck 2015 do not modify this claim. As we are going to see in the next section, this is also the conclusion reached by the Planck measurements of Non-Gaussianities.

5.3 Non-Gaussianties

Let us now turn to the constraints on primordial Non-Gaussianity, see Ade et al. (2013). Before discussing what was measured by the Planck satellite, it is interesting to review how the results are sometimes presented in the literature (Komatsu 2010). In order to visualize the bispectrum, it is convenient to plot the quantity $\mathcal{B}_{\mathcal{R}}(k_1, k_2, k_3)(k_1k_2k_3)^2$ in terms of the ratios $x \equiv k_3/k_1$ and $y \equiv k_2/k_1$ with the

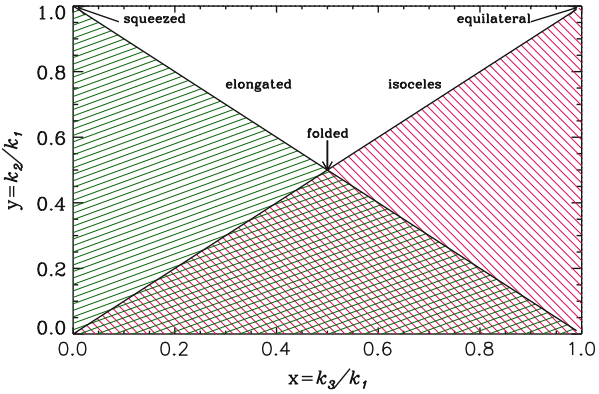

Fig. 12 Visualization of the bispectrum shape. From the fact that the three vectors $\boldsymbol{k}_1$, $\boldsymbol{k}_2$ and $\boldsymbol{k}_3$ form a triangle, it is possible to faithfully represent the bispectrum in the *white triangle*. Then, different configurations correspond to vertices or edges of that triangle

conditions that $\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3 = 0$ and $k_1 \geq k_2 \geq k_3$. This immediately implies that $0 \leq x \leq 1$ and $0 \leq y \leq 1$ and, therefore, the visualization can be restricted to this square, see Fig. 12. The fact that $k_2 \geq k_3$ means that $y > x$ and, as a consequence, the red hatched region is in fact forbidden. Then, since the three vectors $\boldsymbol{k}_1$, $\boldsymbol{k}_2$ and $\boldsymbol{k}_3$ form a triangle, every edge length is smaller than the sum of the length of the two other edges. This means that $y > 1 - x$ and the green hatched region is also forbidden. The conditions $k_2 < k_1 + k_3$ (namely $y < 1 + x$) and $k_3 < k_1 + k_2$ (namely $y > x - 1$) lead to new constraints but outside the square $[0, 1] \times [0, 1]$ and, therefore, are not interesting for us. The previous considerations show that it is sufficient to plot $\mathcal{B}_{\mathcal{R}}(k_1, k_2, k_3)(k_1 k_2 k_3)^2$ in the white, non hatched, region in Fig. 12 in order to have a complete representation of the bispectrum.

It is common practice to single out particular configurations. The squeezed triangle corresponds to $k_1 \sim k_2 \gg k_3$ which means $x \sim 0$ and $y \sim 1$. The equilateral configuration is given by $k_1 \sim k_2 \sim k_3$ or $x \sim y \sim 1$. The folded case is defined by $k_1 \simeq 2k_2 \simeq 2k_3$ or $x \sim y \sim 1/2$. These three configurations correspond to three vertices of the white triangle in Fig. 12. Another configuration is the elongated one for which $k_1 \sim k_2 + k_3$ or $x + y \sim 1$ and is therefore represented by a line in Fig. 12. The same is true for the isosceles triangle $k_1 > k_2 \sim k_3$ or $x \sim y$.

Let us now study how the local bispectrum looks like in this representation. From Eq. (135), one can write

$$\mathcal{B}_{\mathcal{R}}(k_1, k_2, k_3) = -\frac{6 f_{\mathrm{NL}}^{\mathrm{loc}}}{5}(2\pi^2)^2 \frac{A_{\mathrm{S}}^2}{k_1^6} \frac{1}{x^3 y^3}(1 + y^3 + x^3). \tag{240}$$

where $\mathcal{P}(k) = A_{\mathrm{S}}(k/k_*)^{n_{\mathrm{S}}-1}$ and, for simplicity, we have taken $n_{\mathrm{S}} = 1$ [strictly speaking, one should consider the power spectrum of Eq. (75) but, in fact, this does not change significantly the result of this calculation]. It follows, since

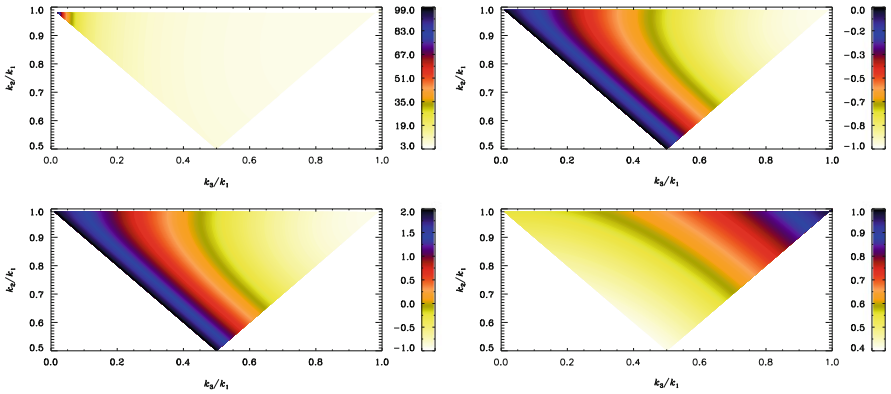

Fig. 13 Bispectrum for different shape configuration: local (*top left panel*), equilateral (*top right panel*), orthogonal (*bottom left panel*). The *bottom right panel* represents the slow-roll prediction computed for a model where $V = m^2\phi^2/2$. Notice that the absolute normalization in these figures is irrelevant

$\mathcal{B}_{\mathcal{R}}(k_1, k_2, k_3)(k_1k_2k_3)^2 = k_1^6 y^2 x^2 \mathcal{B}_{\mathcal{R}}$ that

$$\mathcal{B}_{\mathcal{R}}(k_1, k_2, k_3)(k_1k_2k_3)^2 = -\frac{6f_{\rm NL}^{\rm loc}A_{\rm S}^2}{10}(2\pi^2)^2\frac{1}{xy}(1 + y^3 + x^3). \tag{241}$$

In Fig. 13 (top left panel) we have represented this function (without the overall factor in the above expression). As it is clear from the plot (and also from the analytical expression), the local shape peaks at the squeezed triangle. The local shape has been constrained by the Planck 2013 data and one obtains (Ade et al. 2013) $f_{\rm NL}^{\rm loc} = 2.7 \pm 5.8$ at 68%CL. The Planck 2015 data (Ade et al. 2015b) with temperature only implies $f_{\rm NL}^{\rm loc} = 2.5 \pm 5.7$ and including polarization data, one arrives at

$$f_{\rm NL}^{\rm loc} = 0.8 \pm 5, \tag{242}$$

thus tightening the conclusion that the perturbations are Gaussian.

Another shape that was studied by the Planck team is the equilateral one. It is defined by

$$\mathcal{B}_{\mathcal{R}}(k_1, k_2, k_3) = \frac{18}{5}f_{\rm NL}^{\rm eq}(2\pi^2)^2A_{\rm S}^2\left[\frac{1}{k_1^3k_2^3} + \frac{1}{k_2^3k_3^3} + \frac{1}{k_1^3k_3^3} + \frac{2}{(k_1k_2k_3)^2} - \frac{1}{k_1k_2^2k_3^3}\right.$$
$$\left. - \frac{1}{k_1k_3^2k_2^3} - \frac{1}{k_2k_1^2k_3^3} - \frac{1}{k_2k_3^2k_1^3} - \frac{1}{k_3k_1^2k_2^3} - \frac{1}{k_3k_2^2k_1^3}\right], \tag{243}$$

where, again, one has taken $n_{\rm S} = 1$ for simplicity. One can re-express the bispectrum in terms of our variables x and y and then multiply by $(k_1k_2k_3)^2$. This gives

$$\mathcal{B}_{\mathcal{R}}(k_1,k_2,k_3)(k_1k_2k_3)^2 = \frac{18}{5}f_{\rm NL}^{\rm eq}(2\pi^2)^2A_{\rm S}^2\frac{1}{xy}\big(x^3+1+y^3+2xy-y-x-y^2 \\ -y^2x-x^2-x^2y\big). \tag{244}$$

The corresponding bispectrum has been represented in Fig. 13 (top right panel). The coefficient $f_{\rm NL}^{\rm eq}$ has been constrained by Planck 2013 which finds (Ade et al. 2013) $f_{\rm NL}^{\rm eq} = -42 \pm 75$, a value compatible with zero. With the Planck 2015 data (Ade et al. 2015b) (temperature only), one obtains that $f_{\rm NL}^{\rm eq} = -16 \pm 70$ and, including polarization,

$$f_{\rm NL}^{\rm eq} = -4 \pm 43. \tag{245}$$

Finally, the last shape studied by Planck is the orthogonal one for which the bispectrum can be expressed as

$$\mathcal{B}_{\mathcal{R}}(k_1,k_2,k_3) = \frac{18}{5}f_{\rm NL}^{\rm ortho}(2\pi^2)^2A_{\rm S}^2\Bigg[\frac{3}{k_1^3k_2^3}+\frac{3}{k_2^3k_3^3}+\frac{3}{k_1^3k_3^3}+\frac{8}{(k_1k_2k_3)^2}-\frac{3}{k_1k_2^2k_3^3} \\ -\frac{3}{k_1k_3^2k_2^3}-\frac{3}{k_2k_1^2k_3^3}-\frac{3}{k_2k_3^2k_1^3}-\frac{3}{k_3k_1^2k_2^3}-\frac{3}{k_3k_2^2k_1^3}\Bigg], \tag{246}$$

which leads to

$$\mathcal{B}_{\mathcal{R}}(k_1,k_2,k_3)(k_1k_2k_3)^2 = \frac{18}{5}f_{\rm NL}^{\rm ortho}(2\pi^2)^2A_{\rm S}^2\frac{1}{xy}\big(3x^3+3+3y^3+8xy-3y-3x-3y^2 \\ -3y^2x-3x^2-3x^2y\big), \tag{247}$$

and is plotted in Fig. 13 (bottom left panel). The coefficient $f_{\rm NL}^{\rm ortho}$ has been measured by Planck 2013 and the result reads (Ade et al. 2013): $f_{\rm NL}^{\rm ortho} = -25 \pm 39$. This conclusion is confirmed by the Planck 2015 measurements (Ade et al. 2015b), namely $f_{\rm NL}^{\rm ortho} = -34 \pm 33$ (temperature only). If polarization data are included, then one finds

$$f_{\rm NL}^{\rm ortho} = -26 \pm 21. \tag{248}$$

Once again, the measured value is compatible with Gaussian primordial fluctuations.

It is also interesting to represent explicitly the slow-roll result using the same visualization tools. This bispectrum was derived in Eq. (141). Expressed in terms of

x and y, each term $\mathcal{F}^{(i)}$ reads

$$\mathcal{F}^{(1)} = \frac{H^4}{16M_{\rm Pl}^4\epsilon_1}k_1^3\Bigg[\left(1+\frac{1}{1+x+y}\right)\frac{x^2y^2}{1+x+y}+\left(1+\frac{y}{1+x+y}\right)\frac{x^2}{1+x+y} + \left(1+\frac{x}{1+x+y}\right)\frac{y^2}{1+x+y}\Bigg], \tag{249}$$

$$\mathcal{F}^{(2)} = \frac{H^4}{16M_{\rm Pl}^4\epsilon_1}\times -\frac{k_1^3}{2}(1+y^2+x^2)\Bigg[-(1+x+y)+\frac{y+x+xy}{1+x+y} + \frac{xy}{(1+x+y)^2}\Bigg], \tag{250}$$

$$\mathcal{F}^{(3)} = -\frac{H^4}{16M_{\rm Pl}^4\epsilon_1}k_1^3\Bigg[\frac{1}{2}(-1+x^2-y^2)\frac{x^2}{1+x+y}\left(2+\frac{1+y}{1+x+y}\right) + \frac{1}{2}(-1-x^2+y^2)\frac{y^2}{1+x+y}\left(2+\frac{1+x}{1+x+y}\right) + \frac{1}{2}(-1-x^2-y^2)\frac{1}{1+x+y}\left(2+\frac{x+y}{1+x+y}\right)\Bigg], \tag{251}$$

$$\mathcal{F}^{(7)} = \frac{H^4}{16M_{\rm Pl}^4\epsilon_1}k_1^3\frac{\epsilon_2}{2\epsilon_1}(1+x^3+y^3). \tag{252}$$

If we write $\mathcal{F}^{(i)} \equiv H^4/(16M_{\rm Pl}^4\epsilon_1)k_1^3 f^{(i)} = (2\pi)^2\epsilon_1 A_{\rm S}^2 k_1^3 f^{(i)}$, then we see from the above Eqs. (249)–(252) that the functions $f^{(i)}$ only depend on x and y. In particular, this definition factors out the term k_1^3. As a consequence, using Eq. (141), the quantity $\mathcal{B}_{\mathcal{R}}^{\rm sr}(k_1,k_2,k_3)(k_1k_2k_3)^2$ can be written as (Hazra et al. 2013)

$$\mathcal{B}_{\mathcal{R}}^{\rm sr}(k_1,k_2,k_3)(k_1k_2k_3)^2 = \left(2\pi^2\right)^2\epsilon_1\frac{A_{\rm S}^2}{xy}\sum_{i=1,2,3,7}f^{(i)}(x,y). \tag{253}$$

This bispectrum is represented in Fig. 13 (bottom right panel). We notice that it is similar (up to a sign) to the equilateral shape (243).

Finally, Planck 2013 has also measured the four point correlation function for the local configuration. The corresponding constrain on the $\tau_{\rm NL}$ reads (Ade et al. 2013)

$$\tau_{\rm NL} < 2800, \tag{254}$$

at 95 % confidence level, that is to say, a result compatible with Gaussianity. A recent analysis (Feng et al. 2015) has confirmed this conclusion. Feng et al. (2015)

has indeed found $\tau_{\rm NL} = 0.3 \pm 0.9 \times 10^4$ and $g_{\rm NL} = -1.2 \pm 2.8 \times 10^5$. Finally, Planck 2015 (Ade et al. 2015b) obtained $g_{\rm NL} = (-9.0 \pm 7.7) \times 10^4$ at 68 % confidence level.

We conclude this section on Non-Gaussianity measurements as we concluded the section on isocurvature modes: the fact that we do not detect a signal beyond the vanilla situation is another non-trivial test for single-field slow-roll inflation with a minimal kinetic term. In the remaining part of this review, we therefore focus on this class of models and derive the corresponding implications that can be inferred from the Planck data.

5.4 *Slow-Roll Inflation*

We have seen in Sect. 4.1 that the power spectra of scalar and tensor perturbations can be expressed in terms of the slow-roll parameters. Since the CMB measurements constrain the power spectra, they also constrain the slow-roll parameters (Martin et al. 2014d). In Fig. 14, we show the two dimensional marginalized posterior distributions for the parameters ϵ_1, ϵ_2, ϵ_3 and P_*, where this last quantity represents the overall normalization of the power spectrum[4] obtained from the Planck 2013 data. We see that P_* and ϵ_2 are well constrained while there only exists an upper bound on ϵ_1 and almost no constraints on ϵ_3. Explicitly, one has $3.035 \lesssim \ln\left(10^{10} P_*\right) \lesssim 3.15$, $\log(\epsilon_1) \lesssim -2.01$ and $0.023 \lesssim \epsilon_2 \lesssim 0.063$ at the two sigma level. Planck 2015 (Ade et al. 2015a) has also analyzed this question and found $\epsilon_1 < 0.0068$ and $\epsilon_2 = 0.029^{+0.008}_{-0.007}$ using PlanckTT+lowP and restricting the hierarchy at first order in slow-roll. When high-ℓ polarization data are included in the analysis, one finds $\epsilon_1 < 0.0066$ and $\epsilon_2 = 0.030^{+0.007}_{-0.006}$.

Let us now discuss the physical information on inflation that can be inferred from the above results. Firstly, from Eqs. (75) and (76), one has at next-to-leading order on slow-roll

$$P_* = \mathcal{P}_{\zeta 0} a_0 = \frac{H_*^2}{8\pi^2 \epsilon_{1*} M_{\rm Pl}^2} \left[1 - 2(C+1)\epsilon_1 - C\epsilon_2\right], \tag{256}$$

from which we deduce that, at second order in slow-roll, the Hubble parameter during inflation can be expressed as (Martin et al. 2014d)

$$\frac{H_*^2}{M_{\rm Pl}^2} = 8\pi^2 \epsilon_1 P_* \left[1 + 2(1+C)\epsilon_1 + C\epsilon_2\right]. \tag{257}$$

[4] That is to say, we have re-written Eq. (75) as

$$\mathcal{P}_\zeta = P_* \left[1 + \frac{a_1}{a_0} \ln\left(\frac{k}{k_*}\right) + \frac{a_2}{a_0} \ln^2\left(\frac{k}{k_*}\right) + \cdots\right], \tag{255}$$

which defines the quantity P_*.

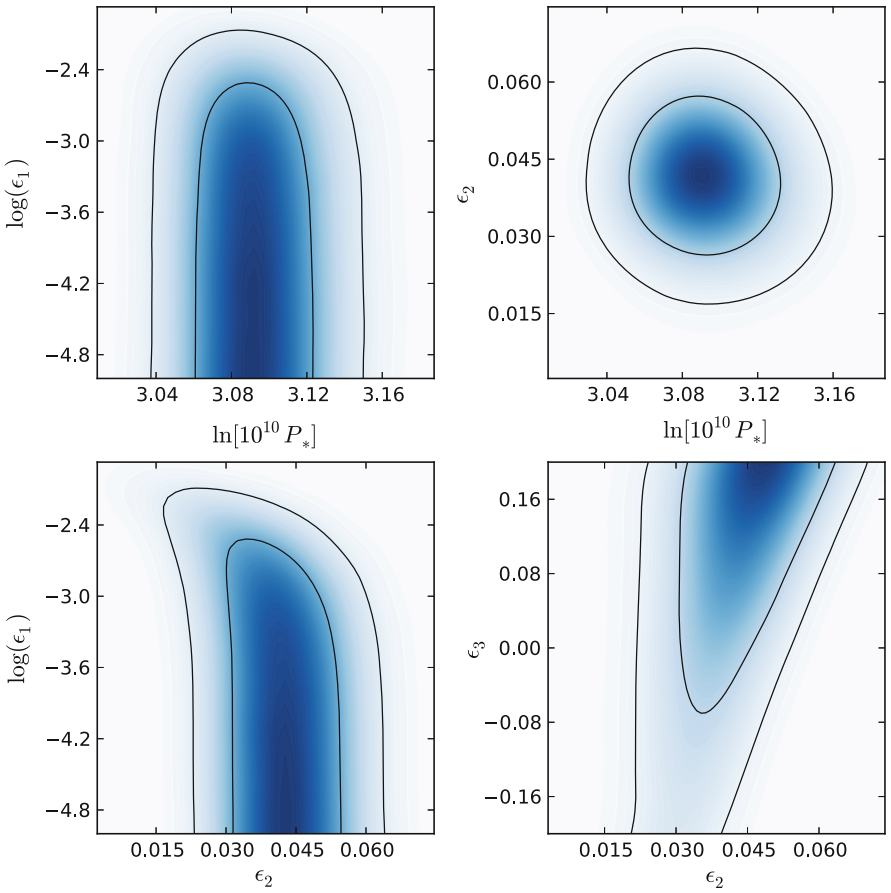

Fig. 14 Two dimensional posterior distributions of the parameters ϵ_1, ϵ_2, ϵ_3 and P_* obtained from the Planck 2013 data. The pivot scale is chosen to be $k_* = 0.05\mathrm{Mpc}^{-1}$ and the priors are taken to be as follows: a Jeffreys' prior (i.e. a flat prior on the logarithm of the corresponding quantity) for P_* such that $\ln(10^{10}P_*) \in [2.7, 4.2]$, a Jeffreys' prior for ϵ_1 such that $\log(\epsilon_1) \in [-5, -0.7]$ (the choice of the upper bound ensures that $\epsilon_1 < 0.2$ and, therefore, that the slow-roll approximation is valid) and flat priors for ϵ_2 and ϵ_3 such that $\epsilon_2 \in [-0.2, 0.2]$ and $\epsilon_3 \in [-0.2, 0.2]$. Figure taken from Martin et al. (2014b)

Since we know the posterior of P_* and ϵ_1, one can derive the corresponding one for H_*. The result is represented in Fig. 15, see the red dashed curves. Clearly, the fact that we only have an upper bound on ϵ_1 implies that we also only have an upper bound on H_*. With the Jeffreys' prior on ϵ_1 (see the left panel in Fig. 15), one obtains $\ln\left(10^5\frac{H_*}{M_{\mathrm{Pl}}}\right) \lesssim 1.6$, that is to say

$$H_* \lesssim 1.2 \times 10^{14}\mathrm{GeV}. \tag{258}$$

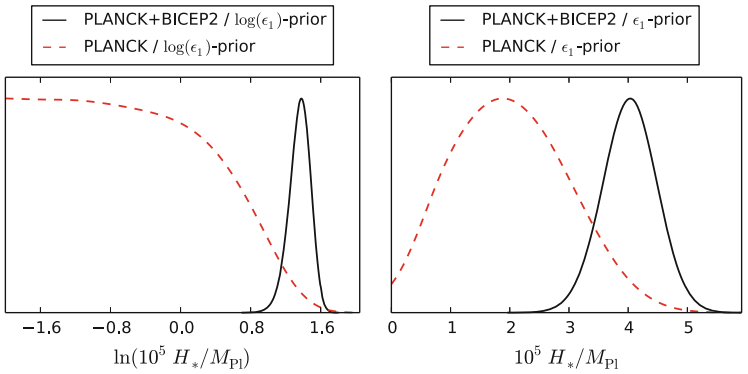

Fig. 15 Marginalized posterior distribution for the inflationary Hubble parameter at the time of pivot crossing with a Jeffreys' prior (*left panel*) and a flat panel (*right panel*) on ϵ_1 (*left panel*). The *dashed red line* represents the distribution obtained from the Planck 2013 data while the *solid black line* corresponds to the case where the Planck 2013 data are combined with the BICEP2 measurement (here, interpreted as a detection of gravity waves) and illustrates how a detection of primordial gravitational waves could allow us to determine the energy scale of inflation. Figure taken from Martin et al. (2014d)

One obtains a similar number if a flat prior on ϵ_1 is assumed (see the right panel in Fig. 15). Those values can be expressed into gravitating energy scales through

$$\rho_*^{1/4} = 3^{1/4}\sqrt{H_* M_{\rm Pl}} \lesssim 2.2 \times 10^{16}{\rm GeV}, \tag{259}$$

where this value assumes a Jeffreys' prior on ϵ_1 (again, a similar result is obtained with a flat prior). If primordial gravity waves are detected, then this would fix the value of r and, hence, the energy scale of inflation. This is illustrated in Fig. 15 where we have also plotted the posterior distribution of H_* obtained when the BICEP2 results (Ade et al. 2014c) are taken into account (assuming, for the sake of illustration, that they correspond to a detection of gravity waves).

Secondly, let us now study what the constraints on the slow-roll parameters mean for the shape of the inflation potential (Martin et al. 2014d). From Eq. (34), we see that this gives an upper bound on the first derivative of the inflation potential, namely

$$\left|V_\phi\right| \lesssim 0.14\frac{V}{M_{\rm Pl}}. \tag{260}$$

Using PlanckTT+lowP, the recent Planck 2015 data (Ade et al. 2015a) implies that $|V_\phi| \lesssim 0.116\,(V/M_{\rm Pl})$. On the other hand, the second Hubble flow parameter gives information about the second derivative of the inflaton potential. From Eq. (35), one sees that

$$M_{\rm Pl}^2\frac{V_{\phi\phi}}{V} = 2\epsilon_1 - \frac{\epsilon_2}{2}. \tag{261}$$

From this expression, we also obtain bounds on the second derivative of the potential. Indeed, one has $M_{\rm Pl}^2 V_{\phi\phi}/V > -\epsilon_{2\rm sup}$ and $M_{\rm Pl}^2 V_{\phi\phi}/V < 2\epsilon_{1\rm sup} - \epsilon_{2\rm min}/2$. Explicitly, one has

$$-0.03 \lesssim M_{\rm Pl}^2 \frac{V_{\phi\phi}}{V} \lesssim 0.008. \tag{262}$$

Planck 2015 (Ade et al. 2015a), using PlanckTT+lowP, finds the following value $M_{\rm Pl}^2 V_{\phi\phi}/V = -0.01^{+0.005}_{-0.009}$ at 95 % confidence level.

Thirdly, although the shape of the power spectrum is entirely characterized by Eq. (75), it is also interesting to derive constraints on the so-called power-law parameters (Martin et al. 2014d). These parameters are in fact simple combinations of the Hubble flow parameters, as exemplified by Eqs. (86). We now investigate this question in more detail. In Eqs. (86), we gave the spectral indices at first order in slow-roll. At second order, they read

$$\begin{aligned} n_{\rm S} &= 1 - (2\epsilon_1 + \epsilon_2) - 2\epsilon_1^2 - (3 + 2C)\epsilon_1\epsilon_2 - C\epsilon_2\epsilon_3, \\ n_{\rm T} &= -2\epsilon_1 - 2\epsilon_1^2 - 2(1 + C)\epsilon_1\epsilon_2, \end{aligned} \tag{263}$$

while the tensor-to-scalar ratio can be expressed as [see also Eq. (87)]

$$r = 16\epsilon_1(1 + C\epsilon_2). \tag{264}$$

One can also define the runnings for scalar and tensor and, in the slow-roll approximation, they are second-order quantities and their expressions read

$$\alpha_{\rm S} = -2\epsilon_1\epsilon_2 - \epsilon_2\epsilon_3, \qquad \alpha_{\rm T} = -2\epsilon_1\epsilon_2, \tag{265}$$

Finally, let us mention that the running of the running for the tensor mode is also completely specified by the first three Hubble flow functions and is given by

$$\beta_{\rm T} = -2\epsilon_1\epsilon_2\left(\epsilon_2 + \epsilon_3\right). \tag{266}$$

One sees that, in general, one has six independent quantities, namely r, $n_{\rm S}$, $n_{\rm T}$, $\alpha_{\rm S}$, $\alpha_{\rm T}$ and $\beta_{\rm T}$. However, the predictions of slow-roll inflation can be expressed in terms of three Hubble flow parameters (at least at this order), ϵ_1, ϵ_2 and ϵ_3. This implies that all the parameters describing the tensor sector can, in fact, be expressed in terms of those characterizing the scalar sector. Explicitly, these so-called consistency relations can be expressed as

$$\begin{aligned} n_{\rm T} &\simeq -\frac{r}{8}, \\ \alpha_{\rm T} &\simeq \frac{r}{8}\left[\frac{r}{8} + (n_{\rm S} - 1)\right], \\ \beta_{\rm T} &\simeq \frac{r}{8}\left[\frac{r}{8} + (n_{\rm S} - 1)\right]\left(1 - n_{\rm S} - \frac{r}{4}\right) + \frac{r}{8}\alpha_{\rm S}. \end{aligned} \tag{267}$$

As before, since we know the posterior distributions of the slow-roll parameters for the Planck 2013 data, we can infer those of the power-law parameters. They are represented in Fig. 16, see the red dashed and dotted green lines. We see that the scalar spectral index $n_{\rm S}$ is very well constrained and is around $n_{\rm S} \simeq 0.96$ (the Planck 2013 value, with WMAP large-angle polarization, reads $n_{\rm S} = 0.9603 \pm 0.0073$). On the other hand, we only have an upper bound on the tensor-to-scalar ratio which is of course expected since $r \propto \epsilon_1$. At two sigmas, one obtains $\log(r) \lesssim -0.88$ which gives

$$r \lesssim 0.13. \tag{268}$$

Notice that this result is obtained assuming a Jeffeys' prior on ϵ_1. If, instead, a flat prior is chosen, one has $\log(r) \lesssim -0.64$, leading to $r \lesssim 0.23$. This is because a flat prior has the tendency to favor large values of r compared to what is obtained with a Jeffrey's prior. In Fig. 16, we have also represented the results obtained by combining Planck 2013 and BICEP2 assuming that this last signal is due to primordial gravity waves. Of course, in that case r is determined and, as a consequence, the tensor spectral index is also fixed. It is now known that the BICEP2 signal can be entirely explained by dust contamination (Ade et al. 2016) but, nevertheless, it is interesting to see what would be the implications for inflation of a detection of primordial gravity waves.

Recently, Planck 2015 (Ade et al. 2015a) has also put constraints on r. As usual, these constraints depend on the data sets used and on the assumptions made about the theoretical frameworks. Here we just quote two numbers. Using PlanckTT,TE,EE+lowP and considering that r is the only extra parameter beyond the base ΛCDM model, one obtains

$$r_{0.002} < 0.1, \tag{269}$$

at 95 % confidence limit. If instead PlanckTT+lowP+WP (we remind that WP means the polarization data on large scales measured by WMAP), this number becomes $r_{0.002} < 0.09$. Here, the subscript "0.002" indicates that the pivot scale is taken to be $0.002\,{\rm Mpc}^{-1}$.

Very recently, a joint analysis by the BICEP2/Keck Array team and the Planck collaboration was released (BICEP2/Keck et al. 2015). The results are presented in Fig. 17. In Ade et al. (2014c), BICEP2 announced the detection of primordial gravity waves at a level corresponding to a tensor-to-scalar ratio of $r \sim 0.16$. The reason for this claim can be seen in Fig. 17. In this plot, the red solid curve is the signal due to the weak lensing of E-mode that produces B-modes on small angular scales. This contribution is necessarily present in the standard model of Cosmology and its amplitude can be inferred unambiguously once we know the value of the cosmological parameters. The black dots represent the signal measured by BICEP2 and Keck Array. As is well visible, in the range $\ell \sim [50, 120]$, there is an excess of power with respect to the red solid line and, hence, there must be another source of B-modes. BICEP2 interpreted this excess as a contribution coming from primordial

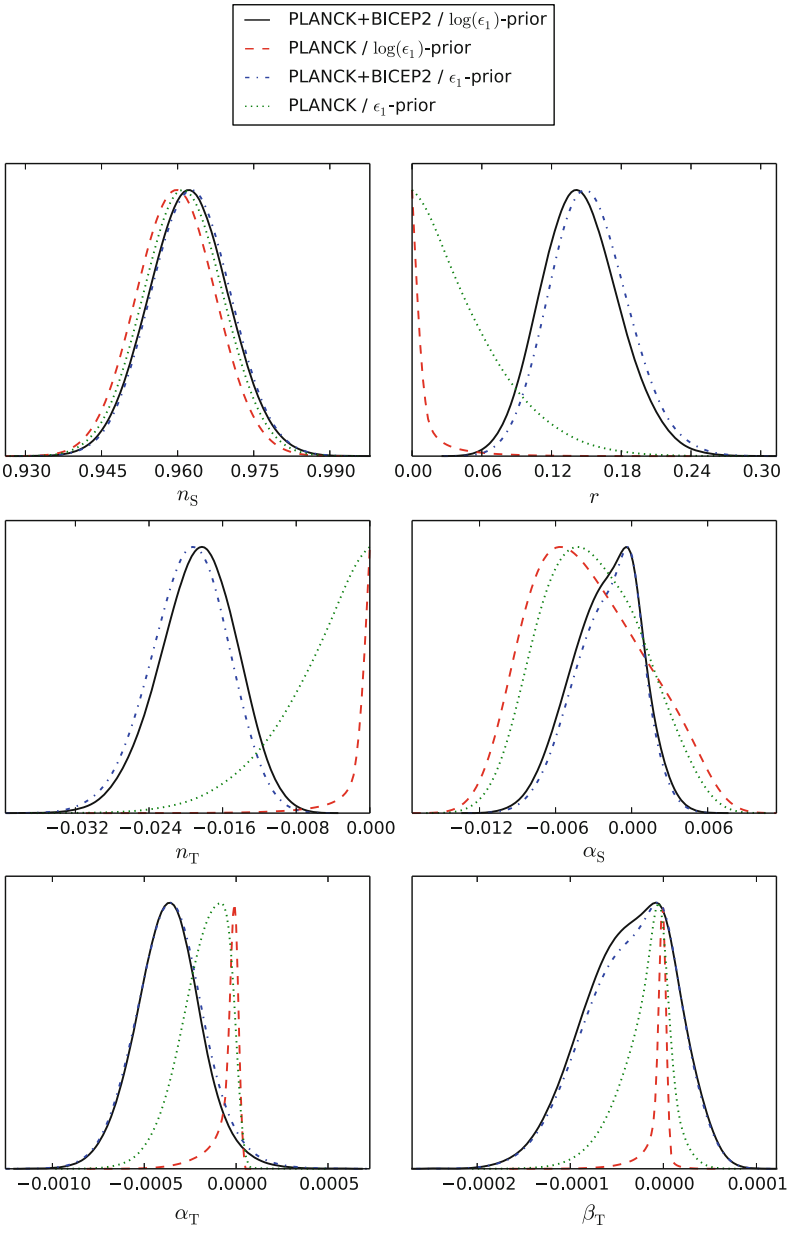

Fig. 16 Marginalized posterior distributions for the derived power law parameters n_S, r, n_T, α_S, α_T and β_T obtained by importance sampling from the distributions of the second order slow-roll parameters. The *dashed red line* and the *dotted green lines* are the distributions obtained from Planck 2013. We have also represented the results obtained by combining Planck 2013 with BICEP2 (see the *solid black line* and the *dotted dashed blue line*). The most striking feature which would follow from a detection of gravity waves (here illustrated by including the BICEP2 results taken at face value) is of course that r, and therefore n_T using the slow-roll consistency relations (267), would now be measured. Figure taken from Martin et al. (2014d)

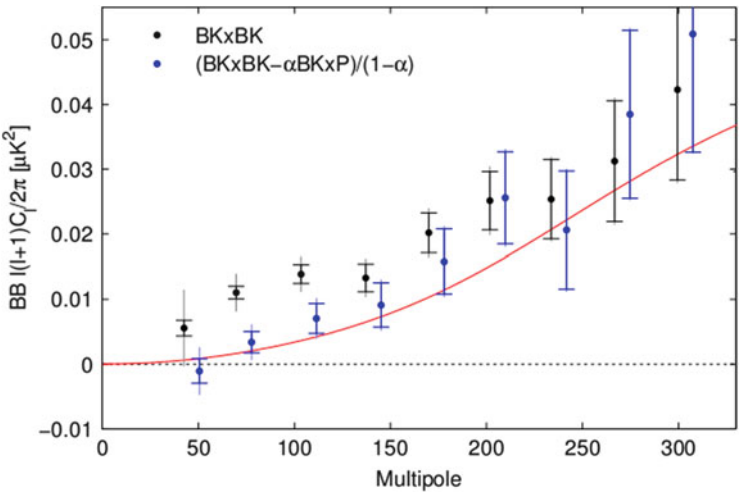

Fig. 17 *B*-mode CMB polarization multipole moments compared to the lensing signal (*solid red curve*). The *black dots* represent the values of C_ℓ^{BB} obtained from the BICEP2/Keck array map. On recognizes the bump in the range $\ell \sim [50, 120]$ that deviates from the lensing curve and that was interpreted as a detection of primordial gravity waves with $r \sim 0.16$ in Ade et al. (2014c). The *blue dots* correspond to the same multipole moments but after subtraction of the dust contribution, estimated from the cross-spectrum with the Planck 353 GHz channel. Clearly, the new data are in good agreement with what is expected from lensing. Figure taken from BICEP2/Keck et al. (2015)

gravity waves. However, there is another known source of contamination: dust. BICEP2 could not measure accurately the dust contribution because it operates at a single frequency only. The BICEP2 team therefore used theoretical models available at that time to remove the dust signal. On the other hand, Planck measures the CMB at different frequencies and, as a consequence, can estimate with good precision the dust contribution. It is therefore clear that a joint analysis between the two teams is the best way to use at the same time the good sensitivity of BICEP2/Keck Array and the good control of the dust signal of the Planck team. The result of this analysis are the blue dots in Fig. 17. We see that the bump has disappeared which means that the excess of power observed was probably entirely due to dust contamination and not to primordial gravity waves. The signal is now compatible with lensing. The new analysis suggests that the best value of r is now $r \sim 0.05$ but with very low significance and $r \sim 0$ cannot be excluded. In other words, there is no longer a detection of primordial gravity waves. In addition, one obtains a new upper limit which is now $r < 0.12$ (at 95 % confidence limit) instead of $r < 0.11$ from the Planck 2013 data. Notice that we obtained before $r < 0.13$ from the Planck 2013 data, see Eq. (268), and not $r < 0.11$, but this is just due to some differences between our analysis and the Planck one (essentially, different priors).

In Martin et al. (2014d), it was demonstrated that the sets of inflationary models preferred by Planck alone and BICEP2 alone are almost disjoint, indicating a clear tension between the two data sets. Using a Bayesian measure of compatibility

between BICEP2 and Planck, it was indeed shown that, for models favored by Planck 2013 the two data sets tended to be incompatible, whereas there was a moderate evidence of compatibility for the BICEP2 preferred models. This means that the three assumptions (1) slow-roll inflation is the correct description of the early Universe (2) Planck 2013 data accurately measure CMB temperature anisotropies and (3) BICEP2 measurement is due to primordial gravity waves are mutually exclusive. In other words, if one has the theoretical prejudice that slow-roll inflation did occur in the early Universe, then Martin et al. (2014d) already proved that the value $r \sim 0.16$ was likely to be overestimated. In some sense, the fact that dust contamination can explain the BICEP2 signal reinforces our trust in inflation!

Let us now turn to the scalar running $\alpha_{\rm S}$. At 95 % confidence level, one finds

$$-0.012 \lesssim \alpha_{\rm S} \lesssim 0.006, \tag{270}$$

that is to say a value consistent with no running. Finally, one notices that the quantities $\alpha_{\rm T}$ and $\beta_{\rm T}$ are well-constrained. It is easy to understand why on the example of $\alpha_{\rm T}$. One has $\alpha_{\rm T} = (r/8)^2 + (n_{\rm S} - 1)(r/8)$, see Eq. (267). In this equation $n_{\rm S} - 1$ is known, one can take $n_{\rm S} - 1 \simeq -0.04$ which means that $\alpha_{\rm T} \simeq (r/8)^2 - 0.04(r/8)$. This parabola has a minimum at $r/8 \simeq 0.02$ which corresponds to $\alpha_{\rm T} \simeq -9 \times 10^{-5}$. The maximum is for $r \simeq 0.13$ and gives $\alpha_{\rm T} \simeq -0.0004$. We therefore expect $-0.0004 \lesssim \alpha_{\rm T} < -9 \times 10^{-5}$ and which (roughly speaking) explains why the distribution of $\alpha_{\rm T}$ in Fig. 16 is peaked (see the red dashed line).

A last remark is in order at this point. Very often, as already pointed out, the power spectrum is parametrized as

$$\mathcal{P}_\zeta(k) = A_{\rm S} \left(\frac{k}{k_*}\right)^{n_{\rm S}-1+\alpha_{\rm S}/2\ln(k/k_*)+\cdots}, \tag{271}$$

and a similar expression for the tensors. Clearly, this is not exactly what inflation predicts since not expanding $n_{\rm S} - 1 = -2\epsilon_1 - \epsilon_2$ (if one works at first order in slow-roll) in the above formula means in fact keeping an infinite number of higher order corrections which is clearly inconsistent since $n_{\rm S}$ is determined at a fixed order. Of course, since $n_{\rm S} - 1$ is small, for all practical purposes, this does not impact a lot the final results.

5.5 *Model Comparison*

Let us now turn to model comparison (Martin et al. 2011, 2014a,b,d). We would like to determine the models of inflation that perform the best being given the current CMB data. From a statistical point of view, this question is subtle. Indeed, suppose we have two models: $\mathcal{M}_1$ characterized by one parameter θ_{11} and $\mathcal{M}_2$ characterized by two parameters θ_{21}, θ_{22}. What does it mean to claim that model $\mathcal{M}_1$ is better

than model $\mathcal{M}_2$ (or the opposite)? Naively, one could compare the likelihoods of the two models for the values of the parameters leading to the best fits. But model $\mathcal{M}_2$ has one extra parameter and, therefore, one expects this model to automatically improve the fit. Therefore, in some sense, it would be "unfair" to claim that $\mathcal{M}_2$ is better than $\mathcal{M}_1$ since it is "more complicated". Moreover, suppose that only for, say, $\theta_{21} \in \left[10^{-20}, 10^{-19}\right]$ does $\mathcal{M}_2$ lead to a good χ^2 while, a priori, θ_{12} could vary, in say $[-1, 1]$. Suppose, in addition, that this does not happen for $\mathcal{M}_1$, namely that for θ_{11} in its natural range of variation, the fit is always "reasonable". How do we take into account this wasted parameter space for model $\mathcal{M}_2$ in our assessment of the respective performance of the two models?

In order to answer these questions, one recalls that if $\mathcal{L}_2(\theta_{21}, \theta_{22}) \equiv p(D|\theta_{21}, \theta_{22}, \mathcal{M}_2)$ is the likelihood of model $\mathcal{M}_2$ (D represents the data, here we have of course CMB data in mind), then the probability of the parameters θ_{21}, θ_{22}, can be expressed as (the Bayes' theorem) (Trotta 2008)

$$p(\theta_{21}, \theta_{22}|D, \mathcal{M}_2) = \frac{1}{\mathcal{E}(D|\mathcal{M}_2)}\mathcal{L}_2(\theta_{21}, \theta_{22})\pi(\theta_{21}|\mathcal{M}_2)\pi(\theta_{22}|\mathcal{M}_2), \tag{272}$$

where π represents the prior distributions and $\mathcal{E}$ is a normalization factor which depends on the data and the model. We would like to calculate the probability $p(\mathcal{M}_2|D)$ of model $\mathcal{M}_2$ and, therefore, we expect a similar equation to hold, namely

$$p(\mathcal{M}_2|D) = \frac{1}{p(D)}p(D|\mathcal{M}_2)\pi(\mathcal{M}_2), \tag{273}$$

where $p(D)$ is a normalization factor depending on the data only and π encodes our a priori information about model $\mathcal{M}_2$. Clearly, Eqs. (272) and (273) have the same structure since they represent two applications of the Bayes's theorem. To make progress we need to know $p(D|\mathcal{M}_2)$. But this quantity is in fact easy to calculate since $\int p(\theta_{21}, \theta_{22}|D, \mathcal{M}_2)\mathrm{d}\theta_{21}\mathrm{d}\theta_{22} = 1$, Eq. (272) leads to

$$\begin{aligned}\mathcal{E}(D|\mathcal{M}_2) &= \int \mathcal{L}_2(\theta_{21}, \theta_{22})\pi(\theta_{21}|\mathcal{M}_2)\pi(\theta_{22}|\mathcal{M}_2)\mathrm{d}\theta_{21}\mathrm{d}\theta_{22} \\ &= \int p(D|\theta_{21}, \theta_{22}, \mathcal{M}_2)\pi(\theta_{21}|\mathcal{M}_2)\pi(\theta_{22}|\mathcal{M}_2)\mathrm{d}\theta_{21}\mathrm{d}\theta_{22} \\ &= p(D|\mathcal{M}_2).\end{aligned} \tag{274}$$

Of course the previous considerations apply in general and the quantity $\mathcal{E}(D|\mathcal{M}_i)$ is called the Bayesian evidence of the model $\mathcal{M}_i$ and its definition reads (Trotta 2008)

$$p(D|\mathcal{M}_i) \equiv \mathcal{E}\left(D|\mathcal{M}_i\right) = \int \mathrm{d}\theta_{ij}\mathcal{L}\left(\theta_{ij}\right)\pi\left(\theta_{ij}|\mathcal{M}_i\right). \tag{275}$$

The Bayesian evidence is often normalized to a reference model $\mathcal{M}_{\rm REF}$ and one defines $B^i_{\rm REF} \equiv \mathcal{E}(D|\mathcal{M}_i)/\mathcal{E}(D|\mathcal{M}_{\rm REF})$. In that case, the posterior probability of the model $\mathcal{M}_i$ (for non-committal model priors) can be re-expressed as

$$p\left(\mathcal{M}_i|D\right) = \frac{B^i_{\rm REF}}{\sum_j B^j_{\rm REF}}\,. \tag{276}$$

In the following, we will give $B^i_{\rm REF}$ since this quantity is in one-to-one correspondence with the probability of the model $\mathcal{M}_i$. In particular, one sees that $p(\mathcal{M}_i) > p(\mathcal{M}_j)$, namely model $\mathcal{M}_i$ is better than $\mathcal{M}_j$ (or more probable), if $\mathcal{E}(\mathcal{M}_i) > \mathcal{E}(\mathcal{M}_j)$ or, equivalently, $B^i_{\rm REF} > B^j_{\rm REF}$.

In order to see why computing the Bayesian evidence answers the questions asked before and can give a fair estimate of the performances of a model, let us consider the idealized following situation. Let us assume that the likelihood function of model $\mathcal{M}_1$ has width $\delta\theta_{11}$ and that the prior is flat and has width $\Delta\theta_{11}$. Since π is normalized, we have $\pi(\theta_{11}) = 1/\Delta\theta_{11}$. We also assume that the likelihood is more informative than the prior, namely $\delta\theta_{11} < \Delta\theta_{11}$. Then, Eq. (275) is approximately given by

$$\mathcal{E}(D|\mathcal{M}_1) \simeq \mathcal{L}_{1,\rm max}\frac{\delta\theta_{11}}{\Delta\theta_{11}}. \tag{277}$$

A similar calculation for $\mathcal{M}_2$ leads to

$$\mathcal{E}(D|\mathcal{M}_2) \simeq \mathcal{L}_{2,\rm max}\frac{\delta\theta_{21}}{\Delta\theta_{21}}\frac{\delta\theta_{22}}{\Delta\theta_{22}}. \tag{278}$$

For simplicity, one can take the reference model to be model $\mathcal{M}_1$ and, of course, one has $B^1_{\rm REF} = 1$. For $B^2_{\rm REF}$, one finds

$$B^2_{\rm REF} = \frac{\mathcal{L}_{2,\rm max}}{\mathcal{L}_{1,\rm max}}\frac{\Delta\theta_{11}}{\delta\theta_{11}}\frac{\delta\theta_{21}}{\Delta\theta_{21}}\frac{\delta\theta_{22}}{\Delta\theta_{22}}. \tag{279}$$

On this last equation, we see that deciding whether model $\mathcal{M}_1$ is better or worse than $\mathcal{M}_2$ does not reduce to the comparison of the likelihood function at the best fit, $\mathcal{L}_{2,\rm max}/\mathcal{L}_{1,\rm max}$ but that this ratio is corrected by a factor which describes how much parameter space has been wasted. So the best model is not the one which has the largest χ^2 but the one which achieves the best compromise between quality of the fit and simplicity of the theoretical description.

As explained before, here, we focus on single-field slow-roll inflationary models (with minimal kinetic term) only. At this stage, the strategy is clear: one must evaluate the Bayesian evidence of each of these models in order to rank them according to their ability to fit the data. This first requires to identify all models of this type and this was recently done in *Encyclopaedia Inflationaris*, see Martin et al. (2014a). In this work, about 200 models have been identified. A model corresponds

to a specific choice of potential and of priors for its parameters. Two different models can have the same potential but different priors. Each model is denoted by an acronym according to the terminology introduced in Martin et al. (2014b) and, in the present lecture, we just make use of this convention. A detailed justification of the priors chosen can also be found in that reference. In Fig. 18, we have represented the Bayesian evidence of the different models (being given the Planck 2013 data) by an horizontal bar the length of which is proportional to $\ln B^i_{\rm REF}$, see also the caption of Fig. 18 and Martin et al. (2014a,b,d), and Ringeval (2014). In order to translate the numerical value of the evidence into strength of belief, we introduce the Jeffrey's scale (Trotta 2008). If $|\ln B^i_{\rm REF}| < 1$, then the model is in the "inconclusive zone", if $1 < |\ln B^i_{\rm REF}| < 2.5$, then it is in the "weak evidence zone", if $2.5 < |\ln B^i_{\rm REF}| < 5$, then it is in the "moderate evidence zone" and, finally, if $|\ln B^i_{\rm REF}| < 5$, it is in the "strong evidence zone". If the reference model is taken to be the best model, then, by definition all $\ln B^i_{\rm REF}$ are negative. In that case, the best models are those in the inconclusive zone and those in the "strong evidence zone" can be considered as ruled out.

In Fig. 18, we see that the best Planck 2013 model is KMIII and that 52 models end up being in the inconclusive zone, namely: KMIII, $\mathrm{ESI}_{\sqrt{2}}$, $\mathrm{BI}_{6\mathrm{s}}$, $\mathrm{MHI}_{\mathrm{s}}$, BI_{s}, ESI, $\mathrm{BI}_{5\mathrm{s}}$, $\mathrm{KKLTI}_{\mathrm{s}}$, $\mathrm{KMII}_{V>0}$, $\mathrm{BI}_{4\mathrm{s}}$, $\mathrm{ESI}_{\mathrm{o}}$, $\mathrm{ESI}_{\sqrt{2/3}}$, KMII, HI, $\mathrm{BI}_{3\mathrm{s}}$, $\mathrm{BI}_{2\mathrm{s}}$, $\mathrm{RGI}_{\mathrm{s}}$, $\mathrm{RGI}_{1/16}$, $\mathrm{BI}_{\mathrm{ph}}$, AI, $\mathrm{BI}_{1\mathrm{s}}$, MHI, $\mathrm{SFI}_{\mathrm{l}}$, SFI, $\mathrm{KKLTI}_{\mathrm{stg}}$, $\mathrm{BI}_{\mathrm{stg}}$, KKLTI, SBI, RGI, $\mathrm{SFI}_{\mathrm{s}}$, $\mathrm{PSNI}_{\mathrm{oA}}$, $\mathrm{SFI}_{4\mathrm{l}}$, $\mathrm{PSNI}_{\mathrm{ft2}}$, $\mathrm{PSNI}_{\mathrm{oB}}$, $\mathrm{PSNI}_{\mathrm{ft1}}$, $\mathrm{PSNI}_{\mathrm{oC}}$, $\mathrm{LI}_{\alpha>0}$, SFI_4, $\mathrm{ESI}_{\mathrm{l}}$, SSBI2, $\mathrm{PSNI}_{\mathrm{ft3}}$, $\mathrm{PSNI}_{\mathrm{epA}}$, SSBI4, TWI_{ϕ_0}, $\mathrm{RGI}_{\mathrm{l}}$, $\mathrm{SFI}_{4\mathrm{s}}$, $\mathrm{MHI}_{\mathrm{l}}$, $\mathrm{PSNI}_{\mathrm{epB}}$, $\mathrm{TWI}^{\mathrm{r}}_{\phi_0}$, $\mathrm{SBI}_{\alpha_{\min}}$, LI, $\mathrm{SFI}_{3\mathrm{l}}$. They represent $\sim$ 26 % of the models analyzed. We also find that 21 % of the models are in the "weak evidence zone", 17 % in the "moderate evidence zone" and 34 % in the "strong evidence zone". Planck 2013 is therefore able to rule out about one third of the inflationary models. Model comparison with Planck 2015 data cannot yet be done since the scientific products are not delivered. However, given the consistency of these two data sets, we do not expect very different results.

We have seen before that the "winner" is KMIII which is a string inspired model with the following potential (Conlon and Quevedo 2006)

$$V(\phi) = M^4\left[1 - \alpha\left(\frac{\phi}{M_{\rm Pl}}\right)^{4/3} \mathrm{e}^{\beta(\phi/M_{\rm Pl})^{4/3}}\right], \tag{280}$$

where α and β are two free parameters. Its heir apparent is $\mathrm{ESI}_{\sqrt{2}}$ the potential of which is given by Stewart (1995), Dvali and Tye (1999) and Cicoli et al. (2009)

$$V(\phi) = M^4\left(1 - \mathrm{e}^{-\sqrt{2}\phi/M_{\rm Pl}}\right). \tag{281}$$

The fourteenth on the list is the Starobinsky model (Starobinsky 1980; Bezrukov and Shaposhnikov 2008), namely

$$V(\phi) = M^4\left(1 - \mathrm{e}^{-\sqrt{2/3}\phi/M_{\rm Pl}}\right)^2. \tag{282}$$

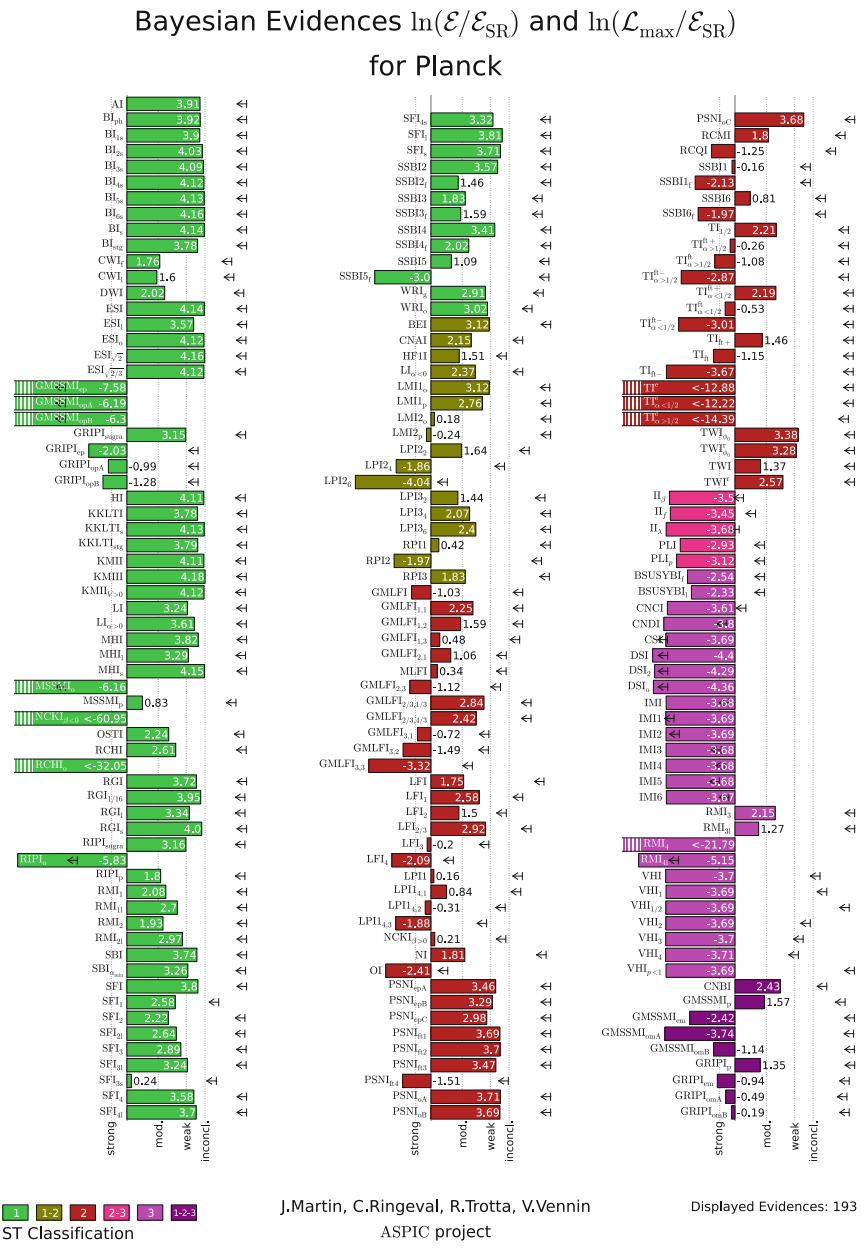

Fig. 18 $\ln B^i_{\rm REF}$ for the all single field slow-roll models with minimal kinetic terms. The reference model is taken to be the one where the priors are directly chosen on the Hubble flow parameters. Each model is represented by a bar, the length of which is directly proportional to $\ln B^i_{\rm REF}$ (the numerical value of $\ln B^i_{\rm REF}$ being indicated on the same line). A bar on the left means that $\ln B^i_{\rm REF} < 0$ and a bar on the right that $\ln B^i_{\rm REF} > 0$. The color code refers to the Schwarz-Cesaro Escalante classification (Schwarz et al. 2001). The *vertical dotted black line* indicates the Jeffreys' categories, see the text for more explanations. Figure taken from Martin et al. (2014b)

Actually, all these models being in the "inconclusive zone", the difference between their Bayesian evidence is not significant. This means that, for instance, one should view the Starobinsky model as good as KMIII. In fact, the main common point between all these scenarios is that they all possess a "plateau-like" potential, meaning that $V_\phi(\phi) \to 0$ as $\phi \to \infty$. We conclude that Planck 2013 has been able to constrain the shape of the inflationary potential, a truly remarkable achievement when one remembers that inflation can take place at 10^{15}GeV, and certainly something impossible to do in an accelerator!

We have seen before that, in order to explain the data, we do not need to consider models more complicated than single field slow-roll inflation with a minimal kinetic term. This does not mean that more complicated models are ruled out (in the frequentist point of view) in the sense that, with a carefully chosen set of parameters, they can lead to good fits. However, from the previous considerations, we see that those models must have a very "bad" Bayesian evidence. Computing the evidence of those more complicated models is certainly a difficult task (for instance, for models predicting a non negligible level of Non-Gaussianities, one would need to take into account the higher order correlation functions). But, in fact, we do not need to carry out such a calculation which, at this stage appears to be useless. Indeed we know in advance that they are much "worse" than single-field models because of their huge wasted parameter space (Giannantonio and Komatsu 2015). It is sufficient to know that they all are in the "strong evidence zone" and, clearly, we are not much interested in knowing the ranking in this Jeffreys category since the models are ruled out (in the Bayesian sense) anyway.

We have seen how the Bayesian evidence allows us to rank the various inflationary models. However, two models with a different number of parameters can have the same evidence if the extra parameters are not constrained by the data. This is certainly not a desirable property as the model with less parameters is clearly simpler and, therefore, should be favored. In order to break this degeneracy, we now introduce the Bayesian complexity (Kunz et al. 2006). For a model $\mathcal{M}_i$, it is defined by (Kunz et al. 2006)

$$C_{\rm b}^i = \left\langle -2\log\mathcal{L}\left(\theta_{ij}\right)\right\rangle + 2\log\mathcal{L}\left(\theta_{ij}^{\rm ML}\right), \tag{283}$$

where $\langle\cdot\rangle$ means averaging over the posteriors and $\theta_{ij}^{\rm ML}$ represents the maximum likelihood estimate of the model's parameters. One can easily show, see for instance (Martin 2013), that the number of unconstrained parameters, given a data set, can be expressed as

$$N_{\rm uc}^i = N_{\rm param}^i - C_{\rm b}^i, \tag{284}$$

where $N_{\rm param}^i$ represents the number of free parameters of model $\mathcal{M}_i$. We see that this gives us a new criterion to discriminate the various models since a model such that $0 < N_{\rm uc}^i < 1$ ought to be preferred. The Bayesian complexities (given the Planck 2013 data) of all the *Encyclopaedia Inflationaris* models have been

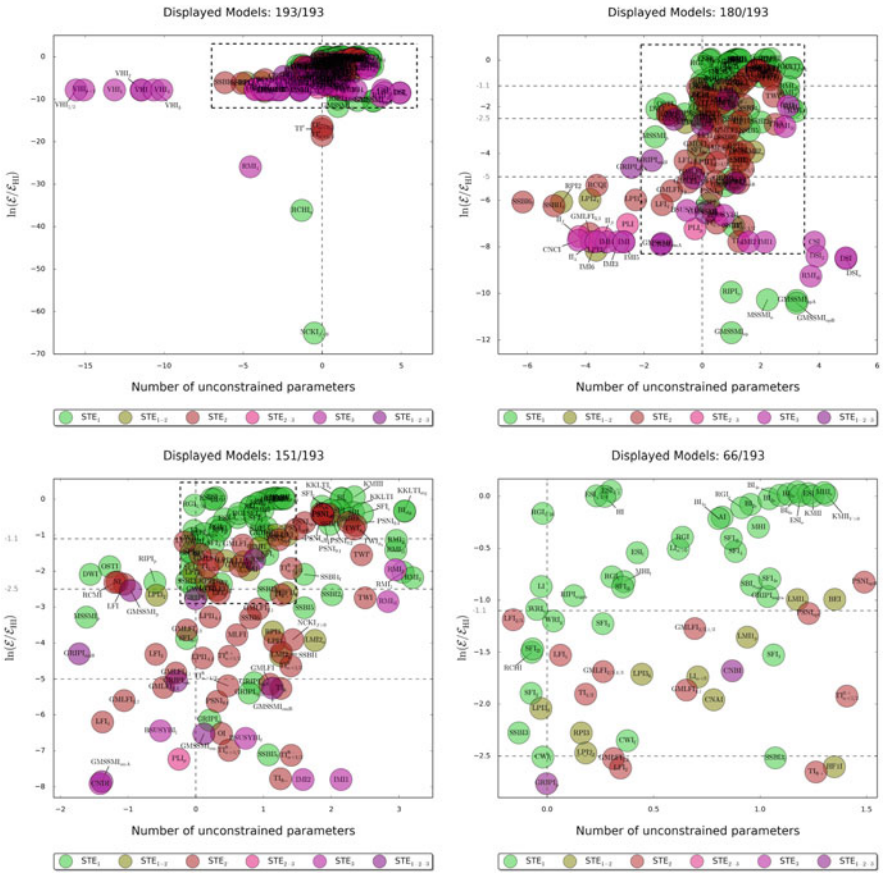

Fig. 19 Inflationary models in the space $\left(N_{\rm uc}, \ln B^i_{\rm REF}\right)$. Each model is represented by a circle (the radius of which has no meaning) with its acronym, taken from Martin et al. (2014b), written inside. The four panels corresponds to successive zooms towards the best region (*indicated by the dashed rectangles*), the one with $0 < N_{\rm uc} < 1$, namely where all the parameters are constrained by the data and a large value of the evidence, namely the model achieves a good fit without wasting parameter space. Figures taken from Martin et al. (2014b)

computed in Martin et al. (2014b). In Fig. 19, we have represented these scenarios in the space $\left(N_{\rm uc}, \ln B^i_{\rm REF}\right)$. It can be noticed that, among the models in the Planck 2013 inconclusive zone, those with a minimal number of unconstrained parameters are: $\mathrm{ESI}_{\sqrt{2}}$, $\mathrm{ESI}_{\sqrt{2/3}}$, HI, $\mathrm{BI}_{2\mathrm{s}}$, $\mathrm{RGI}_{\mathrm{s}}$, AI, $\mathrm{BI}_{1\mathrm{s}}$, MHI, RGI, $\mathrm{SFI}_{4\mathrm{l}}$, $\mathrm{LI}_{\alpha>0}$, SFI_4, $\mathrm{ESI}_{\mathrm{l}}$, $\mathrm{RGI}_{\mathrm{l}}$, $\mathrm{MHI}_{\mathrm{l}}$, $\mathrm{SBI}_{\alpha_{\rm min}}$ and $\mathrm{SFI}_{3\mathrm{l}}$. The number of preferred models is now 17, that is to say $\sim 9\,\%$ of the total number of models analyzed here. Of course, as already remarked, these models are all of the plateau shape. The distribution of models in the four Jeffreys categories versus the number of unconstrained parameters is summarized in Fig. 20.

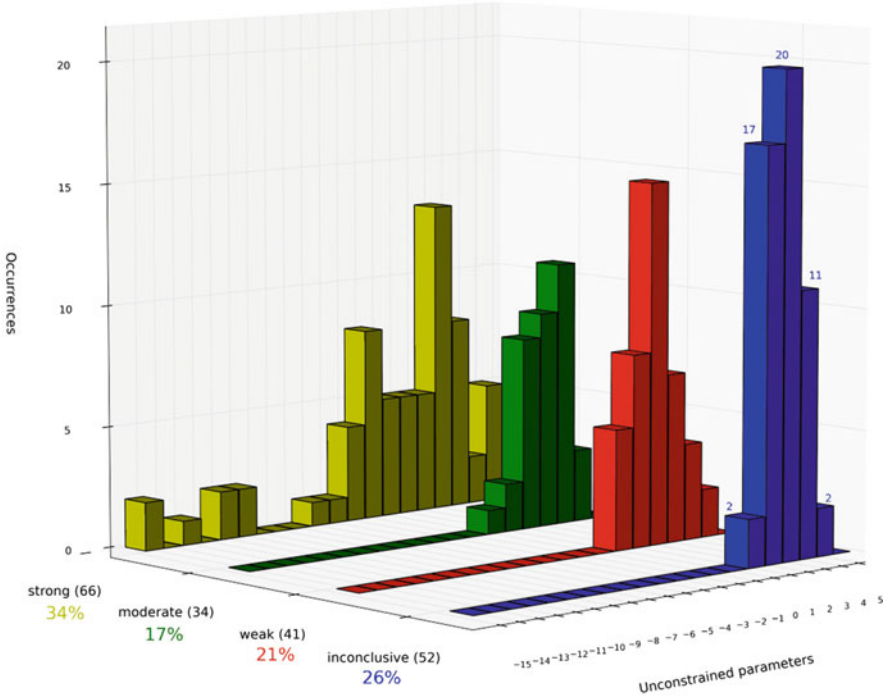

Fig. 20 Occurrences of inflationary models in the four Jeffreys categories for different values of the unconstrained number of parameters. Figure taken from Martin (2013)

5.6 Reheating

We now describe the constraints on reheating that can be inferred from the Planck 2013 data. This question was recently studied in Martin et al. (2015). In Sect. 3.3, we have seen that, as far as CMB data are concerned, reheating can be entirely described by the parameter $R_{\rm rad}$, see Eq. (71) or, equivalently, $R_{\rm reh}$, see Eq. (70). For each inflationary model, Martin et al. (2015) has calculated the posterior distribution of the parameter $\ln R_{\rm reh}$. In order to estimate how much reheating is constrained, it is convenient to introduce the ratio $\Delta\pi_{\ln R_{\rm reh}}/\Delta\mathcal{P}_{\ln R_{\rm reh}}$. In this formula, $\Delta\pi_{\ln R_{\rm reh}}$ is the standard width of the prior while $\Delta\mathcal{P}_{\ln R_{\rm reh}}$ is the standard width of the posterior distribution. Therefore if $\Delta\pi_{\ln R_{\rm reh}}/\Delta\mathcal{P}_{\ln R_{\rm reh}} = 1$, the posterior is as wide as the prior and reheating is not constrained at all. If, however, $\Delta\pi_{\ln R_{\rm reh}}/\Delta\mathcal{P}_{\ln R_{\rm reh}} > 1$, then the posterior distribution is more peaked than the prior and there is information gain. Clearly, the larger the ratio $\Delta\pi_{\ln R_{\rm reh}}/\Delta\mathcal{P}_{\ln R_{\rm reh}}$, the more peaked the posterior.

The prior on $\ln R_{\rm reh}$ has to be chosen carefully and must be justified by physical considerations. Clearly, the energy density at the end of reheating must be smaller than that at the end of inflation and larger than at the BBN time where $\rho_{\rm nuc} = (10\,{\rm MeV})^4$. Therefore, we require $\rho_{\rm nuc} < \rho_{\rm reh} < \rho_{\rm end}$. For the mean equation of

state, we take $-1/3 < \overline{w}_{\rm reh} < 1$ since, by definition, reheating is a non accelerated phase of expansion. As a consequence, one can show that this leads to

$$\ln\left(\frac{\rho_{\rm nuc}^{1/4}}{M_{\rm Pl}}\right) < \ln R_{\rm reh} < \ln\left(\frac{\rho_{\rm nuc}^{1/4}}{M_{\rm Pl}}\right) + \frac{4}{3}\ln\left(\frac{\rho_{\rm end}^{1/4}}{M_{\rm Pl}}\right). \tag{285}$$

The order of magnitude of $R_{\rm reh}$ being unknown, we choose a Jeffreys prior in the above range. Notice that this differs from what was done in the Planck 2013 paper (Ade et al. 2014b). Indeed, in that work, specific reheating scenarios were considered such as instantaneous reheating or "restrictive reheating" where, apparently without a strong justification, the reheating energy density is fixed to 10^9GeV. Moreover, it seems that a prior on the quantity ΔN_* was chosen which is clearly awkward since it does not necessarily guarantee that the two physical conditions on $\rho_{\rm reh}$ and $\overline{w}_{\rm reh}$ discussed previously are valid.[5] In the Planck 2015 paper (Ade et al. 2015a), it seems that this weird approach has been given up. The new method now seems closer to what is done in the present article. Notice, however, that, if the prior on the reheating energy density appears reasonable, only specific values of $\overline{w}_{\rm reh}$ are considered which is, of course, not the most general case. Let us also remark that Ade et al. (2015a) introduces an equation of state parameter during reheating, denoted $w_{\rm int}$, called the "effective equation of state" but without defining it precisely. In particular, it is difficult to know if it is equal to the parameter introduced in Eq. (64), which is the correct parameter that ought to be used and was introduced for the first time in Martin and Ringeval (2010).

In Fig. 21, we have represented each *Encyclopaedia Inflationaris* model in the space $\left(\Delta\pi_{\ln R_{\rm reh}}/\Delta\mathcal{P}_{\ln R_{\rm reh}}, \ln B^i_{\rm REF}\right)$: good models are on the right and models for which reheating is constrained are on the top. The horizontal dashed line $\Delta\pi_{\ln R_{\rm reh}}/\Delta\mathcal{P}_{\ln R_{\rm reh}} = 1$ locates the models for which reheating is not constrained. In order to globally assess the value of the constraints, we can define the following quantity

$$\left\langle\frac{\Delta\pi_{\ln R_{\rm reh}}}{\Delta\mathcal{P}_{\ln R_{\rm reh}}}\right\rangle \equiv \frac{1}{\sum_j \mathcal{E}_j}\sum_i \mathcal{E}_i\left(\frac{\Delta\pi_{\ln R_{\rm reh}}}{\Delta\mathcal{P}_{\ln R_{\rm reh}}}\right)_i, \tag{286}$$

which is the mean value of $\Delta\pi_{\ln R_{\rm reh}}/\Delta\mathcal{P}_{\ln R_{\rm reh}}$ weighted by the Bayesian evidence, i.e. the mean value in the space of models. This is a fair estimate since disfavored models will not contribute a lot to this quantity due to their small evidence.

[5] An additional problem comes from the description of Martin and Ringeval (2010) by Ade et al. (2014b). Indeed it is claimed in this last paper that, for large field models where $V(\phi) \sim \phi^n$, Martin and Ringeval (2010) considered only scenarios of reheating for which $\overline{w}_{\rm reh} = (n-2)/(n+2)$, a wrong claim as can be checked directly by reading Martin and Ringeval (2010).

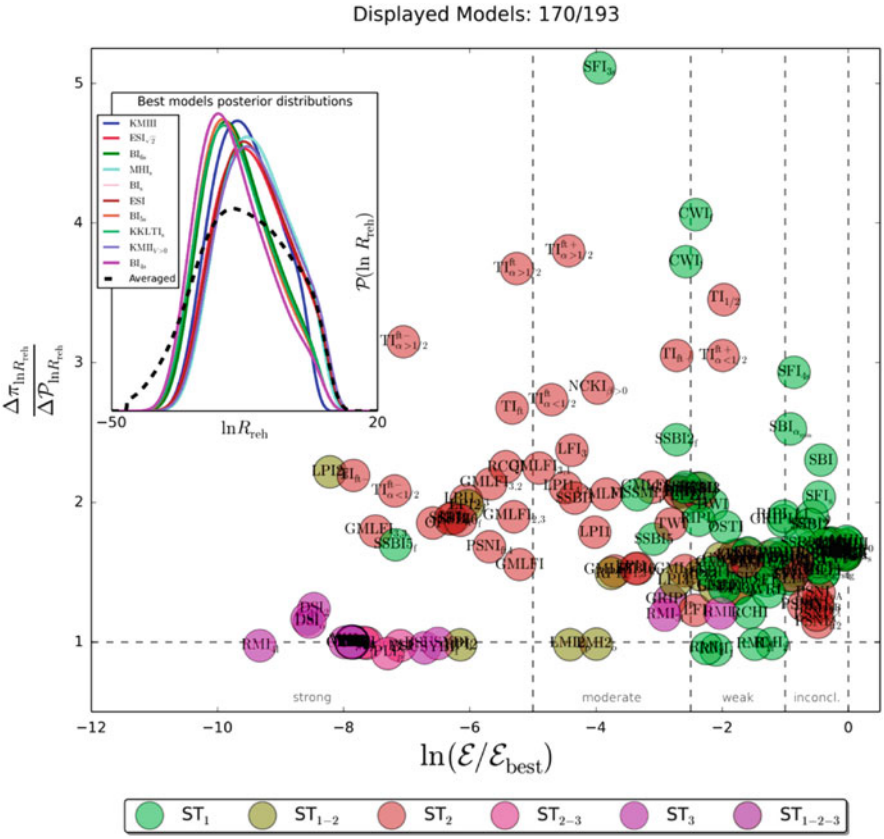

Fig. 21 The quantity $\Delta\pi_{\ln R_{\rm reh}}/\Delta\mathcal{P}_{\ln R_{\rm reh}}$, quantifying how much the reheating is constrained, versus the Bayesian evidence for *Encyclopaedia Inflationaris* models (each model is represented by a circle the size of which has no meaning). The *inset* shows the posterior distribution of the reheating parameter for the ten best Planck 2013 models. Figure taken from Martin et al. (2015)

Numerically, the Planck 2013 data are such that

$$\left\langle\frac{\Delta\pi_{\ln R_{\rm reh}}}{\Delta\mathcal{P}_{\ln R_{\rm reh}}}\right\rangle \simeq 1.66 \tag{287}$$

which, therefore, indicates that reheating is indeed constrained.

It is also interesting to assume that the mean equation of state is known. In that case, the parameter $R_{\rm reh}$ only depends on the energy density at the end of reheating or, equivalently, on the reheating temperature. In Fig. 22, we have represented the similar quantities as in Fig. 21 for four different values of $\overline{w}_{\rm reh}$, namely $\overline{w}_{\rm reh} = -0.3, -0.2, 0, 0.2$. The most striking feature of this plot is that, for positive values of $\overline{w}_{\rm reh}$, the models tend to cluster around the horizontal line $\Delta\pi_{\ln R_{\rm reh}}/\Delta\mathcal{P}_{\ln R_{\rm reh}} = 1$. This indicates that, for those values of the mean equation of state, reheating is not

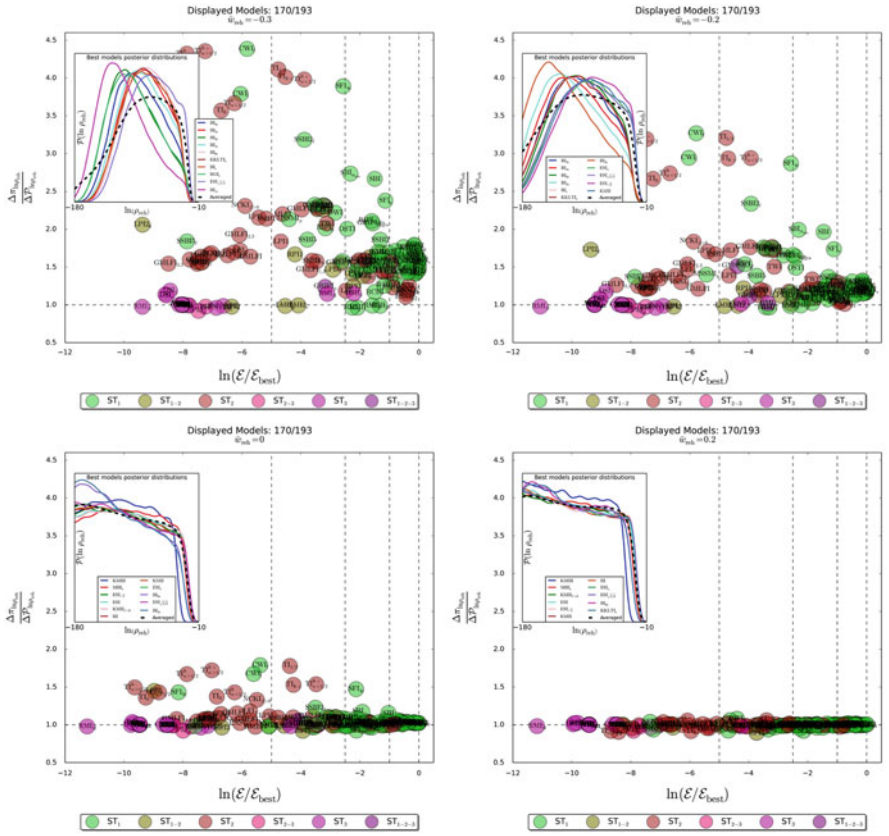

Fig. 22 Same as in Fig. 21 but assuming the mean equation of state during reheating is known. The prior-to-posterior width for the reheating energy density $\ln(\rho_{\rm reh}/M_{\rm Pl}^4)$ is represented assuming four values of the mean equation of state $\overline{w}_{\rm reh}$, namely $\overline{w}_{\rm reh} = -0.3$ (*top left panel*), $\overline{w}_{\rm reh} = -0.2$ (*top right panel*), $\overline{w}_{\rm reh} = 0$ (*bottom left panel*) and $\overline{w}_{\rm reh} = 0.2$ (*bottom right panel*). Figures taken from Martin et al. (2015)

constrained. As a consequence, the number obtained in Eq. (287) comes in fact from a region in parameter space where $\overline{w}_{\rm reh} < 0$. This conclusion makes sense since, for $\overline{w}_{\rm reh} < 0$, the dispersion of the predictions in the $(r, n_{\rm S})$ space is much bigger than for positive equation of state. More details can be found in Martin et al. (2015), in particular concrete bounds on the reheating temperature for different models.

Concluding, the reheating phase is already constrained by the Planck 2013 data. The precise values of the allowed reheating temperatures depend on the model under consideration and on the mean equation of state. If $\overline{w}_{\rm reh} > 0$, the constraints are very mild. It is also worth noticing that two identical models with two different reheating histories can have different Bayesian evidence. This means that, given the accuracy of the CMB measurements, reheating now needs to be properly included in data analysis.

6 Conclusion

In this last section, we briefly summarize what we have learned about inflation in the recent years. Inflation is a "violent" phenomenon since it could occur at energies as high as the Grand Unified Theory scale, i.e. $\sim 10^{16}$ GeV. It is thus quite remarkable to be able to say something about physics at such a high energy scale. The picture that seems to emerge from the recent high accuracy astrophysical measurements is that inflation is realized in its simplest version, namely single-field slow-roll with a minimal kinetic term. Additional features, such as the presence of several fields or non-minimal kinetic term, which may appear as (natural) consequences of embedding inflation in high energy physics, do not seem to be relevant. If, indeed, inflation is really realized in its vanilla version, an important challenge will be (is) to understand, from the high energy point of view, why these extra ingredients are in fact not present. Also, important questions such as the physical nature of the inflaton field remains unanswered.

The shape of the potential is also constrained and appears to be of the "plateau shape", a typical example of this class of scenarios being the Starobinsky model. Popular models such as monomial potentials are now disfavored.

Interestingly enough, inflationary reheating is also constrained by the Planck data. The constraints are model dependent and correspond to an average reduction of the prior-to-posterior of about 40 %.

Given this situation, what should be done to increase our knowledge of inflation? It is clear that in order to measure more precisely the shape of the potential, one needs to constrain the values of the Hubble flow parameters ϵ_n. So far, we only have a good measurement of the scalar spectral index which is a specific combination of ϵ_1 and ϵ_2, namely $n_{\rm S} = 1 - 2\epsilon_1 - \epsilon_2$. To measure ϵ_1 and ϵ_2 separately, one needs another observable. A more accurate measurement of the scalar power spectrum cannot really do the job since it involves an additional parameter, ϵ_3, see Eq. (265). We are therefore left with either the tensor-to-scalar ratio r, which is directly proportional to ϵ_1, see Eq. (87), or the bispectrum which depends on ϵ_1 and ϵ_2 in a different combination than the spectral index, see Eqs. (253) and (249)–(252).

Measuring primordial Non-Gaussianities has one great advantage: we already know in advance where one should find the signal. If one dares an analogy, it is like searching for the Higgs boson. We know that if it is not found in a specific window, the consequences would be drastic. However, the shortcoming is that the amplitude of the signal, $f_{\rm NL} \simeq 0.01$, is so small that it is not clear whether it is technologically feasible. On the other hand, improving the limits on Non-Gaussianities could be very rewarding. Many non-vanilla scenarios predict $f_{\rm NL} \simeq 1$ and reaching this limit could allow us to rule out single field slow-roll models!

Measuring the tensor-to-scalar ratio is the other possibility. It requires to measure the tensor contribution which can be done through a detection of B-mode CMB polarization. At the moment, there are considerable efforts in this direction. The first claim of a detection of primordial gravity waves was of course made by the BICEP2 team (Ade et al. 2014c). The signal corresponds to a tensor-to-scalar ratio

of $r \sim 0.16$. However, as later shown by the Planck team and already discussed before, the signal can probably be entirely explained by dust emission (Ade et al. 2016). Other ground based experiments are currently operating in Antarctica such as BICEP3 & Keck (three channels: 100, 150 and 200 GHz, sky coverage of 1–2 % and resolution of $30'$), SPTPol/SPT$_3$G (90 GHz and 150 GHz, 6 %, $1.2'$), in Chile such as Atacama B-mode Search (ABS) (145 GHz, 2 %, $30'$), Atacama Cosmology Telescope (ACTPol)/AdvACT (30, 40, 90, 150 and 230 GHz, 6 %, $1.4'$), POLARBEAR/SIMONS (90, 150 and 220 GHz, 6 %, 3–$5'$) and in the Canary islands such as QUIJOTE (11–20 and 30 GHz, 65 %, $15'$–$55'$). Soon (2016) in Chile, the experiment Cosmology Large Scale Surveyor (CLASS) (40, 90 and 150 GHz, 70 %) will start taking data. There are also balloon borne experiments such as EBEX (150, 250 and 410 GHz, 8 %, $10'$) and SPIDER (90, 150 and 280 GHz, 8 %, $30'$–$40'$) which are operating in Antarctica and Primordial Inflation Polarization Explorer (PIPER) (200, 270, 350 and 600 GHz, 70 %, $10'$–$20'$) which will be starting in 2016 in Palestine in USA (Texas). The most efficient of these experiments will reach a level corresponding to $r \sim 0.01$ in the following 3–5 years. If one wants to go further, one needs space missions. Two projects appear to be particularly promising: the Lite satellite for the studies of *B*-modes polarization and Inflation from cosmic background Radiation Detection (LiteBIRD) (Matsumura et al. 2014) selected as one of the prioritized projects in the master plan 2014 by the Science Council of Japan and the Cosmic Origins Explorer (COrE+) (The COrE Collaboration 2011) which is a proposal for European Space Agency (ESA) M4 space mission. LiteBIRD has a polarization sensitivity of $\sim 4.5\,\mu\mathrm{K} \times$ arcmin, a resolution of $\theta_{\mathrm{fwhm}} = 38.5'$ and a sky coverage of 70 %. COrE+ can be "light" with a sensitivity of $\sim 2.5\,\mu\mathrm{K}\times$arcmin and a resolution of $\theta_{\mathrm{fwhm}} = 6'$ or "extended" with a sensitivity of $\sim 1.5\,\mu\mathrm{K}\times$arcmin and a resolution of $\theta_{\mathrm{fwhm}} = 4'$ (in both cases, the sky coverage is 70 %). With these space missions, one should be able to gain one order of magnitude on r and reach $r \sim 10^{-3}$ in the next decade, assuming no delensing. With delensing, one might be able to probe even smaller values of r.

Using the above analogy, measuring r is like searching for super-symmetry. We do not know at which level it should show up (we do not know the super-symmetry breaking scale) but it could be around the corner and, hence, technologically realistic. In fact, a determination of r would immediately lead to the inflaton field excursion. An excursion which is just Planckian corresponds to a tensor-to-scalar ratio of $r \sim 10^{-3}$ that is to say precisely the limit reached by future space missions. Therefore, given that $r = 16\epsilon_1$, if $\epsilon_1 \gtrsim 10^{-4}$, then one should be able to measure it in the next decade.

Of course, a detection of primordial gravity waves would also impact model comparison. It was recently shown in Martin et al. (2014e) that this could allow us to rule out almost three-quarters of the inflationary models compared to one-third for Planck 2013.

In conclusion, detecting *B*-mode CMB polarization and, hence, primordial gravity waves, is probably the next challenge for primordial Cosmology. An additional step would then be to check the consistency relation, $r = -8n_{\mathrm{T}}$, which would constitute the final proof that vanilla inflation occurred in the early Universe.

However, if r is very small, this measurement might be too difficult. In any case, at the time of writing, detecting primordial gravity waves appears to be the next frontier for inflation. Only time will tell whether this is true or not.

Acknowledgements It is a pleasure to thank J. Fabris for inviting me to lecture at this school and for his hospitality. I am very grateful to all the participants for very interesting discussions. I also thank P. Peter, C. Ringeval L. Sriramkumar and V. Vennin for careful reading of the manuscript. I thank J. Schwab for having given me the permission to reproduce Figs. 1, 2 and 3, that were originally made and published in Schwab et al. (2008) and Rappaport et al. (2008).

References

Ade, P.A.R., et al.: Planck 2013 Results. XXIV. Constraints on primordial non-Gaussianity (2013). arXiv:1303.5084

Ade, P.A.R., et al.: Planck 2013 results. XVI. Cosmological parameters. Astron. Astrophys. **571**, A16 (2014a). arXiv:1303.5076

Ade, P.A.R., et al.: Planck 2013 results. XXII. Constraints on inflation. Astron. Astrophys. **571**, A22 (2014b). arXiv:1303.5082

Ade, P.A.R., et al.: Detection of *B*-mode polarization at degree angular scales by BICEP2. Phys. Rev. Lett. **112**(24), 241101 (2014c). arXiv:1403.3985, doi:10.1103/PhysRevLett.112.241101

Ade, P.A.R., et al.: Planck 2015. XX. Constraints on inflation (2015a). arXiv:1502.02114

Ade, P.A.R., et al.: Planck 2015 results. XVII. Constraints on primordial non-Gaussianity (2015b). arXiv:1502.01592

Ade, P.A.R., et al.: Planck intermediate results. XXXIII. Signature of the magnetic field geometry of interstellar filaments in dust polarization maps. Astron. Astrophys. **586**, A136 (2016). arXiv:1411.2271

Alishahiha, M., Silverstein, E., Tong, D.: DBI in the sky. Phys. Rev. **D70**, 123505 (2004). arXiv:hep-th/0404084, doi:10.1103/PhysRevD.70.123505

Amendola, L., Gordon, C., Wands, D., Sasaki, M.: Correlated perturbations from inflation and the cosmic microwave background. Phys. Rev. Lett. **88**, 211302 (2002). arXiv:astro-ph/0107089, doi:10.1103/PhysRevLett.88.211302

Amin, M.A., Hertzberg, M.P., Kaiser, D.I., Karouby, J.: Nonperturbative dynamics of reheating after inflation: a review. Int. J. Mod. Phys. **D24**, 1530003 (2015). arXiv:1410.3808, doi:10.1142/S0218271815300037

Arroja, F., Koyama, K.: Non-Gaussianity from the trispectrum in general single field inflation. Phys. Rev. **D77**, 083517 (2008). arXiv:0802.1167, doi:10.1103/PhysRevD.77.083517

Bardeen, J.M.: Gauge invariant cosmological perturbations. Phys. Rev. **D22**, 1882–1905 (1980). doi:10.1103/PhysRevD.22.1882

Bardeen, J.M., Steinhardt, P.J., Turner, M.S.: Spontaneous creation of almost scale - free density perturbations in an inflationary universe. Phys. Rev. **D28**, 679 (1983). doi:10.1103/PhysRevD.28.679

Bartolo, N., Matarrese, S., Riotto, A.: Adiabatic and isocurvature perturbations from inflation: power spectra and consistency relations. Phys. Rev. **D64**, 123504 (2001). arXiv:astro-ph/0107502, doi:10.1103/PhysRevD.64.123504

Berera, A.: Warm inflation. Phys. Rev. Lett. **75**, 3218–3221 (1995). arXiv:astro-ph/9509049, doi:10.1103/PhysRevLett.75.3218

Berera, A., Moss, I.G., Ramos, R.O.: Warm inflation and its microphysical basis. Rep. Prog. Phys. **72**, 026901 (2009). arXiv:0808.1855, doi:10.1088/0034-4885/72/2/026901

Bezrukov, F.L., Shaposhnikov, M.: The standard model Higgs boson as the inflaton. Phys. Lett. **B659**, 703–706 (2008). arXiv:0710.3755, doi:10.1016/j.physletb.2007.11.072

BICEP2/Keck, et al.: A joint analysis of BICEP2/Keck array and planck data. Phys. Rev. Lett. **114**, 101301 (2015). arXiv:1502.00612

Binetruy, P., Kiritsis, E., Mabillard, J., Pieroni, M., Rosset, C.: Universality classes for models of inflation. J. Cosmol. Astropart. Phys. **1504**(04), 033 (2015). arXiv:1407.0820

Bucher, M.: Physics of the cosmic microwave background anisotropy (2015). arXiv:1501.04288, doi:10.1142/S0218271815300049

Bucher, M., Moodley, K., Turok, N.: The general primordial cosmic perturbation. Phys. Rev. **D62**, 083508 (2000). arXiv:astro-ph/9904231, doi:10.1103/PhysRevD.62.083508

Byrnes, C.T.: Lecture notes on non-Gaussianity (2014). arXiv:1411.7002

Byrnes, C.T., Wands, D.: Curvature and isocurvature perturbations from two-field inflation in a slow-roll expansion. Phys. Rev. **D74**, 043529 (2006). arXiv:astro-ph/0605679, doi:10.1103/PhysRevD.74.043529

Byrnes, C.T., Sasaki, M., Wands, D.: The primordial trispectrum from inflation. Phys. Rev. **D74**, 123519 (2006). arXiv:astro-ph/0611075, doi:10.1103/PhysRevD.74.123519

Casadio, R., Finelli, F., Luzzi, M., Venturi, G.: Improved WKB analysis of cosmological perturbations. Phys. Rev. **D71**, 043517 (2005a). arXiv:gr-qc/0410092, doi:10.1103/PhysRevD.71.043517

Casadio, R., Finelli, F., Luzzi, M., Venturi, G.: Higher order slow-roll predictions for inflation. Phys. Lett. **B625**, 1–6 (2005b). arXiv:gr-qc/0506043, doi:10.1016/j.physletb.2005.08.056

Casadio, R., Finelli, F., Luzzi, M., Venturi, G.: Improved WKB analysis of slow-roll inflation. Phys. Rev. **D72**, 103516 (2005c). arXiv:gr-qc/0510103, doi:10.1103/PhysRevD.72.103516

Chen, X.: Primordial Non-Gaussianities from inflation models. Adv. Astron. **2010**, 638979 (2010). arXiv:1002.1416, doi:10.1155/2010/638979

Chen, X., Huang, M.-X., Kachru, S., Shiu, G.: Observational signatures and non-Gaussianities of general single field inflation. J. Cosmol. Astropart. Phys. **0701**, 002 (2007). arXiv:hep-th/0605045, doi:10.1088/1475-7516/2007/01/002

Chialva, D., Mazumdar, A.: Super-Planckian excursions of the inflaton and quantum corrections (2014). arXiv:1405.0513, doi:10.1142/S0217732315400088

Choe, J., Gong, J.-O., Stewart, E.D.: Second order general slow-roll power spectrum. J. Cosmol. Astropart. Phys. **0407**, 012 (2004). arXiv:hep-ph/0405155, doi:10.1088/1475-7516/2004/07/012

Choi, K.-Y., Gong, J.-O., Jeong, D.: Evolution of the curvature perturbation during and after multi-field inflation. J. Cosmol. Astropart. Phys. **0902**, 032 (2009). arXiv:0810.2299, doi:10.1088/1475-7516/2009/02/032

Chongchitnan, S., Efstathiou, G.: Dynamics of the inflationary flow equations. Phys. Rev. **D72**, 083520 (2005). arXiv:astro-ph/0508355, doi:10.1103/PhysRevD.72.083520

Cicoli, M., Burgess, C.P., Quevedo, F.: Fibre inflation: observable gravity waves from IIB string compactifications. J. Cosmol. Astropart. Phys. **0903**, 013 (2009). arXiv:0808.0691, doi:10.1088/1475-7516/2009/03/013

Conlon, J.P., Quevedo, F.: Kahler moduli inflation. J. High Energy Phys. **0601**, 146 (2006). arXiv:hep-th/0509012, doi:10.1088/1126-6708/2006/01/146

Dai, L., Kamionkowski, M., Wang, J.: Reheating constraints to inflationary models. Phys. Rev. Lett. **113**, 041302 (2014). arXiv:1404.6704, doi:10.1103/PhysRevLett.113.041302

Dvali, G.R., Tye, S.H.H.: Brane inflation. Phys. Lett. **B450**, 72–82 (1999). arXiv:hep-ph/9812483, doi:10.1016/S0370-2693(99)00132-X

Feng, C., Cooray, A., Smidt, J., O'Bryan, J., Keating, B., et al.: Planck trispectrum constraints on primordial Non-Gaussianity at cubic order. Phys. Rev. **D92**(4), 043509 (2015). arXiv:1502.00585

Gangui, A.: NonGaussian effects in the cosmic microwave background from inflation. Phys. Rev. **D50**, 3684–3691 (1994). arXiv:astro-ph/9406014, doi:10.1103/PhysRevD.50.3684

Gangui, A., Martin, J.: Cosmic microwave background bispectrum and slow roll inflation. Mon. Not. R. Astron. Soc. **313**, 323 (2000a). arXiv:astro-ph/9908009, doi:10.1046/j.1365-8711.2000.03210.x

Gangui, A., Martin, J.: Best unbiased estimators for the three point correlators of the cosmic microwave background radiation. Phys. Rev. **D62**, 103004 (2000b). arXiv:astro-ph/0001361, doi:10.1103/PhysRevD.62.103004

Gangui, A., Lucchin, F., Matarrese, S., Mollerach, S.: The three point correlation function of the cosmic microwave background in inflationary models. Astrophys. J. **430**, 447–457 (1994). arXiv:astro-ph/9312033, doi:10.1086/174421

Gangui, A., Martin, J., Sakellariadou, M.: Single field inflation and non-Gaussianity. Phys. Rev. **D66**, 083502 (2002). arXiv:astro-ph/0205202, doi:10.1103/PhysRevD.66.083502

Garcia-Bellido, J., Roest, D.: Large-*N* running of the spectral index of inflation. Phys. Rev. **D89**(10), 103527 (2014). arXiv:1402.2059, doi:10.1103/PhysRevD.89.103527

Giannantonio, T., Komatsu, E.: Bayesian evidence of nonstandard inflation: isocurvature perturbations and running spectral index. Phys. Rev. **D91**(2), 023506 (2015). arXiv:1407.4291, doi:10.1103/PhysRevD.91.023506

Gong, J.-O., Stewart, E.D.: The density perturbation power spectrum to second order corrections in the slow roll expansion. Phys. Lett. **B510**, 1–9 (2001). arXiv:astro-ph/0101225

Gong, J.-O., Pi, S., Leung, G.: Probing reheating with primordial spectrum. J. Cosmol. Astropart. Phys. **1505**(05), 027 (2015). arXiv:1501.03604

Gordon, C., Lewis, A.: Observational constraints on the curvaton model of inflation. Phys. Rev. **D67**, 123513 (2003). arXiv:astro-ph/0212248, doi:10.1103/PhysRevD.67.123513

Gordon, C., Wands, D., Bassett, B.A., Maartens, R.: Adiabatic and entropy perturbations from inflation. Phys. Rev. **D63**, 023506 (2001). arXiv:astro-ph/0009131, doi:10.1103/PhysRevD.63.023506

Grishchuk, L.P., Martin, J. Best unbiased estimates for the microwave background anisotropies. Phys. Rev. **D56**, 1924–1938 (1997). arXiv:gr-qc/9702018, doi:10.1103/PhysRevD.56.1924

Guth, A.H.: The inflationary universe: a possible solution to the horizon and flatness problems. Phys. Rev. **D23**, 347–356 (1981). doi:10.1103/PhysRevD.23.347.

Hazra, D.K., Martin, J., Sriramkumar, L.: The scalar bi-spectrum during preheating in single field inflationary models. Phys. Rev. **D86**, 063523 (2012). arXiv:1206.0442, doi:10.1103/PhysRevD.86.063523

Hazra, D.K., Sriramkumar, L., Martin, J.: BINGO: a code for the efficient computation of the scalar bi-spectrum. J. Cosmol. Astropart. Phys. **1305**, 026 (2013). arXiv:1201.0926, doi:10.1088/1475-7516/2013/05/026

Hoffman, M.B., Turner, M.S.: Kinematic constraints to the key inflationary observables. Phys. Rev. **D64**, 023506 (2001). arXiv:astro-ph/0006321, doi:10.1103/PhysRevD.64.023506

Hotchkiss, S., Sarkar, S.: Non-Gaussianity from violation of slow-roll in multiple inflation. J. Cosmol. Astropart. Phys. **1005**, 024 (2010). arXiv:0910.3373, doi:10.1088/1475-7516/2010/05/024

Hu, W., Sugiyama, N.: Small scale cosmological perturbations: an analytic approach. Astrophys. J. **471**, 542–570 (1996). arXiv:astro-ph/9510117, doi:10.1086/177989

Ijjas, A., Steinhardt, P.J., Loeb, A.: Inflationary paradigm in trouble after Planck2013. Phys. Lett. **B723**, 261–266 (2013). arXiv:1304.2785, doi:10.1016/j.physletb.2013.05.023

Jimenez, J.B., Musso, M., Ringeval, C.: Exact mapping between tensor and most general scalar power spectra. Phys. Rev. **D88**, 043524 (2013). arXiv:1303.2788, doi:10.1103/PhysRevD.88.043524

Kinney, W.H.: Inflation: flow, fixed points and observables to arbitrary order in slow roll. Phys. Rev. **D66**, 083508 (2002). arXiv:astro-ph/0206032, doi:10.1103/PhysRevD.66.083508

Kofman, L., Linde, A.D., Starobinsky, A.A.: Towards the theory of reheating after inflation. Phys. Rev. **D56**, 3258–3295 (1997). arXiv:hep-ph/9704452, doi:10.1103/PhysRevD.56.3258

Komatsu, E.: Hunting for primordial non-Gaussianity in the cosmic microwave background. Class. Quant. Grav. **27**(12), 124010 (2010). arXiv:1003.6097, doi:10.1088/0264-9381/27/12/124010

Kunz, M., Trotta, R., Parkinson, D.: Measuring the effective complexity of cosmological models. Phys. Rev. **D74**, 023503 (2006). arXiv:astro-ph/0602378, doi:10.1103/PhysRevD.74.023503

Langlois, D.: Correlated adiabatic and isocurvature perturbations from double inflation. Phys. Rev. **D59**, 123512 (1999). arXiv:astro-ph/9906080, doi:10.1103/PhysRevD.59.123512

Langlois, D., Riazuelo, A. Correlated mixtures of adiabatic and isocurvature cosmological perturbations. Phys. Rev. **D62**, 043504 (2000). arXiv:astro-ph/9912497, doi:10.1103/PhysRevD.62.043504

Leach, S.M., Liddle, A.R., Martin, J., Schwarz, D.J.: Cosmological parameter estimation and the inflationary cosmology. Phys. Rev. **D66**, 023515 (2002). arXiv:astro-ph/0202094, doi:10.1103/PhysRevD.66.023515

Liddle, A.R.: Inflationary flow equations. Phys. Rev. **D68**, 103504 (2003). arXiv:astro-ph/0307286, doi:10.1103/PhysRevD.68.103504

Liddle, A.R., Parsons, P., Barrow, J.D.: Formalizing the slow roll approximation in inflation. Phys. Rev. **D50**, 7222–7232 (1994). arXiv:astro-ph/9408015, doi:10.1103/PhysRevD.50.7222

Linde, A.D.: A new inflationary universe scenario: a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. Phys. Lett. **B108**, 389–393 (1982). doi:10.1016/0370-2693(82)91219-9

Linde, A.D., Mezhlumian, A.: Inflation with Omega not = 1. Phys. Rev. **D52**, 6789–6804 (1995). arXiv:astro-ph/9506017, doi:10.1103/PhysRevD.52.6789

Lorenz, L., Martin, J., Ringeval, C.: Brane inflation and the WMAP data: a Bayesian analysis. J. Cosmol. Astropart. Phys. **0804**, 001 (2008a). arXiv:0709.3758, doi:10.1088/1475-7516/2008/04/001

Lorenz, L., Martin, J., Ringeval, C.; K-inflationary power spectra in the uniform approximation. Phys. Rev. **D78**, 083513 (2008b). arXiv:0807.3037, doi:10.1103/PhysRevD.78.083513

Lucchin, F., Matarrese, S.: Power law inflation. Phys. Rev. **D32**, 1316 (1985). doi:10.1103/PhysRevD.32.1316

Maldacena, J.M.: Non-Gaussian features of primordial fluctuations in single field inflationary models. J. High Energy Phys. **0305**, 013 (2003). arXiv:astro-ph/0210603, doi:10.1088/1126-6708/2003/05/013

Martin, J.: Inflation and precision cosmology. Braz. J. Phys. **34**, 1307–1321 (2004). arXiv:astro-ph/0312492

Martin, J.: Inflationary cosmological perturbations of quantum- mechanical origin. Lect. Notes Phys. **669**, 199–244 (2005). arXiv:hep-th/0406011

Martin, J.: Inflationary perturbations: the cosmological Schwinger effect. Lect. Notes Phys. **738**, 193–241 (2008). arXiv:0704.3540, doi:10.1007/978-3-540-74353-8_6

Martin, J.: Inflation after Planck: and the winners are (2013). arXiv:1312.3720

Martin, J., Ringeval, C.: Inflation after WMAP3: confronting the slow-roll and exact power spectra to CMB data. J. Cosmol. Astropart. Phys. **0608**, 009 (2006). arXiv:astro-ph/0605367, doi:10.1088/1475-7516/2006/08/009

Martin, J., Ringeval, C.: First CMB constraints on the inflationary reheating temperature. Phys. Rev. **D82**, 023511 (2010). arXiv:1004.5525, doi:10.1103/PhysRevD.82.023511
Martin, J., Schwarz, D.J.: The influence of cosmological transitions on the evolution of density perturbations. Phys. Rev. **D57**, 3302–3316 (1998). arXiv:gr-qc/9704049, doi:10.1103/PhysRevD.57.3302
Martin, J., Schwarz, D.J.: WKB approximation for inflationary cosmological perturbations. Phys. Rev. **D67**, 083512 (2003). arXiv:astro-ph/0210090, doi:10.1103/PhysRevD.67.083512
Martin, J., Sriramkumar, L.: The scalar bi-spectrum in the Starobinsky model: the equilateral case. J. Cosmol. Astropart. Phys. **1201**, 008 (2012). arXiv:1109.5838, doi:10.1088/1475-7516/2012/01/008
Martin, J., Ringeval, C., Trotta, R.: Hunting down the best model of inflation with bayesian evidence. Phys. Rev. **D83**, 063524 (2011). arXiv:1009.4157, doi:10.1103/PhysRevD.83.063524
Martin, J., Vennin, V., Peter, P.: Cosmological inflation and the quantum measurement problem. Phys. Rev. **D86**, 103524 (2012). arXiv:1207.2086, doi:10.1103/PhysRevD.86.103524
Martin, J., Ringeval, C., Vennin, V.: K-inflationary power spectra at second order. J. Cosmol. Astropart. Phys. **1306**, 021 (2013). arXiv:1303.2120, doi:10.1088/1475-7516/2013/06/021
Martin, J., Ringeval, C., Vennin, V.: Encyclopædia inflationaris. Phys. Dark Univ. (2014a). arXiv:1303.3787
Martin, J., Ringeval, C., Trotta, R., Vennin, V.: The best inflationary models after planck. J. Cosmol. Astropart. Phys. **1403**, 039 (2014b). arXiv:1312.3529, doi:10.1088/1475-7516/2014/03/039
Martin, J., Sriramkumar, L., Hazra, D.K.: Sharp inflaton potentials and bi-spectra: effects of smoothening the discontinuity. J. Cosmol. Astropart. Phys. **1409**(09), 039 (2014c). arXiv:1404.6093, doi:10.1088/1475-7516/2014/09/039
Martin, J., Ringeval, C., Trotta, R., Vennin, V.: Compatibility of planck and BICEP2 in the light of inflation. Phys. Rev. **D90**, 063501 (2014d). arXiv:1405.7272, doi:10.1103/PhysRevD.90.063501
Martin, J., Ringeval, C., Vennin, V.: How well can future CMB missions constrain cosmic inflation? J. Cosmol. Astropart. Phys. **1410**(10), 038 (2014e). arXiv:1407.4034, doi:10.1088/1475-7516/2014/10/038
Martin, J., Ringeval, C., Vennin, V.: Observing the inflationary reheating. Phys. Rev. Lett. **114**(8), 081303 (2015). arXiv:1410.7958
Matsumura, T., Akiba, Y., Borrill, J., Chinone, Y., Dobbs, M., et al.: Mission design of LiteBIRD. J. Low. Temp. Phys. **176**, 733 (2014). arXiv:1311.2847, doi:10.1007/s10909-013-0996-1
Mukhanov, V.: Physical Foundations of Cosmology. Cambridge University Press, Cambridge (2005)
Mukhanov, V.: Quantum cosmological perturbations: predictions and observations. Eur. Phys. J. **C73**, 2486 (2013). arXiv:1303.3925, doi:10.1140/epjc/s10052-013-2486-7
Mukhanov, V.: Inflation without selfreproduction. Fortschr. Phys. **63**, 36 (2015). arXiv:1409.2335
Mukhanov, V.F., Chibisov, G.V.: Quantum fluctuation and nonsingular universe (in Russian). JETP Lett. **33**, 532–535 (1981)
Mukhanov, V.F., Chibisov, G.V.: The vacuum energy and large scale structure of the universe. Sov. Phys. JETP **56**, 258–265 (1982)
Mukhanov, V.F., Feldman, H.A., Brandenberger, R.H.: Theory of cosmological perturbations. Part 1. Classical perturbations. Part 2. Quantum theory of perturbations. Part 3. Extensions. Phys. Rep. **215**, 203–333 (1992). doi:10.1016/0370-1573(92)90044-Z
Munoz, J.B., Kamionkowski, M.: The equation-of-state parameter for reheating. Phys. Rev. D **91**(4), 043521 (2015). arXiv:1412.0656

O'Meara, J.M., Burles, S., Prochaska, J.X., Prochter, G.E., Bernstein, R.A., et al.: The deuterium to hydrogen abundance ratio towards the QSO SDSS1558-0031. Astrophys. J. **649**, L61–L66 (2006). arXiv:astro-ph/0608302, doi:10.1086/508348

Peskin, M.E., Schroeder, D.V.: An Introduction to Quantum Field Theory. Addison-Wesley, Reading, MA (1995)

Peter, P.: Cosmological perturbation theory (2013). arXiv:1303.2509

Peter, P., Uzan, J.-P.: Primordial Cosmology. Oxford Graduate Texts. Oxford University Press, Oxford (2009)

Peter, P., Polarski, D., Starobinsky, A.A.: Confrontation of double inflationary models with observations. Phys. Rev. **D50**, 4827–4834 (1994). arXiv:astro-ph/9403037, doi:10.1103/PhysRevD.50.4827

Planck 2015 results. XIII. Cosmological parameters (2015). arXiv:1502.01589

Podolsky, D.I., Felder, G.N., Kofman, L., Peloso, M.: Equation of state and beginning of thermalization after preheating. Phys. Rev. **D73**, 023501 (2006). arXiv:hep-ph/0507096, doi:10.1103/PhysRevD.73.023501

Polarski, D., Starobinsky, A.A.: Spectra of perturbations produced by double inflation with an intermediate matter dominated stage. Nucl. Phys. **B385**, 623–650 (1992). doi:10.1016/0550-3213(92)90062-G

Polarski, D., Starobinsky, A.A.: Isocurvature perturbations in multiple inflationary models. Phys. Rev. **D50**, 6123–6129 (1994). arXiv:astro-ph/9404061, doi:10.1103/PhysRevD.50.6123

Polarski, D., Starobinsky, A.A.: Semiclassicality and decoherence of cosmological perturbations. Class. Quant. Grav. **13**, 377–392 (1996). arXiv:gr-qc/9504030, doi:10.1088/0264-9381/13/3/006

Ramirez, E., Liddle, A.R.: Stochastic approaches to inflation model building. Phys. Rev. **D71**, 123510 (2005). arXiv:astro-ph/0502361, doi:10.1103/PhysRevD.71.123510

Rappaport, S., Schwab, J., Burles, S., Steigman, G.: Big bang nucleosynthesis constraints on the self-gravity of pressure. Phys. Rev. **D77**, 023515 (2008). arXiv:0710.5300, doi:10.1103/PhysRevD.77.023515

Renaux-Petel, S.: On the squeezed limit of the bispectrum in general single field inflation. J. Cosmol. Astropart. Phys. **10**, 20 (2010). arXiv:1008.0260, doi:10.1088/1475-7516/2010/10/020

Ringeval, C.: Fast Bayesian inference for slow-roll inflation. Mon. Not. R. Astron. Soc. **439**, 3253 (2014). arXiv:1312.2347, doi:10.1093/mnras/stu109

Roest, D.: Universality classes of inflation. J. Cosmol. Astropart. Phys. **1401**(01), 007 (2014). arXiv:1309.1285, doi:10.1088/1475-7516/2014/01/007

Schwab, J., Hughes, S.A., Rappaport, S.: The self-gravity of pressure in neutron stars (2008). arXiv:0806.0798

Schwarz, D.J., Terrero-Escalante, C.A., Garcia, A.A.: Higher order corrections to primordial spectra from cosmological inflation. Phys. Lett. **B517**, 243–249 (2001). arXiv:astro-ph/0106020, doi:10.1016/S0370-2693(01)01036-X

Seery, D., Lidsey, J.E.: Primordial non-Gaussianities in single field inflation. J. Cosmol. Astropart. Phys. **0506**, 003 (2005). arXiv:astro-ph/0503692, doi:10.1088/1475-7516/2005/06/003

Seery, D., Lidsey, J.E., Sloth, M.S.: The inflationary trispectrum. J. Cosmol. Astropart. Phys. **0701**, 027 (2007). arXiv:astro-ph/0610210, doi:10.1088/1475-7516/2007/01/027

Sreenath, V., Sriramkumar, L.: Examining the consistency relations describing the three-point functions involving tensors. J. Cosmol. Astropart. Phys. **1410**(10), 021 (2014). arXiv:1406.1609, doi:10.1088/1475-7516/2014/10/021

Sreenath, V., Tibrewala, R., Sriramkumar, L.: Numerical evaluation of the three-point scalar-tensor cross-correlations and the tensor bi-spectrum. J. Cosmol. Astropart. Phys. **1312**, 037 (2013). arXiv:1309.7169, doi:10.1088/1475-7516/2013/12/037

Sreenath, V., Hazra, D.K., Sriramkumar, L.: On the scalar consistency relation away from slow roll. J. Cosmol. Astropart. Phys. **1502**(02), 029 (2015). arXiv:1410.0252

Sriramkumar, L.: An introduction to inflation and cosmological perturbation theory (2009). arXiv:0904.4584

Starobinsky, A.A.: Spectrum of relict gravitational radiation and the early state of the universe. JETP Lett. **30**, 682–685 (1979)

Starobinsky, A.A.: A new type of isotropic cosmological models without singularity. Phys. Lett. **B91**, 99–102 (1980). doi:10.1016/0370-2693(80)90670-X.

Starobinsky, A.A.: Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations. Phys. Lett. **B117**, 175–178 (1982). doi:10.1016/0370-2693(82)90541-X

Steigman, G.: Primordial nucleosynthesis in the precision cosmology era. Ann. Rev. Nucl. Part. Sci. **57**, 463–491 (2007). arXiv:0712.1100, doi:10.1146/annurev.nucl.56.080805.140437

Stewart, E.D.: Inflation, supergravity and superstrings. Phys. Rev. **D51**, 6847–6853 (1995). arXiv:hep-ph/9405389, doi:10.1103/PhysRevD.51.6847

Suyama, T., Takahashi, T., Yamaguchi, M., Yokoyama, S.: On classification of models of large local-type Non-Gaussianity. J. Cosmol. Astropart. Phys. **1012**, 030 (2010). arXiv:1009.1979, doi:10.1088/1475-7516/2010/12/030

The COrE Collaboration, Armitage-Caplan, C., Avillez, M., Barbosa, D., Banday, A., Bartolo, N., Battye, R., Bernard, J., de Bernardis, P., Basak, S., Bersanelli, M., Bielewicz, P., Bonaldi, A., Bucher, M., Bouchet, F., Boulanger, F., Burigana, C., Camus, P., Challinor, A., Chongchitnan, S., Clements, D., Colafrancesco, S., Delabrouille, J., De Petris, M., De Zotti, G., Dickinson, C., Dunkley, J., Ensslin, T., Fergusson, J., Ferreira, P., Ferriere, K., Finelli, F., Galli, S., Garcia-Bellido, J., Gauthier, C., Haverkorn, M., Hindmarsh, M., Jaffe, A., Kunz, M., Lesgourgues, J., Liddle, A., Liguori, M., Lopez-Caniego, M., Maffei, B., Marchegiani, P., Martinez-Gonzalez, E., Masi, S., Mauskopf, P., Matarrese, S., Melchiorri, A., Mukherjee, P., Nati, F., Natoli, P., Negrello, M., Pagano, L., Paoletti, D., Peacocke, T., Peiris, H., Perroto, L., Piacentini, F., Piat, M., Piccirillo, L., Pisano, G., Ponthieu, N., Rath, C., Ricciardi, S., Rubino Martin, J., Salatino, M., Shellard, P., Stompor, R., Urrestilla, L.T.J., Van Tent, B., Verde, L., Wandelt, B., Withington. S.: COrE (Cosmic Origins Explorer) a white paper. ArXiv e-prints (2011). arXiv:1102.2181

Traschen, J.H., Brandenberger, R.H.: Particle production during out-of-equilibrium phase transitions. Phys. Rev. **D42**, 2491–2504 (1990). doi:10.1103/PhysRevD.42.2491

Trotta, R.: Bayes in the sky: Bayesian inference and model selection in cosmology. Contemp. Phys. **49**, 71–104 (2008). arXiv:0803.4089, doi:10.1080/00107510802066753

Turner, M.S.: Coherent scalar field oscillations in an expanding universe. Phys. Rev. **D28**, 1243 (1983). doi:10.1103/PhysRevD.28.1243

Vennin, V.: Horizon-flow off-track for inflation. Phys. Rev. **D89**(8), 083526 (2014). arXiv:1401.2926, doi:10.1103/PhysRevD.89.083526

Wands, D.: Multiple field inflation. Lect. Notes Phys. **738**, 275–304 (2008). arXiv:astro-ph/0702187, doi:10.1007/978-3-540-74353-8_8

Wands, D., Bartolo, N., Matarrese, S., Riotto, A.: An observational test of two-field inflation. Phys. Rev. **D66**, 043520 (2002). arXiv:astro-ph/0205253, doi:10.1103/PhysRevD.66.043520

Wang, L.-M., Kamionkowski, M.: The cosmic microwave background bispectrum and inflation. Phys. Rev. **D61**, 063504 (2000). arXiv:astro-ph/9907431, doi:10.1103/PhysRevD.61.063504

Weinberg, S.: Cosmological fluctuations of short wavelength. Astrophys. J. **581**, 810–816 (2002). arXiv:astro-ph/0207375, doi:10.1086/344441

Yokoyama, J., Linde, A.D.: Is warm inflation possible? Phys. Rev. **D60**, 083509 (1999). arXiv:hep-ph/9809409, doi:10.1103/PhysRevD.60.083509

Lecture Notes on Non-Gaussianity

Christian T. Byrnes

Abstract We discuss how primordial non-Gaussianity of the curvature perturbation helps to constrain models of the early universe. Observations are consistent with Gaussian initial conditions, compatible with the predictions of the simplest models of inflation. Deviations are constrained to be at the sub percent level, constraining alternative models such as those with multiple fields, non-canonical kinetic terms or breaking the slow-roll conditions. We introduce some of the most important models of inflation which generate non-Gaussian perturbations and provide practical tools on how to calculate the three-point correlation function for a popular class of non-Gaussian models. The current state of the field is summarised and an outlook is given.

1 Introduction and the Aims of These Lecture Notes

The theory of inflation, a period of quasi-exponential expansion of the universe very shortly after the big bang, is now widely regard as part of the standard cosmological model. The predictions of the simplest inflationary models have passed increasingly stringent tests from observations of the cosmic microwave background (CMB), most recently by the Planck satellite. Remarkably, the apparently crazy idea that the formation of all structures in our universe such as galaxies were caused by quantum perturbations of the field driving inflation, does have strong observational evidence. Inflation therefore provides a mechanism to relate the smallest and largest scales in the universe.

Despite the evidence that inflation occurred, rather little is known about the properties of inflation. The energy scale could be anywhere between the TeV and the GUT scale, an enormous range, notable for stretching far beyond the highest energies we can ever reach with a terrestrial experiment such as a particle collider. This provides cosmologists with the opportunity to provide constraints on extremely high energy physics, as witnessed by the study on the field of embedding inflation

C.T. Byrnes (✉)
Department of Physics and Astronomy, Pevensey II Building, University of Sussex, Brighton BN1 9RH, UK
e-mail: c.byrnes@sussex.ac.uk

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_3

into models of string theory. It is both an opportunity and a challenge that we have limited information about the relevant model of particle physics during inflation. For example, was inflation driven by one or more fields, what form did their potential(s) take, and their kinetic term(s)? What is the energy scale of inflation, and how did the universe become radiation dominated after inflation ended?

Our best way to answer these questions is by studying the statistical properties of the perturbations generated during inflation for different classes of models and then evolving this spectrum of perturbations forwards in time to make predictions for the pattern of temperature perturbations in the CMB, as well as the density perturbations of large scale structure such as galaxy clusters.

Over the past decade, the study of the Gaussianity of the primordial perturbations has become a large field, being the main theme of many focused conferences and workshops every year (Komatsu et al. 2009). The simplest models of inflation are expected to produce perturbations which are extremely close to Gaussian (Maldacena 2003). Any observation of non-Gaussianity would rule out the simplest models. Gaussian perturbations are very tightly constrained by the definition of Gaussianity, for example all information about the correlation functions of the perturbations is encoded in the two-point function alone. By contrast, non-Gaussian perturbations could be anything else. This opens a Pandoras box full of possibilities to search a huge observational data set for anything and everything, which leads to the danger of coincidental patterns in the data being interpreted as signals of real physics. Anomalies of this sort, strange patterns which were not predicted in advance of analysing the data do exist, but all are of a reasonably small statistical significance, especially when taking into account the "look elsewhere" effect, i.e. if 100 people search for different non-Gaussian patterns in a Gaussian data set, then one of these patterns will probably be "detected" at the 99 % confidence level due to the statistical fluctuations in the data.

Fortunately general classes of inflationary models do predict specific shapes of non-Gaussianity, for example certain templates of the bispectrum (the three-point function, which is zero for a Gaussian distribution) would point towards a requirement that multiple fields contributed to the physics of inflation, while a non-canonical kinetic term of the inflaton fields Lagrangian predict a different signature in the bispectrum which is observationally distinct. Classifying different models according to the form of non-Gaussianity they generate, and finding observational constraints on the corresponding scenario has become a major topic within the study of inflation.

The aim of these lectures is to provide both a background knowledge about non-Gaussianity of the primordial perturbations which gave rise to all structures as well as some concrete calculations for reasonably simple scenarios, thereby developing a real working knowledge of the field and providing the tools to perform calculations yourself for some scenarios. The course contents are up to date and much of the content here is not contained in any textbook, at least at the time of publishing these notes.

These lecture proceedings are based on a 4 h lecture course held at the II JBP Cosmology school in Espirito Santo, Brasil. The slides are available from the

school's website (http://www.cosmo-ufes.org/jpbcosmo2-mini-courses-seminars--posters.html). The most closely related courses were on Inflationary Cosmology by Jérôme Martin and CMB theory by David Wands.

Students wishing to learn more may turn to many review articles about non-Gaussianity, here we just list a selection published during or after 2010. However note that all predate the Planck data release which significantly improved the constraints on non-Gaussianity. Reviews which are more focused towards how the observational constraints are made include Komatsu (2010), Yadav and Wandelt (2010), Liguori et al. (2010), Desjacques and Seljak (2010), Verde (2010), Bartolo et al. (2010) and focus on tests relating to isotropy and anomalies are considered in Abramo and Pereira (2010), Copi et al. (2010). Perhaps the most comprehensive review of non-Gaussian inflationary models is given in Chen (2010). Reviews focused on the local model of non-Gaussianity include Wands (2010), Byrnes and Choi (2010).

The plan of the lectures is as follows: We first provide more introduction and motivation for this topic, presented in a novel manner. We then study some specific models of non-Gaussianity, and the classes of models which give rise to them, Sect. 3. In Sect. 4.1 we provide a practical introduction to a simple method of calculation inflationary perturbations to non-linear order, the δN formalism and use this to derive some general formulae which are useful for calculating the amplitude of the three and four-point functions of the inflationary perturbations. We then go into great detail to study a concrete example of a popular inflationary model whose perturbations are non-Gaussian, the curvaton scenario, see Sect. 5. In Sect. 6 we provide some nontechnical question and answers about the subject and its current status after the first major data release from the Planck satellite, which has made by far the most stringent constraints on non-Gaussianity available. Finally we conclude in Sect. 7.

2 Gaussian Distributions

A Gaussian (or normal) distribution is defined by the probability distribution function (pdf)

$$\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-x_0)^2}{2\sigma^2}}, \tag{1}$$

where x_0 is the mean of the distribution and σ^2 the variance. Gaussian distributions are relatively simple and have many neat properties, some of which we will explore in these lectures. It only has two free parameters, and we will see that in cosmology the mean can usually be redefined to be zero, leaving only the variance. In contrast, non-Gaussian distributions can have an arbitrary number of free parameters.

The central limit theorem states that in the limit of a large number of measurements, if all measurements are drawn from independent, identically distributed pdfs then the limiting distribution will the Gaussian. Because many processes in nature depend only on the average of many "small scale" processes, we often find nearly Gaussian perturbations in nature. Therefore a good first guess for an unknown distribution is often that it will be nearly Gaussian.

In quantum mechanics, the ground state of the simple harmonic oscillator is Gaussian distributed. Since inflation predicts that the primordial density perturbation is generated from the quantum fluctuations of a light scalar field, which is quantised analogously to the simple harmonic oscillator, we may expect the curvature perturbation to be Gaussian.

This means, for example, that if you divide the CMB sky into many small patches, there will be the same number of patches which are hotter than average as those which are colder than average, and that if you plotted a histogram of the temperature deviations in each patch, they would form a normal (bell) curve. Similarly if you were to divide the early universe into little cubes, the density distribution would fit a Gaussian curve. However, doing the same in the late time universe, when the density perturbations have become large this changes, since

$$-1 \leq \frac{\rho - \bar{\rho}}{\bar{\rho}} \leq \infty, \tag{2}$$

where $\bar{\rho}$ is the average density over the total volume, and ρ the density in the patch being measured, is not a symmetric distribution, unlike the Gaussian distribution. The asymmetry arises because there is a limit to how empty space can become ($\rho = 0$), whilst there is almost no limit to how overdense it can become.

There are two lessons which can be learnt from this. Firstly that small perturbations can more easily be Gaussian in practice, for example the CMB temperature is given by $T = 2.75 \pm 10^{-5}$ K, and so although the temperature cannot be lower than absolute zero, in practice no perturbation will ever come close to being so large. We will later see that the temperature distribution of the CMB comes extremely close to following a Gaussian distribution. Secondly gravity acts both to make the density perturbation larger (overdense regions become denser due to gravitational attraction, leaving the voids emptier) and less Gaussian. A Gaussian distribution remains Gaussian under a linear transformation (which corresponds to shifting the mean value or changing the variance), but becomes non-Gaussian under any non-linear transformation, for example squaring a Gaussian distribution leads to a chi-squared distribution. In practise, we will always identify the linear perturbation with a Gaussian perturbation, and all higher-order terms as being non-Gaussian corrections. The lowest order corrections, the quadratic corrections, follow a chi-squared distribution.

Since the gravitational equations of motion are non-linear (thats what makes them hard to solve!), the perturbations will not only grow, but they will also become non-Gaussian, even if the initial perturbations were exactly Gaussian. The non-Gaussianity generated if one starts with a Gaussian primordial density perturbation

is known as the secondary non-Gaussianity. Until the time that the CMB formed, this secondary non-Gaussianity was very small and the corrections can be neglected. For large scale structure, the secondary non-Gaussianity becomes large at later times, and at smaller scales which have a larger amplitude and are hence less linear. This means that detecting primordial non-Gaussianity is easier in the CMB than large scale structure; most (or all) of the non-Gaussian signal measured in the clustering of galaxies is secondary non-Gaussianity. These lectures will focus exclusively on primordial non-Gaussianity, which we aim to use as a probe into the physics of the early universe. Fundamental questions which we hope to answer include how many scalar fields were present during inflation, what form their Lagrangian had and how reheating proceeded after inflation.

2.1 Distinct Characteristics of Gaussian Distributions

As described above, a Gaussian distribution has just two free parameters, the mean and variance. The mean can typically be defined to be zero, since it is convenient to define a perturbed quantity as being its deviation from the average value. Physically the mean tells us about the homogeneous universe, but nothing about the primordial perturbations. For example, the mean density of the universe has been measured as being very close to the critical density, which corresponds to a spatially flat universe. The average value of the CMB temperature redshifts with time and hence provides us with no information about inflation. The only additional information we can learn for a Gaussian distribution is how the variance depends on scale, for example whether the variance becomes larger if we divide the CMB sky into larger patches. In practise such a measurement is usually discussed in Fourier space, where the Fourier wavenumbers satisfy $k = |\mathbf{k}| \sim (\text{physical scale})^{-1}$, and the two-point correlator of the curvature perturbation is related to the power spectrum by

$$\langle \zeta_{\mathbf{k}} \zeta_{\mathbf{k}'} \rangle = P_\zeta(k)(2\pi)^3 \delta^3(\mathbf{k} + \mathbf{k}') \,, \tag{3}$$

and the variance per logarithmic interval in k-space is given by

$$\mathcal{P}_\zeta(k) = \frac{4\pi k^3}{(2\pi)^3} P_\zeta(k) = A_s \left(\frac{k}{k_{\text{pivot}}} \right)^{n_s - 1} . \tag{4}$$

A_s denotes the amplitude of the scalar perturbations, it corresponds to the variance of the perturbations at the pivot scale. The spectral index, $n_s - 1$ parametrises a possible scale dependence, where $n_s = 1$ corresponds to scale independence (in which case the power spectrum does not depend on k). Using combined Planck and WMAP data, see Table 2 of Ade et al. (2014c), the values of the two primordial

spectra parameters are

$$\ln\left(10^{10}A_s\right) = 3.089^{+0.024}_{-0.027}, \tag{5}$$

$$n_s - 1 = 0.9603 \pm 0.0073. \tag{6}$$

The assumptions of homogeneity and isotropy imply that the power spectrum is only a function of k. The parametrisation (4) is a simple ansatz for the scaling which has got nothing to do with whether the perturbations follow a Gaussian distribution or not. This simple ansatz is a good match to observations, meaning that we only require two parameters to describe the primordial power spectrum. If the perturbations are Gaussian, then all of the information is contained in the two-point correlation function, or equivalently the power spectrum. All of the odd n-point correlators are zero, while all of the even n-point correlators can be reduced to disconnected products of the two-point function, which contain no new information (this is known as Wick's theorem, we will not prove it here, it is a standard proof in quantum field theory courses). Hence Gaussian statistics are very prescriptive but not very informative, everything can be learnt just by measuring the two-point function. Given that we have not detected primordial non-Gaussianity, and we can parametrise the power spectrum using just two numbers, we face the remarkable fact that observations of millions of pixels on the CMB sky, which lead to over one thousand well measured power spectrum amplitudes (the C_l's), can be described by primordial perturbations which are fully specified by only two parameters. Such a simple state of affairs is perfectly consistent with inflation, a period of quasi de Sitter expansion being driven by a single slowly-rolling scalar field can naturally lead to a spectrum of nearly Gaussian and nearly scale invariant perturbations.

In order to learn more about inflation, we must carefully check how consistent this simple picture is, both theoretically and as a match with observations. The rest of these lectures will hopefully motivate why the difficult search for non-Gaussianity is worthwhile and exciting, despite the lack of any clear observational detection of primordial non-Gaussianity.

3 Different Models of Non-Gaussianity

As described in the last section, Gaussianity is very prescriptive. Non-Gaussianity is anything else, so in principle we should search for all possible patterns in the data when hunting for non-Gaussianity. Apart from the fact that this is computationally unfeasible, it is also clear that there will be random patterns distributed in any large data set, many of which are statistical flukes of no cosmological significance. If the underlying distribution is Gaussian, then we should still expect one in a hundred tests to appear non-Gaussian at the 99 % confidence level. This is the well known issue about anomalies in the CMB and the difficulty of quantifying the unlikeliness of posterior distributions. For example see the review article about large angle

anomalies (Copi et al. 2010) and the WMAP paper (Bennett et al. 2011), which often do not reach the same conclusions.

The path we will take in these lectures is to study which types of non-Gaussianity simple models of inflation predict. A major advance from the last decade in this field is the realisation that different types of extensions of the simplest inflationary models produce specific and predictable types of non-Gaussianity (Komatsu et al. 2009). Models of single-field, slow-roll inflation with canonical kinetic terms and a Bunch-Davies initial vacuum state produce Gaussian perturbations. Breaking any one of these four conditions produce specific shapes of non-Gaussianity. These shapes are then searched for with data and the constraints can be interpreted in terms of model parameters, for example constraining how strongly the kinetic term of the inflaton field is allowed to deviate from a canonical form.

In analogy with the two-point function which defines the power spectrum, (3), we define the bispectrum via the three-point function of the curvature perturbation

$$\langle \zeta(\mathbf{k}_1)\zeta(\mathbf{k}_2)\zeta(\mathbf{k}_3)\rangle = (2\pi)^3\delta^3(\mathbf{k_1}+\mathbf{k_2}+\mathbf{k_3})B_\zeta(k_1,k_2,k_3). \tag{7}$$

The delta function comes from assuming statistical homogeneity, assuming isotropy in addition allows us to write the bispectrum in terms of just the three amplitudes of the wave vectors (i.e. the three side lengths of a triangle in Fourier space). Compared to the power spectrum which was a function of just one amplitude, we see that the bispectrum may contain a lot more information, i.e. information about its shape as well as about its amplitude.

3.1 Local Non-Gaussianity

We have previously described how the linear perturbations can be identified as the Gaussian perturbations. An obvious next step would be to consider the linear term as the first in a Taylor series expansion, in which the second order term is of order the linear term squared, and so on for higher orders. Given that the amplitude of the linear term is observed to be 10^{-5}, we may expect the convergence to be strong, and that we are unlikely to need to include terms up to a high order. This approach is not only mathematically quite simple, but also quite well motivated by many classes of physical models. We first study how the model is defined, and we provide a technique for calculations before providing examples of concrete models. In particular, we will make an in depth study of the curvaton scenario, and use this case both as a justification of the local model, and to motivate extensions to this simple model. We will then see how non-Gaussianity provides an observational probe to distinguish between different models of inflation.

The local model of non-Gaussianity, in its simplest form, is defined by

$$\zeta(\mathbf{x}) = \zeta_G(\mathbf{x}) + \frac{3}{5}f_{\rm NL}(\zeta_G^2(\mathbf{x}) - \langle\zeta_G^2(\mathbf{x})\rangle). \tag{8}$$

The name comes from the fact that it is defined locally in real space, ζ is a local function of position. The factor of $3/5$ comes from the original definition being in terms of the Bardeen potential (Komatsu and Spergel 2001), which is related on large scales to the primordial curvature perturbation in the matter dominated era by $\Phi = (3/5)\zeta$. The variance term has been subtracted from the quadratic part in order that the expectation value satisfies $\langle\zeta\rangle = 0$, any other choice would leave a non-zero expectation value, which would be degenerate with the background term, meaning that ζ would not be a purely perturbed quantity. When working in Fourier space, this constant term is only important for the $k = 0$ mode (zero wavelength corresponds to a homogeneous mode), and is therefore often neglected. Under a Fourier transform, the quadratic term becomes a convolution,

$$\zeta(\mathbf{k}) = \zeta_G(\mathbf{k}) + \frac{3}{5} f_{\rm NL} \frac{1}{(2\pi)^3} \int d^3\mathbf{q} \zeta_G(\mathbf{q}) \zeta_G(\mathbf{k} - \mathbf{q}). \tag{9}$$

In general, the local bispectrum is defined by

$$B_\zeta^{\rm local} = 2\frac{3}{5} f_{\rm NL}^{\rm local} \left(P_\zeta(k_1) P_\zeta(k_2) + 2 \text{ perms} \right), \tag{10}$$

which in the case of a scale-invariant power spectra reduces to

$$B_\zeta^{\rm local} = 2A_s^2 f_{\rm NL}^{\rm local} \left\{ \frac{1}{k_1^3 k_2^3} + 2 \text{ perms} \right\}. \tag{11}$$

Although (8) does imply (10) the reverse is not the case, more general models for ζ can give rise to the same bispectrum. We will study generalisations of the local model in Sect. 4, and we will see how (1) the scale dependence of $f_{\rm NL}$ in Sect. 4.2 and (2) the trispectrum in Sect. 4.3 may be used to break the degeneracy between these different expansions for ζ which generate equal amplitudes (and shapes) of the bispectrum.

3.2 Equilateral and Orthogonal Shapes

If the inflaton field has a non-canonical kinetic term, interaction terms may give rise to a large bispectrum. The bispectrum is maximised for three modes which have the strongest interaction, for this model this corresponds to the time when all three of the modes exit the horizon during inflation, and hence all have similar wavelengths. Two templates which are both maximised in the limit of an equilateral triangle have been commonly used as a test of such models are the equilateral and orthogonal, the

latter was designed to be orthogonal to the equilateral (Senatore et al. 2010), as the name suggests.

$$B_\zeta^{\rm equil} = 6A_s^2 f_{\rm NL}^{\rm equil} \left\{ -\left(\frac{1}{k_1^3 k_2^3} + 2 \text{ perms}\right) - \frac{2}{(k_1 k_2 k_3)^2} + \left(\frac{1}{k_1 k_2^2 k_3^3} + 5 \text{ perms}\right) \right\}, \tag{12}$$

$$B_\zeta^{\rm ortho} = 6A_s^2 f_{\rm NL}^{\rm ortho} \left\{ -3\left(\frac{1}{k_1^3 k_2^3} + 2 \text{ perms}\right) - \frac{8}{(k_1 k_2 k_3)^2} + 3\left(\frac{1}{k_1 k_2^2 k_3^3} + 5 \text{ perms}\right) \right\}. \tag{13}$$

For both cases we ignore the scale dependence of the power spectrum and the intrinsic scale dependence of the bispectrum itself (Chen 2005).

Popular models of inflation with non-canonical kinetic terms include k-inflation (Armendariz-Picon et al. 1999) and Dirac-Born-Infeld (DBI) inflation (Silverstein and Tong 2004; Alishahiha et al. 2004). DBI inflation is one of the most popular string theory inspired models of inflation, for a recent review of the models see Baumann and McAllister (2014), and it generically predicts such a large equilateral non-Gaussianity that some of the "simplest" realisations have already been ruled out, see e.g. Lidsey and Huston (2007).

Single field models can be parametrised as $\mathcal{L} = P(X, \phi)$, where the kinetic term is $X = g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$, and a model with canonical kinetic term satisfies $\mathcal{L} = -X/2 - V(\phi)$, and hence has a speed of sound equal to unity, where

$$c_s^2 = \frac{P_{,X}}{P_{,X} + 2XP_{,XX}}. \tag{14}$$

We note that despite the widespread use of the subscript s in the literature, (14) really defines the phase speed of the perturbations, which is the same as the adiabatic sound speed for classical fluids but not for scalar fields. See Christopherson and Malik (2009) for a clarification of this point. For models with a sound speed much less than one, one has $f_{\rm NL} \sim 1/c_s^2$ for both the equilateral and orthogonal models, and hence the Planck constraints provide a lower bound on this parameter $c_s^2 \gtrsim 0.1$ (Ade et al. 2014b).

3.3 *Feature Models*

It is possible to generate large non-Gaussianity in single-field models with canonical kinetic terms if slow-roll is violated. By definition, inflation requires that $\epsilon < 1$, but higher-order derivatives of this parameter can become large. They can only become large for short periods of time, otherwise typically ϵ will quickly grow to become order unity and end inflation before 60 efoldings of inflation have been achieved.

A temporary break down of slow-roll can be achieved by adding a step like feature into the potential, for example of the form (Chen et al. 2007a)

$$V(\phi) = V(\phi)_{\rm sr}\left[1 + c\tanh\left(\frac{\phi - \phi_s}{d}\right)\right], \tag{15}$$

where $V(\phi)_{\rm sr}$ is the potential which generates slow-roll inflation, c is the height of the step, d the width of the step and ϕ_s determines the position of the step. Only the modes which exit the horizon while the inflaton is traversing the step will have an enhanced non-Gaussianity, hence the bispectrum will have a localised shape in Fourier space around a characteristic scale, determined by the Hubble scale when $\phi = \phi_s$. If this scale does not fit inside the range of about seven efoldings which the CMB probes there is almost no hope of a detection. The bispectrum typically also has fast oscillations imprinted on it, which makes it observationally hard to detect or constrain (Chen et al. 2007a). However a violation of slow roll will generically also produce a feature in the power spectrum at the same scale, so correlating this feature with the bispectrum should aid a detection in such models.

3.4 Other Bispectral Shapes

Although the four shape templates already described above are among the most popular, plenty of others exist and have been searched for. Other well known examples include

1. **Flattened/folded** configuration, which can be generated by models where the initial state of the perturbations is not the usual Bunch-Davies vacuum state, but an excited state (Chen et al. 2007b; Holman and Tolley 2008; Meerburg et al. 2009). The name comes because this shape has the largest signal in the limit of a flattened isosceles triangle, satisfying $k_1 \simeq k_2 \simeq k_3/2$.
2. **Cosmic strings** or other topological defects are strongly non-Gaussian objects which generate a complicated non-Gaussian shape, which has still not been fully calculated. For a review see Ringeval (2010). The Planck constraints on topological defects coming from both the power spectrum and bispectrum are given in Ade et al. (2014a).
3. **Magnetic fields** are ubiquitous in the universe and have been observed to exist in galaxies, galaxy clusters and even in voids. Their origin remains a mystery. Magnetic fields are intrinsically non-Gaussian objects and their abundance is constrained by the observed Gaussianity of the CMB perturbations. Interestingly, the trispectrum is claimed to provide a tighter constraint on the magnetic field abundance than the bispectrum (Trivedi et al. 2012). For a review of this topic see Durrer and Neronov (2013).

3.5 How Similar are the Bispectral Shapes?

Given that a huge number of bispectral shapes can be generated, many of which are very similar to one of the standard shapes considered earlier in this section, it is useful to be able to calculate how correlated two different shapes are. Since these shapes are a function of three variables, this is hard to do analytically or "by eye". For example, DBI inflation predicts a bispectral shape which is highly correlated to the equilateral shape, but it is numerically much harder work with since it is not separable in to a product of the three side lengths. Knowing that they have nearly the same shape, one may use the observation constraint on equilateral non-Gaussianity to constrain the sound speed of DBI inflation.

A scale invariant non-linearity parameter corresponds to a bispectrum which scales as $B \propto P^2 \propto k^{-6}$, so it is helpful to define a shape function which factors out this momentum dependence

$$S(k_1, k_2, k_3) = \frac{1}{f_{\rm NL}} (k_1 k_2 k_3)^2 B_\zeta(k_1, k_2, k_3). \tag{16}$$

The shape correlator is defined as the inner product of two shapes, to which a volume weighting $1/\Sigma k_i$ is applied, designed in order to match the signal to noise which comes from experiments, for details see Fergusson and Shellard (2009). The inner product between the two shapes S and S' is given by

$$F(S, S') = \int_{V_k} S(k_1, k_2, k_3) S'(k_1, k_2, k_3) \frac{1}{k_1 + k_2 + k_3} dV_k, \tag{17}$$

where V_k is the allowed volume in Fourier space of the k modes, which must satisfy the delta function condition $\Sigma \mathbf{k}_i = 0$. It should also satisfy constraints on the largest and smallest modes available to the experiment, defined respectively by the volume and resolution of the survey. In practise, the results are often reasonably insensitive to these choices, and if (17) converges when integrated over all k_i, this value is sometimes used without applying any cut offs. The shape correlator is defined by

$$\mathcal{C}(S, S') = \frac{F(S, S')}{\sqrt{F(S, S) F(S', S')}}. \tag{18}$$

The equilateral shape is about 50 % correlated with the local shape (Fergusson and Shellard 2009) and uncorrelated with the orthogonal shape (Senatore et al. 2010).

4 Local Non-Gaussianity and Its Extensions

We will first provide a convenient method to calculate the curvature perturbation ζ, which is especially convenient for the local model of non-Gaussianity. We will also see how this can be used to study generalisations of the local model to include scale

dependence of $f_{\rm NL}$ in Sect. 4.2, and to higher order in perturbation theory to study the trispectrum in Sect. 4.3.

4.1 The δN Formalism

The flat, unperturbed FRW metric is given by

$$ds^2 = -dt^2 + a(t)^2 \delta_{ij} dx^i dx^j, \tag{19}$$

and neglecting vector and tensor perturbations, the perturbed space-space part is given by

$$g_{ij} = a(t)^2 e^{2\zeta(t,\mathbf{x})} \delta_{ij}. \tag{20}$$

Therefore, the curvature perturbation ζ is the difference between the local expansion rate and the global expansion rate

$$\zeta(t, \mathbf{x}) = \delta N = N(t, \mathbf{x}) - N(t), \tag{21}$$

where

$$N(t) = \ln\left(\frac{a(t)}{a_{\rm initial}}\right) = \int \frac{da}{a} = \int H(t) dt. \tag{22}$$

N should be integrated from a spatially flat hypersurface shortly after horizon crossing, to a final uniform energy density (or equivalently a uniform Hubble) hypersurface. For some references which developed the δN formalism to linear order see Starobinsky (1985), Sasaki and Stewart (1996), Sasaki and Tanaka (1998), Lyth et al. (2005).

During inflation, the scalar fields provide the only contribution to the energy density, and within the slow-roll approximation their time derivatives do not provide a second degree of freedom (i.e. the field value and its derivative are not independent). Therefore

$$\zeta = N(\phi_a + \delta\phi_a) - N(\phi_a), \tag{23}$$

where a labels the scalar fields, and we may expand this as a Taylor series to find the key result

$$\zeta = N_a \delta\phi_a + \frac{1}{2} N_{ab} \delta\phi_a \delta\phi_b + \cdots, \tag{24}$$

where the field perturbations should be evaluated at the initial time (shortly after horizon crossing), summation convention is used and

$$N_a = \frac{\partial N}{\partial \phi_{a*}}. \tag{25}$$

Notice that the derivatives of N depend only on background quantities, so provided that the statistical distribution of the field perturbations is known at horizon crossing, we can do perturbation theory using only background quantities. For some intuition about this remarkable fact, see for example (Wands et al. 2000).

Assuming canonical kinetic terms, Bunch Davies vacuum and slow roll, the initial conditions are very simple. The field perturbations are Gaussian, and

$$\langle \delta\phi_a(\mathbf{k})\delta\phi_b(\mathbf{k}')\rangle = \delta_{ab}P_*(k)(2\pi)^3\delta^3(\mathbf{k}+\mathbf{k}'), \tag{26}$$

where

$$\mathcal{P}_*(k) = \frac{4\pi k^3}{(2\pi)^3}P_*(k) = \left(\frac{H_*}{2\pi}\right)^2. \tag{27}$$

The cross correlation terms in (26) are slow-roll suppressed and hence neglected (Byrnes and Wands 2006). Using these results and the linear part of (24), we may calculate the power spectrum

$$P_\zeta(k) = N_a N_a P_*(k)\,. \tag{28}$$

To calculate $f_{\rm NL}$, we first need the three-point function of ζ. The first non-zero contribution comes from taking two first-order terms and one second-order term from (24), we arbitrarily take the second-order term to correspond to the $\mathbf{k}_3$ term, the other two choices correspond to the two permutations which are added

$$\begin{aligned}
&\langle \zeta(\mathbf{k}_1)\zeta(\mathbf{k}_2)\zeta(\mathbf{k}_3)\rangle \\
&= \frac{1}{2}N_a N_b N_{cd}\langle \delta\phi_a(\mathbf{k}_1)\delta\phi_b(\mathbf{k}_2)\int \frac{d^3\mathbf{q}}{(2\pi)^3}\delta\phi_c(\mathbf{q})\delta\phi_d(\mathbf{k}_3-\mathbf{q})\rangle + 2 \text{ perms} \\
&= \frac{1}{2}N_a N_b N_{cd}2\int \frac{d^3\mathbf{q}}{(2\pi)^3}\langle \delta\phi_a(\mathbf{k}_1)\delta\phi_c(\mathbf{q})\rangle\langle \delta\phi_b(\mathbf{k}_2)\delta\phi_d(\mathbf{k}_3-\mathbf{q})\rangle + 2 \text{ perms} \\
&= N_a N_b N_{cd}\int d^3\mathbf{q}\,\delta_{ac}P_*(k_1)\delta^3(\mathbf{k_1}+\mathbf{q})\langle \delta\phi_b(\mathbf{k}_2)\delta\phi_d(\mathbf{k}_3-\mathbf{q})\rangle + 2 \text{ perms} \\
&= N_a N_b N_{cd}\delta_{ac}\delta_{bd}P_*(k_1)P_*(k_2)\delta^3(\mathbf{k_1}+\mathbf{k_2}+\mathbf{k_3})) + 2 \text{ perms} \\
&= N_a N_b N_{ab}\frac{1}{(N_c N_c)^2}P_\zeta(\mathbf{k}_1)P_\zeta(\mathbf{k}_2) + 2 \text{ perms}.
\end{aligned} \tag{29}$$

In going from the first to the second line, we have applied Wick's theorem to split the four-point function into two two-point functions. Placing the $\delta\phi_c$ and $\delta\phi_d$ in to the same angle bracket results in zero unless $k_3 = 0$, which is not observationally relevant. This leaves us with two ways to get a non-zero result, and both of those cases lead to exactly the same result which explains the factor of 2. We have applied (26) to go to the third line and performed the integration to reach the fourth line. The final line follows by application of (28).

Then using the definition of the bispectrum (7), as well as that of $f_{\rm NL}^{\rm local}$, (10), we find

$$f_{\rm NL} = \frac{5}{6}\frac{N_a N_b N_{ab}}{(N_c N_c)^2} \,. \tag{30}$$

This result was first derived by Lyth and Rodriguez in 2005, and is very useful since it allows us to calculate the bispectrum amplitude using only background quantities (and we know it must have the local shape).

4.1.1 Single-Field Inflation

In the case of single-field inflation, the derivatives of N are given by

$$N' \simeq \frac{\bar{H}}{\dot{\bar{\varphi}}} \simeq \frac{1}{\sqrt{2}}\frac{1}{M_{\rm Pl}}\frac{1}{\sqrt{\epsilon}} \sim \mathcal{O}\left(\epsilon^{-\frac{1}{2}}\right) \,, \tag{31}$$

$$N'' \simeq -\frac{1}{2}\frac{1}{M_{\rm Pl}^2}\frac{1}{\epsilon}(\eta - 2\epsilon) \sim \mathcal{O}\left(1\right) \,, \tag{32}$$

where the slow-roll parameters are defined by

$$\epsilon = \frac{M_{\rm Pl}^2}{2}\left(\frac{V'}{V}\right)^2 \,, \qquad \eta = M_{\rm Pl}^2\frac{V''}{V} \,. \tag{33}$$

This suggests that

$$f_{NL} = \frac{5}{6}\frac{N''}{N'^2} = \frac{5}{6}(\eta - 2\epsilon) \tag{34}$$

but since $f_{\rm NL}$ is slow-roll suppressed for this model, we should have also included the equally small non-Gaussianity of the field perturbations at horizon exit.

The final result, known as the Maldacena consistency relation (Maldacena 2003), states that

$$f_{\rm NL} \equiv \frac{5}{12}\lim_{k_1 \to 0}\frac{B(k_1, k_2, k_3)}{P(k_1)P(k_2)} = \frac{5}{12}(1 - n_s). \tag{35}$$

See Creminelli and Zaldarriaga (2004) for a general proof, valid for any single field model (even with non-canonical kinetic terms, breaking slow roll and a non Bunch-Davies vacuum state). The exciting result is that a detection of the bispectrum in the squeezed limit (similar to local non-Gaussianity) would rule out all single-field models. A detection of non-Gaussianity in any non-squeezed configuration would not do this.

4.1.2 Single-Source Inflation

If we instead assume that a single-field generated the primordial curvature perturbation, which was not the inflaton field, then large local non-Gaussianity is possible (but not required or even generic). Many models in the literature fit into this case, for example

- the curvaton scenario (to be studied in depth in Sect. 5)
- modulated (p)reheating (the duration of reheating varies with position) (Dvali et al. 2004; Zaldarriaga 2004; Suyama and Yamaguchi 2008; Ichikawa et al. 2008)
- inhomogeneous end of inflation (the duration of inflation varies with position) (Bernardeau and Uzan 2002, 2003; Lyth 2005; Huang 2009)
- multiple-field slow-roll inflation can produce large non-Gaussianity for certain trajectories which turn in field space (Alabidi 2006; Byrnes et al. 2008; Peterson and Tegmark 2011; Wang 2010; Elliston et al. 2011)

What they all have in common is that the duration of periods with differing equations of state varies with position. This is required in order that N becomes perturbed, since it only depends on the amount of expansion, i.e. H. In modulated reheating, the equation of state is 0 while the inflaton oscillates in a quadratic potential, but jumps to $1/3$ after the inflaton has decayed into radiation. This means that varying the time of reheating will change the expansion history, and hence N and ζ. For more on how these models are related see e.g. Vernizzi (2004), Alabidi et al. (2010), Elliston et al. (2014).

4.2 Scale Dependence of f_{NL}

As defined in (8), f_{NL} is just a constant modulating the second order term in the expansion. Whilst this is a very good approximation in many models, it is generally not exact. Similarly to how the power spectrum usually has a spectral index of order the slow-roll parameters (reflecting evolution during inflation), the non-Gaussianity typically has a similar sized scale dependence. The analogy is not exact since there are simple models in which the scale invariance is exact (such as the quadratic curvaton model that we will study later), and there are other cases where the scale dependence is large, despite all fields obeying the slow-roll conditions.

In analogy to the spectral index, we define

$$n_{f_{\rm NL}} \equiv \frac{d\ln|f_{\rm NL}(k)|}{d\ln k}. \tag{36}$$

Strictly speaking, this is only defined for an equilateral triangle, for which all three $k's$ are equal. However at lowest order, this spectral index is the same for any shape of triangle provided that one scales all three sides by the same ratio as shown in Byrnes et al. (2010). Corrections do become important in the case that one strongly deforms the triangle while changing its scaling (Shandera et al. 2011; Dias et al. 2013).

The most general formula for $n_{f_{\rm NL}}$ which also allows for a non-trivial field space metric is given in Byrnes and Gong (2013). There it was shown that there are three effects which may give rise to scale dependence: (1) multiple field effects, from fields with different spectral indices contributing to the curvature perturbation (2) non-linearity of the field perturbation equation of motion, which is only absent in the case of a field with a quadratic potential and no backreaction from gravity and (3) a non-trivial field space metric. One example of a case with a scale independent $f_{\rm NL}$ is the quadratic curvaton scenario, where the curvaton field's energy density is subdominant by construction and the perturbations from the inflaton field are neglected. However this is clearly an idealised case, and in general we should expect some scale dependence.

How large can we expect the scale dependence to be? The answer is model dependent, but in some cases relatively simple results are known and may be used to gain intuition. In the case of a single-source inflation model (see the previous sub section), in which the corresponding field direction was an isocurvature mode during inflation (otherwise the non-Gaussianity is very small if the field is the inflaton)

$$f_{\rm NL} = \frac{5}{6}\frac{N_{\sigma\sigma}}{N_\sigma^2}, \tag{37}$$

$$n_{f_{\rm NL}} \simeq \frac{N_\sigma}{N_{\sigma\sigma}}\frac{V'''}{3H^2} \simeq \frac{5}{6}\frac{{\rm sgn}(N_\sigma)}{f_{\rm NL}}\sqrt{\frac{r_{\rm T}}{8}}\frac{V'''}{3H^2}\,. \tag{38}$$

In order for $f_{\rm NL}$ to be observable, the term $1/f_{\rm NL}$ must be less than unity. We also have the observational constraint from the tensor-to-scalar ratio that $\sqrt{r_{\rm T}/8} \lesssim 0.1$. However the last term in (38) may be arbitrarily large, it is not a slow-roll parameter. Clearly it is zero in the case of a quadratic potential. The best studied case with non quadratic potentials, which often requires a numerical treatment, is the curvaton scenario. In this case, the scale-dependence is typically of a similar size to the slow-roll parameters but in some cases it may be much larger, even $n_{f_{\rm NL}} \sim 1$ is possible, at which values the formalism used for the calculation becomes inaccurate. See Sect. 5.2 for details.

If multiple-fields generate the curvature perturbation, e.g. both the inflaton and the curvaton, then $f_{\rm NL}$ will generically be scale dependent even when the non-

Gaussian field has a quadratic potential. This is due to $f_{\rm NL}$ being defined as the ratio of the bispectrum to the square of the power spectrum (10). The bispectrum is only sourced by the non-Gaussian field, while the power spectrum is sourced by both fields. Unless the power spectra of both fields have the same scale-dependence, the relative importance of the two fields perturbations will depend on scale. On the scales where the non-Gaussian field is more important $f_{\rm NL}$ is larger, while in the opposite case where the Gaussian (inflaton) field perturbations dominate $f_{\rm NL}$ will be negligible. An explicit realisation of this case is presented in Sect. 5.1.

4.3 The Trispectrum

The (connected) four point function of the primordial curvature perturbation is defined by

$$\langle \zeta_{\mathbf{k}_1}\, \zeta_{\mathbf{k}_2}\, \zeta_{\mathbf{k}_3} \zeta_{\mathbf{k}_4} \rangle_c \equiv T_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)(2\pi)^3 \delta^3(\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3} + \mathbf{k_4}) \,, \tag{39}$$

which using the δN formalism, and assuming that the fields have a Gaussian distribution at Hubble exit see Sect. 4.1, is given by

$$\begin{aligned} T_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) =& \tau_{NL} \left[P_\zeta(k_{13}) P_\zeta(k_3) P_\zeta(k_4) + (11 \text{ perms}) \right] \\ &+ \frac{54}{25} g_{NL} \left[P_\zeta(k_2) P_\zeta(k_3) P_\zeta(k_4) + (3 \text{ perms}) \right] , \end{aligned} \tag{40}$$

where $k_{13} = |\mathbf{k}_1 + \mathbf{k}_3|$ and the trispectrum non-linearity parameters can be calculated using (Alabidi and Lyth 2006; Seery and Lidsey 2007; Byrnes et al. 2006)

$$\tau_{NL} = \frac{N_{ab} N_{ac} N_b N_c}{(N_d N_d)^3} \,, \tag{41}$$

$$g_{NL} = \frac{25}{54} \frac{N_{abc} N_a N_b N_c}{(N_d N_d)^3} \,. \tag{42}$$

Hence we see that the trispectrum depends on two non-linearity parameters (as opposed to one, $f_{\rm NL}$, for the bispectrum), and they may be observationally distinguishable since they are prefactors of terms with different shape dependences in (40). The observational constraints on $\tau_{\rm NL}$ are tighter than those on $g_{\rm NL}$ because the pre factor for $\tau_{\rm NL}$ is large when either $k_i \to 0$ or $k_{ij} \to 0$, while it is large only in the former case for $g_{\rm NL}$.

No constraint has yet been made with Planck data on $g_{\rm NL}$, from WMAP9 data (Sekiguchi and Sugiyama 2013) found $g_{\rm NL} = (-3.3 \pm 2.2) \times 10^5$ while Regan et al. (2015) found $g_{\rm NL} = (-4.1 \pm 2.3) \times 10^5$. From scale-dependent bias data, Giannantonio et al. found $-5.6 \times 10^5 < g_{\rm NL} < 5.1 \times 10^5$ (2-σ) assuming that $f_{\rm NL} = 0$, which weakens to $-5.9 \times 10^5 < g_{\rm NL} < 4.7 \times 10^5$ when marginalising over $f_{\rm NL}$

(Giannantonio et al. 2014). Using the same technique, Leistedt et al. recently found $-4.0 \times 10^5 < g_{\rm NL} < 4.9 \times 10^5$ (2-σ) when marginalising over $f_{\rm NL}$ (Leistedt et al. 2014). The constraint on (positive definite) $\tau_{\rm NL}$ comes from the Planck collaboration and is significantly tighter than the constraint on $g_{\rm NL}$, being $\tau_{\rm NL} < 2800$ at $2-\sigma$ (Ade et al. 2014b).

The derivation of (41) and (42) is similar to that of $f_{\rm NL}$ in Sect. 4.1, see Byrnes et al. (2006) for details. The different forms of the contraction in the numerator follow due to $\tau_{\rm NL}$ consisting of two first-order terms and two-second order terms, while $g_{\rm NL}$ is made out of three first-order terms and one third-order term [corresponding to the third derivative of N in (42)]. In both cases the total four-point function is sixth order in $\delta\phi$, or third order in the power spectrum. Thats why there is a pre factor proportional to the power spectrum cubed in from of the $\tau_{\rm NL}$ and $g_{\rm NL}$ terms in (40).

4.4 Suyama-Yamaguchi Inequality

Applying the Cauchy-Schwarz inequality to (30) and (41) one may derive the Suyama-Yamaguchi inequality (Suyama and Yamaguchi 2008)

$$\tau_{\rm NL} \geq \left(\frac{6 f_{\rm NL}}{5}\right)^2 . \tag{43}$$

In the single-source limit the inequality saturates to an equality, and this is the most widely considered test of whether the curvature perturbation was generated by a single field (which we stress is not the same as asking whether inflation was driven by a single field, in fact the single field which generates the perturbations must not be the inflaton field in order for their to be any possibility of the trispectrum parameters being large enough to observe). Unfortunately the tight constraints on $f_{\rm NL}$ and difficulty in constraining $\tau_{\rm NL}$ means that we are only likely to be able to detect $\tau_{\rm NL}$ if $\tau_{\rm NL} \gg f_{\rm NL}^2$, while testing the equality is beyond observational reach for the foreseeable future (Biagetti et al. 2013; Grassi et al. 2014).

5 The Curvaton Scenario as a Worked Example

The curvaton scenario is arguably the most popular model for studying non-Gaussianity (Enqvist and Sloth 2002; Lyth and Wands 2002; Moroi and Takahashi 2001) [see also Mollerach (1990), Linde and Mukhanov (1997) for earlier related work]. The scenario is quite minimal in that it requires only the smallest possible extra complication for a model to be able to generate large local non-Gaussianity, and analytical solutions are possible, making this an excellent pedagogical example. We will consider this model in several parts, starting from the simplest case and

then gradually dropping model assumptions and seeing how the picture becomes both richer and more complex. Often, dropping a model assumption also leads to the existence of a new observable, meaning that the different cases of the curvaton scenario are (at least in principle) distinguishable.

The curvaton is an additional scalar field present during inflation. In order to pick up scale invariant perturbations its mass must be light compared to the Hubble scale, in addition it is required to have a subdominant energy density compared to the inflaton by assumption. Until Sect. 5.2 we will assume it has a quadratic potential

$$V = \frac{1}{2} m_\sigma^2 \sigma^2. \tag{44}$$

Due to the assumption that the curvatons energy density is small, its equation of motion for both the background and perturbation are the same during inflation, in the special case of a quadratic potential

$$\ddot{\sigma} + 3H\dot{\sigma} + V_{,\sigma} = 0, \tag{45}$$

$$\ddot{\delta\sigma} + 3H\dot{\delta\sigma} + V_{,\sigma\sigma}\delta\sigma = 0. \tag{46}$$

Neglecting the kinetic energy density of the curvaton, its energy density perturbation is a constant, and is given by

$$\frac{\delta\rho_\sigma}{\rho_\sigma} \simeq \frac{V(\sigma + \delta\sigma) - V(\sigma)}{V(\sigma)} = 2\frac{\delta\sigma}{\sigma} + \left(\frac{\delta\sigma}{\sigma}\right)^2. \tag{47}$$

It is not possible to get the correct numerical factors using this simple treatment, but it is nonetheless a useful approximation to relate the first term to the linear curvature perturbation caused by the curvaton, $\zeta_\sigma^{(1)} = \frac{\delta\sigma}{\sigma}$ and the second order term $\zeta_\sigma^{(2)} = \left(\frac{\delta\sigma}{\sigma}\right)^2$, so that (neglecting numerical factors), $\zeta_\sigma = \zeta_\sigma^{(1)} + \zeta_\sigma^{(2)}$ is a constant. However the total curvature perturbation is different and not conserved, it is proportional to $\Omega_\sigma = \rho_\sigma/\rho_{tot}$, assuming that the curvaton is the only perturbed component of the universe

$$\zeta \sim \Omega_\sigma \zeta_\sigma. \tag{48}$$

We will later consider what changes when all of these assumptions are dropped. We may already learn two important lessons which remain true in more general contexts, $f_{\rm NL} \propto \zeta^{(2)}/\zeta^{(1)2} \propto 1/\Omega_\sigma$, and $g_{\rm NL} = 0$. The first lesson is quite general, and states that even if the perturbations of a field as nearly Gaussian, as is the case for the curvaton, then an inefficient transfer of that fields curvature perturbation to the total curvature perturbation will generate non-Gaussianity. This follows since the transfer multiplies the curvature perturbation equally at all orders, making the second order term larger when compared to the square of the first order term. The second lesson is that quadratic potentials tend to only generate quadratic non-

Gaussianity, notice that it is only for a quadratic potential that (47) can be truncated at second order.

We now study the crucial parameter Ω_χ, which we have seen will affect both the amplitude of perturbations and the level of non-Gaussianity. During inflation, the curvaton will roll slowly due to the assumption that it is light, $m_\chi \ll H$. However the total energy density is also slowly varying (due to the requirement that $\epsilon = -\dot{H}/H^2 \ll 1$), so $\Omega_\chi \simeq \rho_\chi/\rho_\phi \ll 1$ is approximately constant (and very small) during this period. After inflation ends, the inflaton decays into radiation, which redshifts as a^{-4}. The curvaton will initially remain almost stationary, so that $\Omega_\chi \propto a^4$, until its mass become comparable to the Hubble rate at which time it will start to oscillate. Oscillations about a quadratic minimum correspond to a pressureless equation of state, so $\rho_\chi \propto a^{-3}$ and $\Omega_\chi \propto a$. During this period the relative energy density of the curvaton can grow a lot, potentially dominating the energy density of the universe if it either decays late enough or is much lighter than the Hubble parameter at the end of inflation. Finally the curvaton decays into radiation and thereafter $\Omega_\chi = \Omega_\chi|_{\rm decay}$ becomes a constant, and the perturbations in its field (which were initially isocurvature perturbations) have been converted into the primordial density perturbation. For a diagram displaying this process, see Fig. 4 of Dimopoulos (2011).

Including numerical factors, see e.g. Sasaki et al. (2006) for a derivation, the curvature perturbation is given by

$$\zeta = r_{\rm dec}\zeta_\sigma^{(1)} + \left(\frac{3}{2r_{\rm dec}} - 2 - r_{\rm dec}\right)\left(\zeta_\sigma^{(1)}\right)^2, \tag{49}$$

where

$$r_{\rm dec} = \frac{3\rho_\sigma}{4\rho_{\rm radiation} + 3\rho_\sigma} = \frac{3\Omega_\sigma}{4 + 3\Omega_\sigma}\Big|_{\rm decay}. \tag{50}$$

From (49) the full result for $f_{\rm NL}$ is

$$f_{NL} = \frac{5}{4r_{\rm dec}} - \frac{5}{3} - \frac{5}{6}r_{\rm dec}, \tag{51}$$

which in the limit of a subdominant curvaton reduces to

$$f_{\rm NL} \propto \frac{1}{r_{\rm dec}} \propto \frac{1}{\Omega_\sigma}, \qquad \text{for } r_{\rm dec} = \frac{3}{4}\Omega_\sigma \ll 1, \tag{52}$$

in agreement with the arguments made above. In the opposite limit of a dominant curvaton at the decay time we find $f_{\rm NL} = -5/4$ which is much less Gaussian than the slow-roll value predicted by single-field models [the Maldacena consistency relation, see (35)], but observationally indistinguishable from a Gaussian distribution. Observationally probing values of $f_{\rm NL} \sim 1$ is an important target for future experiments. Values of $f_{\rm NL}$ even closer to zero are possible in the curvaton scenario,

but require a finally tuned value of $r_{\rm dec}$, while $r_{\rm dec} = 1$ is the asymptotic value in the limit that the curvaton decays late.

If one had to choose one value of $f_{\rm NL}$ as a prediction from the curvaton, it would be $f_{\rm NL} = -5/4$. If the curvaton instead decayed when it was subdominant, then a large and positive value of $f_{\rm NL}$ is predicted. The Planck constraint of $f_{\rm NL} \lesssim 10$ results in the bound $r_{\rm dec} \gtrsim 0.1$, notice that prior to an $f_{\rm NL}$ constraint being made, $r_{\rm dec} \sim 10^{-5}$ was possible which would generate the correct power spectrum amplitude if $\delta\sigma/\sigma$ was approximately unity. Even without a detection of non-Gaussianity, the constraints have taught us under which circumstances the curvaton scenario can still be a viable model.

5.1 Including the Inflation Field Perturbations

So far we have assumed that the inflation field perturbations are negligible, however the field perturbations of any light scalar field are expected to have the same amplitude at horizon crossing, with the expectation value of their amplitude of perturbations at horizon crossing being

$$\delta\phi = \delta\sigma = \frac{H}{2\pi}. \tag{53}$$

The question is whether the curvature perturbation from the curvaton or the inflaton will be larger.

In this subsection, we will assume that the curvaton is subdominant at decay, $r_{\rm dec} \ll 1$ and we will neglect numerical factors of order unity. The curvature perturbation is given by

$$\zeta \sim \zeta_\phi + r_{\rm dec}(\zeta_\sigma + \zeta_\sigma^2), \tag{54}$$

where

$$\zeta_\phi \sim \frac{1}{M_{\rm Pl}\sqrt{\epsilon_*}}\delta\phi_*, \quad \zeta_\sigma \sim \left.\frac{\delta\sigma}{\sigma}\right|_*, \tag{55}$$

and quantities to be evaluated around horizon crossing are marked with a "$*$". The power spectra due to the two fields are

$$P_\zeta^\phi \sim \frac{1}{\epsilon_*}\left(\frac{H_*}{2\pi}\right)^2, \qquad P_\zeta^\sigma \sim \Omega_\sigma^2 \frac{1}{\sigma_*^2}\left(\frac{H_*}{2\pi}\right)^2, \tag{56}$$

and the total power spectrum is the sum, based on the good approximation that the perturbations of the two fields are uncorrelated (Byrnes and Wands 2006)

$$P_\zeta \equiv P_\zeta^{\rm total} = P_\zeta^\phi + P_\zeta^\sigma. \tag{57}$$

Because the inflaton perturbations are Gaussian, the bispectrum is unchanged from the pure curvaton limit

$$B_\zeta = B_\zeta^\sigma \simeq \frac{1}{r_{\rm dec}} \left(P_\zeta^\sigma\right)^2, \tag{58}$$

but $f_{\rm NL}$ is reduced because the power spectrum is enhanced by the Gaussian inflaton field perturbations

$$f_{\rm NL} = \frac{5}{6}\frac{B_\zeta}{P_\zeta^2} = \frac{5}{6}\frac{B_\zeta^\sigma}{P_\zeta^2} \sim \frac{1}{r_{\rm dec}} \left(\frac{P_\zeta^\sigma}{P_\zeta}\right)^2 \propto k^{2(n_\sigma - n_s)}, \tag{59}$$

where the spectral index n_σ is defined by $\mathcal{P}_\zeta^\sigma \propto k^{n_\sigma - 1}$. In the curvaton limit $n_\sigma \to n_s$ so $f_{\rm NL}$ becomes scale-invariant (assuming that it has a quadratic potential). Alternatively in the limit that the curvaton power spectrum is scale invariant, we have

$$n_{f_{\rm NL}} \to -2(n_s - 1) \simeq 0.1. \tag{60}$$

The tensor-to-scalar ratio is also reduced compared to the single-field inflation limit, whose value is $r_T = 16\epsilon_*$, to

$$r_T = 16\epsilon_* \frac{P_\zeta^\phi}{P_\zeta}. \tag{61}$$

In the curvaton limit $P_\zeta^\phi \ll P_\zeta$, so r_T is expected to be negligible (remember that $\epsilon < 1$ is required in order to have inflation).

The trispectrum parameter $\tau_{\rm NL}$ is instead enhanced compared to $f_{\rm NL}^2$, we have

$$\tau_{\rm NL} = \frac{P_\zeta}{P_\zeta^\sigma}\left(\frac{6f_{\rm NL}}{5}\right)^2 > \left(\frac{6f_{\rm NL}}{5}\right)^2, \tag{62}$$

which obeys the Suyama-Yamaguchi inequality (43). Notice that $\tau_{\rm NL}$ is also reduced by the addition of Gaussian inflaton perturbations, but by a lesser amount than $f_{\rm NL}^2$. Finally $g_{\rm NL}$ will remain negligible under the additional contribution of Gaussian perturbations from the inflation field.

5.2 *The Self-Interacting Curvaton*

Due to our assumption of a quadratic potential, the density and curvature perturbation are only present up to second order. There are subleading corrections in $1/r_{\rm dec}$ which give rise to higher order terms, but they are too small to ever be observed. For

non-quadratic potentials there is a potentially large third order term in (47), which can be constrained or measured by observations of $g_{\rm NL}$. In general we may expect $g_{\rm NL} \sim \tau_{\rm NL} \sim f_{\rm NL}^2$ from the curvaton scenario, unless it has a quadratic potential.

There is no fundamental reason why a quadratic potential is more "natural" or likely than non-quadratic cases, but most of the literature has focused on quadratic potentials due to the great computational simplicity which follows and allows analytical results to be found. When we include self interactions of the curvaton, i.e. extensions beyond a quadratic potential, the equation of motion for the curvaton becomes non-linear [see (45) and (46)] and $\delta\sigma/\sigma$ is no longer conserved. This can lead to very interesting behaviour (Enqvist et al. 2010), such as strongly scale-dependent non-Gaussianity. For certain (finely tuned) initial conditions one has $n_{f_{\rm NL}} \gg \mathcal{O}(\epsilon, \eta)$, turning this into an additional potential observable (Byrnes et al. 2011). Another difference from the quadratic case is that there is no lower bound of $f_{\rm NL}$, and the value of $f_{\rm NL}$ in the limit $r_{\rm dec} \to 1$ depends on the curvaton's potential. Generically the prediction does remain that $f_{\rm NL}$ is of order unity, but not exactly $-5/4$.

5.3 Curvaton Scenario Summary

We close this section with a table demonstrating how the many different realisations of the curvaton scenario could be distinguished using multiple non-Gaussian observables. This has provided us with a concrete realisation of how powerful an observation of non-Gaussianity would be, as well as how even the current constraints without a detection provide interesting information about the curvaton (Table 1).

Table 1 Four different curvaton scenarios, with a quadratic or non-quadratic curvaton potential and either including or neglecting the inflaton perturbations

Scenario	Curvaton potential	$f_{\rm NL}$	$n_{f_{\rm NL}}$	$\tau_{\rm NL}/(6f_{\rm NL}/5)^2$	$g_{\rm NL}$
Pure curvaton scenario	Quadratic	$\geq -\frac{5}{4}$	0	1	Small
	Non-quadratic	Unrestricted	Potentially large	1	Potentially large
Mixed inflaton-curvaton	Quadratic	$\geq -\frac{5}{4}$	Slow-roll$\sim 10^{-2}$	≥ 1	Small
	Non-quadratic	Unrestricted	Potentially large	≥ 1	Potentially large

The three columns of non-Gaussian observables show how we can (in principle) distinguish between the different cases

6 Frequently Asked Questions

Note that these lectures were written after the first Planck cosmology data release in Spring 2013, but before the final data release which is expected in late 2014 or 2015.

- **Do the Planck non-Gaussianity constraints imply that there is negligible non-Gaussianity?**

  Not really. For the local model of non-Gaussianity, they do imply the sky is over 99.9 % Gaussian, which is a remarkable result. For other templates, the constraint could be much weaker. But the constraint $|f_{\rm NL}^{\rm local}| \lesssim 10$ are still two to three orders of magnitude larger than the single-field consistency relation for the squeezed limit of the bispectrum, $f_{\rm NL} \simeq n_s - 1$. Clearly a large window is left for models which strongly deviate from this consistency relation, but have a level of non-Gaussianity which is not yet detectable.
- **Do the Planck non-Gaussianity constraints imply that alternatives to single field inflation are strongly disfavoured?**

  No. Single field inflation remains consistent with the observations, which does suggest they should be preferred from a Bayesian/Occams razor perspective. This was also true before we had Planck results. However it is important to bear two points in mind: (1) A model which is parametrised with the fewest parameters might not be the simplest or most natural from a model building perspective, (we know little about physics at the inflationary energy scale) and (2) there are many multiple field models which predict non-Gaussianity with $|f_{\rm NL}| \ll 1$, and hence are far from being ruled out.
- **Is there a natural target for future non-Gaussianity experiments?**

  Yes, to a certain extent. Several models which convert an isocurvature perturbation present during inflation into the primordial adiabatic perturbation after inflation have a large parameter range in which $|f_{\rm NL}^{\rm local}| \sim 1$. For example, the simplest version of the curvaton scenario, in which the curvaton potential is quadratic and it is dominant at the decay time (which it will be the case if it decays sufficiently late) makes a definite prediction, $f_{\rm NL}^{\rm local} = -5/4$. Similarly, a particularly simple realisation of modulated reheating predicts $f_{\rm NL}^{\rm local} = 5/2$. Hence having an experiment which is capable of discriminating between $f_{\rm NL}^{\rm local} = 1$ and $f_{\rm NL}^{\rm local} = 0$ would have great value in disfavouring popular non-Gaussian models. See the next question for an idea of when we might reach this target.

  For the equilateral model of non-Gaussianity, $f_{\rm NL}^{\rm equil} \sim 1$ is also a natural target for testing models with a non-canonical kinetic term (Baumann et al. 2015).
- **What are the prospects for future non-Gaussianity measurements?**

  The final Planck data release, which will contain double the observation time compared to the first release as well as Planck polarisation data, is expected to only lead to a relatively modest improvement to the $f_{\rm NL}$ constraints, about 20 %, compared to a factor of two for several other cosmological parameters including the spectral index. The next significant improvement in the constraint

for $f_{\rm NL}^{\rm local}$ is expected around 2020 from the Euclid survey, which is forecasted to reach an error bar of around 2–3 (i.e. around a factor of two tighter than Planck) (Giannantonio et al. 2012). Combined with SKA data in perhaps a decade, the constraints are forecasted to be able to distinguish $f_{\rm NL}^{\rm local} = 0$ from $f_{\rm NL}^{\rm local} = \pm 1$ at about $3-\sigma$ confidence (Yamauchi et al. 2014). Beyond this, there is no clear timeline to future experiments which will have even tighter constraints, although several experiments have been proposed, for example Core, Prism and Pixie which would measure the CMB to greater accuracy and to smaller scales.

- **Which forms of non-Gaussianity can we best constrain with future experiments?**

  Currently, the only concrete expectation for a significant improvement in non-Gaussianity constraints comes from the Euclid satellite. The forecasts have mainly been made for the scale dependent halo bias, which is sensitive to the squeezed limit of the bispectrum and hence primarily to local non-Gaussianity. The prospects for the other shapes is weaker, but limited work has been done on studying the galaxy bispectrum and using this as an estimator which could potentially improve sensitivity to all shapes of the bispectrum. This work is very challenging since the secondary signal from non-linear collapse is much larger than the primordial signal (implying observations will have to deal with many potentially large systematic effects). Even starting with Gaussian initial conditions, structure formation is a challenging topic.

7 Conclusions and Future Outlook

Non-Gaussianity is a window on to the physics of the very early universe. The distribution of the primordial perturbations contains much more information if it is non-Gaussian, providing signatures on to the high energy physics of inflation.

We have provided an introduction to the field of primordial non-Gaussianity. In contrast to a Gaussian perturbation, which is simple to describe and has only a variance as a free parameter, a non-Gaussian perturbation could be anything else and have any number of free parameters. Fortunately well motivated models of the early universe tend to predict a reasonably small number of non-Gaussian templates, which may be efficiently parametrised in terms of the three-point function of the curvature perturbation (or temperature perturbation on the CMB). For a Gaussian perturbation all information is included in the two-point function (i.e. the power spectrum) and the three-point function is zero. Hence any detection of the three point function would prove the primordial perturbations were non-Gaussian and the cosmology community has made a great effort to studying the bispectrum. This effort includes calculating the amplitude and shape dependence for classes of inflationary models as well as constraining these templates against observations. Planck has made the tightest ever constraints of the bispectrum, which are significantly tighter than anything which existed before. Although there was no detection of non-

Gaussianity, the constraints already rule out or put some inflationary models under observational pressure, and constrain the primordial perturbations to be more than 99.9 % Gaussian.

The simplest models of single-field slow-roll inflation predict a much smaller deviation from Gaussianity, to a level which is probably too small to ever be tested. However for other models of inflation there is still hope to for a future detection of non-Gaussianity, for example we have seen how $f_{\rm NL} = -5/4$ is a natural prediction of the curvaton scenario, and this value is within an order of magnitude of the current constraints. Improving the constraints so much will not be possible with CMB data, but may be possible with the next generation of large scale structure surveys in about a decade. They probe three dimensional information which allows them to alleviate the cosmic variance limits which the last scattering surface of the CMB suffers from, since we only have one sky to observe. However large scale structure has undergone more evolution through gravitational collapse than the CMB, and the non-linear equations of GR have made the perturbations less linear. The hard task is then to separate the primordial non-Gaussianity from the already detected secondary non-Gaussianities.

A detection of primordial non-Gaussianity would either be a signature of non-linear physics acting during the generation of the primordial perturbations or a non-linear transformation of the primordial perturbation between the initial scalar field perturbation and the curvature perturbation. An example of the former case is inflation with a reduced sound speed of perturbations due to a non-canonical kinetic term, while an in depth study of the latter case was made for the curvaton scenario. A general formalism (δN) was also provided for calculating this non-linear transformation, which for many models allows a study of the perturbations even to non-linear order just by calculating background quantities. This remarkable simplification is possible for the (many) models in which the perturbations are Gaussian at Hubble crossing during inflation and leads to local non-Gaussianity, described by a combination of a Gaussian perturbation and chi-squared non-Gaussianity.

While nature has not provided many clues to the physics of inflation, the search is continuing in a large way. The primordial perturbations have been convincingly demonstrated to deviate from scale invariance at a level consistent with slow-roll inflation. The search for primordial tensor perturbation, non-Gaussianity, isocurvature perturbations and features in the power spectrum, etc, continues. Whether or not these extra signatures are detected, it is only by studying a large range of inflationary models that we learn what to search for with the ever improving data sets, and what the constraints tell us about the first fraction of a second after the big bang.

Acknowledgements Christian T. Byrnes is supported by a Royal Society University Research Fellowship. I wish to thank the organisers of the second JPBCosmo school for the invitation to hold this lecture course in the shadow of the beautiful Pedra Azul mountain in Espírito Santo, Brasil. Special thanks are due to Hermano Endlich Schneider Velten and his family for hospitality before and after the school.

References

Abramo, L.R., Pereira, T.S.: Testing gaussianity, homogeneity and isotropy with the cosmic microwave background. Adv. Astron. **2010**, 378203 (2010). arXiv:1002.3173. doi:10.1155/2010/378203

Ade, P.A.R., et al.: Planck 2013 results. XXV. Searches for cosmic strings and other topological defects. Astron. Astrophys. **571**, A25 (2014a). arXiv:1303.5085

Ade, P.A.R., et al.: Planck 2013 results. XXIV. Constraints on primordial non-Gaussianity. Astron. Astrophys. **571**, A24 (2014b). arXiv:1303.5084

Ade, P.A.R., et al.: Planck 2013 results. XVI. Cosmological parameters. Astron. Astrophys. **571**, A16 (2014c). arXiv:1303.5076. doi:10.1051/0004-6361/201321591

Alabidi, L.: Non-Gaussianity for a two component hybrid model of inflation. J. Cosmol. Astropart. Phys. **0610**, 015 (2006). arXiv:astro-ph/0604611. doi:10.1088/1475-7516/2006/10/015

Alabidi, L., Lyth, D.H.: Inflation models and observation. J. Cosmol. Astropart. Phys. **0605**, 016 (2006). arXiv:astro-ph/0510441. doi:10.1088/1475-7516/2006/05/016

Alabidi, L., Malik, K., Byrnes, C.T., Choi, K.-Y.: How the curvaton scenario, modulated reheating and an inhomogeneous end of inflation are related. J. Cosmol. Astropart. Phys. **1011**, 037 (2010). arXiv:1002.1700. doi:10.1088/1475-7516/2010/11/037

Alishahiha, M., Silverstein, E., Tong, D.: DBI in the sky. Phys. Rev. D **70**, 123505 (2004). arXiv:hep-th/0404084. doi:10.1103/PhysRevD.70.123505

Armendariz-Picon, C., Damour, T., Mukhanov, V.F.: k - inflation. Phys. Lett. **B458**, 209–218 (1999). arXiv:hep-th/9904075. doi:10.1016/S0370-2693(99)00603-6

Bartolo, N., Matarrese, S., Riotto, A.: Non-Gaussianity and the cosmic microwave background anisotropies. Adv. Astron. **2010**, 157079 (2010). arXiv:1001.3957. doi:10.1155/2010/157079

Baumann, D., McAllister, L.: Inflation and String Theory (2014). arXiv:1404.2601

Baumann, D., Green, D., Porto, R.A.: B-modes and the nature of inflation. J. Cosmol. Astropart. Phys. **1501**(01), 016 (2015). arXiv:1407.2621

Bennett, C.L., Hill, R.S., Hinshaw, G., Larson, D., Smith, K.M., et al.: Seven-year Wilkinson microwave anisotropy probe (WMAP) observations: are there cosmic microwave background anomalies? Astrophys. J. Suppl. **192**, 17 (2011). arXiv:1001.4758. doi:10.1088/0067-0049/192/2/17

Bernardeau, F., Uzan, J.-P.: NonGaussianity in multifield inflation. Phys. Rev. D **66**, 103506 (2002). arXiv:hep-ph/0207295. doi:10.1103/PhysRevD.66.103506

Bernardeau, F., Uzan, J.-P.: Inflationary models inducing non-Gaussian metric fluctuations. Phys. Rev. D **67**, 121301 (2003). arXiv:astro-ph/0209330. doi:10.1103/PhysRevD.67.121301

Biagetti, M., Desjacques, V., Riotto, A.: Testing multi-field inflation with galaxy bias. Mon. Not. R. Astron. Soc. **429**, 1774 (2013). arXiv:1208.1616

Byrnes, C.T., Choi, K.-Y.: Review of local non-Gaussianity from multi-field inflation. Adv. Astron. **2010**, 724525 (2010). arXiv:1002.3110. doi:10.1155/2010/724525

Byrnes, C.T., Gong, J.-O.: General formula for the running of fNL. Phys. Lett. **B718**, 718–721 (2013). arXiv:1210.1851. doi:10.1016/j.physletb.2012.11.052

Byrnes, C.T., Wands, D.: Curvature and isocurvature perturbations from two-field inflation in a slow-roll expansion. Phys. Rev. D **74**, 043529 (2006). arXiv:astro-ph/0605679. doi:10.1103/PhysRevD.74.043529

Byrnes, C.T., Sasaki, M., Wands, D.: The primordial trispectrum from inflation. Phys. Rev. D **74**, 123519 (2006). arXiv:astro-ph/0611075. doi:10.1103/PhysRevD.74.123519

Byrnes, C.T., Choi, K-Y., Hall, L.M.H.: Conditions for large non-Gaussianity in two-field slow-roll inflation. J. Cosmol. Astropart. Phys. **0810**, 008 (2008). arXiv:0807.1101. doi:10.1088/1475-7516/2008/10/008

Byrnes, C.T., Gerstenlauer, M., Nurmi, S., Tasinato, G., Wands, D.: Scale-dependent non-Gaussianity probes inflationary physics. J. Cosmol. Astropart. Phys. **1010**, 004 (2010). arXiv:1007.4277. doi:10.1088/1475-7516/2010/10/004

Byrnes, C.T., Enqvist, K., Nurmi, S., Takahashi, T.: Strongly scale-dependent polyspectra from curvaton self-interactions. J. Cosmol. Astropart. Phys. **1111**, 011 (2011). arXiv:1108.2708. doi:10.1088/1475-7516/2011/11/011

Chen, X.: Running non-Gaussianities in DBI inflation. Phys. Rev. D **72**, 123518 (2005). arXiv:astro-ph/0507053. doi:10.1103/PhysRevD.72.123518

Chen, X.: Primordial non-Gaussianities from inflation models. Adv. Astron. **2010**, 638979 (2010). arXiv:1002.1416. doi:10.1155/2010/638979

Chen, X., Easther, R., Lim, E.A.: Large non-Gaussianities in single field inflation. J. Cosmol. Astropart. Phys. **0706**, 023 (2007a). arXiv:astro-ph/0611645. doi:10.1088/1475-7516/2007/06/023

Chen, X., Huang, M.-X. Kachru, S., Shiu, G.: Observational signatures and non-Gaussianities of general single field inflation. J. Cosmol. Astropart. Phys. **0701**, 002 (2007b). arXiv:hep-th/0605045. doi:10.1088/1475-7516/2007/01/002

Christopherson, A.J., Malik, K.A.: The non-adiabatic pressure in general scalar field systems. Phys. Lett. **B675**, 159–163 (2009). arXiv:0809.3518. doi:10.1016/j.physletb.2009.04.003

Copi, C.J., Huterer, D., Schwarz, D.J., Starkman, G.D.: Large angle anomalies in the CMB. Adv. Astron. **2010**, 847541 (2010). arXiv:1004.5602. doi:10.1155/2010/847541

Creminelli, P., Zaldarriaga, M.: Single field consistency relation for the 3-point function. J. Cosmol. Astropart. Phys. **0410**, 006 (2004). arXiv:astro-ph/0407059. doi:10.1088/1475-7516/2004/10/006

Desjacques, V., Seljak, U.: Primordial non-Gaussianity in the large scale structure of the Universe. Adv. Astron. **2010**, 908640 (2010). arXiv:1006.4763. doi:10.1155/2010/908640

Dias, M., Ribeiro, R.H., Seery, D.: Scale-dependent bias from multiple-field inflation. Phys. Rev. D **87**, 107301 (2013). arXiv:1303.6000. doi:10.1103/PhysRevD.87.107301

Dimopoulos, K.: The quantum origin of cosmic structure: theory and observations. J. Phys. Conf. Ser. **283**, 012010 (2011). arXiv:1009.5466. doi:10.1088/1742-6596/283/1/012010

Durrer, R., Neronov, A.: Cosmological magnetic fields: their generation, evolution and observation. Astron. Astrophys. Rev. **21**, 62 (2013). arXiv:1303.7121. doi:10.1007/s00159-013-0062-7

Dvali, G., Gruzinov, A., Zaldarriaga, M.: A new mechanism for generating density perturbations from inflation. Phys. Rev. D **69**, 023505 (2004). arXiv:astro-ph/0303591. doi:10.1103/PhysRevD.69.023505

Elliston, J., Mulryne, D.J., Seery, D., Tavakol, R.: Evolution of fNL to the adiabatic limit. J. Cosmol. Astropart. Phys. **1111**, 005 (2011). arXiv:1106.2153. doi:10.1088/1475-7516/2011/11/005

Elliston, J., Orani, S., Mulryne, D.J.: General analytic predictions of two-field inflation and perturbative reheating. Phys. Rev. D **89**, 103532 (2014). arXiv:1402.4800. doi:10.1103/PhysRevD.89.103532

Enqvist, K., Sloth, M.S.: Adiabatic CMB perturbations in pre - big bang string cosmology. Nucl. Phys. **B626**, 395–409 (2002). arXiv:hep-ph/0109214. doi:10.1016/S0550-3213(02)00043-3

Enqvist, K., Nurmi, S., Taanila, O., Takahashi, T.: Non-Gaussian fingerprints of self-interacting curvaton. J. Cosmol. Astropart. Phys. **1004**, 009 (2010). arXiv:0912.4657. doi:10.1088/1475-7516/2010/04/009

Fergusson, J.R., Shellard, E.P.S.: The shape of primordial non-Gaussianity and the CMB bispectrum. Phys. Rev. D **80**, 043510 (2009). arXiv:0812.3413. doi:10.1103/PhysRevD.80.043510

Giannantonio, T., Porciani, C., Carron, J., Amara, A., Pillepich, A.: Constraining primordial non-Gaussianity with future galaxy surveys. Mon. Not. R. Astron. Soc. **422**, 2854–2877 (2012). arXiv:1109.0958. doi:10.1111/j.1365-2966.2012.20604.x

Giannantonio, T., Ross, A.J., Percival, W.J., Crittenden, R., Bacher, D., et al.: Improved primordial non-Gaussianity constraints from measurements of galaxy clustering and the integrated Sachs-Wolfe effect. Phys. Rev. D **89**, 023511 (2014). arXiv:1303.1349. doi:10.1103/PhysRevD.89.023511

Grassi, A., Heisenberg, L., Byrnes, C.T., Schaefer, B.M.: A test of the Suyama-Yamaguchi inequality from weak lensing. Mon. Not. R. Astron. Soc. **442**(2), 1068 (2014). arXiv:1307.4181

Holman, R., Tolley, A.J.: Enhanced non-Gaussianity from excited initial states. J. Cosmol. Astropart. Phys. **0805**, 001 (2008). arXiv:0710.1302. doi:10.1088/1475-7516/2008/05/001

Huang, Q.-G.: A geometric description of the non-Gaussianity generated at the end of multi-field inflation. J. Cosmol. Astropart. Phys. **0906**, 035 (2009). arXiv:0904.2649. doi:10.1088/1475-7516/2009/06/035

Ichikawa, K., Suyama, T., Takahashi, T., Yamaguchi, M.: Primordial curvature fluctuation and its non-Gaussianity in models with modulated reheating. Phys. Rev. D **78**, 063545 (2008). arXiv:0807.3988. doi:10.1103/PhysRevD.78.063545

Komatsu, E.: Hunting for primordial non-gaussianity in the cosmic microwave background. Classical Quantum Gravity **27**, 124010 (2010). arXiv:1003.6097. doi:10.1088/0264-9381/27/12/124010

Komatsu, E., Spergel, D.N.: Acoustic signatures in the primary microwave background bispectrum. Phys. Rev. D **63**, 063002 (2001). arXiv:astro-ph/0005036. doi:10.1103/PhysRevD.63.063002

Komatsu, E., Afshordi, N., Bartolo, N., Baumann, D., Bond, J.R., et al.: Non-Gaussianity as a probe of the physics of the primordial universe and the astrophysics of the low redshift universe (2009). arXiv:0902.4759

Leistedt, B., Peiris, H.V., Roth, N.: Constraints on primordial non-Gaussianity from 800,000 photometric quasars. Phys. Rev. Lett. **113**(22), 221301 (2014). arXiv:1405.4315

Lidsey, J.E., Huston, I.: Gravitational wave constraints on Dirac-Born-Infeld inflation. J. Cosmol. Astropart. Phys. **0707**, 002 (2007). arXiv:0705.0240. doi:10.1088/1475-7516/2007/07/002

Liguori, M., Sefusatti, E., Fergusson, J.R., Shellard, E.P.S.: Primordial non-Gaussianity and bispectrum measurements in the cosmic microwave background and large-scale structure. Adv. Astron. **2010**, 980523 (2010). arXiv:1001.4707. doi:10.1155/2010/980523

Linde, A.D., Mukhanov, V.F.: Nongaussian isocurvature perturbations from inflation. Phys. Rev. **D56**, 535–539 (1997). arXiv:astro-ph/9610219. doi:10.1103/PhysRevD.56.R535

Lyth, D.H.: Generating the curvature perturbation at the end of inflation. J. Cosmol. Astropart. Phys. **0511**, 006 (2005). arXiv:astro-ph/0510443. doi:10.1088/1475-7516/2005/11/006

Lyth, D.H., Rodriguez, Y.: The inflationary prediction for primordial non-Gaussianity. Phys. Rev. Lett. **95**, 121302 (2005). arXiv:astro-ph/0504045. doi:10.1103/PhysRevLett.95.121302

Lyth, D.H., Wands, D.: Generating the curvature perturbation without an inflaton. Phys. Lett. **B524**, 5–14 (2002). arXiv:hep-ph/0110002. doi:10.1016/S0370-2693(01)01366-1

Lyth, D.H., Malik, K.A., Sasaki, M.: A general proof of the conservation of the curvature perturbation. J. Cosmol. Astropart. Phys. **0505**, 004 (2005). arXiv:astro-ph/0411220. doi:10.1088/1475-7516/2005/05/004

Maldacena, J.M.: Non-Gaussian features of primordial fluctuations in single field inflationary models. J. High Energy Phys. **0305**, 013 (2003). arXiv:astro-ph/0210603. doi:10.1088/1126-6708/2003/05/013

Meerburg, P.D., van der Schaar, J.P., Corasaniti, P.S.: Signatures of initial state modifications on bispectrum statistics. J. Cosmol. Astropart. Phys. **0905**, 018 (2009). arXiv:0901.4044. doi:10.1088/1475-7516/2009/05/018

Mollerach, S.: Isocurvature Baryon perturbations and inflation. Phys. Rev. **D42**, 313–325 (1990). doi:10.1103/PhysRevD.42.313

Moroi, T., Takahashi, T.: Effects of cosmological moduli fields on cosmic microwave background. Phys. Lett. **B522**, 215–221 (2001). arXiv:hep-ph/0110096. doi:10.1016/S0370-2693(01)01295-3

Peterson, C.M., Tegmark, M.: Non-Gaussianity in two-field inflation. Phys. Rev. D **84**, 023520 (2011). arXiv:1011.6675. doi:10.1103/PhysRevD.84.023520

Regan, D., Gosenca, M., Seery, D.: Constraining the WMAP9 bispectrum and trispectrum with needlets. J. Cosmol. Astropart. Phys. **1501**(1), 013 (2015). arXiv:1310.8617

Ringeval, C.: Cosmic strings and their induced non-Gaussianities in the cosmic microwave background. Adv. Astron. **2010**, 380507 (2010). arXiv:1005.4842. doi:10.1155/2010/380507

Sasaki, M., Stewart, E.D.: A general analytic formula for the spectral index of the density perturbations produced during inflation. Prog. Theor. Phys. **95**, 71–78 (1996). arXiv:astro-ph/9507001. doi:10.1143/PTP.95.71

Sasaki, M., Tanaka, T.: Superhorizon scale dynamics of multiscalar inflation. Prog. Theor. Phys. **99**, 763–782 (1998). arXiv:gr-qc/9801017. doi:10.1143/PTP.99.763

Sasaki, M., Valiviita, J., Wands, D.: Non-Gaussianity of the primordial perturbation in the curvaton model. Phys. Rev. D **74**, 103003 (2006). arXiv:astro-ph/0607627. doi:10.1103/PhysRevD.74.103003

Seery, D., Lidsey, J.E.: Non-Gaussianity from the inflationary trispectrum. J. Cosmol. Astropart. Phys. **0701**, 008 (2007). arXiv:astro-ph/0611034. doi:10.1088/1475-7516/2007/01/008

Sekiguchi, T., Sugiyama, N.: Optimal constraint on g_{NL} from CMB. J. Cosmol. Astropart. Phys. **1309**, 002 (2013). arXiv:1303.4626. doi:10.1088/1475-7516/2013/09/002

Senatore, L., Smith, K.M., Zaldarriaga, M.: Non-Gaussianities in single field inflation and their optimal limits from the WMAP 5-year data. J. Cosmol. Astropart. Phys. **1001**, 028 (2010). arXiv:0905.3746. doi:10.1088/1475-7516/2010/01/028

Shandera, S., Dalal, N., Huterer, D.: A generalized local ansatz and its effect on halo bias. J. Cosmol. Astropart. Phys. **1103**, 017 (2011). arXiv:1010.3722. doi:10.1088/1475-7516/2011/03/017

Silverstein, E., Tong, D.: Scalar speed limits and cosmology: acceleration from D-cceleration. Phys. Rev. D **70**, 103505 (2004). arXiv:hep-th/0310221. doi:10.1103/PhysRevD.70.103505

Starobinsky, A.A.: Multicomponent de Sitter (inflationary) stages and the generation of perturbations. JETP Lett. **42**, 152–155 (1985)

Suyama, T., Yamaguchi, M.: Non-Gaussianity in the modulated reheating scenario. Phys. Rev. D **77**, 023505 (2008). arXiv:0709.2545. doi:10.1103/PhysRevD.77.023505

Trivedi, P., Seshadri, P., Subramanian, K.: Cosmic microwave background trispectrum and primordial magnetic field limits. Phys. Rev. Lett. **108**(23), 231301 (2012). arXiv:1111.0744. doi:10.1103/PhysRevLett.108.231301

Verde, L.: Non-Gaussianity from large-scale structure surveys. Adv. Astron. **2010**, 768675 (2010). arXiv:1001.5217. doi:10.1155/2010/768675

Vernizzi, F.: Cosmological perturbations from varying masses and couplings. Phys. Rev. D **69**, 083526 (2004). arXiv:astro-ph/0311167. doi:10.1103/PhysRevD.69.083526

Wands, D.: Local non-Gaussianity from inflation. Classical Quantum Gravity **27**, 124002 (2010). arXiv:1004.0818. doi:10.1088/0264-9381/27/12/124002

Wands, D., Malik, K.A., Lyth, D.H., Liddle, A.R.: A new approach to the evolution of cosmological perturbations on large scales. Phys. Rev. D **62**, 043527 (2000). arXiv:astro-ph/0003278. doi:10.1103/PhysRevD.62.043527

Wang, T.: Note on non-Gaussianities in two-field inflation. Phys. Rev. D **82**, 123515 (2010). arXiv:1008.3198. doi:10.1103/PhysRevD.82.123515

Yadav, A.P.S., Wandelt, B.D.: Primordial non-Gaussianity in the cosmic microwave background. Adv. Astron. **2010**, 565248 (2010). arXiv:1006.0275. doi:10.1155/2010/565248

Yamauchi, D., Takahashi, K., Oguri, M.: Constraining primordial non-Gaussianity via multi-tracer technique with Euclid and SKA. Phys. Rev. D **90**(8), 083520 (2014). arXiv:1407.5453

Zaldarriaga, M.: Non-Gaussianities in models with a varying inflaton decay rate. Phys. Rev. D **69**, 043508 (2004). arXiv:astro-ph/0306006. doi:10.1103/PhysRevD.69.043508

Problems of CMB Data Registration and Analysis

O.V. Verkhodanov

Abstract In this short course, we consider some radio astronomical fundamentals and problems of radio astronomical observations. We discuss the main observational cosmological tests which are investigated with radio astronomy. The most crucial tests are connected with the Cosmic Microwave Background (CMB). Several radio telescopes for CMB study and their basic results are discussed. Some stages of the CMB data analysis pipeline are considered and examples of observational CMB anomalies are discussed. At the end of the course (Appendix 3), the short application of the GLESP package is presented for simulation of the CMB map.

1 Introduction

In the last decade of cosmological research, both ground-based and ambitious space observations of cosmic microwave background (CMB) have been carried out. Among these experiments, WMAP[1] (Wilkinson Microwave Anisotropy Probe) of NASA (Bennett et al. 2003; Hinshaw et al. 2007, 2009; Jarosik et al. 2011; Bennett et al. 2013), which was completed in 2010, and Planck[2] of the ESA (Planck Collaboration 2014a), which was completed in 2013. Among the most discussed ground based experiments in the last 3 years, there are observations and results obtained with the Atacama Cosmology Telescope (ACT) (Das et al. 2014) (see Rolando Dünner's lecture), the South Pole Telescope (SPT) (Schaffer et al. 2011) and BICEP2 (BICEP2 Collaboration 2014). The meaning of these and other CMB experiments is determined by the fact that they provide important insight into our understanding of the model of the Universe and accurately measure the main parameters of the Universe. The achieved progress is, first of all, due to new technologies and instruments in centimeter, millimeter, and sub-millimeter

[1] http://lambda.gsfc.nasa.gov.

[2] http://www.rssd.esa.int/Planck/.

O.V. Verkhodanov (✉)
Special Astrophysical Observatory of RAS, Nizhnij Arkhyz, Russia
e-mail: vo@sao.ru

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_4

astronomy. Modern CMB data processing allows one comparing these observations with theoretical descriptions of the physical processes that were underway in the early Universe.

2 Observational Cosmological Tests of Radio Astronomy

Among the basic astronomy discoveries in the second half of the twentieth century, professional astronomers are sure to mention radio galaxies, quasars, molecular lines, pulsars, cosmic microwave background radiation. These objects and phenomena are interconnected so that they are visible and studied in a separate area of astronomy—radio astronomy. Apparently, it may be said without exaggeration that our modern understanding about cosmology is largely based on observational data obtained by means of radio astronomy observations, as noted once (Peebles 1993). Radio astronomy research in cosmology is devoted to radio galaxies and quasars, their evolutionary properties, as well as the CMB.

Back in the 70s of the last century, Longair (1978) identified three main areas in which radio observations provide an important contribution to cosmology: study of the background radiation at millimeter wavelengths, study of the properties of the intergalactic gas and spatial distribution and cosmological evolution of extragalactic radio sources. Observations of intergalactic gas and different populations of radio sources were necessary in order to implement these projects, which allowed the construction of models for extragalactic sources and to count sources *versus* flux density. In addition, it was proposed to construct and study the relations "luminosity—volume", "redshift—magnitude", "angular size—redshift" and "angular size—flux density' for different populations of radio sources.

Thus, the basic radio cosmology tests which could be considered to study Universe properties are:

- The CMB:
  - power spectrum;
    - properties of low multipoles ($\ell < 300$);
    - properties of high multipoles ($\ell > 300$);
  - the statistical properties of the signal (Gaussian);
  - physics from E-mode polarization (e.g. reionization history);
  - physics from B-mode polarization (e.g. the earliest Universe);
  - mm/submm study of clusters of galaxies.
- Radio galaxies:
  - dependence "size—redshift";
  - dependence "flux density—redshift";
  - dependence "$\log N - \log S$" ("the number of sources—the flux density");
  - gravitational lensing;

  - the formation of large-scale structure;
  - rotational speed of galaxies and dark matter search according HI;
  - the age of stellar systems.

- Pulsars:

  - binary pulsars—emission of gravitational waves;
  - dimensional grid of pulsars—registration of gravitational waves.

This list can be expanded further by adding secondary studies effects occurring in the ionosphere by heavy particles interaction in cosmic rays; research of "normal" radio sources polluting cosmic relic background, and even the radar radio astronomy allowing one to explore the features of the gravitational field in the solar system with asteroid observation.

Here are the main milestones of the radio astronomical studies which significantly contributed to cosmology:

1. identification of radio galaxies (1947–1951);
2. identification of quasars (1961);
3. counts radio $\log N - \log S$ (1964–1997);
4. **the discovery of the CMB** (1965, **1978**);
5. correction theory of the formation of large scale structure (LSS) (1970–1980);
6. **aperture synthesis** (1967–**1974**);
7. discovery of pulsars (1967–**1974**);
8. kinematics of galaxies in HI (21 cm): data on the distribution of dark matter;
9. **study of the binary pulsar** (1974, **1993**);
10. **discovery of CMB fluctuations** (1992, **2006**);
11. definition of the cosmological parameters (2003).

Some results are shown in bold when the achievements marked the Nobel Prize (as well as the award year). Upon receiving these results there were used very different tools to work in different wavelength ranges with different angular resolution and sensitivity. But all of these tools are unified by one definition—they are radio telescopes.

Before we begin to discuss the CMB analysis, we introduce some concepts.

3 Few Fundamentals of Radio Astronomy

What is a radio telescope? Like any other telescope, it is the astronomical instrument to receive radiation (in our case, the radio emission) of celestial objects which has two main elements: an antenna device and a radiometer. Antenna design of radio telescopes are very diverse, because of the very wide range of wavelengths used in astronomy. The radioastronomical range of the electromagnetic scale continues from millimeter to kilometer waves. In the context of the observation of the universe, radioastronomy is determined by the window which is transparent to radio waves in the atmosphere (Fig. 1).

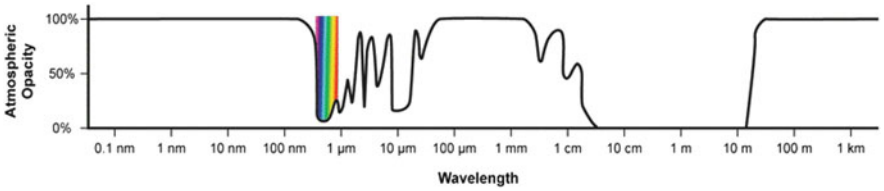

Fig. 1 The windows of atmospheric transparency in the electromagnetic spectrum. X-ray and gamma radiation only reaches heights of 30–40 km, the UV emission is absorbed by the ozone layer at a height of about 30 km. The first narrow window transparency transmits visible light with the adjacent thereto UV and near-IR where the atmosphere has a few narrow windows, caused by mechanisms of radiation absorption by water molecules, and carbon dioxide. Most of the infrared is absorbed by water vapor, carbon dioxide and oxygen, which are contained in the lower atmosphere. Radio astronomy window extends from 1 cm to 30 m. Longer radio waves are already reflected by the ionosphere

The main characteristic of the radio telescope is the antenna beam pattern (BP). The BP is the solid angle covering the area from which the antenna is receiving a signal. In a more correct definition, the BP describes dependence of the gain (directivity factor) of the antenna from the antenna direction in a predetermined plane. The BP of the radio telescope has a main and side lobes, the shape and amplitude of the response which depend on the configuration of the radio telescope. Thus, the antenna of a radio telescope actually receives the signal in the solid angle of 4π (Fig. 2). Typically, the radiation pattern is normalized, i.e. divided by the maximum value of the response (in the main lobe). The shape of the antenna pattern are divided into narrowly focused and broadly directional. Highly directional antennas have one pronounced maximum, which is called the main lobe, and the secondary maxima (side lobes), to reduce the impact of using a variety of hardware and algorithmic methods. The broadly directional antennas have a size of the main lobe of 2π, at least, in one plane. They are used in broadcasting. The BP has the property of reciprocity, i.e. is transmitting and receiving similar. The radio telescope BP can be calculated from the shape of the antenna. In radio astronomy, to determine the BP of the telescope, point radio sources are observed. The radiation patterns are characterized by the width of the main lobe at 0.5 level of the maximum value (abbreviated HPBW—Half Power Beam Width). This value is connected with another parameter of the radio telescope—*resolution*—minimum angular distance between two objects, at which one may distinguish them as separate sources (Wed Rayleigh criterion). This parameter is determined by the diffraction of radio waves at the aperture of the telescope as

$$\theta_{min} = \frac{\lambda}{D}, \tag{1}$$

where λ is the wavelength, D is the aperture diameter. To improve the resolving power, it is necessary either to reduce the wavelength, or to increase the aperture. When using short wavelengths, a higher quality of the mirror surface is demanded.

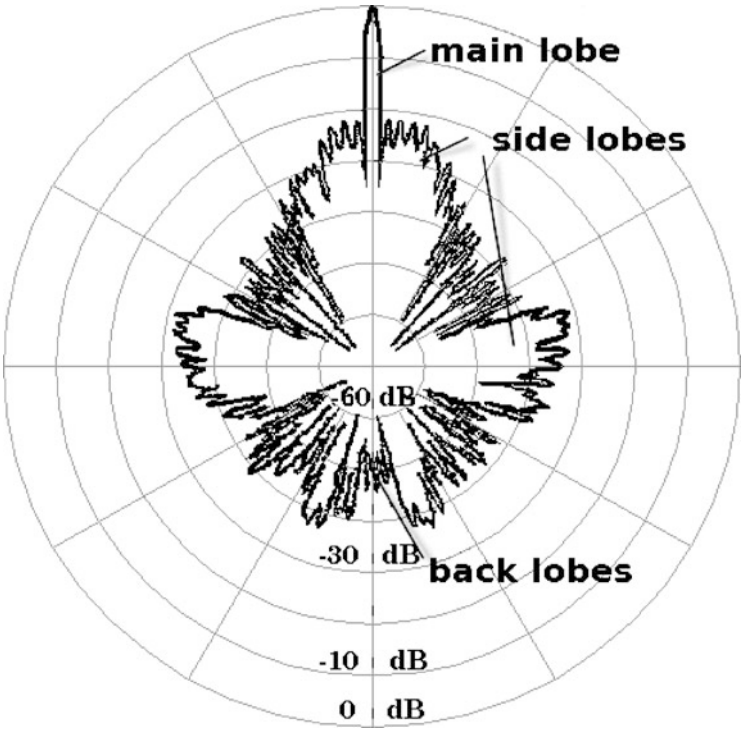

Fig. 2 Example of a directional antenna radiation pattern in polar coordinate system

3.1 Equation of Antenna Smoothing

Other main parameter of the radio telescope is the *sensitivity*, which is determined by the level of flux density fluctuations S:

$$\Delta S = \frac{P}{A_{\text{eff}}\sqrt{\Delta f t}}, \tag{2}$$

where P is the telescope noise power, A_{eff} is the effective area (the collecting surface) antenna, Δf is the bandwidth, t is the time of signal acquisition. To increase the sensitivity of radio telescopes, their collecting surface should be increased, and low-noise receivers should be used.

As well as an optical telescope, a radio telescope plays a role of the linear operator acting on the function describing the object investigated. Each point in the distribution of radio brightness over the investigated object is represented in the focus of the radio telescope in the form of diffraction spots or the circle of confusion. Using only the principle of superposition of these spots from different points of the

object, one can write the equation of the antenna smoothing:

$$I(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(x-x', y-y')G(x', y')dx'dy' \,, \tag{3}$$

where $B(x,y)$ is the radio brightness distribution on the object, $G(x', y')$ is the antenna BP, $I(x,y)$ is the observed power distribution in the telescope focal plane produced by the emitted source. To be absolutely correct, it is necessary to take into account the normalization to the integral under the instrumental function. This integral is equal to 1 from the condition of thermodynamic equilibrium between the antenna and the environment when brightness temperature is constant in the surrounding antenna space. In addition, the output of the system "antenna—radiometer" one also registers random variations in the power $N(x,y)$, caused by instability of the receiver, antenna, and radiation of the earth. Then, the actual observed brightness distribution can be represented in the form:

$$I(x,y) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(x-x', y-y')G(x', y')dx'dy'}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(x', y')dx'dy'} + N(x,y) \tag{4}$$

or in abbreviated form:

$$I = B * A + N \,. \tag{5}$$

Here I is the result of observations, B is the actual distribution of brightness of the subject, A is a hardware function, consisting of a system of 'appliance + eye', N is the multicomponent noise, $N = N_{\rm sky} * A + N_{\,rmsys}$, where $N_{\rm sky}$ is the noise associated with the passage of the signal through the galactic gas and the earth's atmosphere (smoothed by instrumental function) $N_{\,rmsys}$ the system noise "telescope + Receiver + eye". The symbol '*' (asterisk) denotes the convolution operation.

3.2 Sampling Theorem

Another important property, the one used in the analysis of any data, formulated in the sampling theorem (Whittaker 1915; Kotelnikov 1947) (or in another formulation Nyquist-Shannon sampling theorem): Any data (with minor limitations for us) is a continuous function $f(x)$, whose spectrum is limited to a frequency band of 0 to $F_{\rm c}$, completely determined by its sequence of values counted by intervals $\Delta x = \frac{1}{2F_{\,mboxc}}$.

The converse is also true: if the function exists only in the interval $(-x_0/2, x_0/2)$, i.e. limited in space, then for a complete determination of its spectrum is quite measuring the spectral components that are separated from each other at a frequency interval $\Delta F = 1/x_0$.

Here is an example of the use of this theorem in radio astronomy: Space-frequency response of the antennas is limited by frequency ($F_{\mathrm{rm\,c}} = D/\lambda$, D is the maximum size of the antenna) hence measured brightness distribution $I_a(x)$ should have a limited range. Information bearing samples as a function of $I_a(x)$ are located at intervals not less than $\Delta x = 1/2F_c = \lambda/2D$. This interval determines the resolution of the radio telescope. It was named the antenna natural interval.

4 Some Telescopes for CMB Study

Beginning of the 'precision cosmology' era is associated with the study the cosmic microwave background. The relic background carries unique information to restore all the major cosmological parameters with accuracy, almost unattainable in other experiments. For this reason, specialized radio telescope and bolometers massively received funding at the end of the last century and early this one. How many experiments to study the cosmic microwave background radiation have already been done? The answer is "more than 100 !" (see examples at http://lambda.gsfc.nasa.gov/product/expt). Here we consider some telescopes with specialized tools for studying the microwave background. In this short review, we will not describe the Atacama Cosmology Telescope of which the special review was prepared by Rolando Dünner at the School.

4.1 Horn Antenna of the Bell Laboratory

The history of the discovery of the CMB is dramatic. We note four points associated with the observations. (1) As early as in 1941, McKellar et al. (1941) noticed that at the observed population of rotational levels of the cyanogen molecule (CN) J = 1 in the interstellar medium, with transition from the lower level J = 0 (absorption line), requires the presence of radiation with temperature (for the case of black-body) 2.3 K. (2) In 1955, T. A. Shmaonov, under the leadership of Prof. S. E. Haikin in the Main Astronomical Observatory (Pulkovo, St.Petersburg, Soviet Union), measuring the temperature of the background radiation at a wavelength of 3.2 cm, received a temperature value equal to 4 K, which did not depend on the zenith distance z up $z = 30°$ and constant over time (Shamonov 1957; Naselsky et al. 2006). Although the accuracy of the measurement was low, the temperature was 2 K possible above the effective temperature of the troposphere. The explanation for this phenomenon at the time was found. (3) In 1964, Soviet astrophysicists A. G. Doroshkevich and I. D. Novikov, assuming the correctness of the Hot Universe theory, calculated the spectrum of an electromagnetic radiation taking into account the evolution of galaxies, the redshift and the expected temperature of microwave background. They showed the presence of the supervisory window in the spectrum, where temperature of the blackbody microwave background is above interfering

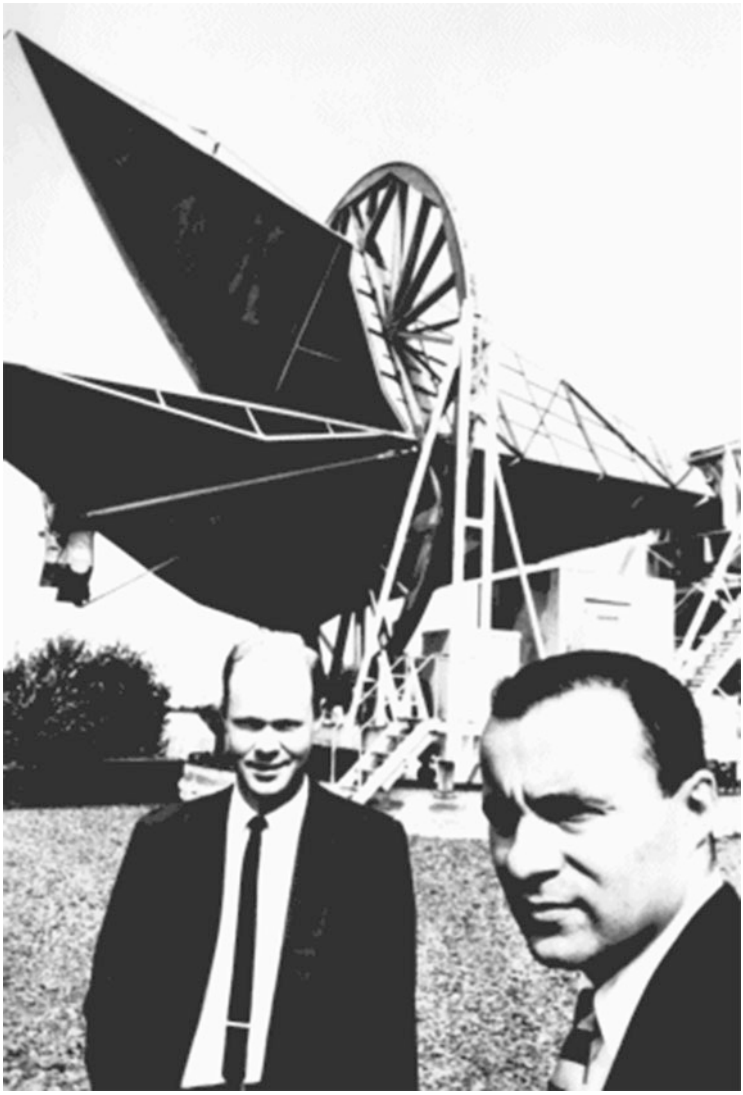

Fig. 3 Robert Wilson (*left*) and Arno Penzias (*right*) on the back of the horn antenna. Figure reproduced from the site http://bell-labs.com

components, and also indicated the type of the horn antenna, which can be observed. An article in the Proceedings of the USSR Academy (Doroshkevich and Novikov 1964; Naselsky et al. 2006) went unnoticed at that moment. (4) In the same year, American researchers 'Bell Labs' Pezias Arno (Arno Penzias) and Robert Wilson (Robert Wilson) studied radio response of balloons using an ultra-sensitive sixth horn antenna (Holmdell horn antenna; Fig. 3) chilled with the liquid helium

(4 K) receiver. A horn antenna (Fig. 3) had a length of 15 m and an aluminum 6-m aperture. The antenna was supplied with 10 m rotary impeller for setting the observation height. In the observation booth, it was installed a receiver created by the Dicke scheme, for observations at a frequency of 4.08 GHz. Struggling with the noise, the researchers found unexplained excess of the noise background equal to 3.5 K, as they have reported, particularly at the Massachusetts Institute of Technology at the seminars organised by Bernard Burke. Earlier in the same seminars, Robert Dicke and James Peebles made a report about possible temperature of the microwave background radiation. Because of these seminars, Bernard Burke put the four speakers in contact with each other and thus they were able to interpret the excess of the background noise as the residual radiation from the moment of recombination in the hot universe theory. Relevant works were published in the same issue of the Astrophysical Journal (Dicke et al. 1965; Penzias and Wilson 1965).

In 1978, A. Penzias and R. Wilson were awarded by the Nobel Prize with the wording "for their discovery of cosmic microwave background radiation".

4.2 COBE

Without dwelling on the theory of the formation of large-scale structure and arising irregularities as well as relevant observational effects, we note that in 1978, astrophysicists from the Zeldovich group (Doroshkevich et al. 1977) calculated possible observable effects 'printed' as fluctuations in the distribution of the CMB, starting physical investigation of the problem for structures formation. To search and study these fluctuations, several experiments were prepared. The most famous among them is COBE (Cosmic Background Explorer) which was launched as NASA space mission (with an orbit 900 km). COBE had a differential microwave radiometer with high accuracy ($\Delta T/T \sim 10^{-5}$). The experiment confirmed the existence of CMB anisotropy (Smoot et al. 1992), anisotropy map at full celestial sphere.

Three instruments were installed on COBE (Fig. 4). They were differential microwave radiometer (Differential Microwave Radiometer—DMR) to search for fluctuations in CMB, the head was George Smoot; spectrophotometer in the far infrared (Far-IR Absolute Spectrophotometer—FIRAS) for the study of Planck (blackbody) CMB spectrum range 0.1–10 mm, the head was John Mather; bolometer for the study of the diffuse infrared background (Diffuse IR Background Experiment—DIRBE) for the study of dust emission percentage between 1.25–240 μm, the head was Mike Hauser. To prepare a CMB anisotropy map, there were used radiometers on three frequencies 31.5, 53 and 90 GHz, which permitted a separation of the signal components. And it was the best way at the time to obtain the signal to the microwave background.

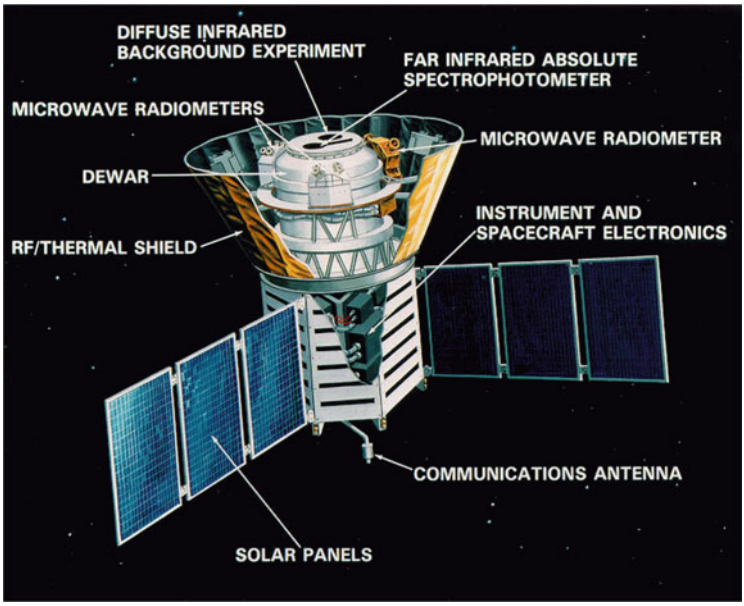

Fig. 4 The COBE satellite with the installed devices. Figure from the site http://lambda.gsfc.nasa.gov

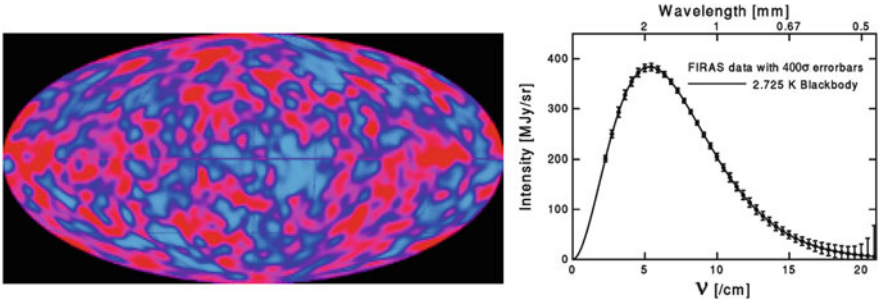

Fig. 5 *Left*: map of CMB fluctuations according to the COBE DMR. *Right*: CMB blackbody radiation spectrum according to COBE FIRAS. Figures from http://lambda.gsfc.nasa.gov

So, for serious cosmological results: definition of the blackbody radiation spectrum and CMB drawing a map of microwave background anisotropy (Fig. 5) John Mather and George Smoot were awarded the Nobel Prize in physics with the wording "for their discovery of the blackbody form and anisotropy of the cosmic microwave background radiation" in 2006.

Fig. 6 A team of researchers of the project against the backdrop of the interferometer DASI. Figure from http://astro.uchicago.edu/dasi/

4.3 DASI

The Degree Angular Scale Interferometer (DASI) is the 13-element interferometer (Fig. 6), created to measure temperature anisotropy and polarization of the CMB to a large range of angular scale with high sensitivity and the establishment of the University of Chicago station Amundsen–Scott South Pole. The instrument used cooled amplifiers HEMP, operating in the frequency

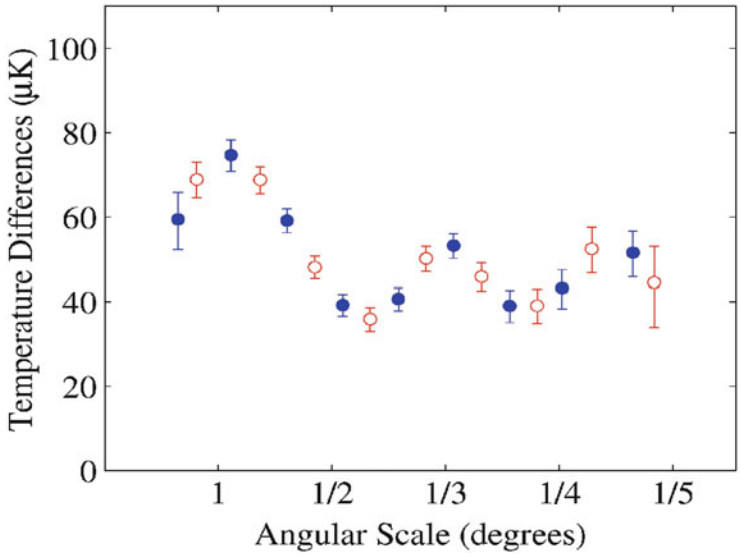

Fig. 7 The polarization cross-spectrum (T–E) according to the DASI. Figure from http://astro.uchicago.edu/dasi/

In addition to confirm and measure the CMB fluctuations at high harmonics (multipoles $\ell > 500$), DASI proved the existence of the E-mode polarization of the CMB (Kovac et al. 2002), and measured the polarization cross-spectrum (Fig. 7).

4.4 CBI

The Cosmic Background Imager (CBI) is the tool for imaging of the cosmic background range 26–36 GHz. The 13-element interferometer (like DASI) with parabolic antennas of diameter of 0.9 m and a 6-m platform (Fig. 8) was established by the international collaboration of universities (uniting California Institute of Technology, Canadian Institute for Theoretical Astrophysics, University of Chicago, NRAO, Max Planck Institute for Radio Astronomy, Oxford and Manchester universities, the University of Chile) at the altitude of 5080 m in the Chajnantor Observatory near San Pedro de Atacama in the Chilean Andes. Interferometer observations began in 1999 and continued until 2008 (with improved characteristics from 2006, including antennas of increased diameters to 1.4 m). Bandwidth of receivers was 26–36 GHz. There were three measured polarization modes: I, Q, U.

The task of CBI was to observe several areas of the sky (Fig. 9) in microwave range for the construction of the angular power spectrum on scales of 5 arcmin to 1° (multipole range $300 < \ell < 3000$). CBI reliably identified the second, third and fourth peaks of the power spectrum (Fig. 10) (Readhead et al. 2004).

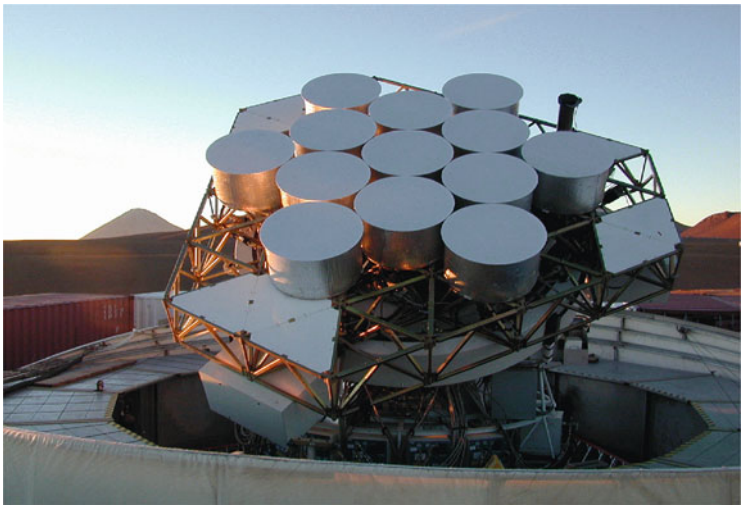

Fig. 8 Interferometer CBI. Figure from http://www.astro.caltech.edu/~tjp/CBI

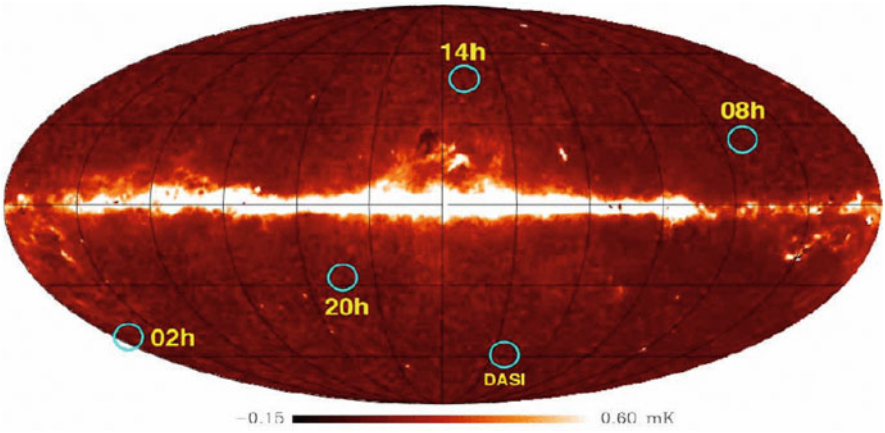

Fig. 9 Zones of observations with CBI. Figure from http://www.astro.caltech.edu/~tjp/CBI

4.5 BOOMERanG

Among the balloon experiments such as BOOMERanG, MAXIMA, Archeops et al., from our point of view, the first one—BOOMERanG (Balloon Observations Of Millimetric Extragalactic Radiation ANd Geophysics) is the most effective, when taking into account the background of interfering components. BOOMERanG observations were carried out at an altitude of 37 km over Antarctica and continued about 12 days from December 1998 to January 1999 and later in 2003. Balloon flight (Fig. 11) was maintained with a stable polar vortex wind around the South Pole, the

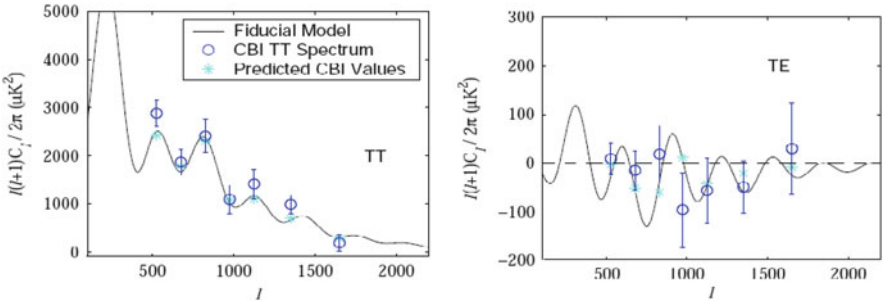

Fig. 10 Correlational power spectra of temperature anisotropy TT (*left*) and E-mode polarization with temperature TE (*right*) according to CBI. Figures from http://www.astro.caltech.edu/~tjp/CBI

Fig. 11 Balloon and telescope of the BOOMERanG experiment. Figure from http://www.astro.caltech.edu/~lgg/boomerang/boomerang_front.htm

effect of which gave the name of the experiment. In observations, there were used the 1.2-m mirror and 16 horns receivers on the focal plane for registration in the three frequency bands: 145, 245 and 345 GHz (de Bernardis et al. 2002). During the flight in 2003, the prototype of polarization Planck HFI receiver at the frequency 143 GHz was established in BOOMERanG.

Among the main results of the experiment, we mention the confirmation of the existence dark energy and the flat geometry of the universe, building maps (Fig. 12), as well as the measurement of the polarization CMB (E-mode) according to the polarization observations of the second run.

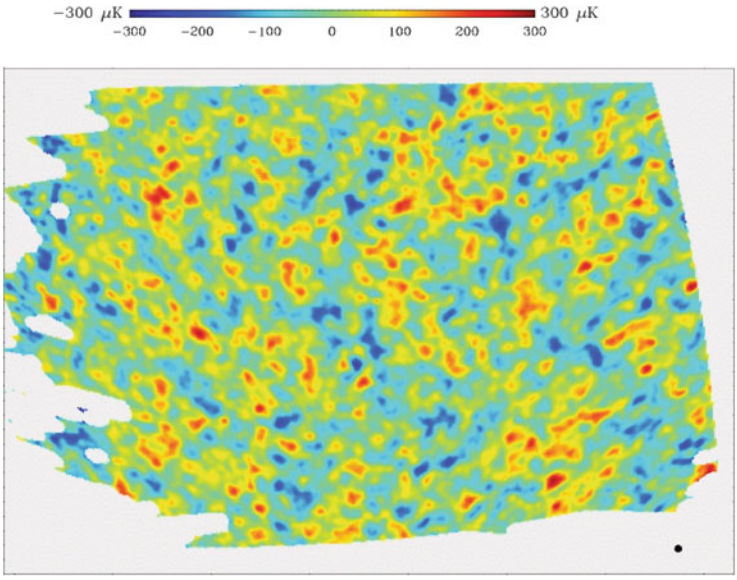

Fig. 12 CMB map from the BOOMERanG experimental data. Figure from http://www.astro.caltech.edu/~lgg/boomerang/boomerang_front.htm

5 WMAP

The launch of the NASA mission WMAP (Wilkinson Microwave Anisotropy Probe) was the crucial step in the CMB science. The WMAP data analysis and results, as well as providing an archive in common use allowed to translate the experimental (and theoretical) cosmological research in the area in a really exact science.

The mission WMAP (Fig. 13) is designed to detect geometry, composition and evolutionary properties of the universe by mapping the cosmic microwave background radiation with a resolution of 13 arcmin. The Satellite orbit (an orbit around the Lagrange point L2 in the Sun–Earth system), observing strategy and the design of the probe are selected so as to obtain non-correlated noise pixel, minimal systematic errors, and accurate calibration. Observations in the WMAP mission were conducted from 2001 to 2010 at five frequency bands: 23 (K), 33 (Ka), 41 (Q), 61 (V), and 94 GHz (W), where in the K and Ka bands used two radiometers, Q and V used four ones, and W used eight ones. Data are presented to the community for the 1st, 3rd, 5th, 7th and 9th years observations. According to the analysis of data, there was produced an anisotropy map with a resolution of $\ell \leq 100$ (Fig. 14) and built a CMB angular power spectrum $1 \leq \ell \leq 600$ for the temperature anisotropy and polarization. All the basic cosmological parameters were independently identified with high precision unattainable in other experiments (Komatsu et al. 2011a).

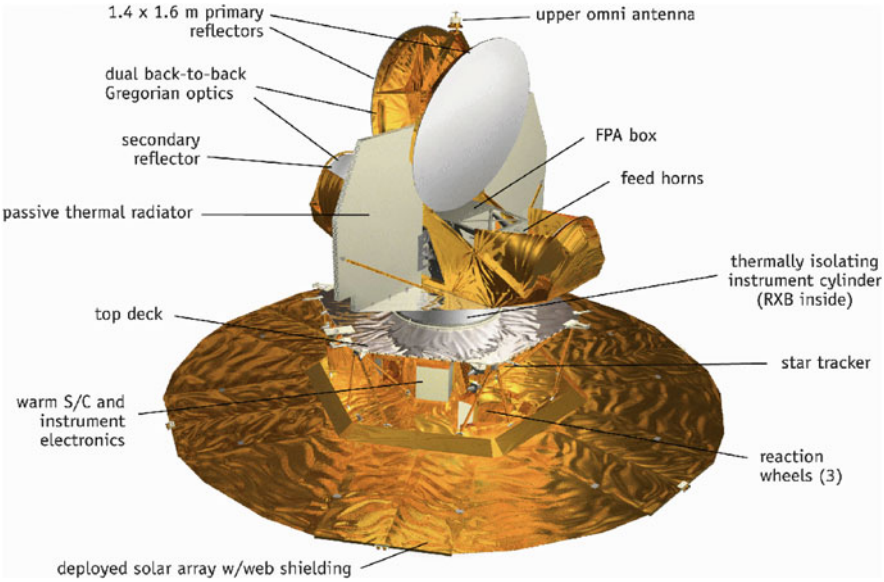

Fig. 13 The CMB probe WMAP. Figure from http://lambda.gsfc.nasa.gov

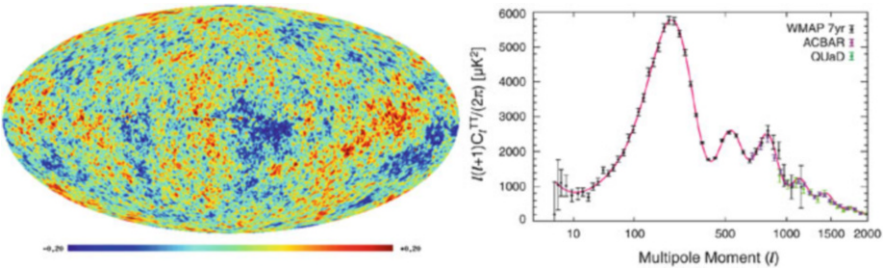

Fig. 14 *Left*: the CMB map by WMAP data of the 7th year. *Right*: the CMB angular power spectrum. From http://lambda.gsfc.nasa.gov

6 Planck Mission

The Planck mission of the European Space Agency with the assistance of NASA (by JPL) was launched on 14 May 2009 and intended for research temperature anisotropy and polarization CMB in a wide wavelength range (from mm to the far-IR) with very high accuracy and sensitivity. Initially, the mission was called COBRAS/SAMBA and after the improvement of the project in 1996, Planck was named in honor of the German physicist Max Planck. Measurements are taken from the Lagrange point L2 as well as in the WMAP mission. The duration of the mission was 4 years. The design consists of two off-axis parabolic mirrors: the primary one has a size of 1.9×1.5 m and the secondary one of 1.1×1.0 m. On the probe, two

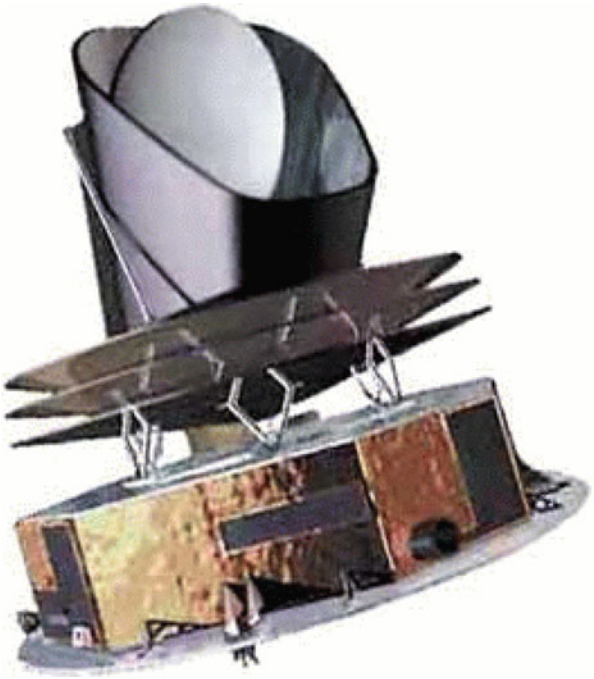

Fig. 15 The probe Planck. Figure from http://www.rssd.esa.int/Planck/

tools are placed (Fig. 15): the low frequency instrument (LFI) with bands at 30, 44, 70 GHz and high frequency instrument (HFI) with bands at 100, 143, 217, 353, 545, 857 GHz. LFI uses the high electron mobility transistors. HFI uses 48 bolometric detectors developed at JPL (Caltech) for photon detection. Planck has advantages over the WMAP by a number of parameters:

- higher resolution (three times) that will measure angular power spectrum to large ℓ,
- higher sensitivity (ten times)
- 9 frequency bands that will improve the model background components.

Cosmological results of the mission were presented in 2013. Among the main scientific objectives, there were the measurement of the total intensity (see Fig. 16) and polarization (including B-modes), building catalogues of clusters of galaxies using data about the Zeldovich–Sunyaev effect (see Fig. 17), observation of gravitational lensing of CMB, as well as observations of bright galactic nuclei, interstellar medium, the magnetic field of the Galaxy and objects in the solar system (Tauber et al. 2010). According to last results (Planck Collaboration 2014b), Planck improved the accuracy of determination of cosmological parameters, prepared maps of CMB, dust, galactic gas CO, y-comptonization, CMB gravitational lensing and other. In September 2014, it was also demonstrated that in B-mode polarization maps, cold magnetic dust corns are dominated practically everywhere on the sky.

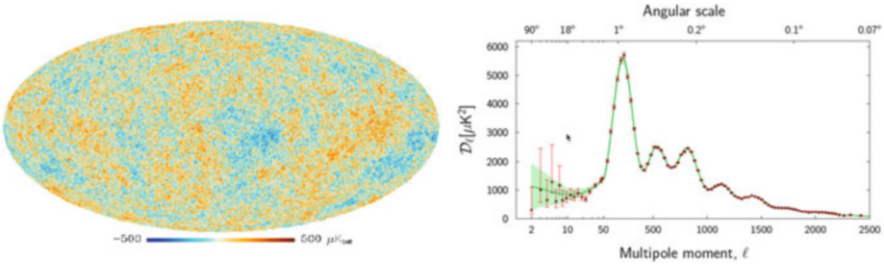

Fig. 16 The CMB map (*left*) and angular the power spectrum (*right*) of the CMB according to the Planck data. Figure from http://www.cosmos.esa.int/web/planck

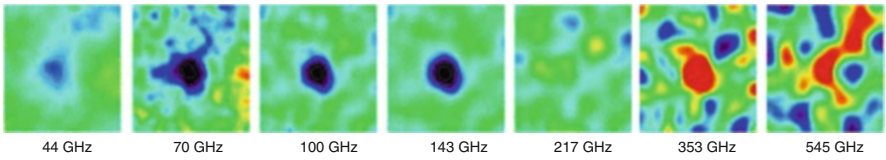

Fig. 17 Example manifestations of the Zeldovich–Sunyaev effect for the cluster of galaxies Abell 2319. From left to right, the same region of the sky at frequencies 44, 70, 100, 143, 217, 353, 545 GHz. At low frequencies (*left*), there is a "dimple" in the image center. At high frequencies (*right*), distribution of radiation has a visible source. Figure from http://www.cosmos.esa.int/web/planck

7 Sky Mapping

The pipeline of the CMB data processing and measurement of cosmological parameters includes several steps: (1) receiving time-ordered data (TOD); (2) map-making, CMB data pixelization and preparation of multifrequency sky maps; (3) cosmological signal rectification using observational data at different wavelengths; (4) statistical analysis of CMB maps; (5) harmonic transformations "map—spherical harmonics"; (6) calculation and analysis of the angular power spectra $C(\ell)$; (7) estimation of cosmological parameters.

Sky pixelization is closely related to map-making that allows one constructing a two-dimensional map of emission distribution on a sphere. For completeness, we outline the map-making algorithm for one-horned experiments (Natoli et al. 2001). We assume that the time-ordered data (TOD) output $\mathbf{d}$ and a map in pixels $\mathbf{m}$ are linearly dependent:

$$\mathbf{d} = \mathbf{P}\mathbf{m} + \mathbf{n}\,, \tag{6}$$

where $\mathbf{n}$ is the vector of random noise and P is the known matrix relating the signals $\mathbf{d}$ and $\mathbf{m}$. The rectangular $N_d \times N_p$ matrix $\mathbf{P}$ is called the pointing matrix. Here, N_d is the number of sky observations and N_p is the number of sky pixels of the size ~FWHM/3, where FWHM is the full width at half maximum of antenna beam. That

is, applying $\mathbf{P}^{\mathrm{T}}$ to the TOD collects them into a map. The value of a pixel of this map is the sum of all the observations of that pixel made at different times according to a given scanning strategy.

The structure of the matrix P depends on what we assume for $\mathbf{m}$. If $\mathbf{m}$ contains a pixelized but unsmeared image of the sky, then $\mathbf{P}$ must account for beam smearing. That is a very general assumption, which, for example, allows properly treating an asymmetric beam profile, such as quasi-Gaussian beams in the WMAP experiments, or possible effects of horn degradation during experiment (Naselsky et al. 2007). In this case, applying the matrix $\mathbf{P}$ to vector $\mathbf{m}$ implies both convolving the sky pattern with the detector beam response and unfolding $\mathbf{m}$ into 'a signal-only' time stream. If, on the other hand, the beam is at least approximately symmetric, then it is possible and certainly more convenient to regard $\mathbf{P}$ as a beam-smeared pixelized sky. The structure of $\mathbf{P}$ for a one-horned experiment would then be very simple. Only one element per row would differ from zero: the one connecting the observation of the jth pixel to the ith element of the TOD. Many methods have been proposed to estimate $\mathbf{m}$ in Eq. (6). Because the problem is linear in $\mathbf{m}$, the generalized least-squares (GLS) method can be used. This involves minimizing the quantity

$$\chi^2 = \mathbf{n}^T\mathbf{V}\mathbf{n} = (\mathbf{d}^T - \mathbf{m}^T\mathbf{P}^T)\mathbf{V}(\mathbf{d} - \mathbf{P}\mathbf{m}) \tag{7}$$

for some non-singular symmetric matrix V. In order to have a 'low noise' algorithm of the χ^2 estimation (a so-called estimator), we have to find a $\mathbf{V}$ that minimizes the variance of $\tilde{\mathbf{m}}$. This is attained if we take $\mathbf{V}$ to be the noise inverse covariance matrix, i.e., $\mathbf{V}^{-1} = N = \langle \mathbf{n}\mathbf{n}^T \rangle$. We can then construct the estimator $\tilde{\mathbf{m}}$ for the map:

$$\tilde{\mathbf{m}} = (\mathbf{P}^T\mathbf{V}\mathbf{P})^{-1}\mathbf{P}^T\mathbf{V}\mathbf{d}\,. \tag{8}$$

The GLS solution of the map-making problem is then

$$\bar{\mathbf{m}} = \Sigma^{-1}\mathbf{P}^T\mathbf{N}^{-1}\mathbf{d}\,, \tag{9}$$

where

$$\Sigma = \mathbf{P}^T\mathbf{N}^{-1}\mathbf{P}\,.$$

The map-making uses the chosen sky pixelization which also provides the integration accuracy when doing harmonic expansion of the signal. The problem of integration on a sphere was already discussed in 1970s. This problem became relevant in astrophysics when all-sky surveys appeared, and it became necessary to expand extended signals in spherical harmonics. Presently, sky map pixelization algorithms, namely, the decomposition of the sky in areas into which observational data are integrated using certain rules, are important ingredients of CMB data processing (Verkhodanov and Doroshkevich 2012).

8 Pixelization Grids

The problem of restoration of the initial CMB distribution is connected with the sky grid selection, since integration in pixel of TOD is required. First, we should note the problem of integration on a sphere, discussed in 1970s (Stroud 1971; Sobolev 1974; Mysovskikh 1976; Konjaev 1979). In the real physical analysis, the problem turns to real digital technologies realized in the celestial grid. To follow this, we should select a pixelization scheme. The simplest one uses equal divisions in latitude and longitude (θ, ϕ). This has been called the Equidistant Cylindrical Projection (ECP). It has the advantages of being both azimuthal and simply hierarchical, in that the data can be easily coarse grained by combining neighboring pixels. The azimuthal symmetry allows for fast spherical harmonic transforms, speeding many operations such as map simulation and inversion (Driscoll and Healy 1994; Muciaccia et al. 1997). The biggest failure of the ECP pixelization is that the pixels near the poles are small and very distorted. In a real experiment, they could be very noisy or even contain no data at all. The ECP scheme can be improved upon by grouping more and more pixels together as one approaches the pole.

Also we should mention the Hierarchical Triangular Mesh[3] (HTM) (Kunszt et al. 2001) developed for the Sloan Digital Sky Survey. The HTM sphere pixelization scheme uses triangular pixels which can recursively be subdivided into four pixels. The base pixels are 8, where 4 are for each hemisphere. They are obtained by the intersection on the sphere of three major big circles. On Earth they can be represented by the equator and two meridians passing at longitudes 0 and 90°. These base spherical triangles all have the same area. Each of them can then be further divided into four spherical triangles, or trixels, by connecting the three sides middle points using great circle segments.

The first all sky CMB maps, produced by the COBE satellite, used a pixelization based on the Quadrilateralized Sky Cube Projection, or "quad cube" (Greisen and Calabretta 1993). The edges of a cube are projected onto a sphere, dividing the sky into six equal areas. These are subdivided into a roughly square, hierarchical lattice. The main drawback of the resulting pixelizations is their lack of azimuthal symmetry, making spherical harmonic transforms time consuming.

At least four methods of the CMB celestial sphere pixelization have been proposed and implemented after the COBE pixelization scheme: the Icosahedron pixelizing by Tegmark (1996), the Igloo pixelization by Crittenden and Turok (1998), the HEALPix[4] method by Górski et al. (1999, 2005), and GLESP[5] (Doroshkevich et al. 2003, 2011).

[3] http://skyserver.org/htm.

[4] Currently http://healpix.jpl.nasa.gov.

[5] http://www.glesp.nbi.dk.

When thinking of possible sky pixelization schemes, Max Tegmark (1996) was the first to ask the question: which sky map pixelization should be considered *good* ? Tegmark focused on two criteria that can be used in choosing the location of N points (or pixel centers) on a sphere: (1) it is necessary to minimize the largest distance d to the nearest pixel (for example, from corner points to the center of a pixel in a square grid); (2) the grid must provide precise integration at the nodes.

For a quadrilateral projection, as in the COBE data analysis, when pixels are located on a rectangular grid on faces of a cube (in the tangent plane), the number of "stretches" increases in going to the face edges. Both the tangential and radial projections somewhat deform the pixel form, such that the further away from the center of an initial face, the more pixels on a sphere differ from those for a regular grid. In this case, it is highly desirable to make faces as small (and hence flat) as possible. A platonic solid with the smallest faces and hence maximum number of such faces is an icosahedron, consisting of 20 triangles (Fig. 18).

Such a partition is advantageous in having the number of faces three times as large as in the cube case, which turn increases the calculation accuracy due to almost the same pixels areas. In addition, the triangles cover the sky in more regular way (with area distortion) (Fig. 18, right) (Tegmark 1996).

The icosahedron pixelization scheme is similar to that used in the COBE experiment: (1) an icosahedron is inscribed into a sphere, and its faces are divided into a regular triangular grid; (2) the points (centers of pixels) are radially projected onto the sphere; (3) the points are somewhat shifted to make all pixels approximately equal in area.

The icosahedron pixelization is advantageous compared with the quadrilaterized scheme in having a larger number of rotational symmetries. The corresponding rotational matrices can be included in the software in advance and can later be called on if necessary in the data analysis.

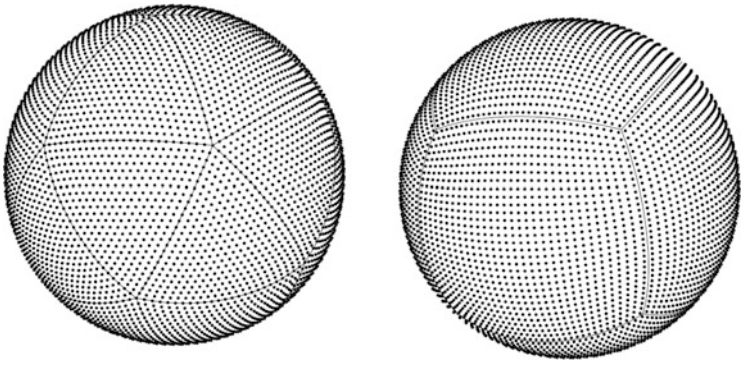

Fig. 18 *Left*: the icosahedron-based pixelization. *Right*: the cube-based pixelization. From Tegmark (1996)

8.1 Igloo Tilings

Crittenden and Turok (1998) suggested another pixelization scheme, called igloo tilings (because the similarity with the Eskimo house). In this scheme, a sphere is divided into rows with edges of constant latitude and each row is divided into identical pixels by lines of constant longitude (Fig. 19). The pixels are roughly trapezoidal, becoming nearly rectangular away form the poles. The northern and southern hemispheres are tiled identically. Igloo tiling have many advantages: (1) they are quite simple; (2) they are naturally azimuthal and can be easily made of equal areas, with most pixels nearly square; (3) most of all, the pixel edges are defined along constant lines of the spherical polar coordinates θ and ϕ, allowing fast and exact spherical harmonic transformations using pixel values. This is essential in constructing exact simulated skies and in optimally recovering the angular power spectrum from the real data in order to evaluate the igloo scheme.

When using the scheme with pixels that are of nearly (or exactly) equal area, the number of pixels in each row must decrease in approaching the poles (see Fig. 19). Igloo pixelizations with either rows equally spaces in latitude, or pixels of uniform area can be constructed. The advantage of equal-latitude spacing is that the pixelization can be created by a simple rebinning. In addition, by letting the pixel areas vary, one can make them less distorted. The equal area pixelization is not exactly equally spaced in latitude, but has the advantage that all the pixels have the same statistical weight.

In standard igloo tilings, the authors of Crittenden and Turok (1998) have chosen to initially divide the cap of each pole into three equal wedges. Higher-resolution pixelizations are found by dividing each wedge into four pieces, one central wedge and three pieces surrounding it. This process is iterated.

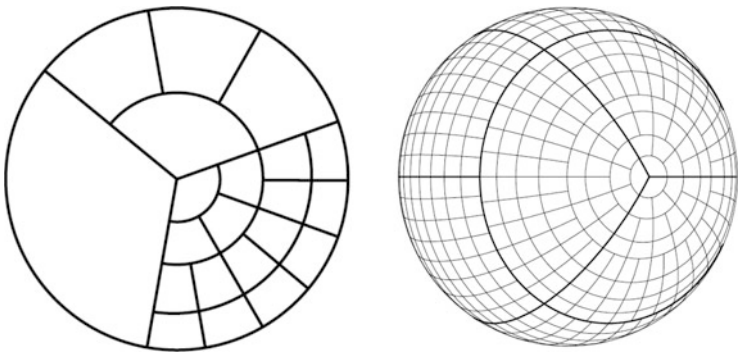

Fig. 19 *Left*: picture of the polar cap region in the Igloo scheme, showing three levels of subdivision to higher resolution pixels. *Right*: picture of the 3:6:3 equal area pixelization, which divides the sky into 12 base patches, three at either cap and six $60^\circ \times 60^\circ$ patches at the equator. Here, each base pixel is broken up into 64 smaller pixels. From Crittenden and Turok (1998)

8.2 HEALPix Hierarchical Pixelization

HEALPix[6] (Hierarchical Equal Area isoLatitude Pixelization) was in fact the first pixelization scheme supplied with a well-documented computer code (Górski et al. 1999).

The authors of Górski et al. (2005) formulated a list of desiderata for the mathematical structure of discrete whole sky maps arranged into three points.

(1) *Hierarchical structure of the data base.* This is recognized as essential for very large databases, and was already postulated in constructing the Quadrilateralized Spherical Cube. A simple argument in favour of these states that the data elements that are neighbours in a multidimensional configuration space (here, on the surface of a sphere) are also neighbours in the tree structure of the database. This property facilitates various topological methods of analysis, and allows easy construction of wavelet transforms on triangular and quadrilateral grids with a fast choice of neighboring pixels.
(2) *Equal areas of discrete elements of partition.* This is advantageous because white noise at the sampling frequency of the instrument is integrated exactly into white noise in the pixel space, and sky signals are sampled without regional dependence, although care must still be taken to choose a pixel size sufficiently small compared to the instrumental resolution to avoid excessive and pixel-shape-dependent signal smoothing.
(3) *Iso-latitude distribution of discrete area elements on a sphere.* This property is essential for computation speed in all operations involving evaluations of spherical harmonics. Because the associated Legendre polynomials are evaluated via slow recursion, any sampling grid deviations from an isolatitude distribution result in a prohibitive loss of computational performance with the growing number of sampling points.

All requirements formulated above are satisfied by a construction with the HEALPix of the sphere, which is shown in Fig. 20.

The HEALPix base-resolution comprised 12 pixel in three rings around the poles and the equator. The next hierarchical level is formed from the previous by dividing each pixel of the previous-resolution level into four equal parts. The resolution of the grid is expressed by the parameter N_{side} (see Fig. 20), which defines the number of divisions along the side of a base-resolution pixel that is needed to reach the desired high-resolution partition, for example, due to a horn beam. All iso-latitude rings located between the upper and lower corners of the equatorial base-resolution pixels ($-2/3 < \cos\theta_* < 2/3$) or in the equatorial zone are divided into the same number of pixels: $N_{\mathrm{eq}} = 4N_{\mathrm{side}}$. The remaining rings are located within the polar cap regions ($|\cos\theta_*| > 2/3$) and contain a varying number of pixels, increasing from ring to ring with increasing distance from poles by one pixel within each quadrant.

[6] http://healpix.jpl.nasa.gov.

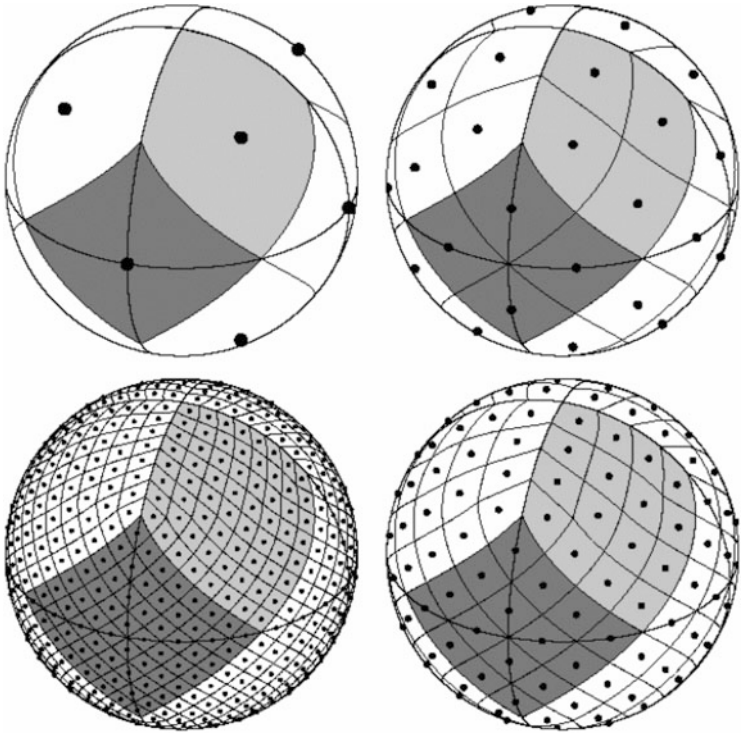

Fig. 20 Orthographic view of HEALPix division of a sphere. Overplot of equator and meridians illustrates octahedral symmetry of the HEALPIx construction. The lowest resolution corresponds to 12 base-resolution pixels. The sphere is hierarchically mosaically partitioned into curvilinear quadrangles. *Light-gray shading* shows one of eight (four north and four south) identical polar base-resolution pixels. *Dark-gray shading* shows one of four identical equatorial base-resolution pixels. The grid resolution of the mosaic increases in three steps from the base level (from Fig. a consecutively to Figs. b, c and, d). The grid resolution is characterized by the parameter $N_{\rm side}$ [equal to 1(2), 2(b), 4(c) and 8(d)] which determines the total number of pixels $N_{rmpix} = 12N_{\rm side}^2$, i.e. the sphere is partitioned into 12(a), 48(b), 192(c) and 768(d) pixels. Pixels are equal in area for each grid resolution. All pixel centers are located on rings of constant latitude, which is important for harmonic analysis and calculation of spherical harmonics

Hence, a HEALPix-pixelized map consists of $N_{\rm pix} = 12N_{\rm side}^2$ pixels with the same area $\Omega_{\rm pix} = \pi/(2N_{\rm side}^2)$.

The authors of Górski et al. (2005) use two system of indexation which are applied to process and store maps in the form of FITS-file (Flexible Image Transport System): with iso-latitude, or ring, index and nested index (Fig. 20). In the first case, we can simply count the pixels moving down from the north to the south pole along each isolatitude ring. In the second case, we can replicate the tree structure of pixel numbering by using a certain algorithm (Górski et al. 2005). The nested indexation allows applying the HEALPix scheme inside a database. Special procedures of the HEALPix package allow using alternative indexation schemes for the analyzed map.

The HEALPix software package, consisting of individual programs and program libraries, includes procedures for harmonic expansion on a sphere of the temperature and polarization anisotropy maps. Among the most frequently used procedures are *synfast*, for construction (and simulating) of maps; *anafast*, for multipole expansion, the angular power spectrum calculation (there is a possibility of data masking; and *map2gif*, for imaging maps on a sphere.

9 GLESP

In this subsection devoted to the GLESP approach we shall focus on the problem of processing on the sphere and then determine the scheme of pixelization. As it was noted above the pixelization of the CMB data on the sphere is only some part of the general problem, which is the determination of the coefficients of the spherical harmonic decomposition of the CMB signal for both anisotropy and polarization. These coefficients, which we call $a_{\ell m}$, are used in subsequent steps in the analysis of the measured signal, and in particular, in the determination of the power spectra, $C(\ell)$, of the anisotropy and polarization, in some special methods for components separation (Stolyarov et al. 2002; Naselsky et al. 2003) and phase analysis (Chiang et al. 2003; Naselsky et al. 2003, 2004a; Coles et al. 2004).

Below we describe a specific method to calculate the coefficients $a_{\ell m}$. It is based on the so-called Gaussian quadratures and is presented in the following subsection. In this specific pixelization scheme corresponds the position of pixel centers along the θ-coordinate to so-called the Gauss–Legendre quadrature zeros and it will be shown that this method increases the accuracy of calculations essentially.

Thus, the method of calculation of the coefficients a_{lm} dictates the method of the pixelization. We call our method GLESP, the Gauss–Legendre Sky Pixelization (Doroshkevich et al. 2003, 2011). We have developed a special code for the GLESP approach and a package of codes which are necessary for the whole investigation of the CMB data including the determination of anisotropy and polarization power spectra, C_ℓ, the Minkowski functionals and other statistics.

This subsection is devoted to description of the main idea of the GLESP method, the estimation of the accuracy of the different steps and of the final results, the description of the GLESP code and its testing. We do not discuss the problem how to make integration over a finite pixel size for the time ordered data. The simplest scheme of integration over pixel area is to use equivalent weight relatively to the center of the pixel. The GLESP code uses this method as e.g. HEALPix and Igloo do.

9.1 Main Ideas and Basic Relations

The standard decomposition of the measured temperature variations on the sky, $\Delta T(\theta, \phi)$, in spherical harmonics is

$$\Delta T(\theta, \phi) = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{m=\ell} a_{\ell m} Y_{\ell m}(\theta, \phi) \,, \tag{10}$$

$$Y_{\ell m}(\theta, \phi) = \sqrt{\frac{(2\ell+1)}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}} P_{\ell}^{m}(x) e^{im\phi} \,, \quad x = \cos\theta \,, \tag{11}$$

where $P_{\ell}^{m}(x)$ are the associated Legendre polynomials. For a continuous $\Delta T(x, \phi)$ function, the coefficients of decomposition, $a_{\ell m}$, are

$$a_{\ell m} = \int_{-1}^{1} dx \int_{0}^{2\pi} d\phi \Delta T(x, \phi) Y_{\ell m}^{*}(x, \phi) \,, \tag{12}$$

where $Y_{\ell m}^{*}$ denotes complex conjugation of $Y_{\ell m}$. For numerical evaluation of the integral Eq. (12) we will use the Gaussian quadratures, a method which was proposed by Gauss in 1814, and developed later by Christoffel in 1877. As the integral over x in Eq. (12) is an integral over a polynomial of x, we may use the following equality (Press et al. 1992):

$$\int_{-1}^{1} dx \Delta T(x, \phi) Y_{\ell m}^{*}(x, \phi) \equiv\equiv \sum_{j=1}^{N} w_j \Delta T(x_j, \phi) Y_{\ell m}^{*}(x_j, \phi) \,,$$

where w_j is a proper Gaussian quadrature weighting function. Here the weighting function $w_j = w(x_j)$ and $\Delta T(x_j, \phi) Y_{\ell m}^{*}(x_j, \phi)$ are taken at points x_j which are the net of roots of the Legendre polynomial

$$P_N(x_j) = 0 \,, \tag{13}$$

where N is the maximal rank of the polynomial under consideration. It is well known that the Eq. (13) has N number of zeros in interval $-1 \leq x \leq 1$. For the Gaussian–Legendre method Eq. (13), the weighting coefficients are Press et al. (1992)

$$w_j = \frac{2}{1 - x_j^2} [P_N'(x_j)]^{-2} \,, \tag{14}$$

where $'$ denotes a derivative. They can be calculated together with the set of x_j with the *"gauleg"* code (Press et al. 1992).

In the GLESP approach, there are the trapezoidal pixels bordered by θ and ϕ coordinate lines with the pixel centers (in the θ direction) situated at points with $x_j = \cos\theta_j$. Thus, the interval $-1 \le x \le 1$ is covered by N rings of the pixels. The angular resolution achieved in the measurement of the CMB data determines the upper limit of summation in Eq. (10), $\ell \le \ell_{\max}$. To avoid the Nyquist restrictions we use a number of pixel rings, $N \ge 2\ell_{\max}$. In order to make the pixels in the equatorial ring (along the ϕ coordinate) nearly squared, the number of pixels in this direction should be $N_\phi^{\max} \approx 2N$. The number of pixels in other rings, N_ϕ^j, must be determined from the condition of making the pixel sizes as equal as possible with the equatorial ring of pixels.

Figure 21 shows the weighting coefficients, w_j, and the position of pixel centers for the case $N = 31$. Figure 22 compares some features of the pixelization schemes used in HEALPix and GLESP. Figure 23 compares pixel shapes and distribution on a sphere in a full sky Mollweide projections of HEALPix and GLESP maps.

In the definition (10) are the coefficients $a_{\ell m}$ complex quantities while ΔT is real. In the GLESP code started from the definition (10) we use the following

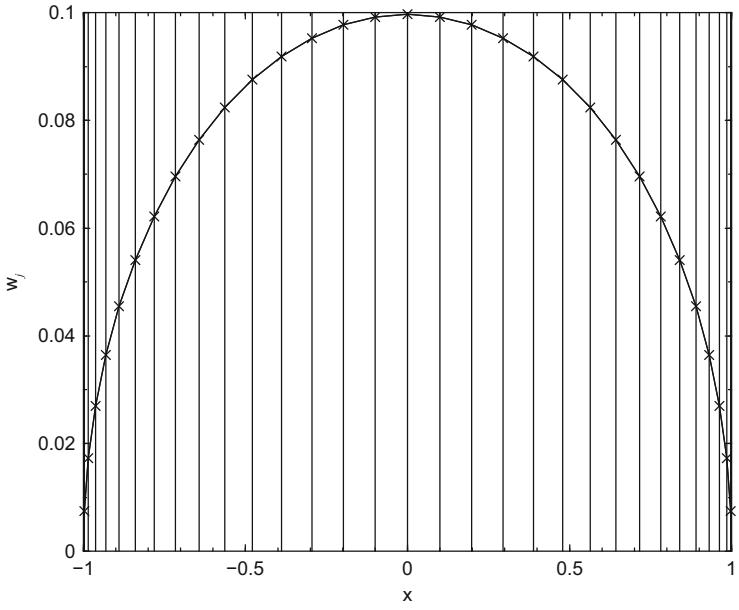

Fig. 21 Gauss–Legendre weighting coefficients (w_j) versus Legendre polynomial zeros ($x_j = \cos\theta_j$) being centers of rings used in GLESP for the case of $N = 31$. Positions of zeros are plotted by *vertical lines*

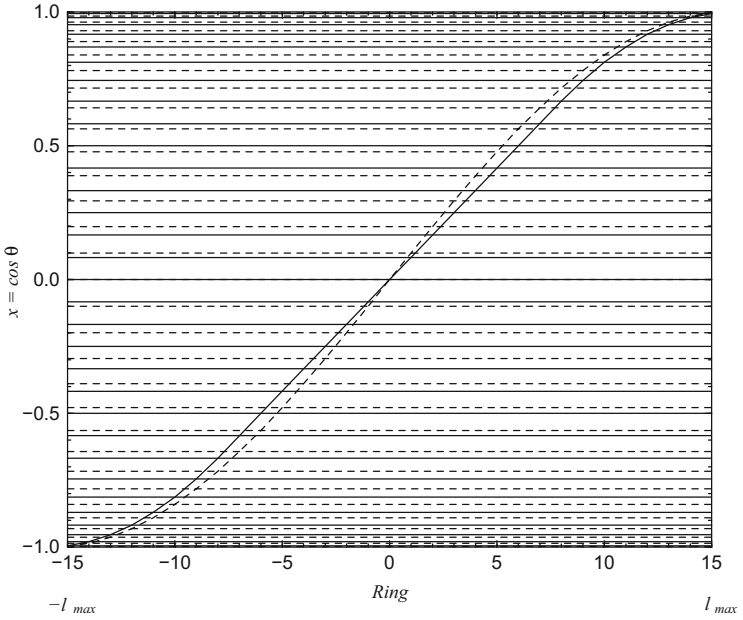

Fig. 22 Ring center position ($x_j = cos\theta_j$) vs ring number for two pixelization schemes, HEALPix (*solid*) and GLESP (*dashed*). Figure demonstrates the case of $N = 31$

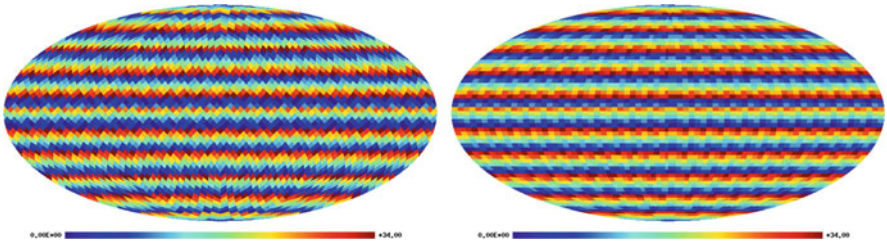

Fig. 23 Schematic representation of two types of pixelization on sphere: HEALPix (*top*) and GLESP (*bottom*). Various color of pixels is used to show their shape

representation of the ΔT

$$\Delta T(\theta,\phi) = \sum_{\ell=2}^{\ell_{\max}} a_{\ell 0} Y_{\ell 0}(\theta,\phi) + \\ + \sum_{\ell=2}^{\ell_{\max}} \sum_{m=1}^{\ell} \left(a_{\ell m} Y_{\ell m}(\theta,\phi) + a_{l,-m} Y_{\ell,-m}(\theta,\phi)\right) , \tag{15}$$

where

$$Y_{l,-m}(\theta,\phi) = (-1)^m Y^*_{l,m}(\theta,\phi), \quad a_{\ell m} = (-1)^m a^*_{l,-m} .$$

Thus,

$$\Delta T(\theta,\phi) = \frac{1}{\sqrt{2\pi}} \sum_{\ell=2}^{\ell_{\max}} Re(a_{\ell,0}) P_\ell^0(\cos\theta) + \\ + \sqrt{\frac{2}{\pi}} \sum_{\ell=2}^{\ell_{\max}} \sum_{m=1}^{l} \sqrt{\frac{2\ell+1}{2} \frac{(\ell-m)!}{(\ell+m)!}} P_\ell^m(\cos\theta) \times \\ \times \left[Re(a_{\ell m}) \cos(m\phi) - Im(a_{\ell m}) \sin(m\phi) \right], \quad (16)$$

where $P_\ell^m(\cos\theta)$ are the well known associated Legendre polynomials. In the GLESP code, we use normalized associated Legendre polynomials whose recurrence relation is given in Appendix 1.

9.2 Properties of GLESP

Following the previous discussion we define the new pixelization scheme GLESP as follows:

- In the polar direction $x = cos\theta$, we define $x_j, j = 1, 2, .., N$, as the net of roots of Eq. (13).
- Each root x_j determines the position of a ring with N_ϕ^j pixel centers with ϕ-coordinates ϕ_i.
- All the pixels have nearly equal area.
- Each pixel has weight w_j [see Eq. (14)].

In our numerical code which realizes the GLESP pixelization scheme we use the following conditions.

- Borders of all pixels are along the coordinate lines of θ and ϕ. Thus with a reasonable accuracy they are trapezoidal.
- The number of pixels along the azimuthal direction ϕ depends on the ring number. The code allows an arbitrary number of these pixels to be chosen. The number of pixels depends on the $\ell_{\max}$ accepted for the CMB data reduction.
- To satisfy the Nyquist's theorem, the number N of the ring along the $x = \cos(\theta)$ axis must be taken as $N \geq 2\ell_{\max} + 1$.
- To make equatorial pixels roughly square, the number of pixels along the azimuthal axis, ϕ, is taken as $N_\phi^{\max} = \text{int}(2\pi/d\theta_k + 0.5)$, where $k = \text{int}(N/2 + 0.5)$, and $d\theta_k = 0.5(\theta_{k+1} - \theta_{k-1})$.
- The nominal size of each pixel is defined as $S_{\text{pixel}} = d\theta_k \times d\phi$, where $d\theta_k$ is the value on the equatorial ring and $d\phi = 2\pi/N_\phi^{\max}$ on equator.
- The number N_ϕ^j of pixels in the jth ring at $x = x_j$ is calculated as $N_\phi^j = \text{int}(2\pi\sqrt{1-x_j^2}/S_{pixel} + 0.5)$;

- Polar pixels are triangular.
- Because the number N_ϕ^j differs from 2^k where k is integer, we use for the Fast Fourier transformation along the azimuthal direction the FFTW code. This code permits one to use not only 2^n-approach, but other base-numbers too, and provide even higher speed.

With this scheme, the pixel sizes are equal inside each ring, and with a maximum deviation between the different rings of $\sim$1.5 % close to the poles (Fig. 24). Increasing resolution decreases an absolute error of an area because of the inequivalence of polar and equator pixels proportionally to N^{-2}.

Figure 25 shows that this pixelization scheme for high resolution maps (e.g. $\ell_{\max} > 500$) produces nearly equal thickness $d\theta$ for most rings.

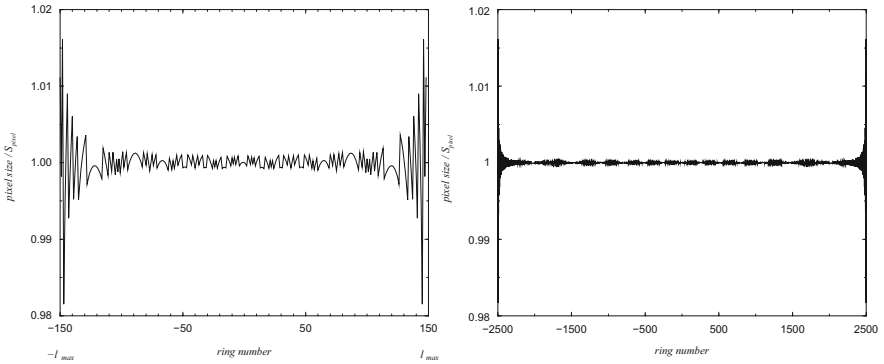

Fig. 24 Pixel size/equator pixel area vs ring-number for GLESP for number of rings $N = 300$ and $N = 5000$

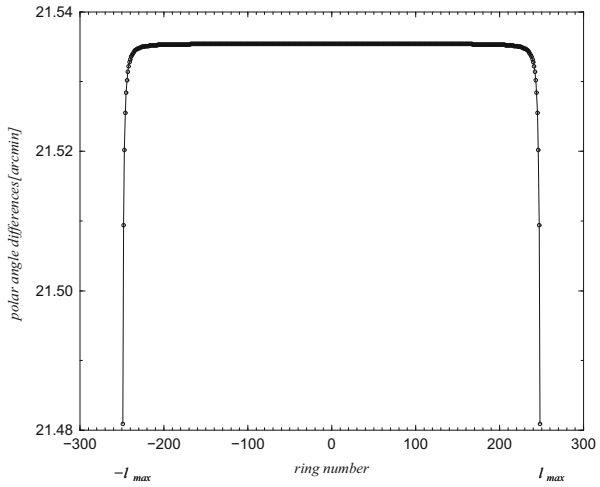

Fig. 25 Pixel size along polar angle ($\ell_{\max} = 250$)

GLESP has not the hierarchical structure, but the problem of the closest pixel selection is on the software level. Despite GLESP is close to the Igloo pixelization scheme in the azimuthal approach, there is a difference between the two schemes in connection with the θ-angle (latitude) pixel step selection. Therefore, we can not unify these two pixelizations. The Igloo scheme applied to the GLESP latitude step will give very different pixel areas. The pixels will be neither equally spaced in latitude, nor of uniform area, like Igloo requires.

9.3 Re-pixelization Problem

Unifying ideas of the non-hierarchical structure of the pixelization, but orienting onto the accuracy of $a_{\ell m}$ calculation, we prepared the GLESP package at http://www.glesp.nbi.dk (see also Appendix 2) giving accuracy limited only by precision of the stored data format. However, there are natural limits of the CMB map accuracy connected with

- level of the system noise,
- level of knowledge about local, galactic and extragalactic foregrounds to be removed,
- instability of reconstruction methods,
- superimposed systematics errors.

The last one connected with some operations often used for map analysis like map rotation in pixel domain and re-pixelization which could be used for map conversion between different pixelization systems.

To transfer a sky distribution map from one pixel grid (e.g. from HEALPix to GLESP) to another one, we should use one of two ways:

1. to calculate $a_{\ell m}$-coefficients and after that to restore a map in a new grid;
2. or to use re-pixelization procedures on the current brightness distribution.

Any re-pixelization procedure will cause loss of information and thereby introduce uncertainties and errors. The GLESP code has procedures for map re-pixelization based on two different methods in the $\Delta T(\theta, \phi)$-domain: the first one consists in averaging input values in the corresponding pixel, the second one is connected with spline interpolation inside the pixel grid.

In the first method, we consider input pixels which fell in our pixel with values $\Delta T(\theta_i, \phi_i)$ to be averaged with a weighting function. The realized weighting function is a function of simple averaging with equal weights. This method is widely used in appropriation of a given values to the corresponding pixel number.

In the second method of re-pixelization, we use a spline interpolation approach. If we have a map $\Delta T(\theta_i, \phi_i)$ recorded in the knots different from the Gauss–Legendre grid, it is possible to repixelize it to our grid $\Delta T(\theta_i', \phi_i')$ using approximately the same number of pixels and the standard interpolation scheme based on the cubic spline approach for the map re-pixelization. This approach is sufficiently fast

because the spline is calculated once for one vector of the tabulated data (e.g. in one ring), and values of interpolated function for any input argument are obtained by one call of separate routine (see routines "*spline*" to calculate second derivatives of interpolating function and "*splint*" to return a cubic spline interpolated value in Press et al. (1992).

Our spline interpolation consists of the three steps:

- we set equidistant knots by the ϕ-axis to reproduce an equidistant grid;
- we change the grid by $x = cos(\theta)$-axis to the required GLESP grid;
- after that, we recalculate ϕ-knots to the rings corresponding to the GLESP x-points.

Figure 26 demonstrates the deviation of accuracy of the power spectrum in a case of re-pixelization from a HEALPix map to a GLESP map with the same resolution. As one can see, for the range $\ell \leq \ell_{\max}/2$, re-pixelization reproduces correctly all properties of the power spectra. For $\ell \geq \ell_{\max}/2$ some additional investigations need to be done to take into account the pixel-window function.

The most standard accurate way is to produce maps just from the timed ordered data or to use transformation from a map of one pixelization to $a_{\ell m}$-coefficients and converse them to a map of another pixelization system.

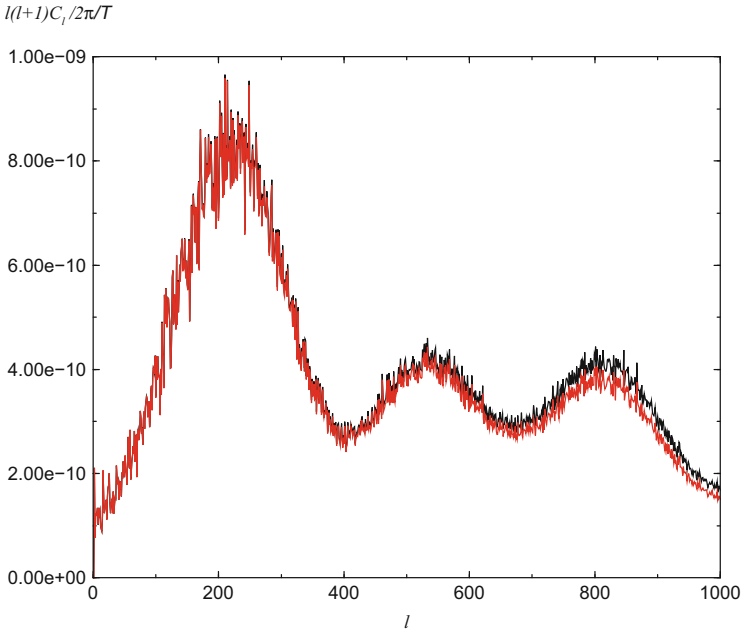

Fig. 26 Power spectra calculated for the initial HEALPix map (*upper curve*) with $\ell_{\max} = 1000$, $N_{side} = 1024$, pixel size = 11.8026$(\prime)^2$, and N_{tot}=12 582 912, and for resulting re-pixelized GLESP map (*lower curve*) with the closest possible pixel size = 11.8038$(\prime)^2$, N_{tot}=12 581 579. Deviations of the power spectra at high ℓ illustrate the ratio of the HEALPix and GLESP window functions

10 Component Separation

Besides CMB, the signal measured in experiments involves contributions from galactic background components and from galactic and extragalactic radio sources. These can be taken into account by superposing masks, i.e., by excluding certain patches of the celestial sphere from consideration [e.g., 22 % in the case of the KQ85y7 mask used for the data collected by the WMAP mission (Gold et al. 2011)]. But there is another way. It is possible, however, to restore the CMB signal over the whole celestial sphere using the results of multifrequency observations. One of the methods for determining the complete CMB temperature map is based on a combination of observational data at different frequencies, additionally multiplied by certain coefficients permitting the exclusion of the galactic signal from the result and, thus, singling out the microwave relic background (Hinshaw et al. 2007). In this approach, the idea used is that the radiation spectra of galactic background components (namely, of synchrotron radiation, free–free radiation, and of the radiation of dust) differ from the CMB spectrum which is produced in random process and has black body emission spectrum. All modern methods of CMB restoration (Stolyarov et al. 2002; Hinshaw et al. 2007; Planck Collaboration 2014c; Doroshkevich and Verkhodanov 2011) use these properties.

As an example, one can consider the WMAP separation method realized in Internal Linear Combination (ILC) of multifrequency observation maps (Hinshaw et al. 2007). Because the combination of channels in the WMAP mission is achieved without using observations from other experiments, this method has been termed Internal Linear Combination. The coefficients can be determined by minimizing the dispersion in the resultant map, equating their sum to unity, so as to preserve the overall normalization of the CMB signal. In describing this procedure, we first note that from the results of simulation (Hinshaw et al. 2007), the instrumental noise has been established to not affect the situation significantly, because it only provides a shift of the order of 10 mK in the estimate of the signal in the galactic plane. In the simple case where the instrumental noise can be neglected and the background components have the same spectrum in the region investigated and differ from each other in different parts of this region only in temperature, the sought ILC temperature can be written as a linear combination of signals from the maps for different frequencies ν_i:

$$T_{\mathrm{ILC}}(p) = \sum_i \zeta_i T_i(p) = \sum_i \zeta_i [T_c(p) + S_i T_f(p)] = T_c(p) + \Gamma T_f(p)\,. \qquad (17)$$

Here, $T_i(p) \equiv T(\nu_i, p)$ is the map of the signal observed at the frequency ν_i, p is a certain pixel of the image (the smallest region of the map with the measured temperature), the map of the signal $T_i(p) = T_c(p) + S_i T_f(p)$ is represented as a sum of the CMB maps $T_c(p)$ and of the background component $S_i T_f(p)$, the coefficient $S_i \equiv S(\nu_i)$ describes the total frequency spectrum of background radiation, and $T_f(p)$ is the distribution of the background radiation temperature. The coefficients

ζ_i that are to be determined satisfy the normalization condition $\sum_i \zeta_i = 1$. The notation $\Gamma \equiv \sum_i \zeta_i S_i$ was introduced in Eq. (17).

The coefficients ζ_i are determined by minimizing the dispersion of $T_{\rm ILC}(p)$. For this dispersion, we have (Hinshaw et al. 2007)

$$\sigma^2_{\rm ILC} = \langle T^2_{\rm ILC}(p)\rangle - \langle T_{\rm ILC}(p)\rangle^2 = \langle T_c^2\rangle - \langle T_c\rangle^2 + 2\Gamma[\langle T_c T_f\rangle - \langle T_c\rangle\langle T_f\rangle] + \Gamma^2\left[\langle T_f^2\rangle - \langle T_f\rangle^2\right]$$
$$= \sigma_c^2 + 2\Gamma\sigma_{cf} + \Gamma^2\sigma_f^2 \qquad (18)$$

where the angular brackets '$\langle\rangle$' denote averaging over the pixels of the selected region. The minimization of $\sigma^2_{\rm ILC}$

$$0 = \frac{\delta\sigma^2_{\rm ILC}}{\delta\zeta_i} = 2\frac{\delta\Gamma}{\delta\zeta_i}\sigma_{cf} + 2\Gamma\frac{\delta\Gamma}{\delta\zeta_i}\sigma_f \qquad (19)$$

yields $\Gamma = -\sigma_{cf}/\sigma_f^2$ and

$$T_{\rm ILC}(p) = T_c(p) - \frac{\sigma_{cf}}{\sigma_f^2}T_f(p)\,.$$

In the ideal case, where no correlation between the CMB and the background exists, i.e., $\sigma_{cf} = 0$, the ILC map coincides with the CMB map. Actually, as emphasized in Hinshaw et al. (2007), ILC map is shifted toward a decrease in the correlation between the CMB signal and the signal from background components.

We note that different versions of the ILC method exist in both pixel space and harmonic space (see the review in Planck Collaboration 2014c). The regions where this method is used can be determined as follows: (1) by dividing the sphere into separate zones (Hinshaw et al. 2007) (for example, in the analysis of WMAP data, the sphere was divided into 12 regions, most of which were situated in the galactic plane); (2) by applying selection rules for averaged pixels (Doroshkevich and Verkhodanov 2011); (3) by fixing a certain set of harmonics (Kim et al. 2009). It is also possible to use other combinations of radio-frequency observations. The modifications result in only a few different maps being obtained. Moreover, there are different versions of the actual procedure for producing a map of an internal linear combination (for instance, the Lagrange ILC, LILC, method (Eriksen et al. 2004a), which yields the same results as ILC). Finally, separation of the signal components and the production of CMB maps is also possible with other methods, such as the Maximum Entropy Method (MEM) (Hinshaw et al. 2007; Stolyarov et al. 2002), fitting templates of background components to other observations (Hinshaw et al. 2007; Stolyarov et al. 2002), Wiener filtration (the Wiener-filtered map, WFM) performed in Tegmark et al. (2003), or weighted removal of the background (the foreground-cleaned map (FCM) in Tegmark et al. 2003). The last was used to produce CMB maps exhibiting a higher resolution ($\ell_{\rm max} = 600$) than the WMAP.

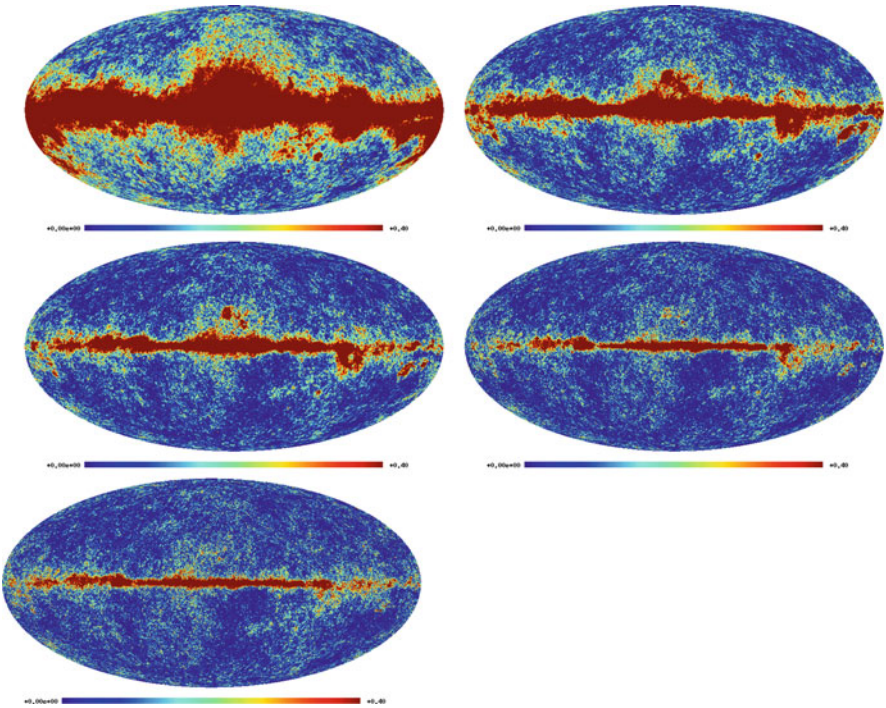

Fig. 27 Maps of the observed microwave radiation in WMAP frequency channels: (**a**) 23 GHz (band K), (**b**) 33 GHz (band Ka), (**c**) 41 GHz (band Q), (**d**) 61 GHz (band V), and (**e**) 94 GHz (band W) from data obtained during the 7th year of WMAP observations. The maps are produced in galactic coordinates. From http://wmap.gsfc.nasa.gov

In what follows, we mainly deal with the ILC WMAP map, although maps obtained by other methods are also mentioned.

Observations were performed by the WMAP within five frequency bands: 23 GHz (band K), 33 GHz (band Ka), 40 GHz (band Q), 61 GHz (band V), and 94 GHz (band W) (Fig. 27), involving intensity and polarization measurements. Data arrays collected by the mission during 1, 2, 5, 7 and 9 years of work were put on a website for general use (Bennett et al. 2003, 2013; Hinshaw et al. 2007, 2009; Jarosik et al. 2011). As a result of observational data processing, which included the registration and storage of time series, map making and sky pixelization, and the separation of signal components and their subsequent analysis, data were obtained on the anisotropy and polarization distributions of the CMB and of background components (synchrotron and free-free radiation, the radiation of dust), and their power spectra were also calculated. The ILC WMAP map produced was smoothed out by a Gaussian-shaped diagram with a 1° resolution. The entire archive of observational and processed data is available and accessible to the scientific community at the WMAP website.

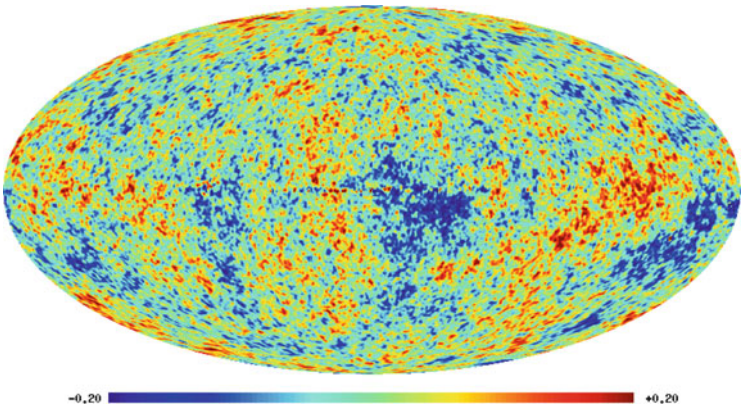

Fig. 28 The ILC CMB map made in galactic coordinates and based on the data obtained by the WMAP during its 7th year of work (shown with resolution of up to $\ell_{\max} = 150$). From http://wmap.gsfc.nasa.gov

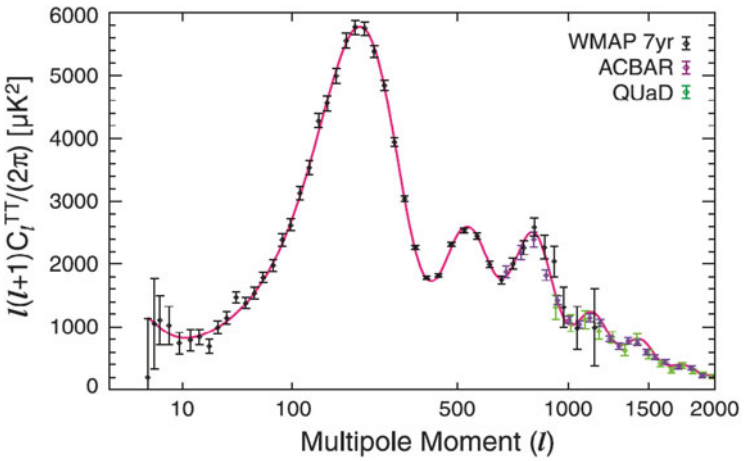

Fig. 29 Angular power spectrum $\ell(\ell + 1)C(\ell)/2\pi$ of the WMAP map for the 7th year of observations (Komatsu, et al. 2011b) and measurement results of the angular power spectrum of temperature fluctuations in ACBAR (Reichardt et al. 2009) and QUaD (Brown et al. 2009) experiments. The results are shown for the multipole range up to $\ell < 2000$, within which the contribution of the Zeldovich–Sunyaev effect and of point sources is not high. The *solid curve* shows the simulated spectrum for ΛCDM–cosmology with parameters determined on the bases of WMAP results. From http://wmap.gsfc.nasa.gov

In Fig. 28, a map of the CMB anisotropy distribution reconstructed by the ILC method is presented for not very high harmonics ($\ell \leq 150$). Figure 29 shows the angular power spectrum produced using data from the WMAP mission and from the ACBAR (Arcminute Cosmology Bolometer Array Receiver)

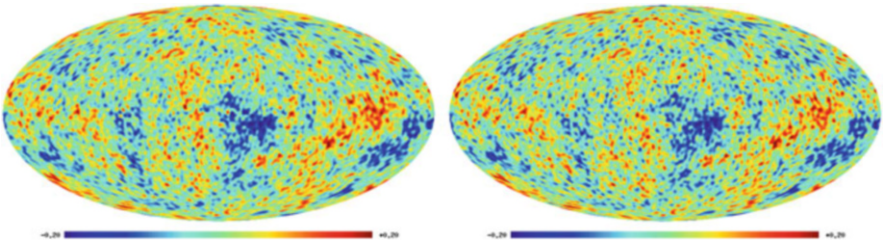

Fig. 30 CMB maps restored from the WMAP (*left*, ILC map) and Planck (*right*, SMICA map) observational data and smoothed up to $\ell_{\max}$=100. From http://wmap.gsfc.nasa.gov

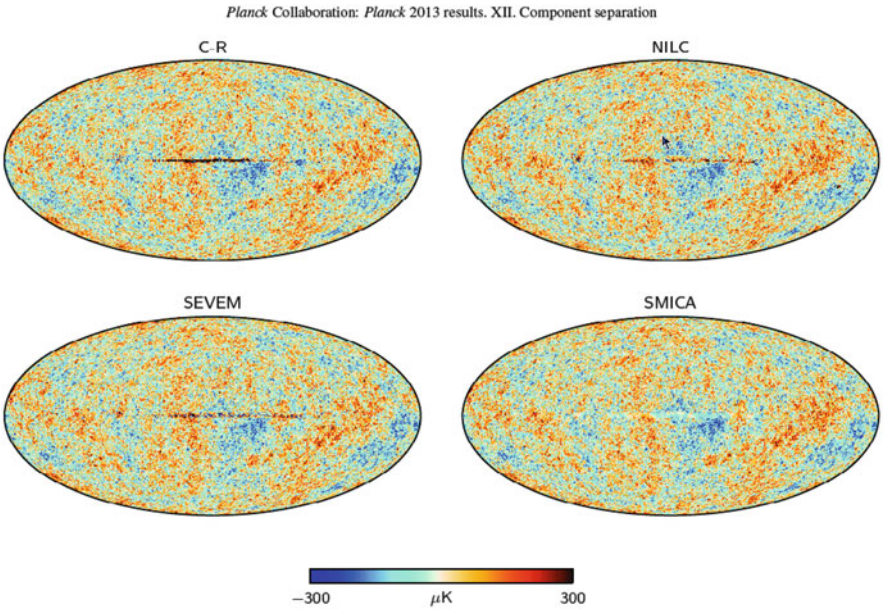

Fig. 31 Four CMB maps (Commander–Ruler, NILC, NILC, SMICA) restored by different methods from the Planck observational data. From http://wmap.gsfc.nasa.gov

(Reichardt et al. 2009) and QUaD (QUEST (Q and U Extragalactic Sub-mm Telescope) at DASI) experiments (Brown et al. 2009).

In Figs. 30 and 31, the restored CMB maps of WMAP and Planck experiments are shown.

For the restored CMB signal, the angular power spectrum is calculated using the so called $a_{\ell m}$-coefficients:

$$C(\ell) = \frac{1}{2\ell + 1}\left[|a_{\ell 0}|^2 + 2\sum_{m=1}^{\ell}|a_{\ell m}|^2\right]. \tag{20}$$

The $a_{\ell m}$-coefficients are obtained in the standard decomposition of the measured temperature variations on the sky, $\Delta T(\theta, \phi)$, in spherical harmonics (multipoles):

$$\Delta T(\theta, \phi) = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{m=\ell} a_{\ell m} Y_{\ell m}(\theta, \phi) \,, \tag{21}$$

$$Y_{\ell m}(\theta, \phi) = \sqrt{\frac{(2\ell+1)}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}} P_{\ell}^{m}(x) e^{im\phi} \,, x = \cos\theta \,,$$

where $P_{\ell}^{m}(x)$ are the associated Legendre polynomials. For a continuous $\Delta T(x, \phi)$ function, the coefficients of decomposition, $a_{\ell m}$, are

$$a_{\ell m} = \int_{-1}^{1} dx \int_{0}^{2\pi} d\phi \Delta T(x, \phi) Y_{\ell m}^{*}(x, \phi) \,, \tag{22}$$

where $Y_{\ell m}^{*}$ denotes complex conjugation of $Y_{\ell m}$. This angular spectrum being the measured characteristics of the CMB, from one hand, is the function of the main cosmological parameters

$$C_{\ell} \equiv C_{\ell}(h, \Omega_b h^2, \Omega_{CDM} h^2, \Omega_{\Lambda} \,, \Omega_{\nu}, n, \ldots) \,,$$

from another hand, which can be calculated using corresponding computational facilities (Seljak and Zaldarriaga 1996; Lewis et al. 2000).

11 Non-Gaussianity

In the standard cosmological scenario with the Big Bang and simple inflation (Starobinsky 1979; Sato 1981; Guth 1981; Albrecht and Steinhardt 1982; Linde 1982), quantum fluctuations of the scalar field generate inhomogeneities in the distribution of visible and dark matter (Mukhanov and Chibisov 1981; Hawking 1982; Starobinsky 1982; Guth and Pi 1982; Bardeen et al. 1983), which leads to fluctuations in the microwave background radiation of the Universe. Temperature and polarization fluctuations of the cosmic microwave background (CMB) are expected, and confirmed at a certain level of precision by observations, to be Gaussian random fields, statistically isotropic in space. Nevertheless, certain models predict small but quite noticeable deviations of the signal from the Gaussian statistic and/or statistical isotropy, which, in principle, may be due to a number of reasons. Within the inflation theory, a relatively strong non-Gaussianity arises in models involving complex inflation (Allen et al. 1987; Linde and Mukhanov 1977; Bernardeau and Uzan 2002; Dvali et al. 2004; Kofman 2003; Barnaby and Cline 2007) (for example, when a nonlinear relation exists between classical fluctuations of the scalar field generated at the inflation stage and the observed

field of matter density fluctuations), while statistical anisotropy may be caused by anisotropic expansion at the inflationary stage (Gumrukcuoglu et al. xxxx, 2007, 2010; Ackerman et al. 2007; Watanabe et al. 2009; Dimopoulos et al. 2010; Watanabe et al. 2010; Dulaney and Gresham 2010), related, for instance, to the presence of classical vector fields. Other sources of non-Gaussianity and of statistical anisotropy, interesting from the standpoint of cosmology, could be a nontrivial topology of space (Starobinsky 1993; de Oliveira-Costa et al. 1996; Inoue et al. 2000; Dineen et al. 2005; Kunz et al. 2008), topological defects (Turok 1991; Turok and Spergel 1990; Durrer 1999; Cruz et al. 2007), anisotropic expansion (Jaffe et al. 2005; Demiański and Doroshkevich 2007), the primordial magnetic field (Durrer et al. 1998; Mack et al. 2002; Naselsky et al. 2004b; Kahniashvili and Ratra 2005; Kahniashvili et al. 2008), etc. Owing to the appearance of new complete-sphere data (Planck Collaboration 2014a), the issue of searching for and explaining the non-Gaussian properties of the CMB has become especially important. Although non-Gaussianity and statistical anisotropy are far from being identical concepts, they are quite close to each other from a cosmological standpoint.

For searching for and analyzing the non-Gaussian properties of the CMB temperature, methods have been developed that use the distribution of temperature fluctuations over the celestial sphere, $\Delta T(\theta, \phi)$, where θ and ϕ are the angles in a polar coordinate system, as well as methods based on the expansion of temperature fluctuations with respect to spherical harmonics. Theoretically, the relation between the primary inhomogeneities, representing adiabatic scalar perturbations, and the coefficients $a_{\ell m}$ is linear (Komatsu and Spergel 2001):

$$a_{\ell m} = (-i)^{\ell} \int \frac{d^3\mathbf{k}}{(2\pi)^3} \Phi(\mathbf{k}) g_{T\ell}(k) Y^*_{\ell m}(\hat{\mathbf{k}}) \,, \tag{23}$$

where $\Phi(\mathbf{k})$ describes the primary perturbation of the density (more precisely, of the gravitational potential) in Fourier space, $g_{T\ell}(k)$ is the transfer function, and $\hat{\mathbf{k}}$ is the unity vector directed along the wave vector $\mathbf{k}$. A linear relation also exists between the temperature fluctuations $\Delta T(\theta, \phi)$ and the primary fluctuations $\Phi(\mathbf{k})$. The total radiation transfer function $g_{T\ell}(k)$ can be computed with the aid of the CMBFAST (Seljak and Zaldarriaga 1996) or CAMB (Lewis et al. 2000) program. A simple linear relation permits, at least theoretically, relating the CMB signal statistics and the statistics of primary perturbations: if the primary fluctuations $\Phi(\mathbf{k})$ are non-Gaussian, the non-Gaussianity can also be observed in the CMB. In sensitive surveys of the entire sky, for instance, the WMAP and Planck missions, it is already possible to search for deviations of the signal from the Gaussian statistics.

We note an issue that must be raised in discussing the problems of CMB non-Gaussianity, related to the calculation of the angular power spectrum $C(\ell)$. By definition,

$$C(\ell) = \frac{1}{2\ell + 1} \left[|a_{\ell 0}|^2 + 2 \sum_{m=1}^{\ell} |a_{\ell m}|^2 \right] , \tag{24}$$

and the coefficients $a_{\ell m}$ are obtained by transforming the map into harmonics:

$$a_{\ell m} = \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} d\phi \Delta T(x,\phi) Y^*_{\ell m}(x,\phi)\,. \tag{25}$$

In this case, the two-point correlator (with the averaging performed over the ensemble of universes) has the form

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell \delta_{\ell\ell'} \delta_{mm'}\,. \tag{26}$$

In the case of averaging at a given ℓ, the meaning of the quantity C_ℓ for non-Gaussian data is no longer obvious.

The simplest and best studied quantity in which the non-Gaussianity of CMB fluctuations may be manifested is the three-point correlation function or its harmonic analog, the bispectrum

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle\,. \tag{27}$$

The bispectrum is quite sensitive to certain forms of non-Gaussianity that are considered 'standard'. These include the so-called local form obtained under the assumption that primary fluctuations exhibit a nonlinearity local in the coordinate space (Komatsu and Spergel 2001):

$$\Phi(x) = \Phi_L(x) + f_{\rm NL}(\Phi^2_L(x) - \langle \Phi^2_L(x) \rangle)\,, \tag{28}$$

where $\Phi_L(x)$ denotes a linear Gaussian field, $\langle \Phi_L(x) \rangle$=0, and $f_{\rm NL}$ is a constant describing the nonlinearity in the form of a quadratic correction to perturbations of the gravitational potential (curvature). Such a form of non-Gaussianity indeed arises in some inflationary models involving an additional scalar field (curvaton) (Linde and Mukhanov 1977; Lyth and Wands 2002; Moroi and Takahashi 2002) and/or the nontrivial dynamics of postinflational modulated reheating (Dvali et al. 2004; Kofman 2003). The 'equilateral' (Creminelli et al. 2006) and 'orthogonal' (Senatore et al. 2010) forms of non-Gaussianity are also considered to be 'standard'. The WMAP mission team devoted some of its work to the investigation of these and of certain other forms of non-Gaussianity, the main instrument being precisely the angular bispectrum of the CMB temperature. The authors of Komatsu, et al. (2011b) established the results of an analysis of 7-year-long observations performed by the WMAP mission to be consistent, at a 95 % confidence level, with the hypothesis that primary fluctuations are Gaussian, and upon combining their results with those of the Sloan Digital Sky Survey (SDSS) [$-29 < f_{\rm NL} < 70$ (Slosar et al. 2008)] they found, in particular, that $-5 < f_{\rm NL} < 59$.

According to the Planck data (Planck Collaboration 2014d), the non-Gaussianity was estimated with unprecedented accuracy. Using three optimal bispectrum estimators, separable template-fitting, binned, and modal, there were obtained consistent values for the primordial local, equilateral, and orthogonal bispectrum amplitudes:

result $f_{\mathrm{NL\,local}} = 2.7 \pm 5.8$, $f_{\mathrm{NL\,equil}} = -42 \pm 75$, and $f_{\mathrm{NL\,orth}} = -25 \pm 39$ (68 % CL statistical). The detected non-Gaussianity was found using skew-C_ℓ statistics in a nonzero bispectrum from residual point sources, and the integrated Sachs–Wolfe lensing bispectrum at a level expected in the ΛCDM scenario.

12 Anomalies

Among the most discussed anomalies violating our expectation from the CMB Gaussian distribution, there are Axis of Evil (Land and Magueijo 2005), Cold Spot (Cruz et al. 2005), violation of parity in the power spectrum (Kim and Naselsky 2010), asymmetry 'North–South' in galactic coordinate system (Eriksen et al. 2004c). And the Planck data added a new unexpected phenomena—too low amplitude of low harmonics (Planck Collaboration 2014f). All these anomalies occur at the largest angular scales ($\theta > 1°$) and demonstrate observation of statistical anisotropy being a sign of non-Gaussianity at low multipoles.

There are two basic approaches in understanding the origin of anomalies. The first one is based on suggestions of complex processes during early stages of Universe. The second one follows the idea of connection of anomalies with foregrounds and/or data analysis procedures. To distinguish the contribution of different effects, we should look attentively to the observations and data analysis of the CMB.

Below we describe the main anomalies registered both in WMAP and Planck data.

12.1 Axis of Evil

The Axis of Evil (Fig. 32) is the most famous among non-Gaussian features of the WMAP CMB data. The Axis unifies some problems which require special explanations. They are the planarity and alignment of the two harmonics,

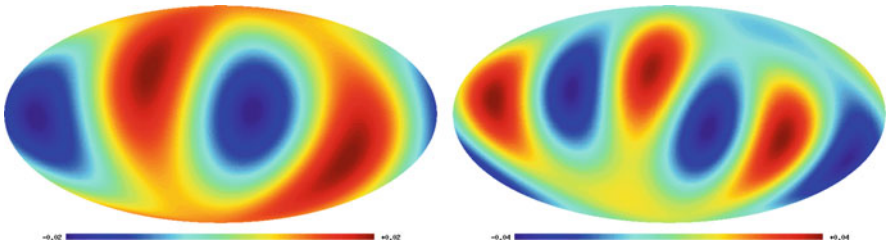

Fig. 32 Axis of Evil: planarity and alignment of the quadrupole (*left*) and the octupole (*right*) on the WMAP CMB map

quadrupole and octupole, and, partly, the problem of extremely low amplitude of the quadrupole. Different estimations of the significance of existence of this axis, and several hypotheses on its origin were made. Various studies, e.g. Copi et al. (2006), Gruppuso and Burigana (2009) investigated the contribution of background components and their influence on the alignment of multipoles ($\ell = 2$ and $\ell = 3$), and indicated a small probability of the background effect on the orientation of the low multipoles. In Copi et al. (2006), where the multipole vectors were used for the estimates of this effect, it was also noted that the positions of the quadrupole and octupole axes correspond to the geometry and direction of motion of the Solar System and are perpendicular to the ecliptic plane and the plane, given by the direction to the dipole. Randomness of such an effect is estimated by the authors as unlikely at the significance level exceeding 98 % and exclude the effect of residual contribution of background components. Continuing the research done, Copi et al. (2009) conclude that the characteristics of low multipoles are abnormally different from random, which may be due to the statistical anisotropy of the universe at large scales, or to the problems of the ILC signal deconvolution method. Park et al. (2007) note that the planarity of the quadrupole and octupole is not statistically significant. They also stress that the residual photon radiation in the ILC map does not affect significantly the level of the effect.

Cosmological models were developed to explain the prominence of the axis in the orientation of multipoles. The alignment of the quadrupole and octupole could be explained within the framework of these models. Various models include the anisotropic expansion of the Universe, rotation and magnetic field (Jaffe et al. 2006; Demiański and Doroshkevich 2007; Koivisto and Mota 2008).

There are some hints demonstrating that the problem of existence of Axis of Evil can be connected with the instability of CMB reconstruction at low multipoles ($2 \leq \ell \leq 10$) in ILC method (Naselsky and Verkhodanov 2007; Naselsky et al. 2008). Another possible solution of the problem is to construct the separation methods on the homogeneous samples of pixels where possible to tune selection of subsample in such a way that the quadrupole amplitude of the restored map grows and phase changes, so, no axis of evil exist (Doroshkevich and Verkhodanov 2011).

The Planck team using new data (Planck Collaboration 2014f) detected the angle between planes of quadrupole and octupole is equal $\sim 13°$ (against $\sim 3°$ or $\sim 9°$ for WMAP data at different observational years) and declared that significance of the quadrupole-octupole alignment is substantially smaller than for the WMAP data, falling to almost 98 % confidence level for the Commander-Ruler and SEVEM maps and 96.7 % confidence level for the NILC map. However, later, Copi et al. (2013) demonstrated that the WMAP and Planck data confirm the alignments of the largest observable CMB modes in the Universe. Using different statistical methods to control the mutual alignment between the quadrupole and octupole, and the alignment of the plane defined by the two harmonics with the dipole direction, authors obtained that both phenomena are at the greater than 3σ level for three Planck maps (SMICA, SEVEM, NILC) studied.

12.2 Cold Spot

The next exited feature discussed in Introduction is the Cold Spot (CS) (Fig. 33). This is a cold region exhibiting a complex structure identified in the CMB using spherical Mexican hat wavelets Cruz et al. (2005). The non-Gaussianity of the signal in the Southern hemisphere was explained precisely by the existence of this region. The galactic coordinates of center of the spot are $b = -57°$, $l = 209°$. The probability of the signal in CS being consistent with the Gaussian model if spherical wavelets are used is about 0.2 % Cruz et al. (2005). After obtaining indication of the signal non-Gaussianity at the CS as well as messages on the reduced density of source (Rudnick et al. 2007) in smoothed maps of radio survey NVSS at 1.4 GHz (Condon et al. 1998), several hypotheses concerning the origin of the Cold Spot were discussed which were related to the integrated Sachs–Wolfe effect (Rudnick et al. 2007), the topological defect (Cruz et al. 2007), anisotropic expansion (Jaffe et al. 2005), the artifact of data analysis (Naselsky et al. 2010), and simply a random deviation (Bennett et al. 2011).

As was noted in Naselsky et al. (2010), the possible galactic foreground residuals in the CMB maps can produce such a type of the spot as a part of non-Gaussianity at low multipoles. We should add that the CS is also manifested in the data of 1982 in maps of a low-frequency survey (Haslam et al. 1982) where synchrotron radiation contributes significantly to the background (Fig. 34). The following fact supports the

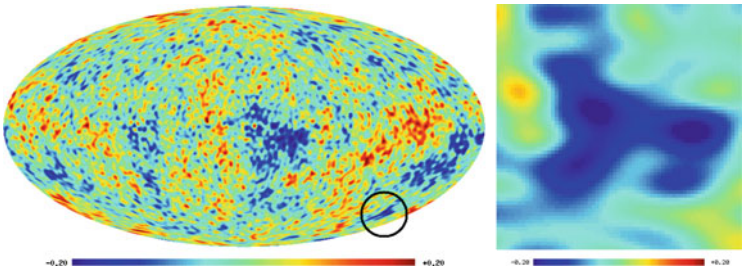

Fig. 33 Cold Spot: position of the Cold Spot on the WMAP CMB map (*left*) and its shape (*right*)

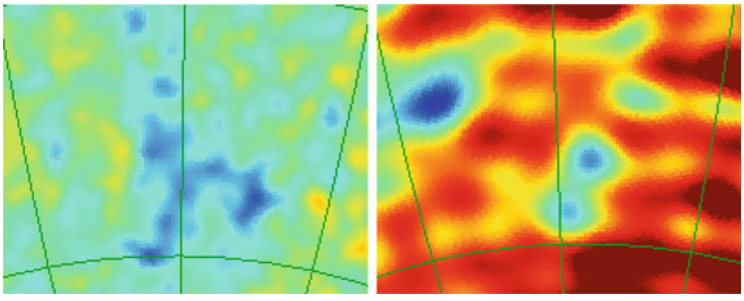

Fig. 34 Cold Spot on the WMAP CMB map (*left*) and 408 MHz map (*right*) with synchrotron emission

hypothesis of the CS being the Galactic phenomena: there exists a high correlation of positions of peaks of CMB fluctuation and galactic magnetic field distribution (Hansen et al. 2012).

12.3 Violation of the Power Spectrum Parity

A remarkable manifestation of non-Gaussian properties of low multipoles consists in parity asymmetry first noticed in Kim and Naselsky (2010) and confirmed in Planck data (Planck Collaboration 2014f). For a Gaussian random field of primary perturbations $\Phi(\mathbf{k})$ with a flat power spectrum, the presence of a plateau in the CMB angular power spectrum is expected at low multipoles, which is due to the Sachs–Wolfe effect, namely, to the fact that $\ell(\ell+1)C_\ell \approx \texttt{const}$. Spherical harmonics change as $Y_{\ell m}(\hat{\mathbf{n}}) = (-1)^\ell Y_{\ell m}(-\hat{\mathbf{n}})$ when the coordinates are reversed. Therefore, an asymmetry in the angular power spectrum for even and odd harmonics can be regarded as the asymmetry of the power of even and odd components of map. The authors Kim and Naselsky (2010) found the power of odd multipoles to systematically exceed the power of even multipoles of low ℓ and termed this phenomenon 'parity asymmetry'. To describe such an asymmetry quantitatively, the following quantities are proposed for consideration:

$$P^+ = \sum_{\text{Even}\,\ell<\ell_{\max}} \ell(\ell+1)C_\ell/2\pi$$

$$P^- = \sum_{\text{Odd}\,\ell<\ell_{\max}} \ell(\ell+1)C_\ell/2\pi\ .$$

Using the data of WMAP power spectrum and the results of Monte Carlo simulations, the authors Kim and Naselsky (2010) calculated the ratio P^+/P^- for the multipole ranges $2 \le \ell \le \ell_{\max}$, where $\ell_{\max}$ lies between 3 and 23. Comparing P^+/P^- for the WMAP data with the similar ratio obtained for simulated maps allows estimating the quantity p equal to the fraction of simulated spectra in which P^+/P^- is less than or equal to the same quantity for the WMAP map. The value of p was found to reach its lower boundary at $\ell_{\max} = 18$, where p equals 0.004 and 0.001 for the data obtained by the WMAP mission during 5 and 3 years of observations, respectively. This fact means that there is a preference for odd multipoles $2 \le \ell \le 18$ in the WMAP data at a confidence level of 99.6 % with a screening mask imposed on the data, and of 99.76 % without any mask. The authors believe the low amplitude of the WMAP CMB quadrupole may be part of the same anomaly as the parity asymmetry. Because the power asymmetry of the CMB signal in the northern and southern hemispheres is manifested more strongly in the case of multipoles with $2 \le \ell \le 19$ than multipoles with $20 \le \ell \le 40$, the authors also believe that the general origin of anomalies (such as the power asymmetry in the hemispheres, the low quadrupole amplitude, and the parity asymmetry) lie in the

region of small ℓ and that the explanation can be either cosmological or related to the presence of systematic errors in observations that were not revealed and/or were additionally introduced in the course of analysis of the data obtained by the WMAP mission.

12.4 Hemispherical Asymmetry

The asymmetry of hemispheres (see e.g. Fig. 30) was detected just after publishing the first year all sky maps of the WMAP (Eriksen et al. 2004c). Then, in Hansen et al. (2004), some calculations based on the angular power spectrum were presented and shown that this spectrum, when estimated locally at different positions on the sphere, appears not to be isotropic. Park (2004) also presented evidence for the existence of such hemispherical asymmetry, in which a particular statistical measure is considered to change discontinuously between two hemispheres on the sky, applying Minkowski functionals to the WMAP data. Since the preferred direction according to Eriksen et al. (2004c) stays close to the ecliptic plane, it was also demonstrated that the large-angular scale N-point correlation functions were different in behaviour when computed on ecliptic hemispheres.

Several studies were focused on the hemisphere difference and its connection with the ecliptic coordinate system (Schwarz et al. 2004; Verkhodanov et al. 2009; Verkhodanov and Khabibullina 2010). Hemispherical asymmetry was also detected with other measures of non-Gaussianity (Eriksen et al. 2004b, 2005; Räth et al. 2007). The Planck team repeated the analysis (Eriksen et al. 2005) on the Planck component separated data. As was shown in Planck Collaboration (2014f), the probabilities of obtaining a value for the χ^2 the Planck fiducial ΛCDM model are different both for North and South. For example, using SMICA map, one obtains 0.932 for Northern ecliptic and 0.592 for Southern ecliptic hemispheres. Thus, the observed properties of the Planck data are consistent with a remarkable lack of power in a direction towards the north ecliptic pole, consistent with the simpler one-point statistics (Planck Collaboration 2014f).

12.5 Difference of the WMAP and Planck Angular Power Spectra

One of the main anomalies first detected in the Planck data was the lack of power at low multipoles detected for angular power spectrum $C(\ell)$. Using the WMAP and Planck officially published spectra, we can compare them via the calculation the difference of maps including only the harmonics with maximum $C(\ell)$ difference (Fig. 35).

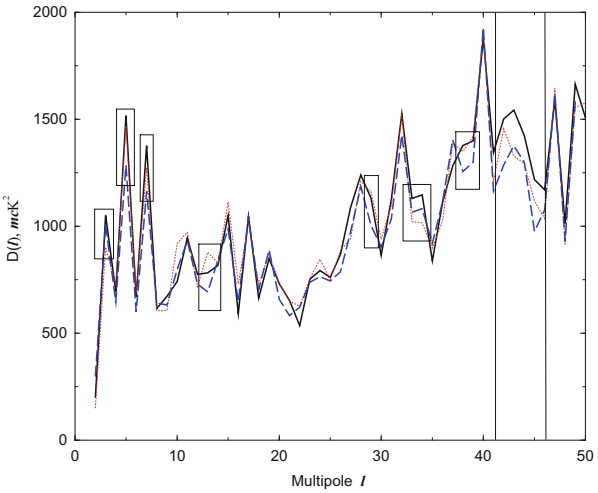

Fig. 35 The angular power spectrum $D(\ell) = \ell(\ell+1)C_\ell/2\pi$ in the multipole $2 \leq \ell < 50$. The *solid line* shows WMAP ILC data of the 7th year of observations. The *dotted line* marks WMAP9 ILC data. The Planck data are marked by the *dashed line*. The *rectangles* show the most different angular moments of distributions. The *vertical lines* mark limits of the multipole range in $\ell \in [41; 46]$

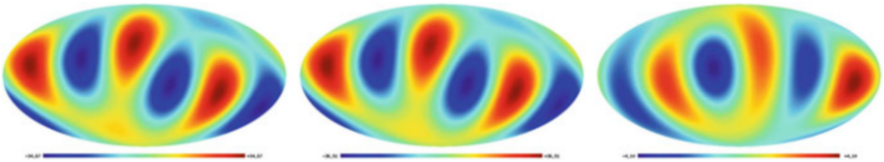

Fig. 36 *Left to right*: the octupole ($\ell = 3$) of the Planck CMB map SMICA, the octupole of the ILC WMAP9 map, and the map of these signals difference

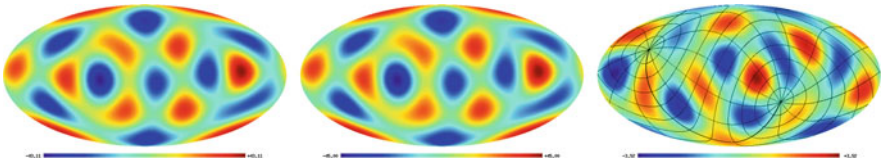

Fig. 37 *Left to right*: the map of $\ell = 5$ of the Planck CMB map SMICA, the $\ell = 5$ of the ILC WMAP9 map, and the map of these signals difference. The equatorial coordinate grid is overlaid on the map of difference

Following Verkhodanov (2014), let us consider the differences of maps corresponding to the harmonics having the maximum difference of power. These ranges are marked by rectangles on Fig. 35. The vertical lines demonstrate limits of the multipole range in $\ell \in [41; 46]$. On Figs. 36, 37, and 38, there are shown maps of harmonic differences at $\ell = 3$, $\ell = 5$ and $\ell = 7$, respectively. Some features of

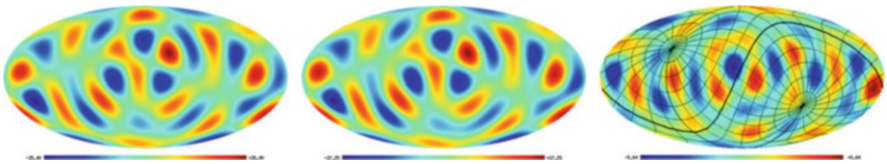

Fig. 38 *Left to right*: the map of $\ell = 7$ of the Planck CMB map SMICA, the $\ell = 7$ of the ILC WMAP9 map, and the map of these signals difference. The ecliptic coordinate grid is overlaid on the map of difference

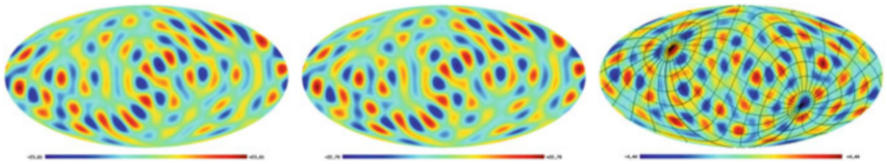

Fig. 39 *Left to right*: the map of $\ell = 13$ of the Planck CMB map SMICA, the $\ell = 13$ of the ILC WMAP9 map, and the map of these signals difference. The ecliptic coordinate grid is overlaid on the map of difference

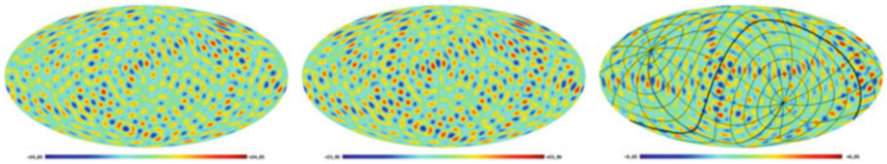

Fig. 40 *Left to right*: the map of $\ell = 29$ of the Planck CMB map SMICA, the $\ell = 29$ of the ILC WMAP9 map, and the map of these signals difference. The equatorial coordinate grid is overlaid on the map of difference

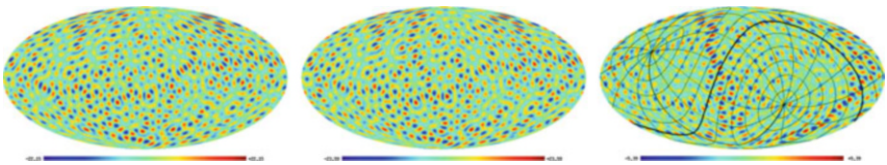

Fig. 41 *Left to right*: the map of $\ell = 37$ of the Planck CMB map SMICA, the $\ell = 37$ of the ILC WMAP9 map, and the map of these signals difference. The equatorial coordinate grid is overlaid on the map of difference

these differences show the position of spots along the Galactic plane, sensitivity of difference map at $\ell = 5$ to the equatorial coordinate system (equatorial poles are placed in singular points—saddles), and the axis of the multipole $\ell = 7$ lies on the Galactic plane and simultaneously, the saddle points of $\ell = 7$ are placed in ecliptic poles.

At the scales less 20°, there are three harmonics $\ell = 13$ (Fig. 39), $\ell = 29$ (Fig. 40), $\ell = 37$ (Fig. 41) which have the maximum difference in power. Following

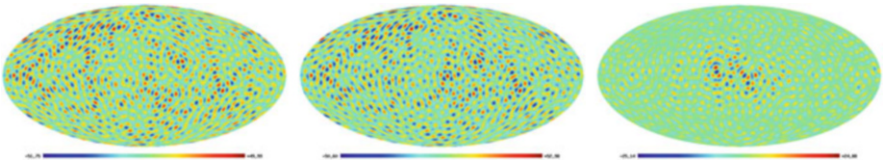

Fig. 42 *Left to right*: the summarized signal of multipole $\ell \in [41; 46]$ for the Planck CMB map SMICA, harmonics $\ell = 41 \div 46$ of WMAP9 ILC, and map of difference of these signals

the difference of signal at the selected multipole maps, we can find the features close to the detected ones earlier.

The map of multipole difference at $\ell = 13$ (angular size of $\sim 6.5°$) contains a feature similar to the harmonic $\ell = 7$ where the ecliptic poles are placed in singular points—local map minima and maxima. The multipole difference at the scales $\ell = 29$ ($\sim 3°$) and $\ell = 37$ ($\sim 2.5°$) contains a similar structure of spots placement. One line drawn by the very contrast spots formed with m-modes combinations of the $\ell = 29$ and $\ell = 37$ coincides with the ecliptic plane. Curiously, that a structure of the bright spots placement for $\ell = 29$ and $\ell = 37$ in the right hemisphere corresponds the anisotropic model $\text{Bianchi}_{\text{VIIh}}$ discussed in Planck Collaboration (2014f). There is the range of multipoles ($\ell \in [41; 46]$) where the spectrum strongly differ for the WMAP and Planck data (Figs. 35 and 42). The map difference for these multipole range shows the extended structure near the Galactic center.

Note, that there are two important moments observed in multipole differences. First, all the maps of multipole difference with high amplitude contain features tied with galactic, ecliptic or/and equatorial (terrestrial) coordinate systems. Second, there is the $\Delta\ell = 8$ period for multipoles numbers having a big difference in amplitude. Peculiar harmonics have numbers $\ell = 5, 13, 29, 37, 45$.

13 Summary

We have considered some radio astronomical fundamentals and problems of radio astronomical observations. Then, we have discussed the main observational cosmological tests which are investigated with radio astronomy, including radio galaxies, radio pulsar and cosmic microwave background radiation. Of course, the most crucial tests among them are connected with the CMB. Tens of radio telescopes operated to study the CMB signal and to measure its fluctuations. We have marked several radio telescopes for CMB study and discussed their basic results.

Also, some stages of the CMB data analysis pipeline are considered. We also described problems of data processing which can distort statistical results of the restored CMB maps. Examples of observational CMB anomalies have been considered.

Staying on the discussions of the CMB anomalies, we can see from their details that most of them manifest the properties sensitive to local environment. Three main environments of the cosmic observatory are displayed in the CMB signal distribution. They are our Galaxy—Milky Way, the Solar (ecliptic) system and some features connected with the equatorial system. The Galaxy is a source of the non-Gaussian residuals visible in positions of CMB spots (see discussion in Verkhodanov 2012). The Cold Spot is a special feature visible in synchrotron emission and on a map of the distribution of the Faraday rotation depth. Thus, it is probably connected with an ionized cloud from Galaxy or its vicinity.

The Solar system objects are considered as a source of additional residuals on the CMB map which are difficult to account in the standard component separation methods. Possible sources of residual signal are the satellite antenna far side-lobes sensitive to the Sun and bright planets, solar wind focusing by the Earth magnetosphere and passing through the Lagrange point L2, the objects at boundary of Solar system like the Kuiper belt.

The equatorial system features detected in some CMB correlation maps or in the single harmonic maps can be due by the influence of the Earth microwave emission also through the antenna back lobes or possible Solar wind emission modulated by the Earth magnetosphere where the magnetic axis is close the Earth rotation axis.

All these explanations may shed light on the origin of CMB low multipoles anomalies. The mismatch of the WMAP and Planck data, from one hand, and BICEP2 (BICEP2 Collaboration 2014) results of B-mode polarization, from another hand, connected with an amplitude of angular power spectrum also can be considered as a problem of low multipoles. For example, authors of Liu et al. (2014) demonstrated that effect of the charged dust emission connected with the Galaxy synchrotron loops can manifest some anomalies at low polarization harmonics and, thus, should be taken into account when microwave components of a signal are separated. This result was later confirmed by the Planck analysis of the cold magnetic dust (Planck Collaboration xxxx).

Also, it is necessary to note that there are some anomalies in the Planck data detected at high ($\ell > 600$) harmonics. There is a disagreement between cosmological parameters determination using the CMB angular power spectrum (including or not other experiments) and using only the Sunyaev-Zeldovich clusters (Planck Collaboration 2014e). Such a disaccordance, as discussed also in this contribution, can be explained by the biased estimates cluster parameters with the X-ray data.

Another reason of anomalies is some kinds of systematics in data analysis. There is a comparatively small difference (about 1.1σ) between cosmological parameters determined with the WMAP and Planck data (Planck Collaboration 2014b). Authors of the paper (Spergel et al. 2015) discussed these differences and repeated the component separation procedure using Planck multifrequency observational maps. They detected that in case of excluding the Planck 217 GHz channel from the data processing, the cosmological parameters measured with the WMAP9 data are restored by the Planck observations with the high accuracy ($< 1.1\sigma$). The authors

Spergel et al. (2015) conclude that there exist some problems in the calibration of this channel data.

Thus, we can say that

1. WMAP and Planck data have practically the same low multipole anomalies;
2. all the visible anomalies can be understood in the frame of the local (galactic and ecliptic) sources of microwave emission;
3. The difference of WMAP and Planck power spectra looks like one due to systematic effects of maps preparation;
4. Planck data are comparatively good (in resolution and sensitivity) when we take into account strangeness, e.g. some "bad" multipoles and 217 GHz data;
5. We are waiting for a new release in the second half of 2014: maps of temperature anisotropy and polarization.

At the end of the course (Appendix 3), the short application of the GLESP package is presented for simulation of the CMB map.

Acknowledgements The author is thankful to the organizers of the II JPB Cosmo School for hospitality and invitation to this nice event. The author is also grateful to NASA for possibility to use NASA Legacy Archive where WMAP maps stored and ESA for open access to observational results in Planck Legacy Archive. The GLESP[7] package (Doroshkevich et al. 2011) was used to process the CMB maps on a sphere. The explorations in this topic was supported by the RFBR grant No 13-02–00027.

Appendix 1: Normalized Associated Legendre Polynomials

In the GLESP code, we use the normalized associated Legendre polynomials f_ℓ^m:

$$f_\ell^m(x) = \sqrt{\frac{2\ell+1}{2}\frac{(\ell-m)!}{(\ell+m)!}}P_\ell^m(x)\,, \tag{29}$$

where $x = \cos\theta$, and θ is the polar angle. These polynomials, $f_\ell^m(x)$, can be calculated using two well known recurrent relations. The first of them gives $f_\ell^m(x)$ for a given m and all $\ell > m$:

$$f_\ell^m(x) = x\sqrt{\frac{4\ell^2-1}{\ell^2-m^2}}f_{\ell-1}^m - \\ - \sqrt{\frac{2\ell+1}{2\ell-3}\frac{(\ell-1)^2-m^2}{\ell^2-m^2}}f_{\ell-2}^m\,. \tag{30}$$

[7] http://www.glesp.nbi.dk.

This relation starts with

$$f_m^m(x) = \frac{(-1)^m}{\sqrt{2}} \sqrt{\frac{(2m+1)!!}{(2m-1)!!}} (1-x^2)^{m/2},$$

$$f_{m+1}^m = x\sqrt{2m+3} f_m^m .$$

The second recurrent relation gives $f_\ell^m(x)$ for a given ℓ and all $m \le l$:

$$\sqrt{(\ell-m-1)(\ell+m+2)} f_\ell^{m+2}(x) + \frac{2x(m+1)}{\sqrt{1-x^2}} f_\ell^{m+1}(x) + \sqrt{(\ell-m)(\ell+m+1)} f_\ell^m(x) = 0 . \quad (31)$$

This relation is started with the same $f_\ell^\ell(x)$ and $f_\ell^0(x)$ which must be found with (30).

As is discussed in Press et al. (1992, Sect. 5.5), the first recurrence relation (30) is formally unstable if the number of iteration tends to infinity. Unfortunately, there are no theoretical recommendations about the maximum iteration one can use in the quasi-stability area. However, it can be used because we are interested in the so-called *dominant* solution (Press et al. 1992, Sect. 5.5), which is approximately stable. The second recurrence relation (31) is stable for all ℓ and m.

Appendix 2: The GLESP Package

Structure of the GLESP Code

The code is developed in two levels of organization. The first one, which unifies F77 FORTRAN and C functions, subroutines and wrappers for C routines to be used for FORTRAN calls, consists of the main procedures: '*signal*', which transforms given values of $a_{\ell m}$ to a map, '*alm*', which transforms a map to $a_{\ell m}$, '*cl2alm*', which creates a sample of $a_{\ell m}$ coefficients for a given C_ℓ and '*alm2cl*', which calculates C_ℓ for $a_{\ell m}$. Procedures for code testing, parameters control, Kolmogorov-Smirnov analysis for Gaussianity of $a_{\ell m}$ and homogeneity of phase distribution, and others, are also included. Operation of these routines is based on a block of procedures calculating the Gauss–Legendre pixelization for a given resolution parameter, transformation of angles to pixel numbers and back.

The second level of the package contains the programs which are convenient for the utilization of the first level routines. In addition to the straight use of the already mentioned four main procedures, they also provide means to calculate map patterns generated by the Y_{20}, Y_{21} and Y_{22} spherical functions, to compare two sets of $a_{\ell m}$

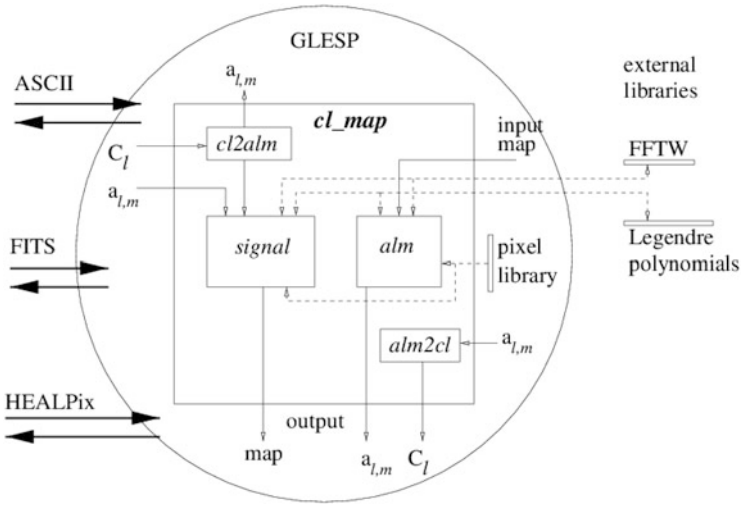

Fig. 43 Structure of the GLESP package

coefficients, to convert a GLESP map to a HEALPix map, to convert a HEALPix map, or other maps, to a GLESP map.

Figure 43 outlines the GLESP package. The circle defines the zone of the GLESP influence based on the pixelization library. It can include several subroutines and operating programs. The basic program *'cl2map'* of the second level, shown as a big rectangle, interacts with the first level subroutines. These subroutines are shown by small rectangles and call external libraries for the Fourier transform and Legendre polynomial calculations. The package reads and writes data both in ASCII table and FITS formats. More than ten programs of the GLESP package operate in the GLESP zone.

The package satisfies the following principles:

- Each program is designed to be easily joined with other modules of a package. It operates both with a given file and standard output.
- Each program can operate separately.
- Each program is accessible in a command string with external parameters. It has a dialogue mode and could be tuned with a resource file in some cases.
- Output format of resulting data is organized in the standard way and is prepared in the FITS format or ASCII table accessible for other packages.
- The package programs can interact with other FADPS procedures and CATS database (http://cats.sao.ru).

Main Operations

There are four types of operations accessible in the GLESP package:

- Operations related to maps:
  1. Spherical harmonic decomposition of a temperature anisotropy map into $a_{\ell m}$ (*cl2map*).
  2. Spherical harmonic decomposition of a temperature anisotropy and Q,U-polarization maps into $a_{\ell m}$ and $e, b_{\ell m}$-coefficients (*polalm*).
  3. Smooth a map with a Gaussian beam (*cl2map*).
  4. Sum/difference/averaging between maps (*difmap*).
  5. Scalar multiplication/division (*difmap*).
  6. Map rotation (*difmap*).
  7. Conversion from Galactic to equatorial coordinates (*difmap*).
  8. Cut temperature values in a map (*mapcut*).
  9. Cut a zone in/from a map (*mapcut*).
  10. Cut out cross-sections from a map (*mapcut*).
  11. Produce simple patterns (*mappat*).
  12. Read ASCII into binary (*mappat*).
  13. Read point sources to binary map (*mappat*).
  14. Print values in map (*mapcut*).
  15. Find min/max values in map sample per pixel (*difmap*).
  16. Simple statistic on a map (*difmap*).
  17. Correlation coefficients of two maps (*difmap*).
  18. Pixel size on a map (*ntot*).
  19. Plot figures (*f2fig*).
- Operations related to $a_{\ell m}$:
  1. Synthesize the temperature anisotropy map from given $a_{\ell m}$ (*cl2map*).
  2. Synthesize the temperature anisotropy and Q,U-polarization maps from given $a_{\ell m}$ and $e, b_{\ell m}$-coefficients (*polmap*).
  3. Sum/difference (*difalm*).
  4. Scalar multiplication/division (*difalm*).
  5. Vector multiplication/division (*difalm*).
  6. Add phase to all harmonics (*difalm*).
  7. Map rotation in harmonics (*difalm*).
  8. Cut out given mode of harmonics (*difalm*).
  9. Calculate angular power spectrum C_ℓ (*alm2dl*).
  10. Calculate phases (*alm2dl*).
  11. Select the harmonics with a given phase (*alm2dl*).
  12. Compare two $a_{\ell m}$ samples (*checkalm*).
  13. Produce $a_{\ell m}$ of map derivatives (*dalm*).

- Operations related to angular power spectrum C_ℓ :
  1. Calculate power spectrum C_ℓ (*alm2dl*).
  2. Simulate a map by a given C_ℓ (*cl2map*).
  3. Simulate $a_{\ell m}$ by C_ℓ (*createalm*).
- Operations related to phases $\phi_{\ell m}$ and amplitudes $|a_{\ell m}|$:
  1. Calculate phases $\phi_{\ell m}$ (*alm2dl*).
  2. Calculate amplitudes $|a_{\ell m}|$ (*alm2dl*).
  3. Simulate $a_{\ell m}$ by phases (*createalm*).
  4. Select harmonics with a given phase (*alm2dl*).
  5. Add a phase to all harmonics (*difalm*).

Main Programs

The following procedures organized as separate programs in the pixel and harmonics domain are realized now:

alm2dl calculates spectra and phases by $a_{\ell m}$-coefficients.
checkalm compares different $a_{\ell m}$-samples.
cmap converts HEALPix format maps to the GLESP package format.
cl2map converts a map to $a_{\ell m}$-coefficients and $a_{\ell m}$-coefficients to a map, simulates a map by a given C_ℓ-spectrum.
createalm creates $a_{\ell m}$-coefficients by phases, amplitudes or/and C_ℓ-spectrum.
dalm calculates the first and second derivatives by $a_{\ell m}$-coefficients
difalm calculates arithmetic operations over $a_{\ell m}$-samples.
difmap calculates arithmetic operations over maps, produces coordinates transformations.
f2fig produces color pictures in GIF-images.
f2map converts a GLESP map to a HEALPix format map.
mapcut cuts amplitude and coordinates in a GLESP map, produces one-dimensional cross-sections.
mappat produces standard map patterns, reads ASCII data to produce a map, reads point sources position from ASCII files.
polalm converts temperature and Q,U-polarization anisotropy maps to $a_{\ell m}$ and $e, b_{\ell m}$-coefficients
polmap converts $a_{\ell m}$ and $e, b_{\ell m}$-coefficients tp temperature and Q,U-polarization anisotropy maps

Data Format

The GLESP data are represented in two formats describing $a_{\ell m}$-coefficients and maps.

$a_{\ell m}$-coefficients data contains index describing number of ℓ and m modes corresponding to the HEALPix, real and imaginary parts of $a_{\ell m}$. These three parameters are described by three-fields records of the Binary Table FITS format.

Map data are described by the three-fields Binary Table FITS format containing a vector of $x_i = \cos\theta$ positions, a vector of numbers of pixels per each layer N_{ϕ_i}, and set of temperature values in each pixel recorded by layers from the North Pole.

Appendix 3: Practical Work "Study of Power Spectrum"

Task

1. Construct map CMB within the ΛCDM.
2. Generate card template with radio and its smooth Gaussian radiation pattern.
3. Find the power spectrum of the sum of these cards.

Necessary Resources

Packet data analysis of background radiation on the sphere GLESP[8] (Doroshkevich et al. 2003), Library of calculating the fast Fourier transform FFTW[9] Version 3.2 not earlier, the Internet, a computer running OS Linux (or any type of Unix).

Running Time 2 h.

Description

Highlighting the CMB, we can, on the one hand, build its power spectrum and its shape estimate cosmological parameters, and, on the other hand, research the statistics of the fluctuations. CMB power spectrum shows how much energy at any angular scales contained in the incoming radiation. For the full sphere, it is determined by as the average value of the squares of harmonic modes

$$C(\ell) = \frac{1}{2\ell + 1}\left[|a_{\ell 0}|^2 + 2\sum_{m=1}^{\ell}|a_{\ell,m}|^2\right].$$

[8] http://www.glesp.nbi.dk.

[9] http://www.fftw.org.

This expression contains double harmonic mode $m > 0$, which is explained by complex conjugation of harmonics and, consequently, the equality of the squares their amplitudes. Note that in the calculation of the power spectrum, we explicitly apply the hypothesis of a random Gaussian distribution primary perturbation amplitudes which is reflected in the distribution of amplitudes of the harmonics of the CMB. This hypothesis is used in the procedure of averaging all modes m for a given multipole ℓ. Variation of harmonic mode amplitudes for a given ℓ occurs inside the confidence interval defined as the cosmic variance.

The power spectrum of $C(\ell)$ reflects the physical conditions in the early Universe and thus is a function of the relevant cosmological parameters (Naselsky et al. 2006),

$$C(\ell) \equiv C_\ell(h, \Omega_b h^2, \Omega_{CDM} h^2, \Omega_\Lambda, \Omega_\nu, n, \ldots) .$$

Here, in particular, it has been indicated the Hubble constant $h = H_0/100\,\mathrm{km/s/Mpc}$, the density of baryonic matter Ω_b, hidden mass density Ω_{CDM}, the density of the "dark energy" Ω_Λ, the density of massive neutrinos Ω_ν, spectral index of adiabatic perturbations n and other parameters. The resulting current values of the main parameters are: Hubble constant $h = 0.674\ pm{+}0.014$, matter density $\Omega_m = 0.314 \pm 0.020$, baryon density $100\Omega_b h^2 = 2.207 \pm 0.033$, age of the universe $t_0 = 13.813 \pm 0.058\,\mathrm{Gyr}$, spectral index $n_s = 0.9616 \pm 0.0094$ at scale $k = 0.05\,\mathrm{Mpc}^{-1}$ (Planck Collaboration 2014b), established in 2013 by the Planck collaboration.

The solution of the functional associated with the fitting cosmological model to the observational data, currently almost automated and executes a program *CAMB* (Lewis et al. 2000) *CMBFast*, (Seljak and Zaldarriaga 1996), receiving from its entrance cosmological parameters and outputs the result in the form of a smooth power spectrum of the microwave background radiation (Fig. 14).

In this practical work on data analysis of CMB the following stages:

1. Simulation maps of the CMB model for the Universe ΛCDM.
2. Adding a model map radio sources with different flux density.
3. Smoothing card directivity pattern size selected in the field of spatial harmonics.
4. Calculation of the power spectrum.

Procedures in GLESP

The first step is to generate a map for a given power spectrum. Angular power spectrum can be calculated using on-line program *CAMB* at

http://lambda.gsfc.nasa.gov/toolbox/tb_camb_form.cfm.

For further analysis and modeling can be used FITS-file table containing encoded ASCII, or pre-calculated file available online

http://sed.sao.ru/~vo/cosmo_school/presentations/vo/LCDM.dat,

which recorded a two-column table with the symbolic numbers of multipoles and the corresponding values power spectrum in the model of the Universe ΛCDM (Ω_Λ=0.693, Ω_{CDM}=0.257, Ω_b=0.0481) (Planck Collaboration 2014b).

To simulate the map, one can use the program '*cl2map*' of the GLESP package, which is run with the following parameters:

cl2map -Dl LCDM.dat -r 6 -lmax -nx 1000 2001 4002 -np -ao alm.fts -o map.fts

In this command, the format contains the following flags: '*-Dl*' indicates that the next parameter '*LCDM.dat*' is the name of the file containing the power spectrum; the flag '*-r*' indicates that there will be generated a random Gaussian noise corresponding to the random Gaussian fields of initial density perturbations, with some starting seed, set the following parameters string (for example, here is 6); the flag '*-lmax 1000*' indicates that the maximum multipole for the generated map is $\ell_{\max}$ =1000; '*-nx 2001 -np 4002*' are the resolution keys: '*-nx*' and '*-np*' set the pixelization grid determining the number of rings and the number of pixels in a equatorial ring respectively (the number of rings must always be less than $n_x = 2\ell_{\max}$, and the number of pixels in the equatorial ring is $2n_x$, to comply with the Nyquist theorem and make the pixels quasisquare); the flag '*-ao*' indicates that the following parameter '*alm.fts*' is the name of output file containing generated spherical harmonic coefficients a_{ellm} in the form of FITS-file; the flag '*-o*' indicates that the following parameter '*map.fts*' is the file of the generated output map of CMB temperature anisotropy.

The map can be visualized by placing in the GIF-image and displayed, for example, by using the program '*xv*':

f2fig map.fts -o map.gif; xv map.gif

To see the coefficients a_{ellm}, one can move them from a binary representation in ASCII-table using the '*alm2dl*':

alm2dl -g alm.fts > alm.dat; less alm.dat

In the second phase of work, one should create a file a list of radio sources, which then should be applied to the sky. The file contains information written in the ASCII-format:

```
hh:mm:ss1 dd:mm:ss1 amplitude1
hh:mm:ss2 dd:mm:ss2 amplitude2
hh:mm:ss3 dd:mm:ss3 amplitude3,
```

wherein the first and second columns show the equatorial coordinates of radio sources, and the third contains the flux density in mJy. Lists of real sources can be obtained using a database CATS.[10] For example, one can construct a sample of radio sources from the NVSS survey (Condon et al. 1998) (Fig. 44), conducted on radio interferometer VLA (USA).

To add sources to the CMB map, one should first generate a source map with the following command:

mappat -fp src.dat -o src.fts -nx 2001 -np 4002

[10] http://cats.sao.ru.

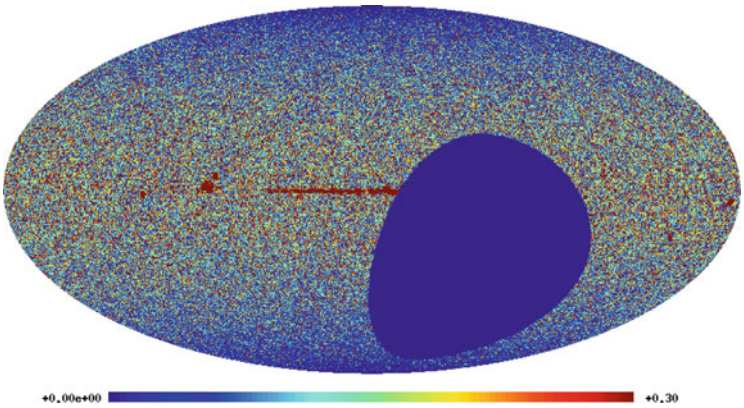

Fig. 44 Map of radio sources in the survey NVSS (Condon et al. 1998)

Resolution of both maps (the number of rings and pixels in the equatorial ring) must be identical. The flag '*–fp*' suggests that the next parameter '*src.dat*' is the file with a list of radio sources. The flag '*-o*' is used to enter the name of the output file '*src.fts*'.

Further, both maps: with CMB and sources can be stacked with the command '*difmap*':

difmap -sum src.fts map.fts -o map_src.fts .

In the third stage of the workshop, must be decomposed into spherical harmonics new map '*map_src.fts*' and smooth it Gauss diagram orientation. To calculate the expansion in spherical harmonics applying the used earlier procedure '*cl2map*', but with a new flag:

cl2map -map map_src.fts -lmax 1000 -ao alm_src.fts > /dev/null

Here, the flag '*–map*' indicates that the map is used to enter '*map_src.fts*' for the harmonic analysis.

Smoothing is done with the procedure '*rsalm*':

rsalm alm_src.fits -fw 10 -o alm_sm.fits

Here, the glag '*-fw 10*' suggests that the smoothing is done with a Gaussian beam pattern of $10'$. The resulting harmonics are written into the FITS-file '*alm_sm.fits*'.

Using the program '*alm2dl*', one can calculate the power spectrum of $C(\ell)$:

alm2dl -lmax 1000 alm_sm.fits -cl > cl.dat,

which is visualized as a two-column ASCII-table, for example, using the '*xmgr*'. Maps '*alm_sm.fits*' can be constructed from harmonics using the procedure '*cl2map*':

cl2map -lmax -nx 1000 2001 4002 -np -ai alm_sm.fts -o map_sm.fts

f2fig map_sm.fts -o map_sm.gif

References

Ackerman, L., Carroll, S.M., Wise, M.B.: Phys. Rev. D **75**, 083502 (2007). [Erratum-ibid. D **80**, 069901 (2009)]. astro-ph/0701357

Albrecht, A., Steinhardt, P.J.: Phys. Rev. Lett. **48**, 1220 (1982)

Allen, T.J., Grinstein, B., Wise, M.B.: Phys. Lett. B **197**, 66 (1987)

Bardeen, J.M., Steinhardt, P.J., Turner, M.S.: Phys. Rev. D **28**, 679 (1983)

Barnaby, N., Cline, J.M.: Phys. Rev. D **75**, 6004 (2007). astro-ph/0611750

Bennett, C.L., Halpern, M., Hinshaw, G., et al.: Astrophys. J. Suppl. **148**, 1 (2003). astro-ph/0302207

Bennett, C.L., Hill, R.S., Hinshaw, G., et al.: Astrophys. J. Suppl. **192**, 17 (2011). arXiv:1001.4758

Bennett, C.L., Larson, D., Weiland, J.L., et al.: Astrophys. J. Suppl. **208**, 20 (2013). arXiv:1212.5225

Bernardeau, F., Uzan, J.-P.: Phys. Rev. D **66**, 103506 (2002)

BICEP2 Collaboration, Ade, P.A.R., Aikin, R.W., Barkats, D., et al.: Phys. Rev. Lett. **112**, 241101 (2014). arXiv:1403.3985

Brown, M.L., et al.: Astrophys. J. **705**, 978 (2009)

Chiang, L.-Y., Naselsky, P.D., Verkhodanov, O.V., Way, M.J.: Astrophys. J. **590**, L65 (2003). astro-ph/0303643

Coles, P., Dineen, P., Earl, J., Wright, D.: Mon. Not. R. Astron. Soc. **350**, 989 (2004). astro-ph/0310252

Condon, J.J., Cotton, W.D., Greisen, E.W., et al.: Astronom. J. **115**, 1693 (1998)

Copi, C.J., Huterer, D., Schwarz, D.J., Starkman, G.D.: Mon. Not. R. Astron. Soc. **367**, 79 (2006). astro-ph/0508047

Copi, C.J., Huterer, D., Schwarz, D.J., Starkman, G.D.: Mon. Not. R. Astron. Soc. **399**, 295 (2009)

Copi, C.J., Huterer, D., Schwarz, D.J., Starkman, G.D.: (2013) arXiv:1311.4562

Creminelli, P., et al.: J. Cosmol. Astropart. Phys. **0605**, 004 (2006). astro-ph/0509029

Crittenden, R.G., Turok, N.G.: Exactly Azimuthal Pixelizations of the Sky, Report-no: DAMTP-1998-78 (1998). astro-ph/9806374

Cruz, M., Martinez-Gonzalez, E., Vielva, P., Cayon, L.: Mon. Not. R. Astron. Soc. **356**, 29 (2005)

Cruz, M., et al.: Science **318**, 1612 (2007). arXiv:0710.5737

Das, S., Louis, T., Nolta, M.R., et al.: J. Cosmol. Astropart. Phys. **04**, 014 (2014). arXiv:1301.1037

de Bernardis, P., et al.: Astrophys. J. **564**, 559 (2002)

de Oliveira-Costa, A., Smoot, G.F., Starobinsky, A.A.: Astrophys. J. **468**, 457 (1996). astro-ph/9705125

Demiański, M., Doroshkevich, A.: Phys. Rev. D **75**, 123517 (2007)

Dicke, R.H., et al.: Astrophys. J. **142**, 414 (1965)

Dimopoulos, K., Karciauskas, M., Wagstaff, J.M.: Phys. Rev. D **81**, 023522 (2010). arXiv:hep-ph/0907.1838

Dineen, P., Rocha, G., Coles, P.: Mon. Not. R. Astron. Soc. **358**, 1285 (2005). astro-ph/0404356

Doroshkevich, A.G., Novikov, I.D.: Dokl. Akad. Nuak USSR **154**, 809 (1964)

Doroshkevich, A.G., Verkhodanov, O.V.: Phys. Rev. D **83**, 3002 (2011). arxiv:1008.4094

Doroshkevich, A.G., Zeldovich, Ya.B., Syunyaev, R.A.: In: Longair, M.S., Einasto, J. (eds.) The Large Scale Structure of the Universe. Proc. Symp., Tallinn, Estonian SSR, p. 393. D. Reidel Publ. Co., Dordrecht (1977)

Doroshkevich, A.G., Naselsky, P.D., Verkhodanov, O.V., et al.: Int. J. Mod. Phys. **14**, 275 (2003). astro-ph/0305537

Doroshkevich, A.G., Verkhodanov, O.V., Naselsky, O.P., et al.: Int. J. Mod. Phys. **20**, 1053 (2011). arXiv:0904.2517

Driscoll, J.R., Healy, D.M.: Adv. Appl. Math. **15**, 202 (1994)

Dulaney, T.R., Gresham, M.I.: Phys. Rev. D **81**, 103532 (2010). arXiv:1001.2301

Durrer, R.: N. Astron. Rev. **43**, 111 (1999)

Durrer, R., Kahniashvili, T., Yates, A.: Phys. Rev. D **58**, 123004 (1998). astro-ph/9807089

Dvali, G., Gruzinov, A., Zaldarriaga, M.: Phys. Rev. D **69**, 023505 (2004). arXiv:astro-ph/0303591
Eriksen, H.K., et al.: Astrophys. J. **605**, 14 (2004a)
Eriksen, H.K., Novikov, D.I., Lilje, P.B., Banday, A.J., Górski, K.M.: Astrophys. J. **612**, 64 (2004b)
Eriksen, H.K., Hansen, F.K., Banday, A.J., et al.: Astrophys. J. **605**, 14 (2004c)
Eriksen, H.K., Banday, A.J., Górski, K.M., Lilje, P.B.: Astrophys. J. **622**, 58 (2005)
Gold, B., et al.: Astrophys. J. Suppl. **192**, 15 (2011). arXiv:1001.4555
Górski, K.M., Hivon, E., Wandelt, B.D.: Evolution of Large–Scale Structure: From Recombination to Garching. PrintPartners Ipskamp, Enschede (1999)
Górski, K.M., Hivon, E., Banday, A.J., Wandelt, B.D., et al.: Astrophys. J. **622**, 759 (2005)
Greisen, E.W., Calabretta, M.: Bull. Am. Astron. Soc. **182**, 09.01 (1993)
Gruppuso, A., Burigana, C.: J. Cosmol. Astropart. Phys. **08**, 004 (2009)
Gumrukcuoglu, A.E., Contaldi, C.R., Peloso, M.: astro-ph/0608405
Gumrukcuoglu, A.E., Contaldi, C.R., Peloso, M.: J. Cosmol. Astropart. Phys. **0711**, 005 (2007). arXiv:0707.4179
Gumrukcuoglu, A.E., Himmetoglu, B., Peloso, M.: Phys. Rev. D **81**, 063528 (2010). astro-ph.CO/1001.4088
Guth, A.: Phys. Rev. D **23**, 347 (1981)
Guth, A.H., Pi, S.Y.: Phys. Rev. Lett. **49**, 1110 (1982)
Hansen, F.K., Banday, A.J., Górski, K.M.: Mon. Not. R. Astron. Soc. **354**, 641 (2004)
Hansen, M., Zhao, W., Frejsel, A.M., et al.: Mon. Not. R. Astron. Soc. **426**, 57 (2012). arXiv:1202.1711
Haslam, C.G.T., Salter, C.J., Stoffel, H., Wilson, W.E.: Astron. Astrophys. **47**, 1 (1982)
Hawking, S.F.: Phys. Lett. B **115**, 295 (1982)
Hinshaw, G., Spergel, D.N., Verde, L., et al.: Astrophys. J. Suppl. **170**, 288 (2007). astro-ph/0603451
Hinshaw, G., Weiland, J.L., Hill , R.S., et al.: Astrophys. J. Suppl. **180**, 225 (2009). astro-ph/0803.073
Inoue, K.T., Tomita, K., Sugiyama, N.: Mon. Not. R. Astron. Soc. **314**, L21 (2000). astro-ph/9906304
Jaffe, T., et al.: Astrophys. J. **629**, L1 (2005). astro-ph/0503213
Jaffe, T.R., Banday, A.J., Eriksen, H.K., et al.: Astron. Astrophys. **460**, 393 (2006)
Jarosik, N., Bennett, C.L., Dunkley, J., et al.: Astrophys. J. Suppl. **192**, 14 (2011). arXiv:1001.4744
Kahniashvili, T., Ratra, B.: Phys. Rev. D **71**, 103006 (2005). astro-ph/0503709
Kahniashvili, T., Lavrelashvili, G., Ratra, B.: Phys. Rev. D **78**, 063012 (2008). arXiv:0807.4239
Kim, J., Naselsky, P.: Astrophys. J. **714**, L265 (2010)
Kim, J., Naselsky, P., Christensen, P.R.: Phys. Rev. D **79**, 023003 (2009)
Kofman, L.: Report # CITA-03-12 (2003). arXiv:astro-ph/0303614
Koivisto, T., Mota, D.F.: J. Cosmol. Astropart. Phys. **06**, 018 (2008)
Komatsu, E., Spergel, D.N.: Phys. Rev. D **63**, 063002 (2001). astro-ph/0005036
Komatsu, E., et.al.: Astrophys. J. Suppl. **192**, 18 (2011a)
Komatsu, E., et.al.: Astrophys. J. Suppl. **192**, 18 (2011b). arXiv:1001.4538
Konjaev, S.I.: Mat. Zametki **25**, 629 (1979)
Kotelnikov, V.A.: Problemy pomehoustoichivoj radisvyazi [Problems of interference-free radio links (in Russian)]. Gosenegoizdat, Moscow (1947)
Kovac, J.M., et al.: Nature **420**, 772 (2002)
Kunszt, P.Z., et al.: Physics and astronomy. In: Mining the Sky. ESO Astrophys. Symp., vol. 631. Springer, Berlin/Heidelberg (2001)
Kunz, M., Aghanim, N., Riazuelo, A., Forni, O.: Phys. Rev. D **77**, 023525 (2008)
Land, K., Magueijo, J.: Phys. Rev. L **95**, 071301 (2005). astro-ph/0502237
Lewis, A., Challinor, A., Lasenby, A.: Astrophys. J. **538**, 473 (2000). astro-ph/9911177
Linde, A.D.: Phys. Lett. **108B**, 389 (1982)
Linde, A., Mukhanov, V.: Phys. Rev. D **56**, 535 (1997)
Liu, H., Mertsch, P., Sarkar, S.: Astrophys. J. **789**, L29 (2014). arXiv:1404.1899

Longair, M.S.: Radio astronomy and cosmology. In: Observational Cosmology, Advanced Course, 8th, Saas-Fee, Switzerland, April 10–15, 1978, Lectures, p. 127. Geneva Observatory, Sauverny
Lyth, D.H., Wands, D.: Phys. Lett. B **524**, 5 (2002). arXiv:hep-ph/0110002
Mack, A., Kahniashvili, T., Kosowsky, A.: Phys. Rev. D **65**, 123004 (2002). astro-ph/0105504
McKellar, A., Kan-Mitchell, J., Conti, P.S.: Publ. Dominion Astrophys. Obs. (Victoria, BC) **7**(6), 251 (1941)
Moroi, T., Takahashi, T.: Phys. Lett. B **522**, 215 (2001). [Erratum-ibid. B **539**, 303 (2002)]. arXiv:hep-ph/0110096
Muciaccia, P.F., Natoli, P., Vittorio, N.: Astrophys. J. **488**, L63 (1997)
Mukhanov, V.F., Chibisov, G.V.: J. Exp. Theor. Phys. Lett. **33**, 532 (1981)
Mysovskikh, I.P.: In: Devore, R.A., Scherer, K. (eds.) Quantitative Approximation. Academic, New York (1976)
Naselsky, P.D., Verkhodanov, O.V.: Astrophys. Bull. **62**, 203 (2007)
Naselsky, P.D., Doroshkevich, A.G., Verkhodanov, O.V.: Astrophys. J. **599**, L53 (2003). astro-ph/0310542
Naselsky, P.D., Doroshkevich, A.G., Verkhodanov, O.V.: Mon. Not. R. Astron. Soc. **349**, 695 (2004a). astro-ph/0310601
Naselsky, P.D., Chiang, L.-Y., Olesen, P., Verkhodanov, O.V.: Astrophys. J. **615**, 45 (2004b). astro-ph/0405181
Naselsky, P.D., Novikov, D.I., Novikov, I.D.: The Physics of the Cosmic Microwave Background. Cambridge Astrophysics Series. Cambridge University Press, New York (2006)
Naselsky, P.D., Verkhodanov, O.V., Christensen, P.R., Chiang, L.-Y.: Astrophys. Bull. **62**, 285 (2007). astro-ph/0211093
Naselsky, P.D., Verkhodanov, O.V., Nielsen, M.T.B.: Astrophys. Bull. **63**, 216 (2008). arXiv:0707.1484
Naselsky, P.D., Christensen, P.R., Coles, P., et al.: Astrophys. Bull. **65**, 101 (2010). arXiv:0712.1118
Natoli, P., de Gasperis, G., Gheller, C., Vittorio, N.: Astron. Astrophys. **372**, 346 (2001). astro-ph/0101252
Park, C.-G.: Mon. Not. R. Astron. Soc. **349**, 313 (2004)
Park, C.-G., Park, C., Gott III, J.R.: Astrophys. J. **660**, 959 (2007). astro-ph/0608129
Peebles, P.J.E.: Principles of Physical Cosmology. Princeton Series in Physics. Princeton University Press, Princeton (1993)
Penzias A.A., Wilson R.W.: Astrophys. J. **142**, 419 (1965)
Planck Collaboration, Ade, P.A.R., et al.: Astron. Astrophys. **571**, A1 (2014a). arXiv:1303.5062
Planck Collaboration, Ade, P.A.R., Aghanim, N., Armitage-Caplan, C., et al.: Astron. Astrophys **571A**, 16 (2014b). arXiv:1303.5076
Planck Collaboration, Ade, P.A.R., et al.: Astron. Astrophys. **571**, 12 (2014c). arXiv:1303.5072
Planck Collaboration, Ade, P.A.R., Aghanim, N., Armitage-Caplan, C., et al.: Astron. Astrophys. **571A**, 24 (2014d). arXiv:1303.5084
Planck Collaboration, Ade, P.A.R., Aghanim, N., Armitage-Caplan, C., et al.: Astron. Astrophys. **571A**, 20 (2014e). arXiv:1303.5080
Planck Collaboration, Ade, P.A.R., Aghanim, N., Armitage-Caplan, C., et al.: Astron. Astrophys. **571A**, 23 (2014f). arXiv:1303.5083
Planck Collaboration, Ade, P.A.R., Aghanim, N., Armitage-Caplan, C., et al.: arXiv:1409.6728.
Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in FORTRAN, 2nd edn. Cambridge University Press, Cambridge (1992). http://www.nr.com
Räth, C., Schuecker, P., Banday, A.J.: Mon. Not. R. Astron. Soc. **380**, 466 (2007)
Readhead, A.C.S., et al.: Science **306**, 836 (2004)
Reichardt, C.L., et al.: Astrophys. J. **694**, 1200 (2009)
Rudnick, L., Brown, S., Williams, L.R.: Astrophys. J. **671**, 40 (2007). arXiv:0704.0908
Sato, K.: Mon. Not. R. Astron. Soc. **195**, 467 (1981)
Schaffer, K.K., Crawford, T.M., Aird, K.A., et al.: Astrophys. J. **743**, 90 (2011). arXiv:1111:7245
Schwarz, D.J., Starkman, G.D., Huterer, D., Copi, C.J.: Phys. Rev. Lett. **93**, 221301 (2004)

Seljak, U., Zaldarriaga, M.: Astrophys. J. **469**, 437 (1996)
Senatore, L., Smith, K.M., Zaldarriaga, M.: J. Cosmol. Astropart. Phys. **1001**, 028 (2010). arXiv:0905.3746
Shamonov, T.A.: Pribory and Tekhnika Eperimenta **1**, 83 (1957)
Slosar, A., et al.: J. Cosmol. Astropart. Phys. **8**, 31 (2008)
Smoot, G.F., et al.: Astrophys. J. Lett. **396**, L1 (1992)
Sobolev, S.L.: Introduction to the Theory of Cubature Formulae. NAUKA, Moscow (1974)
Spergel, D., Flauger, R., Hlozek, R.: Phys. Rev. D **91**(2), 023518 (2015). arXiv:1312.3313
Starobinsky, A.A.: ZhETF Pisma v Redaktsiiu **30**, 719 (1979)
Starobinsky, A.A.: Phys. Lett. B **117**, 175 (1982)
Starobinsky, A.A.: J. Exp. Theor. Phys. Lett. **57**, 622 (1993). arXiv:gr-qc/9305019
Stolyarov, V., Hobson, M.P., Ashdown, M.A.J., Lasenby, A.N.: Mon. Not. R. Astron. Soc. **336**, 97 (2002)
Stroud, A.H.: Approximate Calculation of Multiple Integrals. Prentice-Hall, Englewood Cliffs (1971)
Tauber, J.A., et al.: Astron. Astrophys. **520A**, 1 (2010)
Tegmark, M.: Astrophys. J. **470**, L81 (1996)
Tegmark, M., de Oliveira-Costa, A., Hamilton, A.: Phys. Rev. D **68**, 123523 (2003)
Turok, N.: Phys. Scripta T **36**, 135 (1991)
Turok, N., Spergel, D.: Phys. Rev. Lett. **64**, 2736 (1990)
Verkhodanov, O.V.: Phys. Usp. **55**, 1098 (2012)
Verkhodanov, O.V., Doroshkevich, A.G.: The sky pixelization for CMB mapping, Chapt. 8. In: Way, M., Scargle, J., Ali, K., Srivastava, A. (eds.) Advances in Machine Learning and Data Mining for Astronomy, p. 133. Chapman and Hall/CRC Press, Boca Raton (2012)
Verkhodanov, O.V.: Astrophys. Bull. **69**, 330 (2014)
Verkhodanov, O.V., Khabibullina, M.L., Majorova, E.K.: Astrophys. Bull. **64**, 263 (2009). arXiv:0912.3073
Verkhodanov, O.V., Khabibullina, M.L.: Astrophys. Bull. **65**, 390 (2010)
Watanabe, M.A., Kanno, S., Soda, J.: Phys. Rev. Lett. **102**, 191302 (2009). arXiv:hep-th/0902.2833
Watanabe, M.A., Kanno, S., Soda, J.: Prog. Theor. Phys. **123**, 1041 (2010)
Whittaker, E.T.: Proc. R. Soc. Edinb. **35**, 181 (1915)

Cosmic Microwave Background Observations

Rolando Dünner

Abstract The observations of the CMB.

1 Introduction

Cosmology has intrigued humanity since very early times. Questions like "How did the Universe begin?", "How old is it?" or "How big is it?" are natural questions that accompanied society and religion well before modern Science had a say in our understanding of nature. Despite all this interest, and despite the effort of the greatest minds of their times, it was not until the sixteenth century that science finally proved that it was the Earth that went around the Sun, demolishing many centuries of dogmatic thinking that placed the Earth at the center of the Universe. Among others, there was a key factor that made this possible: the invention of the telescope by Galileo drastically expanded the horizons of astronomical observation, producing the required experimental evidence to clear all doubts around such fundamental question. This event marked an inflection point in human Cosmology, followed by a series of great discoveries, always related to new technologies, allowing for finer and deeper observations. It was like this that four centuries later (1927), Edwin Hubble, observing from the world's largest telescope at Mount Wilson, made one of the most amazing discoveries by showing that the Universe is not steady but expanding, giving a new twist to the newly developed theory of gravity developed by Einstein (General Relativity) and providing the base for the revolutionary ideas that gave birth to the Big Bang Cosmology. Wait another 35 years, and the advances in electronics and telecommunications made possible another astonishing discovery, when two engineers, Arno Penzias and Robert W. Wilson, while testing a new transatlantic communication system, fortuitously detected the afterglow of the Big Bang, or Cosmic Microwave Background (CMB), settling the idea that the Universe experienced a hot and dense era before expanding into its current state. These examples prove how scientific development is tightly linked to the incorporation of

R. Dünner (✉)
Instituto de Astrofísica, Facultad de Física, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile
e-mail: rdunner@astro.puc.cl

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_5

new technologies to improve the observation of nature, providing new experimental evidence to support theoretical ideas.

The study of the Cosmic Microwave Background has revolutionized Cosmology, as it provides a snapshot of the early state of the Universe with detailed information about its composition, structure and physical state. This signal, the oldest in the form of electromagnetic radiation, reaches us with as a 2.76 K blackbody spectrum, carrying most of the relevant information as tiny fluctuations on the order of tenths of micro-Kelvin in temperature and a few micro-Kelvin in polarization. Moreover, the signal is extended on the sky, requiring large high-fidelity maps to extract the relevant information. Measuring this signal is a huge technical challenge which has only been achieved recently thanks to the development of high sensitivity detectors and readout systems, cryogenics, well understood reflective and refractive optical designs and powerful computer systems able to analyze huge volumes of data.

These pages summarize four lectures given at the II JPBCosmo School (2014) about the state of the art CMB Observation, centered around the case of the Atacama Cosmology telescope, and including a description of the instrument, observing techniques, data reduction and analysis, and new prospects for observing the CMB polarization signal.

2 Precision Cosmology

Historically speaking, Cosmology was always considered a very speculative and uncertain science, characterized by brave hypotheses and little evidence. The term "Precision Cosmology" appeared after CMB observations brought in fantastic experimental evidence, allowing us to strongly constrain our cosmological models (to better than 1 % in precision), producing this change of paradigm.

After the Penzias and Wilson's discovery in 1964 (Penzias and Wilson 1965), another 30 years had to pass until two instruments on board of the COBE satellite unambiguously measured the spectrum and anisotropies of the CMB for the first time (Smoot et al. 1992; Mather et al. 1994). COBE's discovery was followed by a rush of ground-based and balloon experiments (for examples see de Bernardis et al. 1999; Hanany et al. 2000; Devlin et al. 1999; Carlstrom & DASI Collaboration 2000) which systematically contributed to the development of new technologies and techniques required to dig into the rich information frozen on the CMB signal. Most of these experiments were intended to measure the tiny angular anisotropies of the CMB, requiring increasing sensitivities and angular resolutions. In 2001 the WMAP satellite was launched by NASA, later producing the most precise measurement of the CMB. After 9 years of observations "the allowed volume of cosmological parameters was reduced by a factor in excess of 68,000" http://map.gsfc.nasa.gov, 2013 justifying the concept of precision cosmology.

WMAP produced a full map of the sky with an angular resolution of 0.2°, meaning multipoles of less than $\ell \approx 500$, which is enough to measure the first three peaks of the CMB power spectrum, and with enough sensitivity to determine those

multipoles to their cosmic variance limit. Further exploration of the CMB required mapping finer resolutions and measuring the polarization of the signal, which was only done to a more limited degree by WMAP. In 2009 the European Space Agency launched Planck (http://www.esa.int/Our_Activities/Space_Science/Planck, 2013), another space observatory with improved WMAP sensitivity and resolution, which reached 5 arcmin, measuring multipoles well down the diffusion damping tail of the CMB power spectrum up to $\ell \approx 2000$. Planck greatly confirmed and improved WMAP results, further constraining cosmological parameters and characterizing the microwave signal across the sky, and continues to release results today. One of the most expected results from Planck is the polarization maps and analysis, which may possibly shed light on the very early epochs of the inflationary universe.

The spacial resolution of CMB observations from space is limited by the size of the optics that can be put on a satellite. This means that finer angular resolutions are better achieved from the ground. This brings another set of great challenges, especially because the atmosphere strongly interacts with millimeter wavelengths, mostly due to its water vapor content. This forces observations to be done in sites where the atmosphere is extremely dry and thin, as can be found in Antarctica or at the highlands of the Atacama Desert in Chile. This complication is balanced by the possibility of implementing large arrays of high-sensitivity detectors at a much lower cost. In 2007, the South Pole Telescope (SPT http://pole.uchicago.edu, 2013) and the Atacama Cosmology Telescope (ACT http://www.princeton.edu/act, 2014), with 10 and 6 m main apertures, began observing from the South Pole and from the Atacama Desert respectively. Their higher resolution, combined with an extreme sensitivity provided by their kilo-pixel class cameras, allowed these experiments to produce super-fine maps of the CMB over thousands of square degrees in the sky. With this finer resolution (order 1 arcmin), these experiments measured the CMB power spectrum all the way down the diffusion tail and into the secondary anisotropy region, dominated by the cosmic infrared background, thermal and kinetic Sunyaev-Zel'dovich clustering signals and radio sources, dramatically expanding the available science to later epochs of the Universe, while improving the "standard" CMB science by measuring and decoupling this foreground signal from the CMB.

Figure 1 compares the maps from COBE, WMAP, Planck and ACT, stressing the effect of increasing the resolution of the map.

In the following we will describe technical aspects of the Atacama Cosmology Telescope, as reference for the technologies and techniques required to observe the CMB from the ground.

3 The Atacama Cosmology Telescope

The Atacama Cosmology Telescope is a 6 m telescope installed at 5200 m on Cerro Toco, a few kilometers away from the ALMA site in the Chajnantor valley. It is a dedicated instrument to measure the CMB at millimeter wavelengths over large

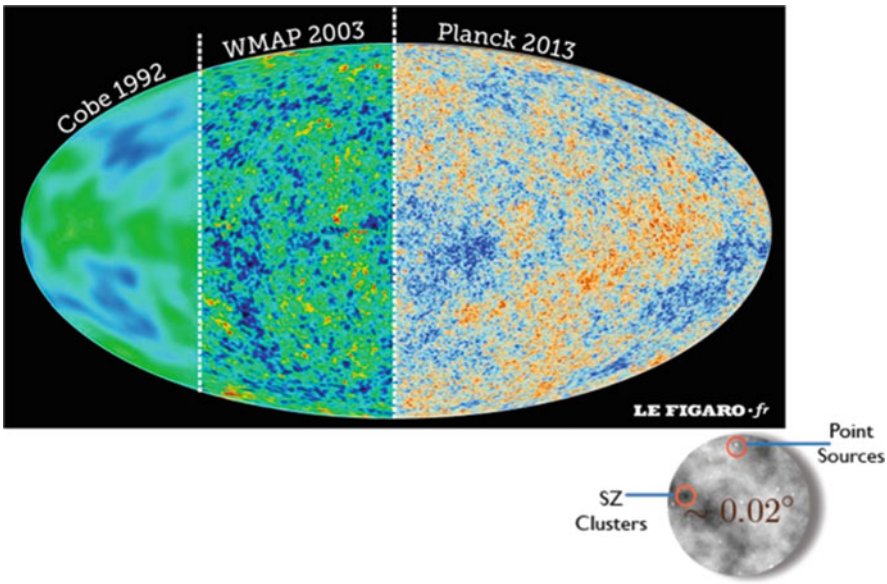

Fig. 1 Direct comparison of the resolution achieved by COBE, WMAP, Planck and ACT maps. The Increase in resolution is evident when moving *from left to right*. The finer resolution of ACT (*lower right circle*) clearly detects foreground galaxies and clusters of galaxies through their thermal Sunyaev-Zel'dovich effect signature, as indicated

areas of the sky. It started operation in 2007 with MBAC, a 3-K TES pixel bolometer camera observing in three bands (150, 220 and 280 GHz) with unprecedented sensitivity (Swetz et al. 2011). At these frequencies, the resolution reaches less than 1 arcmin, permitting the direct detection of a large number of extragalactic point sources. Moreover, the three bands are chose to probe the decrement, null and increment frequencies of the Sunyaev-Zel'dovich (SZ) effect, clearly distinguishing clusters of galaxies from the background, producing a redshift-independent blind survey of these objects only limited by their mass.

MBAC observed for four seasons before being decommissioned in 2011. During those years it mapped nearly 2000 square degrees over two long stripes of the sky. The scientific outcome was a long list of publications (for main results and references see Niemack et al. 2008; Menanteau et al. 2010, 2013; Sherwin et al. 2011; Hand et al. 2012; Dünner et al. 2013; Sifón et al. 2013; Calabrese et al. 2013; Sievers et al. 2013; Hasselfield et al. 2013a,b; Dunkley et al. 2013; Sehgal et al. 2013; Das et al. 2014; Louis et al. 2014; Marsden et al. 2014) ranging from new cosmological parameter constraints, through detection of dozens of galaxy clusters and extragalactic sources, first detection of kinetic SZ effect, first detection of gravitational lensing effects on the CMB, first detection of Dark Energy effects purely using CMB data, to the new observing and data processing techniques required to achieve these results.

At the relevant frequencies, radiation strongly interacts with rotational and vibrational water modes, thus observing the CMB requires very dry atmosphere conditions, measured as Precipitable Water Vapor (PWV). Normal PWV values around de Earth range from 2 to 50 mm, while the median PWV at the ACT site is only 0.5 mm, justifying such an extreme location for the telescope.

To measure temperature fluctuations of only a few micro-Kelvins we need extremely sensitive detectors, which imply having extremely low noise levels. As you may know, any electronic element produces noise proportional to its temperature. Then to measure such a low temperature radiation the detectors must also operate at very low temperature. In the case of MBAC, the detectors operated at only 0.3 K, together with most of the superconducting readout electronics. These temperatures are achieved using cryogenics systems based in He^4 and He^3, which is a Helium isotope, operating in a very similar way as a fridge does. For these the whole refractive optics and detectors are kept under vacuum inside the camera, which is constructed as Matryoshka doll, with layers at decreasing temperatures up to the lowest is reached at the center where the detectors are.

The detectors in MBAC were Transition Edge Sensors (TES), which are superconducting devices that are operated in the transition between being normal and super conductors. This means that tiny changes in temperature of the device produce large changes in resistivity, which is measured by running a current through it. So in practice they are thermometers. When radiation falls on one of these detectors it warms it up, so we can detect very tiny changes in radiation loading. As they do not discriminate between the wavelength of the incoming radiation, but only on the temperature achieved by the device, these detectors are called bolometers.

An important property of TES's is that they can be micro-fabricated using techniques similar to those developed in the electronics industry, so they can be cheaply stacked together in large arrays, increasing this way the telescope sensitivity. For instance MBAC had three cameras, each one containing 1024 detectors.

The optical design of the telescope and camera are very important to achieve a good result. As this is an extended source, and considering that at these wavelengths the images produced are limited by diffraction (as opposed to optical telescopes where the image resolution is normally limited by the turbulences in the atmosphere, or seeing), the shape of the features in the maps produced is directly affected by the properties of the optics. Moreover, at such low temperatures, stray radiation from the surrounding landscape can enter the telescope producing ghost images, motivating the installation of large baffles around the telescope to direct all the stray light to the sky. To reduce diffraction features, the optical design is off-axis, clearing up the optical path. To achieve a large focal plane, necessary to accommodate many detectors, the camera contains refractive optics composed of filters, lenses and a Lyot-stop to limit the illumination of the primary mirror and minimize spillover. These lenses are very similar so optical lenses, but their materials are very different because then operate at these long wavelengths.

The observations are performed by continuously scanning the sky at constant elevation while the Earth rotation slowly moves the sky across the field of view.

This is important to keep the airmass constant, as the water vapor is the main source of radiation as mentioned before. The same patch of the sky is observed while rising and setting, producing a cross-linked observation patterns which is important to reconstruct modes in all directions during map making. Contrarily to optical telescopes, which integrate the radiation falling in their detectors while staring at a single place in the sky, ACT is continuously reading its detectors while scanning, producing long time-streams of data occupying many terabytes of disk space. We call these time-streams Time Ordered Data or TOD. All this data must then be reduced offline to produce the image of the sky.

The raw data from the telescope is very noisy. Most of its power comes from the atmospheric turbulence, which is clearly predominant at low frequencies in the time-streams, forming strong correlations between detectors (a strong common mode). At higher frequencies the TOD is dominated by detectors noise, which is mostly Gaussian noise. The CMB signal is then buried under the noise in a single TOD and can only be recovered after combining many TODs.

The map making procedure consists in estimating the best possible map of the sky given all the data from the time-streams. This implies performing a likelihood minimization given the best noise model we can produce. The map-making equation can be written as

$$\mathbf{M}^T \mathbf{N}^{-1} \mathbf{M} x = \mathbf{M}^T \mathbf{N}^{-1} d \tag{1}$$

where M is a pointing matrix that relates every sample from the TODs with a pixel in the map, **N** is the noise covariance matrix from the TODs, d is the time-space data samples and x is the map. Given the huge amount of data used to produce the map, it is impossible to express and invert this equation in its full form, so iterative methods are used instead. These methods only used a block diagonal version of the noise matrix, grouping TODs obtained simultaneously by different detectors, while modeling the frequency dependence by computing the noise matrix divided in frequency bins. The whole process can only be done in a reasonable time using a supercomputer with thousands of cores.

4 Polarization Sensitive Maps (ACTpol)

MBAC was decommissioned in 2011 and used to build a new polarization sensitive camera called ACTpol. This new camera also used TES detectors, but operating at only 0.1 K using a dilution refrigerator, device that uses the latent heat in a mixture of He^3 and He^4 to reach these super cold temperatures. Also several improvements were introduced to the optics and to the detector coupling, which was now done using a wafer feed-horn array. The polarization was achieved by coupling each feed-horn to two orthogonal microstrip antennas, each one coupled to its own TES detectors, such that each detector is coupled to a single polarization. Antenna-pairs

were distributed across the array in several directions to evenly cover Q and U polarizations on the sky.

One advantage of measuring polarization is that the atmosphere is not polarized, significantly reducing the level of noise in the data. On the other hand, the polarized signal is about an order of magnitude dimmer than the temperature signal, increasing the experimental challenge to measure it correctly.

ACTpol began observing in 2013, mapping nearly 200 degrees on the sky. The polarized maps were used to measure the E-mode CMB polarization (Naess et al. 2014), although the reached sensitivity in a single year of observations was not enough to reach the B-mode polarization signal. ACTpol is expected to achieve the sensitivity to measure B-mode polarization after the next 2 years of observations, ending in early 2016.

5 What is Coming

As mentioned before, the field of CMB observations is developing very rapidly, with its prime focus towards reaching finer angular resolutions and high polarization sensitivities. Associated to these observational goals there is an ambitious set of scientific aims, like characterizing the lensing signal on the CMB to measure the distribution of matter at redshifts peaking at $z \approx 2$, and detecting primordial B-modes produced by gravitational waves formed during the epoch of inflation.

Inflation is one of the largest mysteries still to be resolved, as it implies understanding the very beginning of the Universe. Several experiments are now competing to be the first to detect this signal, some of them being BICEP, ABS, CLASS, Planck, SPTpol and AdvACT, which is the successor of ACTpol on the ACT telescope. The challenge is huge, as primordial B-modes are expected to be a tiny signal over degree scales on the sky, while being strongly affected by foreground signals like synchrotron and polarized dust from our galaxy. Decoupling this signals will require measuring the sky at several frequency bands and over large areas of the sky. Moreover, measuring large features on the sky require mapping large areas, favoring low latitude locations for ground based telescopes.

Wrapping up, measuring large areas of the sky require excellent atmospheric conditions, access to large areas of the sky and very fast telescopes. The latter implies increasing the sensitivity of the detectors, which is now reaching its optical limit. A way to overcome this is adding more and more detectors in the focal plane, and of course using optical designs with large focal planes. In my opinion, the implementation of these new techniques will require at least a few iterations of these lower cost ground experiments before justifying new satellite missions, while the seek for finer resolutions imply large telescopes which are also impractical in orbit. Then the field is probably going to point towards a rapid development of ground and balloon experiments, with a significant amount of technological development.

References

Calabrese, E., Hlozek, R.A., Battaglia, N., et al.: Phys. Rev. D **87**, 103012 (2013)
Carlstrom, J.E., DASI Collaboration: Bull. Am. Astron. Soc. **32**, 1496 (2000)
Das, S., Louis, T., Nolta, M.R., et al.: J. Cosmol. Astropart. Phys. **4**, 014 (2014)
de Bernardis, P., Ade, P.A.R., Artusa, R., et al.: New Astron. Rev. **43**, 289 (1999)
Devlin, M.J., Caldwell, R., Dorwart, W.B., et al.: Bull. Am. Astron. Soc. **31**, 1457 (1999)
Dunkley, J., Calabrese, E., Sievers, J., et al.: J. Cosmol. Astropart. Phys. **7**, 025 (2013)
Dünner, R., Hasselfield, M., Marriage, T.A., et al.: Astrophys. J. **762**, 10 (2013)
Hanany, S., Ade, P., Balbi, A., et al.: Astrophys. J. Lett. **545**, L5 (2000)
Hand, N., Addison, G. E., Aubourg, E., et al.: Phys. Rev. Lett. **109**, 041101 (2012)
Hasselfield, M., Moodley, K., Bond, J.R., et al.: Astrophys. J. Suppl. **209**, 17 (2013a)
Hasselfield, M., Hilton, M., Marriage, T.A., et al.: J. Cosmol. Astropart. Phys. **7**, 008 (2013b)
Louis, T., Addison, G.E., Hasselfield, M., et al.: J. Cosmol. Astropart. Phys. **7**, 016 (2014)
Marsden, D., Gralla, M., Marriage, T.A., et al.: Mon. Not. R. Astron. Soc. **439**, 1556 (2014)
Mather, J.C., Cheng, E.S., Cottingham, D.A., et al.: Astrophys. J. **420**, 439 (1994)
Menanteau, F., González, J., Juin, J.-B., et al.: Astrophys. J. **723**, 1523 (2010)
Menanteau, F., Sifón, C., Barrientos, L.F., et al.: Astrophys. J. **765**, 67 (2013)
Naess, S., Hasselfield, M., McMahon, J., et al.: J. Cosmol. Astropart. Phys. **10**, 007 (2014)
Niemack, M.D., Zhao, Y., Wollack, E., et al.: J. Low Temp. Phys. **151**, 690 (2008)
Penzias, A.A., Wilson, R.W.: Astrophys. J. **142**, 419 (1965)
Sehgal, N., Addison, G., Battaglia, N., et al.: Astrophys. J. **767**, 38 (2013)
Sherwin, B.D., Dunkley, J., Das, S., et al.: Phys. Rev. Lett. **107**, 021302 (2011)
Sievers, J.L., Hlozek, R.A., Nolta, M.R., et al.: J. Cosmol. Astropart. Phys. **10**, 060 (2013)
Sifón, C., Menanteau, F., Hasselfield, M., et al.: Astrophys. J. **772**, 25 (2013)
Smoot, G.F., Bennett, C.L., Kogut, A., et al.: Astrophys. J. Lett. **396**, L1 (1992)
Swetz, D.S., Ade, P.A.R., Amiri, M., et al.: Astrophys. J. Suppl. **194**, 41 (2011)

Part II
Seminars

Physics of Baryons

J.A. de Freitas Pacheco

Abstract In this lecture, different milestones in the cosmological history of baryons are reviewed. First, the appearance of hadrons as a consequence of the confinement of quarks is discussed. Then the era in which nuclei interact to produce light elements like deuterium, helium and lithium is described. The third relevant episode is the decoupling between matter-radiation and the properties of the "last scattering" surface. The next covered aspect refers to the value of the residual ionization fraction when "freezing" occurs and the thermal decoupling of matter from CMB photons. As stars appear in the Universe, their UV radiation begins to reionize the intergalactic medium and such a process is also discussed in this lecture. Finally, results from cosmological simulations are presented, permitting to describe where baryons can be found either in the form of stars or in the form of cold, warm and hot gas.

1 Introduction

In the matter-energy budget of the Universe, baryons contribute with only 4 % , the remaining 96 % representing components of unknown nature dubbed respectively dark matter and dark energy, which dominate the present dynamics of the Universe. Only in the early past the energy density under the form of radiation and relativistic particles surpassed that associated to the dark components and thus, controlled the cosmic expansion.

Even constituting only a small fraction of the matter-energy content of the Universe, baryons play a fundamental role in cosmology since they are able to emit electromagnetic radiation, and hence they can be used as tracers of the evolution of cosmic structures. Moreover, baryons form stars and planets, the pillars for the existence of life in the Universe.

In the present lecture different topics related to the physics of baryons will be discussed. The appearance of hadrons in the Universe, the primordial nucleosynthe-

J.A. de Freitas Pacheco (✉)
Université de Nice-Sophia Antipolis, Observatoire de la Côte d'Azur, F-06340 Nice Cedex 4, France
e-mail: pacheco@oca.eu

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_6

sis, the decoupling of baryons from photons and the consequences for the cosmic microwave background (CMB), as well as the present distribution of baryons, either in the form of stars or in the form of gas (cold and hot), derived from cosmological simulations.

2 The Appearance of Hadrons

Presently baryons are not elementary particles but complex structures constituted by quarks. In the very early Universe the baryonic charge was carried by the elementary constituents of matter, the quarks. These particles are normally confined, forming hadrons, but at the very high pressure and temperature prevailing in the primitive Universe, they are in a state of "asymptotic" freedom. As the Universe expands, the temperature (and the pressure) of the cosmic plasma decreases and the conditions for the existence of a state of "asymptotic" freedom are no more satisfied, leading to the confinement and, consequently, to the appearance of hadrons.

The confining property of quarks and gluons manifests itself in the long range behavior of the "heavy" quark potential. At zero temperature, the potential rises approximately linearly with the particle separation, i.e., $V_{qq}(r) \approx \sigma r$, where σ denotes the string tension. The resulting force obliges quarks and gluons to be confined into a hadronic bag. On the other side, chiral symmetry breaking leads to a non-vanishing quark anti-quark condensate in the vacuum. Inside the hadron bag, however, the condensate vanishes. At high temperatures the individual hadronic bags are expected to merge and quarks and gluons can move freely. This bag picture is closely related to percolation models for the QCD phase transition. It provides an intuitive argument for the occurrence of deconfinement and chiral symmetry restoration. Consequently, two distinct phase transitions are expected to occur—one associated to chiral symmetry restoration at the temperature T_χ, related to models like the SU(3) gauge theory, and another to the deconfinement at the temperature T_d. In general, $T_\chi < T_d$. In "pure" QCD, however, there seems to be only one transition from the low temperature hadronic regime to the high temperature quark-gluon plasma phase.

From a thermodynamical point of view, in a first approximation, the confinement can be treated as a first order phase transition. Under this assumption, the transition occurs when the chemical potential and the pressure of both phases (quark & hadrons) are equal.

The equation of state (EoS) for the quark-gluon plasma is still quite uncertain and here the following (simple and pedagogical) procedure will be adopted. Assuming that the pressure and the energy density of the quark-gluon plasma depend only on the temperature, the first law of thermodynamics can be written as

$$dS = \frac{V}{T}\left(\frac{d\varepsilon}{dT}\right)dT + \frac{(P+\varepsilon)}{T}dV \; . \tag{1}$$

In the equation above V is the volume of the system, T is the temperature, P is the pressure, ε is the energy density and S is the entropy. Since Eq. (1) is an exact differential, using the Schwarz condition the following equation can be obtained

$$T\frac{dP}{dT} - P = \varepsilon \ . \tag{2}$$

This differential equation can be integrated if the energy density as a function of the temperature is specified. Supposing that quarks and gluons in the asymptotic regime are fully relativistic, their energy density can be written as

$$\varepsilon = \frac{\pi^2}{30} g_{\text{eff}} \frac{(kT)^4}{(\hbar c)^3} + B \ , \tag{3}$$

where g_{eff} is the effective number of degrees of freedom, including the contribution of quarks and gluons, and the constant B was introduced phenomenologically to represent the bag constant or the confinement energy density in the context of the so-called MIT model (Shuryak 1980; Cleyman et al. 1986). In this case, the constant B is necessarily positive. However, as we shall see below, this may not be the case and, in this context it should be considered as a phenomenological (free) parameter measuring thermodynamic deviations from an ideal Stephan gas when the temperature approaches the critical transition value.

Replacing Eq. (3) into Eq. (2) and integrating, one obtains

$$P = \frac{\pi^2}{90} g_{\text{eff}} \frac{(kT)^4}{(\hbar c)^3} - B - A(kT) \ , \tag{4}$$

where A is an integration constant that has the dimension of an inverse volume.

Quantum chromodynamics (QCD) calculation in a lattice permits also an estimate of the EoS of the quark-gluon plasma and a rich literature on this subject is available (see, for instance, Karsch 1995, 2002; Brown et al. 1988; Khan et al. 2001 and references therein). EoS derived from the SU(3) gauge theory thermodynamics on lattices of various sizes were discussed in Boyd et al. (1995, 1996). These authors found a deviation from the ideal gas behavior of about 15–20 %, even at high temperatures. Indeed, at high temperatures one expects, according to Eq. (4), that the quantity P/T^4 approaches the Stephan-Boltzmann limit, i.e., $P/T^4 \rightarrow \pi^2 g_{\text{eff}}/90$. However, this is not the case according to these calculations. An explanation for this behavior was proposed by Begun et al. (2010) by considering the variation of the effective mass of the gluon with the temperature, leading to a dispersion relation of the form $\omega^2 = k^2 + m^2(T)$ and a factor less than the unity that multiplies the Stephan constant.

If the quantity $(\varepsilon - 3P)/T^4$ is plotted against the temperature normalized in terms of the critical temperature T_c corresponding to a first order transition in pure SU(3) gluodynamics, a peaked maximum is observed at $T_{max} = 1.1T_c$. Such a temperature

maximum is not predicted by Eq. (4) since

$$\frac{d}{dT}\left[\frac{(\varepsilon - 3P)}{T^4}\right] = -\frac{16B}{T^5} - \frac{9A}{T^4} \, . \tag{5}$$

The equation above indicates that a maximum exists only if B or A is negative.

Assuming that Eq. (4) represents adequately the EoS of the quark-gluon plasma, the effective degree of freedom and the constants A and B can be considered as free parameters, which can be evaluated by fitting QCD lattice data derived from a quark-gluon plasma including three flavors. Figure 1 shows the resulting fit (solid curve) and lattice data taken from Karsch (2002).

From the fitting procedure, it results that $\pi^2 g_{\rm eff}/90 = 4.25$ instead of 5.209 expected for a 3 flavor quark-gluon plasma, $B = -390.8\,{\rm MeV\,fm^{-3}}$ and $A = 4.24\,{\rm fm^{-3}}$. Note that a negative bag constant has resulted from the fitting procedure, a conclusion also reached by the authors in Begun et al. (2010). In this case, the parameter B should not be interpreted as the "bag" constant as mentioned before. Anyway, the usual fuzzy bag model is unable to explain the maximum of the relation $(\varepsilon - 3P)/T^4$ which, using the values of the parameters estimated by the fitting procedure occurs at $T_{max} = 164\,{\rm MeV}$, corresponding to a deconfinement temperature $T_d = 0.909 T_{\rm max} \approx 149\,{\rm MeV}$. Notice that Karsch (2002) obtained from his own calculations $T_d = 154 \pm 8\,{\rm MeV}$ and a similar value was derived by Bazarov et al. (2012).

The deconfinement temperature can be also estimated by comparing the pressure of the hadronic phase, supposed to be constituted by pions (bosons), which are

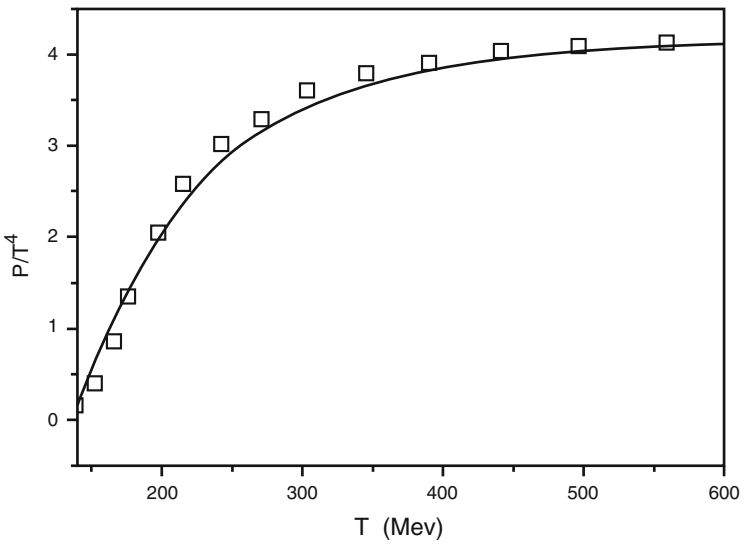

Fig. 1 Best fit of QCD lattice results for a 3 flavor quark-gluon plasma using the Eq. (4) given in the text

carriers of the strong force. The pressure due to pions, neglecting their self-interaction is

$$P_\pi = \frac{g_\pi}{2\pi^2}\frac{(kT)^4}{(\hbar c)^3} I(y) \,, \tag{6}$$

where we have introduced

$$I(y) = -\int_y^\infty x\sqrt{x^2 - y^2}\,\lg(1 - e^{-x})dx \,, \tag{7}$$

with $y = m_\pi c^2/kT$. In the numerical calculations, the number of degrees of freedom was taken as $g_\pi = 3$ (three pions: one neutral and two charged) and the mass m_π of all of them was taken to be equal to $140\,\mathrm{MeV}/c^2$. The transition temperature can be estimated from the equality of both pressures given respectively by Eqs. (4) and (6).

Numerical solution of the aforementioned equations gives for the transition temperature $T_d \approx 142\,\mathrm{MeV}$, a value consistent with the previous estimate. It is interesting to compare these numbers with estimates derived from heavy ion collisions, which indicate deconfinement temperatures in the range 150–180 MeV. The energy density of the quark-gluon plasma at this temperature is 0.44 GeV fm^{-3} also consistent with lattice calculations (see Karsch 2002), while at the hadron phase the energy density is only 2.4 MeV fm^{-3}

The evolution of the temperature of the cosmic plasma after inflation and reheating is given by

$$T \approx \frac{1.423}{\sqrt{t}}\,\mathrm{MeV} \,, \tag{8}$$

where the time is given in seconds. Replacing the temperature $T_d \sim 145\,\mathrm{MeV}$ derived for the quark-hadron phase transition in the equation above, one obtains that hadrons appear when the Universe was approximately 96 μs old.

Another interesting question to be answered is how long was the duration of the phase transition? The process begins with the appearance of small bubbles of hadrons in the cosmic plasma. The bubbles grow and coalesce until all quarks are confined. The process occurs in conditions where both the pressure and the temperature remain constant. Despite the constancy of the pressure and of the temperature, the total energy density varies during the phase transition since the energy density of the deconfined phase is higher than that of the hadron phase. The difference being the "latent heat" of the transition, i.e., $L = (\varepsilon_1 - \varepsilon_2)$, with $\varepsilon_1 \approx 0.68$ GeV fm^{-3} and $\varepsilon_2 \approx 0.25$ GeV fm^{-3} being respectively the total energy density of the quark and of the hadron phases at the transition point. Note that in order to get the total energy density (or total pressure) in both phases at the transition point it is necessary to add the contribution of photons and leptons. When added either to the quark-gluon pressure or to the pion pressure, one obtains for the total pressure during the transition the value of 0.096 GeV fm^{-3}.

If x is the fraction of matter in the quark phase, then the total energy density at any instant of the transition is $\varepsilon = x\varepsilon_1 + (1-x)\varepsilon_2$. The conservation of the energy is given by the equation $T^{\nu}{}_{0;\nu} = 0$ that can be explicitly written as

$$\frac{d\varepsilon}{dt} + 3H(P+\varepsilon) = 0 \, , \tag{9}$$

where H is the Hubble parameter whose evolution is controlled by the equation

$$H^2 = \frac{8\pi G}{3c^2}\varepsilon \, . \tag{10}$$

Equation (9) can be rewritten in terms of the variable x and, in this case one obtains

$$\frac{dx}{dt} + \frac{\sqrt{24\pi G}}{Lc}\left(P + \varepsilon_2 + xL\right)\sqrt{\varepsilon_2 + xL} \, . \tag{11}$$

This equation should be integrated with the conditions: at t_1, beginning of the phase transition, $x = 1$ and at t_2, end of the process, $x = 0$. Hence, the duration $\Delta t = (t_2 - t_1)$ of the phase transition is given by

$$\Delta t = \frac{2c}{\sqrt{24\pi GP}}\left[\arctan\left(\sqrt{\frac{\varepsilon_1}{P}}\right) - \arctan\left(\sqrt{\frac{\varepsilon_2}{P}}\right)\right] \, . \tag{12}$$

Using the values for the (constant) pressure during the transition and the energy densities at the beginning and at the end of the transition, one obtain for the duration of the process $\Delta t \approx 13\,\mu$s.

3 Primordial Nucleosynthesis

Just after the phase transition quarks are confined into protons and neutrons. The abundance ratio between these two species is fixed by the Boltzmann factor, i.e.,

$$\frac{n}{p} = \exp\left(-\frac{Q}{kT}\right) \, , \tag{13}$$

where $Q = 1.294\,$MeV is the mass difference between the neutron and the proton. With $T = 145\,$MeV, it results $n/p \approx 0.991$. Thus, just after the quark-hadron transition, the density of neutrons and protons is nearly the same but as the Universe expands and cools, the neutron-to-proton ratio decreases. Equation (13) fixes the relative abundances of neutrons and protons while they are in statistical equilibrium,

which is maintained by a series of weak interaction reactions as, for instance

$$\begin{aligned} n &\leftrightarrow p + e^- + \bar{\nu}_e \ , \\ e^+ + n &\leftrightarrow p + \bar{\nu}_e \ , \\ n + \nu_e &\leftrightarrow p + e^- \ , \end{aligned} \tag{14}$$

The statistical equilibrium is broken when the rate Γ_ν of the weak interactions that interconvert neutrons and protons becomes comparable to the expansion rate H^{-1} of the Universe. The cross-section for the reaction $n\nu_e \leftrightarrow pe^-$ is

$$\sigma(E_\nu) = \sigma_0 \left(\frac{E_\nu + Q}{m_e c^2} \right)^2 \ , \tag{15}$$

where m_e is the electron mass, Q as before is the mass difference between the neutron and the proton and

$$\sigma_0 = \frac{2\pi^2 \hbar^3}{f \tau_n m_e^3 c^4} = 2.61 \times 10^{-44} \, \mathrm{cm}^2 \ . \tag{16}$$

In order to derive the numerical value above, we have taken $f = 1.636$ for the phase space factor and $\tau_n = 887$ s for the neutron life time. In this case, the weak interaction rate per nucleon is obtained by integrating over the energy distribution of neutrinos, supposed to be still in statistical equilibrium

$$\Gamma_\nu = c \int_0^\infty \sigma(E_\nu) \frac{dn_\nu}{dE_\nu} dE_\nu \ . \tag{17}$$

Numerically, the solution of the equation above gives $\Gamma_\nu \approx 0.92 \ (T/\mathrm{MeV})^5$. On the other side, using Eq. (10) and $g_{\mathrm{eff}} = 10.75$, one obtains $H = 0.68 \ (T/\mathrm{MeV})^2$. Both quantities are comparable when $T \approx 1.1$ MeV, a temperature that corresponds to an age for the Universe of $t \approx 1.7$ s. At this moment, the neutron-to-proton ratio is $n/p \approx 0.308$.

After the decoupling of neutrinos, the expected abundance of neutrons and protons should be "frozen" but neutrons begin to decay and interact with protons to produce deuterium according to the reaction

$$n + p \leftrightarrow {}^2D + \gamma \ . \tag{18}$$

In this evolutionary phase, the density of neutrons varies as

$$\frac{dn_n}{dt} = -3Hn_n - \frac{n_n}{\tau_n} - n_n n_p < \sigma v >_D + n_D \psi_n \ . \tag{19}$$

In this equation, the first term on the right side represents the variation of the neutron density due to the expansion of the universe, the second represents the losses of neutrons due to their decay, the third represents the losses due to deuterium formation and the last one represents the production of neutrons due to the photodisintegration of deuterium. A similar equation can be written for the evolution of the proton density, namely

$$\frac{dn_p}{dt} = -3Hn_p + \frac{n_n}{\tau_n} - n_n n_p < \sigma v >_D + n_D \psi_n \, . \tag{20}$$

Introducing respectively the particle concentration of neutrons $X_n = n_n/n_b$, protons $X_p = n_p/n_b$ and deuterium nuclei $X_D = n_D/n_b$, where n_b is the total baryon density that obeys the conservation equation

$$\frac{dn_b}{dt} + 3Hn_b = 0 \, , \tag{21}$$

Eqs. (19) and (20) can be simplified as

$$\frac{dX_n}{dt} = -\frac{X_n}{\tau_n} - X_n X_p n_b < \sigma v >_D + X_D \psi_n \, , \tag{22}$$

$$\frac{dX_p}{dt} = \frac{X_n}{\tau_n} - X_n X_p n_b < \sigma v >_D + X_D \psi_n \, ,$$

If the reaction (18) occurs in a situation of quasi-equilibrium, we have the relation

$$\frac{X_n X_p}{X_D} = \frac{\psi_n}{n_b < \sigma v >_D} = \frac{2}{3} \left(\frac{mkT}{4\pi \hbar^2} \right)^{3/2} \, , \frac{e^{-B/kT}}{n_b} = F(T, n_b) \tag{23}$$

where $B = 2.225\,\mathrm{MeV}$ is the binding energy of a deuterium nucleus. In terms of the radiation temperature, the density of baryons can be written as

$$n_b = \frac{3H_0^2 \Omega_b}{8\pi G m} \frac{g_{\mathrm{eff}}(T)}{g_{\mathrm{eff}}(T_0)} \left(\frac{T}{T_0} \right)^3 = 8.66 \times 10^{23} (\Omega_b h^2) \frac{g_{\mathrm{eff}}}{g_0} \left(\frac{T}{\mathrm{MeV}} \right)^3 \, . \tag{24}$$

In the equation above Ω_b is the critical baryonic density parameter (ratio between the baryon density and the critical density), m is the mass of a nucleon, T_0 is the present radiation temperature and the g_{eff}'s are the effective degree of freedom at the considered temperature.

Although the nucleosynthesis process does not occur in an equilibrium situation, it is interesting to perform some estimates based on this hypothesis. The critical temperature at which the relative concentration of deuterium satisfies $X_n X_p / X_D \approx 1$ (equivalent to a deuterium concentration of $X_D \approx 0.1$), can be estimated from the

condition $\log F(T, n_b) = 0$ [see Eqs. (23) and (24)]. Numerically, one obtains the equation

$$24.891 - \frac{3}{2}\log T_{\rm MeV} - \frac{2.225}{T_{\rm MeV}} - \log \Omega_b h^2 = 0 \,. \tag{25}$$

Since the critical deuterium concentration is expected to occur after the annihilation of the e^+e^- pairs, we have assumed that $g_{\rm eff}/g_0 = 1$ in order to derived the equation above. Taking $\Omega_b h^2 = 0.0221$, the numerical solution of this equation gives for the critical temperature $T_c \approx 0.068\,{\rm MeV}$. At this temperature the Universe is about 438 s old.

When the deuterium concentration attains the aforementioned critical value, a series of new reactions take place leading to the helium synthesis, namely

$$\begin{aligned} {}^2{\rm D} + {}^2{\rm D} &\rightarrow {}^3{\rm He} + n \leftrightarrow {}^3{\rm H} + p \,, \\ {}^3{\rm H} + {}^2{\rm D} &\rightarrow {}^4{\rm He} + n \,. \end{aligned} \tag{26}$$

Thus, in a first approximation, the resulting abundance of ^{4}He can be estimated if one assumes that all the free available neutrons will be locked into helium nuclei due to its high stability. At the critical point, the conservation equations indicate $X_n \approx 0.13$, $X_p \approx 0.77$ and $X_{\rm D} \approx 0.1$. If all free neutrons go into helium nuclei, then the resulting mass abundance is $Y(^4{\rm He}) \approx 2X_n \approx 0.26$. This is close to observed values of "primordial" helium in extragalactic HII regions, i.e., $Y(^4{\rm He}) = 0.2565 \pm 0.0010$ [see, for instance, Izotov and Thuan (2010)].

In Fig. 2 are plotted the primordial abundances of light elements derived from detailed nucleosynthesis calculations as a function of the baryon-to-photon ratio η given by

$$\eta = \frac{n_b}{n_\gamma} = 2.68 \times 10^{-8} (\Omega_b h^2) \,. \tag{27}$$

The abundances of ^{4}He are by mass while the others are by number. Horizontal lines indicate observed values or upper limits as in the case of ^{3}He. The width of the lines indicates the uncertainties in observations. The vertical line intends to show a common value of η in the intersections between observed and calculated values. Note there is still a tension with respect to the primordial lithium abundance. The resulting value for the baryon-to-photon ratio is $\eta = 6.2 \times 10^{-10}$ and replacing this in Eq. (27) one obtains $\Omega_b h^2 = 0.0221$, in agreement with the value derived from the amplitude of the peaks of the angular power spectrum of the cosmic microwave background (CMB). Polynomial fits representing the abundance curves shown in Fig. 2 can be found in Burles et al. (2001). These fits can be used to estimate the cosmological baryon-to-photon ratio from measured abundances of the primordial elements.

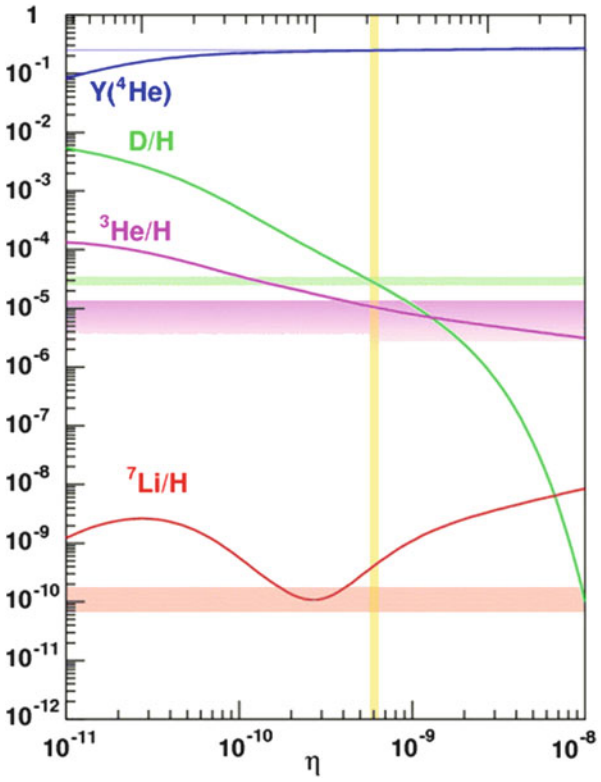

Fig. 2 Primordial abundances of ^{4}He, ^{3}He, ^{2}D and ^{7}Li given as a function of the baryon-to-photon ratio η. The ^{4}He abundance is by mass whereas those of the other isotopes are by number

4 The Recombination Era

After the nucleosynthesis era, the cosmic plasma is constituted mainly by photons, e^+e^- pairs, neutrinos and ionized baryons. These components, excepting neutrinos, are strongly coupled through electromagnetic interactions. The dynamics of the Universe is dominated by the relativistic matter, namely, photons, neutrinos and e^+e^- pairs. Around $kT \sim m_e c^2 = 0.511\,\mathrm{MeV}$, just after neutrino decoupling, the positron-electron pairs are no more in chemical equilibrium and annihilate. The released energy goes to photons but not to neutrinos since they are not coupled to the plasma. By entropy conservation, the temperature of photons increase with respect that of neutrinos in proportion to the variation of the number of degrees of freedom. Hence, before annihilation $g_{\mathrm{eff}} = 11/2$ and after $g_{\mathrm{eff}} = 2$ (only photons are present). Initially the neutrino temperature was equal to that of photons and after the annihilation of positron-electron pairs their relative temperature is $T_\nu = (4/11)^{1/3} T_\gamma$.

The cosmic expansion dominated by radiation ends when the energy density of photons and neutrinos becomes comparable to that under the form of non-relativistic matter. The equality occurs when

$$\frac{\pi^2}{30} g_{\rm eff} \frac{(kT_0)^4}{(\hbar c)^3} (1+z)^4 = \frac{3H_0^2 c^2 \Omega_m}{8\pi G} (1+z)^3 \,, \tag{28}$$

where $g_{\rm eff} = g_\gamma + (T_\nu/T_\gamma)^4 g_\nu = 3.557$ takes into account the fact that neutrinos are colder than photons and Ω_m is the critical total matter density parameter. The equation above can be written numerically

$$(1+z_{\rm eq}) = 22760(\Omega_m h^2) \,. \tag{29}$$

Using results from the Planck mission Ade et al. (2014), i.e., $\Omega_m h^2 = 0.143$, the redshift of matter-radiation equality is $(1+z_{\rm eq}) \approx 3250$, which corresponds to a radiation temperature of $T \approx 8900\,{\rm K}$.

At these temperatures photons and baryons are still coupled, since the photon mean free path due to Thomson scattering on free electrons is less than the Hubble radius c/H. Thus, decoupling occurs when the following condition is satisfied

$$\frac{H(z)}{n_e(z)\sigma_T c} \leq 1 \,. \tag{30}$$

In this equation n_e is the electron density and $\sigma_T = 6.65 \times 10^{-25}\,{\rm cm}^2$ is the Thomson cross section. Assuming ionization equilibrium and neglecting the contribution of helium, the ionization fraction of hydrogen obeys the Saha equation, this is

$$\frac{X_{\rm H}^2}{1-X_{\rm H}} = \frac{G(T)}{n} \,, \tag{31}$$

where we have considered that $n_p = n_e$, i.e., the proton density is equal to the electron density, $n = n_H + n_e$ with $n_{\rm H}$ being the density of neutral hydrogen, $X_{\rm H} = n_e/n$ and

$$G(T) = \left(\frac{m_e kT}{2\pi\hbar^2}\right)^{3/2} e^{-I/kT} \,, \tag{32}$$

where $I = 13.6\,{\rm eV}$ is the ionization potential of hydrogen and the total hydrogen density n is given by

$$n(z) = 1.12 \times 10^{-5} \Omega_b h^2 (1+z)^3 = n_0 (1+z)^3 \,. \tag{33}$$

Numerical solution of Eqs. (10), (30) and (31) gives for the baryon-radiation decoupling redshift $(1+z_{\rm ion}) = 1124$, corresponding to a temperature of $T_{\rm ion} = 3063\,{\rm K}$. When decoupling occurs the residual ionized fraction of hydrogen is $X_{\rm H} \approx 0.0071$.

4.1 The Last Scattering Surface

All the CMB photons do not decouple at the same redshift since scattering is a random process, and thus they have a non-zero probability of being observed today even if they were "last-scattered" in a region with an optical depth with respect to the line-of-sight larger than unity.

The probability for a photon to be "last" scattered in the redshift interval $z, z + dz$ is

$$P(z)dz = e^{-\tau_s(z)} \left| \frac{d\tau_s(z)}{dz} \right| dz \, . \tag{34}$$

In the above equation, the scattering optical depth is given by

$$\tau_s(z) = \int_0^z \sigma_T c n_e(z') \left| \frac{dt}{dz'} \right| dz' = \tag{35}$$

$$= \frac{\sigma_T c}{H_0} \int_0^z \frac{n(z') X_{\rm H}(z') dz'}{(1+z')\sqrt{\Omega_V + \Omega_m (1+z')^3}} \, .$$

Figure 3 shows the last scattering probability as a function of the redshift. The maximum escape probability occurs at $(1 + z) = 1114$. This means that CMB photons observed today have a maximum probability of have being scattered at this redshift for the last time. For lower redshifts the number of scatters (free electrons) decreases as well as the probability for the photon to have an interaction. On the other side, for higher redshifts the probability $P(z)$ also decreases because the Thomson optical depth increases and the photon will probably have other interactions before "escaping". The mean last scattering redshift derived from the computed distribution is $< (1 + z) >= 1097$ with a dispersion of $\sigma_z = 48$. The width of the last scattering surface, measured at half-maximum of the probability distribution is $\Delta z 100$.

4.2 Thermal Decoupling

After the baryon-radiation decoupling, there is still a residual ionization of the order of $X_{\rm H} \approx 0.0071$ as we have seen previously. Since the recombination rate of electrons and protons is higher than the cosmic expansion rate, the ionization fraction decreases continuously until "freezing" occurs. The evolution of the

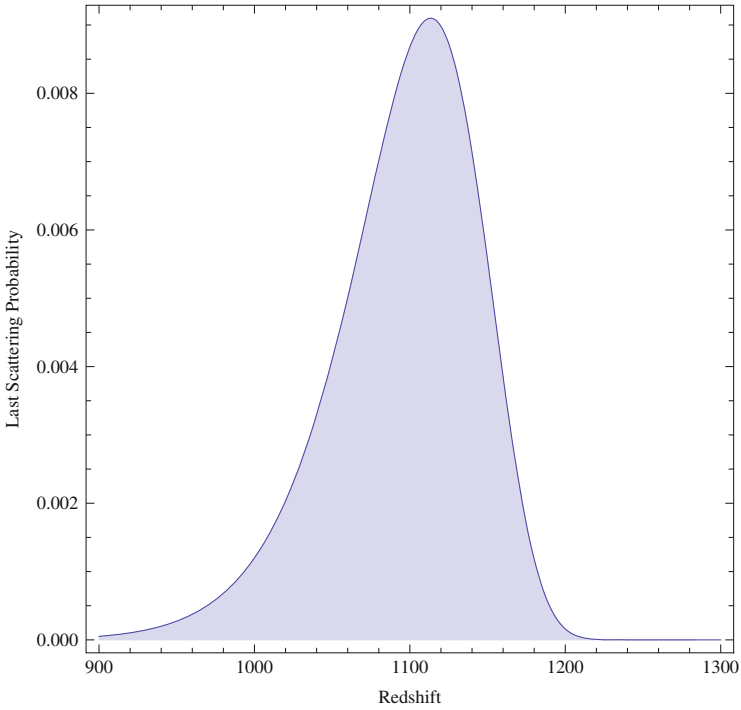

Fig. 3 Photon last scattering probability distribution as a function of the redshift

ionization fraction is governed by the equation

$$\frac{dX_H}{dz} = \alpha_B[T(z)]n(z)\frac{X_{\rm H}^2}{H_0(1+z)\sqrt{\Omega_V + \Omega_m(1+z)^3}}\ , \tag{36}$$

where $\alpha_B(T)$ is the total hydrogen recombination coefficient, excepting to the ground level that produces a photon able to reionize hydrogen. Solution of this equation permits to calculate the evolution of the hydrogen ionization fraction and the recombination timescale. The latter becomes comparable to the Hubble time at $(1+z) \approx 540$ when the ionization fraction is about $X_{\rm H} \approx 0.0003$. For lower redshifts the hydrogen ionization fraction is frozen at such a value.

The residual hydrogen ionization estimated above is rather small but even so the free electrons interact with CMB photons, suffering a drag that keeps the matter temperature close to the radiation temperature. Thus, due to this thermal coupling, the matter temperature evolves as

$$\frac{dT_m}{dt} = \frac{8\sigma_T u_\gamma}{3m_e c}\frac{X_{\rm H}}{(1+X_{\rm H})}(T_\gamma - T_m)\ , \tag{37}$$

where u_γ is the energy density of CMB radiation. This equation permits to define the Compton cooling timescale as

$$t_c = \frac{T_m}{\mid dT_m/dt \mid} = \frac{3.69 \times 10^{19}}{(1+z)^4 X_H} \left(\frac{T_m}{T_\gamma - T_m} \right) . \tag{38}$$

Thermal decoupling occurs when the Compton cooling timescale becomes comparable to the Hubble time, this is $t_c = H^{-1}$. From this condition one obtains

$$(1+z)^{5/2} \approx 3.981 \times 10^5 \sqrt{\Omega_m h^2} , \tag{39}$$

where it was assumed that $(T_\gamma - T_m)/T_m = 1$ and the hydrogen ionization fraction was taken to be equal to the freezing value. The relation above implies that thermal decoupling occurs at $(1+z) \approx 120$. This means that for lower redshifts the matter temperature varies only due to the expansion losses and is given approximately by

$$T_m \approx 160 \left[\frac{(1+z)}{120} \right]^2 \text{ K} . \tag{40}$$

5 Reionization

The spectra of QSOs obtained at different redshifts show dramatic variations. Objects at $z > 6$ indicate that the radiation shortward Lyman-α is completely absorbed, suggesting the presence of neutral hydrogen along the line-of-sight of those quasars. This effect is the so-called Gunn-Peterson (GP) trough. Figure 4, taken from Fan et al. (2000) shows an example of a quasar at $z = 5.8$, where it is possible to observe the GP effect. Note that at shorter wavelengths Lyman-β and OVI emission lines have non-zero fluxes, suggesting that this enhanced transmission occurs thanks to the presence of ionized hydrogen bubbles along the line-of-sight. Moreover, since the optical depth is proportional to the product of the line oscillator strength to the wavelength, i.e., $\tau \propto f\lambda$, the Ly-β optical depth is about 6.2 times smaller than Ly-α, giving a higher flux transmission in that transition.

For closer QSOs, namely, for objects at $z < 5.0$, the flux shortward Lyman-α is clearly detected and the Ly-α absorption is resolved into a multitude of individual absorption lines, which constitute the so-called "Lyman-α forest". The number density of these features increases with the redshift approximately as $N \propto (1+z)^{5/2}$ (Tytler et al. 2004). Some authors have considered that these absorption features are due to "clouds" of neutral hydrogen in the line-of-sight of the quasar and that are located in filaments forming the cosmic web (Chiu et al. 2003; Miralda-Escudé 2000). The present belief is that the absorption features are formed in the low density gas constituting the intergalactic medium (IGM). This gas is probably in thermal equilibrium, heated by photoionization and cooled by the cosmic expansion and

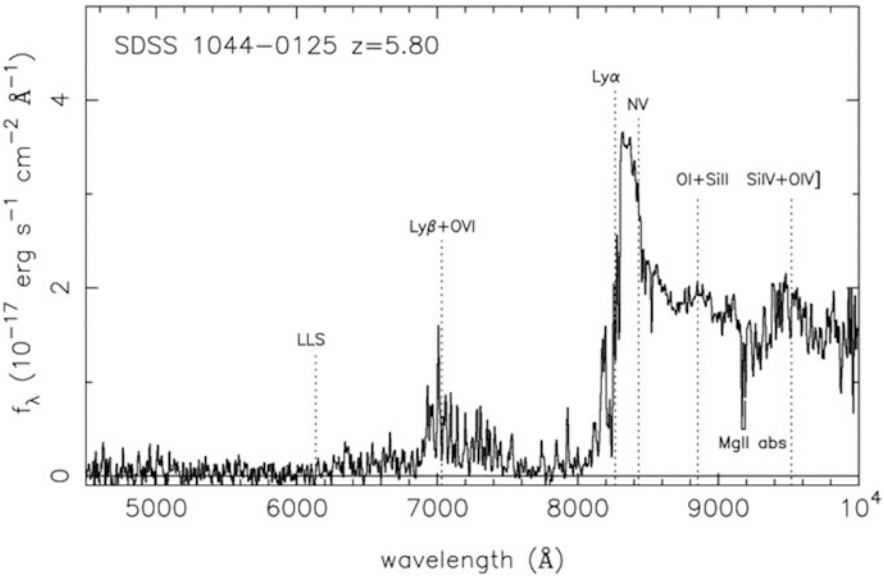

Fig. 4 Spectrum of the quasar SDSS 1044-0125 at z = 5.8. The flux attenuation shortward Lyman-α is clearly observed as well as an increase in the transmission around $\lambda = 7000$ Å, where the emission lines due to Lyman-β and OVI are seen. Figure from Fan et al. (2000)

radiative losses (Fan et al. 2006). Spectra of two QSOs at $z = 0.16$ and $z = 3.6$ are shown in Fig. 5, where the aforementioned absorption features can be seen.

Moreover, a moderate fraction of neutral hydrogen at $z \sim 6$ is required in order to explain the damped wing of Ly-α absorption features seen in the afterglow of some gamma-ray bursts (Totani et al. 2006, 2014). Additionally, the fraction of Ly-α emitting galaxies drops around $z \sim 7$, indicating again an important absorption by neutral hydrogen present in the intervening medium (Treu et al. 2013).

All these observations suggest that as the star formation process begins around z ∼20–30, the UV radiation of the young (and probably massive) stars ionize the nearby gas producing HII bubbles or "Strömgren spheres". These bubbles grow in volume and in number thanks to newly formed stars in recent assembled galaxies, then overlap producing a medium totally ionized. The first phase, when the UV sources isolated inside their own bubbles evolve, is relatively slow but the process accelerates when bubbles begin to overlap, since the mean free path of Lyman-continuum photons increase. Other UV sources may also contribute to the process as quasars formed in massive proto-halos, whose central black hole in a state of accretion, has a massive disk able to emit a large amount of ionizing photons (Montesinos et al. 2011). However, the contribution of quasars is probably not important for $z > 6 - 7$. After the end of the overlapping phase, the intergalactic medium becomes optically thin but still contains a small fraction of neutral hydrogen able to absorb Ly-α photons.

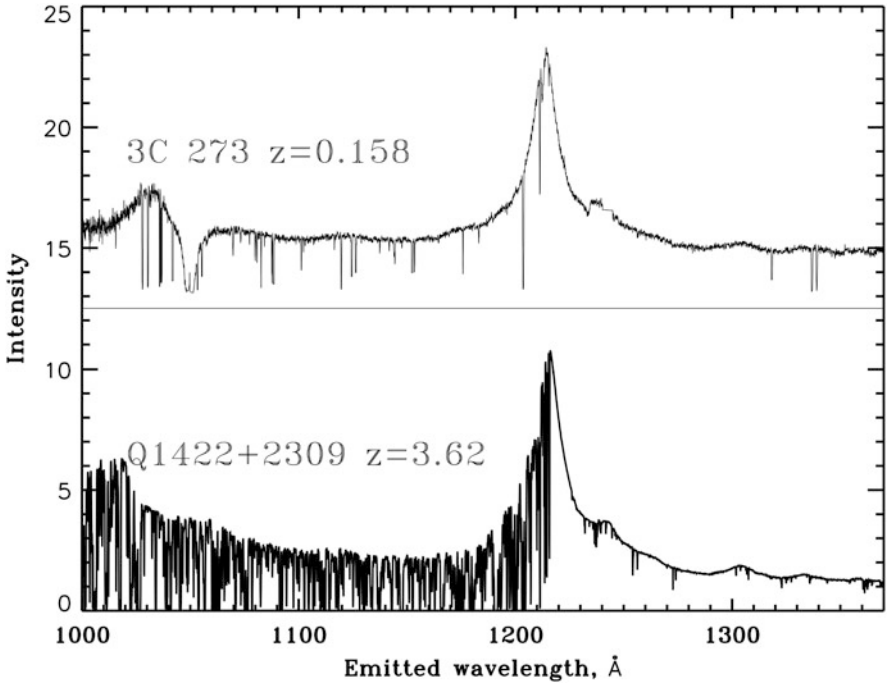

Fig. 5 Spectra of two quasars showing absorption features associated with neutral hydrogen clouds in the line-of-sight of these objects. Note also that the continuum shortward Lyman-α is not absorbed, indicating a transparent medium or that the gas in front of these quasars is ionized. Figure from http://pages.astronomy.ua.edu/keel/agn/forest.html, courtesy of professor William Keel

A simple analytical description of the overlapping process can be obtained in the following way (Barkana and Loeb 2001): imagine a HII bubble of proper volume V_p with an uniform particle density $n_{\rm ion}$. Then, the variation of the total number of ionized atoms depends on the absorption rate of UV photons and on the rate of recombinations occurring inside the bubble, i.e.,

$$\frac{d(n_{\rm ion}V_p)}{dt} = \frac{dN_{\rm uv}}{dt} - \alpha_B(T) n_{\rm ion}^2 V_p \ . \tag{41}$$

In this equation $\alpha_B(T)$, as before, is the total recombination coefficient between protons and electrons, excluding the fundamental level and it was assumed again that the electron density is equal to the proton density. In other words, the contribution of helium as an electron donor was neglected. Recalling that the comoving volume V is related to the proper volume as $V_p = a^3 V$ and that particles are conserved, Eq. (41), after a suitable average over the different bubble volumes, can be recast as

$$\frac{d < V >}{dt} = \frac{1}{n_0}\frac{dN_{\rm uv}}{dt} - \alpha_B(T) C \frac{n_0 < V >}{a^3} \ . \tag{42}$$

In the equation above, n_0 is the present particle density (see Eq. (33)) and $C =< n_{\rm ion}^2 > / < n_{\rm ion} >^2$ is the volume-averaged clumping factor of the gas inside the bubble. Recall also that $n_{\rm ion} a^3 = n_0$.

The maximum possible ionized volume can be estimated from Eq. (42) if recombinations are neglected. In this case, $< V_{\rm max} >= N_{\rm uv}/n_0$. In order to estimate the number of UV photons emitted by the stars that formed the considered HII region, the following assumptions are made: Let $Q_{\rm uv}$ be the number of ionizing photons emitted by unit of mass of formed stars. An average performed using a Salpeter initial mass function gives $Q_{\rm uv} = 6.62 \times 10^{60}$ ph $M_\odot^{-1}$. Consider a proto-halo of mass M and let be f_* the fraction of the baryonic mass converted into stars and f_{esc} the fraction of UV photons that escapes from the star formation region, being able to ionize the environmental gas. In this case,

$$V_{max} = \frac{Q_{\rm uv}}{n_0} M \left(\frac{\Omega_b}{\Omega_m}\right) f_* f_{\rm esc} \ . \tag{43}$$

Typically, $f_* \sim 0.3$ and $f_{\rm esc} \sim 0.1$. If spherical, the radius of such a volume is about $r_{\rm max} \approx 1.13\ (M/M_\odot)^{1/3}$ kpc and for a proto halo with a mass of $10^8\ M_\odot$, the maximum radius of the HII region is around 520 kpc.

Define now the ionization filling factor F that measures the ionized fraction of the Hubble volume V_H. This can be obtained by summing all the bubbles inside V_H, i.e.,

$$F = \sum_{\rm ion} \frac{< V_{\rm ion} >}{V_H} \ . \tag{44}$$

Performing such a sum in Eq. (42) one obtains

$$\frac{d}{dt} \sum_{\rm ion} \frac{< V_{\rm ion} >}{V_H} = \frac{1}{n_0 V_H} \sum_{\rm ion} \frac{dN_{uv}}{dt} - \alpha_B(T) C \frac{n_0}{a^3} \sum_{\rm ion} \frac{< V_{\rm ion} >}{V_H} \ . \tag{45}$$

Working the terms appearing in this equation gives respectively: the term on the left side is simply

$$\frac{d}{dt} \sum_{\rm ion} \frac{< V_{\rm ion} >}{V_H} = \frac{dF}{dt} \ . \tag{46}$$

The first term on the right side can be written as

$$\frac{1}{n_0 V_H} \sum_{\rm ion} \frac{dN_{\rm uv}}{dt} \approx \frac{1}{n_0} \sum_{\rm ion} \frac{Q_{\rm uv}}{\tau_*} \frac{M_*}{V_H} f_{\rm esc} \approx \frac{Q_{\rm uv} R_*}{n_0} f_{\rm esc} \ , \tag{47}$$

where τ_* is the mean life time of the ionizing stars, M_* is the mass of formed stars and R_* is the cosmic star formation rate, given in $M_\odot$ yr^{-1} Mpc^{-3}. The last term

is trivial, since it is sufficient to apply the definition of F. Using these relations and taking the redshift instead of time as the independent variable permit to recast Eq. (45) as

$$\frac{dF}{dz} = (1+z)^2 \frac{\alpha_B C n_0}{H(z)} F - \frac{Q_{\rm uv} R_* f_{\rm esc}}{n_0 (1+z) H(z)} \ . \tag{48}$$

This equation governs the evolution of the ionized volume fraction of the intergalactic medium, assuming that the cosmic star formation rate controls the production of UV photons when $z > 6$.

In order to solve numerically Eq. (48) some assumptions need to be made. The first point concerns the temperature of the ionized region, which will be assumed constant and the same for all bubbles. Studies in the vicinity of quasars Raskutti et al. (2012) suggest that the temperature of the IGM for $z > 8$ is, in fact, approximately constant and equal to 11,000 K. In the redshift interval $6 \leq z \leq 8$ the temperature increases up to 15,800 K. Since we expect that reionization will be completed at $z \sim 6$, this variation will discarded. The second point refers to the adopted star formation rate, which is given by the expression

$$R_*(z) = \frac{(0.0103 + 0.12z)}{[1 + (z/4)^{2.8}]} \ M_\odot \ {\rm yr}^{-1} \ {\rm Mpc}^{-3} \ . \tag{49}$$

This equation provides a good fit of the existing data (Hopkins and Beacom 2006) and is consistent with the cosmological simulations by Filloux et al. (2011, 2010). Figure 6 shows the adopted cosmic star formation rate (solid black curve) compared with data and simulations. Two free parameters remain: the clumping factor of the ionized gas and the escape fraction of photons from the star formation region. Concerning the clumping factor, early simulations yield large C-values of the order of 30. However the volume average performed in these studies included the very dense halo gas, increasing the clumpiness of the IGM. As we shall see in the next section, for $z > 6$, only a small fraction of baryons have collapsed and most of them constitute the intergalactic medium. The clumpiness of the gas is small and C-values of about 1–3 seem to be more reasonable (Shull et al. 2012). It should be emphasized that if $C > 1$ a complete overlap of the HII bubbles should not occur because of mass conservation. The escape fraction of UV photons can be estimated by requiring that the calculated scattering optical depth τ_s in the line-of-sight reproduces the observed value. This is given by Kuhlen and Faucher-Gighère (2012)

$$\tau_s = \int_0^{z=30} n_e(z) \sigma_T c F(z) \frac{dz}{(1+z) H(z)} \ . \tag{50}$$

From Planck's data, $\tau_s = 0.0925$ (Ade et al. 2014). Numerical solution of Eq. (48) for $C = 1$ with the initial condition $F = 0$ at $z = 30$ indicate that $F = 1$, this is complete overlapping of the ionized bubbles, occurs at $z = 5.89$ with $\tau_s = 0.0926$

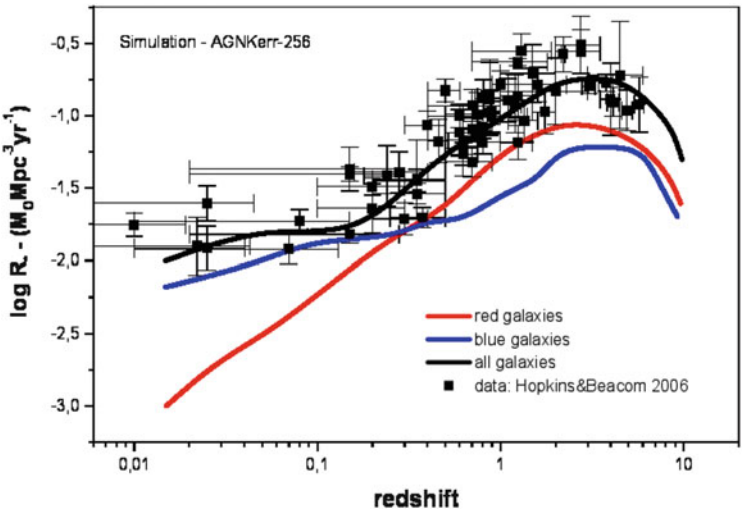

Fig. 6 The cosmic star formation rate: data points are from a compilation by Hopkins and Beacom (2006). The *red* and the *blue curves* represent respectively the star formation rate in red and blue galaxies derived from the simulation AGN Kerr (Filloux et al. 2011, 2010). The *black curve* is the total cosmic star formation rate for which a fit is given in the text

if $f_{\rm esc} = 0.019$. Note that this fraction is in nice agreement with the estimates by Chiu et al. (2003) in this redshift range. Figure 7 shows the evolution of the fraction of the ionized Hubble volume as a function of the redshift. It is worth mentioning that the reionization of the IGM is by no means "instantaneous". Half of the Hubble volume becomes reionized only at $z \approx 7.2$ and at $z \sim 11.5$ only 10 % of the IGM is reionized. This simple analysis suggests that star-forming galaxies are probably the main sources of UV-photons, being able to reionize the universe. The process begins around $z \sim 30$ and is achieved around $z \sim 6$.

5.1 Lyman-α Absorption

As we have seen, the reionization of the Universe by star-forming galaxies is nearly complete around $z \sim 6$. However, the ionization fraction is such that there is still enough neutral hydrogen atoms to produce a substantial absorption in the Lyman series.

In order to estimate the fraction of neutral atoms, we proceed in different steps. Firstly, the mean radiation intensity shortward the Lyman continuum is computed by solving the transfer equation. Considering, in a first approximation, that all ionizing

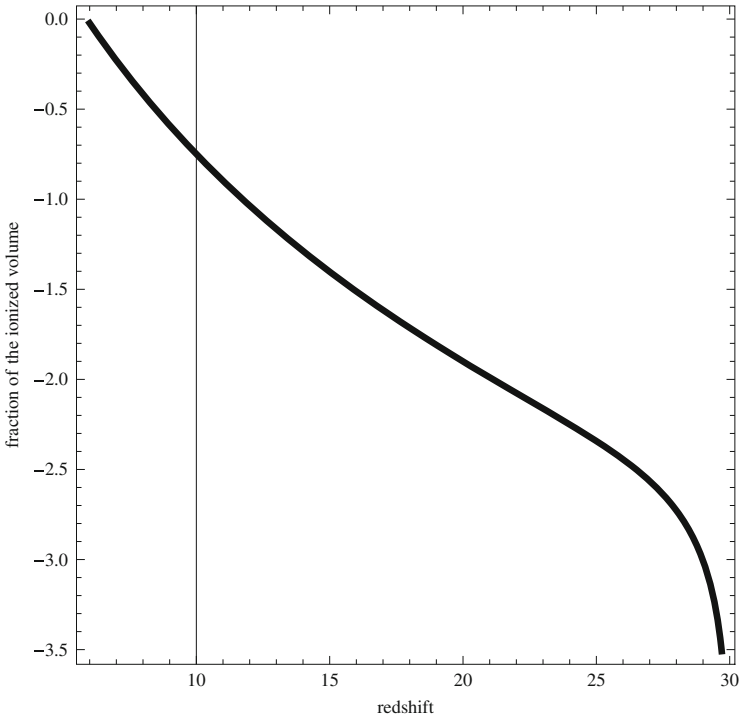

Fig. 7 Fraction of the ionized Hubble volume as a function of the redshift for C = 1 (Note that in ordinates is given the decimal logarithm of ionized volume fraction)

photons are emitted by stars near the Lyman limit, the emissivity coefficient can be written as

$$j_\nu(z) = \frac{Q_{\rm uv} R_*(z) f_{\rm esc} h\nu_L}{4\pi} \delta_D(\nu - \nu_L) \ , \tag{51}$$

where $\delta_D(x)$ stands for the Dirac's delta distribution function and $h\nu_L$ is the energy of a photon at the edge of the Lyman continuum. In the ionized region, the mean free path of a Ly-c photon is larger than the mean distance between sources, namely, the mean distance between galaxies. Thus, absorption can be neglected and, in this case, the solution of the transfer equation is trivial, giving for the intensity

$$I_\nu(z) = (1+z)^3 \int_z^{z_{\rm max}} \frac{c j_\nu(z')}{(1+z')^4 H(z')} dz' = \frac{Q_{\rm uv} hc f_{\rm esc} R_*(z)}{4\pi H(z)} \ . \tag{52}$$

Once the intensity of the ionizing radiation is obtained, the second step is to compute the ionization rate, which is given by

$$\Gamma_{\rm ion} = 4\pi \int_{\nu_L}^{\infty} \frac{I_\nu}{h\nu_L}\sigma_\nu d\nu = \frac{Q_{\rm uv} c\sigma_0}{2H_0} \frac{f_{\rm esc} R_*(z)}{\sqrt{\Omega_V + \Omega_m (1+z)^3}} \,, \tag{53}$$

where $\sigma_\nu = \sigma_0(\nu_L/\nu)^3$ is the approximate ionization cross section for hydrogen with $\sigma_0 = 6.3 \times 10^{-18}\,{\rm cm}^2$. Note that although the Ly-C absorption was neglected in the solution of the transfer equation, it is not possible to guarantee that the ionizing radiation intensity be homogeneous since the sources (galaxies) are strongly clustered.

The next step is to compute the Ly-α optical depth, which as we shall see below, is a "local" quantity due to effects of the cosmic expansion. The Ly-α absorption cross section is

$$\sigma_{12}(\nu) = \frac{\pi e^2}{m_e c} f_{12} \phi(\nu - \nu_L) \,, \tag{54}$$

where $\phi(\nu - \nu_L)$ is the line profile and ν_L is the line frequency in the local standard of rest. The line profile satisfies the normalization condition when integrated over frequencies. For an observer located at a position characterized by a redshift z, the element of optical depth in the radial direction is

$$d\tau_\alpha(\nu) = \frac{\pi e^2}{m_e c} f_{12} n_{\rm H} \phi(\nu - \nu_L) dr \,, \tag{55}$$

where $n_{\rm H}$ stands for the density of hydrogen atoms in the fundamental level and $f_{12} = 0.4162$ is the line oscillator strength. A photon emitted with a frequency ν_L at redshift z' will be detected with a frequency ν at redshift z as a consequence of the cosmic expansion, i.e.,

$$\nu = \nu_L - \nu_L \frac{V}{c} = \nu_L - \nu_L \frac{Hr}{c} \,. \tag{56}$$

Therefore, a variation dr in the proper radial distance corresponds to a variation of the photon frequency given by

$$dr = \frac{c}{\nu_L} \frac{d(\nu - \nu_L)}{\mid dV/dr \mid} = \frac{c}{\nu_L} \frac{d(\nu - \nu_L)}{H(z)} \,. \tag{57}$$

Substituting this result into Eq. (55) one obtains

$$d\tau_\alpha(\nu) = \frac{\pi e^2}{m_e c} f_{12} n_{\rm H} \frac{c}{\nu_L} \phi(\nu - \nu_L) \frac{d(\nu - \nu_L)}{H(z)} \,. \tag{58}$$

The integration along the line-of-sight was transformed in an integration in frequency along the profile (this is simply the so-called Sobolev approximation, very well known of stellar astrophysicists). Taking into account the normalization condition, one finally obtains

$$\tau_\alpha(\nu) = \frac{\pi e^2}{m_e c} f_{12} \lambda_{12} \frac{n_{\rm H}}{H(z)} \ . \tag{59}$$

As already anticipated, this relation depends only on local quantities, consequence of the Sobolev approximation.

The density of neutral H atoms is related to the total density by $n_H = (1 - X_H)n$, where X_H is the ionized fraction. In order to compute X_H (or $1 - X_H$), ionization equilibrium is usually assumed in the literature either in analytical studies or in numerical simulations. Here, the evolution of the ionization fraction is supposed to occur out of equilibrium and this is governed by the equation

$$\frac{d(1 - X_H)}{dz} = (1 - X_H) \frac{\Gamma_{\rm ion}}{(1 + z)H(z)} - \frac{\alpha_B n_0 \zeta(z)(1 + z)^2}{H(z)} \tag{60}$$

where $\zeta(z)$ is the fraction of non-collapsed baryons, still present in the diffuse IGM (see next section).

Numerical solutions of Eq. (60) were obtained for different values of the escape fraction $f_{\rm esc}$, this is 0.25, 0.30 and 0.35, neglecting clumping effects. Here higher values for the escape fraction of UV photons were considered, since the IGM is completely ionized as well as the gas in filaments or that inside halos. This increases dramatically the escape fraction of ionizing photons from star forming regions and the adopted range of values is similar to those adopted in other studies based on ionization equilibrium (Schirber 2003; Robertson et al. 2013). In Fig. 8 is shown the solution of Eq. (60) for $f_{\rm esc} = 0.35$. For comparison, it is also shown the evolution of the neutral hydrogen fraction when ionization equilibrium is assumed. Note that at $z \sim 6.5$ the non-equilibrium solution predicts a neutral hydrogen fraction more than one order of magnitude higher than the value expected from the equilibrium solution. This implies that even without an important clumping, there is enough neutral atoms in the IGM to produce a non-negligible Ly-α absorption.

Figure 9 shows the expected Ly-α flux transmission calculated from $T = \exp(-\tau_{\rm Ly_\alpha})$ where the optical depth was computed from Eq. (59) combined with the solution of Eq. (60). The best agreement with data by Songaila et al. (2004), Fan et al. (2006) is obtained with $f_{\rm esc} = 0.35$.

6 Baryons, Where Are You?

As we have seen previously, primordial nucleosynthesis and CMB data indicate that baryons represent about 4 % of the matter-energy content of the Universe. Then, we may ask how these baryons are distributed throughout the Universe.

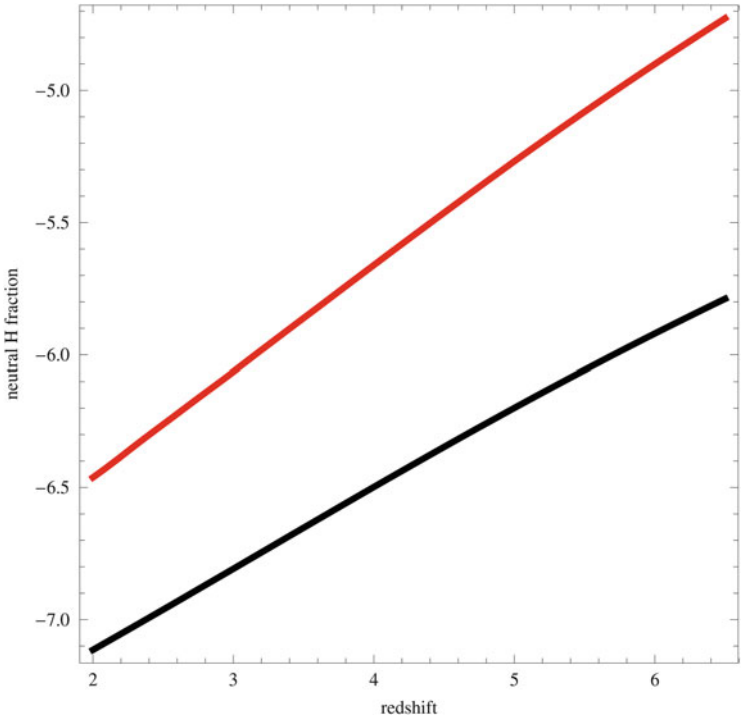

Fig. 8 Evolution of the fraction of neutral hydrogen as a function of the redshift. The *upper* (*red*) *curve* corresponds to a non-equilibrium solution discussed in the text, while the *lower* (*black*) *curve* corresponds to an equilibrium situation (note that ordinates are the decimal logarithm of the neutral hydrogen fraction)

Past investigations (Fukujita et al. 1998) lead to the conclusion that presently stars represent about 12 % of the total amount of baryons with the remaining fraction being under the form of cold, warm and hot gas. The cold gas ($T \leq 15{,}000$ K) and dense gas is present in galaxies, constituting the reservoir from which new stars are formed. The cold, rarefied and photoionized gas constitutes the intergalactic medium while the warm-hot gas can be found in different environments. In the one hand, gas at temperatures around 10^5–10^6 K is found in filaments constituting the cosmic web and halos, and in the other hand, gas at temperatures of about 10^7–10^8 K is found in the intracluster medium.

From the observational point of view, damped Ly-α absorption features (DLAs) observed in spectra of quasars are probably associated to the cold gas phase. Features with line-widths less than 40 km/s (narrow absorption lines or NLAs) are supposed to formed by clouds constituted by cold photoionized gas in the intergalactic medium, while broader features (BLAs) are thought to be formed in a warmer medium with temperatures around 10^5 K. This picture is supported by high-resolution UV spectra of low redshift quasars obtained with the COS

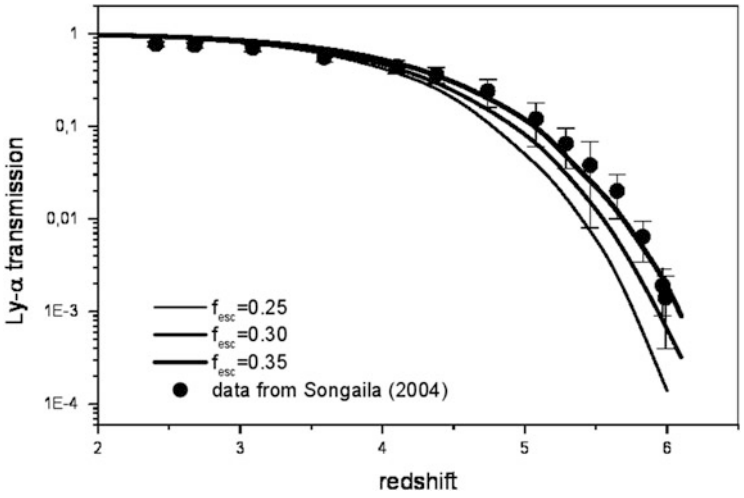

Fig. 9 Theoretical flux transmission shortward Ly-α as a function of the redshift (*solid curves*) for different values of the escape fraction of ionizing photons. From the *lower to the top curve*, the escape fraction f_{esc} is respectively 0.25, 0.30 and 0.35. Data points are from Songaila et al. (2004)

spectrograph on board of the Hubble Space Telescope, which show the presence of OVI absorption lines. About 37 % of the OVI absorbers have velocities coincident with those of the HI features. If these lines trace the same medium, they suggest a temperature $T \leq 10^5$ K. Approximately 53 % of the OVI absorbers have complex structures suggesting higher temperatures. The fraction of baryons in these different phases in rather uncertain because it depends on badly known parameters as the distribution of clouds with the redshift, ionization fraction and abundance. Different estimates suggest that at low redshift 30 % of the baryons are in the cold phase where DLA features are formed and about 20 % are in the warm phase traced by the broad and complex OVI absorption features. Including the contribution of stars, it is easy to conclude that a fraction of baryons is "missing" in the global budget.

Cosmological simulations are certainly an adequate tool to investigate the evolution of the distribution of baryons since they permit to probe a significant volume of the Universe. Moreover, cosmological simulations permit to follow consistently the evolution of dark matter and baryons in different phases, to test star formation conditions and to include feedback effects due to massive stars (stellar winds and UV ionizing radiation), supernovas and AGNs. Past investigations based in this approach (Cen and Ostriker 1999a,b, 2006; Davé 2001) suggest that the "missing" baryons might be in a gaseous phase at temperatures 10^5–10^7 K. According to these studies, the warm-hot (WHIM) gas is primarily heated by shocks resulting of the formation of structures like filaments by gravitational instability, with the feedback of supernovas playing a secondary role. These investigations have also shown that there is a continuous transfer of gas from the diffuse IGM to the WHIM that is accentuated when $z < 3$.

6.1 The Nice Cosmological Code

The Nice cosmological code is based on the structure of the GADGET code, in a formulation that conserves energy and entropy (Springel 2005). The code is fully adaptive in space and time, allowing simulations with an adequate dynamic range required to study both high and low density regions. Different physical mechanisms affecting the dynamics and the thermodynamics of the gas were included such as, cooling including collisional excitation of HI, HeI and He II levels, free-free transitions, radiative recombination, hydrogen molecular emission, atomic fine-structure level excitations of trace elements and Compton interactions with the CMB photons. Local heating and ionization produced by UV radiation of newly formed stars were also included as well as mechanical energy injected either by supernovas or AGNs.

Some particular aspects of the Nice code should be emphasized. For instance, the return of mass to the interstellar medium, consequence of the stellar evolution, implying that masses associated to gas and star particles vary in time. The delay between the onset of type Ia and type II supernovas is also taken into account in the code. Moreover, the chemical enrichment of the environment is made via an adequate algorithm for SPH codes that is able to simulate a turbulent diffusion process.

Concerning the injection of mechanical energy by AGNs in the surrounding medium, this is performed through two opposite jets with an opening angle of about 30° with a power derived from the AGNs simulations (Koide et al. 2002).

The results presented here are from simulations performed in a cube with a size of $50h^{-1}$ Mpc, using a flat ΛCDM cosmology with $H_0 = 70$ km/s/Mpc. The normalization of the matter density fluctuation spectrum was taken to be $\sigma_8 = 0.9$. The mass resolution for "gas particles" is about $1.18 \times 10^8 M_\odot$ and all the computations were performed at the "Center for Numerical Computations of the Cote d'Azur Observatory" at Nice.

Figure 10 shows the evolution of the baryon fraction in different phases. At $z \sim 6$, most of gas in the diffuse intergalactic medium, which as we have seen before is now completely ionized by the UV radiation of stars. Notice that at this redshift only 0.8 % of the baryons are under the form of stars but even this small fraction is enough to reionize the IGM. As the Universe evolves, at $z \sim 2.5$, the collapsed fraction of baryons has increased: stars represent now 4.6 % of the baryon budget while the fraction under the form of cold and dense gas or WHIM is respectively of 14.6 % and 7 %. Most of the baryons (73.6 %) are still in the diffuse IGM. Presently, the major fraction of baryons is in the warm-hot shocked medium (43 %) and in the diffuse IGM (36 %). The remaining fraction constitutes the stars (14 %) and the cold and dense gas present in the interstellar medium of galaxies (7 %).

These simulations permit also to predict the evolution of the metallicity of the cold and dense phase, supposed to be responsible for the formation of DLA features observed in the spectra of QSOs. Figure 11 shows the predicted metallicity of the

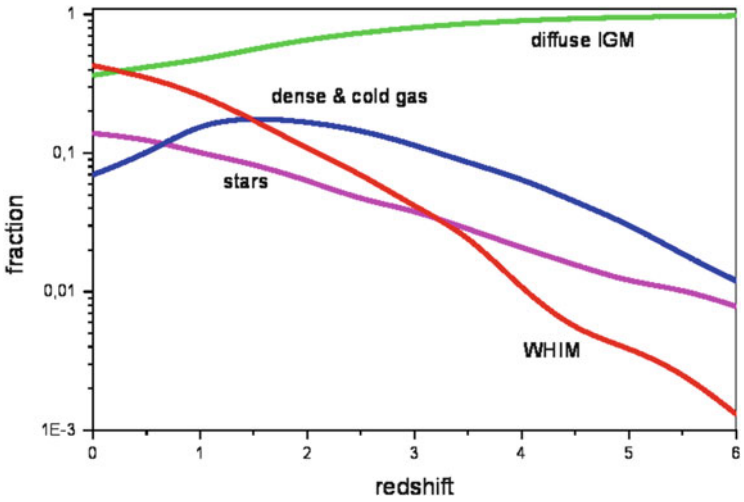

Fig. 10 Evolution of the baryon fraction in different phases derived from cosmological simulations

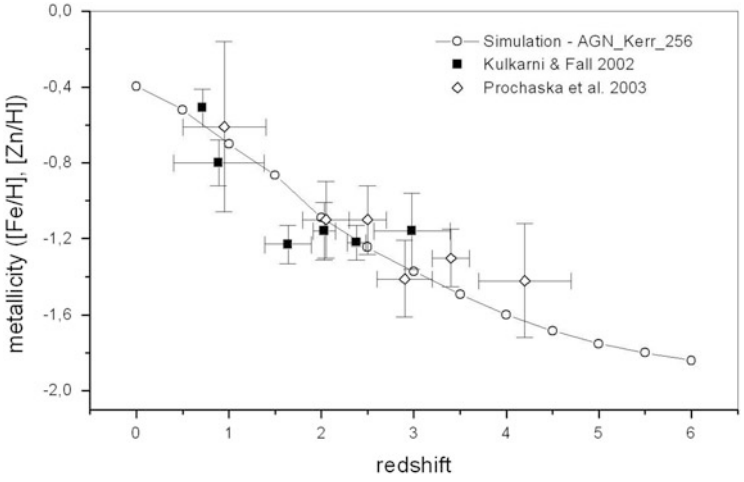

Fig. 11 Predicted evolution of the metallicity of the cold and dense phase compared with metallicities derived from DLA systems

cold and dense phase as a function of the redshift compared with metallicities derived from DLAs systems. A more detailed analysis can be found in Durier and de Freitas Pacheco (2011, 2012).

7 Summary

As we have seen in this lecture, hadrons appear has a consequence of a phase transition in which quarks and gluons in a state of asymptotic freedom become confined. Such a transition occurs when the Universe has cooled to a temperature of about 145 MeV, corresponding to an age of 96 μs. The duration of this transition is about 13 μs.

When hadrons appear, the relative abundance of neutrons and protons is fixed by statistical equilibrium conditions maintained by weak interaction reactions. Just after the quark-hadron transition the neutron-to-proton ratio is about unity, decreasing as the Universe expands and cools.

Neutrinos decouple from the cosmic plasma when the temperature decreased to about 1.1 MeV or when the Universe is aged of ~ 1.7 s. At neutrino decoupling, the neutron-to-proton ratio is $n/p \sim 0.308$. After neutrino decoupling the n/p ratio is no longer given by its statistical equilibrium value. Neutrons begin to decay and interact with protons to produce deuterium. When the cosmic plasma cools to a temperature of ~ 70 keV the neutron fraction is $n_n/n_b \sim 0.13$, a series of reactions take place leading to helium production as well as to small amounts of deuterium and lithium. The resulting abundances depend on the baryon-to-photon ratio, a nearly conserved quantity. From the observed abundances of helium in H II regions present in metal-poor galaxies and deuterium abundances derived from absorption features in the spectra of quasars, it is possible to estimate the abundance of baryons in the universe using the results of primordial nucleosynthesis. This is in nice agreement with determinations issued from the angular power spectrum of the CMB.

Protons are coupled to radiation through the interaction between photons and electrons. When the mean free path due to Thomson scattering becomes less than the Hubble radius, matter decouples from radiation and this occurs around $z \sim 1120$, when $T \sim 3000$ K. Since scattering is a random process, the "last-scattering" surface is not defined by a unique redshift. CMB photons seen today are, on the average, originated from redshift $< (1 + z) >= 1097$ with a dispersion $\sigma_z = 48$.

After matter-radiation decoupling, the residual ionization fraction of hydrogen continues to decrease because recombinations occur in a rate faster than the expansion rate of the Universe. This condition is broken when $(1 + z) \sim 540$. For lower redshifts the ionization fraction remains "frozen" at $X_H \sim 0.0003$. Even with such a low ionization level, matter is thermally coupled to CMB photons and decoupling occurs when the Compton cooling timescale becomes comparable to H^{-1}. This occurs at $(1+z) \approx 120$. After thermal decoupling, the matter temperature decays as the inverse of the square of the scale factor (adiabatic expansion).

As the star formation activity begins around $z \sim$ 20–30, UV photons are produced, escape from their host galaxies, ionizing the intergalactic medium. The HII bubbles formed around star forming galaxies overlap around $z \sim 6$, when the IGM becomes totally ionized. However, the fraction of neutral hydrogen atoms is enough to produce a significant Lyman-α absorption in the spectra of quasars.

Presently, baryons can be found in different phases: most of them are either in the warm-hot medium pervading filaments and the intracluster medium ($\sim 43\,\%$) or in the diffuse photoionized IGM ($\sim 36\,\%$). The remaining fraction is collapsed into stars ($\sim 14\,\%$) or constitutes the cold-dense gas present in galaxies, reservoir of new stars.

References

Barkana, R., Loeb, A.: Phys. Rep. **349**, 125 (2001)
Bazarov A., et al.: Phys. Rev. **D85**, 054503 (2012)
Boyd, G., et al.: Phys. Rev. Lett. **75**, 4169 (1995)
Begun, V.V., Gorenstein, M.I., Mogilevsky, O.A.: Ukr. J. Phys. **55**, 1049 (2010). arXiv:hep-ph/1001.3139
Boyd, G., et al.: Nucl. Phys. B **469**, 419 (1996)
Brown, F.R., et al.: Phys. Rev. Lett. **61**, 2058 (1988)
Burles, S., Nollett, K.M., Turner, M.S.: Astrophys. J. **552**, L1 (2001)
Cen, R., Ostriker, J.P.: Astrophys. J. **519**, L109 (1999a)
Cen, R., Ostriker, J.P.: Astrophys. J. **514**, 1 (1999b)
Cen, R., Ostriker, J.P., Astrophys. J. **650**, 560 (2006)
Chiu, W.A., Fan, X., Ostriker, J.P.: Astrophys. J. **599**, 759 (2003)
Cleyman, J., Gavai, R.V., Suhonen, E.: Phys. Rep. **130**, 217 (1986)
Davé, R., et al.: Astrophys. J. **552**, 473 (2001)
Durier, F., de Freitas Pacheco, J.A.: Int. J. Mod. Phys. **E20**, 44 (2011)
Durier, F., de Freitas Pacheco, J.A.: AIP Conf. Proc. **1471**, 10 (2012)
Fan, X., et al.: AJ **120**, 1167 (2000)
Fan, X., Carilli, C.L., Keating, B.: Annu. Rev. Astron. Astrophys. **44**, 415 (2006)
Filloux, C., Durier, F., de Freitas Pacheco, J.A., Silk, J.: Int. J. Mod. Phys. **D19**, 1233 (2010)
Filloux, C., de Freitas Pacheco, J.A., Durier, F., de Araujo, J.C.N.: Int. J. Mod. Phys. **D20**, 2399 (2011)
Fukujita, M., Hogan, C.J., Peebles, P.J.E.: Astrophys. J. **503**, 518 (1998)
Hopkins, A.M., Beacom, J.F.: Astrophys. J. **651**, 142 (2006)
Izotov, Y.I., Thuan, T.X.: Astrophys. J. **710**, L67 (2010)
Karsch, F.: Nucl. Phys. A **590**, 367C (1995). arXiv:hep-lat/9503010
Karsch, F., Lect. Notes Phys. **583**, 209 (2002). arXiv:hep-lat/0106019v2
Khan, A.A., et al.: Phys. Rev. D **64**, 074510 (2001). arXiv:hep-lat/0103028
Koide, S., Shibata, K., Kudoh, T., Meyer, D.L.: Science **295**, 1688 (2002)
Kuhlen, M., Faucher-Gighère, C.-A.: Mon. Not. R. Astron. Soc. **423**, 862 (2012)
Montesinos, M.A., de Freitas Pacheco, J.A.: Astron. Astrophys. **526**, A146 (2011)
Miralda-Escudé, J., Haehnelt, M., Rees, M.J.: Astrophys. J. **530**, 1 (2000)
Planck collaboration, Ade, P.A.R., et al.: Astron. Astrophys. **571**, A16 (2014). arXiv:1303.5076v3
Raskutti, S., et al.: Mon. Not. R. Astron. Soc. **421**, 1969 (2012)
Robertson, B.E., et al.: Astrophys. J. **768**, 71 (2013)
Shull, J.M., Harness, A., Trenti, M., Smith, B.D.: Astrophys. J. **747**, 100 (2012)
Shuryak, E.V.: Phys. Rep. **61**, 71 (1980)
Schirber, M., Bullock, J.S.: Astrophys. J. **584**, 110 (2003)
Songaila, A.: Astron. J. **127**, 2598 (2004)
Springel, V.: Mon. Not. R. Astron. Soc. **364**, 1105 (2005)
Totani, T., et al.: Publ. Astron. Soc. Jpn. **58**, 485 (2006)
Totani, T., et al.: Publ. Astron. Soc. Jpn. **66**(3), 63 (2014). arXiv:1312.3934
Treu, T., et al.: Astrophys. J. **775**, L29 (2013)
Tytler, D., et al.: Astrophys. J. **617**, 1 (2004)

Peculiar Velocity Effects on the CMB

Miguel Quartin

Abstract The aberration effect due to our peculiar velocity $\boldsymbol{\beta}$ with respect to the CMB induces mixing among multipoles and off-diagonal correlations at all scales which can be used as a measurement of $\boldsymbol{\beta}$, which is independent of the standard measurement using the CMB temperature dipole. Here we review the two techniques that allow this measurement and the forecasts for the achievable precision with which these correlations can be measured in a number of CMB missions. The forecast in the case of Planck was later shown to be very accurate when the collaboration released their first set of data and measured the CMB aberration for the first time. Finally, we show that a proper accounting for the aberration effect solves the mystery of the power asymmetries in the CMB, present since the WMAP era.

1 Introduction

The dipole of the CMB is measured to be much larger than the other multipoles and this is usually attributed to a Doppler effect due to our peculiar velocity $\boldsymbol{\beta}$ with respect to the CMB rest frame. Under this assumption we can infer, by combining the measured temperatures of WMAP dipole (Hinshaw et al. 2009) with the COBE monopole (Lineweaver et al. 1996; Mather et al. 1999), its direction to be $l = 263.99^\circ \pm 0.14^\circ$, $b = 48.26^\circ \pm 0.03^\circ$ in galactic coordinates, and its modulus to be $\beta \equiv |\boldsymbol{\beta}| = (1.231 \pm 0.003) \times 10^{-3}$. These very precise numbers rely however entirely on the above assumption, but generically the CMB dipole is not necessarily due only to a relative velocity. One way to test this assumption was originally proposed in Burles and Rappaport (2006). Recently it has also been pointed out (Challinor and van Leeuwen 2002; Kosowsky and Kahniashvili 2011; Amendola et al. 2011) that *all* the CMB multipoles $a_{\ell m}$ have a correction due to our local peculiar velocity because the primordial anisotropies are distorted by the Doppler and aberration effects. This shows up as a correlation between different

M. Quartin (✉)
Instituto de Física, Universidade Federal do Rio de Janeiro, CEP 21941-972, Rio de Janeiro, RJ, Brazil
e-mail: mquartin@if.ufrj.br

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_7

multipoles ℓ and could be used as an alternative way of measuring our velocity, as an independent consistency check. This offers also a way to test the isotropy of the Universe, not only on very large scales, but as we will discuss below, also at small scales.

The analysis of these CMB correlations has been performed (Challinor and van Leeuwen 2002; Kosowsky and Kahniashvili 2011; Pereira et al. 2010; Amendola et al. 2011) relying upon a Taylor expansion in orders of β of boost effects (Doppler and aberration) on the multipole coefficients $a^X_{\ell m}$, where X stands for T (temperature) or E, B (the E and B modes of polarization). It was shown in Amendola et al. (2011) that one can non-trivially and very efficiently define three different estimators in order to measure directly the three cartesian components of $\boldsymbol{\beta}$, without having to scan all the possible directions in the sky. Moreover it was shown in Chluba (2011), Notari and Quartin (2012), that although a first order Taylor expansion breaks down for individual $a_{\ell m}$'s, it works surprisingly well for the full estimators all the way to $\ell \sim 3000$.

Even though the linear expansion seem to work surprisingly well, the conceptually safer path is to compute the full aberration kernel given by a transformation $a^{[A]}_{\ell m} = \sum_{\ell'=0}^{\infty} K_{\ell' \ell m}\, a^{[P]}_{\ell' m}$, between the aberrated $[A]$ frame and the primordial $[P]$ correlations. One in principle would have to compute the matrix elements $K_{\ell' \ell m}$, sometimes referred to as the *aberration kernel*, which are integrals of spherical harmonics with different arguments, plug the matrix elements into all the possible two-point correlation functions and compare with the data. While in principle straightforward this procedure has some disadvantages: (1) this can be a heavy and delicate numerical task, because of the highly oscillating integrands and also because of the huge number of correlators that one has to consider (Planck alone measures multipoles up to $\ell \simeq 2000$ for both temperature and polarization); (2) it is not obvious to understand in this case whether the three simple estimators can be written explicitly for the three cartesian components of the velocity for any n in $a^*_{\ell m} a_{\ell \pm n m'}$, so perhaps the procedure would have to be carried out scanning the sky in all possible directions which would probably make this approach even more expensive in terms of computational time. This approach was explored in Notari and Quartin (2012), and very simple analytical approximations were empirically found. These solutions were later confirmed when they were re-derived mathematically from first principles in Dai and Chluba (2014).

The ability to measure exotic contributions to the dipole is of great interest to test Cosmology on very large scales, which could hide non-trivial phenomena, as suggested by some reported anomalies on the low-ℓ CMB multipoles itself (Copi et al. 2010; Bennett et al. 2011) [see also Notari and Quartin (2012) and references therein].

Interestingly, it was shown in Notari et al. (2014) that the aberration and Doppler effects provide not only a signal but also a noise in the form of a bias to the CMB. Although such a bias does not affect traditional parameter estimation when measuring the full sky (Catena and Notari 2013) or some common non-Gaussianity measurements (Catena et al. 2013), it does introduce a significant amount of power

anomalies to the CMB (Notari et al. 2014; Quartin and Notari 2015). In fact, it was concluded in Quartin and Notari (2015) that with a proper accounting of boost effects and noise asymmetry no significant power asymmetries seem to be present in the full Planck data set. Interestingly, without proper removal of Doppler and aberration effects one would find spurious anomalies at high ℓ between 3σ and 5σ. Even when considering only $\ell < 600$ we find that the boost is non-negligible and alleviates the discrepancy by roughly half-σ.

2 Fitting Functions for the Aberration Kernel

It was shown in Challinor and van Leeuwen (2002), Kosowsky and Kahniashvili (2011), Amendola et al. (2011) that when subjected to an aberration effect the $a_{\ell m}$ coefficients of the spherical harmonic decomposition of the temperature (and polarization) contrast transform in the following way

$$a^{X\,[A]}_{\ell m} \;=\; \sum_{\ell'=0}^{\infty} K^{X}_{\ell'\,\ell\,m}\, a^{X\,[P]}_{\ell' m}\,, \tag{1}$$

where the superscript X stands for either temperature (T) or one of the two independent modes of polarization (E and B) and where $[A]$ denotes the aberrated coefficients, to be contrasted with $[P]$, the primordial (non-aberrated) ones.

In the case of temperature, the exact coefficients of (1) are given by Kosowsky and Kahniashvili (2011) [we here follow Amendola et al. (2011) for the convention of the sense of $\boldsymbol{\beta}$ which results in an overall sign change in the velocity β]

$$K^{T}_{\ell'\,\ell\,m} = \int_{-1}^{1} \frac{\mathrm{d}x}{\gamma\,(1-\beta x)}\, \tilde{P}^{m}_{\ell'}(x)\, \tilde{P}^{m}_{\ell}\!\left(\frac{x-\beta}{1-\beta x}\right), \tag{2}$$

where $\gamma \equiv 1/\sqrt{1-\beta^2}$ is the standard Lorentz factor and

$$\tilde{P}^{m}_{\ell}(x) \;\equiv\; \sqrt{\frac{2\ell+1}{2}\,\frac{(\ell-m)!}{(\ell+m)!}}\, P^{m}_{\ell}(x)\,, \tag{3}$$

and where P^{m}_{ℓ} are the associated Legendre polynomials. For polarization the formulae are similar if one makes use of spin-weighted spherical harmonics (${}_s\tilde{Y}^{m}_{\ell}$). Following Challinor and van Leeuwen (2002), we get

$$K^{P}_{\ell'\,\ell\,m} \;=\; \frac{1}{2}\Big[{}_2K_{\ell'\,\ell\,m} + {}_{-2}K_{\ell'\,\ell\,m}\Big], \tag{4}$$

in which

$$ {}_sK^P_{\ell'\,\ell\,m} = \int_{-1}^{1} \frac{\mathrm{d}x}{\gamma\,[1-\beta x]}\, {}_s\tilde{P}^m_{\ell'}(x)\, {}_s\tilde{P}^m_{\ell}\left(\frac{x-\beta}{1-\beta x}\right) \tag{5} $$

and where in turn ${}_s\tilde{P}^m_\ell(x) \equiv \sqrt{2\pi}\ {}_sY^m_\ell(x, \phi = 0)$ (see Notari and Quartin 2012).

The oscillatory nature of these integrals poses a numerical challenge, which make their direct computation very slow for high ℓ. In a recent work (Chluba 2011) a recursive method was developed which is claimed to allow fast and accurate evaluation of these integrals. In this section we follow instead a different route and compute some of the integrals just by numerical integration. Surprisingly, we found out numerically that these integrals can be fit *very precisely* by Bessel functions, which greatly simplifies the analysis.

Following Challinor and van Leeuwen (2002) it is convenient to define the quantities

$$ {}_sG_{\ell m} \equiv \sqrt{\frac{\ell^2-m^2}{4\ell^2-1}\left[1-\frac{s^2}{\ell^2}\right]}, \tag{6} $$

where again s represents a spin weight which is 0 for temperature and 2 for the E and B modes of polarization. Note that we always have $0 \leq {}_sG_{\ell m} \lesssim 1/2$, the lower limit being achieved when $|m| = \ell$ and the higher one when $m = 0$. An exquisite fit for small scales to the full non-linear integral for general m is given by

$$ \begin{aligned} K^T_{\ell-1\,\ell\,m} &\simeq J_1\Big(-2\,\beta\,\ell\ {}_0G_{\ell m}\Big), \\ K^T_{\ell+1\,\ell\,m} &\simeq J_1\Big(2\,\beta\,(\ell+1)\ {}_0G_{\ell+1\,m}\Big), \end{aligned} \tag{7} $$

where J_1 is the Bessel function of the first kind. Moreover, we find that similar relations to the above one apply also for non-neighboring correlations (i.e., between any ℓ and $\ell \pm n$, $n \geq 1$) and also for polarization:

$$ \begin{aligned} K^X_{\ell-n\,\ell\,m} &\simeq J_n\left(-2\,\beta\left[\prod_{k=0}^{n-1}\left[(\ell-k)\ {}_sG_{\ell-k\,m}\right]\right]^{1/n}\right), \\ K^X_{\ell+n\,\ell\,m} &\simeq J_n\left(2\,\beta\left[\prod_{k=1}^{n}\left[(\ell+k)\ {}_sG_{\ell+k\,m}\right]\right]^{1/n}\right), \end{aligned} \tag{8} $$

which we find to be accurate to around $0.2\,\%$ for all values of ℓ and m and all values of n. For $n = 0$ one cannot apply Eq. (8) directly, but we find that an

analogous fit is given by (with the same precision)

$$K^X_{\ell\,\ell\,m} \simeq J_0\Bigg(\beta\sqrt{2}\Bigg[-(\ell+1)(\ell+2)\,({}_sG_{\ell+1\,m})^2 - \ell(\ell-1)\,({}_sG_{\ell\,m})^2 + \ell(\ell+1) - m^2 + 1 - s^2 + \frac{s^2\,m^2}{\ell(\ell+1)}\Bigg]^{\frac{1}{2}}\Bigg). \tag{9}$$

As can be seen from the above relations, the temperature and polarization aberration kernels become almost indistinguishable for small scales (usually for any $\ell \gg 1$, or when $|m| \simeq \ell$, for $\ell \gg 5$).

A Taylor expansion of the Bessel functions yields to leading order

$$J_n(x) = \frac{1}{2^n\,n!}x^n + \mathcal{O}\left(x^{n+2}\right), \tag{10}$$

which is valid for positive integer values of n. Expanding the above fits in orders of β we find that the coefficients with the Doppler correction *exactly* up to leading order for any value of n for both temperature and polarization. In other words, for a given n the fits are exact to order $\mathcal{O}\left(\beta^n\right)$. This was confirmed through direct analytic integration of the leading order of the Taylor expansion in β of Eq. (2) for specific values of $\{\ell,\ \ell',\ m\}$. This is an interesting result, as the $\{\ell,\ \ell \pm 3\}$, $\{\ell,\ \ell \pm 4\}$ and so forth leading order coefficients were never derived before in the literature. The $\{\ell,\ \ell - 3\}$ is for instance simply:

$$K^T_{\ell-3\,\ell\,m} = -\frac{1}{6}\beta^3\,\ell(\ell-1)(\ell-2)\;{}_0G_{\ell\,m}\;{}_0G_{\ell-1\,m}\;{}_0G_{\ell-2\,m} + \mathcal{O}\left(\beta^5\right), \tag{11}$$

and similarly simple expressions hold for other $\{\ell,\ \ell \pm n\}$ correlations. Another cross-check of the above formulae is to confront the expansion up to second order with the coefficients in Challinor and van Leeuwen (2002).

One could therefore conceive of the above fitting formula as direct method to estimate the correlators which bypasses the need to pre-deboost the CMB. Nevertheless this is a subtle point since, as shown in Amendola et al. (2011), in the velocity estimator there is a nearly-exact cancelation of the leading-order term in ℓ and it is thus possible that small corrections to the value of the coefficients lead to moderate corrections to the estimators. One would therefore need to carefully to check whether the accuracy provided by the fits here proposed is enough for such analysis. Moreover, as stated in Sect. 1 the usefulness of such method would also be dependent on whether also in this case three simple estimators can be written explicitly for the three cartesian components of the velocity, thus avoiding the need to compute and minimize a χ^2 for all possible sky directions. If both issues can be circumvented, then this would provide a technique to measure our velocity, which could be used as a cross-check for the pre-deboost technique.

As a side note, the fit Eq. (8) gets worse for high values of β and in fact breaks down whenever β is close to 1. This can be understood as one must have on physical grounds that $\beta < 1$ and from Eq. (2) that the integrals must approach zero in the limit $\beta \to 1$; the proposed fit instead does not go to zero in the same limit.[1] This fact is not important for the realistic case, since $\beta \sim 10^{-3}$, but it might be relevant if one tries to use our fitting function to try and guess an exact analytic solution to the integrals Eq. (2). In fact, the very high precision and breadth of applicability of an arguably simple fitting function Eq. (8) (which holds for any n and either temperature or polarization) hints to the existence of an analytical solution involving these Bessel functions, which perhaps can only be derived in some special limit. An analytical derivation of Eq. (2) or Eq. (5) was later found in Dai and Chluba (2014), confirming its validity as a very good approximation for $\beta \lesssim 0.1$.

3 Applications to Current and Future CMB Experiments

3.1 Summary of CMB Experiments

In this section we explore the idea in more detail and estimate the expected signal strength and detection possibilities in a number of CMB experiments. As it turns out, experiments which cover only a few percent of the sky are not favored to detect the proposed correlations, due to extra contribution from cosmic variance (see Eqs. (12) and (24) below). We will focus instead on the experiments which cover a substantial fraction of the sky, in particular WMAP, Planck and SPT as well as some of the proposed future ones (ACTPol, SPTPol, Cosmic Origins Explorer—COrE and The Experimental Probe of Inflationary Cosmology—EPIC).

Estimates in this section refer to statistical noise alone; care must be taken when interpreting these due to the presence of foregrounds and systematic noise.

In Table 1 we list a short summary of the above CMB experiments. In each case we list only the best frequency channel for measuring the spectra. Here, $\theta_{\rm fwhm}$ is the beam size diffraction limit with full width at half maximum (fwhm); σ_T is the thermodynamic temperature sensitivity per pixel (a square the side of which is the fwhm extent of the beam); σ_P likewise for the polarization quantities Q and U; $\ell^T_{\rm cvlim}$ is the multipole at which the temperature noise spectrum equals cosmic variance; $\ell^{T,E}_{\rm max}$ is the multipole for which the TT or EE spectrum equals the instrument (statistical) noise (i.e., $S/N = 1$). For both SPTPol and ACTPol we assume a net 9-h per day observation time.

[1] There is a correction to the proposed fits for $\beta \gtrsim 0.1$: the fits get much better as $\beta \to 1$ (but not exact) if one replaces β by $\beta/(1-\beta^2)^{1/4}$ in the argument of the Bessel functions. Note that in practice, however, such a correction is irrelevant as we know that $\beta \sim 10^{-3}$.

Table 1 Summary of selected CMB experiments

Experiment	Best ν (GHz)	σ_T	σ_P	$\theta_{\rm fwhm}$	$f_{\rm sky}$ (%)	$\ell^T_{\rm cvlim}$	$\ell^T_{\rm max}$	$\ell^E_{\rm max}$
WMAP (9 years)	94	14	20	13.2′	78	600	900	5
Planck (30 months)	143	1.0	1.7	7.2′	80	1800	2500	1700
ACTPol Wider (4 yrs)	150	2.5	3.5	1.4′	40	3600	4600	3300
COrE (4 years)	225	0.2	0.36	4.7′	80	3100	3700	3000
EPIC 4K	220	0.24	0.34	3.8′	80	3400	4300	3300
EPIC 30K	220	0.56	0.79	3.8′	80	3000	3900	3000

Values quoted stand for the optimal frequency band with respect to a combination of temperature and angular sensitivities. σ_T is the thermodynamic temperature sensitivity per pixel (in units of $10^6\mu$K/K); σ_P likewise for the polarization quantities E and B; $\ell^T_{\rm cvlim}$ is the approximate multipole at which the temperature noise spectrum equals cosmic variance; $\ell^{T,E}_{\rm max}$ is the approximate multipole for which the TT or EE spectrum equals the instrument (statistical) noise. Note that σ_T, $\theta_{\rm fwhm}$, $f_{\rm sky}$, $\ell^T_{\rm cvlim}$ and $\ell^X_{\rm max}$ are related by Eq. (13) [Table from Notari and Quartin (2012)]

The quantities $\theta_{\rm fwhm}$, $\ell^T_{\rm cvlim}$, $f_{\rm sky}$ and σ_T are related by the expression for the noise power-spectrum (see e.g. Dodelson 2003):

$$\Delta C_\ell = \sqrt{\frac{2}{f_{\rm sky}(2\ell+1)}}\Big[C_\ell + N_\ell\Big], \tag{12}$$

where the first term stands for the cosmic variance (CV_ℓ) and N_ℓ for the instrumental noise (see Knox 1995; Dodelson 2003; Perotto et al. 2006; Nolta et al. 2009):

$$N_\ell = \theta^2_{\rm fwhm}\sigma^2_T \exp\left[\ell(\ell+1)\frac{\theta^2_{\rm fwhm}}{8\ln 2}\right]. \tag{13}$$

Cosmic variance is predominant at lower and intermediate values of ℓ (to wit for $\ell < \ell^T_{\rm cvlim}$); instrumental noise dominates for $\ell > \ell^T_{\rm cvlim}$ and determines $\ell^X_{\rm max}$. Note that in the case of Earth experiments the estimate (13), sometimes referred to as the Knox formula, is inaccurate for $\ell \lesssim 500$ due to possible atmospheric fluctuations (Niemack et al. 2010). Note also that we list sensitivities per pixel in $\mu K/K$, but some of the references here listed prefer to describe them as either noise-equivalent temperatures (NET) in $\mu K_{\rm CMB}\sqrt{s}$ (for a single detector—e.g. one bolometer) or as sensitivity in $\mu K\cdot$arcmin (for a given frequency band, with all detectors in that band combined). To convert between these quantities one has to make use of the following relations (Bock et al. 2009):

$$\sigma(\mu K\cdot{\rm arcmin}) = \sqrt{\frac{8\pi f_{\rm sky}\big[{\rm NET}(\mu K_{\rm CMB}\sqrt{s})\big]^2}{t_{\rm mission}(s)\,N_{\rm detectors}}}\,\frac{10,800}{\pi}, \tag{14}$$

$$\sigma\left(\frac{\mu K}{K}\right) = \frac{\sigma(\mu K\cdot{\rm arcmin})}{2.725\,K\,\theta_{\rm fwhm}({\rm arcmin})}. \tag{15}$$

Also worth noticing is the relation between σ_T and $\sigma_{E;B}$. If both linear polarization states are given equal integration times (as is usually the case), the total number of photons available for the temperature measurement will be twice the number available for either polarization measurement, and one has $\sigma_{E;B} = \sqrt{2}\sigma_T$ (Kamionkowski et al. 1997). This relation, however, assumes that all detectors (being either coherent receivers, like HEMT amplifiers, or incoherent ones, like bolometers) in the experiment are sensitive to polarization. If that is not the case, then one has instead the relation

$$\sigma_{E;B} = \sqrt{2\frac{N_{\text{total}}}{N_{\text{polar}}}}\,\sigma_T, \tag{16}$$

where N_{total} stands for the total number of detectors and $N_{\text{pol}} \leq N_{\text{total}}$ is the number of polarization detectors.

Finally, the measurements in the different frequency bands are formally independent, so one can combine different channels to get lower instrumental noises (cosmic variance is completely correlated in all frequency bands and therefore cannot be mitigated) and thus probe higher multipoles. For Planck, as an example, combining two channels (143 and 217 GHz) allows us to increase ℓ^T_{max} by around 20 %. However, more important than an increase in ℓ^T_{max}, the lower instrumental noise after combining all channels means that the correlations here addressed can be detected with higher S/N ratios. To wit, combining the different bands one gets

$$N_\ell^{\text{comb}} = \left[\sum_{i=1}^{\#\text{bands}} \left(N_\ell^i\right)^{-2}\right]^{-1/2}, \tag{17}$$

where N_ℓ^i is the instrument noise spectra for the i-th frequency band.

One must nevertheless be aware that different channels are subject to largely different foregrounds and systematics, so that actually the usable f_{sky} differs for each frequency range. In what follows, for simplicity, we will assume a fixed f_{sky} in all frequency bands for each experiment.

3.2 Detectability of Our Proper Motion

In this subsection we seek to answer the following question: for a given CMB experiment, what is the smallest value of β_{res} it can in principle detect (say, $S/N = 1$ or 3)?

Using the induced off-diagonal $a_{\ell m}$ correlations, an estimator for our peculiar velocity was built in Amendola et al. (2011). We summarize below the procedure. We first define the following basic quantities:

$$F^{XY}_{\ell m} \equiv a^{X*}_{\ell\, m} a^{Y}_{\ell+1\, m}\,, \tag{18}$$

in the appropriate frame, where X, Y stands for either temperature (T) or one of the polarization channels (E, B). The useful quantities are the real part of the above (the imaginary part has zero average)

$$f_{\ell m}^{XY} \equiv \frac{1}{2}(F_{\ell m}^{XY} + F_{\ell -m}^{XY})\,. \tag{19}$$

Note that $f_{\ell m}^{XY} = f_{\ell -m}^{XY}$. Given a peculiar velocity $\beta_{\rm res}$, we can predict the average value of the $f_{\ell m}^{XY\prime}s$, which are given by Amendola et al. (2011):

$$\langle f_{\ell m}^{XY}\rangle \;=\; c_{\ell+1m}^{-Y} C_{\ell}^{XY} + c_{\ell m}^{+X} C_{\ell+1}^{XY}\,, \tag{20}$$

The $c_{\ell m}^{\pm X}$ coefficients are different in the case of temperature and polarization and are given by Eq. (7); to wit:

$$\begin{aligned} c_{\ell m}^{+X} &= \beta_{\rm res}(\ell+1)\;{}_sG_{\ell+1\,m}\,,\\ c_{\ell m}^{-X} &= -\beta_{\rm res}\,\ell\;{}_sG_{\ell\,m}\,, \end{aligned} \tag{21}$$

where $s = 0$ for temperature and $s = 2$ for polarization, and where we made use of Eq. (6). Note that in the limit of flat spectra ($C_{\ell}^{XY} = C_{\ell+1}^{XY}$), one has $\langle f_{\ell m}^{XY}\rangle = 0$. One can show, using Eq. (8), that this remains true up to $\mathcal{O}\big(\beta^3\big)$. Therefore $\langle f_{\ell m}^{XY}\rangle$ is generically higher the more wiggled the power spectrum is.

Since $\langle f_{\ell m}^{XY}\rangle$ is proportional to $\beta_{\rm res}$, it is useful to define a related quantity

$$\langle f_{\ell m}^{XY}\rangle \;\equiv\; \beta_{\rm res}\,\langle \hat{f}_{\ell m}^{XY}\rangle\,. \tag{22}$$

From these predictions, we can compute an estimator

$$\hat{\beta} = \left(\sum_{X,\ell,m} \frac{f_{\ell m}^{XY,\,\rm obs}\langle \hat{f}_{\ell m}^{XY}\rangle}{\mathfrak{C}_{\ell}\mathfrak{C}_{\ell+1}}\right)\left(\sum_{X,\ell,m} \frac{\langle \hat{f}_{\ell m}^{XY}\rangle^2}{\mathfrak{C}_{\ell}\mathfrak{C}_{\ell+1}}\right)^{-1}\,, \tag{23}$$

where (see Amendola et al. 2011; Perotto et al. 2006; Verde 2010)

$$\mathfrak{C}_{\ell} \;\equiv\; \frac{1}{\sqrt{f_{\rm sky}}}\big(C_{\ell} + N_{\ell}\big) \tag{24}$$

and the sums are in principle carried out with $-\ell \le m \le \ell$, but can also be simplified to a sum on $0 \le m \le \ell$ since $f_{\ell m}^{XY} = f_{\ell -m}^{XY}$. An approximate value for the variance of $\hat{\beta}_{\rm res}$ can also be written as (Amendola et al. 2011)

$$\delta\beta \;\equiv\; \sqrt{\langle \hat{\beta}^2\rangle} \approx \left[\sum_{X}\sum_{\ell}\sum_{m=-\ell}^{\ell} \frac{\langle \hat{f}_{\ell m}^{XY}\rangle^2}{\mathfrak{C}_{\ell}\mathfrak{C}_{\ell+1}}\right]^{-\frac{1}{2}} \;=\; \beta_{\rm res}\left(\frac{N}{S}\right). \tag{25}$$

This leads to an important conclusion: the estimate of $\delta\beta$ does not depend on the value of $\beta_{\rm res}$ itself. Note that this is physically motivated by the simple fact that the error in β is mainly due to cosmic variance which, to leading order, is independent on β. Therefore, one can determine the achievable precision of a given experiment without knowledge of the residual velocity $\beta_{\rm res}$. Likewise, one can also determine what is the minimal value $\beta_{\rm res}$ that a given experiment could detect (for a given S/N).

Finally, it was also shown in Amendola et al. (2011) that the estimate on the direction of the velocity is directly related to the one of the magnitude. To wit if one chooses the fiducial velocity as the z-axis, for small angles the error on the absolute value of the velocity and the error in the magnitude of the direction are related by

$$\delta\theta = \frac{\delta\beta}{\beta_{\rm res}}. \tag{26}$$

Note that although as stated above the achievable precision in $\delta\beta$ does not depend on the fiducial value of the velocity (and thus does not depend whether one pre-deboosts the CMB or not), the same is not true for $\delta\theta$, so clearly the direction of the residual velocity will be measurable only if $\beta_{\rm res}$ is significantly larger than $\delta\beta$ for a given experiment.

Figure 1 depicts the expected precision of some of the CMB experiments listed in Table 1 as a function of the highest multipole ℓ taken into account. Here we made use of CAMB (Lewis 2000) and took its default cosmological parameters as our

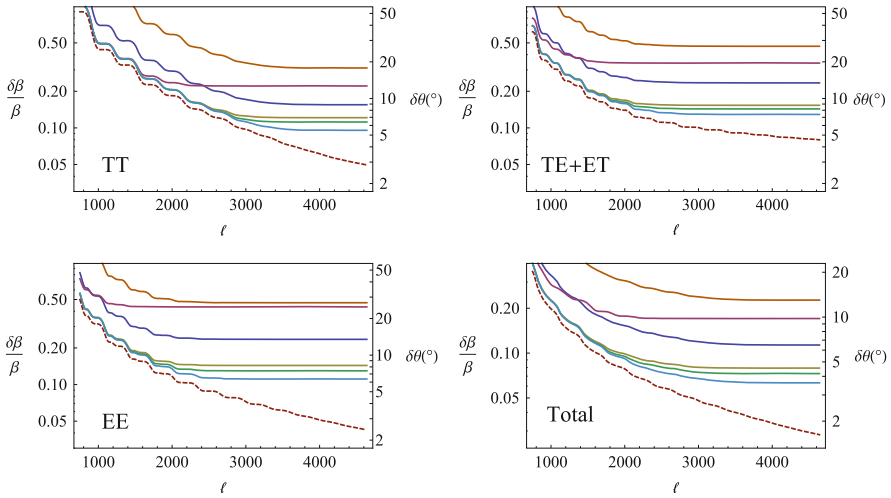

Fig. 1 Precision in the measurement of $\delta\beta$ achieved by different experiments using the temperature T, the E mode of polarization the TE cross power-spectra and the combination of all these. Here we still assume $\beta_{\rm res} = 1.231 \times 10^{-3}$ (i.e., no pre-deboosting). *Top to bottom*: ACTPol Wide (1 year); Planck (30 months); ACTPol Wider (4 years); EPIC (30K); Core (4 years); EPIC (4K) and finally, represented by the *dashed line*, an ideal experiment (no noise) [Figure from Notari and Quartin (2012)]

fiducial ones. Note that even the best experiments (Core and EPIC 4K) fall short of the (arguable) target of $\sim$ 1 % fractional precision on the velocity. Nonetheless, their precision would be able to detect a non-standard signal which amounts to at least 8 % in the CMB dipole. It would take an even better experiment, with sensitivities and beam sizes which are respectively high and small enough to probe temperature and polarization multipoles up to $\ell \simeq 5000$ to achieve the level of 2 %. Note however that going to smaller and smaller scales the signal gets more and more contaminated by local sources, and if they are not taken properly into account this could prevent using this part of the spectrum, because it could introduce an additional preferred direction in the data.

The bottomline is that Planck can measure our peculiar velocity v with a precision of 55 km/s [completely consistent with what was actually measured later in Aghanim et al. (2014)], whereas Core and EPIC 4K/30K could do the same with only 20–25 km/s of error. Finally, an ideal experiment probing temperature and polarization multipoles up to $\ell \simeq 5000$ could achieve $\delta v \simeq 8$ km/s.

The results in Sect. 2 allow us to estimate the accuracy of the first order expansion we use here, as $\beta_{\rm res}\ell$ approaches unity. A direct estimate would be obtained by comparing the non-linear fitting functions (7) with a linear expansion. However, a more careful comparison has to be done at the level of the $\langle f_{\ell m}\rangle$, because of the leading-order cancellations, which could be disrupted by small changes in the coefficients. We therefore evaluated, using Eq. (8), $\langle f_{\ell m}\rangle$ up to third order in β (second order contributions are identically zero for the $\{\ell, \ell + 1\}$ correlations) and compared the result with the first order $\langle f_{\ell m}\rangle$. Amazingly, the leading order cancelation is basically undisturbed, and the correction to $\langle f_{\ell m}\rangle$ is found to be small even for high values of $\beta_{\rm res}$: it is ~ 1 % for $\beta_{\rm res} = 5 \times 10^{-4}$.

4 Power Anomalies

Besides introducing a new signal in the sky, aberration can also introduce a systematic noise, if not corrected for. In fact, it was shown in Notari et al. (2014), Quartin and Notari (2015) that it generates power anomalies in the CMB. In particular, aberration and Doppler were shown to be responsible for both hemispherical power asymmetries and for a dipolar modulation of power of the spectrum. In a nutshell, the hemispherical power asymmetry is related to the fact that in two hemispheres in the sky (usually labelled "north" and "south"), the spectrum looks statistically different (mostly in their overall amplitude). The dipolar power modulation is a related statistical anisotropy, described by an ad-hoc model in which the amplitude of the power spectrum varies smoothly in the sky in a way described by a constant summed to a cosine function. These anomalies have been studied and discussed for a full decade now (see, e.g. Eriksen et al. 2004, 2007; Hansen et al. 2004a,b; Bernui et al. 2008; Hansen 2009; Hoftuft et al. 2009; Paci et al. 2013; Ade et al. 2013; Akrami et al. 2014). See Ade et al. (2013) for a more detailed discussion on these and other "anomalies" of the CMB.

When analysing the CMB data for power anomalies, it is very important to take three factors into account: the effect of the mask used, the asymmetry of the noise and the boost effects. A mask is usually defined in pixel space and it consists of a Boolean map of 1's and 0's. On the edge of the mask this leads to sharp transitions, which result in high-frequency modes in harmonic space. Therefore, such a mask introduces spurious high-ℓ modes in the observed spectrum (the observed spectrum is a convolution of the primordial one with the mask power spectrum).

For each of the above mentioned three directions we finally compute the spectra $D_\ell^{(N,S)} \equiv C_\ell^{(N,S)}\ell(\ell+1)/(2\pi)$, and depict the difference

$$\frac{\delta D_\ell}{D_\ell} \equiv \frac{D_\ell^N - D_\ell^S}{D_\ell^{\text{average}}} \equiv 2\,\frac{D_\ell^N - D_\ell^S}{D_\ell^N + D_\ell^S}\,, \tag{27}$$

in Fig. 2, comparing with mean and standard deviation of the same quantity for the simulations. Note that for simulations the D_ℓ's contain also the boost effects and the noise power in each hemisphere. All results are obtained after binning the spectra in 50-ℓ bins. We also depict (dashed black curve) the average bias due to Doppler and aberration, which is oscillating and has non-zero mean. Such bias was found in Notari et al. (2014) to be extremely well approximated by the analytical expression

$$\frac{\delta D_\ell}{D_\ell} \simeq 4\overline{\beta} + 2\overline{\beta}\,\ell\left(1 - \frac{D_{\ell+1}^{\text{th}}}{D_\ell^{\text{th}}}\right), \tag{28}$$

where D_ℓ^{th} is the fiducial power spectrum and $\overline{\beta}$ is the average

$$\overline{\beta} = \beta\langle\cos\gamma\rangle_R = \beta\int_R d\Omega\cos\gamma\,, \tag{29}$$

where in turn R is the region of interest integrated over the solid angle $d\Omega$ and γ is the angle relative to the boost direction. We can also see that the simulations at $\ell \gtrsim 1300$ start having a net positive bias and this is due to the anisotropic noise which has been added to them and which becomes important at such high ℓ.

In Fig. 2 we show the final results for three directions [respectively defined by the galactic axis, dipole direction and the maximal asymmetry direction found in Ade et al. (2013)] and include also the statistical significance of the asymmetry summing from $\ell = 2$ to different $\ell_{\max}$, in order to illustrate its evolution as we probe ever smaller scales. For a given D_ℓ the uncertainty is given by [see Hivon et al. (2002) or Eq. (11.27) in Dodelson (2003)]

$$\sigma_{D_\ell} = \sqrt{\frac{2}{(2\ell+1)f_{\text{sky}}}}(D_\ell^{\text{th}} + N_\ell^{\text{MC}})\,, \tag{30}$$

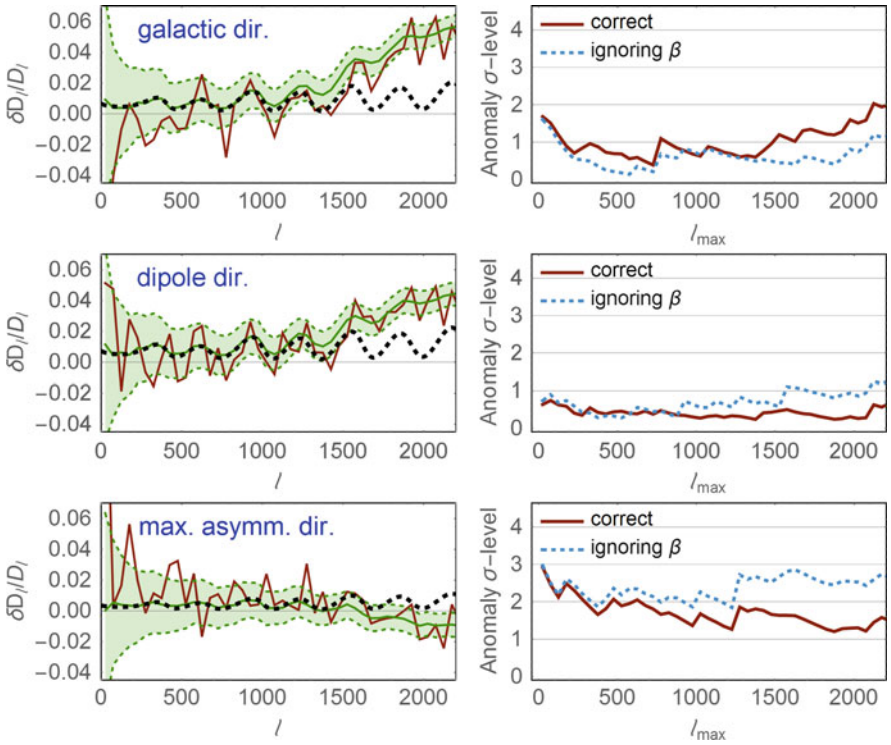

Fig. 2 Hemispherical asymmetry and its statistical significance along three different directions. *Left*: The relative difference between D_ℓ's in two opposite halves of the sky as a function of ℓ. The *brown line* is the Planck data while the *green curve* and band represent the mean and 1σ band from 500 simulations, binned in $50-\ell$ bins. The *black dashed curves* are the analytical estimate of Eqs. (28)–(29), which ignores all noise. *Right*: The corresponding statistical (in)significance of the anomaly, summing all multipoles from 2 to $\ell_{\max}$. The *brown curve* represents the correct estimate; the *dotted blue curve* is the spurious significance if one ignores to boost the simulations. Note that naively ignoring the boost can lead to a spurious anomaly at high ℓ between 2.5 and 2.9σ [Figure from Quartin and Notari (2015)]

where $N_\ell^{\rm MC}$ is the noise power spectrum (obtained as an average through Monte Carlo simulations) multiplied by $\ell(\ell+1)/(2\pi)$. As it can be seen, after taking into account all systematics, the anomaly level is consistently below 2σ.

For the case of the dipolar modulation, it was also shown in Quartin and Notari (2015) that the proper accounting for the boost also reduces the so-called anomaly from 3σ (or even 5σ, in case of a posteriori statistics) to less than 1σ, when all scales are considered.

5 Conclusions

The peculiar velocity of the solar system introduces both Doppler and aberration effects in the sky. For the CMB in particular, this can be used to measure our velocity through the aberration effect, opening a new window to measure the CMB rest frame.

On the other hand, if not properly taking into account, it can introduce spurious anomalies in the sky, which have no relation to any primordial or more fundamental physics. One should therefore either work on the CMB rest frame, or if not, properly and carefully take the peculiar velocity into account.

References

Ade, P.A.R., et al.: Planck 2013 results. XXIII. Isotropy and statistics of the CMB, v3. Astron. Astrophys. **571**, A23 (2014). arXiv:1303.5083

Aghanim, N., et al. Planck 2013 results. XXVII. Doppler boosting of the CMB: eppur si muove. Astron. Astrophys. **571**, A27 (2014). arXiv:1303.5087

Akrami, Y., Fantaye, Y., Shafieloo, A., Eriksen, H.K., Hansen, F.K., et al.: Power asymmetry in WMAP and Planck temperature sky maps as measured by a local variance estimator. Astrophys. J. **784**, L42 (2014). arXiv:1402.0870. doi:10.1088/2041-8205/784/2/L42

Amendola, L., Catena, R., Masina, I., Notari, A., Quartin, M., et al.: Measuring our peculiar velocity on the CMB with high-multipole off-diagonal correlations. J. Cosmol. Astropart. Phys. **1107**, 027 (2011). arXiv:1008.1183. doi:10.1088/1475-7516/2011/07/027

Bock, J., et al.: Study of the experimental probe of inflationary cosmology (EPIC)-intermediate mission for NASA's Einstein inflation probe (2009). arXiv:0906.1188

Bennett, C.L., Hill, R.S., Hinshaw, G., Larson, D., Smith, K.M., et al. Seven-year Wilkinson microwave anisotropy probe (WMAP) observations: are there cosmic microwave background anomalies? Astrophys. J. Suppl. **192**, 17 (2011). arXiv:1001.4758. doi:10.1088/0067-0049/192/2/17

Bernui, A.: Anomalous CMB north-south asymmetry. Phys. Rev. **D78**, 063531 (2008). arXiv:0809.0934. doi:10.1103/PhysRevD.78.063531

Burles, S., Rappaport, S.: Aberration of the cosmic microwave background. Astrophys. J. **641**, L1–L4 (2006). arXiv:astro-ph/0601559. doi:10.1086/503743

Catena, R., Notari, A.: Cosmological parameter estimation: impact of CMB aberration. J. Cosmol. Astropart. Phys. **1304**, 028 (2013). arXiv:1210.2731

Catena, R., Liguori, M., Notari, A., Renzi, A.: Non-Gaussianity and CMB aberration and Doppler. J. Cosmol. Astropart. Phys. **1309**, 036 (2013). arXiv:1301.3777. doi:10.1088/1475-7516/2013/09/036

Challinor, A., van Leeuwen, F.: Peculiar velocity effects in high-resolution microwave background experiments. Phys. Rev. **D65**, 103001 (2002). arXiv:astro-ph/0112457. doi:10.1103/PhysRevD.65.103001

Chluba, J.: Aberrating the CMB sky: fast and accurate computation of the aberration kernel. Mon. Not. R. Astron. Soc. **415**(4), 3227-3236 (2011). arXiv:1102.3415

Copi, C.J., Huterer, D., Schwarz, D.J., Starkman, G.D.: Large angle anomalies in the CMB. Adv. Astron. **2010**, 847541 (2010). arXiv:1004.5602

Dai, L., Chluba, J.: New operator approach to the CMB aberration kernels in harmonic space. Phys. Rev. **D89**, 123504 (2014). arXiv:1403.6117. doi:10.1103/PhysRevD.89.123504

Dodelson, S.: Modern Cosmology. Academic, New York (2003). http://home.fnal.gov/~dodelson/book.html

Eriksen, H.K., Hansen, F.K., Banday, A.J., Gorski, K.M., Lilje, P.B.: Asymmetries in the cosmic microwave background anisotropy field. Astrophys. J. **605**, 14–20 (2004). arXiv:astro-ph/0307507. doi:10.1086/382267

Eriksen, H.K., Banday, A.J., Gorski, K.M., Hansen, F.K., Lilje, P.B.: Hemispherical power asymmetry in the three-year Wilkinson microwave anisotropy probe sky maps. Astrophys. J. **660**, L81–L84 (2007). arXiv:astro-ph/0701089. doi:10.1086/518091

Hansen, F.K., Cabella, P., Marinucci, D., Vittorio, N.: Asymmetries in the local curvature of the WMAP data. Astrophys. J. **607**, L67–L70 (2004a). arXiv:astro-ph/0402396. doi:10.1086/421904

Hansen, F.K., Banday, A.J., Gorski, K.M.: Testing the cosmological principle of isotropy: local power spectrum estimates of the WMAP data. Mon. Not. R. Astron. Soc. **354**, 641–665 (2004b). arXiv:astro-ph/0404206. doi:10.1111/j.1365-2966.2004.08229.x.

Hansen, F.K., Banday, A.J., Gorski, K.M., Eriksen, H.K., Lilje, P.B.: Power asymmetry in cosmic microwave background fluctuations from full sky to sub-degree scales: Is the universe isotropic? Astrophys. J. **704**, 1448–1458 (2009). arXiv:0812.3795. doi:10.1088/0004-637X/704/2/1448

Hinshaw, G., et al.: Five-year Wilkinson microwave anisotropy probe (wmap) observations: data processing, sky maps, and basic results. Astrophys. J. Suppl. **180**, 225–245 (2009). arXiv:0803.0732. doi:10.1088/0067-0049/180/2/225

Hivon, E., Gorski, K.M., Netterfield, C.B., Crill, B.P., Prunet, S., et al.: Master of the cosmic microwave background anisotropy power spectrum: a fast method for statistical analysis of large and complex cosmic microwave background data sets. Astrophys. J. **567**, 2 (2002). arXiv:astro-ph/0105302. doi:10.1086/338126

Hoftuft, J., Eriksen, H.K., Banday, A.J., Gorski, K.M., Hansen, F.K., et al.: Increasing evidence for hemispherical power asymmetry in the five-year WMAP data. Astrophys. J. **699**, 985–989 (2009). arXiv:0903.1229. doi:10.1088/0004-637X/699/2/985

Kamionkowski, M., Kosowsky, A., Stebbins, A.: Statistics of cosmic microwave background polarization. Phys. Rev. **D55**, 7368–7388 (1997). arXiv:astro-ph/9611125. doi:10.1103/PhysRevD.55.7368

Knox, L.: Determination of inflationary observables by cosmic microwave background anisotropy experiments. Phys. Rev. **D52**, 4307–4318 (1995). arXiv:astro-ph/9504054. doi:10.1103/PhysRevD.52.4307

Kosowsky, A., Kahniashvili, T.: The signature of proper motion in the microwave sky. Phys. Rev. Lett. **106**, 191301 (2011). arXiv:1007.4539. doi:10.1103/PhysRevLett.106.191301

Lewis, A., Challinor, A., Lasenby, A.: Efficient computation of CMB anisotropies in closed FRW models. Astrophys. J. **538**, 473–476 (2000). http://camb.info. arXiv:astro-ph/9911177. doi:10.1086/309179

Lineweaver, C.H., et al.: The dipole observed in the cobe dmr four-year data. Astrophys. J. **470**, 38–42 (1996). arXiv:astro-ph/9601151. doi:10.1086/177846

Mather, J.C., Fixsen, D.J., Shafer, R.A., Mosier, C., Wilkinson, D.T.: Calibrator design for the COBE far infrared absolute spectrophotometer (FIRAS). Astrophys. J. **512**, 511–520 (1999). arXiv:astro-ph/9810373. doi:10.1086/306805

Niemack, M.D., Ade, P.A.R., Aguirre, J., Barrientos, F., Beall, J.A., et al.: ACTPol: a polarization-sensitive receiver for the Atacama Cosmology Telescope. Proc. SPIE Int. Soc. Opt. Eng. **7741**, 77411S (2010). arXiv:1006.5049

Nolta, M.R., et al.: Five-year Wilkinson microwave anisotropy probe (WMAP) observations: angular power spectra. Astrophys. J. Suppl. **180**, 296–305 (2009). arXiv:0803.0593. doi:10.1088/0067-0049/180/2/296

Notari, A., Quartin, M.: Measuring our peculiar velocity by 'pre-deboosting' the CMB. J. Cosmol. Astropart. Phys. **1202**, 026 (2012). arXiv:1112.1400. doi:10.1088/1475-7516/2012/02/026

Notari, A., Quartin, M., Catena, R.: CMB aberration and Doppler effects as a source of hemispherical asymmetries. J. Cosmol. Astropart. Phys. **1403**, 019 (2014). arXiv:1304.3506. doi:10.1088/1475-7516/2014/03/019

Paci, F., Gruppuso, A., Finelli, F., De Rosa, A., Mandolesi, N., et al.: Hemispherical power asymmetries in the WMAP 7-year low-resolution temperature and polarization maps. Mon. Not. R. Astron. Soc. **434**, 3071 (2013). arXiv:1301.5195

Pereira, T.S., Yoho, A., Stuke, M., Starkman, G.D.: Effects of a cut, lorentz-boosted sky on the angular power spectrum (2010). arXiv:1009.4937

Perotto, L., Lesgourgues, J., Hannestad, S., Tu, H., Wong, Y.Y.Y.: Probing cosmological parameters with the CMB: forecasts from full Monte Carlo simulations. J. Cosmol. Astropart. Phys. **0610**, 013 (2006). arXiv:astro-ph/0606227. doi:10.1088/1475-7516/2006/10/013

Quartin, M., Notari, A.: On the significance of power asymmetries in Planck CMB data at all scales. J. Cosmol. Astropart. Phys. **1501**(01), 008 (2015). arXiv:1408.5792

Verde, L.: Statistical methods in cosmology. Lect. Notes Phys. **800**, 147–177 (2010). arXiv:0911.3105. doi:10.1007/978-3-642-10598-2_4

Warm Inflation, Cosmological Fluctuations and Constraints from Planck

Rudnei O. Ramos

Abstract Nonisentropic (warm) inflation models are characterized by radiation production due to the decay of fields coupled to the inflaton during inflation. The continuous radiation production might sustain a thermal radiation bath during inflation as a result of the dissipative particle production. The presence of the radiation bath can impact on the dynamics of inflation and, consequently, on the observable quantities measured from the cosmic microwave background radiation (CMBR). Besides of dissipative effects, these are also accompanied by stochastic fluctuations. Both the origin and the impact of these effects on the inflationary dynamics are reviewed here.

1 Introduction

Inflation is the most acceptable paradigm that solves the flatness and horizon problems of the standard Big-Bang cosmological model (Liddle and Lyth 2000). We generically define inflation as an early accelerated expansion dominated by vacuum energy density and driven by a fluid with negative pressure:

$$\text{Inflation} \quad \Rightarrow \quad \ddot{a} > 0, \;\; p < -\rho/3.$$

Besides of solving the flatness and horizon problems of the hot Big-Bang theory, inflation also provides a solution of how inhomogeneities can originate, thus giving a mechanism through which large-scale structures can form. Inflation is typically driven by a scalar field, the inflaton field. Density perturbations are sourced by perturbations of this inflaton field, which can be either of quantum and/or thermal origin. Inflation also solves some outstanding problems related to Grand-Unified theories (GUT), like the problem of dangerous heavy relics (e.g., magnetic monopoles).

R.O. Ramos (✉)
Departamento de Física Teórica, Universidade do Estado do Rio de Janeiro, 20550-013 Rio de Janeiro, RJ, Brazil
e-mail: rudnei@uerj.br

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_8

In the standard scenario of inflation, which we call *cold inflation*, the interactions of the inflaton field with other field degrees of freedom are negligible during inflation. In this case the universe enters in a vacuum dominated phase and any previous (if any) initial radiation energy density, ρ_r, quickly redshifts away. In cold inflation, the inflationary phase must end with a reheating phase, through which the universe enters in the radiation dominated phase, such that the standard Big-Bang cosmological evolution follows. The standard theory of inflation, as a bonus, predicts that the large scale distribution of galaxies can be traced back to quantum vacuum fluctuations of a weakly coupled field, the inflaton, during the inflationary era (Liddle and Lyth 2000).

However, there can be regimes of parameters such that the inflaton interactions with other field degrees of freedom are not negligible. It then happens that these interactions can generate dissipation terms, such that a small fraction of vacuum energy density can be converted to radiation. If the magnitude of these dissipation terms are strong enough to compensate the redshift of the radiation by the expansion, then a steady state can be produced, with the inflationary phase happening in a thermalized radiation bath. This scenario is called *warm inflation* (WI) [for a recent review see Berera et al. 2009]. In warm inflation, the evolution equation for the radiation gets modified to $\dot{\rho}_r + 4H\rho_r = \Upsilon\dot{\phi}^2$, where $H = \dot{a}/a$ and Υ is a dissipation term that can be a function of both the inflaton field ϕ and temperature T. Typically, in warm inflation it is assumed that $T \gtrsim H$, in which case thermal fluctuations dominate over the quantum ones. As a consequence, density fluctuations are now sourced by thermal fluctuations as opposite to quantum fluctuations in the cold inflation scenario.

WI dynamics can be viewed as an analogous of that of open systems. Dissipative dynamics along with stochastic forces are typical manifestations for the dynamics of a system interacting with some large environment. We can think of such similar dynamics to also manifest in the early universe. For instance, when the matter content of the universe can be split into a subsystem interacting with a large energy reservoir, then physical processes may be represented through effective dissipation and stochastic noise terms.

The presence of a radiation bath in WI can impact on the dynamics of inflation and, consequently, on the observable quantities measured from the cosmic microwave background radiation. In particular, the amplitude of primordial curvature perturbations is enhanced and this is particularly significant when a non-trivial statistical ensemble of inflaton fluctuations is also maintained. Since gravitational modes are decoupled from the radiation bath for energies well below the Planck scale, the presence of the thermal radiation bath and/or a non vanishing statistical ensemble for the inflaton generically lowers the tensor-to-scalar ratio and yields a modified consistency relation for warm inflation, as well as changing the tilt of the scalar spectrum. This is able to alter the landscape of observationally allowed inflationary models, with for example the quartic chaotic potential being in very good agreement with the Planck results for nearly-thermal inflaton fluctuations, whilst essentially ruled out for an underlying vacuum state. Besides of dissipative effects, these are also accompanied by stochastic fluctuations. Both the origin and

the impact of these effects on the inflationary dynamics is reviewed in the next sections. We show how the dissipative and stochastic forces associated are able to alter the landscape of observationally allowed inflationary models, with for example the quartic chaotic potential being in very good agreement with the Planck results for nearly-thermal inflaton fluctuations. We also connect the results obtained from warm inflation with the recent BICEP2 measurement of a B-mode in the CMBR, believed to come from gravitational waves of primordial origin, and how warm inflation can conciliate both Planck and BICEP2 results.

2 Warm Inflation Dynamics

Warm inflation is defined in terms of the effective evolution equation for the inflaton field while averaging out (functionally integrating over) the other field degrees of freedom. This produces a stochastic (Langevin-like) equation of motion for the inflaton which is of the form (Ramos and Silva 2013)

$$\ddot{\phi}(\mathbf{x},t)+3H\dot{\phi}(\mathbf{x},t)+\int d^4x'\Sigma_R(x,x')\phi(x')+V_{,\phi}-\frac{1}{a^2}\nabla^2\phi(\mathbf{x},t)=\xi_q+\xi_T\ , \qquad (1)$$

where $\Sigma_R(x,x')$ is a self-energy contribution and ξ_q and ξ_T are stochastic fields, with Gaussian distributions, describing quantum and thermal (noise) fluctuations, respectively. They satisfy appropriate fluctuation and dissipation relations. For example, two-point correlation function for ξ_T is related to thermal part of $\Sigma_R(x,x')$. The self-energy term can be put in the form of a dissipative term in the adiabatic approximation, $\dot{\phi}/\phi, H, \dot{T}/T < \Gamma$, where Γ is a decay rate term. In the most recent implementations of warm inflation (Berera et al. 2009), Γ is the decay width of a heavy scalar field coupled to the inflaton field and that can decay into light radiation fields. Under the adiabatic approximation, we can approximate the self-energy term in Eq. (1) as $\int d^4x'\Sigma_R(x,x')\phi(x') \approx \Upsilon\dot{\phi}(\mathbf{x},t)$ and the two-point correlation function for the stochastic source ξ turns into a local correlation function: $\langle\xi_T(\mathbf{x},t)\xi_T(\mathbf{x}',t')\rangle = a^{-3}\Upsilon T\delta(\mathbf{x}-\mathbf{x}')\delta(t-t')$. Likewise, the two-point correlation function for the quantum noise term can be expressed as $\langle\xi_q(\mathbf{x},t)\xi_q(\mathbf{x}',t')\rangle = H^2\,(1+2\mathcal{N})\,/2\,a^{-3}\delta(\mathbf{x}-\mathbf{x}')\delta(t-t')$, where $\mathcal{N}$ accounts for the possibility of a non-trivial distribution of inflaton particles, produced as a result of the dissipative processes occurring during inflation. For sufficiently fast interactions, this is expected to approach a Bose-Einstein distribution: $\mathcal{N}\approx n_{\rm BE}(k)=1/[\exp(k/aT)-1]$.

The dissipation coefficient Υ is found (Bastero-Gil et al. 2011, 2013) to have the following generic power law dependence with ϕ and T

$$\Upsilon = C_\phi\frac{T^c}{\phi^{c-1}}, \qquad (2)$$

where the value of the power c dependent on the specifics of the model construction for WI and on the temperature regime of the thermal bath. Typically, it is found that $c = 3$ (low temperature), $c = -1$ (high temperature) or $c = 0$ (constant dissipation). For example, $c = 3$ corresponds to the case where the inflaton interacts with a heavy (scalar) boson field, which in turn decays into light scalars. This is the case we use throughout this work, unless otherwise specified. The effectiveness of WI can be parametrized by the ratio $Q \equiv \Upsilon/3H$. The strong dissipative regime for WI is for $Q \gg 1$, while for $Q \ll 1$, it is the weak dissipative regime for WI.

As usual, we can study the dynamics by splitting the inflaton field in a background homogeneous part and perturbations, $\phi(\mathbf{x}, t) = \phi(t) + \delta\varphi(\mathbf{x}, t)$. The background quantities $\phi(t)$ and the radiation energy density $\rho_r(t)$ satisfy

$$\ddot{\phi} + (3H + \Upsilon)\dot{\phi} + V_{,\phi} = 0, \tag{3}$$

$$\dot{\rho}_r + 4H\rho_r = \Upsilon\,\dot{\phi}^2\,, \tag{4}$$

$$3H^2 = 8\pi G\rho\,. \tag{5}$$

Prolonged inflation requires the slow-roll conditions $|\epsilon_X| \ll 1$, where $\epsilon_X = -d\ln X/Hdt$, and X is any of the background field quantities. These slow-roll coefficients in WI satisfy

$$\begin{aligned}
\epsilon &= \frac{m_{\rm P}^2}{2}\left(\frac{V_{,\phi}}{V}\right)^2 \ll 1 + Q\,,\\
\eta &= m_{\rm P}^2\left(\frac{V_{,\phi\phi}}{V}\right) \ll 1 + Q\,,\\
\beta &= m_{\rm P}^2\left(\frac{\Upsilon_{,\phi}V_{,\phi}}{\Upsilon V}\right) \ll 1 + Q\,,
\end{aligned} \tag{6}$$

where $m_{\rm P}$ is the reduced Planck mass, $m_P = 1/\sqrt{8\pi G} = 2.4 \times 10^{18}$GeV.

3 Perturbations and Connection with CMBR Measurable Quantities

From Eq. (1) in the local approximation, the equation of motion for the fluctuations $\delta\varphi(\mathbf{x}, t)$ at linear order (in Fourier momentum space) is given by

$$\ddot{\delta\varphi}(\mathbf{k}, t) + (3H + \Upsilon)\dot{\delta\varphi}(\mathbf{k}, t) + V''(\phi)\delta\varphi(\mathbf{k}, t) + \frac{k^2}{a^2}\delta\varphi(\mathbf{k}, t) = \tilde{\xi}_T(\mathbf{k}, t) + \tilde{\xi}_q(\mathbf{k}, t)\,. \tag{7}$$

The solution of this equation can be expressed in terms of a Green function and from it we define the power spectrum for the inflaton field perturbations,

$$P_{\delta\varphi}(z) = \frac{k^3}{2\pi^2}\int \frac{d^3k'}{(2\pi)^3}\langle\delta\varphi(\mathbf{k},z)\delta\varphi(\mathbf{k}',z)\rangle \ , \tag{8}$$

where $z = k/(aH)$. These perturbations at some scale k get frozen as soon these scales cross the horizon, $z_* = 1$ and are imprinted in the CMBR when they reenter the horizon again at the decoupling era. The power spectrum can then be related to the CMBR quantities like the amplitude Δ_R for the curvature perturbation (defined in terms of the gauge invariant curvature perturbation $\xi = -H\delta\varphi/\dot{\phi}$),

$$\Delta_R^2 = \frac{H^2}{\dot{\phi}^2}P_R = \Delta_R^2(k_0)\left(\frac{k}{k_0}\right)^{n_s-1} \ , \tag{9}$$

and the spectral index n_s, (N_e is the number of e-folds of inflation):

$$n_s - 1 = \frac{d\ln\Delta_R^2}{dN_e} = \frac{d\ln\Delta_R^2}{d\ln k} \ , \tag{10}$$

where N_e is the number of e-folds of inflation.

For a generic inflaton phase-space distribution at the time when observable CMB scales leave the horizon during inflation, $z = z_*$, the dimensionless power spectrum of curvature perturbations in WI found to be given by (Ramos and Silva 2013)

$$\Delta_R^2 = \left(\frac{H_*}{\dot{\phi}_*}\right)^2\left(\frac{H_*}{2\pi}\right)^2\left[1 + 2n_* + \left(\frac{T_*}{H_*}\right)\frac{2\sqrt{3}\pi Q_*}{\sqrt{3+4\pi Q_*}}\right] \ , \tag{11}$$

which yields the standard cold inflation result, $\left(H_*/\dot{\phi}_*\right)^2[H_*/(2\pi)]^2$, in the limit $n_*, Q_*, T_* \to 0$.

Another quantity that can be used to constrain the many different inflation models is the tensor to scalar curvature perturbation ratio, $r = \Delta_T^2/\Delta_R^2$. Inflation in general also predicts a spectrum of gravitational waves. Gravitational wave perturbations also gets amplified during inflation, just like vacuum and thermal (in WI) perturbations of the inflaton field. Gravity waves are weakly coupled to the thermal bath and the spectrum of tensor modes retains its vacuum form, $\Delta_T^2 = (2/\pi^2)(H_*^2/m_P^2)$. This therefore suppresses the tensor-to-scalar ratio, yielding a modified consistency relation in the case of WI,

$$r \simeq \frac{8|n_t|}{1 + 2n_* + 2\pi Q_* T_*/H_*} \ , \tag{12}$$

where $n_t = -2\epsilon_*$ is the tensor index, while the spectral index n_s in WI is (Bartrum et al. 2014)

$$n_s - 1 \simeq 2\eta_* - 6\epsilon_* + \frac{2\kappa_*}{1+\kappa_*}\left(7\epsilon_* - 4\eta_* + 5\sigma_*\right) , \tag{13}$$

where $\sigma = m_P^2 V'/(\phi V) < 1 + Q$ and we have used the slow-roll equations, $3H(1 + Q)\dot{\phi} \simeq -V'(\phi)$ and $\rho_R \simeq (3/4)Q\dot{\phi}^2$, to determine the variation of $\kappa \equiv 2\pi QT/H$ as different scales become super-horizon during inflation.

Modifications are, however, more prominent in the opposite limit of nearly-thermal inflaton fluctuations, with $n_* \simeq n_{BE*}$. For $T_* \gtrsim H_*$ and $Q_* \ll 1$ we then obtain:

$$n_s - 1 \simeq 2\sigma_* - 2\epsilon_* , \tag{14}$$

which is, in particular, independent of the curvature of the potential, which only determines its running:

$$n_s' \simeq 2\sigma_*(\sigma_* + 2\epsilon_* - \eta_*) - 4\epsilon_*(2\epsilon_* - \eta_*) . \tag{15}$$

In this case, a red-tilted spectrum, $n_s < 1$, corresponds to either potentials with a negative slope, such as hill-top models, or large field models where $\epsilon_* > 2(m_P/\phi_*)^2$.

The above results for r and n_s in WI also easily goes to the cold inflation result when $n_*, Q_*, T_* \to 0$. In cold inflation (see, e.g., Liddle and Lyth 2000), it is found that $r = 16\varepsilon$, $n_s = 1 + 2\eta - 6\varepsilon$.

The observed amplitude of curvature perturbations from the recent measurements from Planck (Ade et al. 2013) gives $\Delta^2_{\mathcal{R}} \simeq 2.2 \times 10^{-9}$. The recently released Planck results also give for the tensor to scalar amplitude ratio the result $r < 0.11$ at 95 % CL (when the high-ℓ CMB ACT+SPT data are added) and for the spectral index $n_s = 0.9600 \pm 0.0072$, while when including the Planck lensing likelihood gives $n_s = 0.9653 \pm 0.0069$ and $r < 0.13$, and by also adding BAO data, it gives $n_s = 0.9643 \pm 0.0059$ and $r < 0.12$.

Let us show some of the results that are obtained from the above expressions, which were derived in the context of WI. For that, we specialize in the particular case of a chaotic quartic inflation potential,

$$V(\phi) = \lambda\phi^4 . \tag{16}$$

In the cold inflation case, it is known that such potential for inflation is not consistent with the present data, since it predicts a tensor-to-scalar ratio r and spectral index n_s outside of the allowed region as recently determined by the results from Planck (Ade et al. 2013).

Our results for WI (Bartrum et al. 2014) are shown in Fig. 1.

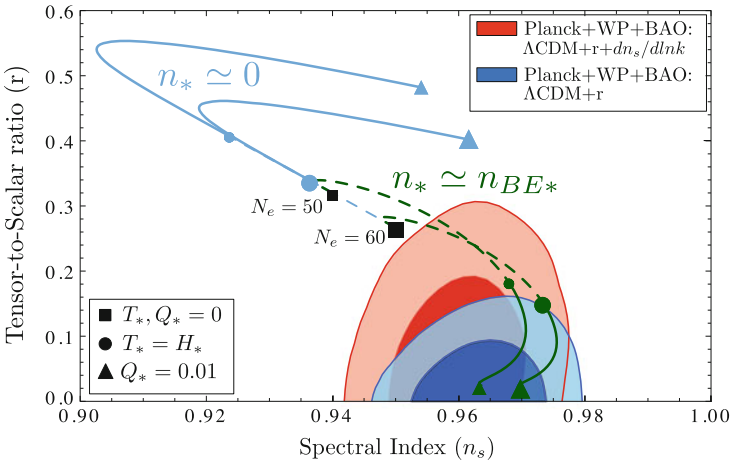

Fig. 1 Trajectories in the (n_s, r) plane for $V(\phi) = \lambda\phi^4$ as a function of the dissipative ratio, $Q_* < 0.01$, 50–60 e-folds before the end of inflation, compared with the Planck results (Ade et al. 2013), for $g_* = 228.75$ relativistic degrees of freedom. The *dark green (light blue) curves* correspond to nearly-thermal (negligible) inflaton occupation numbers n_*, with *dashed branches* for $T_* \leq H_*$. Note that corresponding curves converge in the cold inflation limit, $T_*, Q_* \to 0$. Figure from Bartrum et al. (2014)

The results in Fig. 1 for the trajectories in the (n_s, r) plane show that depending on the values of dissipation ratio Q and the statistical state for the inflatons, particularly in the case where the inflatons are in a quasi-thermalized state with a Bose-Einstein distribution $n_{\rm BE}$, the trajectories fall just easily in the allowed region from Planck.

4 Accounting for the Perturbations of the Radiation Bath: Coupled Two-Fluid System

It should be noticed that the inflaton and the radiation bath can be considered as a coupled two-fluid system. At the same time, inner interactions in the radiation fluid can produce decay and departure from equilibrium in the radiation fluid. These, in turn, can be related to dissipative fluxes in the radiation fluid, which in hydrodynamics can be described in terms of transport coefficients (the shear and bulk viscosities).

From hydrodynamics (Weinberg 1972), the shear and bulk viscosity coefficients can be expressed, respectively, as

$$\eta_s = \frac{4}{15}\rho_r\tau, \quad \eta_b = 4\rho_r\tau\left(\frac{1}{3} - v_s^2\right)^2. \tag{17}$$

where τ is a time scale for equilibration in the radiation fluid and v_s is the speed of sound.

Note that for conformal field theories $v_s^2 = 1/3$, dilatation is a symmetry and the fluid remains always in equilibrium. The same holds for scale invariant field theories, in which we have an ideal equation of state, $\omega_r = 1/3$, implying that $\eta_b = 0$. However, quantum corrections can break scale invariance, which then leads to a nonvanishing bulk viscosity in general, $\eta_b \neq 0$. For example, modeling the radiation bath with $m_\sigma^2\sigma^2/2 + \lambda_\sigma\sigma^4/4!$, ($m_\sigma \ll T, \quad \lambda_\sigma \ll 1$), the bulk and shear viscosity coefficients are found to be given by

$$\eta_b \simeq 8.9 \times 10^{-5}\lambda_\sigma T^3 \ln^2(0.064736\lambda_\sigma), \quad \eta_s \simeq 3.04 \times 10^3 \frac{T^3}{\lambda_\sigma^2} \,. \tag{18}$$

We recall that in the inflaton effective equation of motion both dissipation and fluctuation (stochastic noise) are present as a result of interactions. The same is expected to happen for the radiation fluid (analogous to Landau's theory of random fluids). Random sources and dissipative stresses in the radiation fluid equations are introduced via a stress term Π_{ab} in the stress-energy tensor,

$$T^{(f)}{}_{ab} = (p^{(f)} + \rho^{(f)})\, u_a{}^{(f)} u_b{}^{(f)} + p^{(f)}\, g_{ab} + \Pi_{ab}, \tag{19}$$

where

$$\Pi_{ij} = -\left(\eta_s \nabla_i u_j{}^{(f)} + \eta_s \nabla_j u_i{}^{(f)} + (\eta_b - \frac{2}{3}\eta_s)\delta_{ij}\nabla_k u^{(f)k}\right) - \Sigma_{ij}, \tag{20}$$

where fluctuations terms are generated by a Gaussian noise term Σ_{ab}. The correlation functions of the stochastic noise term Σ_{ij} are assumed to be local and determined by the fluctuation-dissipation relation,

$$\begin{aligned} &\langle \Sigma_{ij}(x,t)\Sigma_{kl}(x',t')\rangle = \\ &2T\left(\eta_s\delta_{ik}\delta_{jl} + \eta_s\delta_{il}\delta_{jk} + (\eta_b - \frac{2}{3}\eta_s)\delta_{ij}\delta_{kl}\right)\, \delta^{(3)}(x-x')\delta(t-t'). \end{aligned} \tag{21}$$

5 Cosmological Fluctuations

Starting now from the perturbed spacetime metric:

$$ds^2 = -(1+2\alpha)dt^2 - 2\beta_{,i}dt\, dx^i + a^2\left(\delta_{ij}(1+2\varphi) + 2\gamma_{,ij}\right) dx^i dx^j, \tag{22}$$

where

$$\chi = a(\beta + a\dot{\gamma}), \tag{23}$$

$$\kappa = 3H\alpha - 3\dot{\varphi} - \nabla^2\chi. \tag{24}$$

The perturbed Einstein equations in gauge-ready form can be expressed as (Hwang and Noh 2002)

$$\nabla^2\varphi + H\kappa = -4\pi G\delta\rho, \tag{25}$$

$$\kappa + \nabla^2\chi = -12\pi G(\rho + p)\delta v, \tag{26}$$

$$\dot{\chi} + H\chi - \alpha - \varphi = 8\pi G\delta\Pi, \tag{27}$$

$$\dot{\kappa} + 2H\kappa + \nabla^2\alpha - 3(\rho + p)\alpha = 4\pi G(\delta\rho + 3\delta p). \tag{28}$$

The inflaton and radiation perturbation equations now become

$$\begin{aligned}&(\delta\dot{\phi} - \alpha\dot{\phi})^{\cdot} + 3H(\delta\dot{\phi} - \alpha\dot{\phi}) - \nabla^2\delta\phi + \Omega_{,\phi\phi}\delta\phi - \kappa\dot{\phi}\\ &+\delta\Upsilon\dot{\phi} + \Upsilon\delta\dot{\phi} - \alpha\ddot{\phi} = (2\Upsilon T)^{1/2}\xi^{(\phi)}\end{aligned} \tag{29}$$

$$\begin{aligned}&\delta\dot{\rho}^{(f)} - \alpha T\dot{s} + 3H(\delta\rho^{(f)} + \delta p^{(f)} - \eta_b\kappa)\\ &+(Ts - 3H\eta_b)(\nabla^2\delta v^{(f)} - \kappa) + s_{,\phi}\delta q = -\delta Q^{(\phi)},\end{aligned} \tag{30}$$

$$\begin{aligned}&a^{-3}\{a^3(Ts - 3H\eta_b)\delta v^{(f)}\}^{\cdot} + \alpha(Ts - 3H\eta_b) + \delta p^{(f)}\\ &-\eta_b\kappa - \eta'\nabla^2(\delta v^{(f)} + \chi) = -\delta J^{(\phi)} + (2\eta' T)^{1/2}\xi^{(f)}.\end{aligned} \tag{31}$$

where we have defined η' as

$$\eta' = \frac{4}{3}\eta_s + \eta_b. \tag{32}$$

In the radiation perturbation equations we have energy and momentum transfer terms that can be expressed as

$$\delta Q^{(\phi)} = -\delta\Upsilon\dot{\phi}^2 - 2\Upsilon\dot{\phi}\,(\delta\dot{\phi} - \alpha\dot{\phi}) + (2\Upsilon T)^{1/2}\dot{\phi}\,\xi^{(\phi)} + \boldsymbol{\nabla}\cdot\boldsymbol{P}, \tag{33}$$

$$\delta J^{(\phi)} = \Upsilon\dot{\phi}\,\delta\phi + \nabla^{-2}\boldsymbol{\nabla}\cdot(\dot{\boldsymbol{P}} + 4H\boldsymbol{P}), \tag{34}$$

where $\boldsymbol{P}$ is a stochastic energy flux added to the stress energy tensor:

$$\boldsymbol{P} = -C_P(2\Upsilon T)^{1/2}\dot{\phi}\,\nabla^{-2}\boldsymbol{\nabla}\xi^{(\phi)}. \tag{35}$$

The two cases $C_P = 0$ and $C_P = 1$ govern whether the noise source $\xi^{(\phi)}$ appears in the energy flux or in the momentum flux.

Recalling the background equations:

$$\ddot{\phi} + (3H + \Upsilon)\dot{\phi} + V_{,\phi} = 0, \tag{36}$$

$$\dot{\rho}_r + 4H\left(\rho_r - \frac{9}{4}H\eta_b\right) = \Upsilon\,\dot{\phi}^2\,, \tag{37}$$

$$3H^2 = 8\pi G\rho\,, \tag{38}$$

prolonged inflation requires the slow-roll conditions $|\epsilon_X| \ll 1$, where $\epsilon_X = -d\ln X/Hdt$, and X is any of the background field quantities.

At leading order in the slow-roll approximation, we have that:

$$3H(1+Q)\dot{\phi} \simeq -V_{,\phi}, \tag{39}$$

$$4\rho_r \simeq 3Q\dot{\phi}^2 + 9H\eta_b, \tag{40}$$

$$3H^2 \simeq 8\pi GV, \tag{41}$$

where $Q = \Upsilon/(3H)$.

In the numerical results that follows, we will study the results for a dissipative parameter with cubic dependence with the temperature, $c = 3$, $\Upsilon = C_\phi \frac{T^3}{\phi^2}$, and a quartic chaotic model with inflationary potential $V = \lambda\phi^4/4$.

For the metric perturbations, we can choose a gauge (e.g. the zero-shear gauge, $\chi = 0$). From the slow-roll equations, the gauge-invariant curvature perturbation (Lukash variable) Φ, is defined as

$$\Phi = -\frac{1}{1+Q}\zeta^\phi - \frac{Q}{1+Q}\zeta^v\,, \tag{42}$$

where

$$\zeta^\phi = -\varphi + H\delta\phi/\dot{\phi}\,, \tag{43}$$

$$\zeta^v = \varphi + H\delta\rho_r/4\rho_r\,. \tag{44}$$

At late times, when $z \equiv k/(aH) \to 0$, we have $\Phi = -\zeta^\phi = -\zeta^v$. The power spectrum is found from

$$\langle\zeta^i(k,t)\zeta^i(k',t)\rangle = P^i(k,t)\,(2\pi)^3\delta^{(3)}(k+k')\,. \tag{45}$$

We consider the following cases:

- $\mathcal{N}_* = 0$ for a nonthermal (vacuum) inflaton fluctuation, or
- $\mathcal{N}_* = n_{BE}(k) = 1/(e^{k/(aT)} - 1)$ for a thermal statistical state for the inflaton
- $C_P = 0$, field stochastic $\xi^{(\phi)}$ term in the energy flux
- $C_P = 1$, no field stochastic $\xi^{(\phi)}$ term in the energy flux
- with and without viscosities

The effect of each term to the power spectrum and for the spectral tilt are shown in Fig. 2.

The effects of the viscosities on the power spectrum are exemplified in Fig. 3.

A more throughout analysis of both bulk and shear viscosities on the power spectrum and the spectral tilt in warm inflation can be found in Bastero-Gil et al. (2014a).

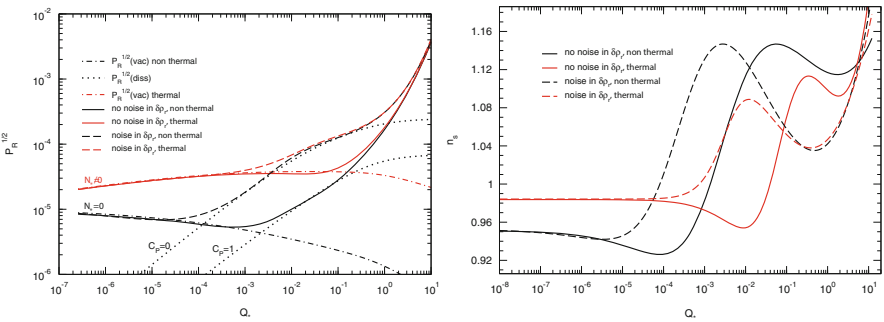

Fig. 2 *LHS plot*: Amplitude of the primordial spectrum versus Q_*. *Solid (dashed) lines* have been obtained with $C_P = 1\,(0)$; *black lines* are for $\mathcal{N}_* = 0$ (non-thermal) and *gray (red) lines* for $\mathcal{N}_* = n_{BE}$ (thermal). For comparison, the *dotted line* shows the analytical solution for $P^{1/2}_{\mathcal{R},diss}$, and *dot-dashed lines* $P_{\mathcal{R},vac}$. The amplitude of the spectrum becomes constant some e-folds after horizon crossing, and here we are plotting the value at $N_e = 20$. The total number of e-folds is 50. *RHS plot*: spectral index versus Q_*. We have taken $\lambda = 10^{-14}$. Figure from Bastero-Gil et al. (2014a)

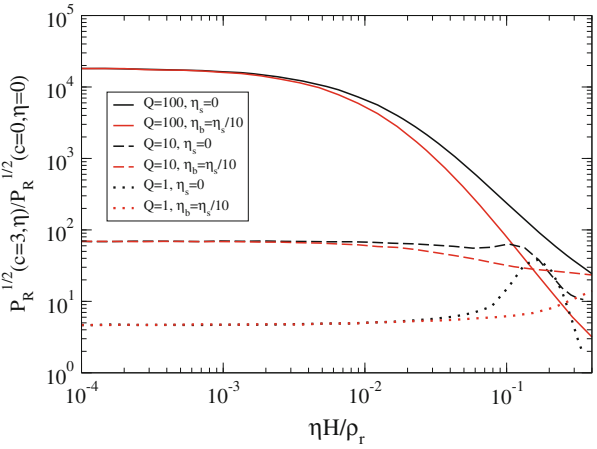

Fig. 3 The total (square root) amplitude for the power spectrum for the case $c = 3$ and normalized by its value for $c = 0$, as a function of viscosity. Three different values for the dissipation ratio Q are used to illustrate the ability of the viscosity to damp the spectrum. *Black curves* include only the effect of the bulk viscosity ($\eta_s = 0$) and the *red* includes the effects of both the shear viscosity and bulk viscosity coefficient given by $\eta_b = \eta_s/10$

6 The BICEP2 Recent Results and Possible Consequences for WI

The BICEP2 experiment (Ade et al. 2014) has recently reported evidence for a large tensor-to-scalar ratio $r = 0.20^{+0.07}_{-0.05}$ (without foreground dust subtraction) from the observation of B-mode polarization in the CMBR at degree angular scales. While this is good news for the inflationary paradigm (Liddle and Lyth 2000), which predicts a primordial tensor component in the CMBR spectrum, BICEP2's value seems to be in tension with the constraint on the tensor-to-scalar ratio reported by the Planck collaboration last year (Ade et al. 2013). The Planck collaboration has, in particular, placed an upper bound $r < 0.11$ (95 % CL), assuming that primordial scalar curvature perturbations are described solely by an adiabatic component with a simple power-law spectrum, i.e. no running of the spectral index.

The constraints obtained by BICEP2 in the (n_s, r) plane are illustrated in Fig. 4, where we superimpose them in the same Fig. 1 that includes the WI trajectories. Note from the figure that WI can also be consistent with the BICEP2 results. The figure also illustrates the discrepancy (at the 2-σ level) of the BICEP2 results from that from Planck in the absence of a running for the spectral index ($n_s' = dn_s/d\ln k$).

On the one hand, Planck has also confirmed a significant deficit of power on large angular scales with respect to their best-fit ΛCDM model, with a primordial spectrum characterized by a constant red-tilted spectral index, so that any additional contributions like gravity waves are naturally rather constrained. On the other hand, any modification of the primordial spectrum that tends to reduce the power on large scales will help relaxing the above constraint on r. Several possibilities were already mentioned by the Planck collaboration, and they have been further

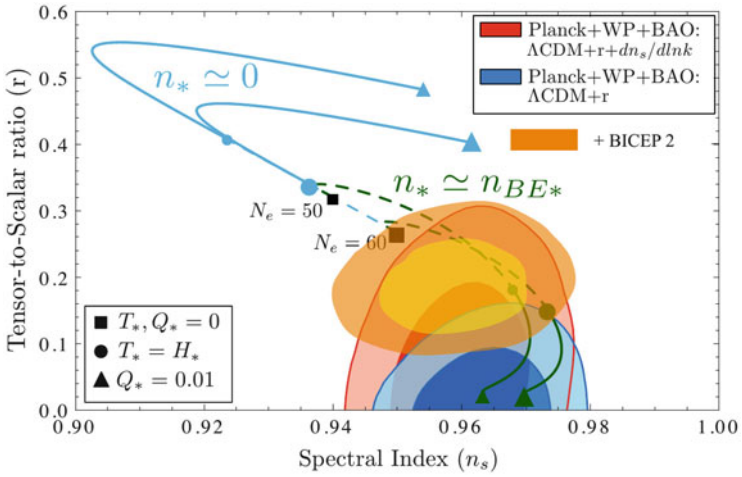

Fig. 4 Trajectories in the (n_s, r) plane. Same as in Fig. 1, but now including the contours from BICEP2

explored in view of the BICEP2 result, for example a negative running of the scalar spectral index, sterile neutrinos as extra relativistic degrees of freedom, a blue-tilted tensor spectrum, or isocurvature perturbations (Kawasaki and Yokoyama 2014). In particular the tension in the bound on the tensor-to-scalar ratio r between Planck and BICEP2 can be resolved by introducing isocurvature perturbations that are anti-correlated with the main adiabatic component.

As shown in Bastero-Gil et al. (2012), a cosmological baryon asymmetry can be produced through dissipative particle production during inflation, a mechanism known as *warm baryogenesis*. The produced baryon-to-entropy ratio $\eta_s = n_b/s$ was shown in that reference to be consistent with the observed cosmological asymmetry $7.2 \times 10^{-11} < \eta_s < 9.2 \times 10^{-11}$. Since the produced asymmetry η_s depends on the inflaton field and temperature, super-horizon fluctuations of the inflaton field will also be imprinted in the CMB temperature anisotropies in the form of baryon isocurvature perturbations. These have the same origin and will thus be fully (anti-)correlated with the main adiabatic curvature perturbations. This is a very distinctive feature of warm baryogenesis and makes it a testable model, which is not the case of most of the baryogenesis mechanisms proposed in the literature. Besides baryon isocurvature perturbations, dark matter isocurvature ones can equally be produced by the same mechanism during WI. These isocurvature perturbations B_m can be produced at a level still consistent with the upper bounds set by Planck, $|B_m| < 0.079$, yet, they can affect the CMB temperature anisotropies $\langle(\Delta T/T)^2\rangle$ strongly,

$$\langle(\Delta T/T)^2\rangle \sim P_\zeta \left(1 + \frac{5}{6} r_{\rm eff}\right) , \tag{46}$$

where

$$r_{\rm eff} = r + \frac{6}{5}\left(4B_m^2 + 4B_m\right) , \tag{47}$$

such that for anti-correlated isocurvature perturbations a smaller effective tensor-to-scalar ratio can be obtained.

It is important to mention that in deriving the bounds on the tensor-to-scalar ratio r from CMB temperature anisotropies, the Planck collaboration has assumed that primordial scalar curvature perturbations are described uniquely by an adiabatic component. The effects of any other component such as baryon isocurvature modes are then necessarily absorbed into an effective tensor-to-scalar ratio, $r_{\rm eff}$, which is smaller than the true tensor contribution if the additional components are anti-correlated with the dominant adiabatic modes. Using this effect of an anti-correlated matter isocurvature perturbation that can be generated during WI, an either partial or even full screening is naturally present and may reconcile the BICEP2 detection of B-mode polarization with the upper bound on the tensor-to-scalar ratio placed by Planck, as shown in details in Bastero-Gil et al. (2014b). A partial screening would, in particular, be interesting if there is future evidence for a non-zero tensor-to-scalar ratio in the temperature power spectrum that is somewhat smaller than the value

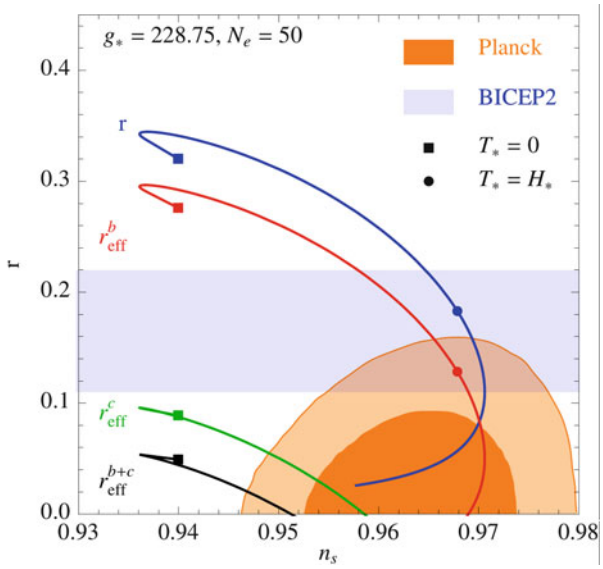

Fig. 5 Trajectories in the $n_s - r$ plane for the quartic model (thermal inflaton occupation numbers) with 50 e-folds of inflation and $g_* = 228.75$, with baryon ($r^b_{\rm eff}$), CDM ($r^c_{\rm eff}$), and full matter isocurvature perturbations ($r^{b+c}_{\rm eff}$). The *shaded regions* show the 68 % and 95 % CL Planck contours (Ade et al. 2013), including the results of WMAP and BAO observations, and the 1σ interval for r derived by the BICEP2 collaboration (Ade et al. 2014) after taking into account dust contributions. These results correspond to the range $T_*/H_* = 0 - 13$ ($Q_* = 0 - 0.01$). Figure from Bastero-Gil et al. (2014b)

inferred from the polarization data. This screening can, in fact, be effective for a wide range of values for the tensor-to-scalar ratio and is not inherent to the large value obtained by the BICEP2 collaboration, which is presently is under scrutiny.

An example of this effect of screening of the tensor-to-scalar ratio is shown in Fig. 5, again for the case of a quartic inflaton potential in WI.

7 Summary and Perspectives

We have seen that warm inflation is able to describe in a concomitantly and natural way the effects of both quantum and thermal fluctuations. While the observational constraints involving the spectral index n_s and the tensor to scalar curvature ratio r already rule out a large class of inflaton polynomial potentials $V \sim \phi^p$, with $p > 3$, dissipative effects and thermal fluctuations can make these higher polynomial inflaton potentials compatible again with the observational constraints, as shown recently in Ramos and Silva (2013) and in Bartrum et al. (2014). We have here shown the results for the particular case of a quartic inflaton potential and seen how

easily WI can make it in concordance with the recent results from Planck. The most important effect of dissipation and/or a non-trivial inflaton particle distribution is the lowering of the tensor-to-scalar ratio in the modified consistency relation in Eq. (12), so we expect next upcoming Planck release and future CMB B-mode polarization searches to shed new light on the nature of inflaton fluctuations.

We have also seen that WI can naturally account for the baryon asymmetry of the universe. As an additional bonus, we predicts the production of anti-correlated isocurvature perturbations as a result of the matter production due to dissipative effects during WI. The obtained results for chaotic models suggest that a large tensor-to-scalar ratio could be accommodated by the current Planck results due to the presence of these matter isocurvature modes associated with an asymmetric dissipation of the inflaton's energy density into baryonic or CDM species. This provides a way of reconciling the Planck results with the recent results from the BICEP2 experiment on B-modes measurements on the CMB.

Acknowledgements The author would like to thank for financial support from research grants from CNPq and FAPERJ (Brazilian funding agencies). I also would like to thank the organizers for the invitation and for putting together a stimulating program.

References

Ade, P.A.R., et al. [Planck Collaboration]: Astron. Astrophys. **571**, A23 (2013/2014). arxiv:1303.5082 [astro-ph.CO]
Ade, P.A.R., et al. [BICEP2 Collaboration]: Phys. Rev. Lett. **112**, 241101 (2014)
Bartrum, S., Bastero-Gil, M., Berera, A., Cerezo, R., Ramos, R.O., Rosa, J.G.: Phys. Lett. B **732**, 116 (2014)
Bastero-Gil, M., Berera, A., Ramos, R.O.: J. Cosmol. Astropart. Phys. **1109**, 033 (2011)
Bastero-Gil, M., Berera, A., Ramos, R.O., Rosa, J.G.: Phys. Lett. B **712**, 425 (2012)
Bastero-Gil, M., Berera, A., Ramos, R.O., Rosa, J.G.: J. Cosmol. Astropart. Phys. **1301**, 016 (2013)
Bastero-Gil, M., Berera, A., Moss, I.G., Ramos, R.O.: J. Cosmol. Astropart. Phys. **1405**, 004 (2014a)
Bastero-Gil, M., Berera, A., Ramos, R.O., Rosa, J.G.: J. Cosmol. Astropart. Phys. **1410** 10, 053 (2014b)
Berera, A., Moss, I.G., Ramos, R.O.: Rep. Prog. Phys. **72**, 026901 (2009)
Hwang, J.-C., Noh, H.: Classical Quantum Gravity **19**, 527 (2002)
Kawasaki, M., Yokoyama, S.: J. Cosmol. Astropart. Phys. **1405**, 046 (2014)
Liddle, A.R., Lyth, D.H.: Cosmological Inflation and Large Scale Structure. Cambridge University Press, Cambridge (2000)
Ramos, R.O., Silva, L.A.: J. Cosmol. Astropart. Phys. **1303**, 032 (2013)
Weinberg, S.: Gravitation and Cosmology. Wiley, New York (1972)

A Brief History of the Brazilian Participation in CMB Measurements

Thyrso Villela

Abstract This contribution is a short report on the Brazilian participation in Cosmic Microwave Background (CMB) observational programs. It includes brief descriptions of the experiments aiming to measure both CMB properties and Galactic microwave signals that hamper these measurements. The work done by Brazilian researchers involved in the development of these experiments and in the subsequent observations is briefly described as well.

1 Introduction

The history of Cosmic Microwave Background (CMB) measurements in Brazil began in the early 1980s, when groups from Princeton University and University of California at Berkeley flew instruments on board stratospheric balloon platforms to map the angular distribution of this radiation in the southern sky. These experiments were carried out with the support from the National Institute for Space Research (Instituto Nacional de Pesquisas Espaciais—INPE). In particular, the collaboration between the University of California and INPE evolved to a fruitful joint work that produced several scientific and technological results related to CMB measurements and to topics as CMB foreground characterization.

The CMB signal can be measured from a few GHz up to a few hundreds GHz. The full characterization of the CMB radiation field requires measurements of its spectrum, polarization and angular distribution. CMB spectrum measurements have to deal with a signal level of the order of a few kelvins, while CMB anisotropy signal is in the range of millikelvins to microkelvins. The CMB polarization signal, on the other hand, is only detected at a few microkelvins level. These signal intensities impose serious experimental and observational constraints.

T. Villela (✉)
Divisão de Astrofísica, Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos, SP, Brazil
e-mail: thyrso.villela@inpe.br

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_9

Depending on the frequency range, this signal can be picked up from the ground, but, in general, it is necessary to use space platforms to avoid the atmosphere during the observations. So, over the years, balloon borne and satellite platforms were employed to allow precise measurements of the spectrum, angular distribution and polarization of the CMB. Besides the atmosphere, other contaminants can hamper CMB observations as well, like the radio and microwave emissions from our own galaxy and extragalactic radio sources. The detailed knowledge of such signals is of pivotal importance to better characterize the CMB data.

In order to measure these weak CMB signals, besides the development of very sensitive experiments, it is necessary to account for the spurious signals that can hamper the CMB observations. Thus, CMB measurement requires a strict control of systematic errors.

Observational efforts in Brazil were carried out over the past three decades to measure all three CMB characteristics and their contaminant foregrounds. This contribution presents a brief report about the history of the work done by Brazilian scientists and technicians in CMB measurements and related programs.

2 CMB Measurements and Related Programs

We show below some of the results related to CMB measurements programs, including hardware development and data analysis, in which Brazilian scientists were involved since 1982. The Brazilian institutions that took part in this effort were INPE and, later, the Federal University of Itajubá (Universidade Federal de Itajubá—Unifei).

2.1 CMB Angular Distribution: 3 mm Experiment

The first Brazilian involvement in CMB measurements was in the 3.3 mm (90 GHz) experiment. This instrument was a test bed and precursor for the Cosmic Background Explorer (COBE) 90 GHz Differential Microwave Radiometer (DMR) channel, which was launched on November 1989. With a noise temperature of 130 K, a bandwidth of 600 MHz and a beam width of 7° FWHM, the 3 mm experiment was mounted on a gondola which rotated with a period of 100 s. It made a very precise measurement of the CMB dipole anisotropy and placed a stringent upper limit for the quadrupole (Villela et al. 1983; Lubin et al. 1985; Lubine and Villela 1985; Villela 1987). Figure 1 shows parts of the experiment and Fig. 2 presents the instrument assembled and taking off for its Brazilian flight.

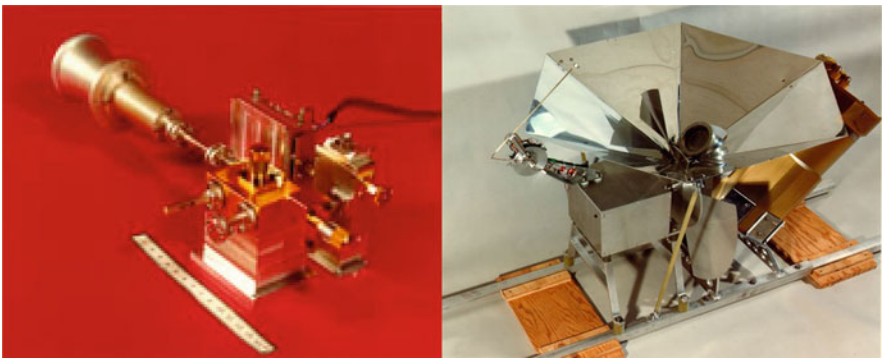

Fig. 1 90 GHz horn (*left*) and radiation shield to avoid spurious ground pick-up (*right*) used in the 3 mm experiment

In this flight, the gondola was launched from INPE's campus in Cachoeira Paulista, state of São Paulo, in November 19, 1982. The instrument performed very well and took CMB data at an average altitude of 30 km for several hours. The CMB dipole could be seen in real time as the gondola rotated at 1 rpm. Unfortunately, due to a failure in the separation mechanism that connects the gondola do the flight ladder and balloon, the flight extended for several hours beyond the estimated duration and the payload was lost, as the batteries ran out of charge causing a communication breakdown between the payload and the ground control station. The gondola was recovered about 2 years later. A history of this flight and the gondola recovery in Tapiraí, SP, in February 1985 can be seen in Villela (1994).

Despite the fact that the payload was lost, the telemetered data were used to analyze the data. A very precise measurement of the CMB dipole was made as well as a stringent CMB quadrupole was set through the combination of these data with data collected in two previous flights from the Northern Hemisphere. The final result of the 3 mm experiment showed a dipole intensity of 3.44 ± 0.17 mK in the direction $\alpha = 11^h.2 \pm 0^h.1, \delta = -6^0.0 \pm 1^0.5$. An rms quadrupole amplitude upper limit was set to a 90 % confidence level as 7×10^{-5} (Lubin et al. 1985; Villela 1987). Figure 3 shows a map of the CMB dipole measured by the 3 mm experiment.

Table 1 shows a comparison of the 90 GHz dipole measurement (Lubin et al. 1985; Villela 1987) with the Relikt (Strukov et al. 1987) and COBE (Smoot et al. 1991) satellite measurements. It can be seen from this table that the 3 mm data were in good concordance with satellite data.

The balloon experiments carried out in the early 1980s (Fixsen et al. 1983; Lubin et al. 1985; Lubin and Villela 1986) paved the way for the COBE satellite, as shown in Fig. 4.

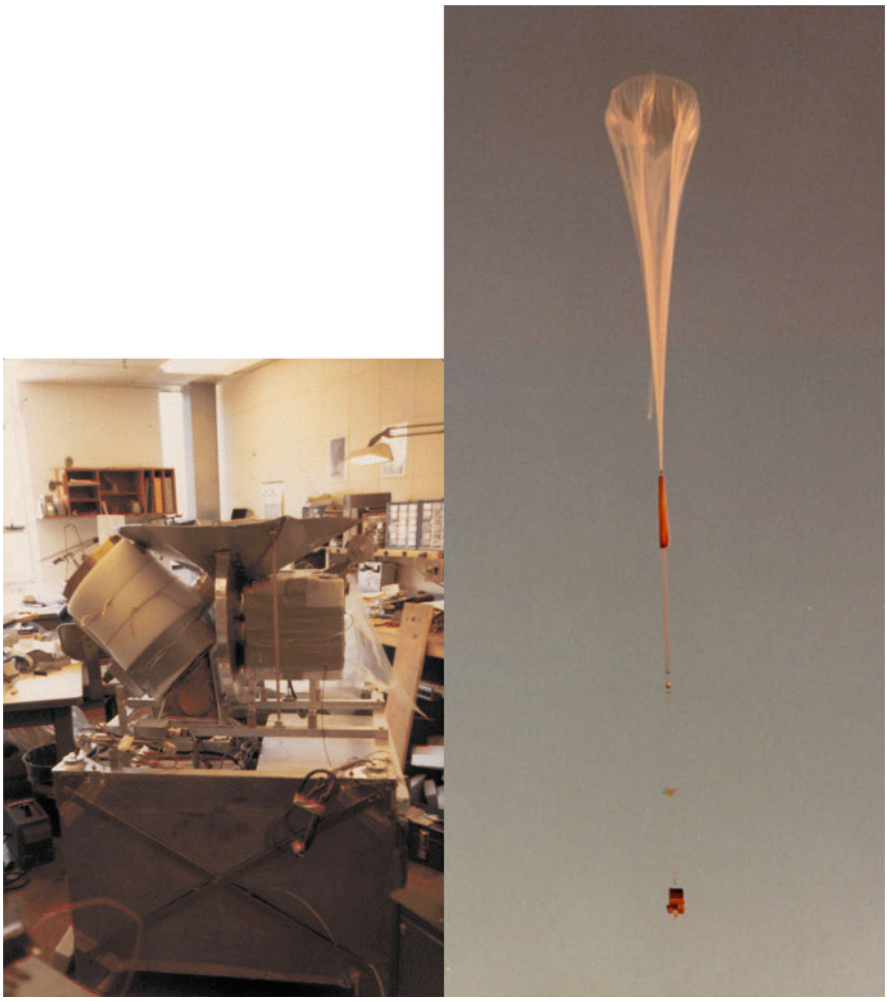

Fig. 2 (*Left*) 3 mm experiment assembled in the laboratory and (*right*) the gondola taking off from Cachoeira Paulista, SP, for its balloon flight in November 19, 1982

With the advent of the COBE satellite, it was possible to use the DMR data to constrain the topology of the universe, work done by an INPE graduate student Angélica de Oliveira-Costa, who went to Berkeley to work with COBE data (de Gouveia dal Pino et al. 1995).

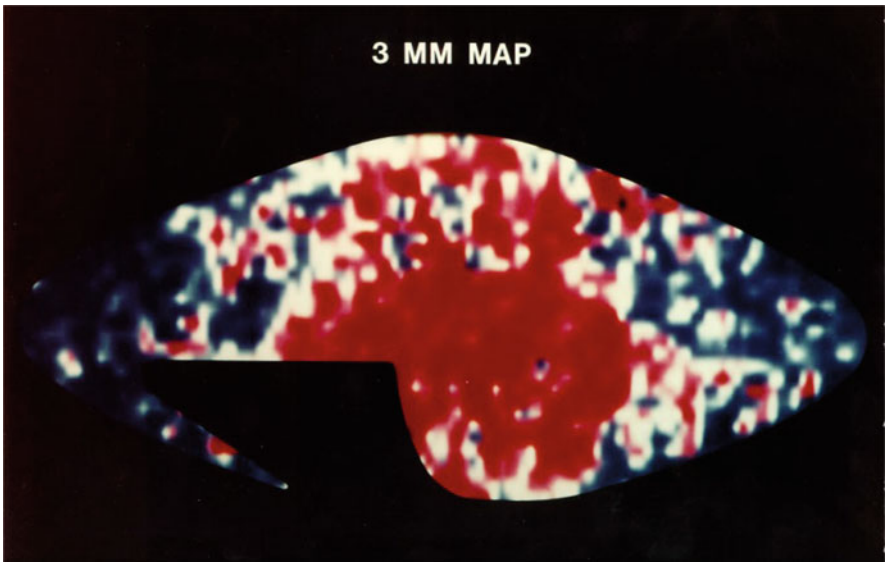

Fig. 3 CMB dipole anisotropy map, in celestial coordinates, covering 80 % of the celestial sphere. The lack of data is evident in the *black region* in the *lower left corner* of the map, which was due to the non-recovery of the experiment during the Southern Hemisphere 1982 campaign

Table 1 CMB dipole measurements

Experiment	Amplitude (mk)	l (^{0})	b (^{0})
Relikt	3.16 ± 0.07	266.4 ± 2.3	48.5 ± 1.6
3 mm	3.44 ± 0.17	264 ± 1.9	$49.2\pm$
COBE	3.3 ± 0.1	265 ± 1	48 ± 1

2.2 ACME, HACME and BEAST Experiments

As a natural extension of the large angular scales experiments, a series of experiments to search for CMB anisotropy at medium angular scales was conducted by groups from the University of California (campuses of Berkeley and Santa Barbara) in which INPE's researchers, including graduate students, were involved. These experiments were performed in the 1990s. The Advanced Cosmic Microwave Explorer (ACME), shown in Fig. 5, was developed at that time (Meinhold et al. 1993).

Four balloon flights were performed in which the detectors were bolometers [the ACME-MAX series: e.g. Gundersen et al. 1993; Clapp et al. 1994; Devlin et al. 1994; Tanaka et al. 1996; Lim 1996], and four observational campaigns at the South Pole, in which the detectors employed were High Electron Mobility Transistors (HEMT) [the ACME-SP series: Santos et al. 2012; Gundersen et al. 1995]. For instance, Carlos Alexandre Wuensche worked on the MAXIMA experiment (Wuensche 1995). Newton Figueiredo also participated in these observations. For

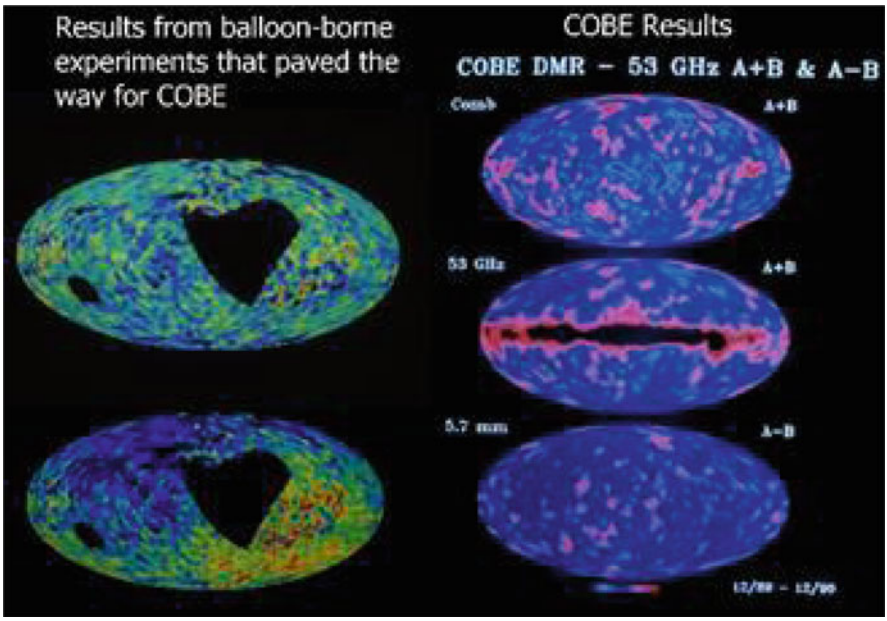

Fig. 4 Results from the balloon-borne experiments that paved the way for COBE. The *left side* of the figure shows a combination of the 12 and 3 mm data obtained from balloon flights, while the *right side* of the picture shows COBE results (P. Lubin)

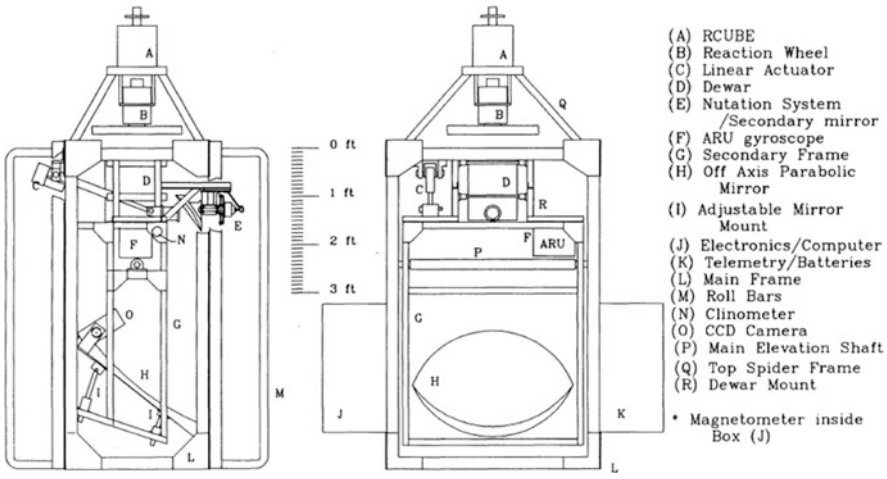

Fig. 5 Balloon gondola of the ACME experiment. This platform was designed to automatically point the telescope during the flights. It used inertial sensors to achieve this goal (Meinhold et al. 1993)

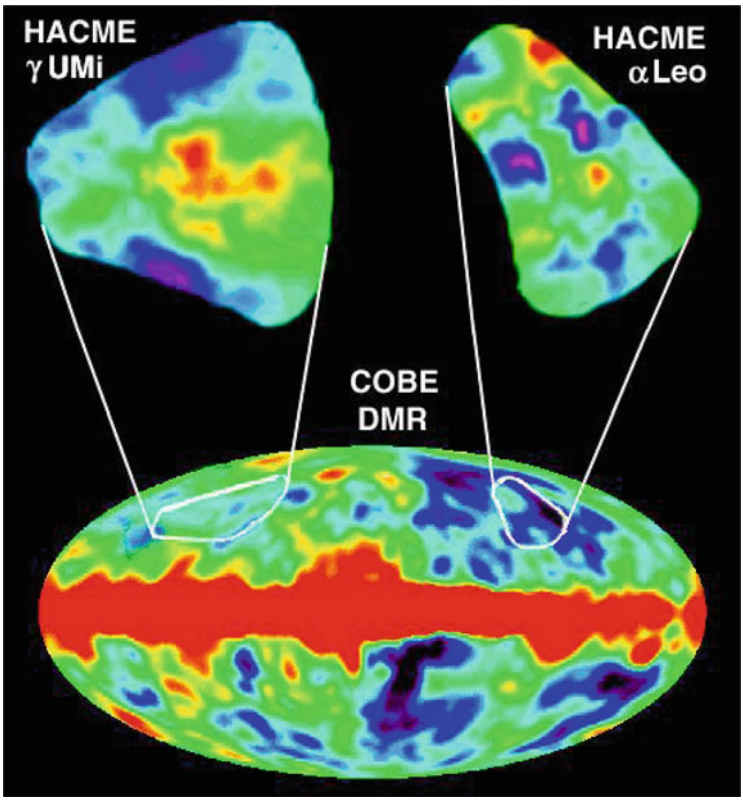

Fig. 6 CMB maps obtained with the HACME experiment

information on the scientific results of these experiments see Wuensche and Villela (2010) and references therein.

The next generation of medium angular scale experiments was the HEMT on ACME (HACME) experiment (Staren et al. 2000). It was flown in 1996 and made observations covering 1150 square degrees of the celestial sphere near the stars γ Ursae Minoris and α Leonis at the frequencies of 39, 41, and 43 GHz, with an $\sim$ 0.077 beam (FWHM). The detected cosmic signal was smaller than 77 μK. Figure 6 shows the CMB maps obtained from these observations (Tegmark et al. 2000).

After ACME and HACME, a new experiment was designed to search for anisotropies in CMB angular distribution with better sensitivity. The Background Emission Anisotropy Scanning Telescope (BEAST) (Childers et al. 2005; Figueiredo 1997) was developed as a collaboration involving the University of California Santa Barbara, INPE, Jet Propulsion Laboratory, Unifei, University of Milan, University of Rome, IASF/CNR, Caltech and University of Illinois. INPE provided several microwave parts for this experiment. The BEAST primary dish is the largest mirror ever flown in a CMB experiment. The BEAST innovative

Table 2 BEAST optical parameters (Figueiredo et al. 2005)

Primary mirror	
Focal distance	1250.0 mm
Semi-major axis	2200.00 mm
Aperture	1966.1 mm
Secondary mirror	
Semi-major axis	600.0 mm
Semi-minor axis	575.4 mm
Focal distance	170.0 mm
Eccentricity	0.2833
Flat mirror	
Diameter	2600.0 mm

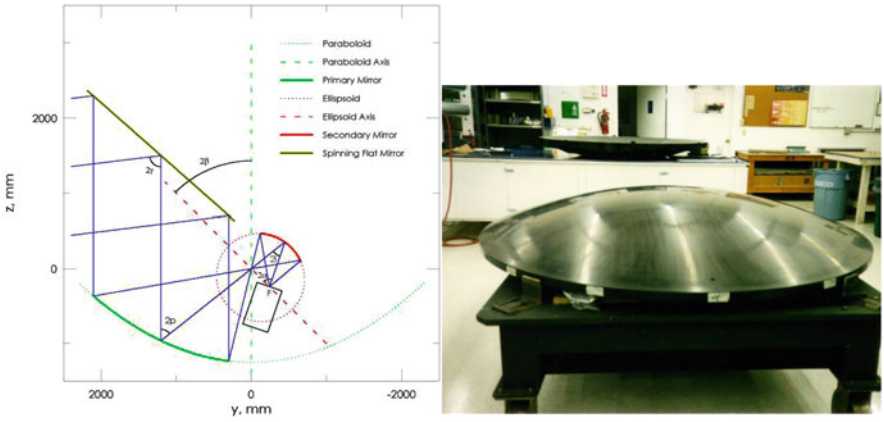

Fig. 7 BEAST optical design (**a**). In (**b**) it is shown the mold for the 2.6-m diameter primary mirror

optical design, which had a large focal plane, was part of the work done by Newton Figueiredo, an INPE graduate student, for his thesis work. Table 2 and Fig. 7 show BEAST optical design parameters.

BEAST, besides data collected in balloon flights, also collected CMB data from the ground at White Mountain Research Station (USA) at an altitude of 3800 m. An astronomical site survey at the Barcroft Facility of the White Mountain Research Station was made (Marvil et al. 2006) in order to better characterize this site for CMB measurements. BEAST maps cover about 2500 square degrees in the celestial sphere in the declination band $33° \leq \delta \leq 42°$ and right ascension band $0\,h \leq \alpha \leq 24\,h$. Most of the observations were in the Ka band (30′ FWHM) and Q band (23′ FWHM). BEAST map results are presented in Meinhold et al. (2005), while Galactic contamination in these maps is described in Mejía et al. (2005), whose author was a post-doc at INPE at the time working on BEAST data analysis (Figs. 8 and 9). The resulting BEAST data CMB power spectrum is presented in O'Dwyer

Fig. 8 BEAST experiment before flight showing the 2.6-m diameter flat mirror

(2005). Figure 10 shows a slice of the BEAST Q band sky map compared to the same slice of the sky observed by WMAP.

The BEAST rms signal level was $57 \pm 5\,\mu$K (noise, without Galaxy, beam smoothed to $30'$). The cosmic signal was $30 \pm 3\,\mu$K. INPE graduate student Agenor Pina worked on the analysis of the BEAST data (Pina 2002). An independent

Fig. 9 BEAST experiment before flight on the crane (a) and the dome where it operated at the White Mountain Research Station

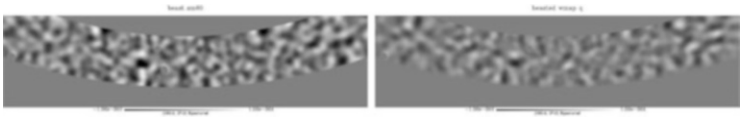

Fig. 10 WMAP map section (*above*) compared to same sky region as observed by BEAST (*below*) in the *Q* band

calculation of BEAST CMB angular power spectrum has been done by Donzelli et al. (2006).

3 CMB Polarization: WMPol Experiment

BEAST was adapted to operate as a polarimeter and became the White Mountain Polarimeter (WMPol). INPE provided several parts for this instrument and Rodrigo Leonardi, an INPE graduate student at that time, worked in its development and operations. BEAST was deployed at the White Mountain Research Station site and operated there for several months taking data at 42 and 90 GHz. These observations

allowed us to set an upper limit of 14 μK for the CMB E-mode polarization in the $170 < l < 240$ multipole interval (Leonardi 2006; Levy et al. 2008) at 42 GHz.

Figures 11 and 12 show some of the microwave parts provided by INPE for the BEAST and WMPol experiments. Some of these parts were jointly developed by INPE and companies in the São José dos Campos and Campinas regions. The know-how acquired in such developments was important for these companies, as they qualified themselves to be providers of precise mechanical parts.

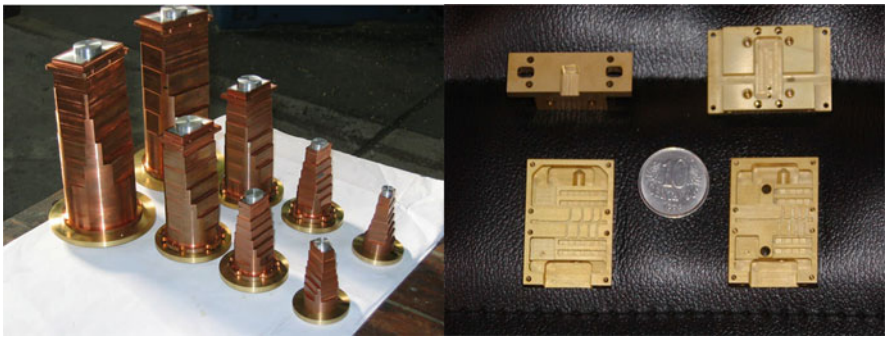

Fig. 11 WMPol transitions (**a**) and amplifiers bodies compared in size to a Brazilian 10 cents coin (**b**)

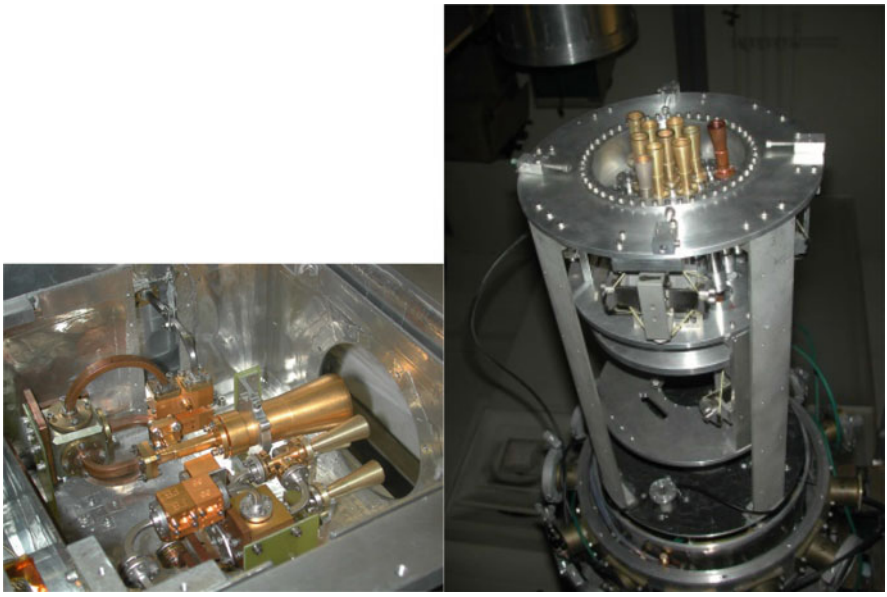

Fig. 12 WMPol microwave hardware (**a**) mounted in the dewar (**b**)

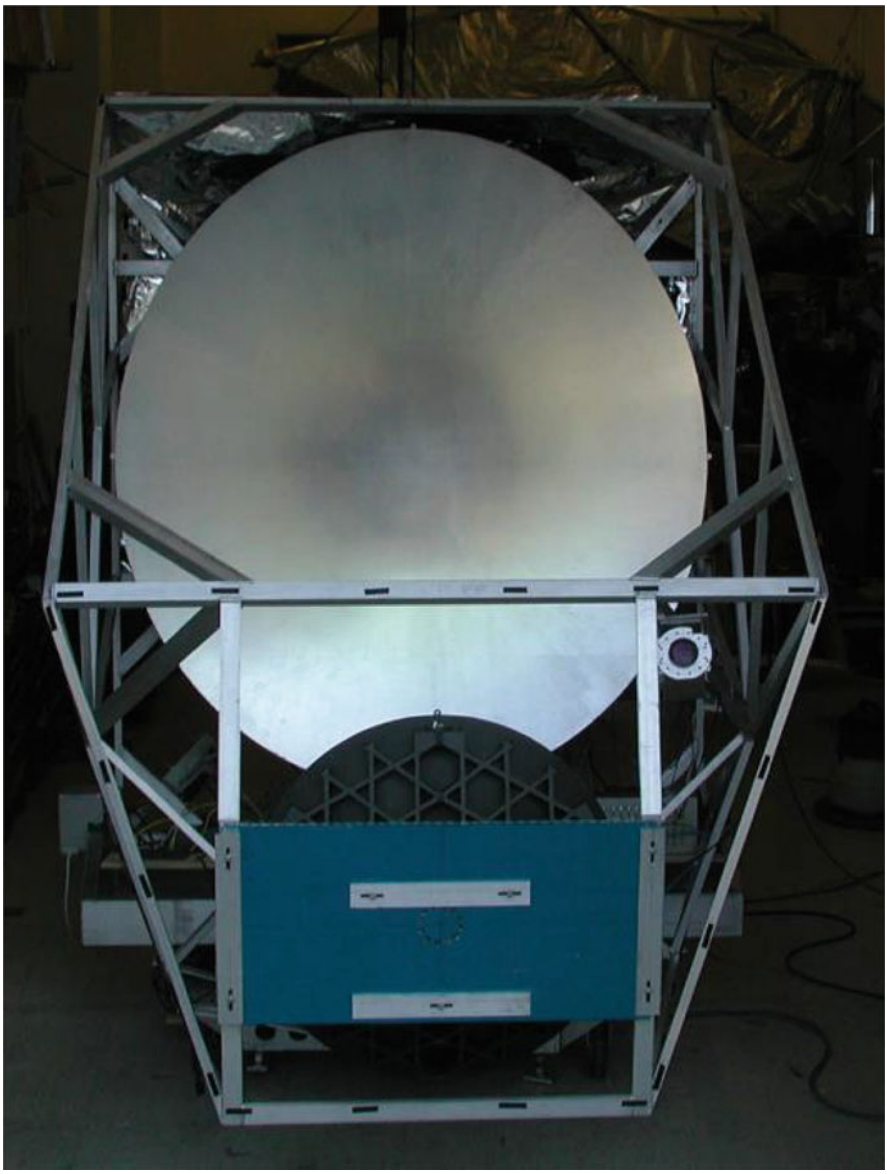

Fig. 13 WMPol fully assembled for observations at its White Mountain Research Station observing site

Figure 13 is a picture of WMPol in its working configuration. It observed the sky at 42 and 90 GHz. These observations allowed us to set an upper limit for CMB polarization at 42 GHz. Figure 14 presents this result.

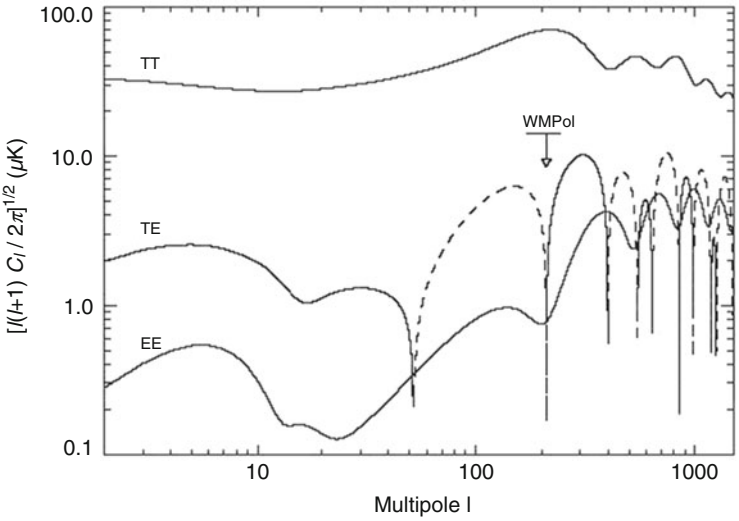

Fig. 14 WMPol upper limit on CMB polarization (Leonardi 2006; Levy et al. 2008)

4 CMB Spectrum: ARCADE Experiment

The search for distortions in the CMB spectrum and the need for precise measurements on the lower frequency part of the this spectrum lead to the development of the Absolute Radiometer for Cosmology, Astrophysics, and Diffuse Emission (ARCADE) experiment, a collaboration involving Goddard Space Flight Center (GSFC/NASA), University of California, Santa Barbara, Jet Propulsion Laboratory (JPL/NASA), University of Maryland, and INPE. ARCADE was designed to operate on board high altitude balloons in order to measure the CMB frequency spectrum at centimeter wavelengths and to search for signals from the first stars to form after the Big Bang. The instrument was cooled down to −270 °C, through a radical thermal design that puts cold components outside the dewar.

INPE provided horns and microwave transitions at 90, 30, 10, 8, 5, and 3 GHz for the ARCADE project. The transitions were fabricated in Copper through electrodeposition process. The corrugated horns were made out of Aluminum. Figure 15 show these horns and transitions, respectively, while Fig. 16a, b show some of them assembled in the instrument.

ARCADE was launched in July 22, 2006 from Palestine, TX, USA. The flight had a total of 4 h at the float altitude of 37 km. Successful thermal operations were carried out during the flight consisting in comparing the sky to an on-board target several times, while maintaining the instrument within 0.1 K of the sky temperature. This procedure allowed 2.5 h of science data to be collected.

The ARCADE data showed a detection of a bright radio background that is six times brighter than the expected combined contribution from all known radio

Fig. 15 Horns (**a, b**) and transitions (**c**) provided by INPE for the ARCADE 2 experiment

galaxies (Fig. 17). This unexpectedly bright radio background has unknown origin up to now. Moreover, this ARCADE result is consistent with existing radio surveys, which makes this detection a very intriguing one, as radio source counts are well known, and their emissions don't even come close to making up the detected background. New sources, too faint (10x) to be directly observed, would have to vastly outnumber (100x) all the galaxies in the universe (Kangas et al. 2005a; Fixsen et al. 2011; Seiffert et al. 2011).

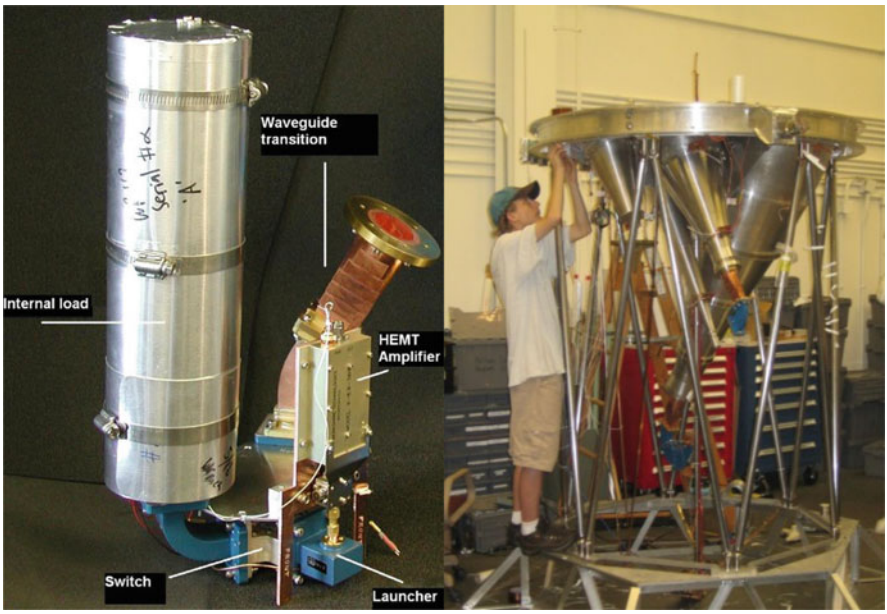

Fig. 16 Transitions (*left*, (**a**)) and horns (*right*, (**b**)) mounted on the ARCADE 2 experiment

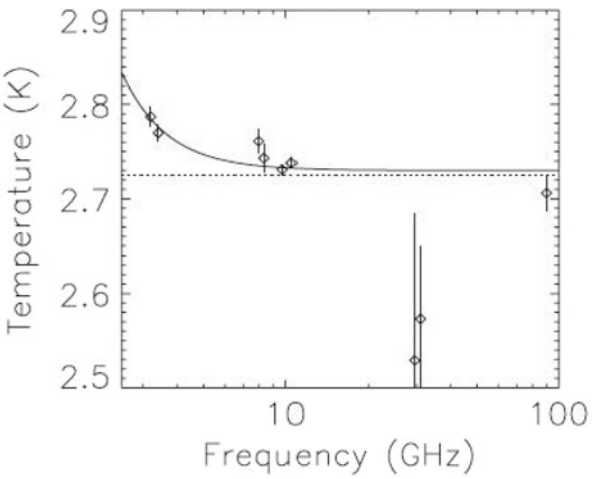

Fig. 17 ARCADE results showing an excess intensity in the lower frequency part of the spectrum compared to the 2.7 K CMB temperature (Fixsen et al. 2011)

5 Radio and Microwave CMB Foregrounds: GEM and COFE Experiments

Brazilian scientists and technicians have also been involved in the characterization of CMB foregrounds. The Galactic Emission Mapping (GEM) project (de Amici et al. 1994; Torres et al. 1996) is a collaboration encompassing the University of California, Berkeley, in the USA, University of Milan, in Italy, Instituto de Telecomunicazioni, in Portugal, and INPE, aiming to map the Galactic synchrotron radiation at 408, 1465, 2.3, 5 and 10 GHz. It employs a 5.5-m diameter dish, which rotates at constant speed to scan the sky, and extension panels to avoid ground emission pick-up (Tello 1997; Tello et al. 1999, 2000). A schematic view of the GEM dish set up, including the fence used to avoid ground emission pick up is shown in Fig. 18. The GEM observational is also presented in Fig. 18.

GEM is currently operating at INPE campus in Cachoeira Paulista, SP (Fig. 19). From this site, it can observe the sky in the declination interval between 52o23'14.1" and +7o 8'50.98", covering about 33 % of the sky. GEM has already produced maps at 1.465 GHz, a work done by Camilo Tello in his Ph.D. thesis (Tello 1997) and 2.3 GHz (Tello et al. 2013). A preliminary map at 408 MHz (Souza 2000) has been produced. Preliminary maps of the polarized Galactic emission at 5 GHz have also been produced by Ivan Ferreira, a former INPE graduate student who worked on the development of the 5 GHz polarimeter (Ferreira 2008). Figure 20 is a picture of GEM in its current observing site in Cachoeira Paulista, SP.

COsmic Foreground Explorer (CoFE) is a balloon-borne microwave polarimeter designed to measure the low-frequency and low-l characteristics of the dominant diffuse polarized foregrounds. Short duration balloon flights from the Northern and Southern Hemispheres will allow the telescope to cover up to 80 % of the sky with an expected sensitivity per pixel better than 100 μK/deg2 from 10 to 20 GHz. This

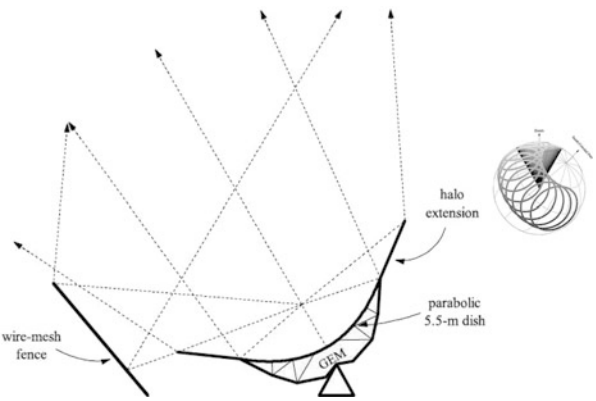

Fig. 18 GEM schematic views of the parabolic dish, extension halo panels, and fence (**a**) and observational strategy (**b**)

Fig. 19 Picture of the GEM experiment in Cachoeira Paulista, SP, Brazil

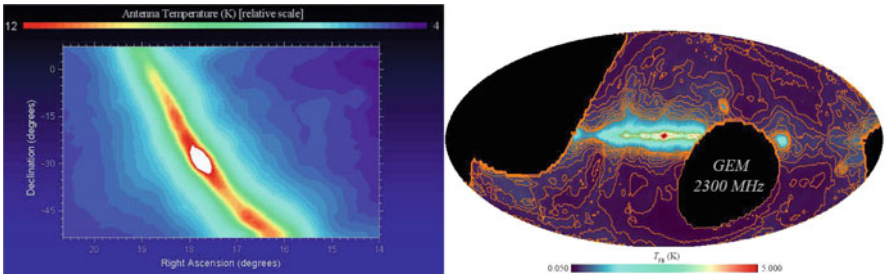

Fig. 20 GEM map of the Galactic Center at 1465 GHz (**a**) and GEM map of the sky at 2.3 GH (**b**)

is an important effort toward characterizing the polarized foregrounds for future CMB experiments, in particular the ones that aim to detect primordial gravity wave signatures in the CMB polarization angular power spectrum (Leonardi et al. 2006).

6 Concluding Remarks

The collaborative works described in this paper benefited a couple of generations of Brazilian and US students. Brazilian students went to the University of California, at Santa Barbara and Berkeley, to work on Lubin's and Smoot's groups. US students visited INPE to work on the GEM project. These collaborations with Philip Lubin and George Smoot are still active.

Regarding instrument developments, all joint projects provided interesting spin-offs for other projects and careers (Kangas et al. 2005a). Camilo Tello quit his position as a researcher at INPE and is working now in the USA in a mobile phone company, where he employs his expertise in microwaves acquired during his thesis work. He is still very involved and active in the GEM project. In particular, the experience acquired with the development of the ACME instrument, where a stabilized balloon-borne platform was built, has been used at INPE to develop a stabilized platform for the Masco X- and gamma-ray imaging telescope (Villela 2000, 2002). A concise compilation of some of the instrumental work done in these collaborations can be found in (Wuensche and Villela 2010; Villela et al. 2011) and a general approach on CMB observations can be found in (Villela and Wuensche 2009).

Besides the instrumental work, emphasis has also been given to the analysis of CMB data obtained with satellites. COBE data were used to investigate possible large scale fractal structure in the universe (de Gouveia dal Pino et al. 1995); investigations of the CMB angular distribution measured by WMAP (Abramo et al. 2006; Bernui et al. 2006a; Santos et al. 2012) and Planck (Santos et al. 2012) were carried out; CMB temperature maps obtained by WMAP have been used to investigate possible Gaussian departures (Bernui et al. 2006a,b). As general approaches for CMB data analysis, an alternative algorithm for the harmonic analysis of CMB distribution has been tested (Wuensche et al. 1994) and simulations of the CMB power spectrum for a class of mixed, non-Gaussian, primordial random fields have been performed (Andrade et al. 2004).

Just for the sake of curiosity, it is interesting to draw a quick line on what happened in CMB measurements after the balloon flights in Brazil in the early 1980s. The Princeton University group leader was David Wilkinson. Dale Fixsen and Edward Cheng completed the team. The University of California, Berkeley, team had Philip Lubin and George Smoot as leaders and Gerald Epstein and myself as graduate students. Cheng, Fixsen, Lubin, Smoot and Wilkinson were part of the COBE satellite team. For their work in this experiment they were awarded the Gruber Prize in Cosmology in 2006, along with John Mather and the COBE team. George Smoot won the 2006 Nobel Prize in Physics for the DMR/COBE results (Smoot et al. 1992), to which he was the principal investigator. He shared this prize with John Mather, principal investigator of the FIRAS/COBE. David Wilkinson became involved in the so-called Microwave Anisotropy Probe (MAP), that later was renamed Wilkinson Microwave Anisotropy Probe (WMAP) in his honor after

he passed away in September 5, 2002. WMAP, as COBE, was a great scientific success, being followed by the Planck satellite.

As far as the Brazilian scientists that have been involved in the collaborations and experiments mentioned in this paper, it is worth to mention that all of them made relevant contributions to the field of CMB and its foregrounds. Carlos Alexandre Wuensche is a senior researcher at INPE, where he works on CMB data analysis and is training new students in CMB science. Angélica de Oliveira-Costa left for the USA, where she worked in CMB data analysis and CMB foregrounds. Newton Figueiredo and Agenor Pina hold positions as professors at the Federal University of Itajubá, where they are training new students in CMB related work. Ivan Ferreira is a professor at the Institute of Physics of the University of Brasília. Larissa Santos went to Italy, where she got a Ph.D. in physics from the University of Rome Tor Vergata. She is now a post-doc in China. Rodrigo Leonardi is deeply involved in the Planck satellite mission. He joined the mission in 2005, and provided support for the integration and testing of the Low Frequency Instrument (LFI), one of the scientific payloads on board the spacecraft. In 2009, just after the satellite launch, he went to work for the Planck Science Office at the European Space Agency, supporting the scientific operations of the Planck payload, and helping with the development of the Planck Legacy Archive, a system which distributes Planck's final scientific products to the public.

In summary, over the past three decades, Brazilian participation in CMB measurements programs has produced some interesting scientific and technological results. These results were used for thesis development for both Brazilian and US students and helped the advancement of observational cosmology.

Acknowledgements I thank INPE for the continuous support to the projects I have been involved since 1982, providing the necessary manpower, laboratory facilities and financial support for them. CNPq and Fapesp provided throughout these years the financial support that allowed the completion of all the experimental and observational activities necessary to these projects. These supports were crucial for the developments described in this paper. I also thank the collaborative work of INPE's scientists, students and technical personnel in these projects. I acknowledge CNPq grant 308113/2010-1 for allowing me to continue the work related to CMB measurements.

References

Abramo, L.R.W., et al.: Phys. Rev. D **74**, 063506 (2006)
Andrade, A.P., Wuensche, C.A., Ribeiro, A.L.B.: Astrophys. J. **602**, 555 (2004)
Bernui, A., et al.: Astron. Astrophys. **454**, 409–414 (2006a)
Bernui, A., Tsallis, C., Villela, T.: Phys. Lett. A **356**, 426–430 (2006b)
Childers, J., et al.: Astrophys. J. Suppl. Ser. **158**, 124–138 (2005)
Clapp, A.C., et al.: Astrophys. J. **433**, L57–L60 (1994)
de Amici, G., et al.: Astrophys. Space Sci. **214**, 151–160 (1994)
de Gouveia dal Pino, E.M., et al.: Astrophys. J. **442**, L45 (1995); de Oliveira-Costa, A.: Ph.D. Thesis (1996)
Devlin, M., et al.: Astrophys. J. **430**, L1–L4 (1994)
Donzelli, S., et al.: Mon. Not. R. Astron. Soc. **369**, 441–448 (2006)

Ferreira, I.S.: Ph.D. Thesis (2008)
Figueiredo, N.: Ph.D. Thesis (1997)
Figueiredo, N., et al.: Astrophys. J. Suppl. Ser. **158**, 118–123 (2005)
Fixsen, D., Cheng. E., Wilkinson, D.T.: Phys. Rev. Lett. **50**, 620 (1983)
Fixsen, D., et al.: Astrophys. J. **734**, 5 (2011)
Gundersen, J.O., et al.: Astrophys. J. Lett. **413**, L1–L5 (1993)
Gundersen, J.O., et al.: Astrophys. J. **443**, L57–L60 (1995)
Kangas, M., et al.: IEEE Antennas Wirel. Propag. Lett. **4**, 245–248 (2005a); Kangas, M., et al.: IEEE Antennas Wirel. Propag. Lett. **4**, 304–307 (2005b); Kogut, A., et al.: Astrophys. J. **734**, 4 (2011)
Leonardi, R.: Ph.D. Thesis (2006)
Leonardi, R., et al.: NewAR **50**, 977–983 (2006)
Levy, A., et al.: Astrophys. J. Suppl. Ser. **177**, 419–430 (2008)
Lim, M., et al.: Astrophys. J. **469**, L69 (1996)
Lubin, P., Villela, T.: Observations of the large scale anisotropy in the cosmic background radiation at 3 mm. In: The Cosmic Background Radiation and Fundamental Physics, vol. 1, pp. 65–81 (1985)
Lubin, P., Villela, T.: Measurements of the cosmic background radiation. In: Madore, B.F., Tully, R.B. (eds.) Galaxies Distances and Deviations from Universal Expansion, vol. 180, pp. 169–175. Reidel, Dordrecht (1986)
Lubin, P., et al.: Astrophys. J. Lett. **298**, L1–L5 (1985)
Marvil, J., et al.: New Astron. **11**, 218–225 (2006)
Meinhold, P., et al.: Astrophys. J. **406**, 12–25 (1993)
Meinhold, P., et al.: Astrophys. J. Suppl. Ser. **158**, 101–108 (2005)
Mejía, J., et al.: Astrophys. J. Suppl. Ser. **158**, 109–117 (2005)
O'Dwyer, I., et al.: Astrophys. J. Suppl. Ser. **158**, 93–100 (2005)
Pina, A.: Ph.D. Thesis (2002)
Santos, L., Villela, T., Wuensche, C.A.: Astron. Astrophys. **544**, A121 (2012); Santos, L., et al.: Astron. Astrophys. **569**, A75 (2014); Schuster, J., et al.: Astrophys. J. **412**, L47–L50 (1993)
Seiffert et al.: Astrophys. J. **734**, 6 (2011)
Smoot, G., et al.: Astrophys. J. (1991)
Smoot, G., et al.: Astrophys. J. (1992)
Souza, R.R.: M.Sc. Dissertation. INPE (2000)
Staren, J., et al.: Astrophys. J. **539**, 52–56 (2000)
Strukov, I.A., Skulachev, D.P.: SvAL **10**, 1–4 (1987)
Tanaka, S., et al.: Astrophys. J. **468**, L81 (1996)
Tegmark, M., et al.: Astrophys. J. **541**, 535–541 (2000)
Tello, C.: Ph.D. Thesis (1997)
Tello, C., et al.: RaSc **34**, 575–586 (1999)
Tello, C., et al.: Astron. Astrophys. Suppl. Ser. **145**, 495–508 (2000)
Tello, C., et al.: Astron. Astrophys. **556**, A1 (2013)
Torres, S., et al.: Astrophys. Space Sci. **240**, 225–234 (1996)
Villela, T.: Ph.D. Thesis (1987)
Villela, T.: Rev. Hum. **10**(2), 171–185 (1994)
Villela, T.: Microwave instrumentation for astrophysical observations: some contributions. In: 2011 IEEE/MTTS International Microwave Symposium MTT (Microwave Theory and Techniques), vol. 1 (2011)
Villela, T., Wuensche, C.A.: Cosmic microwave background physics: observations, AIP Conf. Proc. **1132**, 141 (2009)

Villela, T., et al.: Br. Assoc. Am. Stud. **15**, 963 (1983)
Villela, T., et al.: Adv. Space Res. **26**, 1415–1418 (2000)
Villela, T., et al.: Adv. Space Res. **30**, 1333–1342 (2002)
Wuensche, C.: Ph.D. Thesis (1995)
Wuensche, C.A., Villela, T.: (2010) arXiv:1002.4902
Wuensche, C.A., Lubin, P.M., Villela, T.: An alternative algorithm for cosmic microwave background radiation sky harmonic analysis. In: Astronomical Data Analysis Software and Systems III. A.S.P. Conference Series, vol. 61, pp. 269. Astronomical Society of the Pacific, San Francisco, CA (1994)

Part III
Communications

On Dark Degeneracy

Saulo Carneiro and Humberto A. Borges

Abstract Let DF be a unified dark fluid with equation of state $p = \omega\rho$, with $-1 \leq \omega < 0$. We show that: (1) DF is observationally indistinguishable from a fluid formed by pressureless dark matter DM and a dark energy DE with equation-of-state parameter $\omega = -1$. In general these components interact and DF is non-adiabatic. (2) If the momentum transfer between the components is negligible, DE does not cluster and DM coincides with clustering matter. This is the case if DM is non-relativistic. These results imply the following corollaries. Let us consider a dark energy candidate with equation-of-state parameter $-1 \leq \omega < 0$. (1) Any observational analysis which identifies cold dark matter with clustering matter leads to $\omega \approx -1$. (2) If, in addition, such an analysis assumes that dark matter is conserved, the concordance is given by the ΛCDM model.

The vacuum energy problem is certainly the most difficult challenge in the interface of theoretical cosmology and quantum field theory. The vacuum density in curved backgrounds is derived by an appropriate renormalisation procedure which subtracts the divergent contribution of flat spacetime. In the high energy limit, the result can be used to obtain a non-singular inflationary scenario where a radiation phase is born from a previous de Sitter universe through a continuous and fast transition along which the vacuum density decays producing relativistic matter (Carneiro and Tavakol 2009; Carneiro 2012). Nevertheless, any model with matter creation from a decaying vacuum has phenomenological status because vacuum energy-momentum conservation is one of the conditions behind renormalisation techniques (Wald 1994). In the low energy limit, estimates of the energy density of the QCD vacuum condensate in the FLRW spacetime leads to $\Lambda \approx m^3 H$, where $m \approx 150\,\text{MeV}$ is the energy scale of the QCD phase transition and H is the expansion rate (Schutzhold 2002). In an approximate de Sitter universe with $\Lambda \approx H^2$, this leads to $\Lambda \approx m^6$, which coincides with the observed value of Λ. Needless to say, these estimates are not free of the difficulties inherent to non-perturbative QCD in curved backgrounds.

S. Carneiro (✉) • H.A. Borges
Instituto de Física, Universidade Federal da Bahia, Campus de Ondina, Salvador, BA 40210-340, Brazil
e-mail: saulo.carneiro@pq.cnpq.br; humberto@ufba.br

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_10

These theoretical limitations oblige us to look for an observational route to the late-time cosmological term, usually referred to as dark energy (DE). Observations show that the Universe is in accelerated expansion. This means that the dark energy equation-of-state parameter ω, defined by

$$p = \omega\rho, \tag{1}$$

is a negative function of time.[1] However, as pointed out in Kunz (2009), background observations cannot uniquely fix this function because the effects of DE and dark matter (DM) are degenerated in a way that will be clarified below. We will show that such a degeneracy can be broken at perturbation level, provided that:

- Dark matter is defined as the clustering component observed in galaxies and clusters.
- Dark matter is assumed to be cold, that is, non-relativistic.

On this basis the degeneracy will be reduced to two distinct classes of DE models with $\omega = -1$, namely the ΛCDM model, with a constant Λ, and interacting models with an energy flux from DE to DM. We will also show a specific example of dark degeneracy, namely a unified dark sector modeled by a single scalar field.

Dark degeneracy is formulated here as follows. Let us split dark fluid (1) as

$$\rho = \Lambda + \rho_m, \tag{2}$$

$$p_\Lambda = -\Lambda, \tag{3}$$

$$p_m = 0, \tag{4}$$

with $\Lambda > 0$. From (1)–(4) we have

$$\rho_m = -\frac{\omega + 1}{\omega}\,\Lambda. \tag{5}$$

For $-1 \leq \omega < 0$ we have $\rho_m \geq 0$, and this component can be interpreted as dark matter. This degeneracy in the definitions of DE and DM is unavoidable at the background level. Nevertheless, we will see below that the degeneracy can be broken at perturbation level if we correctly define the observed dark matter and assume that DM is non-relativistic.

With densities and pressures given by (2)–(4), the Friedmann and continuity equations assume the form[2]

$$\rho_m + \Lambda = 3H^2, \tag{6}$$

[1] We are assuming that the dark fluid respects the weak energy condition, i.e. $\rho > 0$ and $-1 \leq \omega < 0$.

[2] We are considering the spatially flat case and using $c = \hbar = 8\pi G = 1$.

$$\dot{\rho}_m + 3H\rho_m = -\dot{\Lambda}, \tag{7}$$

where the dot means derivative with respect to cosmological time. Equation (7) expresses total energy conservation, and it means that the two components may interact in general (Zimdahl et al. 2001). Differentiating (6) and using (7) we derive

$$\rho_m = -2\dot{H}. \tag{8}$$

Using (5) in (6) we have

$$\Lambda = -3\omega H^2. \tag{9}$$

Differentiating (9) and using (8), (6) and (5) we obtain

$$\dot{\Lambda} = \left(3\omega H - \frac{\dot{\omega}}{\omega + 1}\right)\rho_m. \tag{10}$$

Hence, Eq. (7) can be rewritten as

$$\dot{\rho}_m + 3H\rho_m = \Gamma\rho_m, \tag{11}$$

with Γ given by

$$\Gamma = \frac{\dot{\omega}}{\omega + 1} - 3\omega H. \tag{12}$$

Writing $\rho_m = Mn$, where M and n are defined as the DM particle mass and number density respectively, and using $H = \dot{a}/a$, where a is the scale factor, we obtain from (11)

$$\Gamma = \frac{1}{na^3}\frac{d}{dt}(na^3), \tag{13}$$

which shows that Γ defines the rate of DM creation.

The energy-conservation equation (7) can be written in the covariant form

$$T^{\mu\nu}_{m\,;\nu} = Q^\mu, \tag{14}$$

$$T^{\mu\nu}_{\Lambda\,;\nu} = -Q^\mu, \tag{15}$$

where $T^{\mu\nu}_A$ $(A = m, \Lambda)$ are the energy-momentum tensors of each component, and Q^μ is the energy-momentum transfer from DE to DM. The latter can be decomposed as

$$Q^\mu = Qu^\mu + \bar{Q}^\mu, \tag{16}$$

with $u^\mu \bar{Q}_\mu = 0$, where u^μ is the dark fluid 4-velocity. From (3) we see that $T^{\mu\nu}_\Lambda = \Lambda g^{\mu\nu}$, which shows that Λ is an invariant. Hence, from (15) it is easy to obtain

$$Q = -\Lambda_{,\nu} u^\nu, \tag{17}$$

$$\bar{Q}^\mu = \Lambda_{,\nu} \left(u^\mu u^\nu - g^{\mu\nu}\right). \tag{18}$$

In the dark fluid comoving frame we have $\bar{Q}^\mu = 0$, and $Q = -\dot{\Lambda} = \Gamma \rho_m$. The former shows that there is no momentum transfer in the isotropic background, while the latter shows that Q represents the energy transfer between the components. For pressureless DM we can write $Q = \Gamma g_{\mu\nu} T^{\mu\nu}_m$, which shows that Γ is also invariant.

A linear perturbation of (18) leads to

$$\delta \bar{Q}^0 = 0, \tag{19}$$

$$\delta \bar{Q}_i = (\delta\Lambda + \dot{\Lambda}\theta)_{,i}, \tag{20}$$

where θ is the dark fluid velocity potential, $\vec{\nabla}\theta = \delta\vec{u}$. Assuming that dark matter is non-relativistic, the momentum transfer should be negligible compared to the energy transfer, that is, dark matter particles are created with negligible momentum compared to their rest energy. This means that $\bar{Q}^\mu$ is zero at both background and perturbation levels, which also means that DM follows geodesics in the dark fluid rest frame (Koyama et al. 2009). Hence, from (20) we obtain

$$\delta\Lambda^c \equiv \delta\Lambda + \dot{\Lambda}\theta = 0, \tag{21}$$

where $\delta\Lambda^c$ is the (gauge-invariant) comoving perturbation of the DE component (Malik and Wands 2009). Equation (21) means that if dark matter in the decomposition (2) is non-relativistic, the dark energy defined by (2) and (3) does not cluster. Therefore, the DM component defined in (2) and (4) coincides with that observed in galaxies and clusters.

Condition (21) is also necessary to avoid the presence of unobserved oscillations and instabilities in the DM power spectrum. From the perturbed Einstein equations for a perfect fluid we have Borges et al. (2013)

$$\Phi_B'' + 3\mathcal{H}(1 + c_a^2)\Phi_B' + [2\mathcal{H}' + (1 + 3c_a^2)\mathcal{H}^2 + c_s^2 k^2]\Phi_B = 0\,, \tag{22}$$

where Φ_B is the Bardeen gravitational potential, $\mathcal{H} = aH$, k is the perturbation comoving wave-number, $c_a^2 = \dot{p}/\dot{\rho}$ is the squared adiabatic sound speed, $c_s^2 = \delta p^c/\delta\rho^c$ is the squared effective sound speed, and the prime means derivative with respect to conformal time. As in (21), the upper-index c means a gauge-invariant comoving perturbation. The term proportional to k^2 in (22) induces oscillations and instabilities in the gravitational potential, which are reflected in the matter power spectrum through the Poisson equation. In order to avoid these oscillations and instabilities we must have $|c_s| \ll 1$. For the two-component fluid (2) we have

$c_s^2 = -\delta\Lambda^c/\delta\rho^c$, and the last condition is satisfied for $\delta\Lambda^c = 0$, that is, $\bar{Q}^\mu = 0$. As discussed in Borges et al. (2013), $c_s^2 \neq c_a^2$ owing to the presence of late-time entropic perturbations related to the energy flux from dark energy to dark matter.

Let us now examine an example of a unified dark sector modeled by a single scalar field, where dark degeneracy can be made explicit.[3] In the FLRW spacetime, the energy density and pressure of a minimally coupled scalar field ϕ are given by

$$\rho = V + \frac{\dot{\phi}^2}{2}, \tag{23}$$

$$p = -V + \frac{\dot{\phi}^2}{2}, \tag{24}$$

where $V(\phi)$ is the scalar field self-interaction potential. Let us split this field into components

$$\Lambda = V - \frac{\dot{\phi}^2}{2}, \qquad p_\Lambda = -\Lambda, \tag{25}$$

$$\rho_m = \dot{\phi}^2, \qquad p_m = 0. \tag{26}$$

The field equations are given, as usual, by

$$3H^2 = V + 2H'^2, \tag{27}$$

$$\dot{\phi} = -2H', \tag{28}$$

where now the prime means derivative with respect to ϕ. Substituting (25)–(26) into the conservation equation (7) we derive the Klein-Gordon equation

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0. \tag{29}$$

All this is general, but we shall now particularise to the special case of a constant creation rate. For a constant Γ, Eqs. (11), (8) and (7) are consistent with[4]

$$\Lambda = 2\Gamma H. \tag{30}$$

Equations (9), (12) and (30) lead to the evolution equation

$$3H^2 - 2\Gamma H + 2\dot{H} = 0. \tag{31}$$

[3] For a further example, see Borges et al. (2013), Carneiro and Pigozzo (2014), where the non-adiabatic generalised Chaplygin gas is studied.

[4] More generally, (11), (8) and (7) lead to $\Lambda = 2\Gamma H + \lambda$, where λ is an integration constant.

The solution is given by Borges and Carneiro (2005)

$$H = \frac{2\Gamma/3}{1 - e^{-\Gamma t}}. \tag{32}$$

It represents a universe that evolves from an Einstein-de Sitter phase, dominated by matter, to an asymptotically de Sitter era. This solution has been tested against the most precise observations, with very good concordance (Alcaniz et al. 2012). Furthermore, it was explicitly shown in Zimdahl et al. (2011) that $\delta\Lambda^c \ll \delta\rho_m^c$ in the observed scales, as required in order to have a coincidence between our definition of DM and that observed in galaxies and clusters. Now, from (25), (27), (28) and (30) we obtain

$$V = \frac{3H^2}{2} + \Gamma H. \tag{33}$$

Substituting this potential into (27) we have

$$4H'^2 + 2\Gamma H - 3H^2 = 0. \tag{34}$$

The solution is

$$H = \frac{2\Gamma}{3}\cosh^2\left(\frac{\sqrt{3}\phi}{4}\right), \tag{35}$$

where an integration constant was conveniently chosen. The corresponding potential (33) presents a minimum at $\phi = 0$, around which it can be expanded as

$$V \approx \frac{4\Gamma^2}{3} + \frac{1}{2}\left(\frac{3\Gamma^2}{4}\right)\phi^2. \tag{36}$$

We then see that the mass of the scalar field is given by $M = \sqrt{3}\Gamma/2$. On the other hand, at the minimum we have, from (35), $H_{dS} = 2\Gamma/3$, where H_{dS} is the expansion rate in the de Sitter limit. This result can also be obtained from (32).

We have shown that dark degeneracy can be broken with the help of two observationally based statements. The first is an appropriate definition of dark matter as the clustering component observed in large scale structures, a definition that is meaningful only if the defined DE component does not cluster. We have seen that this can be achieved by additionally assuming that dark matter is cold, an assumption corroborated by observations. Therefore, we can reduce the degeneracy to two competitive classes of DE models with $\omega = -1$. In particular, this means that an observational joint analysis which identifies DM as the matter observed in large scale structures should naturally give $\omega \approx -1$. At the same time, these results translate the problem of vacuum density to the following question: Is there DM creation at late-time stages of the expansion? If DM cannot be produced at low

energies, there is a constant DE density. However, the creation of ultralight and cold DM particles cannot be ruled out. A constant-rate creation would correspond to a DE density which decays linearly with H, which would corroborate its association to the vacuum condensate of strongly interacting fields and the corresponding association of DM to condensate fluctuations. Discriminating between these two theoretical possibilities on an observational basis is an exciting endeavour.

References

Alcaniz, J.S., Borges, H.A., Carneiro, S., Fabris, J.C., Pigozzo, C., Zimdahl, W.: Phys. Lett. **B716**, 165 (2012); Pigozzo, C., Dantas, M.A., Carneiro, S., Alcaniz, J.S.: J. Cosmol. Astropart. Phys. **1108**, 022 (2011); Borges, H.A., Carneiro, S., Fabris, J.C., Pigozzo, C.: Phys. Rev. **D77**, 043513 (2008); Velten, H., Montiel, A., Carneiro, S.: Mon. Not. R. Astron. Soc. **431**, 3301 (2013); Devi, N.C., Borges, H.A., Carneiro, S., Alcaniz, J.S.: arXiv:1407.1821 (2014).

Borges, H.A., Carneiro, S.: Gen. Relativ. Gravit. **37**, 1385 (2005)

Borges, H.A., Carneiro, S., Fabris, J.C., Zimdahl, W.: Phys. Lett. **B727**, 37 (2013)

Carneiro, S.: AIP Conf. Proc. **1471**, 61 (2012); Chimento, L.P., Carneiro, S.: to appear in AIP Conf. Proc. arXiv:1402.2311 (2014)

Carneiro, S., Pigozzo, C.: J. Cosmol. Astropart. Phys. **1410**, 060 (2014). arXiv:1407.7812

Carneiro, S., Tavakol, R.: Gen. Relativ. Gravit. **41**, 2287 (2009)

Koyama, K., Maartens, R., Song, Y.S.: J. Cosmol. Astropart. Phys. **0910**, 017 (2009); Wang, Y., Wands, D., Xu, L., De-Santiago, J., Hojjati, A.: Phys. Rev. **D87**, 083503 (2013); Wands, D., De-Santiago, J., Wang, Y.: Class. Quant. Gravit. **29**, 145017 (2012)

Kunz, M.: Phys. Rev. **D80**, 123001 (2009); Wasserman, I.: Phys. Rev. **D66**, 123511 (2002); Rubano, C., Scudellaro, P.: Gen. Relativ. Gravit. **34**, 1931 (2002); Aviles, A., Cervantes-Cota, J.L.: Phys. Rev. D **84**, 083515 (2011)

Malik, K.A., Wands, D.: Phys. Rep. **475**, 1–51 (2009)

Schutzhold, R.: Phys. Rev. Lett. **89**, 081302 (2002); Klinkhamer, F.R., Volovik, G.E.: Phys. Rev. **D79**, 063527 (2009); Urban, F.R., Zhitnitsky, A.R.: Phys. Rev. **D80**, 063001 (2009); Phys. Lett. **B688**, 9 (2009); Nucl. Phys. **B835**, 135 (2010); Ohta, N.: Phys. Lett. **B695**, 41 (2011); Holdom, B.: Phys. Lett. **B697**, 351 (2011); Cai, R.G., et al.: Phys. Rev. **D84**, 123501 (2011)

Wald, R.M.: Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics, Section 4.6. University of Chicago Press, Chicago (1994)

Zimdahl, W., Schwarz, D.J., Balakin, A.B., Pavón, D.: Phys. Rev. **D64**, 063501 (2001); del Campo, S., Herrera, R., Pavón, D.: J. Cosmol. Astropart. Phys. **0901**, 020 (2009); Chimento, L.P.: Phys. Rev. **D81**, 043525 (2010); He, J.H., Wang, B., Abdalla, E.: Phys. Rev. **D83**, 063515 (2011)

Zimdahl, W., Borges, H.A., Carneiro, S., Fabris, J.C., Hipolito-Ricaldi, W.S.: J. Cosmol. Astropart. Phys. **1104**, 028 (2011)

The Quantum-to-Classical Transition of Primordial Cosmological Perturbations

Nelson Pinto-Neto

Abstract There is a widespread belief that the classical small inhomogeneities which gave rise to all structures in the Universe through gravitational instability originated from primordial quantum cosmological fluctuations. However, this transition from quantum to classical fluctuations is plagued with important conceptual issues, most of them related to the application of standard quantum theory to the Universe as a whole. In this contribution, it is shown how these issues can easily be overcome in the framework of the de Broglie-Bohm quantum theory. This theory is an alternative to standard quantum theory that provides an objective description of physical reality, where rather ambiguous notions of measurement or observer play no fundamental role, and which can hence be applied to the Universe as a whole. In addition, it allows for a simple and unambiguous characterization of the classical limit. This contribution is a compilation of the works done by Grasiele Santos, Ward Struyve and myself cited in Pinto-Neto et al. (Phys Rev D 85:083506, 2012; Phys Rev D 89:023517, 2014), where all details can be found.

1 Introduction

Presently, there is a vivid discussion in the literature about conceptual issues concerning the transition from primordial cosmological quantum fluctuations to the small classical inhomogeneities which gave rise to the structures in the Universe, such as stars and galaxies (Guth and Pi 1985; Albrecht et al. 1994; Polarski and Starobinsky 1996; Lesgourgues et al. 1997; Kiefer and Polarski 1998, 2009; Burgess et al. 2008; Perez et al. 2006; Sudarsky 2011).

All cosmological pictures of structure formation, either inflationary (Liddle and Lyth 2000; Mukhanov 2005; Weinberg 2008; Lyth and Liddle 2009; Peter and Uzan 2009) or bouncing models (Novello and Perez Bergliaffa 2008), rely on the fact that in its far past the Universe became so homogeneous and isotropic that only quantum vacuum fluctuations of inhomogeneities could have survived. Initially, the

N. Pinto-Neto (✉)
ICRA - Centro Brasileiro de Pesquisas Físicas – CBPF, rua Xavier Sigaud, 150, Urca 22290-180, Rio de Janeiro, Brazil
e-mail: nelson.pinto@pq.cnpq.br

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_11

modes of the fluctuations of cosmological interest have their physical wavelengths much smaller than the curvature scale of the background, and their quantum state is very close to the Minkowski vacuum. In the course of cosmological evolution, the physical wavelengths of these fluctuations become of the size of the curvature scale, at which point they begin to feel the background gravitational field. Once the wavelengths are much bigger than the curvature scale, the fluctuations are believed to become classical. They then give rise to the structures in the Universe, through gravitational instability.

The quantum-to-classical transition of the fluctuations is often seen as a result of the squeezing of the vacuum state. The squeezing results in an approximately definite momentum, while the spread in the field distribution becomes very large (due to the Heisenberg uncertainty). It is then argued that the squeezed state is somehow equivalent to an element of an ordinary ensemble of classical fields (see e.g. Guth and Pi 1985; Albrecht et al. 1994; Polarski and Starobinsky 1996; Lesgourgues et al. 1997; Kiefer and Polarski 1998, 2009; Burgess et al. 2008; Weinberg 2008). The argument often invokes decoherence, which is assumed to lead to loss of possible quantum interference for degrees of freedom of interest, due to their interaction with other degrees of freedom.

However, as pointed out by Sudarsky and collaborators (Perez et al. 2006; Sudarsky 2011), the usual accounts of the quantum-to-classical transition have serious shortcomings. They argued that no satisfactory justification has been given on why the quantum state, which is translational and rotational invariant, and which remains so during Schrödinger time evolution, results in a non-invariant state. Even when there is suitable decoherence, which suppresses interference between different non-invariant terms into which the quantum state can be decomposed, it is not explained why one of these terms is selected. According to standard quantum theory, a transition to a non-invariant state could only be obtained by collapse. For instance, in a spherically symmetric s-wave decay state of an atom, there is an actual photon which is detected in a particular direction because somewhere in this process there is a halt in the unitary Schrödinger evolution: the wave function collapses, due to the intervention of an external agent, into a non spherically symmetric function corresponding to an actual particular momentum direction of the photon. It is very unclear what causes the collapse: the external agent could be the rest of the world outside the system and its measurement device, a classical domain, observers, macroscopic systems, and many other imprecise examples. In the cosmological context, however, this inaccurate description becomes intolerable: here we are talking about cosmological primordial fluctuations, that is, fluctuations which will give rise to all structures in the Universe. There are no external agents. Hence, as the wave function of the perturbations is always translational and rotational invariant, how these symmetries can be broken without an external agent, like in the s-wave example? This is the problem with the cosmological scenario, already acknowledged in Liddle and Lyth (2000, p. 64), Mukhanov (2005, p. 348) and Lyth and Liddle (2009, pp. 386–387).

Sudarsky and collaborators (Perez et al. 2006; De Unánue and Sudarsky 2008; León and Sudarsky 2010; Sudarsky 2011; Landau et al. 2012; Leon et al. 2013)

proposed to solve this problem in the context of spontaneous collapse theories, in which the collapse is an unambiguous, objective stochastic process. The whole framework is still under construction and may yield possible deviations from the standard predictions.

In this paper, we will address the problem within the de Broglie-Bohm theory (Bohm and Hiley 1993; Holland 1993; Dürr and Teufel 2009). In this theory, the Universe is described by the universal wave function, together with an actual configuration for gravity (e.g. an actual three metric) and a configuration for the matter (e.g. particle positions or fields). The universal wave function should satisfy the appropriate quantum cosmological wave equation. Its role is to determine the dynamics of the actual configuration, which is deterministic.

For the problem at hand, we will not consider the full quantum cosmological framework, but instead restrict ourselves to the effective theory for the perturbations in terms of the Mukhanov-Sasaki variable, which represents a gauge invariant combination of perturbations of the metric and inflaton field.[1] Although the quantum state remains translational and rotational invariant in this framework, the actual de Broglie-Bohm field corresponding to the perturbations breaks this symmetry. The initial field configuration is not determined by the theory, but in quantum equilibrium it is distributed according to the quantum mechanical distribution.

Also the question of the classical limit is straightforward in the de Broglie-Bohm theory: the classical limit is obtained whenever the actual field configuration evolves approximately according to the classical equations. We will see that the actual de Broglie-Bohm perturbations reach the classical limit at the expected stage and that the quantum equilibrium ensemble of these perturbations then corresponds exactly to the classical ensemble that is assumed in Guth and Pi (1985), Albrecht et al. (1994), Polarski and Starobinsky (1996), Lesgourgues et al. (1997), Kiefer and Polarski (1998), and Kiefer and Polarski (2009).

This contribution is a compilation of the works done by Grasiele Santos, Ward Struyve and myself cited in Pinto-Neto et al. (2012, 2014), where all details can be found.

2 Linear Cosmological Perturbations

Let us first consider the classical description of cosmological perturbations. In the next section we will turn to the quantum description.

The perturbations are considered in a background Friedmann-Robertson-Walker model, with scale factor a and uniform total matter distribution with density ρ and pressure p, and are described by the Mukhanov-Sasaki variable $v(\eta, \mathbf{x})$, which

[1] In the de Broglie-Bohm approach, it is possible to obtain the effective theory of perturbations from the full quantization of the background geometry and linear perturbations, see Falciano and Pinto-Neto (2009).

combines both fluctuations of the matter and metric. The parameter η is conformal time defined by $ad\eta = dt$, t being cosmic time. The Lagrangian for the Mukhanov-Sasaki variable, which can be derived from the Einstein-Hilbert action, is given by

$$L_v = \int d^3\mathbf{x}\frac{1}{2}\left[v'^2 + \left(\frac{z'}{z}\right)^2 v^2 - c_s^2\delta^{ij}v_{,i}v_{,j} - 2\frac{z'}{z}vv'\right], \tag{1}$$

where the primes denote derivatives with respect to conformal time and

$$z = \frac{\sqrt{\beta}}{x\mathcal{H}c_s}, \quad \beta = \frac{3}{2}\frac{8\pi G}{3c^4}a^2(\rho + p), \quad c_s^2 = \frac{dp}{d\rho}. \tag{2}$$

$\mathcal{H} = a'/a$ is the conformal Hubble function, which relates to the Hubble function $H = a^{-1}da/dt$ through $\mathcal{H} = aH$ and $x = a_0/a$ is the red-shift function. Subscripts 0 refer to present day values. The Lagrangian yields the following equations of motion for the Fourier modes $v_{\mathbf{k}}(\eta)$,

$$v_{\mathbf{k}}'' + \left(c_s^2k^2 - \frac{z''}{z}\right)v_{\mathbf{k}} = 0. \tag{3}$$

Defining $\Omega = \rho/\rho_c$, where ρ_c is the critical density today, and using the energy conservation equation

$$\frac{d\rho}{dt} + 3H(\rho + p) = 0 \quad \Rightarrow \quad \frac{d\rho}{dx} = \frac{3(\rho + p)}{x}, \tag{4}$$

we obtain

$$\beta = \tfrac{1}{2xR_H^2}\tfrac{d\Omega}{dx}, \qquad z^2 = \tfrac{1}{2c_s^2x\Omega}\tfrac{d\Omega}{dx},$$
$$c_s^2 = \tfrac{x}{3}\tfrac{d}{dx}\ln\left(\tfrac{1}{x^2}\tfrac{d\Omega}{dx}\right), \tag{5}$$

where $R_H = c/(a_0H_0)$ is the co-moving Hubble radius.

In the case the background matter is a single hydrodynamical fluid with $p = w\rho$ we have $\Omega = \Omega_0x^{3(1+w)}$ and

$$c_s^2 = w, \quad z^2 = \frac{3(1+w)}{2wx^2}, \tag{6}$$

and Eq. (3) reduces to

$$v_{\mathbf{k}}'' + \left(wk^2 - \frac{a''}{a}\right)v_{\mathbf{k}} = 0. \tag{7}$$

The general solution of the mode equation (3) can be formally expanded in powers of k^2 as Mukhanov (2005)

$$\frac{v_{\mathbf{k}}}{z} = A_{1,\mathbf{k}}\left[1 - k^2 \int_{\eta_i}^{\eta} \frac{d\bar{\eta}}{z^2(\bar{\eta})} \int^{\bar{\eta}} c_s^2 z^2 \left(\bar{\bar{\eta}}\right) d\bar{\bar{\eta}} + \ldots\right] + A_{2,\mathbf{k}} \int_{\eta_i}^{\eta} \frac{d\bar{\eta}}{z^2(\bar{\eta})}\left[1 - k^2 \int^{\bar{\eta}} c_s^2 z^2 \left(\bar{\bar{\eta}}\right) d\bar{\bar{\eta}} \int^{\bar{\bar{\eta}}} \frac{d\bar{\bar{\bar{\eta}}}}{z^2\left(\bar{\bar{\bar{\eta}}}\right)} + \ldots\right], \tag{8}$$

where we have presented the terms up to order $\mathcal{O}(k^2)$. The lower bounds η_i in the integrals are related to initial conditions that depend on the specific model being analyzed and we consider only the primitives of the indefinite integrals. The coefficients $A_{1,\mathbf{k}}$ and $A_{2,\mathbf{k}}$ are two constants, determined by the initial conditions, which are roughly the same order of magnitude.[2] We will be interested in the situation where $c_s^2 k^2 \ll z''/z$ and hence we will take only the first terms of the series above.

This description of the perturbations is valid, in the case where entropy perturbations are negligible, whenever the dynamics describing the inflationary phase or a bounce is given by the General Relativity Einstein's equations. They are also valid in the case of the quantum bounce which we will be discussed in section IV. For general bounces, it is not clear that solution (8) is valid through the bounce, neither do we have any particular analytic solution in order to evaluate it away from the bounce as we have in the case of the quantum bounce presented in Peter et al. (2007), which will be considered in section IV. However, if the bounce is short enough, an estimate of Eq. (8) away from the bounce, where General Relativity is valid, will be sufficient to evaluate the orders of magnitude of the amplitudes, assuming that a short bounce does not change the mode evolution too much.

Let us first consider the inflationary case. At the onset of inflation, physical modes are assumed to have wavelengths much smaller than the curvature scale, i.e., $c_s^2 k^2 \gg z''/z$, so that their equation of motion approximately reduces to that of a free mode in Minkowski space-time. In many inflationary models [like power-law inflation or slow-roll (Peter and Uzan 2009)], the precise behavior of these physical modes at early times ($\eta \to \eta_i$, where η_i is some initial time, with $|\eta_i| \gg 1$), is given by

$$y_{\mathbf{k}}(\eta) \sim \mathrm{e}^{-\mathrm{i}k\eta}\left(1 + \frac{A_k}{\eta} + \ldots\right). \tag{9}$$

The physical modes will grow larger during inflation and will eventually obtain wavelengths much bigger than the curvature scale, i.e., $c_s^2 k^2 \ll z''/z$. At that stage,

[2] In fact, as A_1 and A_2 depend on $\mathbf{k}$, this assertion depends on the scale we are talking about. For an almost scale invariant spectrum of cosmological perturbations, the A_2 term is larger than the A_1 term for all scales of cosmological interest, which enforces the argumentation described below for the transition of quantum-to-classical behavior in bouncing models. Only for very short wavelengths can the A_1 term be bigger than A_2.

in many models, the modes are approximately given by

$$y_{\mathbf{k}}(\eta) \sim A_k^d \eta^{\alpha_d} + A_k^g \eta^{\alpha_g} \approx A_k^g \eta^{\alpha_g}, \tag{10}$$

where $\alpha_d > 0$ and $\alpha_g < 0$. The first term represents the decaying mode and the second one the growing mode, which dominates in $y_{\mathbf{k}}$.

In order to analyze the bouncing models, we focus on the term

$$A_{2,\mathbf{k}} \int_{-\infty}^{\eta} \frac{d\bar{\eta}}{\bar{z}^2}, \tag{11}$$

which appears in the solution (8) (we are assuming that the contracting phase begins at a very large negative conformal time, which we take to be $\eta_i \to -\infty$). This term grows with time. We can write

$$\int_{-\infty}^{\eta} \frac{d\bar{\eta}}{z^2(\bar{\eta})} = \left(B - \int_{\eta}^{\infty} \frac{d\bar{\eta}}{z^2(\bar{\eta})} \right), \tag{12}$$

where

$$B = \int_{-\infty}^{\infty} d\eta \, z^{-2} \tag{13}$$

is a constant. For the case the bounce is dominated by a single fluid with equation of state parameter w_q, this constant was evaluated in Vitenti and Pinto-Neto (2012) and reads

$$B \approx \frac{4x_b}{3(1-w_q)E(x_b)z^2(x_b)}, \tag{14}$$

where $E = H/H_0 = \sqrt{\Omega(x)}$, a subscript b refers to the values of the physical quantities at the bounce. It must be understood that, although evaluated at x_b, the functions $E(x)$ and $z^2(x)$ in Eq. (14) are the usual general relativistic expressions for them which are valid just before the bounce but maybe not through the bounce itself.

For realistic bounces occurring at energy scales bigger than the nucleosynthesis energy scale we have $x_b = a_0/a_b \gg 10^{10}$. Furthermore, since $w_q \ll 1$, which is needed in order to obtain a scale invariant spectrum (see Peter et al. 2007), it follows that $B \gg 10^{10}$. Hence, the solution for the mode functions $v_{\mathbf{k}}$ around the bounce (where η is of order $\mathcal{O}(1)$) is given by

$$\begin{aligned} v_{\mathbf{k}} &\approx [A_{1,\mathbf{k}} + A_{2,\mathbf{k}} B] z(\eta) - A_{2,\mathbf{k}} z(\eta) \int_{\eta}^{\infty} \frac{d\bar{\eta}}{z^2(\bar{\eta})} \\ &\approx A_{2,\mathbf{k}} z(\eta) \left[B - \int_{\eta}^{\infty} \frac{d\bar{\eta}}{z^2(\bar{\eta})} \right]. \end{aligned} \tag{15}$$

In the last approximation we assumed that A_1 and A_2 are roughly of the same order.

Equation (15) will be used to achieve the classical limit. Remember that it is valid for perturbation modes for which $c_s^2 k^2 \leqslant z''/z$ (i.e., when their physical wavelengths are much larger than the curvature scale of the contracting background) and in case the background space has already contracted enough.

3 The de Broglie-Bohm Approach to Perturbations

Let us now turn to the corresponding quantum theory. The details of the quantization can for example be found in Polarski and Starobinsky (1996). In the functional Schrödinger picture, the state is a wave functional $\Psi(y, y^*, \eta)$ (the symbol y corresponds to the Mukhanov-Sasaki variable v) and satisfies a functional Schrödinger equation determined by the quantum Hamiltonian corresponding to (1). In the special case the wave functional is a product $\Psi = \Pi_{\mathbf{k}\in\mathbb{R}^{3+}} \Psi_{\mathbf{k}}(y_{\mathbf{k}}, y_{\mathbf{k}}^*, \eta)$, like for the vacuum, each mode wave function satisfies the Schrödinger equation (note that we formally treat $y_{\mathbf{k}}, y_{\mathbf{k}}^*$ as independent fields; equivalently, their real and imaginary components can be used, defined by $y_{\mathbf{k}} = (y_{\mathbf{k},\mathbf{r}} + \mathrm{i} y_{\mathbf{k},\mathbf{i}})/\sqrt{2}$):

$$\mathrm{i}\frac{\partial \Psi_{\mathbf{k}}}{\partial \eta} = \left[-\frac{\partial^2}{\partial y_{\mathbf{k}}^* \partial y_{\mathbf{k}}} + c_s^2 k^2 y_{\mathbf{k}}^* y_{\mathbf{k}} - \mathrm{i}\frac{z'}{z}\left(\frac{\partial}{\partial y_{\mathbf{k}}^*} y_{\mathbf{k}}^* + y_{\mathbf{k}} \frac{\partial}{\partial y_{\mathbf{k}}}\right)\right] \Psi_{\mathbf{k}}. \tag{16}$$

For the ground state, the mode wave functions are given by (see Polarski and Starobinsky 1996)

$$\Psi_{\mathbf{k}} = \frac{1}{\sqrt{2\pi}|f_k(\eta)|} \exp\left\{-\frac{1}{2|f_k(\eta)|^2}|y_{\mathbf{k}}|^2 + \mathrm{i}\left[\left(\frac{|f_k(\eta)|'}{|f_k(\eta)|} - \frac{z'}{z}\right)|y_{\mathbf{k}}|^2 - \int^{\eta} \frac{d\tilde{\eta}}{2|f_k(\tilde{\eta})|^2}\right]\right\}, \tag{17}$$

with f_k a solution to the classical mode equation, and $f_k(\eta_i) = 1/\sqrt{2k}$, where $|\eta_i| \gg 1$. This state is homogeneous and isotropic.

In order to pass to the de Broglie-Bohm approach, we write $\Psi_{\mathbf{k}} = R_{\mathbf{k}} \mathrm{e}^{\mathrm{i}S_{\mathbf{k}}}$, with $R_{\mathbf{k}} = |\Psi_{\mathbf{k}}|$, so that the Schrödinger equation (16) yields the two real equations

$$\frac{\partial S_{\mathbf{k}}}{\partial \eta} + \frac{\partial S_{\mathbf{k}}}{\partial y_{\mathbf{k}}^*}\frac{\partial S_{\mathbf{k}}}{\partial y_{\mathbf{k}}} + c_s^2 k^2 y_{\mathbf{k}}^* y_{\mathbf{k}} + \frac{z'}{z}\left(\frac{\partial S_{\mathbf{k}}}{\partial y_{\mathbf{k}}^*} y_{\mathbf{k}}^* + y_{\mathbf{k}}\frac{\partial S_{\mathbf{k}}}{\partial y_{\mathbf{k}}}\right) - \frac{1}{R_{\mathbf{k}}}\frac{\partial^2 R_{\mathbf{k}}}{\partial y_{\mathbf{k}}^* \partial y_{\mathbf{k}}} = 0, \tag{18}$$

$$\frac{\partial R_{\mathbf{k}}^2}{\partial \eta} + \frac{\partial}{\partial y_{\mathbf{k}}}\left[R_{\mathbf{k}}^2\left(\frac{\partial S_{\mathbf{k}}}{\partial y_{\mathbf{k}}^*} + \frac{z'}{z} y_{\mathbf{k}}\right)\right] + \frac{\partial}{\partial y_{\mathbf{k}}^*}\left[R_{\mathbf{k}}^2\left(\frac{\partial S_{\mathbf{k}}}{\partial y_{\mathbf{k}}} + \frac{z'}{z} y_{\mathbf{k}}^*\right)\right] = 0. \tag{19}$$

We can then postulate an actual field $y(\eta, \mathbf{x})$, whose modes obey the de Broglie-Bohm *guidance equations*

$$y'_{\mathbf{k}} = \frac{\partial S_{\mathbf{k}}}{\partial y^*_{\mathbf{k}}} + \frac{z'}{z} y_{\mathbf{k}}, \quad y^{*\prime}_{\mathbf{k}} = \frac{\partial S_{\mathbf{k}}}{\partial y_{\mathbf{k}}} + \frac{z'}{z} y^*_{\mathbf{k}}. \tag{20}$$

By Eq. (19), these equations of motion guarantee that the mode distribution $R^2_{\mathbf{k}} = |\Psi_{\mathbf{k}}|^2$ is preserved over time (and this is exactly the motivation to introduce these equations of motion). That is, if the initial distribution of the field modes is given by $|\Psi_{\mathbf{k}}|^2$, then it will be given by $|\Psi_{\mathbf{k}}|^2$ at any time. The particular distribution $R^2_{\mathbf{k}} = |\Psi_{\mathbf{k}}|^2$ plays the role of an equilibrium distribution and is called the *quantum equilibrium distribution*.

The guidance equations also follow from the Hamilton equation $y'_{\mathbf{k}} = p_{\mathbf{k}} + \frac{z'}{z} y_{\mathbf{k}}$ corresponding to the classical Hamiltonian, with $p_{\mathbf{k}}$ replaced by $\frac{\partial S_{\mathbf{k}}}{\partial y^*_{\mathbf{k}}}$. The Eq. (18) then resembles the Hamilton-Jacobi equation corresponding to the classical Hamiltonian, with an extra potential

$$Q_{\mathbf{k}} \equiv -\frac{1}{R_{\mathbf{k}}} \frac{\partial^2 R_{\mathbf{k}}}{\partial y^*_{\mathbf{k}} \partial y_{\mathbf{k}}}, \tag{21}$$

called the *quantum potential*.

Taking the time derivative of these equations and using (18), we obtain

$$y''_{\mathbf{k}} + \left(c_s^2 k^2 - \frac{z''}{z} \right) y_{\mathbf{k}} = -\frac{\partial Q_{\mathbf{k}}}{\partial y^*_{\mathbf{k}}}. \tag{22}$$

This is just the classical equation of motion (3) (written in terms of y), modified with an additional *quantum force*. The classical limit is obtained when this quantum force is negligible.

For the ground state, the guidance equations of motion are easily integrated and yield

$$y_{\mathbf{k}}(\eta) = y_{\mathbf{k}}(\eta_i) \frac{|f_k(\eta)|}{|f_k(\eta_i)|}. \tag{23}$$

Note that this result is independent of the precise form of $f_k(\eta)$ and hence is quite general. In quantum equilibrium, the probability distribution of the initial configuration $y_{\mathbf{k}}(\eta_i)$ is given by $|\Psi_{\mathbf{k}}\left(y_{\mathbf{k}}(\eta_i), y^*_{\mathbf{k}}(\eta_i), \eta_i\right)|^2$.

For physical wave lengths and $\eta \to \eta_i$, the behavior of $f_k(\eta)$ is given by Eq. (9). As such,

$$y_{\mathbf{k}}(\eta) \sim \left(1 + \frac{\mathrm{Re} A_k}{\eta} + \ldots \right) \tag{24}$$

(in many inflationary scenarios, $\mathrm{Re}A_k = 0$ and the first order term disappears). So $y_{\mathbf{k}}$ tends to be time independent for $|\eta| \gg 1$. (This is compatible with the fact that the de Broglie-Bohm field configuration is stationary for the ground state of a quantized scalar field in Minkowski space-time (Holland 1993).) Hence, the time dependence of the de Broglie-Bohm field configuration is completely different from that of classical solutions, which oscillates for $|\eta| \gg 1$ and $c_s^2 k^2 \gg z''/z$, see Eq. (9).

When the wavelengths become much bigger than the curvature scale ($c_s^2 k^2 \ll z''/z$), the behavior of $f_k(\eta)$ is approximately given by the growing mode, see Eq. (10) for inflation and Eq. (15) for bouncing models, so that $|f_k|$ equals f_k, up to a time-independent complex factor. As such, the de Broglie-Bohm field configuration approximately evolves according to the classical field equation (3), so that the classical limit has been attained.

The classical limit can also be investigated by examining the behavior of the quantum force and leads to the same result. For the ground state (17), the quantum force is given by

$$F_{Q,\mathbf{k}} \equiv -\frac{\partial Q_{\mathbf{k}}}{\partial y_{\mathbf{k}}^*} = \frac{y_{\mathbf{k}}}{4|f_k|^4} \tag{25}$$

for the mode $\mathbf{k}$. The classical force can be read from Eq. (3) and their ratio is

$$\frac{F_{C,\mathbf{k}}}{F_{Q,\mathbf{k}}} = -4|f_k|^4 \left(c_s^2 k^2 - \frac{z''}{z}\right). \tag{26}$$

For $c_s^2 k^2 \gg \frac{z''}{z}$, this ratio is approximately -1, so that the quantum force cancels the classical force and the mode evolves freely. The guidance equations further restrict the velocities to be zero, so that the mode stands still. For $c_s^2 k^2 \ll \frac{z''}{z}$, this ratio becomes very big because of the growing mode, so that the quantum force becomes negligible with respect to the classical force. As a result, the mode will evolve according to the classical equation of motion. In this way, the transition from quantum to classical behavior is clear and simple.

Let us now turn to the statistical predictions. First, let us denote $y(\eta, \mathbf{x}; y_i)$, with y_i a field on space, a solution to the guidance equations such that $y(\eta_i, \mathbf{x}; y_i) = y_i(\mathbf{x})$. As noted before, if the initial field y_i is distributed according to quantum equilibrium, i.e., $|\Psi(y_i, \eta_i)|^2$, then $y(\eta, \mathbf{x}; y_i)$ will be distributed according to $|\Psi(y, \eta)|^2$. For such an equilibrium ensemble, we can consider the two-point correlation function

$$\langle y(\eta, \mathbf{x}) y(\eta, \mathbf{x} + \mathbf{r}) \rangle_{\mathrm{dBB}} \tag{27}$$

$$= \int \mathcal{D}y_i |\Psi(y_i, \eta_i)|^2 y(\eta, \mathbf{x}; y_i) y(\eta, \mathbf{x} + \mathbf{r}; y_i) \tag{28}$$

$$= \int \mathcal{D}y |\Psi(y, \eta)|^2 y(\mathbf{x}) y(\mathbf{x} + \mathbf{r}) \tag{29}$$

which is the usual expression for the correlation function, and can be calculated to yield

$$\langle y(\eta, \mathbf{x}) y(\eta, \mathbf{x}+\mathbf{r})\rangle_{\mathrm{dBB}} = \frac{1}{2\pi^2}\int dk \frac{\sin kr}{r} k|f_k(\eta)|^2, \tag{30}$$

in the case of the ground state. Just as in the usual account, this ensemble average should approximately be equal to the spatial average of an actual field configuration for the universe. This could be justified by adopting the usual assumption that the spatial integral can be taken over a larger volume than that over which the fields are correlated (Liddle and Lyth 2000; Mukhanov 2005; Weinberg 2008).

4 Conclusions

Using the de Broglie-Bohm theory, we have established the quantum-to-classical transition of primordial perturbations for inflationary models and for a large class of bouncing models. We also considered the particular example of a quantum bounce described in Peter et al. (2007), for which there is an analytical solution for perturbations during the bounce.

While the wave function of the perturbations is homogeneous and isotropic, the actual perturbation $v(\mathbf{x}, \eta)$ is not symmetric. It is a superposition of $A_{1,k}$ and $A_{2,k}$ modes. As we have shown, after an inflationary phase or a long contraction, the $A_{1,k}$ mode becomes negligible compared to the $A_{2,k}$ mode [see Eq. (15)], so that the mode $v_{\mathbf{k}}(\eta)$ and hence the field $v(\mathbf{x}, \eta)$ behave classically.

Note that we did not invoke decoherence. If at some stage in this transition there is decoherence in the field basis, this will not destroy the classical behavior of the fields.

Acknowledgements I would like to thank CNPq of Brazil for financial support and Daniel Sudarsky for some illuminating discussions on this problem.

References

Albrecht, A., Ferreira, P., Joyce, M., Prokopec, T.: Phys. Rev. D **50**, 4807 (1994). arXiv:astro-ph/9303001v2
Bohm, D., Hiley, B.J.: The Undivided Universe. Routledge, New York (1993)
Burgess, C.P., Holman, R., Hoover, D.: Phys. Rev. D **77**, 063534 (2008). arXiv:astro-ph/0601646v1
De Unánue, A., Sudarsky, D.: Phys. Rev. D **78**, 043510 (2008). arXiv:0801.4702v2 [gr-qc]
Dürr, D., Teufel, S.: Bohmian Mechanics. Springer, Berlin (2009)
Falciano, F.T., Pinto-Neto, N.: Phys. Rev. D **79**, 023507 (2009). arXiv:0810.3542v2 [gr-qc]
Guth, A.H., Pi, S.-Y.: Phys. Rev. D **32**, 1899 (1985)
Holland, P.R.: The Quantum Theory of Motion. Cambridge University Press, Cambridge (1993)
Kiefer, C., Polarski, D.: Ann. Phys. **7**, 137 (1998). arXiv:gr-qc/9805014v2

Kiefer, C., Polarski, D.: Adv. Sci. Lett. **2**, 164 (2009). arXiv:0810.0087v2 [astro-ph]
Landau, S.J., Scóccola, C.G., Sudarsky, D.: Phys. Rev. D **85**, 123001 (2012)
León, G., Sudarsky, D.: Class. Quantum Gravit. **27**, 225017 (2010). arXiv:1003.5950v2 [gr-qc]
León, G., Landau, S.J., Sudarsky, D.: Phys. Rev. D **88**, 023526 (2013)
Lesgourgues, J., Polarski, D., Starobinsky, A.A.: Nucl. Phys. B **497**, 479 (1997). arXiv:gr-qc/9611019v1
Liddle, A.R., Lyth, D.H.: Cosmological Inflation and Large-Scale Structure. Cambridge University Press, New York (2000)
Lyth, D.H., Liddle, A.R.: The Primordial Density Perturbation. Cambridge University Press, Cambridge (2009)
Mukhanov, V.: Physical Foundations of Cosmology. Cambridge University Press, New York (2005)
Novello, M., Perez Bergliaffa, S.E.: Phys. Rep. **463**, 127 (2008). arXiv:0802.1634v1 [astro-ph]
Perez, A., Sahlmann, H., Sudarsky, D.: Class. Quantum Gravit. **23**, 2317 (2006). arXiv:gr-qc/0508100v3
Peter, P., Uzan, J.-P.: Primordial Cosmology. Oxford University Press, Oxford (2009)
Peter, P., Pinho, E., Pinto-Neto, N.: Phys. Rev. D **75**, 023516 (2007). arXiv:hep-th/0610205
Pinto-Neto, N., Santos, G., Struyve, W.: Phys. Rev. D **85**, 083506 (2012)
Pinto-Neto, N., Santos, G., Struyve, W.: Phys. Rev. D **89**, 023517 (2014)
Polarski, D., Starobinsky, A.A.: Class. Quantum Gravit. **13**, 377 (1996). arXiv:gr-qc/9504030v2
Sudarsky, D.: Int. J. Mod. Phys. D **20**, 509 (2011). arXiv:0906.0315v3 [gr-qc]
Vitenti, S., Pinto-Neto, N.: Phys. Rev. D **85**, 023524 (2012). arXiv:1111.0888 [astro-ph.CO]
Weinberg, S.: Cosmology. Oxford University Press, New York (2008)

A Path-Integral Approach to CMB

Paulo H. Reimberg

Abstract The fluctuations in the temperature and polarization of the cosmic microwave background are described by a hierarchy of Boltzmann equations. In its integral form, this Boltzmann hierarchy can be converted from the usual Fourier-space base into a position-space and causal description. We shall here restrict ourselves to a particular case when the evolution of the temperature is uncoupled to the polarization, and show that probability densities for random flights play a key role in this description. The integral system can be treated as a perturbative series in the number of steps of the random flights, and the properties of random flight probabilities impose constraints on the domains of dependence.

1 Introduction

In the early stages of the evolution of the Universe, radiation and matter (which was fully ionized up to $t_{dec} = 380.000$ years after the Big Bang) were coupled by different kinds of processes. As the universe expands and cools down, these interaction channels progressively close, and eventually (at $t \simeq t_{dec}$) photons decouple from matter, propagating freely ever since—this is, of course an approximation that neglects some astrophysical effects. This is the Cosmic Microwave Background (CMB) radiation.

If we point a telescope at any given direction, most of the CMB photons that we observe have propagated freely ever since they decoupled from matter at $t \sim t_{dec}$. However, their energies and polarizations bear the imprint of the physics during the epoch leading to the time of decoupling. Roughly speaking, that period can be divided in two phases: one of strong coupling, when photons are more energetic, and a late phase, when photons are only barely coupled to matter through elastic low-energy Thomson scatterings with electrons.

The temperature and polarization of the CMB photons are mathematically described by a system of Boltzmann's equations codifying the collisional physics and free propagation (Straumann 2006). Because of metric perturbations due to

P.H. Reimberg (✉)
Instituto de Física, Universidade de São Paulo, CP 66318 05314-970, São Paulo, Brazil
e-mail: phreimberg@usp.br

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_12

overdensities (or underdensities) of matter in the Universe, we don't measure exact equilibrium distributions, but, more interestingly, fluctuations around equilibrium distributions. These fluctuations allow cosmologists to extract information about the matter content and evolution of the Universe, and a large number of experiments have been (and are) devoted to its measurement, such as the satellites COBE, WMAP, PLANCK, and others (Smoot et al. 1992; Bennett et al. 2012; Planck Collaboration et al. 2014).

Since we observe CMB photons coming from all directions in sky, multipolar decompositions for temperature and polarization are usually employed, and have interesting consequences: because of the nature of low energy photon-electrons interactions, temperature and polarization are only coupled by their quadrupoles. The particular form of low energy limit for Thomson scattering S-Matrix implies also that the evolution of the quadrupoles during the phase of weak coupling between radiation and matter can be related to averages of source terms weighted by random flight probability density functions.

Random flights are a classical problem in mathematical physics (Watson 1944), with many applications in physics and astronomy (Chandrasekhar 1943). The problem was first proposed in the beginning of the twentieth century in context of the study of bird migrations. Lord Rayleigh, soon after, applied the same ideas in acoustics. Further contributions on this subject are described in Dutka (1985). In very simple terms, a random flight (in a D-dimensional space) is the trajectory performed by a body which moves at constant speed and changes its direction of motion into another random direction after randomly or deterministic set of time intervals. If the movement has a fixed origin, we may ask, based on the length of the intermediate displacements, as well as on the number of displacements, what is the probability for the moving body to reach a distance r from the origin.

Following the description introduced in Reimberg and Abramo (2013), that we shall revisit here, the number of steps performed during the random flight is related to the scatterings suffered by the photons, which naturally propagate at constant speed; the steps' lengths relate to the displacement of the photons between scatterings; and the isotropic distribution of angles at each change of direction of propagation is due to the particular behavior of the quadrupole scattering at those energy scale. Finally, the dimension of the space where the random flight happens is related to number of degrees of freedom associated with the quadrupole.

This paper is organized as follows: after introducing the Boltzmann hierarchy in the integral form in Sect. 2, we decouple the evolution of the temperature from the polarization, and show the consequences of this simplification for the expression of the temperature and polarization in Sect. 3. We then show, in Sect. 4, how the family of integrals over spherical Bessel functions that appears in that description is related to random flight probability densities functions, and give an interpretation to the process of polarization of CMB photons in a path-integral fashion. We should observe that uncoupled evolution of the temperature is taken here as a simplification, but can be treated in full generality, as shown in Reimberg and Abramo (2013), that serves as basis for the topics presented in the sequence.

2 Boltzmann's Equations

The hierarchy of Boltzmann's equations describing CMB temperature fluctuations and polarization can be written in terms of a set of coupled integral equations—see, e.g., Seljak and Zaldarriaga (1996). Let the temperature anisotropies observed at position $\mathbf{x}_o$ and at (conformal) time η_o along the direction $\hat{\mathbf{o}}$ be given in terms of its momenta as:

$$\Theta(\mathbf{x}_o, \eta_o, \hat{\mathbf{o}}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \, \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}_o} \, 4\pi \sum_{lm} i^l \, \Theta_l(\mathbf{k}, \eta_o) \, \mathrm{Y}^*_{lm}(\hat{\mathbf{k}}) \, \mathrm{Y}_{lm}(\hat{\mathbf{o}}) \; . \tag{1}$$

Similarly, the polarization, in terms of the usual Stokes parameters Q, U and I, is decomposed as:

$$\frac{Q + iU}{4I}(\mathbf{x}_o, \eta_o, \hat{\mathbf{o}}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \, \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}_o} \, 4\pi \sum_{lm} i^l \, \alpha_l(\mathbf{k}, \eta_o) \, \mathrm{Y}^*_{lm}(\hat{\mathbf{k}}) \, {}_2\mathrm{Y}_{lm}(\hat{\mathbf{o}}) \; , \tag{2}$$

where ${}_2\mathrm{Y}_{lm}$ are the spin-weighted spherical harmonics (Straumann 2006).

The momenta of the CMB temperature and polarization are then given by the integral equations:

$$\begin{aligned} \Theta_l(\mathbf{k}, \eta_o) = \int_0^{\eta_o} d\eta \, \mathrm{e}^{-\mu(\eta)} \Bigg\{ \mu'(\eta) \Bigg[& \theta_{SW}(\mathbf{k}, \eta) \, j_l(k\Delta\eta_0) - k \, V_b(\mathbf{k}, \eta) \, j_l'(k\Delta\eta_0) \\ & + \frac{1}{2}\Big[\Theta_2(\mathbf{k}, \eta) - \sqrt{6}\alpha_2(\mathbf{k}, \eta)\Big] \Big[\frac{3}{2} j_l''(k\Delta\eta_0) + \frac{1}{2} j_l(k\Delta\eta_0) \Big] \Bigg] \\ & + (\Psi' + \Phi')(\mathbf{k}, \eta) \, j_l(k\Delta\eta_0) \Bigg\} \; , \end{aligned} \tag{3}$$

and

$$\alpha_l(\mathbf{k}, \eta_o) = -\frac{3}{2} \sqrt{\frac{(l+2)!}{(l-2)!}} \int_0^{\eta_o} d\eta \, \mu'(\eta) \mathrm{e}^{-\mu(\eta)} \, \frac{1}{2}\Big[\Theta_2(\mathbf{k}, \eta) - \sqrt{6}\alpha_2(\mathbf{k}, \eta)\Big] \frac{j_l(k\Delta\eta_0)}{(k\Delta\eta_0)^2} \; . \tag{4}$$

In Eqs. (3)–(4) a prime denotes a derivative with respect to conformal time η, the optical depth to Thomson scattering is $\mu(\eta)$, and we have defined the interval $\Delta\eta_0 = \eta_o - \eta$. It is sometimes convenient to define the ubiquitous source term $P(\mathbf{k}, \eta) = \frac{1}{2}\Big[\Theta_2(\mathbf{k}, \eta) - \sqrt{6}\alpha_2(\mathbf{k}, \eta)\Big]$. The system above is closed once the perturbed Einstein equations are used to determine how the linear scalar cosmological perturbations θ_{SW}, V_b, Φ and Ψ evolve with time. How-

ever, both the precise nature of the perturbed Einstein equations, or of the initial conditions that are used to evolve those equations, are irrelevant for our results.

Equations (3)–(4) show that the primary sources of temperature fluctuations are the Sachs-Wolfe term, θ_{SW}, the baryon velocity, V_b, and the gravitational potentials Φ and Ψ (also known as the Bardeen potentials—we work in the conformal-Newtonian gauge). The primary source of the polarization of the CMB, on the other hand, is the quadrupole of the temperature fluctuations. The integral equations then couple all the momenta of temperature and polarization, mediated by the visibility function $g(\eta) = \mu'(\eta)\mathrm{e}^{-\mu(\eta)}$.

Henceforth we will take $\mathbf{x}_o = 0$, i. e., the observer is taken to be at the origin of the coordinate system employed for the description of the problem.

3 Uncoupling the Temperature Evolution

We can decouple Eqs. (3)–(4) by neglecting the term α_2 in Eq. (3). This truncation represents the approximation whereby deviations from the equilibrium temperature are described by the Sachs-Wolfe (SW) and integrated Sachs-Wolfe (ISW) effects. Within this approximation, then Eq. (3) becomes:

$$\Theta_l(\mathbf{k},\eta_o) = \int_0^{\eta_o} d\eta \left\{ g(\eta)\left[\theta_{SW}(\mathbf{k},\eta)\, j_l(k\Delta\eta_0) - k\, V_b(\mathbf{k},\eta)\, j_l'(k\Delta\eta_0)\right] + \mathrm{e}^{-\mu(\eta)}(\Psi' + \Phi')(\mathbf{k},\eta)\, j_l(k\Delta\eta_0)\right\} . \tag{5}$$

Neglecting polarization as a source term for the temperature anisotropies is in fact a very good approximation, and the reason for this underlies the argument presented in this paper. The visibility function $g(\eta)$ should be regarded as the probability per unit (conformal) time that photons will scatter with some free electron—and, in fact, $g(\eta)$ is defined in such a way that this probability is normalized, $\int_0^\infty d\eta\, g(\eta) = \int_0^\infty d\mu\, e^{-\mu} = 1$. This means that each time a factor of the visibility function intermediates a source term, that source term is damped by a factor ϵ, with $0 < \epsilon < 1$. Since the lowest-order polarization term has at least one factor of the visibility function, it contributes as a source term to the temperature with two factors of the visibility function. Hence, the SW and ISW effects dominate the intensity of the signal, and Eq. (5) accounts for the largest contribution to the temperature anisotropies.

3.1 CMB Temperature in Position Space

Let's now take the lowest-order contribution to the temperature anisotropies, and express it in terms of position space. Expressing Eq. (1) as:

$$\Theta(\eta_o, \hat{\mathbf{o}}) = \sum_{lm} \theta_{lm}(\eta_o) \mathrm{Y}_{lm}(\hat{\mathbf{o}}) \,, \tag{6}$$

the coefficients $\theta_{lm}(\eta_o)$ are given by:

$$\theta_{lm}(\eta_o) = 2\, i^l \int_0^{\eta_o} d\eta \int \frac{dk}{(2\pi)^{1/2}}\, k^2 \int d^2\hat{\mathbf{k}}\, \mathrm{e}^{-\mu(\eta)} \times \left\{ \mu'(\eta) \left[\theta_{SW}(\mathbf{k}, \eta) - V_b(\mathbf{k}, \eta) \frac{\partial}{\partial \eta} \right] + (\Psi' + \Phi')(\mathbf{k}, \eta) \right\} \mathrm{Y}^*_{lm}(\hat{\mathbf{k}})\, j_l(k\Delta\eta_o) \tag{7}$$

We shall then define the primary source term operator as:

$$S_{lm}(k, \eta) = \int d^2\hat{\mathbf{k}}\, \mathrm{e}^{-\mu(\eta)} \left\{ \mu'(\eta) \left[\theta_{SW}(\mathbf{k}, \eta) - V_b(\mathbf{k}, \eta) \frac{\partial}{\partial \eta} \right] + (\Psi' + \Phi')(\mathbf{k}, \eta) \right\} \mathrm{Y}^*_{lm}(\hat{\mathbf{k}}) \,, \tag{8}$$

where we stress the fact that we have included the optical depth to Thomson scattering in the definition of the source term. In terms of Eq. (8), Eq. (7) can be written as:

$$\theta_{lm} = 2\, i^l \int_0^{\eta_o} d\eta \int \frac{dk}{(2\pi)^{1/2}}\, k^2\, S_{lm}(k, \eta)\, j_l(k\Delta\eta_o) \,. \tag{9}$$

In order to obtain a position-space description, we will express the coefficients $S_{lm}(k, \eta)$ in terms of their counterparts in position space, by means of a Hankel transform:

$$S_{lm}(k, \eta) = 2\,(-i)^l \int \frac{dX}{(2\pi)^{1/2}}\, X^2\, S_{lm}(X, \eta)\, j_l(kX) \,. \tag{10}$$

Using now the orthogonality relation for spherical Bessel functions,

$$\int dk\, k^2\, j_L(ak)\, j_L(bk) = \frac{\pi}{2} \frac{b^L}{a^{L+2}} \delta(a - b) \,,$$

we obtain:

$$\theta_{lm} = \int_0^{\eta_o} d\eta\, S_{lm}(X = \Delta\eta_0, \eta) \,. \tag{11}$$

Equation (11) has a straightforward interpretation: it states that, in order for a source term at time η contribute to the CMB signal at time η_o, that source must be located at the spherical shells of radius $\Delta\eta_0 = \eta_o - \eta$ centered at the observer. The set of those spherical shells is a hypersurface which corresponds, of course, to the past light cone of the observer on $\{\mathbf{x}_o, \eta_o\}$. Since the visibility function is highly peaked at the time of decoupling, the primary source term contributes the most to the signal near the epochs when $z(\eta) \simeq 1100$.

3.2 CMB Polarization in Position Space

We shall now decompose the polarization as:

$$\frac{Q + iU}{4I}(\eta_o, \hat{\mathbf{o}}) = \sum_{lm} \pi_{lm}(\eta_o)\, {}_2\mathrm{Y}_{lm}(\hat{\mathbf{o}}) \,, \tag{12}$$

with the aim of determining the coefficients $\pi_{lm}(\eta_o)$. The source terms in Eq. (4) are Θ_2 and α_2, which are built iteratively from an initial temperature quadrupole. We can, therefore, organize the iterative solution as a series into powers of the visibility function. As a first step in the iterative solution, for example, α_2 (which is of higher order in the visibility function) will not be taken into account—only the temperature quadrupole will contribute to generate polarization at this order. The first iteration is, therefore:

$$\pi^{(1)}_{lm}(\eta_o) = -\frac{3}{4}\, 2\, i^l \sqrt{\frac{(l+2)!}{(l-2)!}} \int_0^{\eta_o} d\eta_1\, g(\eta_1) \int_0^{\eta_1} d\eta \int \frac{dk}{(2\pi)^{1/2}}\, k^2\, S_{lm}(k,\eta) \times j_2(k\Delta\eta_1)\, \frac{j_l(k\Delta\eta_0)}{(k\Delta\eta_0)^2} \,, \tag{13}$$

where the time intervals are defined as $\Delta\eta_1 = \eta_1 - \eta$ and $\Delta\eta_0 = \eta_o - \eta_1$. The source term S_{lm} is the same as was defined in Eq. (8). Using once again the Hankel transform of Eq. (10) we can recast Eq. (13) as:

$$\pi^{(1)}_{lm}(\eta_o) = -\frac{3}{2\pi} \sqrt{\frac{(l+2)!}{(l-2)!}} \int_0^{\eta_o} d\eta_1\, g(\eta_1) \int_0^{\eta_1} d\eta \int_0^{\infty} dX\, X^2\, S_{lm}(X,\eta) \times \int_0^{\infty} dk\, k^2\, j_l(kX)\, \frac{j_l(k\Delta\eta_o)}{(k\Delta\eta_o)^2}\, j_2(k\Delta\eta_1) \,. \tag{14}$$

The interpretation of the expression above is the following: at time η a source term generates a temperature quadrupole. That quadrupole then generates, through a scattering at time η_1, the polarization which is finally observed at time η_o. As we shall see, the integral of the second line of Eq. (14) guarantees that the source term,

at a distance X from the origin, is located in the past lightcone of the observer, for all possible η_1 and η. The variable X will also be upper-bounded, and therefore the upper limit in the integration over the source terms will be replaced by a finite value that, as we shall see, corresponds the radius of the observer's past lightcone.

The next step in the iterative solution is to take the $\pi^{(1)}_{2m}$ just computed and use it as a source term for the polarization itself—this means that now the polarization piece of the source term in Eq. (4), α_2, is no longer assumed to vanish. This contribution, which we will call $\pi^{(2)}_{lm}$, is therefore given by:

$$\pi^{(2)}_{lm}(\eta_o) = -\frac{3}{2}\sqrt{\frac{(l+2)!}{(l-2)!}}\int_0^{\eta_o} d\eta_1\, g(\eta_1)\int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}}\, 4\pi\, i^l \left[-\frac{\sqrt{6}}{2}\,\alpha_2(\mathbf{k},\eta_1)\right] Y^*_{lm}(\hat{\mathbf{k}})\,\frac{j_l(k\Delta\eta_o)}{(k\Delta\eta_o)^2}$$

$$= -\frac{3}{2}\sqrt{\frac{(l+2)!}{(l-2)!}}\int_0^{\eta_o} d\eta_1\, g(\eta_1)\, 4\pi\, i^l\, \frac{1}{2\pi}\left[-\frac{\sqrt{6}}{2}\left(-\frac{3}{2}\right)\sqrt{24}\int_0^{\eta_1} d\eta_2\, g(\eta_2)\right.$$

$$\left.\times\frac{1}{2}\int \frac{dk}{(2\pi)^{3/2}}\, k^2 \int d^2\hat{\mathbf{k}}\,\Theta_2(\mathbf{k},\eta_2)\, Y^*_{lm}(\hat{\mathbf{k}})\,\frac{j_2(k\Delta\eta_1)}{(k\Delta\eta_1)^2}\right]\frac{j_l(k\Delta\eta_o)}{(k\Delta\eta_o)^2}\,, \tag{15}$$

where now the interval $\Delta\eta_1 = \eta_1 - \eta_2$. In terms of the primary sources in position space, after using Eq. (11) we obtain:

$$\pi^{(2)}_{lm}(\eta_o) = -\frac{3}{2\pi}\sqrt{\frac{(l+2)!}{(l-2)!}}\int_0^{\eta_o} d\eta_1\, g(\eta_1)\int_0^{\eta_2} d\eta_2\, g(\eta_2)\int_0^{\eta_2} d\eta$$

$$\times\int_0^{\infty} dX X^2 S_{lm}(X,\eta)\left[9\int dk k^2 j_l(kX)\frac{j_l(k\Delta\eta_o)}{(k\Delta\eta_o)^2}\frac{j_2(k\Delta\eta_1)}{(k\Delta\eta_1)^2}j_2(k\Delta\eta_2)\right]. \tag{16}$$

The term $\pi^{(2)}_{lm}(\eta_o)$ is weighted twice by the visibility function and corresponds, as we will show in details, to the contribution to the total polarization coming from photons that have Thomson scattered twice during the recombination.

At this point it is important to clarify our notation. We will always count the photon scatterings backwards in time: the time the photons are observed is always taken to be η_o; the last time that the photons scattered before being observed is η_1; and so on. By convention, we will always evaluate the primary source term S_{lm} at the instant η, so the sequence of scatterings ends with η. Hence if, as in the case described by Eq. (16), there are two scatterings between the generation of the signal at η and its observation at η_o, then we have $\eta_o \geq \eta_1 \geq \eta_2 \geq \eta$, and the time intervals always express the differences between one time and the previous one, so in that case $\Delta\eta_0 = \eta_o - \eta_1$, $\Delta\eta_1 = \eta_1 - \eta_2$, and $\Delta\eta_2 = \eta_2 - \eta$.

The general term in the iterative series expansion with n intermediate scatterings can be written as Reimberg and Abramo (2013):

$$\pi_{lm}^{(n)}(\eta_o) = -\frac{3}{2\pi}\sqrt{\frac{(l+2)!}{(l-2)!}}\int_0^{\eta_o} d\eta_1\, g(\eta_1) \times \frac{1}{(n-1)!}\int_0^{\eta_1} d\eta_2 \ldots d\eta_n\, \mathrm{T}\{g(\eta_2)\ldots g(\eta_n)\}$$
$$\times \int_0^{\eta_n} d\eta \int_0^{\infty} dX\, X^2\, S_{lm}(X,\eta)$$
$$\times 9^{(n-1)} \int dk\, k^2\, j_l(kX)\, \frac{j_l(k\Delta\eta_o)}{(k\Delta\eta_o)^2} \underbrace{\frac{j_2(k\Delta\eta_1)}{(k\Delta\eta_1)^2} \cdots \frac{j_2(k\Delta\eta_{n-1})}{(k\Delta\eta_{n-1})^2}}_{(n-1)\,times} j_2(k\Delta\eta_n) \tag{17}$$

Here, T stands for the time-ordered product of the sub-intervals, whose purpose is to reproduce the chain of integrations mediated by visibility functions.

The coefficients π_{lm} appearing in Eq. (12) can be expressed, therefore, as:

$$\pi_{lm}(\eta_o) = \sum_{n=0}^{\infty} \pi_{lm}^{(n)}(\eta_o)\,, \tag{18}$$

with $\pi_{lm}^{(n)}(\eta_o)$ given by Eq. (17).

An important condition for the validity of this perturbative expansion of the CMB temperature and polarization is that all the terms in the expansion of Eq. (12), with an arbitrary number n of intermediate scatterings, must be expressed in position space. However, this can only be true if the k integrals over products of spherical Bessel functions appearing in Eq. (17) can be in fact performed, and are well-behaved. In the next Section we will show that, in fact, these integrals are probability densities for random flights with n steps, in a space of suitable dimensionality.

4 Random Flights and the CMB

From the previous Section—specially from Eq. (17)—it is evident that a complete treatment of the CMB in position space requires that some specific integrals of products of spherical Bessel functions should be computed. In order to fulfill this requirement, we will proceed in the following way: first, we will present a simplified version of those integrals, and we will show that they give rise to probability densities associated with random flights. Next, we will show how the integrals we have to solve can be expressed in terms of the random flight integrals.

Let's recall the well-known identities satisfied by spherical Bessel functions:

$$z^{L+\frac{3}{2}} J_{L+\frac{1}{2}}(z) = \frac{d}{dz}\left[z^{L+\frac{3}{2}} J_{L+\frac{3}{2}}(z)\right]\,.$$

Since the spherical Bessel functions are defined as:

$$j_n(z) = \sqrt{\frac{\pi}{2\,z}}\,J_{n+\frac{1}{2}}(z)\ ,$$

we are able to write:

$$\int_0^\infty dk\,k^2\,j_L(k\,r)\prod_{i=1}^{n-1}\frac{j_L(k\,r_i)}{(k\,r_i)^L}\,j_L(k\,r_n) = \frac{\left(\frac{\pi}{2}\right)^{(n+1)/2}}{\left[\Gamma\left(L+\frac{3}{2}\right)\right]^{n-1}}\,\frac{r_n^L}{r^{L+2}}$$
$$\times\frac{d}{dr}\left\{\left[\Gamma\left(L+\frac{3}{2}\right)\right]^{n-2}\int_0^\infty dk\,r\left(\frac{k\,r}{2}\right)^{L+\frac{1}{2}}J_{L+\frac{3}{2}}(k\,r)\prod_{i=1}^{n}\frac{J_{L+\frac{1}{2}}(k\,r_i)}{(k\,r_i)^{L+\frac{1}{2}}}\right\}\ . \tag{19}$$

The derivative of the second line in Eq. (19) is, in fact, the probability density associated with a random flight—see, e.g., Watson (1944). This is the probability density that a particle which moves with a constant (and finite) speed, and which starts from a given position in space, will be at a distance r from the point of origin, after changing randomly directions n times during its trajectory. The length of the intermediate steps are denoted by r_i, $i = 1, \ldots, n$. The order L of the spherical Bessel functions in these integrals is related to the dimensionality of the space where the flight takes place: namely, the dimension D of that space is given by $D = 2L+3$.

Following the notation employed by Watson (1944) we shall denote:

$$p_n(r; r_1, \ldots, r_n \,|\, 2L+3) := \frac{d}{dr}\left\{\left[\Gamma\left(L+\frac{3}{2}\right)\right]^{n-1}\right.$$
$$\left.\times\int_0^\infty dk\,r\left(\frac{kr}{2}\right)^{L+\frac{1}{2}}J_{L+\frac{3}{2}}(kr)\prod_{i=1}^{n}\frac{J_{L+\frac{1}{2}}(kr_i)}{(kr_i)^{L+\frac{1}{2}}}\right\}\ . \tag{20}$$

However, the integral (19) is still not what we need in order to solve the momentum integrals that appear in our iterations—see, e.g., the Fourier integration of Eq. (17).

We will now extend the random flight integrals to include the scenario that appears in the context of the CMB. If $l \geq L$ (which is always the case in our iterative solutions), then the product of two spherical Bessel functions of order l can be written in terms of a single spherical Bessel function of order L. This is a consequence of Gegenbauer's relation (Watson 1944; Talman 1968) and of the orthogonality of associated Legendre polynomials—see Abramo et al. (2010) for a derivation:

$$j_l(kX)\,j_l(kR) = \frac{(-1)^L}{2}\int_{|X-R|}^{X+R}dr\,k^L\left(\frac{XR}{r}\right)^{L-1}P_l^{-L}(\cos\alpha)\ (\sin\alpha)^L\,j_L(kr) \tag{21}$$

where r, R and X must form a triangle, with the angle α being given implicitly in terms of the relation $r^2 = R^2 + X^2 - 2RX\cos\alpha$, and $P_l^{-L}(\cos\alpha)$ is an associated Legendre polynomial. Applying Eqs. (21) and (19) with $L = 2$ we can recast the Fourier integral in Eq. (17) as:

$$\int_0^\infty dk\, k^2\, j_l(kX)\, \frac{j_l(k\Delta\eta_0)}{(k\Delta\eta_0)^2} \prod_{i=1}^{n-1} \frac{j_2(k\Delta\eta_i)}{(k\Delta\eta_i)^2}\, j_2(k\Delta\eta_n)$$

$$= \frac{1}{2} \frac{\left(\frac{\pi}{2}\right)^{(n+1)/2}}{\left[\Gamma\left(\frac{7}{2}\right)\right]^{n-1}} \int d(\cos\alpha) \left(\frac{X\Delta\eta_n}{r^3}\right)^2 P_l^{-2}(\cos\alpha)\, \sin^2\alpha$$

$$\times p_n(r; \Delta\eta_1, \ldots, \Delta\eta_n \,|\, 7)\ , \tag{22}$$

where we used the notation introduced in Eq. (20).

We should examine Eq. (22) more carefully. As anticipated, the presence of the term $p_n(r; \Delta\eta_1, \ldots, \Delta\eta_n|7)$ should not be surprising, due to the interpretation of a random flight process and its validity with respect to the physics of the recombination. The dimensionality ($D = 2 \times 2 + 3 = 7$) of the space associated with the random flight, however, is not yet fully understood. That dimension is determined by the order of the spherical Bessel functions which mediate the sources of anisotropies and the final CMB signal, but since only the quadrupole of the temperature fluctuation contributes to the polarization, the spherical Bessel function of order 2 is the one that characterizes the random flight for the CMB. A possible explanation for this dimension is that, after separating the angular dependence of the CMB from its radial and time dependence through the spherical harmonic decomposition, the light cone has only two dimensions left. Since the multipole L has $2L + 1$ degrees of freedom, we end up with $2L + 3$ dimensions where our relevant variables can perform random flights. However, a more refined argument to explain the dimension 7 is not yet known.

Looking back now at Eq. (17), we recognize that for the $\pi_{lm}^{(n)}$ term in the polarization expansion, the probability density $p_n(r; \Delta\eta_1, \ldots, \Delta\eta_n \,|\, 7)$ appears clearly. The interpretation of an expansion in the number of interactions during recombination is therefore strengthened.

As discussed above, the intervals $\Delta\eta_1, \ldots, \Delta\eta_n$ express the times elapsed between consecutive scatterings. All these subintervals are elements of a partition of the time interval $\eta_1 - \eta$, which is therefore the total time elapsed since the instant the photon leaves equilibrium with matter, at time η, until the instant η_1 when the photon has last scattered prior to its observation. This time interval represents, therefore, the effective duration of recombination for a given photon. The lengths of each subinterval are weighted by the visibility functions, and integrated in order to contemplate all possible histories for photons during recombination. The interval $\eta_o - \eta_1$ expresses the time elapsed since the photon scattered for the last time, before it is observed at the time η_o (we are assuming that no further scatterings take place during this interval). We can represent a photon's history by means of Fig. 1: on the

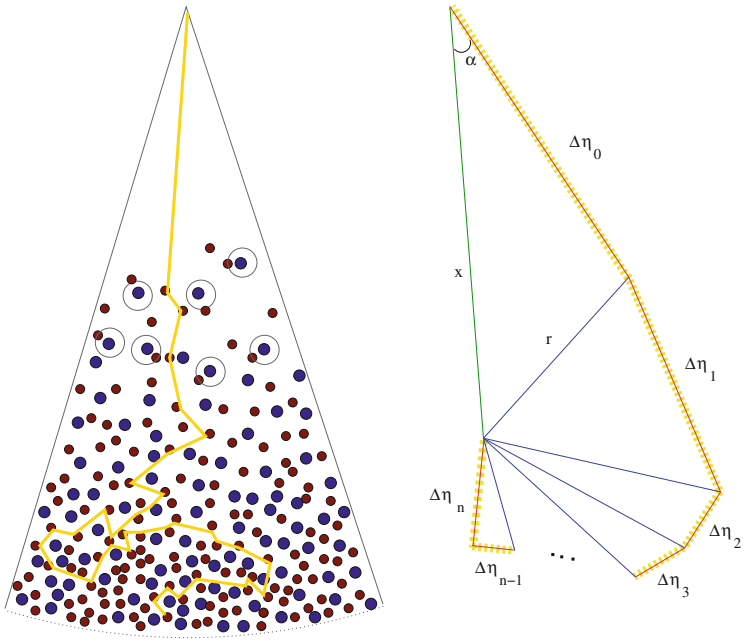

Fig. 1 Connection between scatterings of a photon during recombination (*left*) with a diagrammatic representation of the random flight (*right*). The steps $\Delta\eta_n, \ldots, \Delta\eta_1$ correspond to the lengths of the photon's trajectory between successive scatterings. $\Delta\eta_0$ corresponds to the propagation since the photon's last scattering during recombination, at time η_1, and the observation at time η_o. In both diagrams the observation takes place at the upper vertex

left we show the photon's interactions prior to observation at the vertex of the cone, while on the right we show a diagrammatic representation of that history, with the relevant elements that appear in Eq. (22).

We call special attention to the diagram represented on the right of Fig. 1. This diagram represents the extended random flight performed by a photon during recombination. The steps $\Delta\eta_n, \ldots, \Delta\eta_1$ (going forward in time) belong to a standard random flight, and describe the trajectory of a photon that has left equilibrium with matter at an instant η, then propagated freely for a distance (or a time interval) $\Delta\eta_n$, then Thomson-scattered with an electron at time η_n, then propagated freely for a distance $\Delta\eta_{n-1}$, and so on until the instant η_1, when it scattered for the last time. The standard random flight ends at the instant η_1. The photon, at that moment, is a radius r away from the point where the flight started. The steps indicated by $\Delta\eta_0$ and X do not belong to the standard random flight, but are present in the extended random flight, and are introduced through the spherical Bessel functions of different order in the k integral of Eq. (22). This is necessary because those two steps are not associated with any movement between successive scatterings, but are in fact associated with the distance from the observer to the origin of the photon, and to the end-point of the random flight. It should be strengthened that $\Delta\eta_0$, X and r are

related by $r^2 = \Delta\eta_0^2 + X^2 - 2\,\Delta\eta_0\, X\,\cos\alpha$. Since $0 \leq r \leq \Delta\eta_1 + \Delta\eta_2 + \ldots + \Delta\eta_n$, it follows that $0 \leq X \leq \Delta\eta_0 + \Delta\eta_1 + \Delta\eta_2 + \ldots + \Delta\eta_n$, which then determines the domain of dependence of the problem.

5 The Polarization in Position Space: The 'Path-Integral'ness

We shall now go back to Eq. (17). In terms of the extended random flight just introduced, the polarization coefficient to n^{th} order, $\pi_{lm}^{(n)}(\eta_o)$, can be written as:

$$\pi_{lm}^{(n)} = -\frac{3}{4\pi}\sqrt{\frac{(l+2)!}{(l-2)!}}\int_0^{\eta_o} d\eta_1\, g(\eta_1)\,\frac{1}{(n-1)!}\int_0^{\eta_1} d\eta_2 \ldots d\eta_n\, \mathrm{T}\{g(\eta_2)\,\ldots\,g(\eta_n)\}$$
$$\times\frac{9^{n-1}\left(\frac{\pi}{2}\right)^{(n+1)/2}}{\left[\Gamma\left(\frac{7}{2}\right)\right]^{n-1}}\int_0^{\eta_n} d\eta \int_0^{\eta_o-\eta} dX\, X^2\, S_{lm}(X,\eta)$$
$$\times\int d(\cos\alpha)\left(\frac{X\Delta\eta_n}{r^3}\right)^2 P_l^{-2}(\cos\alpha)\,\sin^2\alpha\; p_n(r;\Delta\eta_1,\ldots,\Delta\eta_n\,|\,7)\;, \qquad (23)$$

where we have already used the aforementioned upper bound for the variable X.

The Eq. (23) can be understood as the combination of three procedures:

- The integration over α corresponds to a marginalization over all possible paths composed of n steps of lengths $\Delta\eta_1 + \ldots + \Delta\eta_n$ that have a net displacement r determined by X and $\Delta\eta_0$, as shown in Fig. 2. This "average over paths" is a function of X, $\Delta\eta_0$, $\Delta\eta_1, \ldots, \Delta\eta_n$.
- The contribution from the primary source term, $S_{lm}(X,\eta)$, is then mediated by this "average over paths" that was just described, for all possible values of X. The maximum value that X may reach is $\Delta\eta_0 + \Delta\eta_1 + \ldots + \Delta\eta_n = \eta_o - \eta$, which is nothing but the radius of the observer's past light cone up to the time η. After computing the contribution of the source terms, we end up with an expression that is a function of $\Delta\eta_0, \Delta\eta_1, \ldots, \Delta\eta_n$.
- The last step is to let the intervals $\Delta\eta_0, \Delta\eta_1, \ldots, \Delta\eta_n$ assume any values through the integrations, each one weighted by its corresponding factor of the visibility function to take into account the probability that the photon will scatter at that instant of time. This accomplishes the goal of accounting for the contribution from sources at all distances, and over any possible number of intermediate steps of the extended random flights.

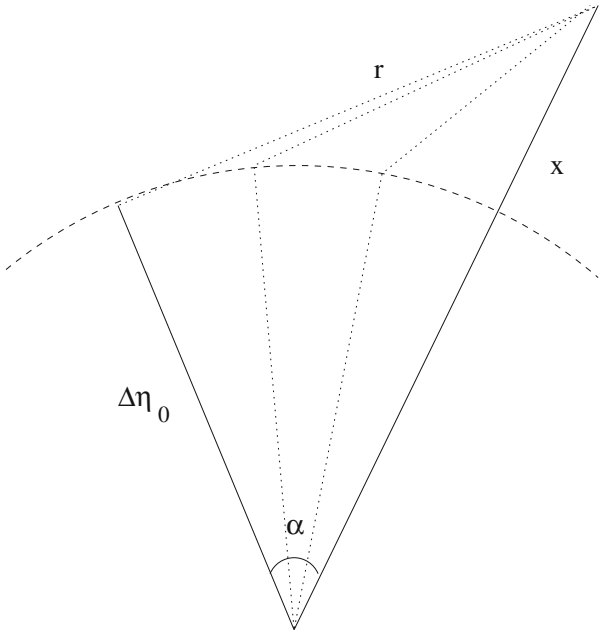

Fig. 2 Marginalization over all paths with n steps, composed of the intermediate displacements $\Delta\eta_1, \ldots, \Delta\eta_n$, which lead to a fixed displacement r with respect to the origin of the flight. The distance r is determined by X and $\Delta\eta_0$ for all possible angles α

Adding the contributions from all $\pi_{lm}^{(n)}$ we obtain:

$$\pi_{lm} = -\frac{3}{4\pi}\sqrt{\frac{(l+2)!}{(l-2)!}}\int_0^{\eta_o} d\eta_1\, g(\eta_1) \sum_{n=1}^{\infty} \frac{1}{(n-1)!}\int_0^{\eta_1} d\eta_2 \ldots d\eta_n\, \mathrm{T}\{g(\eta_2)\ \ldots\ g(\eta_n)\}$$
$$\times \frac{9^{n-1}\left(\frac{\pi}{2}\right)^{(n+1)/2}}{\left[\Gamma\left(\frac{7}{2}\right)\right]^{n-1}} \int_0^{\eta_n} d\eta \int_0^{\infty} dX\, X^2\, S_{lm}(X,\eta)$$
$$\times \int d(\cos\alpha)\left(\frac{X\Delta\eta_n}{r^3}\right)^2 P_l^{-2}(\cos\alpha)\, \sin^2\alpha\; p_n(r;\Delta\eta_1,\ldots,\Delta\eta_n \mid 7)\ . \qquad (24)$$

6 Discussion

We have obtained in Eq. (24) the general expression for the expansion coefficients for the CMB polarization, that should be inserted in Eq. (12). Together with Eqs. (6) and (11), they complete the description in position space of the simplified version

of the Boltzmann's equations for the CMB temperature and polarization we took in consideration. The treatment of the complete version of the Boltzmann's equations can also be given in the same language, as shown in Reimberg and Abramo (2013).

The simplicity brought by uncoupling the temperature from polarization illustrates the geometrical nature of the problem, and justifies the choice of the term 'path-integral' to the approach shown here. The CMB temperature fluctuations are due to the nature of regions from where photons have emerged (over/underdensities, baryons velocity, gravitational potentials), and the time variation of the gravitational potentials along the photons paths (the integrated Sachs-Wolfe effect). The temperature fluctuations, therefore, carry information about the instant when photons decouple, and their travel toward us. The polarization signal, however, accumulates information about the history of scatterings suffered by the photons prior to their decoupling.

Frequent interactions of photons and electrons would bring the system to thermal equilibrium, where no net polarization is present, and the temperature distribution follows the equilibrium distribution determined by the nature of the interaction, and the structure of the space-time where the equilibrium is set. Less frequent scatterings produced as the recombination proceeds, move the system toward a slightly non-equilibrium configuration, where most of the energy distribution is that determined by equilibrium condition [what can be seen when the complete Boltzmann hierarchy is taken into consideration (Reimberg and Abramo 2013)], but the polarization signal is a clear signature that some scatterings happen in this non-equilibrium phase prior to complete decoupling.

We learn from Eq. (24) that all possible histories of photons scatterings during recombination contribute to the final CMB polarization. We add all possible number of scatterings, happening in all possible orders, for all possible time interval between them allowed by the visibility function, and weight each of these histories by the probability density for the random flight with corresponding number of steps, and intermediate displacements.

Because each scattering carries an additional power of the visibility function, the signal coming from a large number of scatterings term is suppressed. Also, properties of random flight processes assure that probability density function for random flights with large number of steps are highly picked around the origin, making their contribution to vanish in the expansion given in Eq. (24), what leads to an illustration of Boltzmann's H-theorem, as discussed in Reimberg and Abramo (2013). This is parallel to the quantum version of path integrals where, beyond the suppression introduced by small parameter in terms of which perturbative expansions are performed, diminishing weights are associated to paths far form the classical solution.

Finally, we can foresee some possible applications of this work. The series expansion in terms of the number of scatterings can be used for numerical simulations of constrained maps of temperature and polarization. Due to the general vanishing property of the probability density functions for the extended random flight if intermediate displacements do not form a polygon, and the decreasing of the visibility function for $z >> 10^3$, we can in practice take all the sources to vanish

outside of a sphere of radius R sufficiently large, and calculate the temperature and polarization corrections using Fourier-Bessel expansions, as shown in Abramo et al. (2010). In Fourier-Bessel basis only a discretized tower of modes contribute to each observable at each multipole, and the computational advantages of this approach are described in Leistedt et al. (2012). In what concerns the convergence of the iterative process, depending on the desired accuracy, application or the angular scale that one wishes to examine, it may be sufficient to consider only the first couple of scatterings of the photons, since going further in the expansion would bring only contributions from terms highly suppressed by powers of the visibility function.

Acknowledgements The authors thank the organizers of the II JPB School of Cosmology, and FAPESP for financial support.

References

Abramo, L.R., Reimberg, P.H., Xavier, H.S.: CMB in a box: causal structure and the Fourier-Bessel expansion. Phys. Rev D **82**, 043510 (2010)

Bennett, C.L., Larson, D., Weiland, J.L., Jarosik, N., Hinshaw, G., Odegard, N., Smith, K.M., Hill, R.S., Gold, B., Halpern, M., Komatsu, E., Nolta, M.R., Page, L., Spergel, D.N., Wollack, E., Dunkley, J., Kogut, A., Limon, M., Meyer, S.S., Tucker, G.S., Wright, E.L.: Nine-year Wilkinson microwave anisotropy probe (WMAP) observations: final maps and results. ArXiv e-prints (2012). arXiv:1212.5225

Chandrasekhar, S.: Stochastic problems in physics and astronomy. Rev. Mod. Phys. **15**(1), 1–89 (1943)

Dutka, J.: On the problem of random flights. Arch. Hist. Exact Sci. **32**(3), 351–375 (1985)

Leistedt, B., Rassat, A., Réfrégier, A., Starck, J.-L.: 3DEX: a code for fast spherical Fourier-Bessel decomposition of 3D surveys. Astron. Astrophys. **540**, A60 (2012)

Planck Collaboration, Ade, P.A.R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A.J., et al.: Planck 2013 results. I. Overview of products and scientific results. Astron. Astrophys. **571**, A1 (2014). arXiv:1303.5062

Reimberg, P.H., Abramo, L.R.: J. Cosmol. Astropart. Phys. **06**, 043 (2013)

Seljak, U., Zaldarriaga, M.: A line of sight approach to cosmic microwave background anisotropies. Astrophys. J. **469**, 437–444 (1996). arXiv:astro-ph/9603033

Smoot, G.F., et al.: Structure in the COBE differential microwave radiometer first year maps. Astrophys. J. **396**, L1–L5 (1992)

Straumann, N.: From primordial quantum fluctuations to the anisotropies of the cosmic microwave background radiation. Ann. Phys. **15**, 701–845 (2006)

Talman, J.D.: Special Functions: A Group Theoretic Approach. W. A. Benjamin, New York (1968)

Watson, G.N.: A Treatise on the Theory of Bessel Functions, 2nd edn. Cambridge University Press, Cambridge (1944)

Geometric Scalar Theory of Gravity

Júnior Diniz Toniato

Abstract The present article introduces a new scalar theory of gravity based on the Einstein's assumption that gravitation is an expression of the geometrical structure of the spacetime. In the geometric scalar theory of gravity all kind of matter and energy interacts with the gravitational (scalar) field only through a metric structure that naturally arises with the non linear dynamics of the scalar field. This allows us to overcome the problems from the previous scalar theories and construct a new scalar theory for gravitation which is in accordance at least with the observational data coming from our solar system.

1 Introduction

Since its formulation until the present days, the Einstein's theory of general relativity (GR) remains consistent with all experimental tests performed, the so called classical tests of gravitation (Turyshev 2009). Notwithstanding, over all these years, there have always been open questions that led physicists to seek alternative paths in the description of gravitational phenomena. Alternative theories of gravitation exist in large numbers and in the most diverse formulations, whereas those following Einstein's ideas, choosing describe gravitation as a geometric phenomenon, are those that obtained greatest success. Inside this extensive group, scalar-tensor theories and $f(R)$ theories are the ones that most currently stand (Clifton 2006).

In the class of the purely scalar metric theories, i.e. where the gravitational field is represented by one or more scalar functions that generate a gravitational metric, much was done up to mid-seventies, but all formulations failed to comply with all classical tests. In 1972, Wei-Tou Ni wrote a compendium of metric theories containing a broad review and analysis of scalar theories (Ni 1972).

J.D. Toniato (✉)
Instituto de Cosmologia Relatividade Astrofísica - ICRA, Centro Brasileiro de Pesquisas Físicas - CBPF, Rio de Janeiro, Brazil
e-mail: toniato@cbpf.br

J.C. Fabris et al. (eds.), *The Cosmic Microwave Background*, Astrophysics and Space Science Proceedings 45, DOI 10.1007/978-3-319-44769-8_13

Table 1 Different proposals for scalar theories of gravitation according to Eqs. (1) and (2)

Scalar theories of gravitation	
Author (year)	Basic functions
Nordström (1912)	$f = \Phi$
	$k = 1$
Nordström (1913–1914)	$f = -\ln \Phi$
	$k = \Phi$
Littlewood (1953)	$f = -2 \ln (1 - \Phi)$
Bergmann (1956)	$k = 1$

Following Ni, these various proposals have the common property of being conformally flat. Its gravitational metrics have the general form,

$$g_{\mu\nu} = e^{-2f(\Phi)}\, \eta_{\mu\nu}\,, \tag{1}$$

where Φ is the gravitational potential and $\eta_{\mu\nu}$ is the Minkowski metric. The field equations of these theories can be summarized in the expression,

$$\Box\Phi \propto k(\Phi)\, T\,, \tag{2}$$

with the $\Box$ being the d'Alembertian operator constructed with the metric (22) and T the trace of the energy-momentum tensor of the source of the gravitational field. The $f(\Phi)$ and $k(\Phi)$ functions have distinct forms according to the theory which one wants to describe. The table below shows the main scalar theories and its correspondent functions (Table 1).

The fact that all these theories are conformally flat is the main cause why one can not couple gravity and electromagnetism, since the Maxwell equations are conformally invariants. Also, with the source of the gravitational field being the trace of the energy-momentum tensor, which is zero for the electromagnetic field, shows that this fields can not produce gravitation. Thus none of the theories in the table above are in agreement with the measurement of the bending of light. Further, all these theories fail to provide the correct advance of the perihelion of Mercury. However, Ni's paper does not cite the theory proposed by Dowker in 1965, which although not predicting the bending of light, gives the right answer for the Mercury's perihelion precession (Dowker 1965).

Though, the role of the scalar field representing the gravitational potential is not fully determined, as I will show here. A recent study of effective metrics in non linear scalar theories shows that is possible to establish a metric structure, not conformally flat, which describe the dynamic of the field itself (Goulart et al. 2011). In the next section I show how this mathematical property emerges. The physical aspects of such property can only be determined if one introduces a way by which this metric will interact with the other fields of nature. In other words, in order to interpret the scalar field as the gravitational potential and the metric generated by

it as the physical metric, one needs to say how matter/energy interacts with it. This will constitute the grounds of the geometric scalar gravity (GSG).

2 Geometrization of a Nonlinear Scalar Theory

Consider a relativistic scalar field Φ with a nonlinear dynamic in the Minkowski spacetime. The action describing its dynamic is given by,

$$S = \int L(\Phi, w)\sqrt{-\eta}\, d^4x\,, \tag{3}$$

where η is the determinant of the Minkowski metric and,

$$w \equiv \eta^{\mu\nu}\,\partial_\mu\Phi\,\partial_\nu\Phi\,. \tag{4}$$

The notation ∂_μ indicates a simple derivative in relation with the coordinate x^μ . The minimal action principle returns the equation of motion of the scalar field,

$$\frac{1}{\sqrt{-\eta}}\,\partial_\mu\Big(\sqrt{-\eta}\,L_w\,\eta^{\mu\nu}\partial_\nu\Phi\Big) - \frac{1}{2}\,L_\Phi = 0\,, \tag{5}$$

where L_X indicates a derivative in relation with the variable X .

Introducing the metric tensor,

$$q^{\mu\nu} = \alpha\,\eta^{\mu\nu} + \frac{\beta}{w}\,\partial^\mu\Phi\,\partial^\nu\Phi\,, \tag{6}$$

with α and β being functions of Φ and w, and the correspondent covariant expression, defined by $q^{\mu\alpha}\,q_{\alpha\nu} = \delta^\mu_\nu$, given by

$$q_{\mu\nu} = \frac{1}{\alpha}\,\eta_{\mu\nu} - \frac{\beta}{\alpha\,(\alpha+\beta)\,w}\,\partial_\mu\Phi\,\partial_\nu\Phi\,, \tag{7}$$

Eq. (5) is rewritten as

$$\frac{L_w}{\alpha+\beta}\left[\Box_q\Phi + \frac{(\alpha+\beta)^{3/2}}{\alpha^{3/2}L_w}\,\partial_\mu\left(\frac{\alpha^{3/2}L_w}{\sqrt{\alpha+\beta}}\right)\gamma^{\mu\nu}\partial_\nu\Phi - \frac{L_\Phi}{2\,L_w}\,(\alpha+\beta)\right] = 0, \tag{8}$$

where the subscript in the d'Alembertian operator indicates that it is constructed with the metric $q_{\mu\nu}$.

Note that by a simple choice of the coefficients α and β is possible to describe the nonlinear dynamic of Φ as if it were embedded in a curved spacetime (generated by

the field itself) where it interacts minimally with $q_{\mu\nu}$. In order to do this we restrict the second order derivatives of Φ to appear only in the $\Box_q\Phi$ term of the above equation.

The simplest manner is the imposition

$$\frac{\alpha^{3/2}L_w}{\sqrt{\alpha+\beta}} = C\,, \tag{9}$$

where C is a constant. The resultant equation is,

$$\Box_q\Phi = j(\Phi\,,\partial\Phi)\,, \tag{10}$$

where we have defined

$$j(\Phi\,,\partial\Phi) \equiv \frac{\alpha^3}{2C}L_\Phi\,L_w\,. \tag{11}$$

Equations (5) and (10) are equivalents, allowing us to interpret the dynamic of Φ as (1) nonlinear in the Minkowski spacetime or (2) "linear" in the metric $q_{\mu\nu}$ with a source $j(\Phi,\partial\Phi)$. Important to emphasize that the use of the word "linear" is made here in a metaphoric sense, given that, since the metric $q_{\mu\nu}$ depends on Φ , the dynamic remains nonlinear.

A second possibility of geometrization consist in relax the condition (9) by substituting the constant C by a function of Φ only,

$$\frac{\alpha^{3/2}L_w}{\sqrt{\alpha+\beta}} = F(\Phi)\,. \tag{12}$$

Using this in the Eq. (8) we get,

$$\frac{L_w}{\alpha+\beta}\left[\Box_q\Phi + (\alpha+\beta)\left(\frac{F_\Phi}{F}\,w - \frac{L_\Phi}{2L_w}\right)\right] = 0\,, \tag{13}$$

and, by appropriately choosing the function F , we can write the dynamic of Φ as "free field" (again in a metaphoric way) without the source of the previous case. Thus, we have,

$$\Box_q\Phi = 0\,. \tag{14}$$

If the function $F(\Phi)$ satisfies the condition

$$\frac{F_\Phi}{F}\,w - \frac{L_\Phi}{2L_w} = 0\,. \tag{15}$$

Note that these two cases are equal when $L_\Phi = 0$.

In GR, matter/energy curves the spacetime where it propagates and, in this sense that we understand how the metric structure $q_{\mu\nu}$ can be associated with a gravitational process. The scalar field itself curves the spacetime around it. But if we want to assign to Φ the role of a gravitational potential, with $q_{\mu\nu}$ being the gravitational metric, we need to determine how it will interact with other fields in the nature. The next section is occupied of this task.

We will use the second geometrization method present in this section to describe the dynamic of Φ in the q-spacetime. The hypothesis postulated and the observational data should help us to determine the Lagrangian of the scalar field and the functional dependence of the metric coefficients α and β .

3 The Fundamentals of the GSG

In order to propose the main properties of GSG we will follow the basic ideas of Einstein's theory. Field formulation of GR describe the gravitational metric as sum of a flat metric (Minkowski) plus a perturbation $h^{\mu\nu}$ (not necessarily small),

$$g^{\mu\nu} = \eta^{\mu\nu} + h^{\mu\nu} . \tag{16}$$

Although the above expression be exact, its covariant version is indeed an infinity series (Feynman et al. 1995),

$$g_{\mu\nu} = \eta_{\mu\nu} - h_{\mu\nu} + h_{\mu\alpha}h^{\alpha}{}_{\nu} - \ldots \tag{17}$$

According to this formulation we can cite the basic properties of GR as follows.

- Gravitational interaction is described by a second order tensor field $h_{\mu\nu}$ that satisfies a non linear dynamic equation (Einstein's equation);
- The theory reproduces Newton's gravity in a weak field approximation;
- Any kind of matter and energy interacts with the gravitational field only through the metric $g_{\mu\nu}$;
- Test particles and electromagnetic waves follows geodesics in the curved spacetime described by $g_{\mu\nu}$;
- The $g_{\mu\nu}$ metric interacts universally with all fields in the nature following the minimum couple principle.

Now, we postulate the basic properties of the GSG.

- Gravitational interaction is described by scalar field Φ that satisfies a non linear dynamic equation;
- The theory reproduces Newton's gravity in a weak field approximation;
- Any kind of matter and energy interacts with the gravitational field only through the metric $q^{\mu\nu}$ [cf. (6)];

- Test particles and electromagnetic waves follows geodesics in the curved spacetime described by $q_{\mu\nu}$;
- The $q_{\mu\nu}$ metric interacts universally with all fields in the nature following the minimum couple principle.

Note that, different from GR, the covariant version of the gravitational metric in GSG is not an infinite series, as shown in Eq. (7).

Immediately, as it is in GR, the coupling between gravitation and electromagnetism in GSG is granted by this hypothesis. The Maxwell's field, under the influence of gravity, will be described by the action,

$$S_E = -\frac{1}{16\pi c}\int F\,\sqrt{-g}\,d^4x\,, \tag{18}$$

where $F = F_{\mu\nu}F^{\mu\nu}$, and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the Maxwell tensor. When varying S_E in relation with A_μ we get precisely the Maxwell's equations in a curved spacetime, $q_{\mu\nu}$ in this case.

Assuming that the test particles follow geodesics relative to the geometry $q_{\mu\nu}$, and the Newtonian limit in the static weak field approximation and low velocities, we have that

$$\frac{d^2x^i}{dt^2} = -c^2\Gamma^i_{00} = -\partial^i\,\Phi_N\,, \qquad \text{with} \quad i = 1,2,3. \tag{19}$$

The last equality is relating the particle acceleration with the Newtonian gravitational force, where Φ_N represents Newton's potential.

From Eq. (7), we have

$$\Gamma^i_{00} \approx -\frac{1}{2}\partial^i\,\ln\alpha. \tag{20}$$

It follows that the Newtonian potential Φ_N is approximately given by

$$\frac{\Phi_N}{c^2} \approx -\frac{1}{2}\,\ln\alpha\,, \tag{21}$$

which yields the relation between the q_{00} component and the Newtonian potential,

$$q_{00} = \frac{1}{\alpha} \approx 1 + 2\,\frac{\Phi_N}{c^2} = 1 - 2\,\frac{GM}{c^2 r}\,, \tag{22}$$

where G is the Newtonian constant and M is the mass of the source.

However, this relation is determined up to a first order approximation in Φ, and in the development of GSG we will extrapolate the above relation by considering a more general expression for the α coefficient,

$$\alpha = e^{-2\Phi}\,. \tag{23}$$

The theory that we are constructing here presents three functions that have to be entirely determined by the end, the Lagrangian of the scalar field and the functions α and β. Since the geometrization method of the previous section gives a condition between them, and with α now being fixed, only remains to determine the Lagrangian of Φ.

4 Field Equation

Let us consider the following shape for the scalar field Lagrangian,

$$L = V(\Phi)\, w\,. \tag{24}$$

Following the second geometrization method in Sect. 2 we have that, in absence of other fields, the field equation is

$$\Box\Phi = 0\,, \tag{25}$$

and conditions (12) and (15) reduce to the expression

$$\alpha + \beta = \alpha^3\, V\,. \tag{26}$$

Important to note that we are not using the subscript q in the d'Alembertian operator anymore. Since in GSG Minkowski metric appears only as an auxiliary structure, we assume that all relevant quantities are constructed with the gravitational metric $q_{\mu\nu}$.

To select among all possible Lagrangians of the above form we look for indications from the various circumstances in which reliable experiments have been performed. In this vein, we initiate the discussion by analyzing the consequences of GSG for the solar system.

4.1 The Static and Spherically Symmetric Solution

Any theory of gravity must account for planetary orbits. In general relativity this motion is described by geodesics of the Schwarzschild geometry. In the GSG particles follow geodesics in the $q_{\mu\nu}$ metric.

Let us start by rewriting the auxiliary Minkowski background metric in spherical coordinates

$$ds_M^2 = dt^2 - dR^2 - R^2\, d\Omega^2. \tag{27}$$

Changing the radial coordinate to $R = \sqrt{\alpha}\, r$, where $\alpha = \alpha(r)$ we get

$$ds_M^2 = dt^2 - \alpha \left(\frac{r}{2\alpha} \frac{d\alpha}{dr} + 1 \right)^2 dr^2 - \alpha\, r^2\, d\Omega^2. \tag{28}$$

Since we are looking for static spherically symmetric solution we assume that the field depends only on the radial variable $\Phi = \Phi(r)$. Then the gravitational metric (7) takes the form

$$ds^2 = \frac{1}{\alpha} dt^2 - B\, dr^2 - r^2\, d\Omega^2, \tag{29}$$

where we have defined

$$B \equiv \frac{\alpha}{\alpha^3 V} \left(\frac{r}{2\alpha} \frac{d\alpha}{dr} + 1 \right)^2. \tag{30}$$

From the PPN analysis of the classical tests of gravitation (Will 2006) we know that the agreement with observations will be satisfied if we have

$$q_{00} \approx 1 - 2GM/c^2 r - 2(GM/c^2 r)^2 \quad \text{and} \quad q_{11} \approx 1 + 2GM/c^2 r\,. \tag{31}$$

Then, looking to Eq. (22), we can guarantee the correspondence between GSG and observations if we assume $B \approx \alpha$. However, we will again extrapolate this condition choosing a more general expression where $B = \alpha$. Using this the field equation can be easily solved, returning

$$\Phi = \frac{1}{2} \ln \left(1 - 2\frac{GM}{c^2 r} \right), \tag{32}$$

where we have used the asymptotic behavior to determine the integration constants and, from Eq. (30), we get

$$V(\Phi) = \frac{(\alpha - 3)^2}{4\,\alpha^3}\,. \tag{33}$$

With these results the line element of the static and spherically symmetric vacuum solution in GSG is given by

$$ds^2 = \left(1 - \frac{r_H}{r} \right) dt^2 - \left(1 - \frac{r_H}{r} \right)^{-1} dr^2 - r^2 d\Omega^2. \tag{34}$$

This geometry has the same form as in general relativity and yields the observed regime for solar tests. Thus, the present geometric scalar gravity is a good description of planetary orbits and also for light rays trajectories that follow geodesics (time-like and null-like, respectively) in the $q_{\mu\nu}$ geometry. If new observations

would require a modification of the metric in the neighborhood of a massive body this should be made by adjusting the form of the potential $V(\Phi)$.

4.2 Action Principle

Now that we have defined all functions for the theory we are in position to write its dynamical equation. Let us start by the action of the scalar field written in the auxiliary Minkowski background. From variational principle

$$\delta S_\Phi = \frac{1}{\kappa c}\,\delta \int \sqrt{-\eta}\, V(\Phi) w\, d^4x\,, \tag{35}$$

we get,

$$\delta S_\Phi = -\frac{2}{\kappa c}\int \sqrt{-\eta}\left(\frac{1}{2}\,V' w + V\,\eta^{\mu\nu}\partial_\mu\partial_\nu\Phi\right)\delta\Phi\, d^4x\,, \tag{36}$$

where κ is a constant with dimensions of distance/energy and the prime indicates a derivative in relation to Φ . The expression in parentheses above is just the left hand side of Eq. (5) and, by comparing with (8) using (26), it returns $\Box\Phi/\alpha^3$. Rewriting η in terms of q we finally get,

$$\delta S_\Phi = -2\int \sqrt{-q}\,\sqrt{V}\,\Box\Phi\,\delta\Phi\, d^4x\,. \tag{37}$$

In presence of matter we add a corresponding term L_m to the total action,

$$S_m = \frac{1}{c}\int \sqrt{-q}\, L_m\, d^4x\,. \tag{38}$$

The first variation of this term as usual yields

$$\delta S_m = -\frac{1}{2}\int \sqrt{-q}\, T^{\mu\nu}\delta q_{\mu\nu}\, d^4x\,, \tag{39}$$

where we have defined the energy-momentum tensor in the standard way

$$T_{\mu\nu} \equiv \frac{2}{\sqrt{-q}}\,\frac{\delta(\sqrt{-q}\,L_m)}{\delta q^{\mu\nu}}.$$

General covariance leads to conservation of the energy-momentum tensor ${T^{\mu\nu}}_{;\nu} = 0$.

The equation of motion is obtained by the action principle

$$\delta S_1 + \delta S_m = 0\,. \tag{40}$$

However, in the GSG theory, the metric $q_{\mu\nu}$ is not the fundamental quantity. We have to write the variation $\delta q_{\mu\nu}$ as function of $\delta\Phi$. After some calculation we get

$$\delta S_m = -\frac{1}{c}\int\left[T + \left(2 - \frac{V'}{2V}\right)E + C^{\lambda}{}_{;\lambda}\right]\delta\Phi\,\sqrt{-q}\,d^4x\,, \tag{41}$$

where we have defined some quantities as follows,

$$T \equiv T^{\mu\nu}\,q_{\mu\nu}, \quad E \equiv \frac{T^{\mu\nu}\,\partial_\mu\Phi\,\partial_\nu\Phi}{\Omega}\,, \tag{42}$$

$$C^{\lambda} \equiv \frac{\beta}{\alpha\,\Omega}\left(T^{\lambda\mu} - E\,q^{\lambda\mu}\right)\,\partial_\mu\Phi\,, \tag{43}$$

and " ; " means the covariant derivative in respect to the q-metric.

Finally, the equation of motion for the gravitational field Φ takes the form

$$\sqrt{V}\,\Box\Phi = \kappa\,\chi, \tag{44}$$

with the notation simplified by writing

$$\chi = -\frac{1}{2}\left[T + \left(2 - \frac{V'}{2V}\right)E + C^{\lambda}{}_{;\lambda}\right]. \tag{45}$$

Equation (44) describes the dynamics of GSG in presence of matter, under the assumptions (23) and (33). The quantity χ involves a non-trivial coupling between the gradient of the scalar field and the complete energy-momentum tensor of the matter, and not uniquely its trace. This property allows the electromagnetic field to interact with the gravitational field. The Newtonian limit gives the identification

$$\kappa \equiv \frac{8\pi G}{c^4}\,. \tag{46}$$

5 Final Comments

GSG is an alternative propose to describe the gravitational process using a single scalar field, but it still treats gravity as a geometrical effect and all kind of matter and energy interact with gravitational potential only through metric $q_{\mu\nu}$ in Eq. (7). With different premises from that previous scalars theories, GSG overcomes the problems surrounding the scalar gravity.

Guided by observations too, we develop GSG choosing the Lagrangian of Φ as $L = Vw$, with

$$V = \frac{(3-\alpha)^2}{4\alpha^3}\,, \tag{47}$$

where the Newtonian limit of the theory led us to work with

$$\alpha = e^{-2\Phi}\,. \tag{48}$$

The geometrization technique is what gives the relation between the β coefficient of the metric and these two other functions, namely

$$\alpha + \beta = \alpha^3\, V\,. \tag{49}$$

Therewith, the field equation of the theory is given by

$$\sqrt{V}\,\Box\Phi = \kappa\,\chi\,, \tag{50}$$

with χ defined in (45).

Even so, the GSG can be seen as a little more than an unique theory in the sense that it represents a way in which is possible to develop scalar theories of gravitation. Relaxing the expressions for α and V can still be in agreement with observations while given a very different gravitational theory.

GSG is a result of a wonderful work with Mario Novello, Ugo Moschella, Eduardo Bittencourt and others. The ideas here can be found with more details in Novello et al. (2013). Also, in Bittencourt et al. (2014), there is the consequences of this theory for the cosmology.

References

Bittencourt, E., Moschella, U., Novello, M., Toniato, J.D.: Phys. Rev. D **90**, 123540 (2014)
Clifton, T.: Alternative theories of gravity. PhD Thesis, Cambridge University (2006)
Dowker, J.S.: Proc. Phys. Soc. **85**, 595–600 (1965)
Feynman, R.P., Moringo, F.B., Wagner, W.G.: Feynman Lectures on Gravitation. Addison-Wesley, Massachusetts (1995)
Goulart, E., Novello, M., Falciano, F., Toniato, J.D.: Class. Quantum Gravit. **28**, 245008 (2011)
Ni, W.-T.: Astrophys. J. **176**, 769–796 (1972)
Novello, M., et al.: J. Cosmol. Astropart. Phys. **06**, 014 (2013)
Turyshev, S.G.: Phys. Usp. **52**, 1–27 (2009)
Will, C.M.: The confrontation between general relativity and experiment. Living Rev. Relativ. **9**, 3 (2006)