Lecture Notes in Physics

The Editorial Policy for Edited Volumes

The series Lecture Notes in Physics reports new developments in physical research and teaching - quickly, informally, and at a high level. The type of material considered for publication includes monographs presenting original research or new angles in a classical field. The timeliness of a manuscript is more important than its form, which may be preliminary or tentative. Manuscripts should be reasonably self-contained. They will often present not only results of the author(s) but also related work by other people and will provide sufficient motivation, examples, and applications.

Acceptance

The manuscripts or a detailed description thereof should be submitted either to one of the series editors or to the managing editor. The proposal is then carefully refereed. A final decision concerning publication can often only be made on the basis of the complete manuscript, but otherwise the editors will try to make a preliminary decision as definite as they can on the basis of the available information.

Contractual Aspects

Authors receive jointly 30 complimentary copies of their book. No royalty is paid on Lecture Notes in Physics volumes. But authors are entitled to purchase directly from Springer other books from Springer (excluding Hager and Landolt-Börnstein) at a $33\frac{1}{3}$% discount off the list price. Resale of such copies or of free copies is not permitted. Commitment to publish is made by a letter of interest rather than by signing a formal contract. Springer secures the copyright for each volume.

Manuscript Submission

Manuscripts should be no less than 100 and preferably no more than 400 pages in length. Final manuscripts should be in English. They should include a table of contents and an informative introduction accessible also to readers not particularly familiar with the topic treated. Authors are free to use the material in other publications. However, if extensive use is made elsewhere, the publisher should be informed. As a special service, we offer free of charge LaTeX macro packages to format the text according to Springer's quality requirements. We strongly recommend authors to make use of this offer, as the result will be a book of considerably improved technical quality. The books are hardbound, and quality paper appropriate to the needs of the author(s) is used. Publication time is about ten weeks. More than twenty years of experience guarantee authors the best possible service.

LNP Homepage (springerlink.com)

On the LNP homepage you will find:
—The LNP online archive. It contains the full texts (PDF) of all volumes published since 2000. Abstracts, table of contents and prefaces are accessible free of charge to everyone. Information about the availability of printed volumes can be obtained.
—The subscription information. The online archive is free of charge to all subscribers of the printed volumes.
—The editorial contacts, with respect to both scientific and technical matters.
—The author's / editor's instructions.

E. Papantonopoulos (Ed.)

The Physics of the Early Universe

Springer

Editor

E. Papantonopoulos
National Technical University of Athens
Physics Department
Zografou
15780 Athens
Greece

E. Papantonopoulos (Ed.), *The Physics of the Early Universe*, Lect. Notes Phys. **653** (Springer, Berlin Heidelberg 2005), DOI 10.1007/b99562

Library of Congress Control Number: 2004116343

ISSN 0075-8450
ISBN 3-540-22712-1 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by the authors/editor
Data conversion: PTP-Berlin Protago-TEX-Production GmbH
Cover design: *design & production*, Heidelberg

Printed on acid-free paper
54/3141/ts - 5 4 3 2 1 0

Lecture Notes in Physics

For information about Vols. 1–606
please contact your bookseller or Springer
LNP Online archive: springerlink.com

Vol.607: R. Guzzi (Ed.), Exploring the Atmosphere by Remote Sensing Techniques.

Vol.608: F. Courbin, D. Minniti (Eds.), Gravitational Lensing:An Astrophysical Tool.

Vol.609: T. Henning (Ed.), Astromineralogy.

Vol.610: M. Ristig, K. Gernoth (Eds.), Particle Scattering, X-Ray Diffraction, and Microstructure of Solids and Liquids.

Vol.611: A. Buchleitner, K. Hornberger (Eds.), Coherent Evolution in Noisy Environments.

Vol.612: L. Klein, (Ed.), Energy Conversion and Particle Acceleration in the Solar Corona.

Vol.613: K. Porsezian, V.C. Kuriakose (Eds.), Optical Solitons. Theoretical and Experimental Challenges.

Vol.614: E. Falgarone, T. Passot (Eds.), Turbulence and Magnetic Fields in Astrophysics.

Vol.615: J. Büchner, C.T. Dum, M. Scholer (Eds.), Space Plasma Simulation.

Vol.616: J. Trampetic, J. Wess (Eds.), Particle Physics in the New Millenium.

Vol.617: L. Fernández-Jambrina, L. M. González-Romero (Eds.), Current Trends in Relativistic Astrophysics, Theoretical, Numerical, Observational

Vol.618: M.D. Esposti, S. Graffi (Eds.), The Mathematical Aspects of Quantum Maps

Vol.619: H.M. Antia, A. Bhatnagar, P. Ulmschneider (Eds.), Lectures on Solar Physics

Vol.620: C. Fiolhais, F. Nogueira, M. Marques (Eds.), A Primer in Density Functional Theory

Vol.621: G. Rangarajan, M. Ding (Eds.), Processes with Long-Range Correlations

Vol.622: F. Benatti, R. Floreanini (Eds.), Irreversible Quantum Dynamics

Vol.623: M. Falcke, D. Malchow (Eds.), Understanding Calcium Dynamics, Experiments and Theory

Vol.624: T. Pöschel (Ed.), Granular Gas Dynamics

Vol.625: R. Pastor-Satorras, M. Rubi, A. Diaz-Guilera (Eds.), Statistical Mechanics of Complex Networks

Vol.626: G. Contopoulos, N. Voglis (Eds.), Galaxies and Chaos

Vol.627: S.G. Karshenboim, V.B. Smirnov (Eds.), Precision Physics of Simple Atomic Systems

Vol.628: R. Narayanan, D. Schwabe (Eds.), Interfacial Fluid Dynamics and Transport Processes

Vol.629: U.-G. Meißner, W. Plessas (Eds.), Lectures on Flavor Physics

Vol.630: T. Brandes, S. Kettemann (Eds.), Anderson Localization and Its Ramifications

Vol.631: D. J. W. Giulini, C. Kiefer, C. Lämmerzahl (Eds.), Quantum Gravity, From Theory to Experimental Search

Vol.632: A. M. Greco (Ed.), Direct and Inverse Methods in Nonlinear Evolution Equations

Vol.633: H.-T. Elze (Ed.), Decoherence and Entropy in Complex Systems, Based on Selected Lectures from DICE 2002

Vol.634: R. Haberlandt, D. Michel, A. Pöppl, R. Stannarius (Eds.), Molecules in Interaction with Surfaces and Interfaces

Vol.635: D. Alloin, W. Gieren (Eds.), Stellar Candles for the Extragalactic Distance Scale

Vol.636: R. Livi, A. Vulpiani (Eds.), The Kolmogorov Legacy in Physics, A Century of Turbulence and Complexity

Vol.637: I. Müller, P. Strehlow, Rubber and Rubber Balloons, Paradigms of Thermodynamics

Vol.638: Y. Kosmann-Schwarzbach, B. Grammaticos, K.M. Tamizhmani (Eds.), Integrability of Nonlinear Systems

Vol.639: G. Ripka, Dual Superconductor Models of Color Confinement

Vol.640: M. Karttunen, I. Vattulainen, A. Lukkarinen (Eds.), Novel Methods in Soft Matter Simulations

Vol.641: A. Lalazissis, P. Ring, D. Vretenar (Eds.), Extended Density Functionals in Nuclear Structure Physics

Vol.642: W. Hergert, A. Ernst, M. Däne (Eds.), Computational Materials Science

Vol.643: F. Strocchi, Symmetry Breaking

Vol.644: B. Grammaticos, Y. Kosmann-Schwarzbach, T. Tamizhmani (Eds.) Discrete Integrable Systems

Vol.645: U. Schollwöck, J. Richter, D.J.J. Farnell, R.F. Bishop (Eds.), Quantum Magnetism

Vol.646: N. Bretón, J. L. Cervantes-Cota, M. Salgado (Eds.), The Early Universe and Observational Cosmology

Vol.647: D. Blaschke, M. A. Ivanov, T. Mannel (Eds.), Heavy Quark Physics

Vol.648: S. G. Karshenboim, E. Peik (Eds.), Astrophysics, Clocks and Fundamental Constants

Vol.649: M. Paris, J. Rehacek (Eds.), Quantum State Estimation

Vol.650: E. Ben-Naim, H. Frauenfelder, Z. Toroczkai (Eds.), Complex Networks

Vol.651: J.S. Al-Khalili, E. Roeckl (Eds.), The Euroschool Lectures of Physics with Exotic Beams, Vol.I

Vol.652: J. Arias, M. Lozano (Eds.), Exotic Nuclear Physics

Vol.653: E. Papantonoupoulos (Ed.), The Physics of the Early Universe

Preface

This book is an edited version of the review talks given in the Second Aegean School on the Early Universe, held in Ermoupolis on Syros Island, Greece, in September 22-30, 2003. The aim of this book is not to present another proceedings volume, but rather an advanced multiauthored textbook which meets the needs of both the postgraduate students and the young researchers, in the field of Physics of the Early Universe.

The first part of the book discusses the basic ideas that have shaped our current understanding of the Early Universe. The discovering of the Cosmic Microwave Background (CMB) radiation in the sixties and its subsequent interpretation, the numerous experiments that followed with the enumerable observation data they produced, and the recent all-sky data that was made available by the Wilkinson Microwave Anisotropy Probe (WMAP) satellite, had put the hot big bang model, its inflationary cosmological phase and the generation of large scale structure, on a firm observational footing.

An introduction to the Physics of the Early Universe is presented in K. Tamvakis' contribution. The basic features of the hot Big Bang Model are reviewed in the framework of the fundamental physics involved. Shortcomings of the standard scenario and open problems are discussed as well as the key ideas for their resolution.

It was an old idea that the large scale structure of our Universe might have grown out of small initial fluctuations via gravitational instability. Now we know that matter density fluctuations can grow like the scale factor and then the rapid expansion of the universe during inflation generates the large scale structure of our Universe. R. Durrer's review offers a systematic treatment of cosmological perturbation theory. After the introduction of gauge invariant variables, the Einstein and conservation equations are written in terms of these variables. The generation of perturbations during inflation is studied. The importance of linear cosmological perturbation theory as a powerful tool to calculate CMB anisotropies and polarisation is explained.

The linear anisotropies in the temperature of CMB radiation and its polarization provide a clean picture of fluctuations in the universe after the big bang. These fluctuations are connected to those present in the ultra-high-energy universe, and this makes the CMB anisotropies a powerful tool for constraining the fundamental physics that was responsible for the generation of structure. Late time effects also leave their mark, making the CMB tem-

perature and polarization useful probes of dark energy and the astrophysics of reionization. A. Challinor's contribution discusses the simple physics that processes primordial perturbations into the linear temperature and polarization anisotropies. The role of the CMB in constraining cosmological parameters is also described, and some of the highlights of the science extracted from recent observations and the implications of this for fundamental physics are reviewed.

It is of prime interest to look for possible systematic uncertainties in the observations and their interpretation and also for possible inconsistencies of the standard cosmological model with observational data. This is important because it might lead us to new physics. Deviations from the standard cosmological model are strongly constrained at early times, at energies on the order of 1 MeV. However, cosmological evolution is much less constrained in the post-recombination universe where there is room for deviation from standard Friedmann cosmology and where the more classical tests are relevant. R. Sander's contribution discusses three of these classical cosmological tests that are independent of the CMB: the angular size distance test, the luminosity distance test and its application to observations of distant supernovae, and the incremental volume test as revealed by faint galaxy number counts.

The second part of the book deals with the missing pieces in the cosmological puzzle that the CMB anisotropies, the galaxies rotation curves and microlensing are suggesting: dark matter and dark energy. It also presents new ideas which come from particle physics and string theory which do not conflict with the standard model of the cosmological evolution but give new theoretical alternatives and offer a deeper understanding of the physics involved.

Our current understanding of dark matter and dark energy is presented in the review by V. Sahni. The review first focusses on issues pertaining to dark matter including observational evidence for its existence. Then it moves to the discussion of dark energy. The significance of the cosmological constant problem in relation to dark energy is discussed and emphasis is placed upon dynamical dark energy models in which the equation of state is time dependent. These include Quintessence, Braneworld models, Chaplygin gas and Phantom energy. Model independent methods to determine the cosmic equation of state are also discussed. The review ends with a brief discussion of the fate of the universe in dark energy models.

The next contribution by A. Lukas provides an introduction into time-dependent phenomena in string theory and their possible applications to cosmology, mainly within the context of string low energy effective theories. A major problem in extracting concrete predictions from string theory is its large vacuum degeneracy. For this reason M-theory (the largest theory that includes all the five string theories) at present, cannot provide a coherent picture of the early universe or make reliable predictions. In this contribution particular emphasis is placed on the relation between string theory and inflation.

In an another development of theoretical ideas which come from string theory, the universe could be a higher-dimensional spacetime, with our observable part of the universe being a four-dimensional "brane" surface. In this picture, Standard Model particles and fields are confined to the brane while gravity propagates freely in all dimensions. R. Maartens' contribution provides a systematic and detailed introduction to these ideas, discussing the geometry, dynamics and perturbations of simple braneworld models for cosmology.

The last part of the book deals with a very important physical process which hopefully will give us valuable information about the structure of the Early Universe and the violent processes that followed: the gravitational waves. One of the central predictions of Einsteins' general theory of relativity is that gravitational waves will be generated as masses are accelerated. Despite decades of effort these ripples in spacetime have still not been observed directly.

As several large scale interferometers are beginning to take data at sensitivities where astrophysical sources are predicted, the direct detection of gravitational waves may well be imminent. This would (finally) open the long anticipated gravitational wave window to our Universe. The review by N. Andersson and K. Kokkotas provides an introduction to gravitational radiation. The key concepts required for a discussion of gravitational wave physics are introduced. In particular, the quadrupole formula is applied to the anticipated source for detectors like LIGO, GEO600, EGO and TAMA300: inspiralling compact binaries. The contribution also provides a brief review of high frequency gravitational waves.

Over the last decade, advances in computer hardware and numerical algorithms have opened the door to the possibility that simulations of sources of gravitational radiation can produce valuable information of direct relevance to gravitational wave astronomy. Simulations of binary black hole systems involve solving the Einstein equation in full generality. Such a daunting task has been one of the primary goals of the numerical relativity community. The contribution by P. Laguna and D. Shoemaker focusses on the computational modelling of binary black holes. It provides a basic introduction to the subject and is intended for non-experts in the area of numerical relativity.

The Second Aegean School on the Early Universe, and consequently this book, became possible with the kind support of many people and organizations. We received financial support from the following sources and this is gratefully acknowledged: National Technical University of Athens, Ministry of the Aegean, Ministry of the Culture, Ministry of National Education, the Eugenides Foundation, Hellenic Atomic Energy Committee, Metropolis of Syros, National Bank of Greece, South Aegean Regional Secretariat.

We thank the Municipality of Syros for making available to the Organizing Committee the Cultural Center, and the University of the Aegean for providing technical support. We thank the other members of the Organizing Committee of the School, Alex Kehagias and Nikolas Tracas for all

their efforts in resolving many issues that arose in organizing the School. The administrative support of the School was taken up with great care by Mrs. Evelyn Pappa. We acknowledge the help of Mr. Yionnis Theodonis who designed and maintained the webside of the School. We also thank Vasilis Zamarias for assisting us in resolving technical issues in the process of editing this book.

Last, but not least, we are grateful to the staff of Springer-Verlag, responsible for the Lecture Notes in Physics, whose abilities and help contributed greatly to the appearance of this book.

Athens, May 2004 *Lefteris Papantonopoulos*

Contents

List of Contributors

Nils Andersson
School of Mathematics,
University of Southampton,
Southampton SO17 1BJ, UK
N.Andersson@maths.soton.ac.uk

Anthony Challinor
Astrophysics Group,
Cavendish Laboratory,
Madingley Road,
Cambridge, CB3 0HE, UK
a.d.challinor@mrao.cam.ac.uk

Ruth Durrer
Université de Genève,
Département de Physique Théorique,
24 Quai E. Ansermet,
1211 Genève, Switzerland
ruth.durrer@physics.unige.ch

Kostas D. Kokkotas
Department of Physics,
Aristotle University of Thessaloniki,
541 24 Thessaloniki, Greece and
Center for Gravitational Wave
Physics, 104 Davey Laboratory,
University Park, PA 16802, USA
kokkotas@auth.gr

Pablo Laguna
Department of Astronomy and
Astrophysics, Institute for Gravitational Physics and Geometry,
Center for Gravitational Wave
Physics, Penn State University,
University Park, PA 16802, USA
pablo@astro.psu.edu

André Lukas
Department of Physics
and Astronomy,
University of Sussex,
Brighton BN1 9QH, UK
a.lukas@sussex.ac.uk

Roy Maartens
Institute of Cosmology
and Gravitation,
University of Portsmouth,
Portsmouth PO1 2EG, UK
roy.Maartens@port.ac.uk

Varun Sahni
Inter-University Center
for Astronomy and Astrophysics,
Puné 411 007, India
varun@iucaa.ernet.in

Robert H. Sanders
Kapteyn Astronomical Institute,
Groningen, The Netherlands
sanders@astro.rug.nl

Deirdre M. Shoemaker
Center for Radiophysics and Space
Research, Cornell University,
Ithaca, NY 14853, USA
deirdre@astro.cornell.edu

Kyriakos Tamvakis
Physics Department,
University of Ioannina,
451 10 Ioannina, Greece
tamvakis@cc.uoi.gr

1 An Introduction to the Physics of the Early Universe

Kyriakos Tamvakis

Physics Department, University of Ioannina, 451 10 Ioannina, Greece

Abstract. We present an elementary introduction to the Early Universe. The basic features of the hot Big Bang are reviewed in the framework of the fundamental physics involved. Shortcomings of the standard scenario and open problems are discussed as well as the key ideas for their resolution.

1.1 The Hubble Law

In a restricted sense Cosmology is the study of the large scale structure of the universe. In a modern, much wider, sense it seeks to assemble all our knowledge of the Universe into a unified picture [1]. Our present view of the Universe is based on the observational evidence and a few theoretical concepts. Central in the established theoretical framework is *Einstein's General Theory of Relativity (GR)* [2] and the dominant role of gravity in the evolution of the Universe. The discovery of the Expansion of the Universe provided the most important established feature of the modern cosmological picture. In addition, the observation of the *Cosmic Microwave Background Radiation (CMB)* provided a strong connection of the present cosmological picture to fundamental Particle Physics.

In 1929 Edwin Hubble [3] announced his discovery that the *redshifts* of galaxies tend to increase with distance. According to the Doppler shift phenomenon, the wavelength of light from a moving source increases according to the formula $\lambda' = \lambda(1 + V/c)$. This formula is modified for relativistic velocities. The quantity $z \equiv \Delta\lambda/\lambda$ is called the redshift. The non-relativistic Doppler formula reads $z = V/c$. The relation discovered by Hubble is

$$z = \frac{\Delta\lambda}{\lambda} \propto L \ . \tag{1.1}$$

Subsequent measurements by him and others established beyond doubt the *Velocity-Distance Law*

$$V \sim H \times L \ . \tag{1.2}$$

Usually the name *Hubble Law* is reserved for the redshift-distance proportionality.

K. Tamvakis, An Introduction to the Physics of the Early Universe, Lect. Notes Phys **653**, 3–29 (2005)
http://www.springerlink.com/

The parameter H is called the Hubble parameter and it has today a value of the order of $100\,km(sec)^{-1}(Mpc)^{-1} = (9.778 \times Gyr)^{-1}$. The Hubble Law established the idea that the Universe consists of expanding space. The light from distant galaxies is redshifted because their separation distance increases due to the expansion of space. The Hubble parameter is constant throughout space at a common instant of time but it is not constant in time. The expansion may have been faster in the past. Observational data support the picture of a Universe that is to a very good approximation *homogeneous* (all places are alike) and *isotropic* (all directions are alike). The hypotheses of homogeneity and isotropy are referred to as the *Cosmological Principle*. Such a Universe is called *uniform*. A uniform Universe remains uniform if its motion is uniform. Thus, the expansion corresponds only to *dilation*, being almost entirely shear-free and irrotational. The Hubble Law can be easily deduced from these facts.

1.2 Comoving Coordinates and the Scale Factor

Homogeneity of the Universe implies also all clocks agree in their intervals of time. Universal time is also refered to as cosmic time. Considering only uniform expansion we introduce a *comoving coordinate system*. All distances between comoving points increase by the same factor. In a comoving coordinate system there exists a universal scale factor R, that increases in time if the Universe is uniformly expanding (or decreases with time if the Universe is uniformly contracting). The scale factor $R(t)$ is a function of cosmic time and has the same value throughout space. All lengths increase with time in proportion to R, all surfaces in proportion to R^2 and all volumes in proportion to R^3.

If R_0 is the value of the scale factor at the present time and L_0 the distance between two comoving points, the corresponding distance at any other time t will be $L(t) = (L_0/R_0)\,R(t)$. If an expanding volume V contains N particles, we can write for the particle number density $n = n_0(R/R_0)^3$. As an application of the last formula, from the present (average) density of matter in the Universe of about one hydrogen atom per cubic meter, we can estimate the average density of matter at an earlier time. At the time at which the scale factor was 1% of what it is today the average matter density was one hydrogen atom per cubic centimeter.

Consider now a comoving body at a fixed coordinate distance. Its actual distance will be proportional to the scale factor, namely $L = R \times$ (coordinate distance). The recession velocity of the comoving body will be proportional to the rate of increase of the scale factor $\dot{R}$, namely $V = \dot{R} \times$ (coordinate distance). Dividing the two relations, we obtain

$$V = L\frac{\dot{R}}{R}\,, \tag{1.3}$$

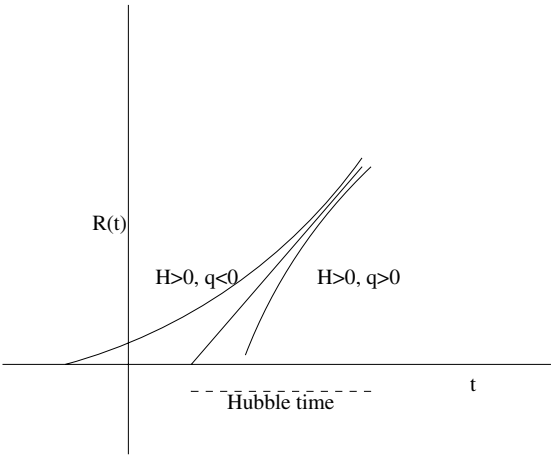

Fig. 1.1. The age of the Universe and Hubble time.

which is the Velocity-Distance Law in another form. The two expressions coincide if we identify the Hubble parameter with the rate of change of the scale factor

$$H = \frac{\dot{R}}{R} \ . \tag{1.4}$$

The Hubble parameter is a time-dependent quantity. Note again that the Velocity-Distance is a simple consequence of uniform expansion. The existence of a scale factor, that is the same throughout space and varies in time, leads directly to the Velocity-Distance Law.

If the Hubble parameter was constant, or if, equivalently, the rate of expansion of the Universe was constant, the inverse of the Hubble parameter would give the time of expansion. This time is $t_H \equiv H_0^{-1}$ and it is called the Hubble time. Although in almost all cosmological models that are being studied the Hubble parameter is not a constant, the Hubble time, thus defined, gives a (rough) measure of the age of the Universe (see Fig. 1.1). Numerically, the Hubble time comes out to be $t_H \sim 10\,h^{-1}$ billion years, where the dimensionless parameter h is called *normalized Hubble parameter* and is a number between 0.5 and 0.8.

Acceleration is by definition the rate of increase of the velocity, namely $\dot{V} = \ddot{R}\times$ (coordinate distance). As before, the coordinate distance of a comoving body is constant. On the other hand, we know that $L = R\times$ (coordinate distance). Thus,

$$\dot{V} = L\frac{\ddot{R}}{R} \ . \tag{1.5}$$

We can define a *deceleration parameter*, independent of the particular body at comoving distance L, as the dimensionless parameter

$$q \equiv -\frac{\ddot{R}}{RH^2} \ . \tag{1.6}$$

When q is positive, it corresponds to deceleration, while, when it is negative, it corresponds to acceleration and should properly be refered to as *acceleration parameter*. We can actually classify uniform Universes according to their values of H and q. Such a classification should be called kinematic classification, in contrast to a classification in terms of the curvature, which is a geometric classification. Kinematically, uniform Universes fall into the following classes:

a) $(H > 0,\, q > 0)$ expanding and decelerating
b) $(H > 0,\, q < 0)$ expanding and accelerating
c) $(H < 0,\, q > 0)$ contracting and decelerating
d) $(H < 0,\, q < 0)$ contracting and accelerating
e) $(H > 0,\, q = 0)$ expanding with zero deceleration
f) $(H < 0,\, q = 0)$ contracting with zero deceleration
g) $(H = 0,\, q = 0)$ static.

There is little doubt that only (a), (b) and (e) are possible candidates for our Universe at present. Extrapolating an expanding scenario backwards, we arrive at a very high density state at $R \to 0$. Evidence from CMB radiation suggests that such a state, described by the suggestive name *Big Bang*[1] could have occurred in the Early Universe.

1.3 The Cosmic Microwave Background

The Hubble expansion can be understood as a natural consequence of homogeneity and isotropy. Nevertheless, an expanding Universe must necessarily have a much denser and, therefore, hotter past. Matter in the Early Universe, at times much before the development of any structure, should be viewed as a gas of relativistic particles in thermodynamic equilibrium. The expansion cannot upset the equilibrium, since the characteristic rate of particle processes is of the order of the characteristic energy, namely T, while the rate of expansion is given by the much smaller scale $H \sim \sqrt{G}\, T^2 \sim (T/M_P)\, T$. In order to be convinced for this, one has to invoke the Friedmann equation (see next chapter) and consider the temperature dependence of the energy-density $\rho \sim T^4$ characteristic of radiation. The model of the Early Universe as a gas of relativistic matter and electromagnetic radiation in equilibrium was first considered [4] by G. Gamow and his collaborators R. Alpher and R. Herman for the purpose of explaining nucleosynthesis. As a byproduct, the existence of relic black body radiation was predicted with wavelength in the range of microwaves corresponding to temperature of a few degrees Kelvin.

[1] This term was first used by Fred Hoyle in a series of BBC radio talks, published in *The Nature of the Universe* (1950). Fred Hoyle was the main proponent of the rival Steady State Theory [9] of the Universe.

This radiation, now known as Cosmic Microwave Background (CMB), was discovered in 1965 by A. Penzias and R. Wilson [5] (see A. Challinor's contribution). The radiation, once extremely hot, has been cooled over billions of years, redshifted by the expansion of the Universe and has today a temperature of a few degrees Kelvin. Black body radiation of a temperature T reaches a maximum at a characteristic wavelength $\lambda_{max} \sim (1.26\,\hbar c/k_B)\,T$. The average wavelength is of that order. Very accurate observations by the Cosmic Background Explorer (COBE) [6] have shown that the intensity of the CMB follows the blackbody curve of thermal radiation with a deviation of only one part in 10^4. Also, after the subtraction of a 24-hour anisotropy that has to do with the motion of the Galaxy at a speed $V = 600\,km/sec$ ($\Delta T/T \sim V/c \sim 0.01$), the radiation is surprising isotropic with only very small anisotropies of order $\sim 10^{-5}$. Very recently [7], $WMAP$ has pushed the accuracy with which these anisotropies are determined down to 10^{-9}. These anisotropies, surviving from the time of decoupling, are the imprint of density fluctuations that evolved into galaxies and clusters of galaxies. The accuracy with which CMB obeys the Planck spectrum is a very strong physical constraint in favour of an expanding Universe that passes through a hot stage. The COBE estimate of the CMB temperature is

$$T_{CMB} = 2.725 \pm 0.002\ ^oK\ .$$

It is possible to get a qualitative idea of the central event related to the relic CMB without going into to much detail. The required quantitative relations can easily be met in the framework of specific cosmological models to be discussed later. We could start at some time in the history of our Universe when the temperature was greater than $10^{10}\ ^oK$. This corresponds roughly to energy of about $1\,MeV$. The abundant particles, i.e. those with masses smaller than the characteristic energy $k_B\,T$, apart from the massless photon are the electrons, neutrinos and their antiparticles. The energy is dominated by the radiation of these particles, which are, at these energies practically massless as the photon. Reactions such as $e + e^+ \rightleftharpoons \gamma + \gamma$ are in thermodynamic equilibrium, not affected at all by the much slower expansion. The very important effect of the expansion is to lower the temperature, which decreases inversely proportional to the scale factor. No qualitative change occurs until the temperature drops below the characteristic threshold energy $k_B\,T \sim m_e c^2$ at which photons can achieve electron-positron pair creation. Below that temperature all electrons and positrons disappear from the plasma. The photon radiation decouples and the Universe becomes essentially transparent to it. It is exactly these photons which, redshifted, we observe as CMB.

The Hubble expansion by itself does not provide sufficient evidence for a Big Bang type of Cosmology. It is only after the observation of the Cosmic Microwave Background and subsequent work on Nucleosynthesis that the Big Bang Model was established as the basic candidate for a Standard Cosmological Model.

1.4 The Friedmann Models

A Cosmological Model is a (very) simplified model of the Universe with a geometrical description of spacetime and a smoothed-out matter and radiation content. The simplest interesting set of cosmological models is provided by the homogeneous and isotropic *Friedmann-Lemaitre* spacetimes (FL) [8] which are a set of solutions of GR incorporating the Cosmological Principle. The line element of a FL model reads

$$ds^2 = dt^2 - R^2(t)d\sigma^2 \ . \tag{1.7}$$

The spatial line element $d\sigma^2$ describes a three-dimensional space of constant curvature independent of time. It is [2]

$$d\sigma^2 = d\chi^2 + f^2(\chi)\left(d\theta^2 + \sin^2\theta\, d\phi^2\right) = d\chi^2 + f^2(\chi)\, d\Omega^2 \ . \tag{1.8}$$

These coordinates are comoving. That means that the actual spatial distance of two points (χ, θ, ϕ) and (χ_0, θ, ϕ) will be $d = R(t)(\chi - \chi_0)$. There are three choices for $f(\chi)$, each corresponding to a different spatial curvature k. That is the value of the Ricci scalar (to be defined below) calculated from $d\sigma^2$ with the scale factor divided out. They are

$$f(\chi) = \begin{cases} \sin\chi & (k=+1) \quad 0 < \chi < \pi \\ \chi & (k=0) \quad 0 < \chi < \infty \\ \sinh\chi & (k=-1) \quad 0 < \chi < \infty \end{cases} \ . \tag{1.9}$$

The case $k = +1$ corresponds to a closed spacetime with a spherical spatial geometry. The case $k = 0$ corresponds to an infinite (flat) spacetime with Euclidean spatial geometry. Finally, the case $k = -1$ corresponds to an open spacetime with hyperbolic spatial geometry. Sometimes the Robertson-Walker metric is written in terms of $r \equiv f(\chi)$ as

$$d\sigma^2 = \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \ .$$

The above metric comes out as a solution of Einstein's Equations

$$\mathcal{R}_{\mu\nu} - \frac{1}{2}\mathcal{R}\, g_{\mu\nu} - \Lambda\, g_{\mu\nu} = 8\pi G\, T_{\mu\nu} \ , \tag{1.10}$$

$R_{\mu\nu}$ is the *Riemann Curvature Tensor* and $\mathcal{R}$ is the *Ricci Scalar* defined as $\mathcal{R} = g^{\mu\nu}\mathcal{R}_{\mu\nu}$. G stands for Newton's Constant of Gravitation. The constant Λ is called the *Cosmological Constant* and $T_{\mu\nu}$ is the *Matter Energy-Momentum Tensor*. A usual choice is that of a fluid

[2] This is the so called Robertson-Walker metric. A more complete name for these spacetime solutions is Friedmann-Lemaitre-Robertson-Walker or just FLRW models.

$$T_\mu^\nu = (-\rho,\, p,\, p\, p)\ , \tag{1.11}$$

with ρ the energy density and p the momentum density, related through some Equation of State.

In the framework of the Robertson-Walker metric, light emitted from a source at the point χ_S at time t_S, propagating along a null geodesic $d\sigma^2 = 0$, taken radial ($d\Omega^2 = 0$) without loss of generality, will reach us at $\chi_0 = 0$ at time t_0 given by

$$\int_{t_S}^{t_0} \frac{dt}{R(t)} = \chi_S\ .$$

A second signal emitted at $t_S + \delta t_S$ will satisfy

$$\int_{t_S+\delta t_S}^{t_0+\delta t_0} \frac{dt}{R(t)} = \chi_S \Rightarrow \frac{\delta t_S}{R(t_S)} = \frac{\delta t_0}{R(t_0)}\ .$$

The ratio of the observed frequencies will be

$$\frac{\omega_0}{\omega_S} = \frac{\delta t_S}{\delta t_0} = \frac{R(t_S)}{R(t_0)}\ .$$

This implies

$$z \equiv \frac{\lambda_0 - \lambda_S}{\lambda_S} = \frac{R(t_0)}{R(t_S)} - 1 \sim -1 + \frac{R(t_0)}{R(t_0) - (t_0 - t_S)\dot{R}(t_0)}$$

$$z \sim (t_0 - t_S)H(t_0) \Rightarrow z = H\, d\ . \tag{1.12}$$

This is the *Hubble Law*. The Velocity-Distance Law is a simple consequence of uniformity, namely

$$V = \dot{d} = \dot{R}\,\frac{d}{R} = H\, d\ . \tag{1.13}$$

Inserting the Robertson-Walker metric into Einstein's Equations, we arrive at the two equations

$$\ddot{R} = -\frac{4\pi G}{3}(\rho + 3p)\, R + \frac{\Lambda}{3}\, R \tag{1.14}$$

$$(\dot{R})^2 = \frac{8\pi G}{3}\,\rho\, R^2 + \frac{\Lambda}{3}\, R^2 - k\ . \tag{1.15}$$

Multiplying the first of these equations by $\dot{R}$ and using the second, we arrive at the equivalent pair of two first order equations, namely

$$\dot{\rho} + 3(\rho + p)\frac{\dot{R}}{R} = 0 \tag{1.16}$$

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{k}{R^2}\ . \tag{1.17}$$

The first of these equations is the Continuity Equation expressing the conservation of energy for the comoving volume R^3. This interpretation is more transparent if we write it in the form

$$\frac{d}{dt}\left(\frac{4\pi R^3}{3}\rho\right) = p\left(\frac{4\pi R^3}{3}\right) \Leftrightarrow \frac{d\mathcal{E}}{dt} = p\mathcal{V} \ .$$

The other equation is purely dynamical and determines the evolution of the scale factor. It is called *The Friedmann Equation*.

At the present epoch we have to a very good approximation $p_0 \approx 0$. We can write (1.15) and (1.14) in terms of the present Hubble parameter H_0 and the present deceleration parameter q_0. It is convenient to introduce a critical density ρ_c defined as

$$\rho_c \equiv \frac{3H^2}{8\pi G} \ . \tag{1.18}$$

At the present time $\rho_{c,0} = 1.05 \times 10^{-5}\, h^2\, GeV\, cm^{-3}$. The name and the meaning of ρ_c will become clear shortly. We also introduce the dimensionless ratio

$$\Omega \equiv \frac{\rho_0}{\rho_c} \tag{1.19}$$

in terms of which the Friedmann equations are written as

$$\frac{k}{R_0^2} = H_0^2\left(\Omega_0 - 1 + \frac{\Lambda}{3H_0^2}\right) \ , \quad q_0 = \frac{1}{2}\Omega_0 - \frac{\Lambda}{3H_0^2} \ . \tag{1.20}$$

In the case of vanishing cosmological constant $\Lambda = 0$, we have

$$q_0 = \frac{1}{2}\Omega_0 \ , \ \ \frac{k}{R_0^2} = H_0^2(\Omega_0 - 1) \tag{1.21}$$

and, therefore

$$\begin{array}{rcl} \rho_0 > \rho_{c,0} & \Rightarrow & k = +1 \\ \rho_0 = \rho_{c,0} & \Rightarrow & k = 0 \\ \rho_0 < \rho_{c,0} & \Rightarrow & k = -1 \ . \end{array} \tag{1.22}$$

Thus, the measurable quantity $\Omega_0 = \rho_0/\rho_{c,0}$ determines the sign of k, *i.e.* whether the present Universe is a hyperbolic or a spherical spacetime. Note that for $\Lambda = 0$, H_0 and q_0 determine the spacetime and the present age completely.

It is often necessary to distinguish different contributions to the density, like the present-day density of pressureless matter Ω_m, that of relativistic particles Ω_r, plus the quantity $\Omega_\Lambda \equiv \Lambda/3H^2$. In addition to these, in models with a variable present-day contribution of the vacuum, one can add a term Ω_v. Thus, in the general case, we have

$$\frac{k}{R_0^2} = H_0^2(\Omega_m + \Omega_r + \Omega_\Lambda + \Omega_v - 1) \ . \tag{1.23}$$

1.5 Simple Cosmological Solutions

1.5.1 Empty de Sitter Universe

In the case of the absence of matter ($\rho = p = 0$) and for $k = 0$, the Einstein-Friedmann equations take the very simple form

$$H^2 = \frac{\Lambda}{3} \tag{1.24}$$

$$q = -\frac{\Lambda}{3H^2} = -1 \ . \tag{1.25}$$

For positive Cosmological Constant $\Lambda > 0$ we have a solution with an exponentially increasing scale factor

$$R(t) = R(t_0) e^{\sqrt{\frac{\Lambda}{3}}\,(t-t_0)} \ . \tag{1.26}$$

This solution describes an expanding Universe *(de Sitter space)* which expands with a constant Hubble parameter and with a constant acceleration parameter. The force that causes the expansion arises from the non-zero cosmological constant. The de Sitter Universe is curved with a constant positive Curvature proportional to Λ.

1.5.2 Vacuum Energy Dominated Universe

In the case that the dominant contribution to the Energy-Momentum Tensor comes from the Vacuum Energy (for example the vacuum expectation value of a Higgs field), the Energy-Momentum Tensor has the form

$$T^{\nu}_{\mu} = -\sigma \delta^{\nu}_{\mu} \ , \tag{1.27}$$

with $\sigma > 0$ a constant. The Equation of State is

$$p = -\rho = -\sigma \tag{1.28}$$

which corresponds to the existence of Negative Pressure. The negative pressure of the vacuum can lead to an accelerated exponential expansion, just as in the previous case of the empty de Sitter space.

For $\Lambda = k = 0$, we obtain the Friedmann-Einstein equations

$$H^2 = \frac{8\pi G}{3}\sigma \tag{1.29}$$

$$q = -\frac{8\pi G\sigma}{3H^2} = -1 \ , \tag{1.30}$$

with the scale factor

$$R(t) = R(t_0)\, e^{(t-t_0)\sqrt{\sigma \frac{8\pi G}{3}}} \ . \tag{1.31}$$

An Exponentially Expanding Vacuum Dominated Universe is a key ingredient of Inflation [10]. The Vacuum Dominated Universe and the Empty de Sitter Universe are physically indistinguishable. This is a consequence of the simple fact that a constant part of the Energy-Momentum Tensor, attributed to matter, is equivallent to a constant of the opposite sign in the left hand side of Einstein's Equations playing the role of a Cosmological Constant, traditionally attributed to geometry.

In a more general case that $p = w\,\rho$, the acceleration parameter is $q = (1+3w)\Omega_v/2$. This shows that for an equation of state parameter

$$w < -\frac{1}{3} \ , \tag{1.32}$$

we are led to accelerated expansion. Current data may indicate that we are at presently undergoing such a phase of accelerated expansion. The vacuum energy seems indeed to be a dominant contributor to the cosmological density budget with $\Omega_v \sim 0.7$, while $\Omega_m \sim 0.3$. Nevertheless, the nature of such a vacuum term is presently uncertain.

1.5.3 Radiation Dominated Universe

The appropriate description of a hot and dense early Universe is that of a gas of relativistic particles in thermodynamic equilibrium. A relativistic gas of temperature T consists of particles with masses $m << T$. Particles with masses $m > T$ are decoupled. The energy density for such a relativistic gas is

$$\rho = \frac{\pi^2}{30}\, Q\, T^4 \ , \tag{1.33}$$

where Q is the number of degrees of freedom of different particle species

$$Q = \sum_B g_B + \frac{7}{8} \sum_F g_F \ , \tag{1.34}$$

where g_B, g_F are the numbers of degrees of freedom for each boson (B) or fermion (F). For example, $Q = g_\gamma = 2$ for photons, as they have two spin states. The pressure of the relativistic gas is given by

$$p = \frac{\pi^2}{90}\, Q\, T^4 = \frac{1}{3}\rho \ . \tag{1.35}$$

As the temperature decreases and crosses the particle mass-thresholds the decoupling particles are subtracted from the effective number of degrees of freedom. Thus, $g_B(T)$, $g_F(T)$ and $Q(T)$ are temperature-dependent.

For a freely expanding gas, the expansion redshifts the wavelength by a factor f as $\lambda \to \lambda' = \lambda f$. The blackbody formula gives

$$\int \frac{d\lambda}{\lambda^5} \frac{1}{e^{2\pi/\lambda T}-1} \propto T^4 \Rightarrow f^4 \int \frac{d\lambda'}{\lambda'^5} \frac{1}{e^{2\pi f/\lambda' T}-1} \propto T^4 \ ,$$

which implies that $\frac{T'}{T} = \frac{\lambda}{\lambda'} = \frac{R}{R'}$. The relation between temperature and the scale factor is,

$$T\,R = const. \tag{1.36}$$

The Friedmann equation, for $\Lambda = k = 0$ gives

$$H^2 = \frac{8\pi G}{3}\rho = \frac{8\pi^3 Q}{90} T^4 \ . \tag{1.37}$$

Note that even if Λ or k were present, they would be irrelevant in early times, when R is small, comparing to the radiation term $\rho \propto R^{-4}$. At late times the situation is reversed and they are more important than the radiation term. From (1.36), (1.37) we obtain

$$\frac{\dot{R}}{R} = -\frac{\dot{T}}{T} = \frac{C^2}{2}\, T^2 \ ,$$

with $C \equiv \left\{\frac{32\pi^3 Q}{90}\right\}^{1/4}$. Taking the initial time $t_0 = 0$ to be a time of infinite temperature $T(0) \to \infty$ and, therefore, vanishing scale factor $R(0) \to 0$, we get

$$R(t) = \overline{C}\,\sqrt{t} \ , \quad T(t) = \frac{C^{-1}}{\sqrt{t}} \ , \quad H = \frac{1}{2t} \ , \tag{1.38}$$

where $\overline{C} = (R(t_1)T(t_1))\,C$, with t_1 any finite time. On the other hand, the deceleration parameter is

$$q = 1 \ . \tag{1.39}$$

The Radiation Dominated Universe is under decelerated expansion.

1.5.4 Matter Dominated Universe

At relatively late times, non-relativistic matter dominates the energy density over radiation. A pressurless gas of non-relativistic particles has the equation of state

$$p = 0 \ . \tag{1.40}$$

The energy density can be written as

$$\rho(T) = \rho(T_0)\left(\frac{T}{T_0}\right)^3 = \hat{\rho}_0\, R^{-3} \ , \tag{1.41}$$

with $\hat{\rho}_0 = \rho(T_0)R_0^3$. The Einstein-Friedmann equations (for $\Lambda = k = 0$) give

$$H^2 = \frac{8\pi G\hat{\rho}_0}{3}\, R^{-3} \tag{1.42}$$

and

$$q = \frac{1}{2} \ . \tag{1.43}$$

The time dependence of the scale factor is

$$R(t) = (6\pi G\hat{\rho}_0)^{1/3} \ t^{2/3} \tag{1.44}$$

and that of the Hubble parameter

$$H = \frac{2}{3\,t} \ . \tag{1.45}$$

Thus, the Matter Dominated Universe with vanishing cosmological constant undergoes a decelerated expansion.

1.5.5 General Equation of State

In certain cosmological settings it is conceivable that matter is not described by a gas of particles, like the ones we considered, but by fields effectively described as a fluid with equation of state

$$p = w\,\rho \ . \tag{1.46}$$

It is not difficult to show that the continuity equation, for arbitrary but constant w, has the solution

$$\rho = \hat{\rho}_0\, R^{-3(1+w)} \tag{1.47}$$

An expanding behaviour corresponds to

$$1 + w > 0 \ . \tag{1.48}$$

Setting again $\Lambda = k = 0$ in the Friedmann equation, we obtain

$$H^2 = C\, R^{-3(1+w)} \ , \quad q = \frac{1}{2}(1+3w) \ . \tag{1.49}$$

The scale factor is solved to be

$$R(t) = \hat{C}(w)\,(t)^{\frac{2}{3(w+1)}} \ , \tag{1.50}$$

with $\hat{C}(w) = (3(1+w)/2)^{2/3(w+1)}\, C^{1/3(w+1)}$.

The above expansion is accelerated provided that

$$1 + 3w < 0 \Rightarrow -1 < w < -\frac{1}{3} \ . \tag{1.51}$$

In the case that the matter is represented by the Energy-Momentum Tensor of a homogeneous scalar field, i.e. that depends only on time, we have

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi) \ , \quad p = \frac{1}{2}\dot{\phi}^2 - V(\phi) \ . \tag{1.52}$$

The resulting "equation of state" $p = w\rho$ involves a ϕ-dependent, and, consequently, time-dependent w, namely

$$w(\phi) = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)} \ . \tag{1.53}$$

1.5.6 The Effects of Curvature

In the expanding solutions for the Early Universe that we considered above, the effects of the curvature term $-k/R^2$ have been neglected. This term becomes important at late times ($R >>$) when the radiation ($\sim R^{-4}$) and matter ($\sim R^{-3}$) terms are smaller. Let us consider the previously described Matter-Dominated Universe inserting the curvature term into the Friedmann Equations. We have

$$H^2 = \frac{C}{R^3} - \frac{k}{R^2} \tag{1.54}$$

$$(2q-1)H^2 = \frac{k}{R^2} \ , \tag{1.55}$$

where $C \equiv 8\pi G\hat{\rho}_0/3$.

- **Open [3], Flat Space ($\mathbf{k=0}$).**

  This case, already studied previously, has

  $$R(t) = \left(\frac{9C}{4}\right)^{1/3} t^{2/3} \ , \ H = \frac{2}{3t} \ , \ q = \frac{1}{2} \ . \tag{1.56}$$

  The energy density satisfies $\rho = \rho_c \equiv 3H^2/8\pi G$ or, equivalently $\Omega = 1$, and it is characterized as critical.

- **Closed, Spherical Space ($\mathbf{k=1}$).**

  In this case we obtain

  $$R(t) = C\sin^2\phi(t) \ , \tag{1.57}$$

  where $\phi(t)$ is the solution of

  $$t = C\left(\phi - \sin\phi\cos\phi\right) \ . \tag{1.58}$$

  It is clear that these equations imply a maximal radius of expansion $R_{max} = C$ reached at time $C\pi/2$. At this time the Hubble parameter becomes zero. Beyond this time the Universe is contracting until the scale factor becomes zero at time $T = C\pi$. Always the deceleration parameter is $q > 1/2$ and $\rho > \rho_c$ (or, equivalently, $\Omega > 1$).

- **Open, Hyperbolic Space ($\mathbf{k=-1}$).**

  In this case we get

  $$R(t) = C\sinh^2\phi(t) \tag{1.59}$$

  $$t = C\left(\sinh\phi\cosh\phi - \phi\right) \ . \tag{1.60}$$

[3] The characterizations *Open, Closed* and *Flat* given below refer to the three-dimensional spatial geometry obtained by setting $t = const.$ in the four-dimensional spacetime manifold.

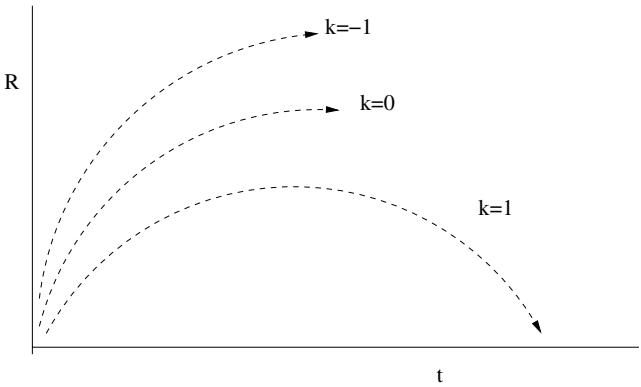

Fig. 1.2. Friedmann Universes beginning with Big Bangs.

The scale factor grows indefinitely. We have $\rho < \rho_c$ (or, equivalently, $\Omega < 1$) and $q < 1/2$. The behaviour of the scale factor in these three cases is shown in Fig. 1.2.

1.5.7 The Effects of a Cosmological Constant

Consider the case of a closed $k = +1$ Universe with a non-zero Cosmological Constant $\Lambda \neq 0$. Assume that matter is present either as radiation or pressureless matter with density $\rho = \hat{\rho}_0 R^{-3} + \hat{\rho}_1 R^{-4}$. The Friedmann equation

$$H^2 = \frac{8\pi G}{3}\rho(R) + \frac{\Lambda}{3} - \frac{1}{R^2} \tag{1.61}$$

shows that the Hubble parameter decreases until it reaches a minimum and then starts increasing again until it reaches an asymptotic value $\Lambda/3$. The scale factor after the Big Bang follows, first radiation dominated and later matter dominated, decelerated expansion, then reaches a plateau at the value $R_0 = 2\pi G\hat{\rho}_0(1 + \sqrt{1 + 2\hat{\rho}_1/3\pi G\hat{\rho}_0^2})$ and finally increases again following an accelerated expansion (Lemaitre Universe).

1.6 The Matter Density in the Universe

From the discussion at the end of the previous chapter it is evident that the fate of the Universe at late times ($p = 0$) is dictated by the Friedmann equation

$$H^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{k}{R^2} \tag{1.62}$$

which can be put into the form

$$\frac{k}{R^2} = H^2 \left(\Omega - 1\right) \ , \tag{1.63}$$

with

$$\Omega = \Omega_m + \Omega_\Lambda \ . \tag{1.64}$$

having denoted $\Omega_\Lambda = \Lambda/3H^2$. The cosmological constant contribution stands for a general effective vacuum contribution which could have a, for the moment unknown, dynamical origin. For $\Omega > 1$, the Universe is closed and, in the absence of a cosmological constant, the expansion would change into contraction. This is not necessarily true in the presence of a non-zero cosmological constant. In the case $\Omega < 1$ the Universe is open and the expansion continues forever. This is true also for the critical case $\Omega = 1$.

A lower bound for Ω is supplied by the observed Visible Matter

$$\Omega \ > \ 0.03 \tag{1.65}$$

Arguments based on Primordial Nucleosynthesis support this value. We can denote $\Omega_{vm} \sim 0.03$. Thus, it seems that most of the mass in the Universe is in an unknown non-baryonic form. This matter is called *Dark Matter*. In general, such matter can only be observed indirectly through its gravitation. Doing that, one arrives at an estimate $\Omega_{dm} \sim 0.3$.

What is the origin of the remaining contribution to Ω? Since it cannot be attributed to matter, visible or dark, it is represented with an effective vacuum term and has been given the name *Dark Energy* . For theoretical reasons (*i.e.* Inflation), the value $\Omega = 1$ is particularly attractive. In that case, the Dark Energy contribution is $\Omega_{de} \sim 0.7$. This estimate is supported by current data[11][12]. In particular, current data support the value $\Omega_\Lambda = \Lambda/3H_0^2 \sim 0.7$ or $\Lambda \sim O(10^{-56})\, cm^{-2}$. The estimated small cosmological constant is sometimes represented by a scale $\overline{\Lambda}^4 = \Lambda/M_P^2 \sim (10^{-3}\, eV)^4$.

Thus, in the case of critical density, the various contributions are

$$\Omega_{vm} \sim 0.03 \ , \ \ \Omega_{dm} \sim 0.27 \ , \ \ \Omega_{de} \sim 0.7 \ . \tag{1.66}$$

Although it seems unavoidable, it is surprising that at least 90% of the matter in the Universe is of unknown form.

1.7 The Standard Cosmological Model

The present Universe seems to be described by a Matter Dominated Friedmann model ($p = 0$) with a possible vacuum contribution in order to account for acceleration. For any time smaller than 10^4 years from the beginning the dominant part of the energy density was relativistic matter (electromagnetic radiation, neutrinos, etc). Thus, the Universe corresponded to a Radiation Dominated Friedmann model ($p = \rho/3$). The relativistic gas description is valid down to times $t \sim 10^{-10}\, sec$, corresponding to energies of the order

of $100\,GeV$. For smaller times, or larger energies, the description depends on the assumed theoretical framework beyond [13] the Standard Model of Particle Physics. If a Quantum Field Theory description of Particle Physics remains valid up to energies of the order of $10^{18}\,GeV$, then, the relativistic gas description of the Early Universe can be extrapolated down to times of the order of $10^{-42}\,sec$.

1.7.1 Thermal History

During the Radiation Dominated epoch the Friedmann equation is $H^2 \sim 8\pi G\rho/3$, since the curvature term is irrelevant at small values of the scale factor. Thus, the energy density has the critical value

$$\rho \sim \frac{3H^2}{8\pi G} = \rho_c \ . \tag{1.67}$$

The solution (1.38), for each interval of constant effective number of degrees of freedom $Q(T)$, gives

$$T(t) = \frac{C^{-1}}{\sqrt{t}} \quad , \quad H = \frac{1}{2t} \quad , \quad \rho = \frac{3}{32\pi G t^2} \ , \tag{1.68}$$

with

$$C \equiv \left\{ \frac{16\pi^3 Q G}{45} \right\}^{1/4} \ . \tag{1.69}$$

The value of $Q(T)$ at any given temperature depends on the Particle Physics model valid in the given temperature/energy range. In the following table we give the values of $Q(T)$ up to temperatures of $O(100\,GeV)$ in the framework of the $SU(3)_C \times SU(2)_L \times U(1)_Y$ Standard Model.

We assume that the relativistic gas is in a state of thermodynamic equilibrium. This is a reasonable assumption since the rate of expansion is much smaller than the rate of interactions that can restore the equilibrium. The rate of these interactions is given by the cross section $\sigma \propto T^{-2} \propto t$ times the particle number density $n \propto T^3 \propto t^{-3/2}$. Thus, the rate of reactions goes as $\sigma n \sim t^{-1/2}$, while the rate of expansion goes as $H = 1/2t$, guaranteeing that $\sigma n > H$ as the Universe expands and cools down.

Let us now attempt a bottom-up description of the expansion starting from the relatively late time of $t \sim 1\,sec$, equivalent to $T \sim 1\,MeV \sim 10^{10}\,K$ and move backwards in time. Below $1\,MeV$, the plasma consists of photons and neutrinos. At temperatures $T \sim 1\,MeV > m_e$ electron-positron pairs should appear thanks to the process

$$\gamma + \gamma \rightleftharpoons e^- + e^+ \ .$$

Temperature	New Particle	4Q(T)
$T < m_e$	$\gamma,\ \nu$	29
$m_e < T < m_\mu$	$e^{\pm}$	43
$m_\mu < T < m_\pi$	$\mu^{\pm}$	57
$m_\pi < T < T_c$	π	69
$T_c < T < m_s$	$\pi,\ u, \overline{u}, d, \overline{d},\ g$	205
$m_s < T < m_c$	$s, \overline{s}$	247
$m_c < T < m_\tau$	$c, \overline{c}$	289
$m_\tau < T < m_b$	$\tau^{\pm}$	303
$m_b < T < m_{W,Z}$	$b, \overline{b}$	345
$m_{W,Z} < T < m_H$	$W^{\pm}, Z$	381
$m_H < T < m_t$	H^0	385
$m_t < T$	$t, \overline{t}$	427

Protons and neutrons play no role in the energy density. Their number is of the order of 10^{-9} of the number of light particles (γ, ν, e). At smaller times $t \sim 10^{-3} - 10^{-4}\, sec$, muons and π-mesons participate in the plasma. Near this range lies the so-called *deconfinement temperature* T_c at which a phase transition between the hadron phase and the quark-gluon phase occurs. Above T_c, at times $t < 10^{-4}\, sec$, gluons and free quarks, u and d and, later, s, join the plasma. At higher temperatures charmed quarks, τ-leptons and bottom quarks appear. At times $t \sim 10^{-10}\, sec$, corresponding to temperatures $T \sim 100\, GeV$, the $W^{\pm}$ and Z bosons of Weak Interactions become abundant. At even higher temperatures, the Higgs boson and top-quark appear. At these temperatures, the full Electroweak symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$ is restored.

1.7.2 Nucleosynthesis

The period from 1 *sec* to 200 *sec* from the Big Bang plays an important role in the history of the Universe [14]. During this period light nuclei have been produced and the usual matter started to appear. This is the time that the abundances of light nuclei were fixed, namely He^4 (.25), H^2 (3×10^{-5}), He^3 (2×10^{-5}), Li^7 (10^{-9}), etc. Heavier nuclei were produced much later in stars. It is remarkable that the primordial Helium abundance of 25% has been modified only by a few per cent during the billions of years of converting hydrogen into helium in stars. Yet only 200 *sec* of the early radiation era sufficed to convert hydrogen into almost all of the helium abundance. The amount of helium produced can be estimated in the following way. For $t < 1\, sec$, or $T > 1\, MeV$, protons and neutrons move freely in the primordial plasma. Their relative number can be expressed through the Boltzmann formula

$$\frac{N_n}{N_p} = e^{-\frac{(m_n - m_p)}{T}} \sim e^{-(1.3\, MeV)/T}\ . \tag{1.70}$$

The equilibrium is maintained by the processes $\overline{\nu} + p \rightleftharpoons e + n$, $n + \nu \rightleftharpoons p + e$, etc. At a temperature $T_f \sim 0.7\, MeV$, these reactions become too slow and the ratio freezes out at the value $(N_n/N_p)_f \sim 0.16$. Thus, there is one neutron to about $5 - 6$ protons. Free neutron decay ($\tau \sim 15\, min$) is too slow to change that. Protons and neutrons collide together to form *deuterium* nuclei or deuterons through the process

$$p + n \rightarrow H^2 + \gamma \ . \tag{1.71}$$

The deuterons break apart through the inverse process giving back to the plasma protons and neutrons. Only beyond $t \sim 100\, sec$ the temperaure drops to a point that it is energetically possible for deuterons to be stable. By this time that protons and neutrons have been able to combine, the abundance of neutrons has decreased to about two neutrons in every 14 protons. Out of 16 nucleons we get two deuterons and 12 protons. The, now stable, deuterons can combine and produce a He^4 nucleus

$$H^2 + H^2 \leftrightarrow He^4 + \gamma \ . \tag{1.72}$$

Actually, one has to consider all the two-body processes, like $p + H^2 \leftrightarrow He^3 + \gamma$, $n + He^3 \leftrightarrow He^4 + \gamma$, etc. The whole process is over in roughly $200\, sec$, and in that time $25\,\%$ of matter is converted into helium (four out of sixteen nucleons form a heliun nucleus) and the remainder consists predominantly of protons. Slight amounts of deuterium, He^3 and Li are also produced.

1.8 Problems of Standard Cosmology

The Standard Cosmological Model described in the previous section incorporates GR, CMB, the Hubble law and the light nuclei abundance. Needless to say that its successes are compatible and intimately connected with the Standard Model of fundamental Particle Physics. Nevertheless, it faces a number of serious problems having to do mostly with the lack of understanding of initial conditions. Modifications are needed, which, however, should leave its successes intact.

1.8.1 The Horizon Problem

The maximum size of a region in which causal relations can be established is given by the horizon

$$r_H(t) = R(t) \Delta\chi = R(t) \int_0^t \frac{dt'}{R(t')} \ . \tag{1.73}$$

During the Radiation Dominated phase, $R(t) \sim t^{1/2}$ and $r_H(t) = 2t$. For $t \rightarrow 0$, r_H shrinks much faster than $R(t)$. Thus, at every epoch, most of

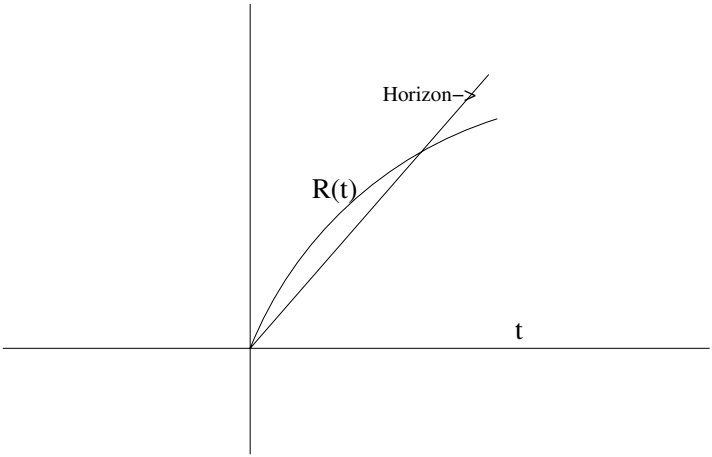

Fig. 1.3. Horizon growth in the Standard Cosmological Model.

the regions within a typical dimension $R(t)$ are causally unrelated despite the extreme isotropy of the Standard Cosmological Model established by the CMB data, as is shown in Fig. 1.3. Radiation and matter were in thermal equilibrium until the time t_R of hydrogen recombination after which the Universe became transparent to radiation. The present isotropy of the CMB implies a similar isotropy at time t_R. Nevertheless, what we see today is the same radiation-temperature from regions that had not established causal contact at the epoch t_R. The coordinate distance between our epoch t_0 and t_R (we take $r = 0$ to be our position), is

$$\Delta\chi(t_0, t_R) = \int_{t_R}^{t_0} \frac{dt'}{R(t')} \ . \tag{1.74}$$

Since the horizon at t_R was

$$\Delta\chi_H = \int_{t_P}^{t_R} \frac{dt'}{R(t')} \tag{1.75}$$

the number of horizon lengths contained in the distance $2\Delta\chi(t_0, t_R)$ between two opposite directions in the sky will be

$$\mathcal{N} = \int_{t_R}^{t_0} \frac{dt'}{R(t')} \Big/ \int_{t_P}^{t_R} \frac{dt''}{R(t'')} \ . \tag{1.76}$$

Applying this formula in the case of the Radiation Dominated expansion, we obtain a very large number. In contrast, an altered expansion law for the scale factor, as in the case of inflation, could change that dramatically. For $R \sim e^{Ht}$, we obtain for $\mathcal{N}$ a very small number.

This is the so-called horizon problem of the Standard Cosmological Model. This problem is solved and the observed homogeneity and isotropy is explained in the framework of Inflation which predicts a period of exponential growth for the Universe.

1.8.2 The Coincidence Puzzle and the Flatness Problem

The Friedmann equation for the present epoch has the form

$$H_0^2 = \frac{8\pi G}{3}\rho_0 - \frac{k}{R_0^2} + \frac{\Lambda}{3} \ . \tag{1.77}$$

Observations indicate that all three terms of the right hand side can be roughly of the same order of magnitude

$$\frac{8\pi G}{3H_0^2}\rho_0 \sim \frac{|k|}{R_0^2 H_0^2} \sim \frac{\Lambda}{3H_0^2} \sim O(1) \ . \tag{1.78}$$

At very early times these terms are of greatly different magnitudes. Since $\rho \propto R^{-4}$, this term dominates over the others which become relevant at very late times. This very near balance of the three different terms seems coincidentally very beneficial for our existence and for the existence of the world around us. For instance, a balance for the first two terms only, for a $k = +1$ model would be disastrous. In a few Planck-times[4] the Universe would collapse. On the other hand, if we have a balance of these two terms in a $k = -1$ Universe, the resulting expansion would be so rapid that at the present epoch Ω would be catastrophically small. The coincidence of the magnitudes of the different terms is often refereed to as the Coincidence Puzzle.

The balance between the different terms can be best formulated in terms of the Entropy of the Universe. During the Radiation Dominated epoch the entropy density s and the *entropy* S of a comoving volume R^3 are given by

$$s = \frac{2\pi^2 Q}{45}T^3 \ , \quad S \equiv s\,R^3 = \frac{2\pi^2 Q}{45}(RT)^3 \ . \tag{1.79}$$

Estimating the present time entropy density from the background of photons and neutrinos as $s_0 \sim n_\gamma \sim 10^3\, cm^{-3}$, we obtain for the entropy the huge number

$$S \sim 10^{87} \ . \tag{1.80}$$

This number is an initial condition of the Standard Cosmological Model. The fact that there are so much more photons than baryons is something determined at the beginning.

[4] The characteristic scale of gravitation, Newton's gravitational constant G defines a characteristic mass, the *Planck mass* $M_P \sim 10^{18}\, GeV$, a characteristic length, the *Planck length*, and a characteristic time, the *Planck time*.

Rewriting the Friedman equation in terms of temperature and entropy, we obtain

$$\left(\frac{\dot{T}}{T}\right)^2 = \left(\frac{4\pi^3 QG}{45}\right) T^4 - \frac{k}{S^{2/3}} \left(\frac{2\pi^2 Q}{45}\right)^{2/3} . \tag{1.81}$$

It is clear that the curvature term at high temperature is negligible since S is a large number. The Friedmann equation can also be written as (Λ is negligible at early times)

$$H^2 R^2 |1-\Omega| = |k| \;\; \Rightarrow \frac{|1-\Omega|}{\Omega} = \frac{1}{S^{2/3}} \left(\frac{45}{2Q\pi^2}\right)^{1/3} \frac{1}{2\pi G\, T^2} . \tag{1.82}$$

Inserting numbers, one finds

$$\frac{|1-\Omega|}{\Omega} \sim 10^{-59} \left(\frac{M_P}{T}\right)^2 . \tag{1.83}$$

This shows in a dramatic way that Ω must have been terribly close to 1 at early epochs. For instance

$$T = 1\, MeV \;\; \rightarrow \;\; \frac{|1-\Omega|}{\Omega} \leq 10^{-15}$$

$$T = 10^{14}\, GeV \;\; \rightarrow \;\; \frac{|1-\Omega|}{\Omega} \leq 10^{-49} .$$

This feature of the Standard Cosmological Model, that Ω is close to 1 at all times, is called the Flatness Problem or, sometimes, the Entropy Problem, referring to the large value of the entropy. It is not a problem in an ordinary sense. It relates however the specific properties of our present Universe to rather special initial data, like the very large value of the entropy, or having $\Omega \sim 1$ at early times. A theory of the Early Universe that could start with S of order 1 and arrive, via physical processes, to the present number, would be considered an improvement because it would not require very specific initial data.

1.9 Phase Transitions in the Early Universe

The $SU(3)_C \times SU(2)_L \times U(1)_Y$ Standard Model of Strong and Electroweak interactions incorporates the concept of Spontaneous Symmetry Breaking according to which, although, the Laws of Nature are symmetric under a given (local) gauge symmetry, the vacuum state is not. As a result, the vacuum expectation values of certain operators in the theory violate the symmetry. The way this is achieved in the Standard Model is through the vacuum expectation value of a scalar (Higgs) field that is an $SU(2)_L$-doublet and carries weak hypercharge. In the broken $SU(3)_C \times U(1)_{em}$ vacuum three out of the

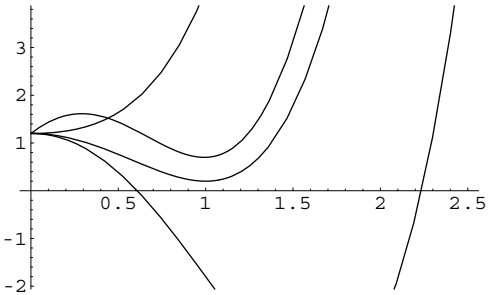

Fig. 1.4. Finite Temperature Effective Potential.

four gauge bosons ($W^{\pm}$, Z^0) of $SU(2)_L \times U(1)_Y$ obtain a mass, while the fourth (photon) remains massless, corresponding to the intact electromagnetic $U(1)_{em}$ gauge interaction.

In the Early Universe matter corresponds to a system in thermodynamic equilibrium with a heat bath. The thermodynamics of this system is described by the Hamiltonian of the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge field theory. The vacuum energy of the system is determined by the minimization of the Free Energy, roughly corresponding to the so-called Effective Potential, which depends on the temperature. At very high temperatures, the global vacuum state is the symmetric one, in contrast to low temperatures, where the global vacuum is the broken one. As the Universe cools down during the Radiation-Dominated epoch it makes a transition from the high temperature symmetric phase to the broken low temperature phase, or it undergoes a *phase transition*. This is shown in Fig. 1.4 where the effective potential at finite temperature is plotted in terms of the Higgs field vev. This behaviour is in agreement with what happens in certain condensed matter systems. For example, a ferromagnet, when heated loses its magnetism, while at zero temperature it is characterized by a non-vanishing magnetization that breaks rotational symmetry. A more appropriate analogue is that of the phase transition from water to ice. Normally, the water-ice phase transition occurs at the freezing point of 0^0C. Nevertheless, undisturbed pure water supercools to a temperature lower than the freezing point before it transforms into ice. When the transition finally occurs, after the supercooling period, the Universe is reheated due to the release of the false vacuum latent heat. Depending on the details of the theory, symmetry breaking will occur via a first order phase transition in which the field tunnels through a potential barrier, or via a second order phase transition in which the field evolves smoothly from one state to the other.

1.10 Inflation

In a phase transition of the type discussed in the previous section where the Universe spends a lot of time in the false vacuum with $<\phi>\approx 0$, the energy density ($\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi)$) can be dominated by the vacuum contribution

$$\rho \approx V(0) \ . \tag{1.84}$$

A scalar potential that could lead to such a behaviour is shown in Fig. 1.5.

During the period of the evolution in the false vacuum state the equation of state will be that of negative pressure

$$p = -\rho = V(0) \ , \tag{1.85}$$

and the Friedmann equation predicts a constant Hubble parameter and an exponentially growing scale factor

$$H^2 \approx \frac{8\pi G}{3}\rho \Rightarrow R(t) = R(t_i)\, e^{H(t-t_i)} \ . \tag{1.86}$$

Assuming that the transition to the true vacuum $<\phi>\neq 0$ is completed, the latent energy stored in the false vacuum will be released and the Universe will be heated up to a temperature comparable to the initial temperature. The product RT would increase during this period proportionally to the scale factor

$$R_f T_f \approx R_f T_i = \left(\frac{R_f}{R_i}\right) R_i T_i = e^{\eta} R_i T_i \ , \tag{1.87}$$

where

$$e^{\eta} \equiv \frac{R_f}{R_i} \ .$$

Consequently the entropy would increase by a factor $e^{3\eta}$. Thus, for a value of this parameter $\eta \sim 60-70$, the presently huge magnitude of the entropy 10^{87} could have arisen from an initial entropy magnitude of $O(1)$. This essentially

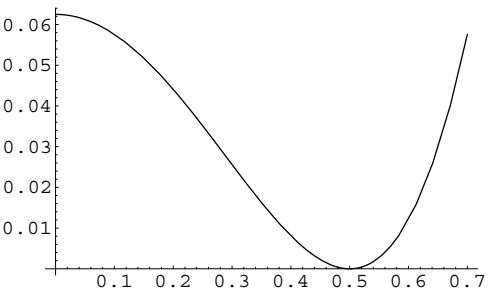

Fig. 1.5. Slow evolution in the "false vacuum".

would explain the entropy puzzle. The, equivalent, flatness problem is seen to be explained in a straightforward fashion by considering the Friedmann equation written as $k = R^2H^2(\Omega - 1)$. Note that Λ plays no role in the Early Universe but can always be included in Ω as $\Omega_\Lambda = \Lambda/3H^2$. Thus, we obtain

$$k = R_f^2 H^2(\Omega_f - 1) = R_i^2 H^2(\Omega_i - 1) \ \Rightarrow \Omega_f = 1 + e^{-2\eta}(\Omega_i - 1) \ . \quad (1.88)$$

For the choice of η that explains away entropy we get $\Omega_f = 1 + 10^{-58}(\Omega_i - 1)$ which is ridiculously close to 1, whatever the initial value Ω_i is. Note that, as we remarked when discussing the horizon problem, exponential expansion is sufficient to explain this puzzle as well.

This is the main idea behind the Inflationary Scenario proposed in 1982 by A. Guth [10]. Nevertheless, this so-called *Old Inflation* does not represent a viable scenario. The way the transition proceeds is through the creation of bubbles of the true (broken) vacuum in the inflating background of the false (unbroken) vacuum. The rate at which these bubbles coalesce cannot keep up pace with the expansion of the surrounding Universe. Concentrations of bubbles form which are finally dominated by one bubble. In this way a very inhomogeneous picture appears.

Another scenario [15] (*New Inflation*), which stems from the same basic idea but follows a modified line of events, was proposed by A. Linde and, independently, P. Steindhard. In the New Inflationary Scenario the whole observable Universe evolves out of a single fluctuation region. Two ingredients of the original inflationary scenario were abandoned, namely, the assumption that the Universe spends a long time in the supercooled $< \phi > \approx 0$ phase and that the phase transition is completed through bubble nucleation. Instead, inflation occurs during the time of the slow growth of the so-called *inflaton* from its initial value to its equilibrium value. This time must be much longer than H^{-1}, something that could be achieved with a suitably flat potential near the origin. In this scenario the Universe is heated up after inflation not because of bubble wall collisions but by particle creation from the oscillations of the classical inflaton. Bubbles, if formed, will be separated by such distances that forbid any causal interaction and the observable part of the Universe will not be in danger from inhomogeneities. Implementation of the new inflationary scenario [16] in a realistic Particle Physics theory has faced a number of problems that have forced cosmologists to abandon the framework of high temperature phase transitions and formulate inflation assuming that the Universe is filled with a chaotically distributed inflaton (*Chaotic Inflation*) . The requirements on the classical theory for the dilaton that can lead to inflation ($q < 0$ or, equivalently, $\rho + 3p < 0$) are

$$H^2 = \frac{8\pi G}{3}\left(V(\phi) + \frac{1}{2}\dot{\phi}^2\right) \approx \frac{8\pi G}{3}V(\phi)$$

$$\ddot{\phi} + 3\dot{\phi}H + V'(\phi) = 0 \approx 3\dot{\phi}H + V'(\phi) \ .$$

This is the *slow-roll approximation* quantified in terms of the parameters

$$\epsilon(\phi) \equiv \frac{1}{16\pi G}\left(\frac{V'}{V}\right)^2 << 1 \quad , \quad |\eta(\phi)| \equiv \frac{1}{8\pi G}\left|\frac{V''}{V}\right| << 1 \ .$$

Note that the last two conditions (slow-roll conditions) are necessary but not sufficient. The slow-roll approximation requires $\dot{\phi}$ to satisfy $3H\dot{\phi} \sim -V'$. The amount of inflation is $N = \ln \frac{R(t_{end})}{R(t)} \sim 8\pi G \int_{\phi_{end}}^{\phi} d\phi \, \frac{V}{V'}$ ..

The historical motivation for inflation arose mostly on the questions concerning the initial conditions required for a Hot Big Bang phase. Nevertheless, in the modern view, by far the most important property of inflation is that it can generate irregularities in the Universe which may lead to the formation of structure.

1.11 The Baryon Asymmetry in the Universe

Relativistic Quantum Theory predicts that for each elementary particle state there is another, in general different, state characterizing its antiparticle. Antiparticles have the same spacetime properties (mass, spin) of particles and opposite electric charge. Global quantum numbers, like Baryon and Lepton Number, are also of opposite sign. Thus, the antiparticle of the electron e is the positron e^+, with positive charge, while the antiparticle of the neutrino is a distinct neutral particle the antineutrino. The two particle wave functions are related by a symmetry operator

$$\psi_e(x) \ \Rightarrow \ \mathcal{C}\left\{\psi_e(x)\right\} = \overline{\psi}_{e^+}(x) \ .$$

The term *Antimatter* refers to the collection of antiparticles. The Universe is almost exclusively filled with matter while existing antimatter is of secondary origin resulting from relativistic collisions of matter. This is true not only of the antiparticles produced in accelerators but also of the antiparticles encountered in Cosmic Rays. The evidence against primary forms of antimatter in the Universe is quite strong. The observed number of Baryons in the Universe, over the number of photons, is

$$\eta = \frac{N_B}{N_\gamma} \sim 10^{-10} \ . \tag{1.89}$$

As first pointed out by A. D. Sakharov [17], the explanation of this asymmetry of matter (Baryon Asymmetry) requires interactions that violate Baryon Number $\mathcal{B}$, the particle-antiparticle symmetry $\mathcal{C}$ (Charge Conjugation) and Parity $\mathcal{P}$. In addition, there should be a departure from thermodynamic equilibrium. $\mathcal{P}$ and $\mathcal{CP}$ non-conservation is a well established fact of the Standard Model. Baryon violation can be found either in Grand Unified Theories (GUTs) or in the non-perturbative sector of the Standard Model. The last requirement can be realized in the expanding Universe where the various interactions come in and out of equilibrium. The oldest proposed scenario is

that of the out of equilibrium decays of superheavy Higgs bosons of GUTs [18]. Another scenario uses the non-perturbative effects (sphalerons) of the Standard Model and generates the Baryon Asymmetry at the Electroweak scale [19]. Since $\mathcal{B}+\mathcal{L}$ is conserved by these modes, it is possible to generate first a Lepton Asymmetry by the out of equilibrium decay of a superheavy right-handed neutrino [20] (Leptogenesis).

Acknowledgements

The author wishes to thank the organizing committee of the 2nd Aegean Summer School on the Early Universe for inviting him and giving him the opportunity to present the above introduction to the Cosmology of the Early Universe. The author also acknowledges the financial support of the EU RTN *"Supersymmetry and the Early Universe"*, contract No. HPRN-CT-2000-00152.

References

1. S. Weinberg, *Gravitation and Cosmology*, J. Wiley and Sons (1972);
   P. J. E. Peebles, *Principles of Physical Cosmology*, Princeton University Press (1993);
   G. Börner, *The Early Universe: Facts and Fiction*, Springer-Verlag (1988);
   E. W. Kolb and M. S. Turner, *The Early Universe*, Addison-Wesley (1990).
2. C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation*, W. H. Freeman & Co. (1973);
   L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, Pergamon Press, 4th edition (1979);
   R. M. Wald, *General Relativity*, Univ. of Chicago Press (1984);
   W. Rindler *Essential Relativity*, Van Nostrand Reinhold (1969).
3. E. Hubble and M. L. Humason, Ap. J. **74**, 43 (1931).
4. R. A. Alpher, R. C. Herman and G. Gamow, Phys. Rev. **73**, 803 (1948).
5. A. A. Penzias and R. W. Wilson, Ap. J. **142**, 419 (1965).
6. J. B. Mather et al., Ap. J. **420**, 439 (1994);
   D. J. Fixen et al., Ap. J. **420**, 445 (1994);
   D. J. Fixen et al., Ap. J. **473**, 576 (1996);
   J. B. Mather et al., Ap. J. **512**, 511 (1999).
7. M. R. Nolta et al., astroph/0305097 (March 2003);
   H. V. Peiris et al., astro-ph/0302225 (Feb. 2003).
8. A. Friedmann, Zeitschrift für Physik, **10**, 377 (1922) and **21**, 326 (1924). Both translated in *Cosmological Constants*, edited by J. Bernstein and G. Feinberg, Columbia Univ. Press (1986).
9. F. Hoyle, *Monthly Notices of the Royal Astronomical Society*, **108**, 372 (1948).
10. A. H. Guth, Phys. Rev. **D 23**, 347 (1981).
11. A. G. Riess et al., PASP, **116**, 1009 (1998);
    S. Perlmutter et al., Ap. J. **517**, 565 (1999).

12. D. N. Spergel et al., Ap. J. Suppl. , **148**, 175 (2003).
13. A. D. Linde, Rep. Prog. Phys. **42**, 389 (1979);
K. A. Olive, Phys. Rep. **190**, 3345 (1990);
C. Vayonakis, Surveys in High Energy Physics **5**, 87 (1986).
14. K. A. Olive, G. Steigman and T. P. Walker, Phys. Rep. **333**, 389 (2000) and references therein.
15. A. D. Linde, Phys. Lett. **108 B**, 389 (1982);
A. Albrecht and P. J. Steindhardt Phys. Rev. Lett. **48**, 1220 (1982).
16. A. Linde *Particle Physics and Inflationary Cosmology*, Harwood (1990);
A. R. Liddle and D. H. Lyth, *Cosmological Inflation and Large Scale Structure*, Cambridge University Press (2000).
17. A. D. Sakharov, JETP Lett. **5**, 24 (1967).
18. S. Weinberg, Phys. Rev. Lett. **42**, 850 (1979).
19. V. Kuzmin, V. Rubakov and M. Shaposhnikov Phys. Lett. **B 155**, 36 (1985).
20. M. Fukugita and T. Yanagida, Phys. Lett. **B 174**, 45 (1986).

2 Cosmological Perturbation Theory

Ruth Durrer

Université de Genève, Département de Physique Théorique, 24 Quai E. Ansermet, 1211 Genève, Switzerland

Abstract. This is a review on cosmological perturbation theory. After an introduction, it presents the problem of gauge transformation. Gauge invariant variables are introduced and the Einstein and conservation equations are written in terms of these variables. Some examples, especially perfect fluids and scalar fields are presented in detail. The generation of perturbations during inflation is studied. Lightlike geodesics and their relevance for CMB anisotropies are briefly discussed. Perturbation theory in braneworlds is also introduced.

2.1 Introduction

The idea that the large scale structure of our Universe might have grown out of small initial fluctuations via gravitational instability goes back to Newton (letter to Bentley, 1692 [1]).

The first relativistic treatment of linear perturbations in a Friedmann-Lemaître universe was given by Lifshitz (1946) [2]. There he found that the gravitational potential cannot grow within linear perturbation theory and he concluded that galaxies have not formed by gravitational instability.

Today we know that it is sufficient that matter density fluctuations can grow. Nevertheless, considerable initial fluctuations with amplitudes of the order of 10^{-5} are needed in order to reproduce the cosmic structures observed today. These are much larger than typical statistical fluctuations on scales of galaxies and we have to propose a mechanism to generate them. Furthermore, the measurements of anisotropies in the cosmic microwave background show that the amplitude of fluctuations is constant over a wide range of scales, the spectrum is scale independent. As we shall see, standard inflation generically produces such a spectrum of nearly scale invariant fluctuations.

In this review I present gauge invariant cosmological perturbation theory. I shall start by defining gauge invariant perturbation variables. Then I present the basic perturbation equations. As examples for the matter equations we shall consider perfect fluids and scalar fields. Then we briefly discuss lightlike geodesics and CMB anisotropies (this section will be very brief since it is complemented by the review on CMB anisotropies by A. Challinor). Finally, I shall make some brief comments on perturbation theory for braneworlds, a topic which is still wide open in my opinion.

R. Durrer, Cosmological Perturbation Theory, Lect. Notes Phys. **653**, 31–69 (2005)
http://www.springerlink.com/

2.2 The Background

I shall not come back to the homogeneous universe which has been discussed in depth by K. Tamvakis. I just specify our notation which is as follows:

A Friedmann-Lemaître universe is a homogeneous and isotropic solution of Einstein's equations. The hyper-surfaces of constant time are homogeneous and isotropic, *i.e.* spaces of constant curvature with metric $a^2(\eta)\gamma_{ij}dx^i dx^j$, where γ_{ij} is the metric of a space with constant curvature κ. This metric can be expressed in the form

$$\gamma_{ij}dx^i dx^j = dr^2 + \chi^2(r)\left(d\vartheta^2 + sin^2\vartheta d\varphi^2\right) \tag{2.1}$$

where $\chi(r)$ was defined in (1.9) and we have rescaled $a(\eta)$ such that $\kappa = \pm 1$ or 0. (With this normalization the scale factor a has the dimension of a length and η and r are dimensionless for $\kappa \neq 0$.) The four-dimensional metric is then of the form

$$g_{\mu\nu}dx^\mu dx^\nu = -a^2(\eta)d\eta^2 + a^2(\eta)\gamma_{ij}dx^i dx^j . \tag{2.2}$$

Here η is called the *conformal time*. The *physical* or *cosmological time* is given by $dt = ad\eta$.

Einstein's equations reduce to ordinary differential equations for the function $a(\eta)$ (with $\dot{} \equiv d/d\eta$):

$$\left(\frac{\dot{a}}{a}\right)^2 + \kappa = \mathcal{H}^2 + \kappa = \frac{8\pi G}{3}a^2\rho + \frac{1}{3}\Lambda a^2 \tag{2.3}$$

$$\left(\frac{\dot{a}}{a}\right)^{\cdot} = \dot{\mathcal{H}} = -\frac{4\pi G}{3}a^2\left(\rho + 3p\right) + \frac{1}{3}\Lambda a^2 = \left(\frac{\ddot{a}}{a}\right) - \left(\frac{\dot{a}}{a}\right)^2 , \tag{2.4}$$

where $\rho = -T^0_0$, $p = T^i_i$ (no sum!) and all other components of the energy momentum tensor have to vanish by the requirement of isotropy and homogeneity. Λ is the cosmological constant. We have introduced $\mathcal{H} = \dot{a}/a$. The Hubble parameter is defined by

$$H = \frac{da/dt}{a} = \frac{\dot{a}}{a^2} = \mathcal{H}/a \ .$$

Energy momentum "conservation" (which is also a consequence of (2.3) and (2.4) due to the contracted Bianchi identity) reads

$$\dot{\rho} = -3\left(\frac{\dot{a}}{a}\right)(\rho + p) = -3(1+w)\mathcal{H}, \tag{2.5}$$

where $w \equiv p/\rho$. Later we will also use $c_s^2 \equiv \dot{p}/\dot{\rho}$. From the definition of w and ρ together with (2.5) one finds

$$\dot{w} = 3(1+w)(w - c_s^2)\mathcal{H} \ . \tag{2.6}$$

From the Friedmann equations one easily concludes that for $\kappa = \Lambda = 0$ and $w =$const. the scale factor behaves like a power law,

$$a \propto \eta^{\frac{2}{1+3w}} \propto t^{\frac{2}{3(1+w)}} \ . \tag{2.7}$$

Important examples are

$$a \propto \eta^q \quad \text{with} \begin{cases} q = 2 & \text{for dust} & w = 0 \\ q = 1 & \text{for radiation} & w = 1/3 \\ q = -1 & \text{for inflation (or a cosm. const.)} & w = -1 \end{cases} \tag{2.8}$$

We also define

$$\begin{aligned} \Omega_\rho &= \left(\frac{8\pi G \rho a^2}{3\mathcal{H}^2} \right)_{\eta=\eta_0} \\ \Omega_\Lambda &= \left. \frac{\Lambda a^2}{3\mathcal{H}^2} \right|_{\eta=\eta_0} \\ \Omega_\kappa &= \left. \frac{-\kappa}{\mathcal{H}^2} \right|_{\eta=\eta_0} , \end{aligned} \tag{2.9}$$

where the index $_0$ indicates the value of a given variable today. Friedmann's equation (2.3) then requires

$$1 = \Omega_\rho + \Omega_\Lambda + \Omega_\kappa . \tag{2.10}$$

One often also uses $\Omega = \Omega_\rho + \Omega_\Lambda = 1 - \Omega_\kappa$.

2.3 Gauge Invariant Perturbation Variables

The observed Universe is not perfectly homogeneous and isotropic. Matter is arranged in galaxies and clusters of galaxies and there are large voids in the distribution of galaxies. Let us assume, however, that these inhomogeneities lead only to small variations of the geometry which we shall treat in first order perturbation theory. For this we define the perturbed geometry by

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \varepsilon a^2 h_{\mu\nu} \ , \tag{2.11}$$

$\bar{g}_{\mu\nu}$ being the unperturbed Friedmann metric. We conventionally set (absorbing the "smallness" parameter ε into $h_{\mu\nu}$)

$$\begin{aligned} &g_{\mu\nu} = \bar{g}_{\mu\nu} + a^2 h_{\mu\nu}, \quad \bar{g}_{00} = -a^2, \quad \bar{g}_{ij} = a^2 \gamma_{ij} \quad |h_{\mu\nu}| \ll 1 \\ &T^\mu_\nu = \overline{T}^\mu_{\ \nu} + \theta^\mu_\nu, \qquad \overline{T}^0_{\ 0} = -\bar{\rho}, \quad \overline{T}^i_{\ j} = \bar{p}\delta^i_j \qquad |\theta^\mu_\nu|/\bar{\rho} \ll 1. \end{aligned} \tag{2.12}$$

2.3.1 Gauge Transformation, Gauge Invariance

The first fundamental problem we want to discuss is the choice of gauge in cosmological perturbation theory:

For linear perturbation theory to apply, the spacetime manifold $\mathcal{M}$ with metric g and the energy momentum tensor T of the real, observable universe must be in some sense close to a Friedmann universe, *i.e.* the manifold $\mathcal{M}$ with a Robertson–Walker metric $\bar{g}$ and a homogeneous and isotropic energy momentum tensor $\overline{T}$. It is an interesting, non–trivial unsolved problem how to construct 'the best' $\bar{g}$ and $\overline{T}$ from the physical fields g and T in practice. There are two main difficulties: First, spatial averaging procedures depend on the choice of a hyper–surface of constant time and they do not commute with derivatives, so that averaged fields $\bar{g}$ and $\overline{T}$ will in general not satisfy Einstein's equations. Secondly, averaging is in practice impossible over super–horizon scales.

Even though we cannot give a constructive prescription, we now assume that there exists an averaging procedure which leads to a Friedmann universe with spatially averaged tensor fields $\overline{Q}$, such that the deviations $(T_{\mu\nu}-\overline{T}_{\mu\nu})/\max_{\{\alpha\beta\}}\{|\overline{T}_{\alpha\beta}|\}$ and $(g_{\mu\nu}-\overline{g}_{\mu\nu})/\max_{\{\alpha\beta\}}\{\overline{g}_{\alpha\beta}\}$ are small, and $\bar{g}$ and $\overline{T}$ satisfy Friedmann's equations. Let us call such an averaging procedure 'admissible'. There may be many different admissible averaging procedures (e.g. over different hyper–surfaces) leading to slightly different Friedmann backgrounds. But since $|g-\bar{g}|$ is small of order ϵ, the difference of the two Friedmann backgrounds must also be small of order ϵ and we can regard it as part of the perturbation.

We consider now a fixed admissible Friedmann background $(\bar{g},\bar{T})$ as chosen. Since the theory is invariant under diffeomorphisms (coordinate transformations), the perturbations are not unique. For an arbitrary diffeomorphism ϕ and its push forward ϕ^*, the two metrics g and $\phi^*(g)$ describe the same geometry. Since we have chosen the background metric $\bar{g}$ we only allow diffeomorphisms which leave $\bar{g}$ invariant *i.e.* which deviate only in first order form the identity. Such an 'infinitesimal' diffeomorphism can be represented as the infinitesimal flow of a vector field X, $\phi=\phi^X_\epsilon$. Remember the definition of the flow: For the integral curve $\gamma_x(s)$ of X with starting point x, i.e., $\gamma_x(s=0)=x$ we have $\phi^X_s(x)=\gamma_x(s)$. In terms of the vector field X, to first order in ϵ, its pullback is then of the form

$$\phi^* = id + \epsilon L_X$$

(L_X denotes the Lie derivative in direction X [1]. The transformation $g\to\phi^*(g)$ is equivalent to $\bar{g}+\epsilon a^2 h \to \bar{g}+\epsilon(a^2h+L_X\bar{g})+\mathcal{O}(\epsilon^2)$, *i.e.* under

[1] The Lie derivative is geometrically just the derivative in the direction of the vector field. To define it in a coordinate independent manner, we consider a vector field X and its flux $\phi_s(x)$, i.e. the solution of the equation $dxds=X(x)$ with starting point x. For small enough values of s this is well defined. The Lie derivative of a tensor field T is then defined by

an 'infinitesimal coordinate transformation' the metric perturbation h transforms as

$$h \to h + a^{-2} L_X \bar{g} \ . \tag{2.13}$$

In the context of cosmological perturbation theory, infinitesimal coordinate transformations are called 'gauge transformations'. The perturbation of a arbitrary tensor field $Q = \bar{Q} + \epsilon Q^{(1)}$ obeys the gauge transformation law

$$Q^{(1)} \to Q^{(1)} + L_X \bar{Q} \ . \tag{2.14}$$

Since every vector field X generates a gauge transformation $\phi = \phi_\epsilon^X$, we can conclude that only perturbations of tensor fields with $L_X \overline{Q} = 0$ for all vector fields X, *i.e.* with vanishing (or constant) 'background contribution' are gauge invariant. This simple result is sometimes referred to as the 'Stewart-Walker Lemma' [4].

The gauge dependence of perturbations has caused many controversies in the literature, since it is often difficult to extract the physical meaning of gauge dependent perturbations, especially on super–horizon scales. This has led to the development of gauge invariant perturbation theory which we are going to use throughout this review. The advantage of the gauge invariant formalism is that the variables used have simple geometric and physical meanings and are not plagued by gauge modes. Although the derivation requires somewhat more work, the final system of perturbation equations is usually simple and well suited for numerical treatment. We shall also see, that on sub-horizon scales, the gauge invariant matter perturbation variables approach the usual, gauge dependent ones. Since one of the gauge invariant geometrical perturbation variables corresponds to the Newtonian potential, the Newtonian limit can be performed easily.

First we note that since all relativistic equations are covariant (i.e. can be written in the form $Q = 0$ for some tensor field Q), it is always possible to express the corresponding perturbation equations in terms of gauge invariant variables [5, 6, 7].

2.3.2 Harmonic Decomposition of Perturbation Variables

Since the $\{\eta = \text{const}\}$ hyper-surfaces are homogeneous and isotropic, it is sensible to perform a harmonic analysis: A (spatial) tensor field Q on these hyper-surfaces can be decomposed into components which transform irreducibly under translations and rotations. All such components evolve independently. For a scalar quantity f in the case $\kappa = 0$ this is nothing else than its Fourier decomposition:

$$f(\mathbf{x}, \eta) = \int d^3k \hat{f}(\mathbf{k}) e^{i\mathbf{kx}}. \tag{2.15}$$

$L_X T = lim_{s \to 0} \left(\phi_s^* T - T\right)/s.$

Here ϕ_s^* denotes the pullback with the (local) diffeomorphism ϕ_s [3].

(The exponentials $Y_{\mathbf{k}}(\mathbf{x}) = e^{i\mathbf{kx}}$ are the unitary irreducible representations of the Euclidean translation group.) For $\kappa = 1$ such a decomposition also exists, but the values k are discrete, $k^2 = \ell(\ell+2)$ and for $\kappa = -1$, they are bounded from below, $k^2 > 1$. Of course, the functions $Y_{\mathbf{k}}$ are different for $\kappa \neq 0$.

They form the complete orthogonal set of eigenfunctions of the Laplacian,

$$\Delta Y^{(S)}_{\mathbf{k}} = -k^2 Y^{(S)}_{\mathbf{k}}. \tag{2.16}$$

In addition, a tensorial variable (at fixed position $\mathbf{x}$) can be decomposed into irreducible components under the rotation group $SO(3)$.

For a vector field, this is its decomposition into a gradient and a rotation,

$$V_i = \nabla_i \varphi + B_i \ , \tag{2.17}$$

where

$$B^i_{|i} = 0 \ , \tag{2.18}$$

where we used $X_{|i}$ to denote the three–dimensional covariant derivative of X. Here φ is the spin 0 and $\mathbf{B}$ is the spin 1 component of the vector field V.

For a symmetric tensor field we have

$$H_{ij} = H_L \gamma_{ij} + \left(\nabla_i\nabla_j - \frac{1}{3}\Delta\gamma_{ij}\right) H_T + \frac{1}{2}\left(H^{(V)}_{i|j} + H^{(V)}_{j|i}\right) + H^{(T)}_{ij}, \tag{2.19}$$

where

$$H^{(V)|i}_i = H^{(T)^i}_i = H^{(T)^j}_{i|j} = 0. \tag{2.20}$$

Here H_L and H_T are spin 0 components, $H^{(V)}_i$ is a spin 1 component and $H^{(T)}_{ij}$ is a spin 2 component.

We shall not need higher tensors (or spinors). As a basis for vector and tensor modes we use the vector and tensor type eigenfunctions of the Laplacian,

$$\Delta Y^{(V)}_j = -k^2 Y^{(V)}_j \quad \text{and} \tag{2.21}$$

$$\Delta Y^{(T)}_{ji} = -k^2 Y^{(T)}_{ji} \ , \tag{2.22}$$

where $Y^{(V)}_j$ is a transverse vector, $Y^{(V)|j}_j = 0$ and $Y^{(T)}_{ji}$ is a symmetric transverse traceless tensor, $Y^{(T)j}_j = Y^{(T)|i}_{ji} = 0$.

According to (2.17) and (2.19) we can construct scalar type vectors and tensors and vector type tensors. To this goal we define

$$Y^{(S)}_j \equiv -k^{-1} Y^{(S)}_{|j} \tag{2.23}$$

$$Y^{(S)}_{ij} \equiv k^{-2} Y^{(S)}_{|ij} + \frac{1}{3}\gamma_{ij} Y^{(S)} \tag{2.24}$$

$$Y^{(V)}_{ij} \equiv -\frac{1}{2k}(Y^{(V)}_{i|j} + Y^{(V)}_{j|i}) \ . \tag{2.25}$$

In the following we shall extensively use this decomposition and write down the perturbation equations for a given mode k. The decomposition of a vector field is then of the form

$$B_i = BY_i^{(S)} + B^{(V)}Y_i^{(V)}. \tag{2.26}$$

The decomposition of a tensor field is given by (compare (2.19))

$$H_{ij} = H_L Y^{(S)}\gamma_{ij} + H_T Y_{ij}^{(S)} + H^{(V)}Y_{ij}^{(V)} + H^{(T)}Y_{ij}^{(T)}. \tag{2.27}$$

Here B, $B^{(V)}$, H_L, H_T, $H^{(V)}$ and $H^{(T)}$ are functions of η and $\mathbf{k}$.

2.3.3 Metric Perturbations

Perturbations of the metric are of the form

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + a^2 h_{\mu\nu}. \tag{2.28}$$

We parameterize them as

$$h_{\mu\nu}dx^\mu dx^\nu = -2Ad\eta^2 - 2B_i d\eta dx^i + 2H_{ij}dx^i dx^j, \tag{2.29}$$

and we decompose the perturbation variables B_i and H_{ij} according to (2.26) and (2.27).

Let us consider the behavior of $h_{\mu\nu}$ under gauge transformations. We set the vector field defining the gauge transformation to

$$\mathbf{X} = T\partial_\eta + L^i\partial_i. \tag{2.30}$$

Using simple identities from differential geometry like $L_{\mathbf{X}}(df) = d(L_{\mathbf{X}}f)$ and $(L_{\mathbf{X}}\gamma)_{ij} = X_{i|j} + X_{j|i}$, we obtain

$$\begin{aligned} L_{\mathbf{X}}\bar{g} = a^2 \Bigg[&-2\left(\frac{\dot{a}}{a}T + \dot{T}\right)d\eta^2 + 2\left(\dot{L}_i - T_{,i}\right)d\eta dx^i \\ &+ \left(2\frac{\dot{a}}{a}T\gamma_{ij} + L_{i|j} + L_{j|i}\right)dx^i dx^j \Bigg]. \end{aligned} \tag{2.31}$$

Comparing this with (2.29) and using (2.13) we obtain the following behaviour of our perturbation variables under gauge transformations (we decompose the vector $L_i = LY_i^{(S)} + L^{(V)}Y_i^{(V)}$):

$$A \to A + \frac{\dot{a}}{a}T + \dot{T} \tag{2.32}$$

$$B \to B - \dot{L} - kT \tag{2.33}$$

$$B^{(V)} \to B^{(V)} - \dot{L}^{(V)} \tag{2.34}$$

$$H_L \to H_L + \frac{\dot{a}}{a}T + \frac{k}{3}L \tag{2.35}$$

$$H_T \to H_T - kL \tag{2.36}$$

$$H^{(V)} \to H^{(V)} - kL^{(V)} \tag{2.37}$$

$$H^{(T)} \to H^{(T)}. \tag{2.38}$$

Two scalar and one vector variable can be set to zero by a cleverly chosen gauge transformations.

One often chooses $kL = H_T$ and $kT = B + \dot{L}$, so that the variables H_T and B vanish. In this gauge (longitudinal gauge), scalar perturbations of the metric are of the form ($H_T = B = 0$):

$$h^{(S)}_{\mu\nu} = -2\Psi d\eta^2 - 2\Phi\gamma_{ij}dx^i dx^j. \tag{2.39}$$

Ψ and Φ are the so called *Bardeen* potentials. In a generic gauge they are given by

$$\Psi = A - \frac{\dot{a}}{a}k^{-1}\sigma - k^{-1}\dot{\sigma} \tag{2.40}$$

$$\Phi = -H_L - \frac{1}{3}H_T + \frac{\dot{a}}{a}k^{-1}\sigma, \tag{2.41}$$

with $\sigma = k^{-1}\dot{H}_T - B$. A short calculation using (2.32) to (2.36) shows that Ψ and Φ are gauge invariant.

In a Friedmann universe the Weyl tensor vanishes. It therefore is a gauge invariant perturbation. For scalar perturbations one finds

$$C^0{}_{i0j} = \frac{1}{2}\left[(\Psi+\Phi)_{|ij} - \frac{1}{3}\triangle(\Psi+\Phi)\gamma_{ij}\right]. \tag{2.42}$$

All other components vanish.

For vector perturbations it is convenient to set $kL^{(V)} = H^{(V)}$ so that $H^{(V)}$ vanishes and we have

$$h^{(V)}_{\mu\nu}dx^\mu dx^\nu = 2\sigma^{(V)}Y^{(V)}_i d\eta dx^i. \tag{2.43}$$

We shall call this gauge the "vector gauge". In general $\sigma^{(V)} = k^{-1}\dot{H}^{(V)} - B^{(V)}$ is gauge invariant[2]. The Weyl tensor from vector perturbation is given by

$$C^0{}_{i0j} = \frac{1}{2}\dot{\sigma}^{(V)}Y^{(V)}_{ij} \tag{2.44}$$

$$C^0{}_{jlm} = \frac{1}{2}\sigma^{(V)}[Y^{(V)}_{jl|m} - Y^{(V)}_{jm|l} - \frac{1}{3}\gamma_{jl}Y^{(V)k}_{m|k} + \frac{1}{3}\gamma_{jm}Y^{(V)k}_{l|k}]. \tag{2.45}$$

Clearly there are no tensorial (spin 2) gauge transformation and hence $H^{(T)}_{ij}$ is gauge invariant. The expression for the Weyl tensor from tensor perturbation is identical to the one for vector perturbation upon replacement of $\sigma^{(V)}_{ij}$ by $\dot{H}^{(T)}_{ij}$.

[2] $Y^{(V)}_{ij}\sigma^{(V)}$ is the shear of the hyper-surfaces of constant time.

2.3.4 Perturbations of the Energy Momentum Tensor

Let $T^\mu_\nu = \overline{T}^\mu_\nu + \theta^\mu_\nu$ be the full energy momentum tensor. We define its energy density ρ and its energy flow 4-vector u as the time-like eigenvalue and eigenvector of T^μ_ν:

$$T^\mu_\nu u^\nu = -\rho u^\mu, \quad u^2 = -1. \tag{2.46}$$

We then parameterize their perturbations by

$$\rho = \bar{\rho}\,(1+\delta)\,, \quad u = u^0\partial_t + u^i\partial_i. \tag{2.47}$$

u^0 is fixed by the normalization condition,

$$u^0 = \frac{1}{a}(1-A). \tag{2.48}$$

We further set

$$u^i = \frac{1}{a}v^i = vY^{(S)i} + v^{(V)}Y^{(V)i}. \tag{2.49}$$

We define $P^\mu_\nu \equiv u^\mu u_\nu + \delta^\mu_\nu$, the projection tensor onto the part of tangent space normal to u and set the stress tensor

$$\tau^{\mu\nu} = P^\mu_\alpha P^\nu_\beta T^{\alpha\beta}. \tag{2.50}$$

In the unperturbed case we have $\tau^0_0 = 0, \tau^i_j = \bar{p}\delta^i_j$. Including perturbations, to first order we still obtain

$$\tau^0_0 = \tau^0_i = \tau^i_0 = 0. \tag{2.51}$$

But τ^i_j contains in general perturbations. We set

$$\tau^i_j = \bar{p}\left[(1+\pi_L)\,\delta^i_j + \Pi^i_j\right], \quad \text{with} \quad \Pi^i_i = 0. \tag{2.52}$$

We decompose Π^i_j as

$$\Pi^i_j = \Pi^{(S)}Y^{(S)\,i}_j + \Pi^{(V)}Y^{(V)\,i}_j + \Pi^{(T)}Y^{(T)\,i}_j. \tag{2.53}$$

We shall not derive the gauge transformation properties of these perturbation variables in detail, but just state some results which can be obtained as an exercise (see also [6]):

- Of the variables defined above only the $\Pi^{(S,V,T)}$ are gauge invariant; they describe the anisotropic stress tensor, $\Pi^\mu_\nu = \tau^\mu_\nu - {}^1\!/_3\tau^\alpha_\alpha\delta^\mu_\nu$. They are gauge invariant due to the Stewart–Walker lemma, since $\bar{\Pi} = 0$. For perfect fluids $\Pi^\mu_\nu = 0$.
- A second gauge invariant variable is

$$\Gamma = \pi_L - \frac{c_s^2}{w}\delta, \tag{2.54}$$

  where $c_s^2 \equiv \dot{p}/\dot{\rho}$ is the adiabatic sound speed and $w \equiv p/\rho$ is the enthalpy. One can show that Γ is proportional to the divergence of the entropy flux of the perturbations. Adiabatic perturbations are characterized by $\Gamma = 0$.

– Gauge invariant density and velocity perturbations can be found by combining δ, v and $v_i^{(V)}$ with metric perturbations.

We shall use

$$V \equiv v - \frac{1}{k}\dot{H}_T = v^{(\rm long)} \tag{2.55}$$

$$D_s \equiv \delta + 3(1+w)\frac{\dot{a}}{a}(k^{-2}\dot{H}_T - k^{-1}B) \equiv \delta^{(\rm long)} \tag{2.56}$$

$$D \equiv \delta^{(\rm long)} + 3(1+w)\left(\frac{\dot{a}}{a}\right)\frac{V}{k} \tag{2.57}$$

$$D_g \equiv \delta + 3(1+w)\left(H_L + \frac{1}{3}H_T\right) = \delta^{(\rm long)} - 3(1+w)\Phi \tag{2.58}$$

$$V^{(V)} \equiv v^{(V)} - \frac{1}{k}\dot{H}^{(V)} = v^{(\rm vec)} \tag{2.59}$$

$$\Omega \equiv v^{(V)} - B^{(V)} = v^{(\rm vec)} - B^{(V)} \tag{2.60}$$

$$\Omega - V^{(V)} = \sigma^{(V)}. \tag{2.61}$$

Here $v^{(\rm long)}$, $\delta^{(\rm long)}$ and $v_i^{(\rm vec)}$ are the velocity (and density) perturbations in the longitudinal and vector gauge respectively, and $\sigma^{(V)}$ is the metric perturbation in vector gauge (see (2.43)). These variables can be interpreted nicely in terms of gradients of the energy density and the shear and vorticity of the velocity field [8].

Here I just want to show that on scales much smaller than the Hubble scale, $k\eta \gg 1$, the metric perturbations are much smaller than δ and v and we can thus "forget them" (which will be important when comparing experimental results with calculations in this formalism): The perturbations of the Einstein tensor are given by second derivatives of the metric perturbations. Einstein's equations yield the following order of magnitude estimate:

$$\mathcal{O}\left(\frac{\delta T}{T}\right) \underbrace{\mathcal{O}\left(8\pi G T\right)}_{\mathcal{O}\left(\frac{\dot{a}}{a}\right)^2 = \mathcal{O}(\eta^{-2})} = \mathcal{O}\left(\frac{1}{\eta^2}h + \frac{k}{\eta}h + k^2 h\right) \tag{2.62}$$

$$\mathcal{O}\left(\frac{\delta T}{T}\right) = \mathcal{O}\left(h + k\eta h + (k\eta)^2 h\right). \tag{2.63}$$

For $k\eta \gg 1$ this gives $\mathcal{O}(\delta, v) = \mathcal{O}\left(\frac{\delta T}{T}\right) \gg \mathcal{O}(h)$. On sub-horizon scales the difference between δ, $\delta^{(\rm long)}$, D_g and D is negligible as well as the difference between v and V or $v^{(V)}$, $V^{(V)}$ and $\Omega^{(V)}$.

2.4 Einstein's Equations

We do not derive the first order perturbations of Einstein's equations. This can be done by different methods, for example with Mathematica. We just write down those equations which we shall need later.

2.4.1 Constraint Equations

$$\left.\begin{aligned} 4\pi G a^2 \rho D &= -(k^2 - 3\kappa)\Phi \quad (00) \\ 4\pi G a^2 (\rho + p) V &= k\left(\left(\tfrac{\dot a}{a}\right)\Psi + \dot\Phi\right) \quad (0i) \end{aligned}\right\} \quad \text{(scalar)} \tag{2.64}$$

$$8\pi G a^2 (\rho + p)\Omega = \frac{1}{2}\left(2\kappa - k^2\right)\sigma^{(V)} \quad (0i) \quad \text{(vector)} \tag{2.65}$$

2.4.2 Dynamical Equations

$$k^2\left(\Phi - \Psi\right) = 8\pi G a^2 p \Pi^{(S)} \quad \text{(scalar)} \tag{2.66}$$

$$k\left(\dot\sigma^{(V)} + 2\left(\frac{\dot a}{a}\right)\sigma^{(V)}\right) = 8\pi G a^2 p \Pi^{(V)} \quad \text{(vector)} \tag{2.67}$$

$$\ddot H^{(T)} + 2\left(\frac{\dot a}{a}\right)\dot H^{(T)} + \left(2\kappa + k^2\right) H^{(T)} = 8\pi G a^2 p \Pi_{ij}^{(T)} \quad \text{(tensor)} \tag{2.68}$$

There is a second dynamical scalar equation, which is somewhat cumbersome and not really needed, since we may use one of the conservation equations given below instead. Note that for perfect fluids, where $\Pi_j^i \equiv 0$, we have $\Phi = \Psi$, $\sigma^{(V)} \propto 1/a^2$ and $H^{(T)}$ obeys a damped wave equation. The damping term can be neglected on small scales (over short time periods) when $\eta^{-2} \lesssim 2\kappa + k^2$, and H_{ij} represents propagating gravitational waves. For vanishing curvature, these are just the sub-horizon scales, $k\eta \gtrsim 1$. For $\kappa < 0$, waves oscillate with a somewhat smaller frequency, $\omega = \sqrt{2\kappa + k^2}$, while for $\kappa > 0$ the frequency is somewhat larger than k.

2.4.3 Energy Momentum Conservation

The conservation equations, $T^{\mu\nu}_{\ ;\nu} = 0$ lead to the following perturbation equations

$$\left.\begin{aligned} \dot D_g + 3\left(c_s^2 - w\right)\left(\tfrac{\dot a}{a}\right) D_g + (1+w)kV + 3w\left(\tfrac{\dot a}{a}\right)\Gamma = 0 \\ \dot V + \left(\tfrac{\dot a}{a}\right)\left(1 - 3c_s^2\right) V = k\left(\Psi + 3c_s^2\Phi\right) + \tfrac{c_s^2 k}{1+w} D_g \\ + \tfrac{wk}{1+w}\left[\Gamma - \tfrac{2}{3}\left(1 - \tfrac{3\kappa}{k^2}\right)\Pi\right] \end{aligned}\right\} \quad \text{(scalar)} \tag{2.69}$$

$$\dot\Omega_i + \left(1 - 3c_s^2\right)\left(\frac{\dot a}{a}\right)\Omega_i = \frac{p}{2(\rho + p)}\left(k - \frac{2\kappa}{k}\right)\Pi_i^{(V)} \quad \text{(vector)} \tag{2.70}$$

These can of course also be obtained from the Einstein equations since they are equivalent to the contracted Bianchi identities. For scalar perturbations we have 4 independent equations and 6 variables. For vector perturbations we have 2 equations and 3 variables, while for tensor perturbations we have 1 equations and 2 variables. To close the system we must add some matter equations. The simplest prescription is to set $\Gamma = \Pi_{ij} = 0$. This matter equation, which describes adiabatic perturbations of a perfect fluid gives us exactly two additional equations for scalar perturbations and one each for vector and tensor perturbations.

Another simple example is a universe with matter content given by a scalar field. We shall discuss this case in the next section. More complicated examples are those of several interacting particle species of which some have to be described by a Boltzmann equation. This is the actual universe at late times, say $z \lesssim 10^7$.

2.4.4 A Special Case

Here we want to rewrite the scalar perturbation equations for a simple but important special case. We consider adiabatic perturbations of a perfect fluid. In this case $\Pi = 0$ since there are no anisotropic stresses and $\Gamma = 0$. Equation (2.66) then implies $\Phi = \Psi$. Using the first equation of (2.64) and (2.58,2.57) to replace D_g in the second of (2.69) by Ψ and V, finally replacing V by (2.64) one can derive a second order equation for Ψ, which is, in this case the only dynamical degree of freedom

$$\ddot{\Psi} + 3\mathcal{H}(1+c_s^2)\dot{\Psi} + [(1+3c_s^2)(\mathcal{H}^2-\kappa) - (1+3w)\mathcal{H}^2 + c_s^2k^2]\Psi = 0\ . \quad (2.71)$$

Another interesting case (especially when discussing inflation) is the scalar field case. There, as we shall see in Sect. 2.6, $\Pi = 0$, but in general $\Gamma \neq 0$ since $\delta p/\delta\rho \neq \dot{p}/\dot{\rho}$. Nevertheless, since this case again has only one dynamical degree of freedom, we can express the perturbation equations in terms of one single second order equation for Ψ. In Sect. 2.6 we shall find the following equation for a perturbed scalar field cosmology

$$\ddot{\Psi} + 3\mathcal{H}(1+c_s^2)\dot{\Psi} + [(1+3c_s^2)(\mathcal{H}^2-\kappa) - (1+3w)\mathcal{H}^2 + k^2]\Psi = 0\ . \quad (2.72)$$

The only difference between the perfect fluid and scalar field perturbation equation is that the latter is missing the factor c_s^2 in front of the oscillatory k^2 term. Note also that for $\kappa = 0$ and $w = c_s^2 =$ constant, the time dependent mass term $m^2(\eta) = -(1+3c_s^2)(\mathcal{H}^2-\kappa) + (1+3w)\mathcal{H}^2$ vanishes. It is useful to define also the variable [12]

$$u = a\left[4\pi G(\mathcal{H}^2 - \dot{\mathcal{H}} + \kappa)\right]^{-1/2}\Psi, \quad (2.73)$$

which satisfies the equation

$$\ddot{u} + (\Upsilon k^2 - \ddot{\theta}/\theta)u = 0, \tag{2.74}$$

where $\Upsilon = c_s^2$ or $\Upsilon = 1$ for a perfect fluid or a scalar field background respectively, and

$$\theta = \frac{3\mathcal{H}}{2a\sqrt{\mathcal{H}^2 - \dot{\mathcal{H}} + \kappa}} \ . \tag{2.75}$$

Another interesting variable is

$$\zeta \equiv \frac{2(\mathcal{H}^{-1}\dot{\Psi} + \Psi)}{3(1+w)} + \Psi \ . \tag{2.76}$$

Using (2.71) and (2.72) respectively one obtains

$$\dot{\zeta} = -k^2 \frac{\Upsilon\mathcal{H}}{\mathcal{H}^2 - \dot{\mathcal{H}}}\Psi \ , \tag{2.77}$$

hence on super horizon scales, $k/\mathcal{H} \ll 1$, this variable is conserved.

The evolution of ζ is closely related to the canonical variable v defined by

$$v = -\frac{a\sqrt{\mathcal{H}^2 - \dot{\mathcal{H}}}}{\sqrt{4\pi G}\Upsilon\mathcal{H}}\zeta \ . \tag{2.78}$$

which satisfies the equation

$$\ddot{v} + (\Upsilon k^2 - \ddot{z}/z)v = 0 \ , \tag{2.79}$$

for

$$z = \frac{a\sqrt{\mathcal{H}^2 - \dot{\mathcal{H}} + \kappa}}{\Upsilon\mathcal{H}} \ . \tag{2.80}$$

More details on the significance of the canonical variable v will be found in Sects. 2.6 and 2.7.

2.5 Simple Examples

We first discuss two simple applications which are important to understand the CMB anisotropy spectrum.

2.5.1 The Pure Dust Fluid for $\kappa = 0, \Lambda = 0$

We assume the dust to have $w = c_s^2 = p = 0$ and $\Pi = \Gamma = 0$. (2.71) then reduces to

$$\ddot{\Psi} + \frac{6}{\eta}\dot{\Psi} = 0 \ , \tag{2.81}$$

with the general solution

$$\Psi = \Psi_0 + \Psi_1 \frac{1}{\eta^5} \ , \tag{2.82}$$

with arbitrary constants Ψ_0 and Ψ_1. Since the perturbations are supposed to be small initially, they cannot diverge for $\eta \to 0$, and we have therefore to choose the growing mode, $\Psi_1 = 0$. Another way to argue is as follows: If the mode Ψ_1 has to be small already at some early initial time $\eta_{\rm in}$, it will be even much smaller later and may hence be neglected at late times. But also the Ψ_0 mode is only constant. This fact led Lifshitz who was the first to analyze cosmological perturbations to the conclusions that linear perturbations do not grow in a Friedman universe and cosmic structure cannot have evolved by gravitational instability [2]. However, the important point to note here is that, even if the gravitational potential remains constant, matter density fluctuations do grow on sub-horizon scales and therefore inhomogeneities, structure can evolve on scales which are smaller than the Hubble scale. To see that we consider the conservation equations (2.69), (2.66) and the Poisson equation (2.64). For the pure dust case, $w = c_s^2 = \Pi = \Gamma = 0$, they reduce to

$$\dot D_g = -kV \quad \text{(energy conservation)} \tag{2.83}$$

$$\dot V + \left(\frac{\dot a}{a}\right) V = k\Psi \quad \text{(gravitational acceleration)} \tag{2.84}$$

$$-k^2\Psi = 4\pi G a^2 \rho \left(D_g + 3 \left(\Psi + \left(\frac{\dot a}{a}\right) k^{-1} V \right) \right) \text{(Poisson)}, \tag{2.85}$$

where we have used the relation

$$D = D_g + 3(1+w) \left(-\Phi + \left(\frac{\dot a}{a}\right) k^{-1} V \right). \tag{2.86}$$

The Friedmann equation for dust gives $4\pi G\rho a^2 = {}^3\!/_2 (\dot a/a)^2 = 6/\eta^2$. Setting $k\eta = x$ and $' = d/dx$, the system (2.83-2.85) becomes

$$D_g' = -V \tag{2.87}$$

$$V' + \frac{2}{x} V = \Psi \tag{2.88}$$

$$\frac{6}{x^2} \left(D_g + 3 \left(\Psi + \frac{2}{x} V \right) \right) = -\Psi. \tag{2.89}$$

We use (2.89) to eliminate Ψ and (2.87) to eliminate D_g, leading to

$$\left(18 + x^2\right) V'' + \left(\frac{72}{x} + 4x \right) V' - \left(\frac{72}{x^2} + 4 \right) V = 0. \tag{2.90}$$

The general solution of (2.90) is

$$V = V_0 x + \frac{V_1}{x^4}. \tag{2.91}$$

The V_1 mode is the decaying mode (corresponding to Ψ_1) which we neglect. The perturbation variables are then given by

$$V = V_0 x \tag{2.92}$$

$$D_g = -15V_0 - \frac{1}{2}V_0 x^2 \tag{2.93}$$

$$V_0 = \Psi_0/3 \ . \tag{2.94}$$

We distinguish two regimes:
i) super-horizon, $x \ll 1$ where we have

$$V = \frac{1}{3}\Psi_0 x \tag{2.95}$$

$$D_g = -5\Psi_0 \tag{2.96}$$

$$\Psi = \Psi_0 \ . \tag{2.97}$$

Note that even though V is growing, it always remains much small than Ψ or D_g on super-horizon scales. Hence the largest fluctuations are of order Ψ which is constant.
ii) Sub-horizon, $x \gg 1$ where the solution is dominated by the terms

$$V = \frac{1}{3}\Psi_0 x \tag{2.98}$$

$$D_g = -\frac{1}{6}\Psi_0 x^2 \tag{2.99}$$

$$\Psi = \Psi_0 = \text{constant}. \tag{2.100}$$

Note that for dust

$$D = D_g + 3\Psi + \frac{6}{x}V = -\frac{1}{6}\Psi_0 x^2 \ .$$

In the variable D the constant term has disappeared and we have $D \ll \Psi$ on super-horizon scales, $x \ll 1$.

On sub-horizon scales, the density fluctuations grow like the scale factor $\propto x^2 \propto a$. Nevertheless, Lifshitz' conclusion [2] that pure gravitational instability cannot be the cause for structure formation has some truth: If we start from tiny thermal fluctuations of the order of 10^{-35}, they can only grow to about 10^{-30} due to this mild, power law instability during the matter dominated regime. Or, to put it differently, if we want to form structure by gravitational instability, we need initial fluctuations of the order of at least 10^{-5}, much larger than thermal fluctuations. One possibility to create such fluctuations is quantum particle production in the classical gravitational field during inflation. The rapid expansion of the universe during inflation quickly expands microscopic scales at which quantum fluctuations are important to cosmological scales where these fluctuations are then "frozen in" as classical perturbations in the energy density and the geometry. We will discuss the induced spectrum on fluctuations in Sect. 2.7.

2.5.2 The Pure Radiation Fluid, $\kappa = 0, \Lambda = 0$

In this limit we set $w = c_s^2 = 1/3$ and $\Pi = \Gamma = 0$ so that $\Phi = -\Psi$. We conclude from $\rho \propto a^{-4}$ that $a \propto \eta$. For radiation, the u–equation (2.74) becomes

$$\ddot{u} + (\frac{1}{3}k^2 - \frac{2}{\eta^2})u = 0, \tag{2.101}$$

with general solution

$$u(x) = A\left(\frac{\sin(x)}{x} - \cos(x)\right) + B\left(\frac{\cos(x)}{x} - \sin(x)\right) \ , \tag{2.102}$$

where we have set $x = k\eta/\sqrt{3} = c_s k\eta$. For the Bardeen potential we obtain with (2.73), up to constant factors,

$$\Psi(x) = \frac{u(x)}{x^2} \ . \tag{2.103}$$

On super-horizon scales, $x \ll 1$, we have

$$\Psi(x) \simeq \frac{A}{3} + \frac{B}{x^3} \ . \tag{2.104}$$

We assume that the perturbations have been initialized at some early time $x_{\rm in} \ll 1$ and that at this time the two modes have been comparable. If this is the case then $B \ll A$ and we may neglect the B-mode at later times.

To determine the density and velocity perturbations and for illustration, we also solve the radiation equations using the conservation and Poisson equations like the dust case. In the radiation case the perturbation equations become (with the same notation as above, $x = c_s k\eta$)

$$D_g' = -\frac{4}{\sqrt{3}}V \tag{2.105}$$

$$V' = 2\sqrt{3}\Psi + \frac{\sqrt{3}}{4}D_g \tag{2.106}$$

$$-2x^2\Psi = D_g + 4\Psi + \frac{4}{\sqrt{3}x}V \ . \tag{2.107}$$

The general solution of this system is

$$\begin{aligned} D_g = {} & D_2\left[\cos(x) - \frac{2}{x}\sin(x)\right] \\ & + D_1\left[\sin(x) + \frac{2}{x}\cos(x)\right] \end{aligned} \tag{2.108}$$

$$V = -\frac{\sqrt{3}}{4}D_g' \tag{2.109}$$

$$\Psi = -\frac{D_g + \frac{4}{\sqrt{3}x}V}{4 + 2x^2} \ . \tag{2.110}$$

Again, regularity at $x = 0$ requires $D_1 = 0$. Comparing with (2.102,2.103) gives $D_2 = 2A$. In the **super-horizon regime**, $x \ll 1$, we obtain

$$\Psi = \frac{A}{3}, \quad D_g = -2A - \frac{A}{3\sqrt{3}}x^2, \quad V = \frac{A}{2\sqrt{3}}x \ . \tag{2.111}$$

On **sub-horizon scales**, $x \gg 1$, we find oscillating solutions with constant amplitude and with frequency of $k/\sqrt{3}$:

$$V = \frac{\sqrt{3}A}{2}\sin(x) \tag{2.112}$$

$$D_g = 2A\cos(x) \ , \quad \Psi = -A\cos(x)/x^2 \ . \tag{2.113}$$

Note that also for radiation perturbations

$$D = -\frac{A}{3\sqrt{3}}x^2 \ll \Psi$$

is small on super horizon scales, $x \ll 1$. The perturbation amplitude is given by the largest gauge invariant perturbation variable. We conclude therefore that perturbations outside the Hubble horizon are frozen to first order. Once they enter the horizon they start to collapse, but pressure resists the gravitational force and the radiation fluid fluctuations oscillate at constant amplitude. The perturbations of the gravitational potential oscillate and decay like $1/a^2$ inside the horizon.

2.5.3 Adiabatic Initial Conditions

Adiabaticity requires that the perturbations of all contributions to the energy density are initially in thermal equilibrium. This fixes the ratio of the density perturbations of different components. There is no entropy flux and thus $\Gamma = 0$. Here we consider as a simple example non relativistic matter and radiation perturbations. Since the matter and radiation perturbations behave in the same way on super-horizon scales,

$$D_g^{(r)} = A + Bx^2, \quad D_g^{(m)} = A' + B'x^2, \quad V^{(r)} \propto V^{(m)} \propto x, \tag{2.114}$$

we may require a constant ratio between matter and radiation perturbations. As we have seen in the previous section, inside the horizon ($x > 1$) radiation perturbations start to oscillate while matter perturbations keep following a power law. On sub-horizon scales a constant ratio can thus no longer be maintained. There are two interesting possibilities: adiabatic and isocurvature perturbations. Here we concentrate on adiabatic perturbations which seem to dominate the observed CMB anisotropies.

From $\Gamma = 0$ one easily derives that two components with $p_i/\rho_i = w_i$ =constant, $i = 1, 2$, are adiabatically coupled if $(1 + w_1)D_g^{(2)} = (1 +$

$w_2)D_g^{(1)}$. Energy conservation then implies that also their velocity fields agree, $V^{(1)} = V^{(2)}$. This result is also a consequence of the Boltzmann equation in the strong coupling regime. We therefore require

$$V^{(r)} = V^{(m)}, \tag{2.115}$$

so that the energy flux in the two fluids is coupled initially.

We restrict ourselves to a matter dominated backgrounds, the situation relevant in the observed universe after equality. We first have to determine the radiation perturbations during a *matter dominated era*. Since Ψ is dominated by the matter contribution (it is proportional to the background density of a given component), we have $\Psi \simeq$ const. $= \Psi_0$. We neglect the contribution from the sub-dominant radiation to Ψ. Energy–momentum conservation for radiation then gives, with $x = k\eta$,

$$D_g^{(r)\prime} = -\frac{4}{3}V^{(r)} \tag{2.116}$$

$$V^{(r)\prime} = 2\Psi + \frac{1}{4}D_g^{(r)}. \tag{2.117}$$

Now Ψ is just a constant given by the matter perturbations, and it acts like a constant source term. The general solution of this system is then

$$D_g^{(r)} = A\cos(c_s x) - \frac{4}{\sqrt{3}}B\sin(c_s x) + 8\Psi\left[\cos(c_s x) - 1\right] \tag{2.118}$$

$$V^{(r)} = B\cos(c_s x) + \frac{\sqrt{3}}{4}A\sin(c_s x) + 2\sqrt{3}\Psi\sin(c_s x), \tag{2.119}$$

where $c_s = 1/\sqrt{3}$ is the sound speed of radiation. Our adiabatic initial conditions require

$$\lim_{x\to 0}\frac{V^{(r)}}{x} = V_0 = \lim_{x\to 0}\frac{V^{(m)}}{x} < \infty. \tag{2.120}$$

Therefore $B = 0$ and $V_0 = A/4 - 2\Psi$. Using in addition $\Psi = 3V_0$ (see (2.100)) we obtain

$$D_g^{(r)} = \frac{4}{3}\Psi\cos\left(\frac{x}{\sqrt{3}}\right) - 8\Psi \tag{2.121}$$

$$V^{(r)} = \frac{1}{\sqrt{3}}\Psi\sin\left(\frac{x}{\sqrt{3}}\right) \tag{2.122}$$

$$D_g^{(m)} = -\Psi(5 + \frac{1}{6}x^2) \tag{2.123}$$

$$V^{(m)} = \frac{1}{3}\Psi x\ . \tag{2.124}$$

On super-horizon scales, $x \ll 1$ we have

$$D_g^{(r)} \simeq -\frac{20}{3}\Psi \quad \text{and} \quad V^{(r)} \simeq \frac{1}{3}x\Psi \ , \tag{2.125}$$

note that $D_g^{(r)} = (4/3)D_g^{(m)}$ and $V^{(r)} = V^{(m)}$ for adiabatic initial conditions.

Another possibility for the initial condition would be isocurvature initial conditions, where you have non-vanishing $D^{(r)}$, $D^{(m)}$ and $V^{(r)}$, $V^{(m)}$ which compensate each other in such a way that $\Psi = 0$ on super-horizon scales. The simplest inflationary models do not lead to such perturbations and the observations imply that they are not dominating the observed anisotropies in the CMB even though they may contribute which could seriously hamper the determination of cosmological parameters with CMB anisotropies (see e.g. [9, 10]).

2.6 Scalar Field Cosmology

We now consider the special case of a Friedmann universe filled with self interacting scalar field matter. The action is given by

$$S = \frac{1}{16\pi G}\int d^4x\sqrt{|g|}R + \int d^4x\sqrt{|g|}\left(\frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - W(\varphi)\right), \tag{2.126}$$

where φ denotes the scalar field and W is the potential. The energy momentum tensor is obtained by varying the action in respect of the metric $g_{\mu\nu}$,

$$T_{\mu\nu} = \partial_\mu\varphi\partial_\nu\varphi - \left[\frac{1}{2}\partial_\lambda\varphi\partial^\lambda\varphi + W\right] g_{\mu\nu} \ . \tag{2.127}$$

The energy density ρ and the energy flux u are defined by

$$T^\mu_\nu u^\nu = -\rho u^\mu \ . \tag{2.128}$$

For the Friedmann background this gives

$$\rho = \frac{1}{2a^2}\dot{\varphi}^2 + W \ , \qquad (u^\mu) = \frac{1}{a}(1, \mathbf{0}) \ . \tag{2.129}$$

The pressure is given by

$$T^i_j = p\delta^i_j \ , \qquad p = \frac{1}{2a^2}\dot{\varphi}^2 - W \ . \tag{2.130}$$

We now define the scalar field perturbation,

$$\varphi = \bar{\varphi} + \delta\varphi \ . \tag{2.131}$$

Clearly, the scalar field only generates scalar perturbations (to first order). Inserting (2.131) in the definition of the energy velocity perturbation v,

$$(u^\mu) = \frac{1}{a}(1 - A, -v_{,i}) \tag{2.132}$$

and the energy density perturbation $\delta\rho$,

$$\rho = \bar\rho + \delta\rho \ , \tag{2.133}$$

we obtain

$$\delta\rho = \frac{1}{a^2}\dot{\bar\varphi}\delta\dot\varphi - \frac{1}{a^2}\dot{\bar\varphi}^2 A + W_{,\varphi}\,\delta\varphi, \tag{2.134}$$

and

$$v = \frac{k}{\dot{\bar\varphi}}\left(\delta\varphi + \dot{\bar\varphi}k^{-1}B\right) \ . \tag{2.135}$$

From the stress tensor, $T_{ij} = \varphi_{,i}\,\varphi_{,j} - \left[\frac{1}{2}\partial_\lambda\varphi\partial^\lambda\varphi + W\right] g_{ij}$ we find

$$p\pi_L = \frac{1}{a^2}\dot{\bar\varphi}\delta\dot\varphi - \frac{1}{a^2}\dot{\bar\varphi}^2 A - W_{,\varphi}\,\delta\varphi \qquad \text{and} \quad \Pi = 0 \ . \tag{2.136}$$

We now define the gauge invariant scalar field perturbation as the value of $\delta\varphi$ in longitudinal gauge

$$\delta\varphi^{(gi)} = \delta\varphi^{(long)} = \delta\varphi + \dot{\bar\varphi}\left(B - k^{-1}\dot H_T\right) = \delta\varphi - \dot{\bar\varphi}\sigma \ . \tag{2.137}$$

The last expression gives $\delta\varphi^{(gi)}$ in a generic gauge. It is clear that this combination is gauge–invariant. This variable is very simply related to the other gauge–invariant scalar variables. Short calculations give

$$V = k\delta\varphi^{(gi)}/\dot{\bar\varphi} \tag{2.138}$$

$$D_g = (1+w)\left[2\frac{\dot a}{a}\delta\varphi^{(gi)}/\dot{\bar\varphi} + \frac{d}{d\eta}\delta\varphi^{(gi)}/\dot{\bar\varphi}\right] \tag{2.139}$$

$$D_s = D_g + (1+w)\Psi \tag{2.140}$$

$$\Gamma = \frac{2W_{,\varphi}}{p\dot\rho}\left[\dot{\bar\varphi}\rho D_s - \dot\rho\delta\varphi^{(gi)}\right] \tag{2.141}$$

$$\Pi = 0 \ . \tag{2.142}$$

The last equation shows that the two Bardeen potentials are equal for scalar field perturbations, $\Phi = \Psi$. With this we can write the perturbed Einstein equations fully in terms of the Bardeen potential Ψ and V. Since we will need them mainly to discuss inflation where curvature plays a minor role, we write them down here only for the case of vanishing spatial curvature. From (2.64) and (2.66) one can easily generalize to the case with curvature.

$$-3\mathcal{H}\dot\Psi - (\dot{\mathcal{H}} + \mathcal{H}^2 - k^2)\Psi = 4\pi G\dot{\bar\varphi}k^{-1}(\dot{\bar\varphi}V + a^2 W_{,\varphi}\,V) \tag{2.143}$$

$$\dot\Psi + \mathcal{H}\Psi = 4\pi G\dot{\bar\varphi}^2 k^{-1}V \tag{2.144}$$

$$\ddot\Psi + 3\mathcal{H}\dot\Psi + (\dot{\mathcal{H}} + 2\mathcal{H}^2)\Psi = 4\pi G\dot{\bar\varphi}k^{-1}(\dot{\bar\varphi}\dot V - \ddot{\bar\varphi}V - a^2 W_{,\varphi}\,V). \tag{2.145}$$

These lead to the following second order equation for the Bardeen potential which we have discussed above:

$$\ddot{\Psi} + 3(\mathcal{H} - \ddot{\varphi}/\dot{\varphi})\dot{\Psi} + (2\dot{\mathcal{H}} - 2\mathcal{H}\ddot{\varphi}/\dot{\varphi} + k^2)\Psi = 0 \tag{2.146}$$

or, using the definition $c_s^2 = \dot{p}/\dot{\rho}$,

$$\ddot{\Psi} + 3\mathcal{H}(1 + c_s^2)\dot{\Psi} + (2\dot{\mathcal{H}} + (1 + 3c_s^2)\mathcal{H}^2 + k^2)\Psi = 0 \ . \tag{2.147}$$

As already mentioned above, this equation differs from the Ψ equation for a perfect fluid only in the last term proportional to k^2. This comes from the fact that the scalar field is not in a thermal state with fixed entropy, but it is in a fully coherent state ($\Gamma \neq 0$) and field fluctuations propagate with the speed of light. On large scales, $k\eta \ll 1$ this difference is not relevant, but on sub–horizon scales it does play a certain role.

2.7 Generation of Perturbations During Inflation

So far we have simply assumed some initial fluctuation amplitude A, without investigating where it came from or what the k–dependence of A might be. In this section we discuss the most common idea about the generation of cosmological perturbations, namely their production from the quantum vacuum fluctuations during an inflationary phase. The treatment here is focused mainly on getting the correct result with as little effort as possible; we ignore several subtleties related, *e.g.* to the transition from quantum fluctuations of the field to classical fluctuations in the energy momentum tensor. The idea is of course that the source of metric fluctuations are the expectation values of the energy momentum tensor operator of the scalar field.

The basic idea is simple: A time dependent gravitational field very generically leads to particle production, analogously to the electron positron production in a classical, time dependent electromagnetic field.

2.7.1 Scalar Perturbations

The main result is the following: During inflation, the produced particles induce a perturbed gravitational field with a (nearly) scale invariant spectrum,

$$k^3|\Psi(k,\eta)|^2 = k^{n-1} \times \text{const.} \quad \text{with} \quad n \simeq 1 \ . \tag{2.148}$$

The quantity $k^3|\Psi(k,\eta)|^2$ is the squared amplitude of the metric perturbation at comoving scale $\lambda = \pi/k$. To insure that this quantity is small on a broad range of scales, so that neither black holes are formed on small scales nor there are large deviation from homogeneity and isotropy on large scales, we must require $n \simeq 1$. These arguments have been put forward for the first time by Harrison and Zel'dovich [11] (still before the advent of inflation),

leading to the name 'Harrison-Zel'dovich spectrum' for a scale invariant perturbation spectrum.

To derive the above result we consider a scalar field background dominated by a potential, hence $a \propto |\eta|^q$ with $q \sim -1$. Looking at the action of this system,

$$S = \int dx^4 \sqrt{|g|} \left(\frac{R}{16\pi G} + \frac{1}{2}(\nabla\varphi)^2 \right) ,$$

it can be shown (see [12]) that the second order perturbation of this action around the Friedmann solution is given by

$$\delta S = \int dx^4 \sqrt{|\overline{g}|} \frac{1}{2} (\partial_\mu v)^2 , \tag{2.149}$$

up to some total differential. Here v is the perturbation variable

$$v = -\frac{a\sqrt{\mathcal{H}^2 - \dot{\mathcal{H}}}}{\sqrt{4\pi G}\mathcal{H}} \zeta \tag{2.150}$$

introduced in (2.78). Via the Einstein equations, this variable can also be interpreted as representing the fluctuations in the scalar field. Therefore, we quantize v and assume that initially, on small scales, $k|\eta| \ll 1$, v is in the (Minkowski) quantum vacuum state of a massless scalar field with mode function

$$v_{\rm in} = \frac{v_0}{\sqrt{k}} \exp(ik\eta) . \tag{2.151}$$

The pre-factor v_0 is a k-independent constant which depends on convention, but is of order unity. From (2.77) we can derive

$$(v/z)^{\cdot} = \frac{k^2 u}{z} ,$$

where $z \propto a$ is defined in (2.80) and $u \propto a\eta\Psi$ is given in (2.73). On small scales, $k|\eta| \ll 1$, this results in the initial condition for u

$$u_{\rm in} = \frac{-iv_0}{k^{3/2}} \exp(ik\eta) . \tag{2.152}$$

The evolution equation for u, (2.101), reduces in the case of power law expansion, $a \propto |\eta|^q$ to

$$\ddot{u} + (k^2 - \frac{q(q+1)}{\eta^2})u = 0 . \tag{2.153}$$

The solutions to this equation are of the form $(k|\eta|)^{1/2} H_\mu^{(i)}(k\eta)$, where $\mu = q + 1/2$ and $H_\mu^{(i)}$ is the Hankel function of the ith kind ($i = 1$ or 2) of order μ. The initial condition (2.152) requires that only $H_\mu^{(2)}$ appears, so that we obtain

$$u = \frac{\alpha}{k^{3/2}}(k|\eta|)^{1/2} H_\mu^{(2)}(k\eta) \ ,$$

where again α is a constant of order unity. We define the value of the Hubble constant during inflation, which is nearly constant by H_i. With $H = \mathcal{H}/a \simeq 1/(|\eta|a)$ we then obtain $a \sim 1/H_i|\eta|$. (2.73) with the Planck mass defined by $8\pi G = M_P^{-2}$, then gives

$$\Psi = \frac{H_i}{2M_P} u \simeq \frac{H_i}{M_P} k^{-3/2} (k|\eta|)^{1/2} H_\mu^{(2)}(k\eta) \ . \tag{2.154}$$

On small scales this is a simple oscillating function while on large scales $k|\eta| \ll 1$ it can be approximated by a power law,

$$\Psi \simeq \frac{H_i}{M_P} k^{-3/2} (k|\eta|)^{1+q} \simeq \frac{H_i}{M_P} k^{-3/2} \ , \quad \text{for} \quad k|\eta| \ll 1 \ . \tag{2.155}$$

Here we have used $\mu = 1/2 + q < 0$ and $q \sim -1$. This yields

$$k^3 |\Psi|^2 \simeq \left(\frac{H_i}{M_P} \right)^2 \ , \tag{2.156}$$

hence $n = 1$. Detailed studies have shown that even though the amplitude of Ψ can still be severely affected by the transition from inflation to the subsequent radiation era, the obtained spectrum is very stable. Simple deviations from de Sitter inflation (like *e.g.* power law inflation), $q > -1$ lead to slightly blue spectra, $n \gtrsim 1$.

2.7.2 Vector Perturbations

In the simplest models of inflation where the only degrees of freedom are the scalar field and the metric, no vector perturbations are generated. But even if they are, subsequent evolution after inflation will lead to their decay. In a perfect fluid background, $\Pi_{ij} = 0$, vector perturbations evolve according to (2.70) which implies

$$\Omega \propto a^{3c_s^2 - 1}. \tag{2.157}$$

For a radiation fluid, $\dot{p}/\dot{\rho} = c_s^2 \leq 1/3$, this leads to a non–growing vorticity. The dynamical Einstein equation (2.67) gives

$$\sigma^{(V)} \propto a^{-2} \ , \tag{2.158}$$

and the constraint (2.65) reads (at early times, so that we can neglect curvature)

$$\Omega \sim (k\eta)^2 \sigma^{(V)}. \tag{2.159}$$

Therefore, even if they are created in the very early universe on super–horizon scales during an inflationary period, vector perturbations of the metric decay

and become soon entirely negligible. Even if Ω_i remains constant in a radiation dominated universe, it has to be so small on relevant scales at formation ($k\eta_{in} \ll 1$) that we may safely neglect it.

Vector perturbations are irrelevant if perturbations have been created at some early time, *e.g.* during inflation. This result changes completely when considering 'active perturbations' like for example topological defects where vector perturbations contribute significantly to the CMB anisotropies on large scales, see [13]. Furthermore, it is interesting to note that vector perturbations do not satisfy a wave equation and therefore will in no case show oscillations. Vorticity simply decays with time.

2.7.3 Tensor Perturbations

The situation is different for tensor perturbations. Again we consider the perfect fluid case, $\Pi_{ij}^{(T)} = 0$. Equation (2.68) implies, if κ is negligible,

$$\ddot{H}_{ij} + \frac{2\dot{a}}{a}\dot{H}_{ij} + k^2 H_{ij} = 0 \ . \tag{2.160}$$

If the background has a power law evolution, $a \propto \eta^q$ this equation can be solved in terms of Bessel or Hankel functions. The less decaying mode solution to (2.160) is $H_{ij} = e_{ij} x^{1/2-\beta} J_{1/2-q}(x)$, where J_ν denotes the Bessel function of order ν, $x = k\eta$ and e_{ij} is a transverse traceless polarization tensor. This leads to

$$H_{ij} = \text{const.} \qquad \text{for } x \ll 1 \tag{2.161}$$

$$H_{ij} = \frac{1}{a} \qquad \text{for } x \gtrsim 1 \ . \tag{2.162}$$

One may also quantize the tensor fluctuations which represent gravitons. Doing this, one obtains (up to log corrections a scale invariant spectrum of tensor fluctuations from inflation: For tensor perturbations the canonical variable is simply given by $h_{ij} = M_P a H_{ij}$. The evolution equation for $h_{ij} = h e_{ij}$ is of the form

$$\ddot{h} + (k^2 + m^2(\eta))h = 0 \ , \tag{2.163}$$

where $m^2(\eta) = -\ddot{a}/a$. During inflation $m^2 = -q(q-1)$ is negative, leading to particle creation. Like for scalar perturbations, the vacuum initial conditions are given on scales which are inside the horizon $k^2 \gg |m^2|$,

$$h_{\rm in} = \frac{1}{\sqrt{k}} \exp(k\eta) \quad \text{for} \quad k|\eta| \gg 1.$$

Solving (2.163) with this initial condition, gives

$$h = \frac{1}{\sqrt{k}} (k|\eta|)^{1/2} H^{(2)}_{q-1/2}(k\eta) \ ,$$

where $H_\nu^{(2)}$ is the Hankel function of degree ν of the second kind. On super horizon scales, $H^{(2)}_{q-1/2}(k\eta) \propto (k|\eta|)^{q-1/2}$ this results in $|h|^2 \simeq |\eta|(k|\eta|)^{2q-1}$. Using the relation between $h_{ij} = he_{ij}$ and H_{ij} one obtains the spectrum of tensor perturbation generated during inflation. For exponential inflation, $q \simeq -1$ one finds again a scale invariant spectrum for H_{ij} on super-horizon scales

$$k^3|H_{ij}H^{ij}| \simeq (H_{\rm in}/M_P)^2 \propto k^{n_T} \quad \text{with } n_T \simeq 0 \ . \tag{2.164}$$

2.8 Lightlike Geodesics and CMB Anisotropies

After decoupling, $\eta > \eta_{\rm dec}$, photons follow to a good approximation light-like geodesics. The temperature shift of a Planck distribution of photons is equal to the energy shift of any given photon, which is independent of the photon energy (gravity is 'achromatic').

The unperturbed photon trajectory follows

$$(x^\mu(\eta)) \equiv (\eta, \int_\eta^{\eta_0} \mathbf{n}(\eta')d\eta' + \mathbf{x}_0) \ ,$$

where $\mathbf{x}_0$ is the photon position at time η_0 and $\mathbf{n}$ is the (parallel transported) photon direction. With respect to a geodesic basis $(\mathbf{e})_{i=1}^3$, the components of $\mathbf{n}$ are constant. If $\kappa = 0$ we may choose $\mathbf{e}_i = \partial/\partial x^i$; if $\kappa \neq 0$ these vector fields are no longer parallel transported and therefore do not form a geodesic basis ($\nabla_{\mathbf{e}_i}\mathbf{e}_j = 0$).

Our metric is of the form $d\tilde{s}^2 = a^2 ds^2$, with

$$ds^2 = (\gamma_{\mu\nu} + h_{\mu\nu})\, dx^\mu dx^\nu, \quad \gamma_{00} = -1,\, \gamma_{i0} = 0,\, \gamma_{ij} = \gamma_{ji}, \tag{2.165}$$

as before.

We make use of the fact that light-like geodesics are conformally invariant. More precisely, ds^2 and $d\tilde{s}^2$ have the same light-like geodesics, only the corresponding affine parameters are different. Let us denote the two affine parameters by λ and $\tilde{\lambda}$ respectively, and the tangent vectors to the geodesic by

$$n = \frac{dx}{d\lambda}, \quad \tilde{n} = \frac{1}{a}n = \frac{dx}{d\tilde{\lambda}} \ , \quad n^2 = \tilde{n}^2 = 0 \ , \ \ n^0 = 1 \ , \ \ \mathbf{n}^2 = 1. \tag{2.166}$$

We set $n^0 = 1 + \delta n^0$. The geodesic equation for the perturbed metric

$$ds^2 = (\gamma_{\mu\nu} + h_{\mu\nu})dx^\mu dx^\nu \tag{2.167}$$

yields, to first order,

$$\frac{d}{d\lambda}\delta n^\mu = -\delta\Gamma^\mu_{\alpha\beta} n^\alpha n^\beta. \tag{2.168}$$

For the energy shift, we have to determine δn^0. Since $g^{0\mu} = -1 \cdot \delta_{0\mu} +$ first order, we obtain $\delta\Gamma^0_{\alpha\beta} = -1/2(h_{\alpha 0|\beta} + h_{\beta 0|\alpha} - \dot{h}_{\alpha\beta})$, so that

$$\frac{d}{d\lambda}\delta n^0 = h_{\alpha 0|\beta} n^\beta n^\alpha - \frac{1}{2}\dot{h}_{\alpha\beta} n^\alpha n^\beta. \tag{2.169}$$

Integrating this equation we use $h_{\alpha 0|\beta} n^\beta = \frac{d}{d\lambda}(h_{\alpha 0} n^\alpha)$, so that the change of n^0 between some initial time η_i and some final time η_f is given by

$$\delta n^0|_i^f = \left[h_{00} + h_{0j} n^j\right]_i^f - \frac{1}{2}\int_i^f \dot{h}_{\mu\nu} n^\mu n^\nu d\lambda \,. \tag{2.170}$$

On the other hand, the ratio of the energy of a photon measured by some observer at t_f to the energy emitted at t_i is

$$\frac{E_f}{E_i} = \frac{(\tilde{n}\cdot u)_f}{(\tilde{n}\cdot u)_i} = \frac{T_f}{T_i}\frac{(n\cdot u)_f}{(n\cdot u)_i} \,, \tag{2.171}$$

where u_f and u_i are the four-velocities of the observer and emitter respectively, and the factor T_f/T_i is the usual (unperturbed) redshift, which relates n and $\tilde{n}$. The velocity field of observer and emitter is given by

$$u = (1-A)\partial_\eta + v^i \partial_i \,. \tag{2.172}$$

An observer measuring a temperature T_0 receives photons that were emitted at the time η_{dec} of decoupling of matter and radiation, at the fixed temperature T_{dec}. In first order perturbation theory, we find the following relation between the unperturbed temperatures T_f, T_i, the measurable temperatures $T_0 = T_f + \delta T_f$, $T_{dec} = T_i + \delta T_i$, and the photon density perturbation:

$$\frac{T_f}{T_i} = \frac{T_0}{T_{dec}}\left(1 - \frac{\delta T_f}{T_f} + \frac{\delta T_i}{T_i}\right) = \frac{T_0}{T_{dec}}\left(1 - \frac{1}{4}\delta^{(r)}|_i^f\right) \,, \tag{2.173}$$

where $\delta^{(r)}$ is the intrinsic density perturbation in the radiation and we have used $\rho^{(r)} \propto T^4$ in the last equality. Inserting the above equation and (2.170) into (2.171), and using (2.29) for the definition of $h_{\mu\nu}$, one finds, after integration by parts [7] the following result for scalar perturbations:

$$\frac{E_f}{E_i} = \frac{T_0}{T_{dec}}\left\{1 - \left[\frac{1}{4}D_g^{(r)} + V_j^{(b)} n^j + \Psi + \Phi\right]_i^f + \int_i^f (\dot{\Psi} + \dot{\Phi}) d\lambda\right\} \,. \tag{2.174}$$

Here $D_g^{(r)}$ denotes the density perturbation in the radiation fluid, and $V^{(b)}$ is the peculiar velocity of the baryonic matter component (the emitter and observer of radiation).

Evaluating (2.174) at final time η_0 (today) and initial time η_{dec}, we obtain the temperature difference of photons coming from different directions $\mathbf{n}$ and $\mathbf{n}'$

$$\frac{\Delta T}{T} \equiv \frac{\Delta T(\mathbf{n})}{T} - \frac{\Delta T(\mathbf{n}')}{T} \equiv \frac{E_f}{E_i}(\mathbf{n}) - \frac{E_f}{E_i}(\mathbf{n}'). \tag{2.175}$$

Direction independent contributions to $\frac{E_f}{E_i}$ do not contribute to this difference. We also do not want to include the term $V_j(\eta_0)n^j$ which simply describes the dipole due to our motion with respect to the emission surface and which is much larger than the contributions from the higher multipoles. Therefore we can set

$$\frac{\Delta T(\mathbf{n})}{T} = \left[\frac{1}{4}D_g^{(r)} + V_j^{(b)}n^j + \Psi + \Phi\right](\eta_{dec}, \mathbf{x}_{dec}) + \int_{\eta_{dec}}^{\eta_0} (\dot{\Psi} + \dot{\Phi})(\eta, \mathbf{x}(\eta))d\eta \ , \tag{2.176}$$

where $\mathbf{x}(\eta)$ is the unperturbed photon position at time η for an observer at $\mathbf{x}_0$, and $\mathbf{x}_{dec} = \mathbf{x}(\eta_{dec})$ (If $\kappa = 0$ we simply have $\mathbf{x}(\eta) = \mathbf{x}_0 - (\eta_0 - \eta)\mathbf{n}$.). The first term in (2.176) is the one we have discussed in the previous chapter. It describes the intrinsic inhomogeneities on the surface of last scattering, due to acoustic oscillations prior to decoupling. Depending on the initial conditions, it can contribute significantly also on super-horizon scales. This is especially important in the case of adiabatic initial conditions. As we have seen in (2.125), in a dust + radiation universe with $\Omega = 1$, adiabatic initial conditions imply $D_g^{(r)}(k,\eta) = -\frac{20}{3}\Psi(k,\eta)$ and $V^{(b)} = V^{(r)} \ll D_g^{(r)}$ for $k\eta \ll 1$. With $\Phi = \Psi$ the the square bracket of (2.176) therefore gives for adiabatic perturbations

$$\left(\frac{\Delta T(\mathbf{n})}{T}\right)^{(OSW)}_{\text{adiabatic}} = \frac{1}{3}\Psi(\eta_{dec}, \mathbf{x}_{dec})$$

on super-horizon scales. The contribution to $\frac{\delta T}{T}$ from the last scattering surface on very large scales is called the 'ordinary Sachs–Wolfe effect' (OSW). It has been derived for the first time by Sachs and Wolfe [14] in 1967. For isocurvature perturbations, the initial condition $D_g^{(r)}(k,\eta) \to 0$ for $\eta \to 0$ is satisfied and the contribution of D_g to the ordinary Sachs–Wolfe effect can be neglected

$$\left(\frac{\Delta T(\mathbf{n})}{T}\right)^{(OSW)}_{\text{isocurvature}} = 2\Psi(\eta_{dec}, \mathbf{x}_{dec}) \ .$$

The second term in (2.176) describes the relative motion of emitter and observer. This is the Doppler contribution to the CMB anisotropies. It appears on the same angular scales as the acoustic term; we call the sum of the acoustic and Doppler contributions "acoustic peaks".

The last two terms are due to the inhomogeneities in the spacetime geometry; the first contribution determines the change in the photon energy due to the difference of the gravitational potential at the position of emitter and observer. Together with the part contained in $D_g^{(r)}$ they represent the "ordinary" Sachs-Wolfe effect. The integral accounts for red-shift or blue-shift caused by the time dependence of the gravitational field along the path of the photon, and represents the so-called integrated Sachs-Wolfe (ISW) effect. In a $\Omega = 1$, pure dust universe, the Bardeen potentials are constant and there is no integrated Sachs-Wolfe effect; the blue-shift which the photons

acquire by falling into a gravitational potential is exactly cancelled by the redshift induced by climbing out of it. This is no longer true in a universe with substantial radiation contribution, curvature or a cosmological constant.

The sum of the ordinary Sachs–Wolfe term and the integral is the full Sachs-Wolfe contribution (SW).

For **vector** perturbations $\delta^{(r)}$ and A vanish and (2.171) leads to

$$(E_f/E_i)^{(V)} = (a_i/a_f)[1 - V_j^{(m)} n^j|_i^f + \int_i^f \dot{\sigma}_j n^j d\lambda] \ . \tag{2.177}$$

We obtain a Doppler term and a gravitational contribution. For **tensor** perturbations, i.e. gravitational waves, only the gravitational part remains:

$$(E_f/E_i)^{(T)} = (a_i/a_f)[1 - \int_i^f \dot{H}_{lj} n^l n^j d\lambda] \ . \tag{2.178}$$

Equations (2.174), (2.177) and (2.178) are the manifestly gauge invariant results for the energy shift of photons due to scalar, vector and tensor perturbations. Disregarding again the dipole contribution due to our proper motion, (2.177,2.178) imply the vector and tensor temperature fluctuations

$$\left(\frac{\Delta T(\mathbf{n})}{T}\right)^{(V)} = V_j^{(m)}(\eta_{dec}, \mathbf{x}_{dec}) n^j + \int_i^f \dot{\sigma}_j(\eta, \mathbf{x}(\eta)) n^j d\lambda \tag{2.179}$$

$$\left(\frac{\Delta T(\mathbf{n})}{T}\right)^{(T)} = -\int_i^f \dot{H}_{lj}(\eta, \mathbf{x}(\eta)) n^l n^j d\lambda \ . \tag{2.180}$$

Note that for models where initial fluctuations have been laid down in the very early universe, vector perturbations are irrelevant as we have already pointed out. In this sense (2.179) is here mainly for completeness. However, in models where perturbations are sourced by some inherently inhomogeneous component (*e.g.* topological defects, see [13]) vector perturbation can be important.

2.9 Power Spectra

One of the basic tools to compare models of large scale structure having stochastic initial fluctuations with observations are power spectra. They are the "harmonic transforms" of the two point correlation functions. If the perturbations of the model under consideration are Gaussian (a relatively generic prediction from inflationary models), then the power spectra contain the full statistical information of the model.

Let us first consider the power spectrum of dark matter,

$$P_D(k) = \left\langle \left| D_g^{(m)}(\mathbf{k}, \eta_0) \right|^2 \right\rangle \ . \tag{2.181}$$

Here $\langle\ \rangle$ indicates a statistical average, ensemble average, over "initial conditions" in a given model. $P_D(k)$ is usually compared with the observed power spectrum of the galaxy distribution. This is clearly problematic since it is by no means evident what the ratio of these two spectra should be. This problem is known under the name of 'biasing' and it is very often simply assumed that the dark matter and galaxy power spectra differ only by a constant factor. The hope is also that on sufficiently large scales, since the evolution of both, galaxies and dark matter is governed by gravity, their power spectra should not differ too much. This hope seems to be reasonably justified [15].

The power spectrum of velocity perturbations satisfies the relation

$$P_V(k) = \left\langle |\mathbf{V}(\mathbf{k}, \eta_0)|^2 \right\rangle \simeq H_0^2 \Omega^{1.2} P_D(k) k^{-2} \ . \tag{2.182}$$

For $\simeq$ we have used that $|kV|(\eta_0) = \dot{D}_g^{(m)}(\eta_0) \sim H_0 \Omega^{0.6} D_g$ on sub-horizon scales (see *e.g.* [16]).

The spectrum we can be both, measured and calculated to the best accuracy is the CMB anisotropy power spectrum. It is defined as follows: $\Delta T/T$ is a function of position $\mathbf{x}_0$, time η_0 and photon direction $\mathbf{n}$. We develop the $\mathbf{n}$–dependence in terms of spherical harmonics. We will suppress the argument η_0 and often also $\mathbf{x}_0$ in the following calculations. All results are for today (η_0) and here ($\mathbf{x}_0$). By statistical homogeneity statistical averages over an ensemble of realisations (expectation values) are supposed to be independent of position. Furthermore, we assume that the process generating the initial perturbations is statistically isotropic. Then, the off-diagonal correlators of the expansion coefficients $a_{\ell m}$ vanish and we have

$$\frac{\Delta T}{T}(\mathbf{x}_0, \mathbf{n}, \eta_0) = \sum_{\ell,m} a_{\ell m}(\mathbf{x}_0) Y_{\ell m}(\mathbf{n}), \quad \langle a_{\ell m} \cdot a^*_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m m'} C_\ell. \tag{2.183}$$

The C_ℓ's are the CMB power spectrum.

The two point correlation function is related to the C_ℓ's by

$$\left\langle \frac{\Delta T}{T}(\mathbf{n}) \frac{\Delta T}{T}(\mathbf{n}') \right\rangle_{\mathbf{n}\cdot\mathbf{n}'=\mu} = \sum_{\ell,\ell',m,m'} \langle a_{\ell m} \cdot a^*_{\ell' m'} \rangle\, Y_{\ell m}(\mathbf{n}) Y^*_{\ell' m'}(\mathbf{n}') =$$

$$\sum_\ell C_\ell \underbrace{\sum_{m=-\ell}^{\ell} Y_{\ell m}(\mathbf{n}) Y^*_{\ell m}(\mathbf{n}')}_{\frac{2\ell+1}{4\pi} P_\ell(\mathbf{n}\cdot\mathbf{n}')} = \frac{1}{4\pi} \sum_\ell (2\ell+1) C_\ell P_\ell(\mu) \ , \tag{2.184}$$

where we have used the addition theorem of spherical harmonics for the last equality; the P_ℓ's are the Legendre polynomials.

Clearly the a_{lm}'s from scalar, vector and tensor perturbations are uncorrelated,

$$\left\langle a^{(S)}_{\ell m} a^{(V)}_{\ell' m'} \right\rangle = \left\langle a^{(S)}_{\ell m} a^{(T)}_{\ell' m'} \right\rangle = \left\langle a^{(V)}_{\ell m} a^{(T)}_{\ell' m'} \right\rangle = 0 \ . \tag{2.185}$$

Since vector perturbations decay, their contributions, the $C_\ell^{(V)}$, are negligible in models where initial perturbations have been laid down very early, *e.g.* , after an inflationary period. Tensor perturbations are constant on super-horizon scales and perform damped oscillations once they enter the horizon.

Let us first discuss in somewhat more detail scalar perturbations. We specialize to the case $\kappa = 0$ for simplicity. We suppose the initial perturbations to be given by a spectrum,

$$\left\langle |\Psi|^2 \right\rangle k^3 = A^2 k^{n-1} \eta_0^{n-1}. \tag{2.186}$$

We multiply by the constant η_0^{n-1}, the actual comoving size of the horizon, in order to keep A dimensionless for all values of n. A then represents the amplitude of metric perturbations at horizon scale today, $k = 1/\eta_0$.

On *super-horizon scales* we have, for *adiabatic* perturbations:

$$\frac{1}{4} D_g^{(r)} = -\frac{5}{3}\Psi + \mathcal{O}((k\eta)^2), \quad V^{(b)} = V^{(r)} = \mathcal{O}(k\eta) \ . \tag{2.187}$$

The dominant contribution on super-horizon scales (neglecting the integrated Sachs–Wolfe effect $\int \dot\Phi - \dot\Psi$) is then

$$\frac{\Delta T}{T}(\mathbf{x}_0, \mathbf{n}, \eta_0) = \frac{1}{3}\Psi(x_{\rm dec}, \eta_{\rm dec}). \tag{2.188}$$

The Fourier transform of (2.188) gives

$$\frac{\Delta T}{T}(\mathbf{k}, \mathbf{n}, \eta_0) = \frac{1}{3}\Psi(k, \eta_{\rm dec})\, e^{i\mathbf{kn}(\eta_0 - \eta_{\rm dec})} \ . \tag{2.189}$$

Using the decomposition

$$e^{i\mathbf{kn}(\eta_0 - \eta_{\rm dec})} = \sum_{\ell=0}^{\infty} (2\ell+1) i^\ell j_\ell(k(\eta_0 - \eta_{\rm dec})) P_\ell(\widehat{\mathbf{k}} \cdot \mathbf{n}) \ ,$$

where j_ℓ are the spherical Bessel functions, we obtain

$$\begin{aligned}
&\left\langle \frac{\Delta T}{T}(\mathbf{x}_0, \mathbf{n}, \eta_0) \frac{\Delta T}{T}(\mathbf{x}_0, \mathbf{n}', \eta_0) \right\rangle \qquad (2.190)\\
&= \frac{1}{V}\int d^3x_0 \left\langle \frac{\Delta T}{T}(\mathbf{x}_0, \mathbf{n}, \eta_0) \frac{\Delta T}{T}(\mathbf{x}_0, \mathbf{n}', \eta_0) \right\rangle \\
&= \frac{1}{(2\pi)^3}\int d^3k \left\langle \frac{\Delta T}{T}(\mathbf{k}, \mathbf{n}, \eta_0) \left(\frac{\Delta T}{T}\right)^*(\mathbf{k}, \mathbf{n}', \eta_0) \right\rangle \\
&= \frac{1}{(2\pi)^3 9}\int d^3k \left\langle |\Psi|^2 \right\rangle \sum_{\ell,\ell'=0}^{\infty} (2\ell+1)(2\ell'+1) i^{\ell-\ell'} \\
&\quad \cdot j_\ell(k(\eta_0 - \eta_{\rm dec})) j_{\ell'}(k(\eta_0 - \eta_{\rm dec})) P_\ell(\hat{\mathbf{k}} \cdot \mathbf{n}) \cdot P_{\ell'}(\hat{\mathbf{k}} \cdot \mathbf{n}') \ . \qquad (2.191)
\end{aligned}$$

In the second equal sign we have used the unitarity of the Fourier transformation. Inserting

$$P_\ell(\hat{\mathbf{k}}\mathbf{n}) = \frac{4\pi}{2\ell+1}\sum_m Y^*_{\ell m}(\hat{\mathbf{k}})Y_{\ell m}(\mathbf{n})$$
$$P_{\ell'}(\hat{\mathbf{k}}\mathbf{n}') = \frac{4\pi}{2\ell'+1}\sum_{m'} Y^*_{\ell' m'}(\hat{\mathbf{k}})Y_{\ell' m'}(\mathbf{n}') \ ,$$

integration over the directions $d\Omega_{\hat{k}}$ gives $\delta_{\ell\ell'}\delta_{mm'}\sum_m Y^*_{\ell m}(\mathbf{n})Y_{\ell m}(\mathbf{n}')$. Using as well $\sum_m Y^*_{\ell m}(\mathbf{n})Y_{\ell m}(\mathbf{n}') = \frac{2\ell+1}{4\pi}P_\ell(\mu)$, where $\mu = \mathbf{n}\cdot\mathbf{n}'$, we find

$$\left\langle \frac{\Delta T}{T}(\mathbf{x}_0,\mathbf{n},\eta_0)\frac{\Delta T}{T}(\mathbf{x}_0,\mathbf{n}',\eta_0)\right\rangle_{\mathbf{n}\mathbf{n}'=\mu} =$$
$$\sum_\ell \frac{2\ell+1}{4\pi}P_\ell(\mu)\frac{2}{\pi}\int\frac{dk}{k}\left\langle\frac{1}{9}|\Psi|^2\right\rangle k^3 j_\ell^2(k(\eta_0-\eta_{\rm dec})). \qquad (2.192)$$

Comparing this equation with (2.184) we obtain for *adiabatic perturbations* on scales $2 \le \ell \ll \chi(\eta_0-\eta_{\rm dec})/\eta_{\rm dec} \sim 100$

$$C_\ell^{(SW)} \simeq C_\ell^{(OSW)} \simeq \frac{2}{\pi}\int_0^\infty \frac{dk}{k}\left\langle\left|\frac{1}{3}\Psi\right|^2\right\rangle k^3 j_\ell^2\left(k\left(\eta_0-\eta_{\rm dec}\right)\right). \qquad (2.193)$$

If Ψ is a pure power law as in (2.186) and we set $k(\eta_0-\eta_{\rm dec}) \sim k\eta_0$, the integral (2.193) can be performed analytically. For the ansatz (2.186) one finds

$$C_\ell^{(SW)} = \frac{A^2}{9}\frac{\Gamma(3-n)\Gamma(\ell-\frac{1}{2}+\frac{n}{2})}{2^{3-n}\Gamma^2(2-\frac{n}{2})\Gamma(\ell+\frac{5}{2}-\frac{n}{2})} \quad \text{for } -3<n<3\ . \qquad (2.194)$$

Of special interest is the *scale invariant* or Harrison–Zel'dovich spectrum, $n=1$ (see Sect. 2.7). It leads to

$$\ell(\ell+1)C_\ell^{(SW)} = \text{const.} \simeq \left\langle\left(\frac{\Delta T}{T}(\vartheta_\ell)\right)^2\right\rangle\ , \quad \vartheta_\ell \equiv \pi/\ell\ . \qquad (2.195)$$

This is precisely (within the accuracy of the experiment) the behavior observed by the DMR experiment aboard the satellite COBE [17], as shown in Fig. 2.1.

Inflationary models predict very generically a HZ spectrum (up to small corrections). The DMR discovery has therefore been regarded as a great success, if not a proof, of inflation. There are other models like topological defects [18, 19, 20] or certain string cosmology models [21] which also predict scale–invariant, *i.e.* Harrison Zel'dovich spectra of fluctuations. These models do however not belong to the class investigated here, since in these models perturbations are induced by seeds which evolve non–linearly in time.

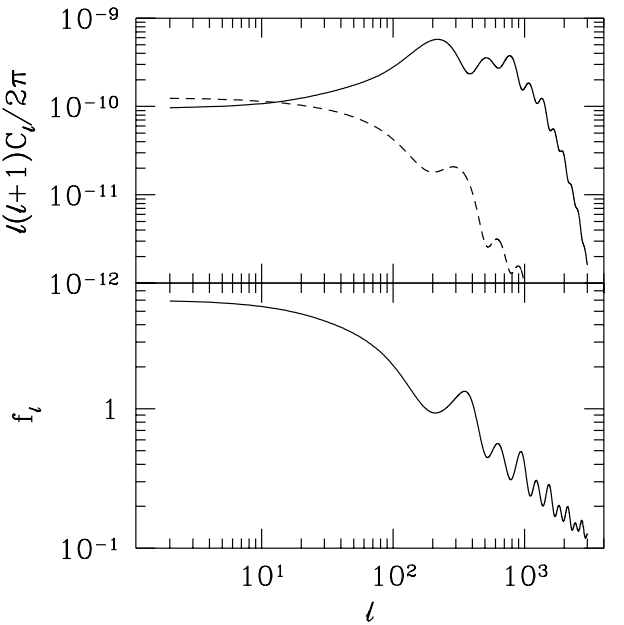

Fig. 2.1. A COBE normalized sample adiabatic (solid line) and isocurvature (dashed line) CMB anisotropy spectrum, $\ell(\ell+1)C_\ell$, are shown on the top panel. The quantity shown in the bottom panel is the ratio of temperature fluctuations for fixed value of A (From Kanazawa et al. [22]).

For isocurvature perturbations, the main contribution on large scales comes from the integrated Sachs–Wolfe effect and (2.193) is replaced by

$$C_\ell^{(ISW)} \simeq \frac{8}{\pi}\int \frac{dk}{k}k^3 \left\langle \left| \int_{\eta_{\rm dec}}^{\eta_0} \dot\Psi(k,\eta) j_\ell^2(k(\eta_0-\eta))d\eta \right|^2 \right\rangle . \qquad (2.196)$$

Inside the horizon Ψ is roughly constant (matter dominated). Using the ansatz (2.186) for Ψ inside the horizon and setting the integral in (2.196) $\sim 2\Psi(k,\eta=1/k)j_\ell^2(k\eta_0)$, we obtain again (2.194), but with $A^2/9$ replaced by $4A^2$. For a fixed amplitude A of perturbations, the Sachs–Wolfe temperature anisotropies coming from isocurvature perturbations are therefore about a factor of 6 times larger than those coming from adiabatic perturbations.

On smaller scales, $\ell \gtrsim 100$ the contribution to $\Delta T/T$ is usually dominated by acoustic oscillations, the first two terms in (2.176). Instead of (2.196) we then obtain

$$C_\ell^{(AC)} \simeq \frac{2}{\pi}\int_0^\infty \frac{dk}{k}k^3 \left\langle \left| \frac{1}{4}D^{(r)}(k,\eta_{\rm dec})j_\ell(k\eta_0) + V^{(r)}(k,\eta_{\rm dec})j_\ell'(k\eta_0) \right|^2 \right\rangle \qquad (2.197)$$

To remove the SW contribution from $D_g^{(r)}$ we have simply replaced it by $D^{(r)}$ which is much smaller than Ψ on super-horizon scales and therefore does not contribute to the SW terms. On sub-horizon scales $D^{(r)} \simeq D_g^{(r)}$ and $V^{(r)}$ are oscillating like sine or cosine waves depending on the initial conditions. Correspondingly the $C_\ell^{(AC)}$ will show peaks and minima. On very small scales they are damped by the photon diffusion which takes place during the recombination process (see A. Challinor's contribution).

For gravitational waves (tensor fluctuations), a formula analogous to (2.194) can be derived,

$$C_\ell^{(T)} = \frac{2}{\pi} \int dk k^2 \left\langle \left| \int_{\eta_{\rm dec}}^{\eta_0} d\eta \dot{H}(\eta, k) \frac{j_\ell(k(\eta_0 - \eta))}{(k(\eta_0 - \eta))^2} \right|^2 \right\rangle \frac{(\ell+2)!}{(\ell-2)!}. \tag{2.198}$$

To a very crude approximation we may assume $\dot{H} = 0$ on super-horizon scales and $\int d\eta \dot{H} j_\ell(k(\eta_0 - \eta)) \sim H(\eta = 1/k) j_\ell(k\eta_0)$. For a pure power law,

$$k^3 \left\langle |H(k, \eta = 1/k)|^2 \right\rangle = A_T^2 k^{n_T} \eta_0^{-n_T}, \tag{2.199}$$

one obtains

$$\begin{aligned} C_\ell^{(T)} &\simeq \frac{2}{\pi} \frac{(\ell+2)!}{(\ell-2)!} A_T^2 \int \frac{dx}{x} x^{n_T} \frac{j_\ell^2(x)}{x^4} \\ &= \frac{(\ell+2)!}{(\ell-2)!} A_T^2 \frac{\Gamma(6 - n_T)\Gamma(\ell - 2 + \frac{n_T}{2})}{2^{6-n_T}\Gamma^2(\frac{7}{2} - n_T)\Gamma(\ell + 4 - \frac{n_T}{2})} \ . \end{aligned} \tag{2.200}$$

For a scale invariant spectrum ($n_T = 0$) this results in

$$\ell(\ell+1) C_\ell^{(T)} \simeq \frac{\ell(\ell+1)}{(\ell+3)(\ell-2)} A_T^2 \frac{8}{15\pi} \ . \tag{2.201}$$

The singularity at $\ell = 2$ in this crude approximation is not real, but there is some enhancement of $\ell(\ell+1)C_\ell^{(T)}$ at $\ell \sim 2$ (see Fig. 2.2).

Since tensor perturbations decay on sub-horizon scales, $\ell \gtrsim 60$, they are not very sensitive to cosmological parameters.

Again, inflationary models (and topological defects) predict a scale invariant spectrum of tensor fluctuations ($n_T \sim 0$).

On very small angular scales, $\ell \gtrsim 800$, fluctuations are damped by collisional damping (Silk damping). This effect has to be discussed with the Boltzmann equation for photons which is presented in detail by A. Challinor in his contribution to this volume.

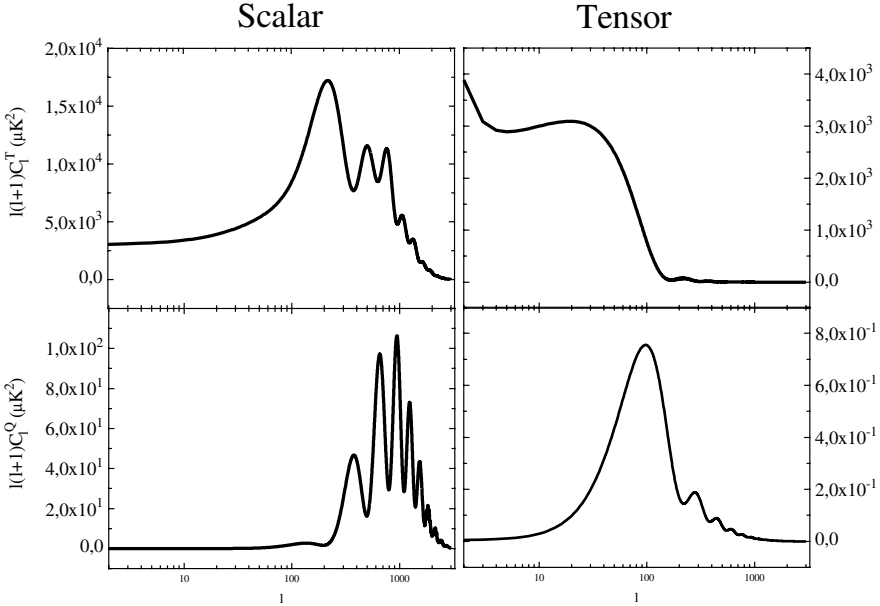

Fig. 2.2. Adiabatic scalar and tensor CMB anisotropy spectra are shown (top panels). The bottom panels show the corresponding polarization spectra (see A. Challinor's contribution). (From [23]).

2.10 Some Remarks on Perturbation Theory in Braneworlds

Since there has been so much interest in them recently, let me finally make some remarks on perturbation theory of five dimensional braneworlds. I shall just present some relatively simple aspects without derivation. A thorough discussion of braneworlds is given by R. Maartens in this volume. Different aspects of the perturbation theory of braneworlds can be found in the growing literature on the subject [24, 25].

The bulk background metric of a five dimensional braneworld has three dimensional spatial slices which are homogeneous and isotropic, hence spaces of constant curvature. Its line element is therefore of the form

$$ds^2 = -n^2(t,y)dt^2 + a^2(t,y)\gamma_{ij}dx^i dx^j + b^2(t,y)dy^2 = g_{AB}dx^A dx^B. \quad (2.202)$$

Perturbations of such a spacetime can be decomposed into scalar, vector and tensor modes with respect to the three dimensional spatial slices of constant curvature. One can always choose the so-called **generalized longitudinal gauge** such that the perturbations of the metric are given as follows:

$$\begin{aligned} ds^2 = &-n^2(1+2\Psi)dt^2 + a^2\left[(1-2\Phi)\gamma_{ij} + 2H_{ij}\right]dx^i dx^j + b^2(1+2C)dy^2 \\ &-2nbBdtdy - 2na\Sigma_i dx^i dt + ab2\mathcal{E}_i dx^i dy\ . \end{aligned} \quad (2.203)$$

Here Σ_i and $\mathcal{E}_i$ are divergence free vector fields, vector perturbations and H_{ij} is a divergence free traceless symmetric tensor, the tensor perturbation, while Ψ, Φ, C and B are four scalar perturbations.

It can be shown that this choice determines the gauge completely. One can actually define gauge invariant perturbation variables which reduce to the ones above in the generalized longitudinal gauge [25]. Writing down the perturbed Einstein equations for these variables in the most general case is quite involved. These equations can be found in [25], but I don't want to repeat them here. I just discuss their general structure in the case of an empty bulk. It is clear that Ψ and Φ correspond to the Bardeen potentials of four dimensional cosmology, Σ_i is the four dimensional vector perturbation and H_{ij} represents four dimensional gravitational waves. C and B as well as $\mathcal{E}_i$ are new degrees of freedom which are not present in the four dimensional theory.

If we assume vanishing perturbations of the bulk energy momentum tensor e.g. if the bulk is anti de Sitter like in the Randall Sundrum model [26] called RSII in what follows (see R. Maartens' contribution), the perturbation equations reduce to

$$\Box_5(\Psi + \Phi) = 0 \tag{2.204}$$

$$\Box_5 \Sigma_i = 0 \tag{2.205}$$

$$\Box_5 H_{ij} = 0 \ , \tag{2.206}$$

and all the other perturbation variables are determined by constraint equations. Here $\Box_5$ is the five dimensional d'Alembertian with respect to the background metric (2.202). This structure of the equations is to be expected: Gravitational waves in d spacetime dimensions are a spin 2 field with respect to the group of rotations $SO(d-2)$ since they are massless (see e.g. [27]). For $d = 5$, $d - 2 = 3$ they therefore have 5 degrees of freedom. These correspond exactly to the one scalar (2.204), two vector (2.205) and two tensor (2.206) degrees of freedom with respect to the 3-dimensional slices of constant curvature. These free massless degrees of freedom obey the wave equations above.

The perturbed Israel junction conditions (see R. Maartens' contribution) then determine boundary conditions for the behaviour of the perturbations at the brane position(s).

As an example, I write down the vector and tensor perturbation equations and their bulk solutions for the RSII model which has only one brane. There the bulk is a five dimensional anti-de Sitter spacetime, and we can choose coordinates so that the background metric has the form

$$ds^2 = \left(\frac{L}{y}\right)^2 \left[-dt^2 + \delta_{ij} dx^i dx^j + dy^2\right] \tag{2.207}$$

and

$$\Box_5 = -\partial_t^2 + \triangle + \partial_y^2 - \frac{3}{y}\partial_y \ , \tag{2.208}$$

where $\triangle$ denotes the three dimensional spatial Laplacian. For an arbitrary mode which satisfies this wave equation we make the ansatz

$$\phi(t, \mathbf{x}, y) = \exp(i(\mathbf{k}\cdot\mathbf{x} - \omega t))\phi(\omega, \mathbf{k}, y) \ .$$

We then obtain a Bessel differential equation for $\phi(\omega, \mathbf{k}, y)$ with general solution

$$\begin{aligned} \phi &= A(\omega, \mathbf{k})(ym)^2 J_2(my) + B(\omega, \mathbf{k})(ym)^2 Y_2(my) \ , \quad m^2 = \omega^2 - k^2 \ , \\ &= \phi_A + \phi_B \ . \end{aligned} \tag{2.209}$$

These modes are normalizable in the sense that

$$\int_{y_b}^{\infty} |\phi|^2 \sqrt{-g} dy < \infty \ .$$

Here $y_b > 0$ is the brane position. The Israel junction condition for the vector and tensor modes for RSII become

$$-\partial_y H_{ij}(y = y_b) = \kappa_5^2 \Pi_{ij}(y_b) \tag{2.210}$$

$$-\Sigma_i(y_b) = \kappa_5^2 \Pi_i(y_b) \ , \tag{2.211}$$

where Π_{ij} respectively Π_i are the tensor respectively vector contribution to the anisotropic stresses on the brane. If the latter vanish, the junction condition simple requires

$$B(\omega, \mathbf{k}) = A(\omega, \mathbf{k}) J_2(my_b)/Y_2(my_b) \ .$$

This result has been derived for the tensor mode by Randall and Sundrum [26]. For scalar perturbations the situation is somewhat more complicated since there is an additional degree of freedom which is the perturbed position of the brane, $y_b \to y_b + \epsilon$. It is rather subtle to take this brane bending correctly into account. A very interesting work showing that this effect is actually most relevant to obtain the correct Newtonian limit in the RSII model can be found in [28].

Let us also briefly discuss the zero–mode. From the brane point of view, the modes discussed here represent waves (particles) which couple only to the energy momentum tensor of the brane and which obey a dispersion relation $\omega^2 - k^2 = m^2$, hence the parameter m of the solutions (2.211) is their mass. In the limit $m \to 0$ the solutions turn into power laws in y,

$$\phi_0(y) = Ay^4 + B \ . \tag{2.212}$$

Of these modes, for tensor and vector perturbations, the B mode is normalizable. For scalar perturbations the situation is more complicated. For a mode to be 'normalizable' we want all the perturbations to be normalizable,

$$\int_{y_b}^{\infty} |\phi|^2 \sqrt{-g} dy < \infty \qquad \text{for } \phi = \Phi \ , \ \Psi \ , \ C \text{ and } B \ .$$

But from the constraints one obtains for $m = 0$ $C \propto y^2$ which is not normalizable. This mode diverges logarithmically and only converges due to the oscillations of the Bessel functions if $m \neq 0$.

If all the five graviton zero–modes are normalizable, like in all compact braneworlds, e.g. in models with two branes, the vector and scalar mode lead to three additional degrees of freedom (to the usual two four-dimensional graviton modes) which couple via the junction conditions to the brane energy momentum tensor and spoil the phenomenology of the model. The resulting four dimensional gravity is not Einstein gravity but Brans–Dicke or even more complicated. As an example, in [29] the contribution from the scalar zero–mode (2.204) to the gravity wave emission from a binary pulsar is calculated and shown to be in contradiction with observations. This problem is well-known from Kaluza Klein theories, it is the so called moduli problem. The way out of it is usually to render the modes massive. There are different suggestions how this can be achieved for the scalar mode. One possibility is the so called Goldberger–Wise mechanism [30] which shows that under certain circumstances a bulk scalar field can do the job. But certainly, a physically acceptable braneworld is defined only together with its mechanism how to get rid of such unwanted modes (see also [31]).

The advantage of the RSII model is that the scalar gravity wave zero–mode is not normalizable in this model and therefore it does not contribute. This very promising property of the RSII model has let to its popularity. It seems to induce the correct four dimensional Einstein gravity on the brane, in the cosmological context this property is still maintained at sufficiently low energies.

2.11 Conclusions

In this contribution I have given an introduction to cosmological perturbation theory. Perturbation theory is an important tool especially to calculate CMB anisotropies and polarisation since these are very small and can be determined reliably within linear cosmological perturbation theory. To determine the evolution of the cosmic matter density, linear perturbation theory has to be complemented with the theory of weakly non-linear Newtonian gravity and with N-body simulations. To finally understand the formation of galaxies, non-gravitational highly non-linear physics, like heating and cooling mechanisms, dissipation, nuclear reactions etc. have to be taken into account. This very difficult subject is still in its infancy.

To make progress in our understanding of braneworlds, linear perturbation theory can also be most helpful. We can use it to determine e.g. the propagating modes of the gravitational field on the brane, light deflection

and redshift in weak gravitational fields and the Newtonian limit. The condition that linear perturbations on the brane at low energy and large distances reduce to those resulting from Einstein gravity is non–trivial and has, to my knowledge, not yet been fully explored to limit braneworld models.

Acknowledgements

I thank the organizers for a well structured school in a most beautiful environment.

References

1. I. Newton, Letters from Sir Isaac Newton to Dr. Bentley, Letter I, 203ff quoted by A. Koyré, *From the Classical World to the Infinite Universe*, Harper and Row (New York 1958).
2. E. Lifshitz, JETP **10**, 116 (1946).
3. D. Barden and C. Thomas, An *Introduction to Differential Manifolds*, Imperial College Press (2003).
4. J.M. Stewart and M. Walker, Proc. R. Soc. London **A341**, 49 (1974).
5. J. Bardeen, Phys. Rev. **D22**, 1882 (1980).
6. H. Kodama and M. Sasaki Prog. Theor. Phys. Suppl **78**, 1 (1984).
7. R. Durrer, Fund. of Cosmic Physics **15**, 209 (1994).
8. M. Bruni, P. Dunsby and G. Ellis, Astrophys. J **395**, 34 (1992).
9. R. Trotta, A. Riazuelo and R. Durrer, Phys. Rev. Lett. **87**, 231301 (2001).
10. M. Bucher, K. Moodley and N. Turok, Phys. Rev. **D 66**, 023528 (2002)
11. E. Harrison, Phys. Rev. **D1** 2726 (1970);
    Ya. B. Zel'dovich, Mont. Not. R. Astr. Soc. **160**, P1 (1972).
12. V. F. Mukhanov, H. A. Feldman and R. H. Brandenberger, Phys. Rept. **215**, 203 (1992).
13. R. Durrer, M. Kunz and A. Melchiorri, Phys. Rept. **364**, 1 (2002).
14. R.K. Sachs and A.M. Wolfe, Astrophys. J. **147**, 73 (1967).
15. M. Tegmark et al., astro-ph/0310725, (2003).
16. P.J.E. Peebles, *Principles of Physical Cosmology*, Princeton University Press (1993).
17. G.F. Smoot et al., Astrophys. J. **396**, L1 (1992).
18. U. Pen, D. Spergel and N. Turok, Phys Rev. D **49**, 692 (1994).
19. R. Durrer and Z. Zhou, Phys. Rev. D **53**, 5394 (1996).
20. B. Allen *et al.*, Phys. Rev. Lett. **79**, 2624 (1997).
21. R. Durrer *et al.*, Phys. Rev. D **59**, 043511 (1999)
22. T. Kanazawa, M. Kawasaki, N. Sugiyama, and T. Yanagida, Prog. Theor. Phys. **100** 1055 (1998).
23. A. Melchiorri and N. Vittorio, in: Proceedings of the NATO Advanced Study Institute 1996, Strasbourg, astro-ph/9610029, (1996).
24. H. Kodama, A. Ishibashi, and O. Seto, Phys. Rev. D **62**, 064022 (2000); D. Langlois, Phys. Rev. D **62**, 126012 (2000); S. Mukohyama, Phys. Rev. D **62**, 084015 (2000); D. Langlois, Phys. Rev. Lett. **86**, 2212 (2001); H. Bridgman, K. Malik, and D. Wands, Phys. Rev. D **65**, 043502 (2002); K. Koyama and J. Soda, Phys. Rev. D **65**, 023514 (2002).

25. A. Riazuelo, F. Vernizzi, D. Steer and R. Durrer, hep-th/0205220, (2002).
26. L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 4690 (1999).
27. S. Weinberg, *The Quantum Theory of Fields*, Vol I p62ff, Cambridge University Press (1995).
28. J. Garriga and T. Tanaka, Phys. Rev. Lett. **84**, 2778 (2000).
29. R. Durrer and P. Kocian, hep-th/0305181, (2003).
30. W. Goldberger and M. Wise, Phys. Lett. **B475**, 275 (2000).
31. J. Lesgourges and L. Sorbo, hep-th/0310007, (2003).

3 Cosmic Microwave Background Anisotropies

Anthony Challinor

Astrophysics Group, Cavendish Laboratory, Madingley Road, Cambridge, CB3 0HE, UK

Abstract. The linear anisotropies in the temperature of the cosmic microwave background (CMB) radiation and its polarization provide a clean picture of fluctuations in the universe some 370 kyr after the big bang. Simple physics connects these fluctuations with those present in the ultra-high-energy universe, and this makes the CMB anisotropies a powerful tool for constraining the fundamental physics that was responsible for the generation of structure. Late-time effects also leave their mark, making the CMB temperature and polarization useful probes of dark energy and the astrophysics of reionization. In this review we discuss the simple physics that processes primordial perturbations into the linear temperature and polarization anisotropies. We also describe the role of the CMB in constraining cosmological parameters, and review some of the highlights of the science extracted from recent observations and the implications of this for fundamental physics.

3.1 Introduction

The cosmic microwave background (CMB) radiation has played an essential role in shaping our current understanding of the large-scale properties of the universe. The discovery of this radiation in 1965 by Penzias and Wilson [1], and its subsequent interpretation as the relic radiation from a hot, dense phase of the universe [2] put the hot big bang model on a firm observational footing. The prediction of angular variations in the temperature of the radiation, due to the propagation of photons through an inhomogeneous universe, followed shortly after [3], but it was not until 1992 that these were finally detected by the Differential Microwave Radiometers (DMR) experiment on the Cosmic Background Explorer (COBE) satellite [4]. The fractional temperature anisotropies are at the level of 10^{-5}, consistent with structure formation in cold dark matter (CDM) models [5, 6], but much smaller than earlier predictions for baryon-dominated universes [3, 7]. Another experiment on COBE, the Far InfraRed Absolute Spectrophotometer (FIRAS), spectacularly confirmed the black-body spectrum of the CMB and determined the (isotropic) temperature to be 2.725 K [8, 9].

In the period since COBE, many experiments have mapped the CMB anisotropies on a range of angular scales from degrees to arcminutes (see [10] for a recent review), culminating in the first-year release of all-sky data from the Wilkinson Microwave Anisotropy Probe (WMAP) satellite in February

A. Challinor, Cosmic Microwave Background Anisotropies, Lect. Notes Phys. **653**, 71–103 (2005)
http://www.springerlink.com/

2003 [11]. The observed modulation in the amplitude of the anisotropies with angular scale is fully consistent with predictions based on coherent, acoustic oscillations [7], derived from gravitational instability of initially adiabatic density perturbations in a universe with nearly-flat spatial sections. The amplitude and scale of these acoustic features has allowed many of the key cosmological parameters to be determined with unprecedented precision [12], and a strong concordance with other cosmological probes has emerged.

In this review article we describe the essential physics of the temperature anisotropies of the CMB, and its recently-detected polarization [13], and discuss how these are used to constrain cosmological models. For reviews that are similar in spirit, but from the pre-WMAP era see e.g. [14, 15]. We begin in Sect. 3.2 with the fundamentals of CMB physics, presenting the kinetic theory of the CMB in an inhomogeneous universe, and the various physical mechanisms that process initial fluctuations in the distribution of matter and spacetime geometry into temperature anisotropies. Sect. 3.3 discusses the effect of cosmological parameters on the power spectrum of the temperature anisotropies, and the limits to parameter determination from the CMB alone. The physics of CMB polarization is reviewed in Sect. 3.4, and the additional information that polarization brings over temperature anisotropies alone is considered. Finally, in Sect. 3.5 we describe some of the scientific highlights that have emerged from recent CMB observations, including the detection of CMB polarization, implications for inflation, and the direct signature of dark energy through correlations between the large-scale anisotropies and tracers of the mass distribution in the local universe. Throughout, we illustrate our discussion with computations based on ΛCDM cosmologies, with baryon density $\Omega_{\rm b}h^2 = 0.023$ and cold dark matter density $\Omega_{\rm c}h^2 = 0.111$. For flat models we take the dark-energy density parameter to be $\Omega_\Lambda = 0.75$ giving a Hubble parameter $H_0 = 73\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. We adopt units with $c = 1$ throughout, and use a spacetime metric signature $+---$.

3.2 Fundamentals of CMB Physics

In this section we aim to give a reasonably self-contained review of the essential elements of CMB physics.

3.2.1 Thermal History and Recombination

The high temperature of the early universe maintained a low equilibrium fraction of neutral atoms, and a correspondingly high number density of free electrons. Coulomb scattering between the ions and electrons kept them in local kinetic equilibrium, and Thomson scattering of photons tended to maintain the isotropy of the CMB in the baryon rest frame. As the universe expanded and cooled, the dominant element hydrogen started to recombine when the temperature fell below $\sim 4000\,\mathrm{K}$ – a factor of 40 lower than might

be anticipated from the 13.6-eV inoization potential of hydrogen, due to the large ratio of the number of photons to baryons. The details of recombination are complicated since the processes that give rise to net recombination occur too slowly to maintain chemical equilibrium between the electrons, protons and atoms during the later stages of recombination [16, 17] (see [18] for recent refinements). The most important quantity for CMB anisotropy formation is the visibility function – the probability that a photon last scattered as a function of time. The visibility function peaks around $\sim 370\,\mathrm{kyr}$ after the big bang, and has a width $\sim 115\,\mathrm{kyr}$, a small fraction of the current age $\sim 13.5\,\mathrm{Gyr}$ [12]. After recombination, photons travelled mostly unimpeded through the inhomogeneous universe, imprinting fluctuations in the radiation temperature, the gravitational potentials, and the bulk velocity of the radiation where they last scattered, as the temperature anisotropies that we observe today. A small fraction of CMB photons (current results from CMB polarization measurements [19] indicate around 20 per cent; see also Sect. 3.5.1) underwent further scattering once the universe reionized due to to the ionizing flux from the first non-linear structures (see also R. Sander's review).

3.2.2 Statistics of CMB Anisotropies

The spectrum of the CMB brightness along any direction $\hat{\boldsymbol{n}}$ is very nearly thermal with a temperature $T(\hat{\boldsymbol{n}})$. The temperature depends only weakly on direction, with fluctuations $\Delta T(\hat{\boldsymbol{n}})$ at the level of 10^{-5} of the average temperature $T = 2.725\,\mathrm{K}$. It is convenient to expand the temperature fluctuation in spherical harmonics,

$$\Delta T(\hat{\boldsymbol{n}})/T = \sum_{lm} a_{lm} Y_{lm}(\hat{\boldsymbol{n}}) \,, \tag{3.1}$$

with $a^*_{lm} = (-1)^m a_{l-m}$ since the temperature is a real field. The sum in (3.1) runs over $l \geq 1$, but the dipole ($l = 1$) is usually removed explicitly when analysing data since it depends linearly on the velocity of the observer. Multipoles at l encode spatial information with characteristic angular scale $\sim \pi/l$ (see R. Durrer's contribution).

The statistical properties of the fluctuations in a perturbed cosmology can be expected to respect the symmetries of the background model. In the case of Robertson–Walker models, the rotational symmetry of the background ensures that the multipoles a_{lm} are uncorrelated for different values of l and m:

$$\langle a_{lm} a^*_{l'm'} \rangle = C_l \delta_{ll'} \delta_{mm'} \,, \tag{3.2}$$

which defines the power spectrum C_l. The angle brackets in this equation denote the average over an ensemble of realisations of the fluctuations. The simplest models of inflation predict that the fluctuations should also be Gaussian at early times, and this is preserved by linear evolution of the small fluctuations. If Gaussian, the a_{lm}s are also independent, and the power spectrum

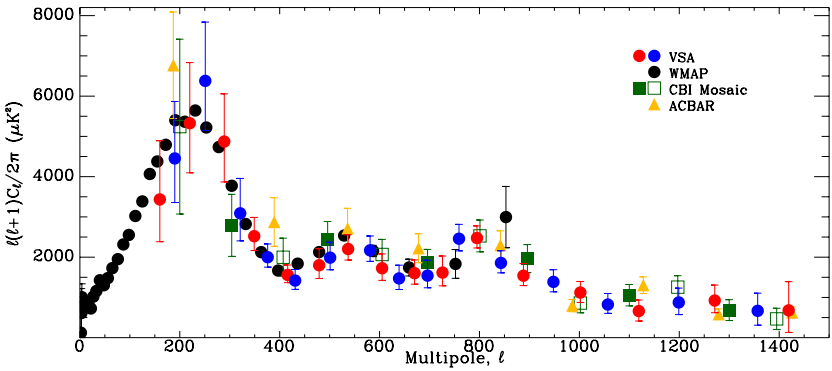

Fig. 3.1. Compilation of CMB anisotropy measurements (as of February 2004) from WMAP (black filled circles), the Very Small Array (VSA [20]; shaded circles representing two interleaving binning schemes), the Cosmic Background Imager (CBI [21, 22]; open and filled squares for two different binning schemes) and the Arcminute Cosmology Bolometer Array Receiver (ACBAR [23]; triangles). (Figure reproduced, with permission, from [20].)

provides the complete statistical description of the temperature anisotropies. For this reason, measuring the anisotropy power spectrum has, so far, been the main goal of observational CMB research. Temperature anisotropies have now been detected up to l of a few thousand; a recent compilation of current data as of February 2004 is given in Fig. 3.1.

The correlation between the temperature anisotropies along two directions evaluates to

$$\langle \Delta T(\hat{\boldsymbol{n}}_1)\Delta T(\hat{\boldsymbol{n}}_2)\rangle = T^2 \sum_l \frac{2l+1}{4\pi} C_l P_l(\cos\theta) \,, \tag{3.3}$$

which depends only on the angular separation θ as required by rotational invariance. Here, $P_l(x)$ are the Legendre polynomials. The mean-square temperature anisotropy is

$$\langle \Delta T^2\rangle = T^2 \sum_l \frac{2l+1}{4\pi} C_l \approx T^2 \int \frac{l(l+1)}{2\pi} C_l \,\mathrm{d}\ln l \,, \tag{3.4}$$

so that the quantity $l(l+1)C_l/2\pi$, which is conventionally plotted, is approximately the power per decade in l of the temperature anisotropies.

3.2.3 Kinetic Theory

The CMB photons can be described by a one-particle distribution function $f(x^a, p^a)$ that is a function of the spacetime position x^a and four-momentum p^a of the photon. It is defined such that the number of photons contained

in a proper three-volume element $\mathrm{d}^3\boldsymbol{x}$ and with three-momentum in $\mathrm{d}^3\boldsymbol{p}$ is $f\mathrm{d}^3\boldsymbol{x}\mathrm{d}^3\boldsymbol{p}$. The phase-space volume element $\mathrm{d}^3\boldsymbol{x}\mathrm{d}^3\boldsymbol{p}$ is Lorentz-invariant and is conserved along the photon path through phase space (see, e.g. [24]). It follows that f is also frame-invariant, and is conserved in the absence of scattering. To calculate the anisotropies in the CMB temperature, we must evolve the photon distribution function in the perturbed universe.

To avoid over-complicating our discussion, we shall only consider spatially-flat models here, and, for the moment, ignore the effects of polarization. For a more complete discussion, including these complications, see e.g. [25, 26]. Curvature mostly affects the CMB through the geometrical projection of linear scales at last scattering to angular scales on the sky today, but has a negligible impact on pre-recombination physics and hence much of the discussion in this section. The subject of cosmological perturbation theory is rich in methodology, but, for pedagogical reasons, we adopt here the most straightforward approach which is to work directly with the metric perturbations. This is also the most prevalent in the CMB literature. The (1+3)-covariant approach [27] is a well-developed alternative that is arguably more physically-transparent than metric-based techniques. It has also been applied extensively in the context of CMB physics [26, 28, 29, 30, 31, 32]. The majority of our discussion will be of scalar perturbations, where all perturbed three-tensors can be derived from the spatial derivatives of scalar functions, although we discuss tensor perturbations briefly in Sect. 3.2.5.

For scalar perturbations in spatially-flat models we can choose a gauge such that the spacetime metric is [33]

$$\mathrm{d}s^2 = a^2(\eta)[(1+2\psi)\mathrm{d}\eta^2 - (1-2\phi)\mathrm{d}\boldsymbol{x}^2] \,, \tag{3.5}$$

where η is conformal time (related to proper time t by $\mathrm{d}t = a\mathrm{d}\eta$), a is the scale factor in the background model and, now, $\boldsymbol{x}$ is comoving position. This gauge, known as the conformal Newtonian or longitudinal gauge, has the property that the congruence of worldlines with constant $\boldsymbol{x}$ have zero shear. The two scalar potentials ϕ and ψ constitute the scalar perturbation to the metric, with ϕ playing a similar role to the Newtonian gravitational potential. In the absence of anisotropic stress, ϕ and ψ are equal. We parameterise the photon four-momentum with its energy ϵ/a and direction $\boldsymbol{e}$ (with $\boldsymbol{e}^2 = 1$), as seen by an observer at constant $\boldsymbol{x}$, so that

$$p^\mu = a^{-2}\epsilon[1-\psi, (1+\phi)\boldsymbol{e}] \,. \tag{3.6}$$

Free photons move on the geodesics of the perturbed metric, $p^\mu\nabla_\mu p^\nu = 0$, so the energy and direction evolve as

$$\mathrm{d}\epsilon/\mathrm{d}\eta = -\epsilon\mathrm{d}\psi/\mathrm{d}\eta + \epsilon(\dot{\phi}+\dot{\psi}) \tag{3.7}$$

$$\mathrm{d}\boldsymbol{e}/\mathrm{d}\eta = -\boldsymbol{\nabla}_\perp(\phi+\psi) \,, \tag{3.8}$$

where dots denote $\partial/\partial\eta$ and $\boldsymbol{\nabla}_\perp$ is the three-gradient projected perpendicular to $\boldsymbol{e}$. We see immediately that ϵ is conserved in the absence of perturbations,

so that the energy redshifts in proportion to the scale factor in the background model. The change in direction of the photon due to the projected gradient of the potentials in the perturbed universe gives rise to gravitational lensing (see e.g. [34] for a review).

The dominant scattering mechanism to affect CMB anisotropies is classical Thomson scattering off free electrons, since around recombination the average photon energy is small compared to the rest mass of the electron. Furthermore, the thermal distribution of electron velocities can be ignored due to the low temperature. The evolution of the photon distribution function in the presence of Thomson scattering is

$$\frac{\mathrm{d}f}{\mathrm{d}\eta} = -a(1+\psi)n_\mathrm{e}\sigma_\mathrm{T}f + \frac{3}{16\pi}a(1+\psi)n_\mathrm{e}\sigma_\mathrm{T}\int f(\epsilon,\boldsymbol{e}')[(1+(\boldsymbol{e}\cdot\boldsymbol{e}')^2]\,\mathrm{d}\boldsymbol{e}' - an_\mathrm{e}\sigma_\mathrm{T}\boldsymbol{e}\cdot\boldsymbol{v}_\mathrm{b}\epsilon\frac{\partial f}{\partial\epsilon}\,, \tag{3.9}$$

where n_e is the electron (proper) number density, σ_T is the Thomson cross section, and the electron peculiar velocity is $\boldsymbol{v}_\mathrm{b} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\eta$. The derivative on the left of (3.9) is along the photon path in phase space:

$$\frac{\mathrm{d}f}{\mathrm{d}\eta} = \frac{\partial f}{\partial\eta} + \boldsymbol{e}\cdot\boldsymbol{\nabla}f + (\dot{\phi} - \boldsymbol{e}\cdot\boldsymbol{\nabla}\psi)\epsilon\frac{\partial f}{\partial\epsilon} \tag{3.10}$$

to first order, where we have used (3.7) and (3.8) and the fact that the anisotropies of f are first order. The first term on the right of (3.9) describes scattering out of the beam, and the second scattering into the beam. The final term arises from the out-scattering of the additional dipole moment in the distribution function seen by the electrons due to the Doppler effect. In the background model f is isotropic and the net scattering term vanishes, so that f is a function of the conserved ϵ only: $f = \bar{f}(\epsilon)$. Thermal equilibrium ensures that $\bar{f}$ is a Planck function.

The fluctuations in the photon distribution function inherit an energy dependence $\epsilon\partial\bar{f}/\partial\epsilon$ from the source terms in the Boltzmann equation (3.9). Separating out the background contribution to f, and its energy dependence, we can write

$$f(\eta,\boldsymbol{x},\epsilon,\boldsymbol{e}) = \bar{f}(\epsilon)[1-\Theta(\eta,\boldsymbol{x},\boldsymbol{e})\mathrm{d}\ln\bar{f}/\mathrm{d}\ln\epsilon], \tag{3.11}$$

so that the CMB spectrum is Planckian but with a direction-dependent temperature $\Delta T/T = \Theta$. Using the Lorentz invariance of f, it is not difficult to show that the quadrupole and higher moments of Θ are gauge-invariant. If we now substitute for f in (3.9), we find the Boltzmann equation for Θ:

$$\frac{\partial(\Theta+\psi)}{\partial\eta} + \boldsymbol{e}\cdot\boldsymbol{\nabla}(\Theta+\psi) = -an_\mathrm{e}\sigma_\mathrm{T}\Theta + \frac{3}{16\pi}an_\mathrm{e}\sigma_\mathrm{T}\int\Theta(\boldsymbol{e}')[(1+(\boldsymbol{e}\cdot\boldsymbol{e}')^2]\,\mathrm{d}\boldsymbol{e}' + an_\mathrm{e}\sigma_\mathrm{T}\boldsymbol{e}\cdot\boldsymbol{v}_\mathrm{b} + \dot{\phi} + \dot{\psi}\,. \tag{3.12}$$

The formal solution of this equation is an integral along the line of sight $\hat{\boldsymbol{n}} = -\boldsymbol{e}$,

$$[\Theta(\hat{\boldsymbol{n}}) + \psi]_R = \mathrm{e}^{-\tau}[\Theta(\hat{\boldsymbol{n}}) + \psi)]_E + \int_E^R \mathrm{e}^{-\tau} S \, \mathrm{d}\eta \, , \tag{3.13}$$

where R is the reception event, E is the emission event, and $\tau \equiv \int a n_\mathrm{e} \sigma_\mathrm{T} \, \mathrm{d}\eta$ is the optical depth back from R. The source term S is given by the right-hand side of (3.12), but with Θ replaced by $-\psi$ in the first term.

We gain useful insight into the physics of anisotropy formation by approximating the last scattering surface as sharp (which is harmless on large angular scales), and ignoring the quadrupole CMB anisotropy at last scattering. In this case (3.13) reduces to

$$[\Theta(\hat{\boldsymbol{n}}) + \psi]_R = \Theta_0|_E + \psi|_E - \hat{\boldsymbol{n}} \cdot \boldsymbol{v}_\mathrm{b}|_E + \int_E^R (\dot{\psi} + \dot{\phi}) \, \mathrm{d}\eta \, , \tag{3.14}$$

where Θ_0 is the isotropic part of Θ, and is proportional to the fluctuation in the photon energy density. The various terms in this equation have a simple physical interpretation. The temperature received along direction $\hat{\boldsymbol{n}}$ is the isotropic temperature of the CMB at the last scattering event on the line of sight, Θ_0, corrected for the gravitational redshift due to the difference in potential between E and R, and the Doppler shift $\boldsymbol{e} \cdot \boldsymbol{v}_\mathrm{b}|_E$ resulting from scattering off moving electrons. Finally, there is an additional gravitational redshift contribution arising from evolution of the gravitational potentials [3].

Machinery for an Accurate Calculation

An accurate calculation of the CMB anisotropy on all scales where linear perturbation theory (see also R. Durrer's contribution) is valid requires a full numerical solution of the Boltzmann equation. The starting point is to expand $\Theta(\theta, \boldsymbol{x}, \boldsymbol{e})$ in appropriate basis functions. For scalar perturbations, these are the contraction of the (irreducible) trace-free tensor products $e^{\langle i_1} \dots e^{i_l \rangle}$ (the angle brackets denoting the trace-free part) with trace-free (spatial) tensors derived from derivatives of scalars [28, 31, 35]. Fourier expanding the scalar functions, we end up forming contractions between $e^{\langle i_1} \dots e^{i_l \rangle}$ and $\hat{k}_{\langle i_1} \dots \hat{k}_{i_l \rangle}$ where $\hat{\boldsymbol{k}}$ is the wavevector. These contractions reduce to Legendre polynomials of $\hat{\boldsymbol{k}} \cdot \boldsymbol{e}$, and so the normal-mode expansion of Θ for scalar perturbations takes the form

$$\Theta(\eta, \boldsymbol{x}, \boldsymbol{e}) = \sum_{l \geq 0} \int \frac{\mathrm{d}^3 \boldsymbol{k}}{(2\pi)^{3/2}} (-\mathrm{i})^l \Theta_l(\eta, \boldsymbol{k}) P_l(\hat{\boldsymbol{k}} \cdot \boldsymbol{e}) \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \, . \tag{3.15}$$

It is straightforward to show that the implied azimuthal symmetry about the wavevector is consistent with the Boltzmann equation (3.12). Inserting the expansion of Θ into this equation gives the Boltzmann hierarchy for the moments Θ_l:

$$\dot{\Theta}_l + k\left(\frac{l+1}{2l+3}\Theta_{l+1} - \frac{l}{2l-1}\Theta_{l-1}\right) = an_\mathrm{e}\sigma_\mathrm{T}\left[(\delta_{l0}-1)\Theta_l - \delta_{l1}v_\mathrm{b} + \frac{1}{10}\Theta_2\right] + \delta_{l0}\dot{\phi} + \delta_{l1}k\psi\ , \tag{3.16}$$

where $\boldsymbol{v}_\mathrm{b} = \int \mathrm{i}\hat{\boldsymbol{k}}v_\mathrm{b}(\boldsymbol{k})\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\,\mathrm{d}^3\boldsymbol{k}/(2\pi)^{3/2}$, and ϕ and ψ are the Fourier transforms of the potentials. This system of ordinary differential equations can be integrated directly with the linearised Einstein equations for the metric perturbations, and the fluid equations governing perturbations in the other matter components, as in the publically-available COSMICS code [33]. Careful treatment of the truncation of the hierarchy is necessary to avoid unphysical reflection of power back down through the moments.

A faster way to solve the Boltzmann equation numerically is to use the line-of-sight solution (3.13), as in the widely-used CMBFAST code [36] and its parallelised derivative CAMB [37]. Inserting the expansion (3.15) gives the integral solution to the hierarchy

$$\Theta_l|_{\eta_0} = (2l+1)\int_0^{\eta_0}\mathrm{d}\eta\,\mathrm{e}^{-\tau}\Big[(\dot{\phi}+\dot{\psi})j_l(k\Delta\eta) - \dot{\tau}(\Theta_0+\psi)j_l(k\Delta\eta) + \dot{\tau}v_\mathrm{b}j_l'(k\Delta\eta) - \frac{1}{20}\dot{\tau}\Theta_2(3j_l''+j_l)(k\Delta\eta)\Big] \tag{3.17}$$

where $\Delta\eta \equiv \eta_0 - \eta$, j_l is a spherical Bessel function, and primes denote derivatives with respect to the argument. Using the integral solution, it is only necessary to evolve the Boltzmann hierarchy to modest l to compute accurately the source terms that appear in the integrand. The integral approach is thus significantly faster than a direct solution of the hierarchy.

The spherical multipoles a_{lm} of the temperature anisotropy can be extracted from (3.15) as

$$a_{lm} = 4\pi\mathrm{i}^l\int\frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^{3/2}}\,\frac{\Theta_l}{2l+1}Y^*_{lm}(\hat{\boldsymbol{k}})\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\ . \tag{3.18}$$

Statistical homogeneity and isotropy imply that the equal-time correlator

$$\langle\Theta_l(\eta,\boldsymbol{k})\Theta_l^*(\eta,\boldsymbol{k}')\rangle = \frac{2\pi^2}{k^3}\Theta_l^2(\eta,k)\delta(\boldsymbol{k}-\boldsymbol{k}')\ , \tag{3.19}$$

so forming the correlation $\langle a_{lm}a^*_{l'm'}\rangle$ gives the power spectrum

$$C_l = \frac{4\pi}{(2l+1)^2}\int\Theta_l^2(k)\,\mathrm{d}\ln k\ . \tag{3.20}$$

If we consider (pure) perturbation modes characterised by a single independent stochastic amplitude per Fourier mode (such as the comoving curvature for the adiabatic mode; (see Sect. 3.2.4), the power $\Theta_l^2(k)$ is proportional to the power spectrum of that amplitude. The spherical Bessel functions in (3.17) peak sharply at $k\Delta\eta = l$ for large l, so that multipoles l are mainly

probing spatial structure with wavenumber $k \sim l/\Delta\eta$ at last scattering. The oscillatory tails of the Bessel functions mean that some power from a given k does also enter larger scale anisotropies. Physically, this arises from Fourier modes that are not aligned with their wavevector perpendicular to the line of sight. As we discuss in the next section, the tightly-coupled system of photons and baryons undergoes acoustic oscillations prior to recombination on scales inside the sound horizon. For the pure perturbation modes, all modes with a given wavenumber reach the maxima or minima of their oscillation at the same time, irrespective of the direction of $\boldsymbol{k}$, and so we expect modulation in the C_ls on sub-degree scales. The first three of these acoustic peaks have now been measured definitively; see Fig. 3.1.

3.2.4 Photon–Baryon Dynamics

Prior to recombination, the mean free path of CMB photons is $\sim 4.9 \times 10^4(\Omega_{\rm b}h^2)^{-1}(1+z)^{-2}$ Mpc. On comoving scales below this length the photons and baryons behave as a tightly-coupled fluid, with the CMB almost isotropic in the baryon frame. In this limit, only the $l = 0$ and $l = 1$ moments of the distribution function are significant.

The stress-energy tensor of the photons is given in terms of the distribution function by

$$T^{\mu\nu} = a^{-2} \int f(\eta, \boldsymbol{x}, \epsilon, \boldsymbol{e}) p^\mu p^\nu \epsilon \, {\rm d}\epsilon {\rm d}\boldsymbol{e} \, , \tag{3.21}$$

so that the Fourier modes of the fractional over-density of the photons are $\delta_\gamma = 4\Theta_0$ and the photon (bulk) velocity $v_\gamma = -\Theta_1$. The anisotropic stress is proportional to Θ_2. In terms of these variables, the first two moment equations of the Boltzmann hierarchy become

$$\dot\delta_\gamma - \frac{4}{3} k v_\gamma - 4\dot\phi = 0 \tag{3.22}$$

$$\dot v_\gamma + \frac{1}{4} k \delta_\gamma - \frac{2}{5} k \Theta_2 + k\psi = \dot\tau (v_\gamma - v_{\rm b}) \, . \tag{3.23}$$

Here, the derivative of the optical depth $\dot\tau = -a n_{\rm e} \sigma_{\rm T}$ (and so is negative). The momentum exchange between the photons and baryons due to the drag term in (3.23) gives rise to a similar term in the Euler equation for the baryons:

$$\dot v_{\rm b} + \mathcal{H} v_{\rm b} + k\psi = R^{-1} \dot\tau (v_{\rm b} - v_\gamma) \, , \tag{3.24}$$

where we have ignored baryon pressure. The ratio of the baryon energy density to the photon enthalpy is $R \equiv 3\rho_{\rm b}/4\rho_\gamma$ and is proportional to the scale factor a, and $\mathcal{H} \equiv \dot a/a$ is the conformal Hubble parameter.

In the tightly-coupled limit $|\dot\tau^{-1}| \ll k^{-1}$ and $\mathcal{H}^{-1}$. In this limit, we can treat the ratios of the mean-free path to the wavelength and the Hubble time as small perturbative parameters. Equations (3.23) and (3.24) then imply

that $v_\gamma = v_\mathrm{b}$ to first order in the small quantities $k/|\dot\tau|$ and $\mathcal{H}/|\dot\tau|$. Comparing the continuity equation for the baryons,

$$\dot\delta_\mathrm{b} - kv_\mathrm{b} - 3\dot\phi = 0\,, \tag{3.25}$$

with that for the photons, we see that $\dot\delta_\gamma = 4\dot\delta_\mathrm{b}/3$, so the *evolution* of the photon–baryon fluid is adiabatic, preserving the local ratio of the number densities of photons to baryons. Combining (3.23) and (3.24) to eliminate the scattering terms, and then using $v_\gamma = v_\mathrm{b}$, we find the evolution of the photon velocity to leading order in tight coupling:

$$\dot v_\gamma + \frac{R}{1+R}\mathcal{H}v_\gamma + \frac{1}{4(1+R)}k\delta_\gamma + k\psi = 0\,. \tag{3.26}$$

The $l > 1$ moments of the photon distribution function arise from the balance between isotropisation by scattering and their generation by photons free streaming over a mean free path; these moments are suppressed by factors $(k/|\dot\tau|)^{l-1}$. In particular, during tight coupling $\Theta_2 \approx (20/27)k\dot\tau^{-1}v_\gamma$ ignoring polarization. (The factor $20/27$ rises to $8/9$ if we correct for polarization [38].)

Combining (3.26) with the photon continuity equation (3.22) shows that the tightly-coupled dynamics of δ_γ is that of a damped, simple-harmonic oscillator driven by gravity [39]:

$$\ddot\delta_\gamma + \frac{\mathcal{H}R}{1+R}\dot\delta_\gamma + \frac{1}{3(1+R)}k^2\delta_\gamma = 4\ddot\phi + \frac{4\mathcal{H}R}{1+R}\dot\phi - \frac{4}{3}k^2\psi\,. \tag{3.27}$$

The damping term arises from the redshifting of the baryon momentum in an expanding universe, while photon pressure provides the restoring force which is weakly suppressed by the additional inertia of the baryons. The WKB solutions to the homogeneous equation are

$$\delta_\gamma = (1+R)^{-1/4}\cos kr_\mathrm{s}\,, \quad \text{and} \quad \delta_\gamma = (1+R)^{-1/4}\cos kr_\mathrm{s}\,, \tag{3.28}$$

where the *sound horizon* $r_\mathrm{s} \equiv \int_0^\eta \mathrm{d}\eta'/\sqrt{3(1+R)}$. Note also that for static potentials, and ignoring the variation of R with time, the mid-point of the oscillation of δ_γ is shifted to $-4(1+R)\psi$. The dependence of this shift on the baryon density produces a baryon-dependent modulation of the height of the acoustic peak in the temperature anisotropy power spectrum; see Sect. 3.3.

The driving term in (3.27) depends on the evolution of the gravitational potentials. If we ignore anisotropic stress, ϕ and ψ are equal, and their Fourier modes evolve as

$$\ddot\phi + 3\mathcal{H}\left(1+\frac{\dot p}{\dot\rho}\right)\dot\phi + \left[2\dot{\mathcal{H}} + \left(1+3\frac{\dot p}{\dot\rho}\right)\mathcal{H}^2\right]\phi + \frac{\dot p}{\dot\rho}k^2\phi = \frac{1}{2}\kappa a^2\left(\delta p - \frac{\dot p}{\dot\rho}\delta\rho\right) \tag{3.29}$$

in a flat universe, which follows from the perturbed Einstein field equations. Here, ρ and p are the total density and pressure in the background model,

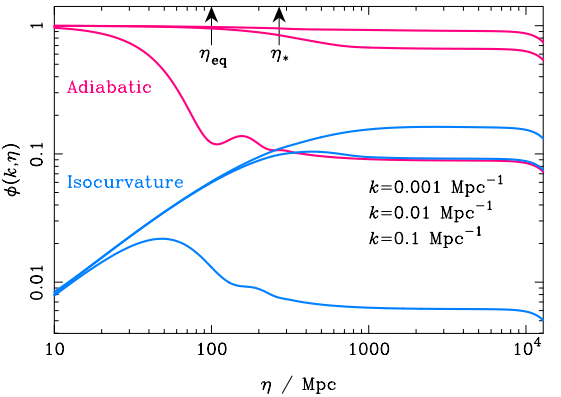

Fig. 3.2. Evolution of the potential ϕ in adiabatic and CDM-isocurvature models for wavenumbers $k = 0.001$, 0.01 and $0.1\,\mathrm{Mpc}^{-1}$ (top to bottom respectively in matter domination). The conformal time at matter–radiation equality $\eta_{\rm eq}$ and last scattering η_* are marked by arrows.

$\delta\rho$ and δp are the Fourier modes of their perturbations, and $\kappa \equiv 8\pi G$. The source term is gauge-invariant; it vanishes for mixtures of barotropic fluids $[p_i = p_i(\rho_i)]$ with $\delta\rho_i/(\rho_i + p_i)$ the same for all components. For *adiabatic* perturbations, this latter condition holds initially and is preserved on super-Hubble scales. It is also preserved in the tightly-coupled photon–baryon fluid as we saw above. For adiabatic perturbations, the potential is constant on scales larger than the sound horizon when p/ρ is constant, but decays during transitions in the equation of state, such as from matter to radiation domination. Above the sound horizon in flat models, it can be shown that the quantity

$$\mathcal{R} \equiv -\phi - 2\frac{\mathcal{H}\dot{\phi} + \mathcal{H}^2}{\kappa a^2(\rho + p)} \tag{3.30}$$

is conserved even through such transitions. The perturbation to the intrinsic curvature of comoving hypersurfaces (*i.e.* those perpendicular to the the four-velocity of observers who see no momentum density) is given in terms of $\mathcal{R}$ as $4(k^2/a^2)\mathcal{R}$. Using the constancy of $\mathcal{R}$ on large scales, the potential falls by a factor of 9/10 during the transition from radiation to matter domination. The evolution of the potential is illustrated in Fig. 3.2 in a flat ΛCDM model with parameters given in Sect. 3.1. The potential oscillates inside the sound horizon during radiation domination since the photons, which are the dominant component at that time, undergo acoustic oscillations on such scales.

The behaviour of the potentials for *isocurvature* perturbations is quite different on large scales during radiation domination [40], since the source term in (3.29) is then significant. In isocurvature fluctuations, the initial perturbations in the energy densities of the various components compensate each

other in such a way that the comoving curvature $\mathcal{R} = 0$. Figure 3.2 shows the evolution of CDM-isocurvature modes, in which there is initially a large fractional perturbation in the dark matter density, with a small compensating fractional perturbation in the radiation. (The full set of possibilities for regular isocurvature modes are discussed in [41].) On large scales in radiation domination the potential grows as a, the scale factor.

Adiabatic Fluctuations

For adiabatic fluctuations, the photons are initially perturbed by $\delta_\gamma(0) = -2\psi(0) = 4\mathcal{R}(0)/3$, *i.e.* they are over-dense in potential wells, and their velocity vanishes $v_\gamma(0) = 0$. If we consider super-Hubble scales at last scattering, there has been insufficient time for v_γ to grow by gravitational infall and the action of pressure gradients and it remains small. The photon continuity equation (3.22) then implies that $\delta_\gamma - 4\phi$ remains constant, and the decay of ϕ through the matter–radiation transition leaves $(\delta_\gamma/4 + \psi)(\eta_*) \approx \phi(\eta_*)/3 = -3\mathcal{R}(0)/5$ on large scales ($k < 3 \times 10^{-3}\,\mathrm{Mpc}^{-1}$) at last scattering. The combination $\delta_\gamma/4 + \psi = \Theta_0 + \psi$ is the dominant contribution to the large-scale temperature anisotropies produces at last scattering; see (3.14). The evolution of the photon density and velocity perturbations for adiabatic initial conditions are show in Fig. 3.3, along with the scale dependence of the fluctuations at last scattering. The plateau in $(\delta_\gamma/4 + \psi)(\eta_*)$ on large scales ensures that a scale-invariant spectrum of curvature perturbations translates into a scale-invariant spectrum of temperature anisotropies, $l(l+1)C_l = \text{constant}$, for small l.

On scales below the sound horizon at last scattering, the photon–baryon fluid has had time to undergo acoustic oscillation. The form of the photon

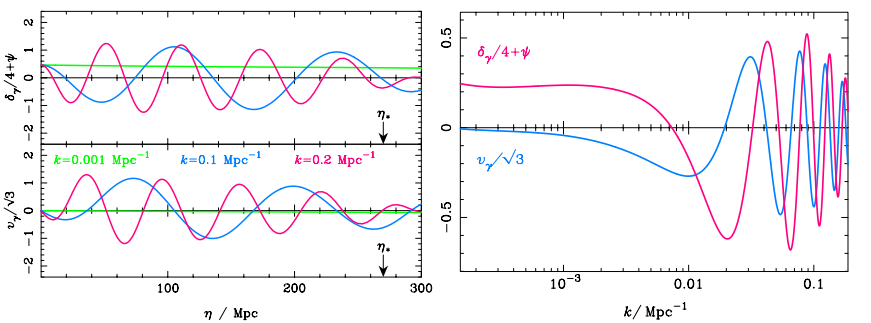

Fig. 3.3. Evolution of the combination $\delta_\gamma/4 + \psi$ (top left) and the photon velocity v_γ (bottom left) which determine the temperature anisotropies produced at last scattering (denoted by the arrow at η_*). Three modes are shown with wavenumbers $k = 0.001$, 0.1 and $0.2\,\mathrm{Mpc}^{-1}$, and the initial conditions are adiabatic. The fluctuations at the time of last scattering are shown as a function of linear scale in the right-hand plot.

initial condition, and the observation that the driving term in (3.27) mimics the cosine WKB solution of the homogeneous equation (see Fig. 3.2), set the oscillation mostly in the $\cos kr_\mathrm{s}$ mode. The midpoint of the oscillation is roughly at $\delta_\gamma/4 = -(1+R)\psi$. This behaviour is illustrated in Fig. 3.3. Modes with $kr_\mathrm{s}(\eta_*) = \pi$ have undergone half an oscillation at last scattering, and are maximally compressed. The large value of $\Theta_0 + \psi$ at this particular scale gives rise to the first acoustic peak in Fig. 3.1, now measured to be at $l = 220.1 \pm 0.8$ [42]. The subsequent extrema of the acoustic oscillation at $kr_\mathrm{s}(\eta_*) = n\pi$ give rise to the further acoustic peaks. The angular spacing of the peaks is almost constant and is set by the sound horizon at last scattering and the angular diameter distance to last scattering. The acoustic part of the anisotropy spectrum thus encodes a wealth of information on the cosmological parameters; see Sect. 3.3. The photon velocity v_γ oscillates as $\sin kr_\mathrm{s}$, so the Doppler term in (3.14) tends to fill in power between the acoustic peaks. The relative phase of the oscillation of the photon velocity has important implications for the polarization properties of the CMB as discussed in Sect. 3.4. The contributions of the various terms in (3.14) to the temperature-anisotropy power spectrum are shown in Fig. 3.4 for adiabatic perturbations.

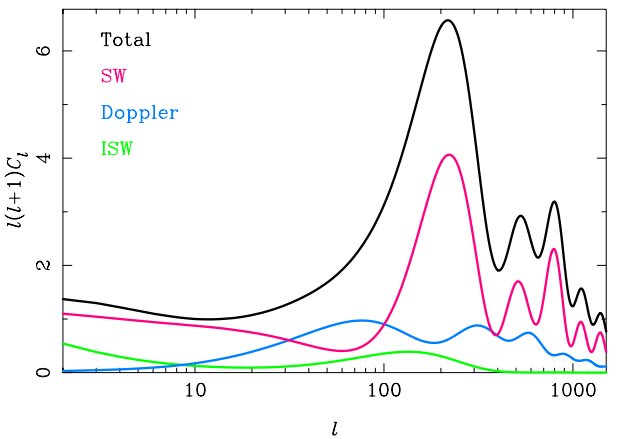

Fig. 3.4. Contribution of the various terms in (3.14) to the temperature-anisotropy power spectrum from adiabatic initial conditions. At high l, the contributions are (from top to bottom): total power; $\delta_\gamma/4 + \psi$ (denoted SW for Sachs–Wolfe [3]); Doppler effect from v_b; and the integrated Sachs–Wolfe effect (ISW) coming from evolution of the potential along the line of sight.

Isocurvature Fluctuations

For the CDM-isocurvature mode[1] the photons are initially unperturbed, as is the geometry: $\delta_\gamma(0) = 0 = \phi(0)$ and $v_\gamma = 0$. On large scales $\delta_\gamma/4 = \phi$ is preserved, so the growth in ϕ during radiation domination is matched by a growth in δ_γ and the photons are under-dense in potential wells. It follows that at last scattering $(\delta_\gamma/4 + \psi)(\eta_*) \approx 2\phi(\eta_*)$ for $k < 3 \times 10^{-3}\,\mathrm{Mpc}^{-1}$. Note that the redshift climbing out of a potential well *enhances* the intrinsic temperature fluctuation due to the photon under-density there. The evolution of the photon fluctuations for isocurvature initial conditions are shown in Fig. 3.5.

The evolution of the potential for isocurvature modes makes the driving term in (3.27) mimic the sine solution of the homogeneous equation, and so δ_γ follows suit oscillating as $\sin \sim kr_\mathrm{s}$ about the equilibrium point $-4(1+R)\psi$. The acoustic peaks are at $kr_\mathrm{s}(\eta_*) \sim n\pi/2$, and the photons are under-dense in the potential wells for the odd-n peaks, while over-dense in the even n. The various contributions to the temperature-anisotropy power spectrum for isocurvature initial conditions are shown in Fig. 3.6. The different peak positions for isocurvature initial conditions allow the CMB to constrain their relative contribution to the total fluctuations. Current constraints are rather dependent on whether one allows for correlations between the adiabatic and isocurvature modes (as are generic in the multi-field inflation models that might have generated the initial conditions), and the extent to which additional cosmological constraints are employed; see [44] for a recent analysis allowing for the most general correlations but a single power-law spectrum.

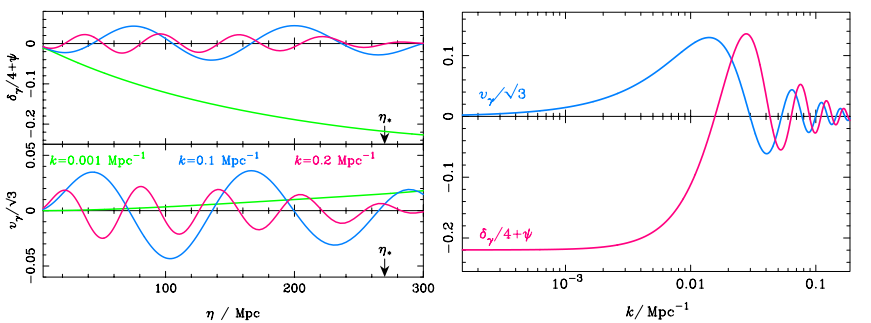

Fig. 3.5. As Fig. 3.3 but for CDM-isocurvature initial conditions.

[1] It is also possible to have the dominant fractional fluctuation in the baryon density rather than the cold dark matter. However, this mode is nearly indistinguishable from the CDM mode since, in the absence of baryon pressure, they differ only by a constant mode in which the radiation and the geometry remain unperturbed, but the CDM and baryon densities have compensating density fluctuations [43].

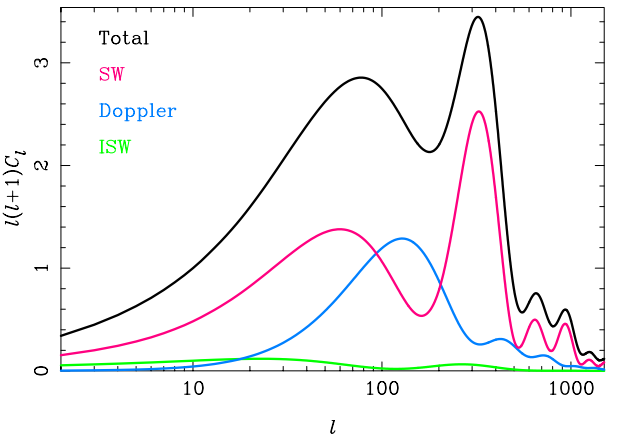

Fig. 3.6. As Fig. 3.4 but for CDM-isocurvature initial conditions. The initial spectrum of entropy perturbations is scale-invariant.

Beyond Tight-Coupling

On small scales it is necessary to go beyond tight-coupling of the photon–baryon system since the photon diffusion length can become comparable to the wavelength of the fluctuations. Photons that have had sufficient time to diffuse of the order of a wavelength can leak out of over-densities, thus damping the acoustic oscillations and generating anisotropy [45]. A rough estimate of the comoving scale below which diffusion is important is the square root of the geometric mean of the particle horizon (or conformal age) and the mean-free path of the photons, *i.e.* $\sqrt{\eta/|\dot{\tau}|}$. Converting this to a comoving wavenumber defines the damping scale

$$k_{\rm D}^{-2} \sim 0.3(\Omega_{\rm m}h^2)^{-1/2}(\Omega_{\rm b}h^2)^{-1}(a/a_*)^{5/2}\,{\rm Mpc}^2 \tag{3.31}$$

when the scale factor is a. Here, a_* is the scale factor at last scattering, and the expression is valid well after matter–radiation equality but well before recombination. The effect of diffusion is to damp the photon (and baryon) oscillations exponentially by the time of last scattering on comoving scales smaller than $\sim 3\,{\rm Mpc}$. The resulting damping effect on the temperature power spectrum has now been measured by several experiments [20, 22, 23].

To describe diffusion damping more quantitatively, we consider scales that were already sub-Hubble during radiation domination. The gravitational potentials will then have been suppressed during their oscillatory phase when the photons (which are undergoing acoustic oscillations themselves) dominated the energy density, and so we can ignore gravitational effects. Furthermore, the dynamical timescale of the acoustic oscillations is then short compared to the expansion time and we can ignore the effects of expansion. In this limit, the Euler equations for the photons and the baryons can be iterated to give the relative velocity between the photons and baryons to first

order in $k/|\dot{\tau}|$:

$$(1+R^{-1})(v_\gamma - v_\mathrm{b}) = \frac{1}{4}k\dot{\tau}^{-1}\delta_\gamma \,. \tag{3.32}$$

Using momentum conservation for the total photon–baryon system gives

$$\dot{v}_\gamma + R\dot{v}_\mathrm{b} + \frac{1}{4}k\delta_\gamma - \frac{2}{5}k\Theta_2 = 0 \,, \tag{3.33}$$

which can be combined with the derivative of (3.32) to give a new Euler equation for the photons correct to first order in tight coupling:

$$(1+R)\dot{v}_\gamma \approx -\frac{1}{4}k\delta_\gamma + \frac{R^2}{4(1+R)}k\dot{\tau}^{-1}\dot{\delta}_\gamma + \frac{16}{45}k^2\dot{\tau}^{-1}v_\gamma \,. \tag{3.34}$$

Here, we have used $\Theta_2 \approx 8k\dot{\tau}^{-1}v_\gamma/9$ which includes the correction due to polarization. In the limit of perfect coupling, (3.34) reduces to (3.26) on small scales. The continuity equation for the photons, $\dot{\delta}_\gamma = 4kv_\gamma/3$ $(+4\dot{\phi})$, shows that the last two terms on the right of (3.34) are drag terms, and on differentiating gives

$$\ddot{\delta}_\gamma - \frac{k^2\dot{\tau}^{-1}}{3(1+R)}\left(\frac{16}{15} + \frac{R^2}{1+R}\right)\dot{\delta}_\gamma + \frac{k^2}{3(1+R)}\delta_\gamma = 0 \,. \tag{3.35}$$

The WKB solution is

$$\delta_\gamma \propto \mathrm{e}^{\pm\mathrm{i}kr_\mathrm{s}}\mathrm{e}^{-k^2/k_\mathrm{D}^2} \,, \quad \text{where} \quad \frac{1}{k_\mathrm{D}^2} \equiv \frac{1}{6}\int_0^\eta \frac{|\dot{\tau}^{-1}|}{1+R}\left(\frac{16}{15} + \frac{R^2}{1+R}\right)\mathrm{d}\eta' \tag{3.36}$$

is the damping scale.

The finite mean-free path of CMB photons around last scattering has an additional effect on the temperature anisotropies. The visibility function $-\dot{\tau}\mathrm{e}^\tau$ has a finite width $\sim 80\,\mathrm{Mpc}$ and so along a given line of sight photons will be last scattered over this interval. Averaging over scattering events will tend to wash out the anisotropy from wavelengths short compared to the width of the visibility function. This effect is described mathematically by integrating the oscillations in the spherical Bessel functions in (3.17) against the product of the visibility function and the (damped) perturbations.

Boltzmann codes such as CMBFAST [36] and CAMB [37] use the tight-coupling approximation at early times to avoid the numerical problems associated with integrating the stiff Euler equations in their original forms (3.23) and (3.24).

3.2.5 Other Features of the Temperature-Anisotropy Power Spectrum

We end this section on the fundamentals of the physics of CMB temperature anisotropies by reviewing three additional effects that contribute to the linear anisotropies.

Integrated Sachs–Wolfe Effect

The integrated Sachs–Wolfe (ISW) effect is described by the last term on the right of (3.14). It is an additional source of anisotropy due to the temporal variation of the gravitational potentials along the line of sight: if a potential well deepens as a CMB photon crosses it then the blueshift due to infall will be smaller than redshift from climbing out of the (now deeper) well. (The combination $\phi + \psi$ has a direct geometric interpretation as the potential for the electric part of the Weyl tensor [46].) The ISW receives contributions from late times as the potentials decay during dark-energy domination, and at early times around last scattering due to the finite time since matter–radiation equality.

The late-time effect contributes mainly on large angular scales since there is little power in the potentials at late times on scales that entered the Hubble radius during radiation domination. The late ISW effect is the only way to probe late-time structure growth (and hence e.g. distinguish between different dark-energy models) with linear CMB anisotropies, but this is hampered by cosmic variance on large angular scales. The late ISW effect produces correlations between the large-scale temperature fluctuations and other tracers of the potential in the local universe, and with the advent of the WMAP data these have now been tentatively detected [47, 48, 49]; see also Sect. 3.5.

In adiabatic models the early-time ISW effect adds coherently with the contribution $\delta_\gamma/4 + \psi$ to the anisotropies near the first peak, boosting this peak significantly [39]; see Fig. 3.4. The reason is that the linear scales that contribute here are maximally compressed with $\delta_\gamma/4 + \psi \sim -\psi/2$ which has the same sign as $\dot{\phi}$ for decaying ϕ.

Reionization

Once structure formation had proceeded to produce the first sources of ultraviolet photons, the universe began to reionize. The resulting free electron density could then re-scatter CMB photons, and this tended to isotropise the CMB by averaging the anisotropies from many lines of sight at the scattering event. Approximating the bi-modal visibility function as two delta functions, one at last scattering[2] η_* and one at reionization $\eta_{\rm re}$, if the optical depth through reionization is $\tau_{\rm re}$, the temperature fluctuation at $\boldsymbol{x} = 0$ at η_0 is

$$[\Theta(\hat{\boldsymbol{n}}) + \psi]_{\eta_0} \approx (1 - \mathrm{e}^{-\tau_{\rm re}})(\Theta_0 + \psi - \hat{\boldsymbol{n}} \cdot \boldsymbol{v}_{\rm b})[-\hat{\boldsymbol{n}}(\eta_0 - \eta_{\rm re}), \eta_{\rm re}] + \mathrm{e}^{-\tau_{\rm re}}(\Theta_0 + \psi - \hat{\boldsymbol{n}} \cdot \boldsymbol{v}_{\rm b})[-\hat{\boldsymbol{n}}(\eta_0 - \eta_*), \eta_*] \,. \quad (3.37)$$

Here, we have used (3.13), neglected the ISW effect, and approximated the scattering as isotropic. The first term on the right describes the effect of

[2] We continue to refer to the last scattering event around recombination as last scattering, even in the presence of re-scattering at reionization.

blending the anisotropies from different lines of sight (to give Θ_0) and the generation of new anisotropies by re-scattering off moving electrons at reionization; the second term is simply the temperature anisotropy that would be observed with no reionization, weighted by the fraction of photons that do not re-scatter. Since $\Theta_0 + \psi$ at the re-scattering event is the average of $\Theta_0 + \psi - \hat{\boldsymbol{n}}' \cdot \boldsymbol{v}_\mathrm{b}$ on the electron's last scattering surface, on large scales $k(\eta_\mathrm{re} - \eta_*) \ll 1$ it reduces to $\Theta_0 + \psi$ at $[-\hat{\boldsymbol{n}}(\eta_0 - \eta_*), \eta_*]$, while on small scales it vanishes. It follows that for scales that are super-horizon at reionization, the observed temperature anisotropy becomes

$$\Theta(\hat{\boldsymbol{n}}) \to \Theta(\hat{\boldsymbol{n}}) - (1 - \mathrm{e}^{-\tau_\mathrm{re}})\hat{\boldsymbol{n}} \cdot \Delta\boldsymbol{v}_\mathrm{b} \,, \tag{3.38}$$

where $\Delta\boldsymbol{v}_\mathrm{b}$ is the difference between the electron velocity at the reionization event and the preceding last scattering event on the line of sight. On such scales the Doppler terms do not contribute significantly and the temperature anisotropy is unchanged. For scales that are sub-horizon at reionization,

$$\Theta(\hat{\boldsymbol{n}}) \to \mathrm{e}^{-\tau_\mathrm{re}}\Theta(\hat{\boldsymbol{n}}) - (1 - \mathrm{e}^{-\tau_\mathrm{re}})\hat{\boldsymbol{n}} \cdot \boldsymbol{v}_\mathrm{b} \,, \tag{3.39}$$

where the Doppler term is evaluated at reionization. In practice, the visibility function is not perfectly sharp at reionization and the integral through the finite re-scattering distance tends to wash out the Doppler term since only plane waves with their wavevectors near the line of sight contribute significantly to $\hat{\boldsymbol{n}} \cdot \boldsymbol{v}_\mathrm{b}$. Figure 3.7 shows the resulting effect $C_l \to \mathrm{e}^{-2\tau_\mathrm{re}} C_l$ on the anisotropy power spectrum on small scales. Recent results from WMAP [19] suggest an optical depth through reionization $\tau_\mathrm{re} \sim 0.17$. Such early reionization cannot have been an abrupt process since the implied redshift $z_\mathrm{re} \sim 15$ is at odds with the detection of traces of smoothly-distributed neutral hydrogen at $z \sim 6$ via Gunn-Peterson troughs in the spectra of high-redshift quasars [50, 51].

Tensor Modes

Tensor modes, describing gravitational waves, represent the transverse trace-free perturbations to the spatial metric:

$$\mathrm{d}s^2 = a^2(\eta)[\mathrm{d}\eta^2 - (\delta_{ij} + h_{ij})\mathrm{d}x^i \mathrm{d}x^j] \,, \tag{3.40}$$

with $h^i_i = 0$ and $\partial_i h^i_j = 0$. A convenient parameterisation of the photon four-momentum in this case is

$$p^\mu = \frac{\epsilon}{a^2}\left[1, e^i - \frac{1}{2}h^i_j e^j\right] \,, \tag{3.41}$$

where $\boldsymbol{e}^2 = 1$ and ϵ is a times the energy of the photon as seen by an observer at constant $\boldsymbol{x}$. The components of $\boldsymbol{e}$ are the projections of the photon

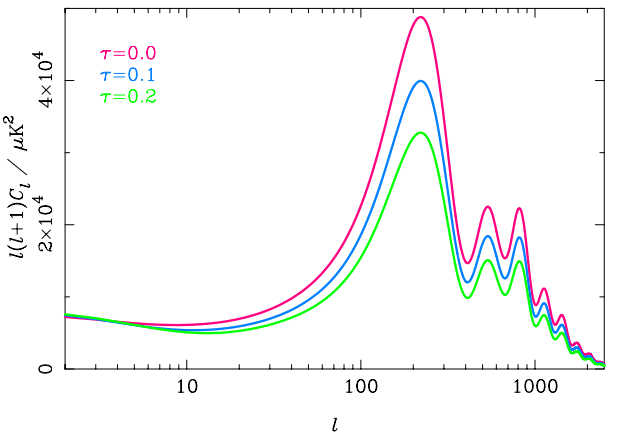

Fig. 3.7. Effect of reionization on the temperature-anisotropy power spectrum. The spectra are (from top to bottom) for no reionization, $\tau_{\rm re} = 0.1$ and 0.2.

direction for this observer on an orthonormal spatial triad of vectors $a^{-1}(\partial_i - h_i^j \partial_j/2)$. In the background $\boldsymbol{e} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\eta$ and is constant. The evolution of the comoving energy ϵ in the perturbed universe is

$$\frac{1}{\epsilon}\frac{\mathrm{d}\epsilon}{\mathrm{d}\eta} + \frac{1}{2}\dot{h}_{ij}e^i e^j = 0\,, \tag{3.42}$$

and so the Boltzmann equation for $\Theta(\eta, \boldsymbol{x}, \boldsymbol{e})$ is

$$\frac{\partial\Theta}{\partial\eta} + \boldsymbol{e}\cdot\boldsymbol{\nabla}\Theta = -an_{\rm e}\sigma_{\rm T}\Theta + \frac{3}{16\pi}an_{\rm e}\sigma_{\rm T}\int\Theta(\boldsymbol{e}')[(1+(\boldsymbol{e}\cdot\boldsymbol{e}')^2]\,\mathrm{d}\boldsymbol{e}' - \frac{1}{2}\dot{h}_{ij}e^i e^j\,. \tag{3.43}$$

Neglecting the anisotropic nature of Thomson scattering, the solution of this equation is an integral along the unperturbed line of sight:

$$\Theta(\hat{\boldsymbol{n}}) = -\frac{1}{2}\int_0^{\eta_0} \mathrm{e}^{-\tau}\dot{h}_{ij}\hat{n}^i\hat{n}^j\,\mathrm{d}\eta\,. \tag{3.44}$$

The time derivative $\dot{h}_{ij}$ is the shear induced by the gravitational waves. This quadrupole perturbation to the expansion produces an anisotropic redshifting of the CMB photons and an associated temperature anisotropy.

Figure 3.8 compares the power spectrum due to gravitational waves (for a review on gravitational waves see the article by N. Andersson and K. Kokkotas) with that from scalar perturbations for a tensor-to-scalar ratio $r = 1$ corresponding to an energy scale of inflation 3.3×10^{16} GeV. The constraints on gravitational waves from temperature anisotropies are not very constraining since their effect is limited to large angular scales where cosmic variance

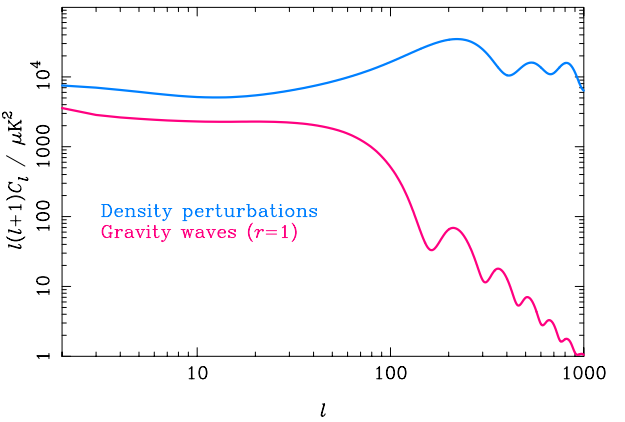

Fig. 3.8. The temperature-anisotropy power spectrum from scalar perturbations (density perturbations; top) and tensor perturbations (gravity waves; bottom) for a tensor-to-scalar ratio $r = 1$.

from the dominant scalar perturbations is large. Gravitational waves damp as they oscillate inside the horizon, so the only significant anisotropies are from wavelengths that are super-horizon at last scattering, corresponding to $l \sim 60$. The current 95-per cent upper limit on the tensor-to-scalar ratio is 0.68 [20]. Fortunately, CMB polarization provides an alternative route to detecting the effect of gravitational waves on the CMB which is not limited by cosmic variance [52, 53]; see also Sect. 3.4.

3.3 Cosmological Parameters and the CMB

The simple, linear physics of CMB temperature anisotropies, reviewed in the previous section, means that the CMB depends sensitively on many of the key cosmological parameters. For this reason, CMB observations over the past decade have been a significant driving force in the quest for precision determinations of the cosmological parameters. It is not our intention here to give a detailed description of the constraints that have emerged from such analyses, e.g. [10], but rather to provide a brief description of how the key parameters affect the temperature-anisotropy power spectrum. More details can be found in the seminal papers on this subject, e.g. [39, 40, 54] and references therein.

3.3.1 Matter and Baryons

The curvature of the universe and the properties of the dark energy are largely irrelevant for the pre-recombination physics of the acoustic oscillations. Their main contribution is felt geometrically through the angular diameter distance to last scattering, $D_{\rm A}$, which controls the projection of linear scales there to angular scales on the sky today. In contrast, those parameters that determine the energy content of the universe before recombination, such as the physical densities in (non-relativistic) matter $\Omega_{\rm m}h^2$, and radiation $\Omega_{\rm r}h^2$ (determined by the CMB temperature and the physics of neutrinos), play an important role in acoustic physics by determining the expansion rate and hence the behaviour of the perturbations. In addition, the physical density in baryons, $\Omega_{\rm b}h^2$, affects the acoustic oscillations through baryon inertia and the dependence of the photon mean-free path on the electron density. The effect of variations in the physical densities of the matter and baryon densities on the anisotropy power spectrum is illustrated in Fig. 3.9 for adiabatic initial conditions.

The linear scales at last scattering that have reached extrema of their oscillation are determined by the initial conditions (*i.e.* adiabatic or isocurvature) and the sound horizon $r_{\rm s}(\eta_*)$. Increasing the baryon density holding the total matter density fixed reduces the sound speed while preserving the expansion rate (and moves last scattering to slightly earlier times). The effect is to reduce the sound horizon at last scattering and so the wavelength of those modes that are at extrema of their oscillation, and hence push the acoustic peaks to smaller scales. This effect could be confused with a change in the angular diameter distance $D_{\rm A}$, but fortunately baryons have another distinguishing effect. Their inertia shifts the zero point of the acoustic oscillations to $\sim -(1+R)\psi$, and enhances the amplitude of the oscillations. In adiabatic models for modes that enter the sound horizon in matter domi-

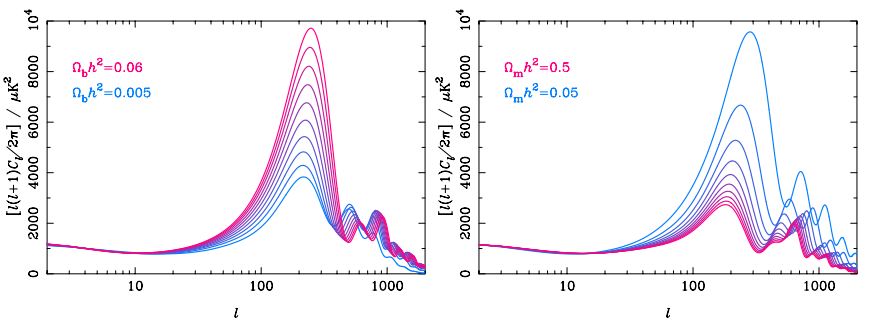

Fig. 3.9. Dependence of the temperature-anisotropy power spectrum on the physical density in baryons (left) and all non-relativistic matter (right). From top to bottom at the first peak, the baryon densities vary linearly in the range $\Omega_{\rm b}h^2 = 0.06$–0.005 (left) and the matter densities in $\Omega_{\rm m}h^2 = 0.05$–0.5 (right). The initial conditions are adiabatic.

nation, $\delta_\gamma/4$ starts out at $-2\psi/3$, and so the amplitude of the oscillation is $-\psi(1+3R)/3$. The combination of these two effects is to enhance the amplitude of $\Theta_0+\psi$ at maximal compression by a factor of $1+6R$ over that at minimal compression. The effect on the power spectrum is to enhance the amplitude of the 1st, 3rd etc. peaks for adiabatic initial conditions, and the 2nd, 4th etc. for isocurvature. Current CMB data gives $\Omega_\mathrm{b}h^2 = 0.023 \pm 0.001$ for power-law ΛCDM models [12], beautifully consistent with determinations from big bang nucleosynthesis. Other effects of baryons are felt in the damping tail of the power spectrum since increasing the baryon density tends to inhibit diffusion giving less damping at a given scale.

The effect of increasing the physical matter density $\Omega_\mathrm{m}h^2$ at fixed $\Omega_\mathrm{b}h^2$ is also two-fold (see Fig. 3.9): (i) a shift of the peak positions to larger scales due to the increase in D_A; and (ii) a scale-dependent reduction in peak height in adiabatic models. Adiabatic modes that enter the sound horizon during radiation domination see the potentials decay as the photon density rises to reach maximal compression. This decay tends to drive the oscillation, increasing the oscillation amplitude. Raising $\Omega_\mathrm{m}h^2$ brings matter–radiation equality to earlier times, and reduces the efficiency of the gravitational driving effect for the low-order peaks. Current CMB data gives $\Omega_\mathrm{m}h^2 = 0.13 \pm 0.01$ for adiabatic, power-law ΛCDM models [12].

3.3.2 Curvature, Dark Energy and Degeneracies

The main effect of curvature and dark energy (see V. Sahni's contribution) on the linear CMB anisotropies is through the angular diameter distance and the late-time integrated Sachs–Wolfe effect; see Fig. 3.10 for the case of adiabatic fluctuations in cosmological-constant models. The ISW contribution is limited to large scales where cosmic variance severely limits the precision of power

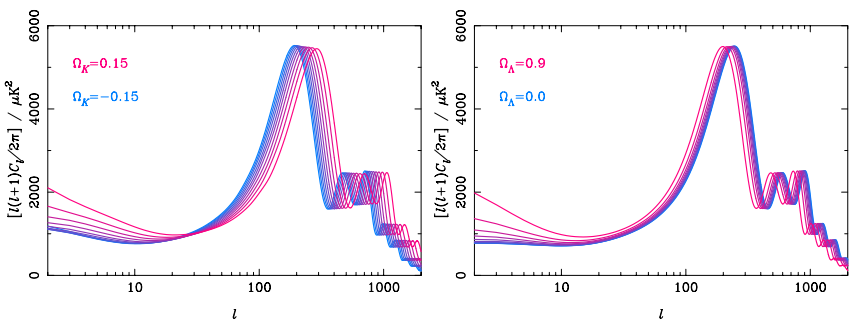

Fig. 3.10. Dependence of the temperature-anisotropy power spectrum on the curvature Ω_K (left) and cosmological constant Ω_Λ (right) in adiabatic models. In both cases, the physical densities in baryons and matter were held constant, thus preserving the conditions on the last scattering surface. The curvature varies (left to right) in the range -0.15–0.15 and the cosmological constant in the range 0.9–0.0.

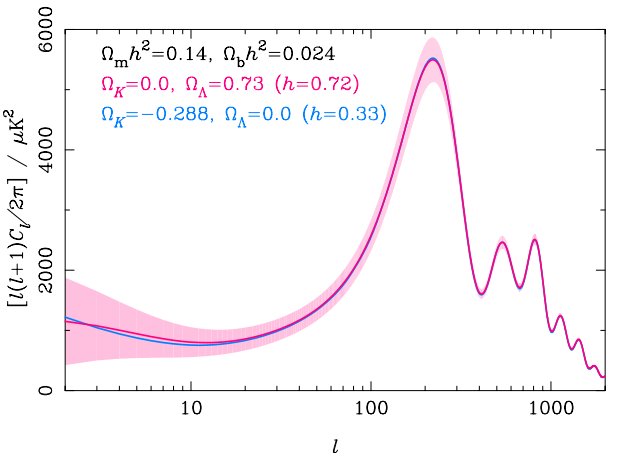

Fig. 3.11. The geometric degeneracy. A scale-invariant adiabatic ΛCDM model with $\Omega_b h^2 = 0.024$, $\Omega_m h^2 = 0.14$ and $\Omega_\Lambda = 0.73$ and $\Omega_K = 0$ (close to the WMAP best-fit values [12]) produces an almost identical spectrum to a closed model $\Omega_K = -0.288$ with vanishing cosmological constant. However, the Hubble constants are very different – $h = 0.72$ in the flat model and 0.33 in the closed model – and so the latter is easily ruled out by external constraints. The shaded region shows the 1σ cosmic variance errors $\Delta C_l/C_l = \sqrt{2/(2l+1)}$ on the power spectrum.

spectrum estimates. There is an additional small effect due to quantisation of the allowed spatial modes in closed models (e.g. [55]), but this is also confined to large scales (*i.e.* near the angular projection of the curvature scale). Most of the information that the CMB encodes on curvature and dark energy is thus locked in the angular diameter distance to last scattering, D_A.

With the physical densities $\Omega_b h^2$ and $\Omega_m h^2$ fixed by the acoustic part of the anisotropy spectrum, D_A can be considered a function of Ω_K and the history of the energy density of the dark energy (often modelled through its current density and a constant equation of state). In cosmological constant models D_A is particularly sensitive to the curvature: the 95-per cent interval from WMAP alone (with the weak prior $H_0 > 50\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$) is $-0.08 < \Omega_K < 0.02$, so the universe is close to being spatially flat. The fact that the impact of curvature and the properties of the dark energy on the CMB is mainly through a single number D_A leads to a geometrical degeneracy in parameter estimation [56], as illustrated in Fig. 3.11. Fortunately, this is easily broken by including other, complementary cosmological datasets. The constraint on curvature from WMAP improves considerably when supernovae measurements [57, 58], or the measurement of H_0 from the Hubble Space Telescope Key Project [59] are included. Other examples of near-perfect degeneracies for the temperature anisotropies include the addition of gravity waves and a reduction in the amplitude of the initial fluctuations mimicing

the effect of reionization. This degeneracy is broken very effectively by the polarization of the CMB.

3.4 CMB Polarization

The growth in the mean-free path of the CMB photons during recombination allowed anisotropies to start to develop. Subsequent scattering of the radiation generated (partial) linear polarization from the quadrupole anisotropy. This linear polarization signal is expected to have an r.m.s. $\sim 5\,\mu$K, and, for scalar perturbations, to peak around multipoles $l \sim 1000$ corresponding to the angle subtended by the mean-free path around last scattering. The detection of CMB polarization was first announced in 2002 by the Degree Angular Scale Interferometer (DASI) team [13]; WMAP has also detected the polarization indirectly through its correlation with the temperature anisotropies [19]. A direct measurement of the polarization power from two-years of WMAP data is expected shortly. Polarization is only generated by scattering, and so is a sensitive probe of conditions at recombination. In addition, large-angle polarization was generated by subsequent re-scattering as the universe reionized, providing a unique probe of the ionization history at high redshift.

3.4.1 Polarization Observables

Polarization is conveniently described in terms of Stokes parameters I, Q, U and V, where I is the total intensity discussed at length in the previous section. The parameter V describes circular polarization and is expected to be zero for the CMB since it is not generated by Thomson scattering. The remaining parameters Q and U describe linear polarization. They are the components of the trace-free, (zero-lag) correlation tensor of the electric field in the radiation, so that for a quasi-monochromatic plane wave propagating along the z direction

$$\begin{pmatrix} \langle E_x^2 - E_y^2\rangle & 2\langle E_x E_y\rangle \\ 2\langle E_x E_y\rangle & -\langle E_x^2 - E_y^2\rangle \end{pmatrix} = \frac{1}{2}\begin{pmatrix} Q & U \\ U & -Q \end{pmatrix}, \tag{3.45}$$

where the angle brackets represent an average on timescales long compared to the period of the wave. For diffuse radiation we define the polarization brightness tensor $\mathcal{P}_{ab}(\hat{\boldsymbol{n}})$ to have components given by (3.45) for plane waves within a bundle around the line of sight $\hat{\boldsymbol{n}}$ and around the specified frequency. The polarization tensor is transverse to the line of sight, and, since it inherits its frequency dependence from the the quadrupole of the total intensity, has a spectrum given by the derivative of the Planck function (see (3.11)).

The polarization tensor can be decomposed uniquely on the sphere into an electric (or gradient) part and a magnetic (or curl) part [52, 53]:

$$\mathcal{P}_{ab} = \nabla_{\langle a}\nabla_{b\rangle} P_E - \epsilon^c{}_{\langle a}\nabla_{b\rangle}\nabla_c P_B\,, \tag{3.46}$$

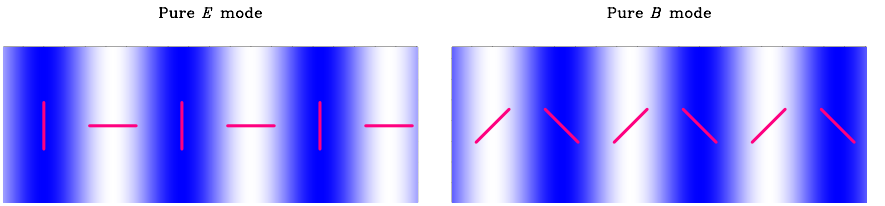

Fig. 3.12. Polarization patterns for a pure-electric mode (left) and pure-magnetic mode (right) on a small patch of the sky for potentials that are locally Fourier modes. The shading denotes the amplitude of the potential. For the electric pattern the polarization is aligned with or perpendicular to the Fourier wavevector depending on the sign of the potential; for the magnetic pattern the polarization is at 45 degrees.

where angle brackets denote the symmetric, trace-free part, ∇_a is the covariant derivative on the sphere, and ϵ_{ab} is the alternating tensor. The divergence $\nabla^a \mathcal{P}_{ab}$ is a pure gradient if the magnetic part $P_B = 0$, and a curl if the electric part $P_E = 0$. The potential P_E is a scalar under parity, but P_B is a pseudo-scalar. For a given potential P, the electric and magnetic patterns it generates (*i.e.* with $P_E = P$ and $P_B = P$ respectively) are related by locally rotating the polarization directions by 45 degrees. The polarization orientations on a small patch of the sky for potentials that are locally Fourier modes are shown in Fig. 3.12. The potentials can be expanded in spherical harmonics (only the $l \geq 2$ multipoles contribute to $\mathcal{P}_{ab}$) as

$$P_E(\hat{\boldsymbol{n}}) = \sum_{lm} \sqrt{\frac{(l-2)!}{(l+2)!}} E_{lm} Y_{lm}(\hat{\boldsymbol{n}}) \,, \quad P_B(\hat{\boldsymbol{n}}) = \sum_{lm} \sqrt{\frac{(l-2)!}{(l+2)!}} B_{lm} Y_{lm}(\hat{\boldsymbol{n}}) \,. \tag{3.47}$$

(The normalisation is conventional.) Under parity $E_{lm} \to (-1)^l E_{lm}$ but $B_{lm} \to -(-1)^l B_{lm}$. Assuming rotational and parity invariance, B is not correlated with E or the temperature anisotropies T, leaving four non-vanishing power spectra: C_l^T, C_l^E, C_l^B and the cross-correlation C_l^{TE}, where e.g. $\langle E_{lm} T^*_{lm} \rangle = C_l^{TE}$.

3.4.2 Physics of CMB Polarization

For scalar perturbations, the quadrupole of the temperature anisotropies at leading order in tight coupling is $\Theta_2 \sim k\dot{\tau}^{-1} v_\gamma$. Scattering of this quadrupole into the direction $-\hat{\boldsymbol{n}}$ generates linear polarization parallel or perpendicular to the projection of the wavevector $\boldsymbol{k}$ onto the sky, i.e. $\mathcal{P}_{ij} \sim \Theta_2 [\hat{k}_{\langle i} \hat{k}_{j \rangle}]^{\mathrm{TT}}$, where TT denotes the transverse (to $\hat{\boldsymbol{n}}$), trace-free part. In a flat universe the polarization tensor is conserved in the absence of scattering; for non-flat models this is still true if the components are defined on an appropriately-propagated basis (e.g. [26]). For a single plane wave perturbation, the polarization on the sky is thus purely electric (see Fig. 3.12). For tensor perturba-

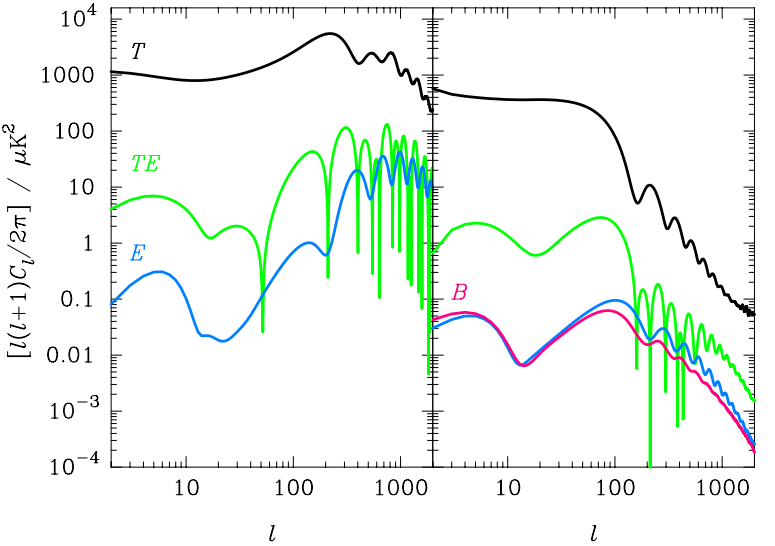

Fig. 3.13. Power spectra produced by adiabatic scalar perturbations (left) and tensor perturbations (right) for a tensor-to-scalar ratio $r = 1$. On large scales the spectra from scalar perturbations are (from top to bottom) C_l^T, C_l^{TE} and C_l^E. For tensor perturbations, they are C_l^T, C_l^{TE}, C_l^B and C_l^E.

tions, the polarization $\mathcal{P}_{ij} \sim \dot{\tau}^{-1}[\dot{h}_{ij}]^{\mathrm{TT}}$ since the tightly-coupled quadrupole is proportional to the shear $\dot{h}_{ij}$. The gravitational wave defines additional directions on the sky when its shear is projected, and the polarization pattern is not purely electric. Thus density perturbations do not produce magnetic polarization in linear perturbation theory, while gravitational waves produce both electric and magnetic [52, 53].

The polarization power spectra produced by scalar and tensor perturbations are compared in Fig. 3.13. The scalar C_l^E spectrum peaks around $l \sim 1000$ since this corresponds to the projection of linear scales at last scattering for which diffusion generates a radiation quadrupole most efficiently. The polarization probes the photon bulk velocity at last scattering, and so C_l^E peaks at the troughs of C_l^T, while C_l^{TE} is zero at the peaks and troughs, and has its extrema in between. For adiabatic perturbations, the large-scale cross-correlation changes sign at $l \sim 50$, and, with the conventions adopted here[3] is positive between $l = 50$ and the first acoustic peak in C_l^T. Isocur-

[3] The sign of E_{lm} for a given polarization field depends on the choice of conventions for the Stokes parameters and their decomposition into electric and magnetic multipoles. We follow [60], which produces the same sign of C_l^{TE} as [25], but note that the Boltzmann codes CMBFAST [36] and CAMB [37] have the opposite sign.

vature modes produce a negative correlation from $l = 2$ to the first acoustic trough.

Tensor modes produce similar power in electric and magnetic polarization. As gravitational waves damp inside the horizon, the polarization peaks just shortward of the horizon size at last scattering $l \sim 100$ despite these large scales being geometrically less efficient at transferring power to the quadrupole during a mean-free time than smaller scales.

For both scalar and tensor perturbations, the polarization would be small on large scales were it not for reionization, since a significant quadrupole is only generated at last scattering when the mean-free path approaches the wavelength of the fluctuations. However, reionization does produce significant large-angle polarization [61] (see Fig. 3.13). The temperature quadrupole at last scattering peaks on linear scales with $k(\eta_{\rm re} - \eta_*) \sim 2$, which then reprojects onto angular scales $l \sim 2(\eta_0 - \eta_{\rm re})/(\eta_{\rm re} - \eta_*)$. The position of the reionization feature is thus controlled by the epoch of reionization, and the height by the fraction of photons that scatter there *i.e.* $\tau_{\rm re}$. The measurement of $\tau_{\rm re}$ with large-angle polarization allows an accurate determination of the amplitude of scalar fluctuations from the temperature-anisotropy power spectrum. In addition, the fine details of the large-angle polarization power can in principle distinguish different ionization histories with the same optical depth, although this is hampered by the large cosmic variance at low l [62].

3.5 Highlights of Recent Results

In this section we briefly review some of the highlights from recent observations of the CMB temperature and polarization anisotropies. Analysis of the former have entered a new phase with the release of the first year data from the WMAP satellite [11]; a further three years worth of data are expected from this mission. Detections of CMB polarization are still in their infancy, but here too we can expect significant progress from a number of experiments in the short term.

3.5.1 Detection of CMB Polarization

The first detection of polarization of the CMB was announced in September 2002 [13]. The measurements were made with DASI, a compact interferometric array operating at 30 GHz, deployed at the South Pole. The DASI team constrained the amplitude of the E and B-mode spectra with assumed spectral shapes derived from a concordant ΛCDM model. They obtained a ~ 5-σ detection of a non-zero amplitude for E with a central value perfectly consistent with that expected from the amplitude of the temperature anisotropies. DASI also detected the temperature–polarization cross-correlation at 95-per cent significance, but no evidence for B-mode polarization was found. The

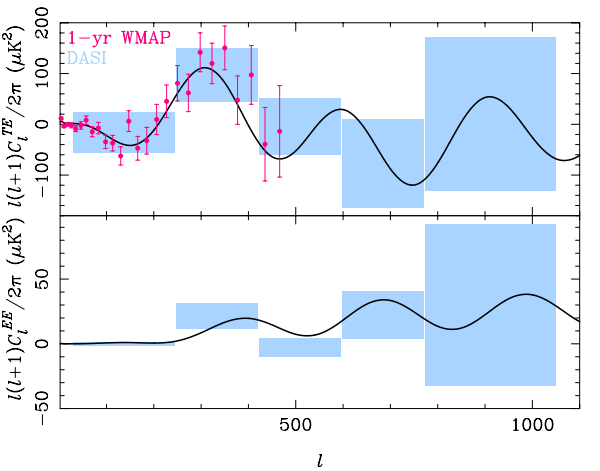

Fig. 3.14. Current measurements (as of February 2004) of C_l^{TE} (top) and C_l^E (bottom). The points with 1-σ errors are from the first one-year data release from WMAP [63]. The error boxes are the flat band-power results from DASI [13] centred on the maximum-likelihood band power and spanning the 68-per cent interval. The solid lines are the predicted power from the best-fit model to all the WMAP data.

DASI results of a maximum-likelihood band-power estimation of the E and TE power spectra are given in Fig. 3.14.

Measurements of C_l^{TE} were also provided in the first-year data release from WMAP, although polarization data itself was not released. These results are also shown in Fig. 3.14. The existence of a cross-correlation between temperature and polarization on degree angular scales provides evidence for the existence of super-horizon fluctuations on the last scattering surface at recombination. This is more direct evidence for such fluctuations than from the large-scale temperature anisotropies alone, since the latter could have been generated gravitationally all along the line of sight. The sign of the cross-correlation and the phase of its acoustic peaks relative to those in the temperature-anisotropy spectrum is further strong evidence for adiabatic fluctuations. The one surprise in the WMAP measurement of C_l^{TE} is the behaviour on large scales. A significant excess correlation over that expected if polarization were only generated at recombination is present on large scales ($l < 20$). The implication is that reionization occurred early, $11 < z_{\rm re} < 30$, giving a significant optical depth for re-scattering: $\tau_{\rm re} = 0.17 \pm 0.04$ at 68-per cent confidence. As mentioned in Sect. 3.2.5, reionization at this epoch is earlier than that expected from observations of quasar absorption spectra and suggests a complex ionization history.

3.5.2 Implications of Recent Results for Inflation

The generic predictions from simple inflation models are that: (i) the universe should be (very nearly) spatially flat; (ii) there should be a nearly scale-invariant spectrum of Gaussian, adiabatic density perturbations giving apparently-super-horizon fluctuations on the last scattering surface; and (iii) there should be a stochastic background of gravitational waves with a nearly scale-invariant (but necessarily not blue) spectrum. The amplitude of the latter is a direct measure of the Hubble rate during inflation, and hence, in slow-roll models, the energy scale of inflation.

As discussed in Sect. 3.3.2, the measured positions of the acoustic peaks constrains the universe to be close to flat. The constraint improves further with the inclusion of other cosmological data. There is no evidence for isocurvature modes in the CMB, although the current constraints are rather weak if general, correlated modes are allowed in the analysis [44]. Several of the cosmological parameters for the isocurvature models most favoured by CMB data are violently at odds with other probes, most notably the baryon density which is pushed well above the value inferred from the abundances of the light-elements. There is also no evidence for primordial non-Gaussianity in the CMB (see e.g. [64])[4].

Within flat ΛCDM models with a power-law spectrum of curvature fluctuations, the spectral index is constrained by the CMB to be close to scale invariant [12], although the inclusion of the latest data from small-scale experiments, such as CBI [70] and VSA [71], tends to pull the best fit from WMAP towards redder power-law spectra: e.g. $n_{\rm s} = 0.97^{+0.06}_{-0.03}$ at 68-per cent confidence combining WMAP and VSA [71]. Slow-roll inflation predicts that the fluctuation spectrum should be close to a power law, with a run in the spectral index that is second order in slow roll: $\mathrm{d}n_{\rm s}/\mathrm{d}\ln k \sim (n_{\rm s}-1)^2$. The WMAP team reported weak evidence for a running spectral index by including small-scale data from galaxy redshift surveys and the Lyman-α forest, but modelling uncertainties in the latter have led many to question the reliability of this result (e.g. [72]). New data from CBI and VSA now provide independent evidence for running in flat ΛCDM models at the 2-σ level from the CMB alone. This reflects the tension between the spectral index favoured by the low-l CMB data (which is anomalously low for $l < 10$, favouring bluer spectra) and the high-l data from the interferometers. The evidence for running is weakened considerably with the inclusion of external priors from large-scale structure data. The best-fit values for the run in $n_{\rm s}$ obtained with the CMB alone are uncomfortably large for slow-roll inflation models, and give low power on small scales that is difficult to reconcile with the early reionization implied by the WMAP polarization data. However, a recent analysis [73]

[4] The WMAP data does appear to harbour some statistically-significant departures from rotational invariance [65, 66, 67, 68, 69]. The origin of these effects, *i.e.* primordial or systematic due to instrument effects or imperfect foreground subtraction, is as yet unclear.

argues that the evidence for running depends crucially on the techniques employed to estimate the low-l power from WMAP data, and that the running is strongly suppressed if exact likelihood techniques are adopted. A definitive answer on whether departures from power-law spectra are significant must probably await further data on both large and small scales.

The final prediction of slow-roll inflation – the generation of nearly scale-invariant background of gravitational waves – is yet to be verified. The current limits on the tensor-to-scalar ratio are only weak: [71] report $r < 0.68$ at 95-per cent confidence from all CMB data in general, non-flat, adiabatic ΛCDM models. Despite this, observations are beginning to place interesting constraints on specific models of inflation in the r–n_s plane [74, 75]. Already, large-field models with power-law potentials steeper than $V \propto \phi^6$ are ruled out due to their red scalar spectra and comparatively large tensor-to-scalar ratio. Future programmes targeting B-mode polarization may ultimately be able to detect gravitational waves down to an inflationary energy scale of a few $\times\, 10^{15}$ GeV. Such observations will sharpen constraints in the r–n_s plane considerably, and should allow fine selection amongst the many proposed models of inflation.

3.5.3 Detection of Late-Time Integrated Sachs–Wolfe Effect

The late-time ISW effect arises from the decay of the gravitational potentials once the universe becomes dark-energy dominated, and so should produce large-angle (positive) correlations between the CMB temperature anisotropies and other tracers of the potential in the local universe. With the advent of the WMAP data, a number of groups have reported the detection of such a correlation. In [47], WMAP data was cross-correlated with data on the hard X-ray background (which is dominated by emission from active galaxies) from the HEAO-1 satellite, and the number density of radio sources from the NVSS catalogue. In each case a positive correlation was detected at significance 3σ and 2.5σ respectively. The correlation with NVSS has also been carried out independently by the WMAP team [48], who also note that the observed positive correlation can be used to rule out the closed, $\Lambda = 0$ model that is a good fit to the CMB data in isolation (see Fig. 3.11). Several groups have now also detected the cross-correlation on large scales between the CMB and optical galaxy surveys, e.g. [49].

3.6 Conclusions

The linear anisotropies of the cosmic microwave background have been studied theoretically for over three decades. The physics, which is now well understood, employs linearised radiative transfer, general relativity, and hydrodynamics to describe the propagation of CMB photons and the evolution of the fluid constituents in a perturbed Friedmann-Robertson-Walker universe.

A number of bold predictions have emerged from this theoretical activity, most notably the existence of acoustic peaks in the anisotropy power spectrum due to oscillations in the photon–baryon plasma prior to recombination. Observers have risen to the challenge of verifying these predictions, and their detection is proceeding at a staggering rate. The large-scale Sachs–Wolfe effect, acoustic peak structure, damping tail, late-time integrated Sachs–Wolfe effect, polarization and reionization signature have all been detected, and the first three have been measured in considerable detail. Already, the size and scale of these effects is allowing cosmological models to be constrained with unprecedented precision. The results are beautifully consistent with almost-scale-invariant adiabatic initial conditions evolving passively in a spatially flat, ΛCDM universe.

Much work still remains to be done to exploit fully the information contained in the CMB anisotropies. The Planck satellite, due for launch in 2007, should provide definitive mapping of the linear CMB anisotropies, and a cosmic-variance limited measurement of the power spectrum up to multipoles $l \sim 2000$. This dataset will be invaluable in assessing many of the issues hinted at in the first-year release of WMAP data, such as the apparent lack of power on large scales and possible violations of rotational (statistical) invariance. Prior to Planck, a number of ground-based programmes should shed further light on the issue of whether departures from a power-law primordial spectrum are required on cosmological scales, and the implications of this for slow-roll inflation. In addition, these small-scale observations will start to explore the rich science of secondary anisotropies, due to e.g. scattering in hot clusters [76] or bulk flows modulated by variations in the electron density in the reionized universe [77, 78], and the weak lensing effect of large-scale structure [79].

Detections of CMB polarization are in their infancy, but we can expect rapid progress on this front too. Accurate measurements of the power spectra of E-mode polarization, and its correlation with the temperature anisotropies, can be expected from a number of ground and balloon-borne experiments, as well as from Planck. The ultimate goal for CMB polarimetry is to detect the B-mode signal predicted from gravitational waves. This would give a direct measure of the energy scale of inflation, and, when combined with measurements of the spectrum density perturbations, place tight constraints on the dynamics of inflation. Plans are already being made for a new generation of polarimeters with the large numbers of detectors and exquisite control of instrument systematics needed to detect the gravity-wave signal if the energy scale of inflation is around 10^{16} GeV. Ultimately, confusion due to imperfect subtraction of astrophysical foregrounds and the effects of weak lensing on the polarization limit will limit the energy scales that we can probe with CMB polarization; see [80] and references therein.

Acknowledgments

The author acknowledges a Royal Society University Research Fellowship.

References

1. A.A. Penzias, R.W. Wilson, Astrophys. J. **142**, 419 (1965).
2. R.H. Dicke et al, Astrophys. J. **142**, 414 (1965).
3. R.K. Sachs, A.M. Wolfe, Astrophys. J. **147**, 73 (1967).
4. G.F. Smoot et al, Astrophys. J. Lett. **396**, 1 (1992).
5. P.J.E. Peebles, Astrophys. J. Lett. **263**, 1 (1982).
6. J.R. Bond, G. Efstathiou, Astrophys. J. Lett. **285**, 45 (1984).
7. P.J.E. Peebles J.T. Yu, Astrophys. J. **162**, 815 (1970).
8. J.C. Mather et al, Astrophys. J. **420**, 439 (1994).
9. J.C. Mather et al, Astrophys. J. **512**, 511 (1999).
10. J.R. Bond, C.R. Contaldi, D. Pogosyan, Phil. Trans. Roy. Soc. Lond. A **361**, 2435 (2003).
11. C.L. Bennett et al, Astrophys. J. Suppl. **148**, 1 (2003).
12. D.N. Spergel et al, Astrophys. J. Suppl. **148**, 175 (2003).
13. J.M. Kovac et al, Nature **420**, 772 (2002).
14. W. Hu, S. Dodelson, Ann. Rev. Astron. Astrophys. **40**, 171 (2002).
15. W. Hu, Ann. Phys. **303**, 203 (2003).
16. P.J.E. Peebles, Astrophys. J. **153**, 1 (1968).
17. Y.B. Zeldovich, V.G. Kurt, R.A. Syunyaev, Journal of Experimental and Theoretical Physics **28**, 146 (1969).
18. S. Seager, D.D. Sasselov, D. Scott, Astrophys. J. Suppl. **128**, 407 (2000).
19. A. Kogut et al, Astrophys. J. Suppl. **148**, 161 (2003).
20. C. Dickinson et al, preprint astro-ph/0402498, (2004).
21. B.S. Mason et al, Astrophys. J. **591**, 540 (2003).
22. T.J. Pearson et al, Astrophys. J. **591**, 556 (2003).
23. C.L. Kuo et al, Astrophys. J. **600**, 32 (2004).
24. C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation*, (W. H. Freeman and Company, San Francisco 1973) pp 583–590.
25. W. Hu et al, Phys. Rev. D **57**, 3290 (1998).
26. A. Challinor, Phys. Rev. D **62**, 043004 (2000).
27. G.F.R. Ellis, J. Hwang, M. Bruni, Phys. Rev. D **40**, 1819 (1989).
28. A. Challinor, A. Lasenby, Astrophys. J. **513**, 1 (1999).
29. A. Challinor, Class. Quantum Grav. **17**, 871 (2000).
30. R. Maartens, T. Gebbie, G.F.R. Ellis, Phys. Rev. D **59**, 083506 (1999).
31. T. Gebbie, G.F.R. Ellis, Ann. Phys. **282**, 285 (2000).
32. T. Gebbie, P.K.S. Dunsby, G.F.R. Ellis, Ann. Phys. **282**, 321 (2000).
33. C. Ma, E. Bertschinger, Astrophys. J. **455**, 7 (1995)
34. M. Bartelmann, P. Schneider: Phys. Rep. **340**, 291 (2001).
35. M.L. Wilson, Astrophys. J. **273**, 2 (1983).
36. U. Seljak, M. Zaldarriaga, Astrophys. J. **469**, 437 (1996).
37. A. Lewis, A. Challinor, A. Lasenby, Astrophys. J. **538**, 473 (2000).
38. N. Kaiser, Mon. Not. R. Astron. Soc. **202**, 1169 (1983).
39. W. Hu, N. Sugiyama, Astrophys. J. **444**, 489 (1995).

40. W. Hu, N. Sugiyama, Phys. Rev. D **51**, 2599 (1995).
41. M. Bucher, K. Moodley, N. Turok, Phys. Rev. D **62**, 083508 (2000).
42. L. Page et al, Astrophys. J. Suppl. **148**, 233 (2003).
43. C. Gordon, A. Lewis, Phys. Rev. D **67**, 123513 (2003).
44. M. Bucher et al, preprint astro-ph/0401417, (2004).
45. J. Silk, Astrophys. J. **151**, 459 (1968).
46. J.M. Stewart, Class. Quantum Grav. **7**, 1169 (1990).
47. S. Boughn, R. Crittenden, Nature **427**, 45 (2004).
48. M.R. Nolta et al, preprint astro-ph/0305097, (2003).
49. P. Fosalba, E. Gaztañaga and F.J. Castander, Astrophys. J. Lett. **597**, 89 (2003).
50. R.H. Becker et al, Astron. J. **122**, 2850 (2001).
51. S.G. Djorgovski et al, Astrophys. J. Lett. **560**, 5 (2001).
52. U. Seljak, M. Zaldarriaga, Phys. Rev. Lett. **78**, 2054 (1997).
53. M. Kamionkowski, A. Kosowsky, A. Stebbins, Phys. Rev. Lett. **78**, 2058 (1997).
54. J.R. Bond et al, Phys. Rev. Lett. **72**, 13 (1994).
55. L.F. Abbott, R.K. Schaefer, Astrophys. J. **308**, 546 (1986).
56. G. Efstathiou, J.R. Bond, Mon. Not. R. Astron. Soc. **304**, 75 (1999).
57. A.G. Riess et al, Astron. J. **116**, 1009 (1998).
58. S. Perlmutter et al, Astrophys. J. **517**, 565 (1999).
59. W.L. Freedman et al, Astrophys. J. **553**, 47 (2001).
60. A. Lewis, A. Challinor, N. Turok, Phys. Rev. D **65**, 023505 (2002).
61. M. Zaldarriaga, Phys. Rev. D **55**, 1822 (1997).
62. G. Holder et al, Astrophys. J. **595**, 13 (2003).
63. G. Hinshaw et al, Astrophys. J. Suppl. **148**, 135 (2003).
64. E. Komatsu et al, Astrophys. J. Suppl. **148**, 119 (2003).
65. A. de Oliveria-Costa et al, preprint astro-ph/0307282, (2003).
66. P. Vielva et al, preprint astro-ph/0310273, (2003).
67. C.J. Copi, D. Huterer, G.D. Starkman, preprint astro-ph/0310511, (2003).
68. H.K. Erikse et al, preprint astro-ph/0401276, (2004).
69. F.K. Hansen et al, preprint astro-ph/0402396, 2004.
70. A.C.S. Readhead et al, preprint astro-ph/0402359, (2004).
71. R. Rebolo et al, preprint astro-ph/0402466, (2004).
72. U. Seljak, P. McDonald, A. Makarov, Mon. Not. R. Astron. Soc. **342**, L79 (2003).
73. A. Slosar, U. Seljak, A. Makarov, preprint astro-ph/0403073, (2004).
74. S.M. Leach, A.R. Liddle, Mon. Not. R. Astron. Soc. **341**, 1151 (2003).
75. M. Tegmark et al, preprint astro-ph/0310723, (2003).
76. R.A. Sunyaev, Y.B. Zeldovich, Comm. Astrophys. Space Phys. **4**, 173 (1972).
77. R.A. Sunyaev, Y.B. Zeldovich, Mon. Not. R. Astron. Soc. **190**, 413 (1980).
78. J.P. Ostriker, E.T. Vishniac, Astrophys. J. Lett. **306**, 51 (1986).
79. A. Blanchard, J. Schneider, Astron. Astrophys. **184**, 1 (1987).
80. C.M. Hirata, U. Seljak, Phys. Rev. D **68**, 083002 (2003).

4 Observational Cosmology

Robert H. Sanders

Kapteyn Astronomical Institute, Groningen, The Netherlands

Abstract. I discuss the classical cosmological tests, *i.e.,* angular size-redshift, flux-redshift, and galaxy number counts, in the light of the cosmology prescribed by the interpretation of the CMB anisotropies. The discussion is somewhat of a primer for physicists, with emphasis upon the possible systematic uncertainties in the observations and their interpretation. Given the curious composition of the Universe inherent in the emerging cosmological model, I stress the value of searching for inconsistencies rather than concordance, and suggest that the prevailing mood of triumphalism in cosmology is premature.

4.1 Introduction

The traditional cosmological tests appear to have been overshadowed by observations of the anisotropies in the cosmic microwave background (CMB). We are told that these observations accurately measure the geometry of the Universe, its composition, its present expansion rate, and the nature and form of the primordial fluctuations [1]. The resulting values for these basic parameters are very similar to those deduced earlier from a variety of observations, the so-called "concordance model", with about 30% of the closure density of the Universe comprised of matter (mostly a pressureless, non-baryonic dark matter), the remainder being in negative pressure dark energy [2]. Given the certainty and precision of these assertions, any current discussion of observational cosmology must begin with the question: Is there any room for doubt? Why should we bother with lower precision cosmological tests when we know all of the answers anyway?

While the interpretation of the CMB anisotropies has emerged as the single most important cosmological tool, we must bear in mind that the conclusions drawn do rest upon a number of assumptions, and the results are not altogether as robust as we are, at times, led to believe. One such assumption, for example, is that of adiabatic initial fluctuations, that is 100% adiabatic. A small admixture of correlated isocurvature fluctuations, an aspect of braneworld scenarios [3], can affect peak amplitudes and thus, the derived cosmological parameters. A more fundamental assumption is that of the validity of traditional Friedmann-Robertson-Walker (FRW) cosmology in the post-decoupling universe. Is the expansion of the universe described by the Friedmann equation? Even minimal changes to the right-hand-side, such

R.H. Sanders, Observational Cosmology, Lect. Notes Phys. **653**, 105–137 (2005)
http://www.springerlink.com/

as the equation of state of the dark energy component, can alter the angular size distance to the last scattering surface at z=1000 and the luminosity distance to distant supernovae. But even more drastic changes to the Friedmann equation, resulting from modified gravitational physics, have been proposed in attempts to remove the unattractive dark energy [4, 5].

Such suggestions reflect a general unease with the concordance model, a model that presents us with a universe that is strange in its composition. The most abundant form of matter consists of, as yet, undetected non-baryonic particles originally postulated to solve the problems of structure formation and of the missing mass in bound gravitational systems such as galaxies and clusters of galaxies. In this second respect, it is fair to say that it has failed or, to be generous, not yet succeeded, because the predicted density distribution of dark halos which emerge from cosmic N-body [6] simulations appears to be inconsistent with observations of spiral galaxies [7] or with strong lensing in clusters of galaxies [8].

Even more mysterious is the "dark energy", the pervasive homogeneous fluid with a negative pressure which may be identified with the cosmological constant, the zero-point energy density of the vacuum. The problem of this unnaturally low energy density, 10^{-122} in Planck units, is well-known, as is the cosmic coincidence problem: why are we observing the Universe at a time when the cosmological constant has, fairly recently, become dynamically important [9]? To put it another way, why are the energy densities of matter and dark energy so comparable at the present epoch? This is strange because the density of matter dilutes with the expanding volume of the Universe while the vacuum energy density does not. It is this problem which has led to the proposal of dynamic dark energy, quintessence (a dark energy, possibly associated with a light scalar field), with an energy density that evolves with cosmic time possibly tracking the matter energy density [10, 11]. Here the difficulty is that the field would generally be expected to have additional observational consequences, such as violations of the equivalence principle at some level, possibly detectable in fifth force experiments [9].

For these reasons, it is even more important to pursue cosmological tests that are independent of the CMB, because one might expect new physics to appear as observations inconsistent with the concordance model. In this sense, discord is more interesting than concord; to take a Hegelian point of view– ideas progress through dialectic, not through concordance. It is with this in mind that I will review observational cosmology with emphasis upon CMB-independent tests.

Below I argue that the evolution of the early, pre-recombination universe is well-understood and tightly constrained by considerations of primordial nucleosynthesis. If one wishes to modify general relativity to give deviations from Friedmann expansion, then such modifications are strongly constrained at early times, at energies on the order of 1 MeV. However, cosmological evolution is much less constrained in the post-recombination universe where there is room for deviation from standard Friedmann cosmology and where

the more classical tests are relevant. I will discuss three of these classical tests: the angular size distance test where I am obliged to refer to its powerful modern application with respect to the CMB anisotropies; the luminosity distance test and its application to observations of distant supernovae; and the incremental volume test as revealed by faint galaxy number counts.

These classical tests yield results that are consistent, to lower precision, with the parameters deduced from the CMB. While one can make minimal changes to standard cosmology, to the equation of state of the dark energy for example, which yield different cosmological parameters, there is no compelling observational reason to do so. It remains the peculiar composition and the extraordinary coincidences embodied by the concordance model that call for deeper insight. Such motivations for questioning a paradigm are not unprecedented; similar worries led to the inflationary scenario which, unquestionably, has had the dominant impact on cosmological thought in the past 25 years and which has found phenomenological support in the recent CMB observations.

I am not going to discuss tests, such as the form of the luminous matter power spectrum [12] or the amplitude of the present mass fluctuations [13], that are based upon the standard model for structure formation. I do not mean to imply that such such tests are unimportant; it is only that I restrict myself here to the classical tests because these address more directly the issues of the global curvature and expansion history of the Universe.

I am also going to refrain, in so far as possible, from discussion of theory of new gravitational physics or of any other sort. The theoretical issues presented by dark matter that can only be detected gravitationally or by an absurdly small but non-zero cosmological constant are essentially not problems for the interpretive astronomer. The primary task is to realistically access the reliability of conclusions drawn from the observations, and that is what I intend to do.

4.2 Astronomy Made Simple (for Physicists)

I think that it is fair to assume that most of you are physicists, so I begin by defining some of the units and terminology used by astronomers. I do this because much of this terminology is arcane for those not in the field.

First of all there is the peculiar logarithmic scale of flux magnitudes, whereby a factor of 100 in flux is divided into five equal logarithmic intervals. The system is ancient and has its origin in the logarithmic response of the human eye. The ratio of the flux of two objects is then given by a difference in magnitudes, *i.e.,*

$$m_2 - m_1 = -2.5\, log(F_2/F_1) \tag{4.1}$$

where, one will notice, smaller magnitude means larger flux. The zero-point of this logarithmic scale is set by some standard star such as Vega. Because

this is related to the flux, and not the luminosity of an object, it is called the "apparent" magnitude. Distant galaxies have apparent magnitudes, in visible light, of greater than 20, and the galaxies in the Hubble Deep Field, go down to magnitudes of 30. The magnitude is typically measured over a specified wavelength range or color band, such as blue (B), visual (V), or infrared (K), and these are designated m_B, m_V, and m_K, or sometimes just B,V, and K. This is made more confusing by the fact that there are several competing photometric systems (or sets of filters) and conversion between them is not always simple.

With a particular photometric system one can measure the color of an astronomical object, expressed as difference in magnitudes in two bands, or color index; e.g.,

$$B - V = 2.5\, log(F_V/F_B)\ . \tag{4.2}$$

Here a larger B-V color index means that an object is relatively redder; a smaller B-V that the object is bluer. Unlike the apparent magnitude, this is an intrinsic property of the object. Or rather, it is intrinsic once the astronomer corrects the magnitudes in the various bands to the zero-redshift ($z = 0$) frame. This is called the "K-correction" and requires a knowledge of the intrinsic spectral energy distribution (SED) of the source, which could be a galaxy or a distant supernova.

The luminosity of an object is also an intrinsic property and is usually expressed by astronomers as an "absolute" magnitude. This is the apparent magnitude an object would have if it were placed at a standard distance, taken to be 10 parsecs, *i.e.,* 3×10^{17} m (more on parsecs below). Because this distance is small by extragalactic standards the absolute magnitudes of galaxies turn out to be rather large negative numbers: $M_G \approx$ -18 to -21. The luminosity of a galaxy L_G in units of the solar luminosity $L_\odot$ can be determined from the relation

$$M_G - M_\odot = -2.5\, log(L_G/L_\odot)\ , \tag{4.3}$$

where the absolute magnitude of the sun (in the V band) is 5.5. The luminosities of galaxies typically range from 10^8 to 10^{11} $L_\odot$. The peak absolute magnitude of a type I supernova (SNIa) is about -19.5, or comparable to an entire galaxy. This is one reason why these objects are such ideal extragalactic distance probes.

The unit of distance used by astronomers is also archaic: the parsec which is about 3×10^{16} m or about 3 light years. This is the distance to a star with an semi-annual parallax of 1 arc second and is not a bad unit when one is discussing the very local region of the galaxy. Our galaxy has a diameter between 10 to 20 kiloparsecs, so the kiloparsec is an appropriate unit when discussing galactic structure. The appropriate unit of extragalactic distance, however, is the "megaparsec" or Mpc, with nearby galaxies being those at distances less than 10 Mpc. The nearest large cluster of galaxies, the Virgo cluster, is at a distance of 20 Mpc, and very distant galaxies are those further

than 100 Mpc, although here one has to be careful about how distance is operationally defined.

We all know that the Universe is uniformly expanding and the Hubble parameter, H, is the recession velocity of galaxies per unit distance, with H_o being its value in the present Universe. It is typically measured in units of km s^{-1}Mpc^{-1} or inverse time. A number of observations point to $H_o \approx 70$ km s^{-1}Mpc^{-1} . The Hubble time is defined as $t_H = {H_o}^{-1}$ which is about $9.8 \times 10^9\ h^{-1}$ years, and this must be comparable to the age of the Universe. The definition $h = H_o/100$ km s^{-1}Mpc^{-1} is a relic of the recent past when the Hubble parameter was less precisely determined, but I keep using it below because it remains convenient as a unit-less quantity. We can also define a characteristic scale for the universe which is the Hubble radius or $r_H = c/H_o$ and this is 3000 h^{-1} Mpc. This would be comparable to the "distance" to the horizon.

Just for interest, one could also define a Hubble acceleration or $a_H = cH_o \approx 7 \times 10^{-10}$ m/s^2. This modest acceleration of 7 angstroms/second squared is, in effect, the acceleration of the Hubble flow at the horizon if we live in a Universe dominated by a cosmological constant as observations seem to suggest. It is also comparable to the acceleration in the outer parts of galaxies where the need for dark matter first becomes apparent [14]. In some sense, it is remarkable that such a small acceleration has led to a major paradigm shift.

4.3 Basics of FRW Cosmology

The fundamental assumption underlying the construction of cosmological models is that of the cosmological principle: The Universe appears spatially isotropic in all its properties to all observers. The only metric which is consistent with this principle is the Robertson-Walker metric:

$$ds^2 = c^2dt^2 - \frac{a^2(t)dr^2}{[1 - r^2/{R_o}^2]} - a^2(t)r^2(d\theta^2 + sin^2(\theta)d\phi^2), \tag{4.4}$$

where r is the radial comoving coordinate, $a(t)$ is the dimensionless scale factor by which all distances vary as a function of cosmic time, and ${R_o}^{-2}$ is a parameter with dimensions of inverse length squared that describes the curvature of the Universe and may be positive, zero, or negative (see [15] for a general discussion).

This is the geometry of the Universe, but dynamics is provided by General Relativity, the Einstein field equations (see K. Tamvakis' review), which yield ordinary differential equations for $a(t)$. The time-time component leads to a second order equation:

$$\ddot{a} = -\frac{4\pi G}{3}a(\rho + 3p/c^2), \tag{4.5}$$

where ρ is the density, p is the pressure and the quantity in parenthesis is the active gravitational mass density. Considering conservation of energy for a perfect fluid

$$d(\rho V) = -pdV/c^2, \tag{4.6}$$

with an equation of state

$$p = w\rho c^2, \tag{4.7}$$

we have $\rho \propto a^{-1(1+w)}$. The equation of state combined with (4.5) tells us that the Universe is accelerating if $w < -1/3$.

The space-space components combined with the time-time component yield the usual first-order Friedmann equation

$$\left(\frac{H}{H_o}\right)^2 - \frac{\Omega_k}{a^2} = \sum_i \Omega_i a^{-3(1+w_i)} \ , \tag{4.8}$$

where $H = \dot{a}/a$ is the running Hubble parameter, the summation is over the various fluids comprising the Universe and

$$\Omega_i = \frac{8\pi G\rho_i}{3{H_o}^2} \tag{4.9}$$

with $\Omega_k = -(r_H/R_o)^2$. We often see (4.8) written in terms of redshift where $a = (1+z)^{-1}$. Each component has its own equation of state parameter, w_i: $w = 0$ for non-relativistic matter (baryons, CDM); $w = 1/3$ for radiation or other relativistic fluid; $w = -1$ for a cosmological constant; and $-1 < w < -1/3$ for "quintessence", dynamic dark energy resulting in ultimate acceleration of the universal expansion. I will not consider $w < -1$ which has been termed "phantom" dark energy [16]; here the effective density increases as the Universe expands (this could be realized by a ghost field, a scalar with a kinetic term in the Lagrangian having the wrong sign so it rolls up rather than down a potential hill).

Given a universe composed of radiation, non-relativistic matter, and quintessence, the Friedmann equation takes its familiar form:

$$\left(\frac{H}{H_o}\right)^2 - \frac{\Omega_k}{a^2} = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_Q a^{-3(1+w)}. \tag{4.10}$$

Here it is evident that radiation drives the expansion at early times ($a << 1$), non-relativistic matter at later times, a non-vanishing curvature ($\Omega_k \neq 0$) at later times still, and, if $w < -1/3$, the vacuum energy density ultimately dominates. I refer to (4.10) with $w = -1$ (the usual cosmological constant) as standard FRW cosmology, while $0 > w \neq -1$ would represent a minimal modification to FRW cosmology. Moreover, when $w = -1$, I replace Ω_Q by Ω_Λ. I will not consider changes to the Friedmann equation which might result from modified gravitational physics.

Because the subject here is observational cosmology we must discuss the operational definitions of distance in an FRW Universe. If there exists a

standard meter stick, an object with a known fixed linear size d which does not evolve with cosmic time, then one could obviously define an angular size distance:

$$D_A = \frac{d}{\theta} \,, \tag{4.11}$$

where θ would be the observed angle subtended by this object. If there exists a standard candle, an object with a known fixed luminosity L which does not vary with cosmic time, then one could also define a luminosity distance:

$$D_L = \big(\frac{L}{4\pi F}\big)^{\frac{1}{2}} \tag{4.12}$$

where F is the measured flux of radiation.

For a RW universe both the angular size distance and the luminosity distance are related to the radial comoving coordinate,

$$r = |R_o|\chi\Big[\frac{r_H}{|R_o|}\int_{\tau_o}^{\tau}\frac{d\tau}{a(\tau)}\Big] \tag{4.13}$$

where $\tau = tH_o$, ${R_o}^2 = -{r_H}^2/\Omega_k$, and $\chi(x)$ is defined by (1.9) in Tamvakis' introductory contribution. Then it is the case that

$$D_A = r\,a(\tau) = r/(1+z) \tag{4.14}$$

and

$$D_L = r/a(\tau) = r(1+z). \tag{4.15}$$

It is evident that both the angular size distance and the luminosity distance depend upon the expansion history (through $\int d\tau/a(\tau)$) and the curvature (through $\chi(x)$).

The same is true of a comoving volume element:

$$dV = r^2 dr d\Omega \,, \tag{4.16}$$

where here $d\Omega$ is an incremental solid angle. Therefore, if there exists a class of objects with a non-evolving comoving density, then this leads to another possible cosmological test: simply count those objects as a function of redshift or flux.

Below, I am going to consider these measures of distance and volume in the form of three classical cosmological tests:

- Angular size tests which essentially involve the determination of $D_A(z)$. Here one measure θ for objects with a known and (hopefully) standard linear size (such as compact radio sources).
- Luminosity distance tests which involve the measurement of $F(z)$ for presumably standard candles (such as supernova type Ia, SNIa).
- dV/dz test which involve the counts of very faint galaxies as a function of flux and redshift.

But before I come to these classic tests, I want to discuss the evidence supporting the validity of the standard hot Big Bang, as an appropriate description of the early pre-recombination Universe.

4.4 Observational Support for the Standard Model of the Early Universe

The discovery 40 years ago of the cosmic microwave background radiation (CMB) ended, for most people, the old debate about Steady-State vs. the Hot Big Bang. Ten years ago, support for the Hot Big Bang was fortified by the COBE satellite which demonstrated that the CMB has a Planck spectrum to extremely high precision; it is, quite literally, the most perfect black body observed in nature [17]. This makes any model in which the CMB is produced by some secondary process, such as thermal re-radiation of starlight by hot dust, seem extremely difficult, if not impossible, to contrive.

Not only does the background radiation have a thermal spectrum, it is now evident that this radiation was hotter in the past than now as expected for adiabatic expansion of the Universe. This is verified by observations of neutral carbon fine structure lines as well as molecular hydrogen rotational transitions in absorption line systems in the spectra of distant quasars. Here, the implied population of different levels, determined primarily by the background radiation field, is an effective thermometer for that radiation field. One example is provided by a quasar with an absorption line system at z = 3.025 which demonstrates that the temperature of the CMB at this redshift was $12.1^{+1.7}_{-8.2}$ K, consistent with expectations ($T \propto 1+z$) [18].

However, the most outstanding success story for the Hot Big Bang is generally considered to be that of Big Bang Nucleosynthesis (BBN) which, for a given number of relativistic particle species, predicts the primordial abundances of the light isotopes with, effectively, one free parameter: the ratio of baryons-to-photons, η [19]. I want to review this success story, and point out that there remains one evident inconsistency which may be entirely observational, but which alternatively may point to new physics.

We saw above in the Friedmann equation (4.10) that radiation, if present, will always dominate the expansion of the Universe at early enough epochs (roughly at $z \approx 2\times10^4\Omega_m$.) This makes the expansion and thermal history of the Universe particularly simple during this period. The Friedmann equation becomes

$$H^2 = \frac{4\pi G a T^4 N(T)}{3c^2} \ , \tag{4.17}$$

here a is the radiation constant and $N(T)$ is the number of degrees of freedom in relativistic particles. The scale factor is seen to grow as $t^{1/2}$ which means that the age of the Universe is given by $t = 1/2H$. This implies, from (4.17), an age-temperature relation of the form $t \propto T^{-2}$. Putting in numbers, the precise relation is

$$t = \frac{2.5}{{T_{MeV}}^2 N(T)^{\frac{1}{2}}} \ \mathrm{s} \ , \tag{4.18}$$

where the age is given in seconds and T_{MeV} is the temperature measured in MeV. It is only necessary to count the number of relativistic particle species:

$$N(T) = \sum g_B + \frac{7}{8} \sum g_F \ , \tag{4.19}$$

where the sums are over the number of bosonic degrees of freedom (g_B) and fermionic degrees of freedom (g_F). The factor 7/8 is due to the difference in Bose-Einstein and Fermi-Dirac statistics. Adding in all the known species, photons, electrons-positrons (when $T_{MeV} > 0.5$), three types of neutrinos and anti-neutrinos, we find

$$t \approx {T_{MeV}}^{-2} \, \mathrm{s} \tag{4.20}$$

for the age-temperature relation in the early Universe.

When the Universe is less than one second old ($T > 1$ MeV) the weak interactions

$$\begin{aligned} p + e^- &\leftrightarrow n + \nu_e \\ n + e^+ &\leftrightarrow p + \nu_e \\ n &\leftrightarrow p + e^- + \nu_e \ , \end{aligned}$$

are rapid enough to establish equilibrium between these various species. But when T falls below 1 MeV, the reaction rates become slower than the expansion rate of the Universe, and neutrons "freeze out", they fall out of thermal equilibrium, as do the neutrinos. This means the equilibrium ratio of neutrons to protons at $T \approx 1$ MeV is frozen into the expanding soup: $n/p \approx 0.20 - 0.25$. You all know that neutrons outside of an atomic nucleus are unstable particles and decay with a half-life of about 15 minutes. But before that happens there is a possible escape route:

$$n + p \leftrightarrow D + \gamma$$

that is to say, a neutron can combine with a proton to make a deuterium nucleus and a photon. However, so long as the mean energy of particles and photons is greater than the binding energy of deuterium, about 86 KeV, the inverse reaction happens as well; as soon as a deuterium nucleus is formed it is photo-dissociated. This means that it is impossible to build up a significant abundance of deuterium until the temperature of the Universe has fallen below 86 KeV or, looking back at (4.20), until the Universe has become older than about 2.5 minutes. Then all of the remaining neutrons are rapidly processed into deuterium. But the deuterium doesn't stay around for long either. Given the temperature and particle densities prevailing at this epoch, there are a series of two-body reactions by which two deuterons combine to make He^4 and trace amounts of lithium and He^3. These reactions occur at a rate which depends upon the overall abundance of baryons, the ratio of baryons to photons:

$$\eta = n_b/n_\gamma = 274 \, \Omega_b h^2 \times 10^{-10} \ . \tag{4.21}$$

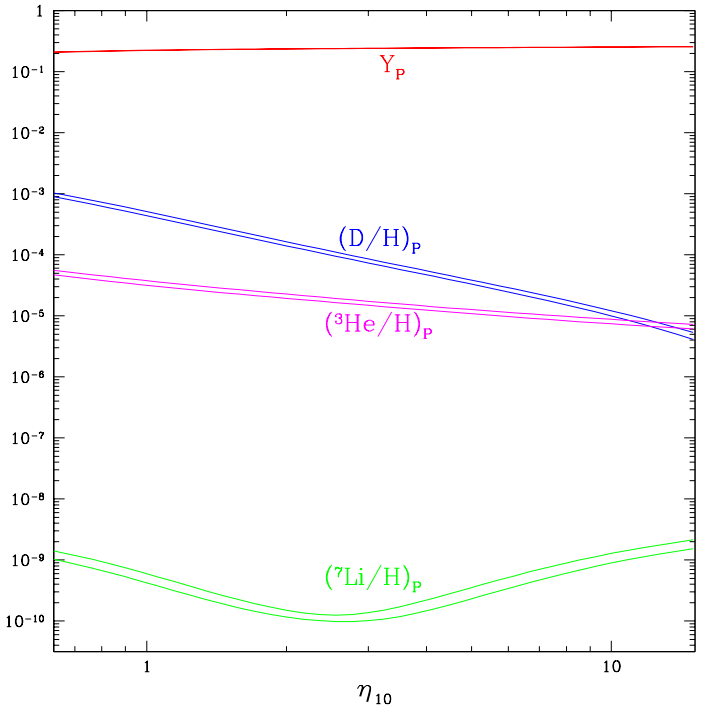

Fig. 4.1. The predicted abundances of the light isotopes as a function of η [19]. Here Y_p is the predicted mass fraction of helium and is based upon the assumption of three neutrino types. The widths of the bands show the theoretical uncertainty.

So essentially all neutrons which survive until T = 86 KeV become locked up in He^4. Therefore, the primordial abundance of helium depends primarily upon the expansion rate of the Universe: the faster the expansion (due, say, to more neutrino types or to a larger constant of gravity) the more helium. The abundance of remaining deuterium, however, depends upon the abundance of baryons, η: the higher η the less deuterium. This is why it is sometimes said [19] that the abundance of primordial helium is a good chronometer (it measures the expansion rate), while the abundance of deuterium is a good baryometer (it measures Ω_b). This is evident in Figs. 4.1 and 4.2 where we see first the predicted abundances of various light isotopes as a function of η, and secondly, the predicted abundance of He vs. that of deuterium for two, three and four neutrino types.

The determination of primordial abundances is not a straightforward matter because the abundance of these elements evolves due to processes occurring within stars ("astration"). In general, the abundance of helium increases (hydrogen is processed to helium providing the primary energy source for stars), while deuterium is destroyed by the same process. This means that astronomers, when trying to estimate primordial abundances of deuterium

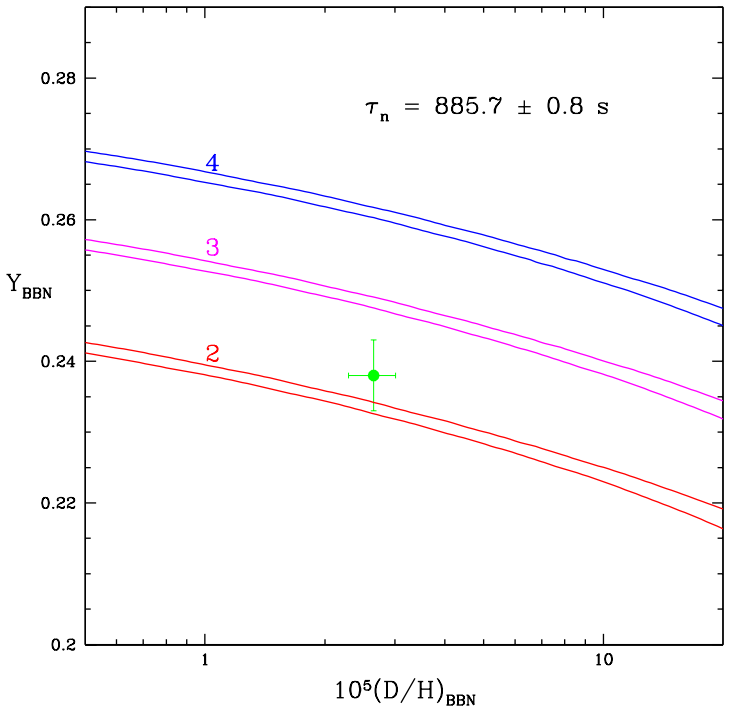

Fig. 4.2. The predicted abundance, Y_p, of helium (the mass fraction) as a function of the predicted deuterium abundance for two, three, and four neutrino types [19]. The point with error bars is the observed abundances of helium and deuterium.

or helium, must try to find pristine, unprocessed material, in so far as possible. One way to find unprocessed material is to look back at early times, or large redshift, before the baryonic material has been recycled through generations of stars. This can be done with quasar absorption line systems, where several groups of observers have been attempting to identify very shallow absorption lines of deuterium at the same redshift as the much stronger hydrogen Lyman alpha absorption line systems [20, 21, 22, 23]. It is a difficult observation requiring the largest telescopes; the lines identified with deuterium might be mis-identified weak hydrogen or metal lines (incidentally, for an astronomer, any element heavier than helium is a metal). Taking the results of various groups at face value, the weighted mean value [19] is D/H $\approx 2.6 \pm 0.3 \times 10^{-5}$. Looking back at Fig. 4.1, we see that this would correspond to $\eta = 6.1 \pm 0.6 \times 10^{-10}$ or $\Omega_b h^2 = 0.022 \pm 0.003$.

A word of caution is necessary here: the values for the deuterium abundance determined by the different groups scatter by more than a factor of two, which is considerably larger than the quoted statistical errors ($\approx 25\%$). This indicates that significant systematic effects are present. But it is noteworthy that the angular power spectrum of the CMB anisotropies also yields

an estimate of the baryon abundance; this is encoded in the ratio of the amplitudes of the second to first peak. The value is $\Omega_b h^2 = 0.024 \pm 0.001$. In other words, the two determinations agree to within their errors. This is quite remarkable considering that the first determination involves nuclear processes occurring within the first three minutes of the Big Bang, and the second involves oscillations of a photon-baryon plasma on an enormous scale when the Universe is about 500,000 years old. If this is a coincidence, it is truly an astounding one.

So much for the baryometer, but what about the chronometer helium? Again astronomers are obliged to look for unprocessed material in order to estimate the primordial abundance. The technique of looking at quasar absorption line systems doesn't work for helium because the absorption lines from the ground state are far in the ultraviolet, about 600 $\mathcal{A}$ for neutral helium and, more likely, 300 $\mathcal{A}$ from singly ionized helium. This is well beyond the Lyman limit of hydrogen, where the radiation from the background quasar is effectively absorbed [24]. Here the technique is to look for He emission lines from HII regions (ionized gas around hot stars) in nearby galaxies and compare to the hydrogen emission lines. But how does one know that the gas is unprocessed? The clue is in the fact that stars not only process hydrogen into helium, but they also, in the late stages of their evolution, synthesize heavier elements (metals) in their interiors. Therefore the abundance of heavier elements, like silicon, is an indicator of how much nuclear processing the ionized gas has undergone. It is observed that the He abundance is correlated with the metal abundance; so the goal is to find HII regions with as low a metal abundance as possible, and then extrapolate this empirical correlation to zero metal abundance [25, 26]. The answer turns out to be He/H ≈ 0.24, which is shown by the point with error bars in Fig. 4.2.

This value is embarrassingly low, given the observed deuterium abundance. It is obviously more consistent with an expansion rate provided by only two neutrino types rather than three, but we know that there are certainly three types. Possible reasons for this apparent anomaly are:

1) Bad astronomy: There are unresolved systematic errors in determination of the relative He abundance in HII regions indicated by the fact that the results of different groups differ by more than the quoted statistical errors [19]. The derivation of the helium to hydrogen ratio from the observed He^+/H^+ ratio requires some understanding of the structure of the HII regions. If there are relatively cool ionizing stars ($T < 35000$ K) spatially separated from the hotter stars, there may be relatively less He^+ associated with a given abundance of H^+. Lines of other elements need to be observed to estimate the excitation temperature; it is a complex problem.

2) New neutrino physics: There may be an asymmetry between neutrinos and anti-neutrinos (something like the baryon- antibaryon asymmetry which provides us with the observed Universe). This would manifest itself as a chemical potential in the Boltzmann equation giving different equilibrium ratios of the various neutrino species [27].

3) New gravitational physics: any change in the gravitational interaction which is effective at early epochs (braneworld effects?) could have a pronounced effect on nucleosynthesis. For example, a lower effective constant of gravity would yield a lower expansion rate and a lower He abundance. The standard minimal braneworld correction term, proportional to the square of the density [28], goes in the wrong direction.

It is unclear if the low helium abundance is a serious problem for the standard Big Bang. But it is clear that the agreement of the implied baryon abundance with the CMB determination is an impressive success, and strongly supports the assertion that the Hot Big Bang is the correct model for the pre-recombination Universe.

4.5 The Post-recombination Universe: Determination of H_o and t_o

Certainly the most basic of the cosmological parameters is the present expansion rate, H_o , because this sets the scale of the Universe. Until a few years ago, there was a factor of two uncertainty in H_o ; with two separate groups claiming two distinct values, one near 50 km $s^{-1}Mpc^{-1}$ and the other nearer 100 km $s^{-1}Mpc^{-1}$, and the errors quoted by both groups were much smaller than this factor of two difference. This points out a problem which is common in observational cosmology (or indeed, astronomy in general). Often the indicated statistical errors give the impression of great precision, whereas the true uncertainty is dominated by poorly understood or unknown systematic effects. That was true in the Hubble constant controversy, and there is no less reason to think that this problem is absent in modern results. I will return to this point several times below.

The great leap forward in determination of H_o came with the Hubble Space Telescope (HST) program on the distance scale. Here a particular kind of variable stars, Cepheid variables, were observed in twenty nearby spiral galaxies. Cepheids exhibit periodic variations in luminosity by a factor of two on timescales of 2-40 days. There is a well-determined empirical correlation between the period of Cepheids and their mean luminosity– the longer the period the higher the luminosity. Of course, this period-luminosity relation must be calibrated by observing Cepheids in some object with a distance known by other techniques and this remains a source of systematic uncertainty. But putting this problem aside, the Hubble Space telescope measured the periods and the apparent magnitudes, without confusion from adjacent bright stars, of a number of Cephieds in each of these relatively nearby galaxies, which yielded a distance determination (4.12). These galaxies are generally too close (less than 15 Mpc) to sample the pure Hubble flow, (the Hubble flow on these scales is contaminated by random motion of the galaxies and systematic cosmic flows) but these determinations do permit a calibration of other secondary distance indicators which reach further out, such as su-

pernovae type Ia (SNIa) and the Tully-Fisher relation (the observed tight correlation between the rotation velocities of a spiral galaxies and their luminosities). After an enormous amount of work by a number of very competent astronomers [29], the answer turned out to be $h = 0.72 \pm .10$.

As I mentioned there is the known systematic uncertainty of calibrating the period-luminosity relation, but there are other possible systematic effects that are less well-understood: How can we be certain that the period-luminosity relation for Cepheids is the same in all galaxies? For example, is this relation affected by the concentration of elements heavier than helium (the metallicity)? In view of such potential problems, other more direct physical methods, which by-pass the traditional "distance ladder" are of interest. Chief among these is the Sunyaev-Zeldovich (S-Z) effect which is relevant to clusters of galaxies [30]. The baryonic mass of clusters of galaxies is primarily in the form of hot gas, which typically exceeds the mass in the visible galaxies by more than a factor of two. This gas has a temperature between 10^7 and 10^8 K (*i.e.,* the sound speed is comparable to the one-dimensional velocity dispersion of the galaxies) and is detected by satellite X-ray telescopes with detectors in the range of several KeV. The S-Z effect is a small change in the intensity of the CMB in the direction of such clusters due to Compton scattering of CMB photons by thermal electrons (classical electron scattering would, of course, produce no intensity change). Basically, CMB photons are moved from the Rayleigh-Jeans part of the black body spectrum to the Wien part, so the effect is observable as a spectral distortion of the black body spectrum in the range of 100 to 300 GHz. It is a small effect (on the order of 0.4 milli Kelvin) but still 5 to 10 times larger than the intrinsic anisotropies in the CMB.

By measuring the amplitude of the S-Z effect one determines an optical depth

$$\tau = \sigma n_e l \ , \tag{4.22}$$

where σ is the frequency dependent cross section, l is the path length, and n_e is the electron density. Because these same clusters emit X-rays via thermal bremsstrahlung, we may also determine, from the observed X-ray intensity, an emission measure:

$$E = {n_e}^2 l \ , \tag{4.23}$$

Here we have two equations for two unknowns, n_e and l. This is simplifying the actual calculation because n_e is a function of radius in the cluster. Knowing l and the angular diameter of the cluster θ we can then calculate the angular size distance to the cluster via (4.11). Hence, the Hubble parameter is given by $H_o = v/D_A$ where v is the observed recession velocity of the cluster. All of this assumes that the clusters have a spherical shape on average, so the method needs to be applied to a number of clusters. Even so biases are possible if clusters have more typically a prolate shape or an oblate shape, or if the X-ray emitting gas is clumpy. Overall, for a number of clusters [31]

the answer turns out to be $h = 0.6$, somewhat smaller than the HST distance ladder method, but the systematic uncertainties remain large.

A second direct method relies on time delays in gravitational lenses [32]. Occasionally, a distant quasar (the source) is lensed by an intervening galaxy (the lens) into multiple images; that is to say, we observe two or more images of the same background object separated typically by one or two seconds of arc. This means that there are two or more distinct null geodesics connecting us to the quasar with two or more different light travel times. Now a number of these quasars are intrinsically variable over time scales of days or months (not periodic but irregular variables). Therefore, in two distinct images we should observe the flux variations track each other with a time delay. This measured delay is proportional to the ratio $D_l D_s / D_{ls}$ where these are the angular size distances to the lens, the source, and the lens to the source. Since this ratio is proportional to ${H_o}^{-1}$, the measured time delay, when combined with a mass model for the lens (the main source of uncertainty in the method), provides a determination of the Hubble parameter. This method, applied to several lenses [33, 34], again tends to yield a value of h that is somewhat smaller than the HST value, i.e., ≈ 0.6. In a recent summary [35] it is claimed that, for four cases where the lens is an isolated galaxy, the result is $h = 0.48 \pm .03$, if the overall mass distribution in each case can be represented by a singular isothermal sphere. On the other hand, in a well-observed lens where the mass distribution is constrained by observations of stellar velocity dispersion [36], the implied value of h is $0.75^{+.07}_{-.06}$. Such supplementary observations are important because the essential uncertainty with this technique is in the adopted mass model of the lens.

It is probably safe to say that $h \approx 0.7$, with an uncertainty of 0.10 and perhaps a slight bias toward lower values, but the story is not over as S-Z and gravitational lens determinations continue to improve. This is of considerable interest because the best fit to the CMB anisotropies observed by WMAP implies that $h = 0.72 \pm .05$ in perfect agreement with the HST result. With the S-Z effect and lenses, there remains the possibility of a contradiction.

With $h = .70$, we find a Hubble time of $t_H = 14$ Gyr. Now in FRW cosmology, the age of the Universe is $t_o = f t_H$ where f is a number depending upon the cosmological model. For an Einstein-de Sitter Universe (*i.e.,* $\Omega_k = 0$, $\Omega_Q = 0$, $\Omega_m = 1$) $f = 2/3$ which means that $t_o = 9.1$ Gyr. For an empty negatively curved Universe, $f = 1$ which means that the age is the Hubble time. Generally, models with a dominant vacuum energy density ($\Omega_Q \approx 1$, $w \approx -1$) are older ($f \geq 1$) and for the concordance model, $f = 0.94$. Therefore, independent determinations of the age of the Universe are an important consistency test of the cosmology.

It is reasonable to expect that the Universe should be older than the oldest stars it contains, so if we can measure the ages of the oldest stars, we have, at least, a lower limit on the age of the Universe. Globular star clusters are old stellar systems in the halo of our own galaxy; these systems are distributed in a roughly spherical region around the galactic disk and have low abundances

of heavy elements suggesting they were formed before most of the stars in the disk. If one can measure the luminosity of the most luminous, L_u, unevolved stars in a globular cluster (that is, stars still burning hydrogen in their cores), then one may estimate the age. That is because this luminosity is correlated with age: a higher L_u means a lower age. Up to about five years ago, this method yielded globular cluster ages of $t_{gc} \approx 14 \pm 2$ Gyr, which, combined with the Hubble parameter discussed above, would be in direct contradiction with the Einstein-de Sitter $\Omega_m = 1$ Universe. But about five years ago the Hipparchus satellite began to return accurate parallaxes for hundreds of nearby stars which led to a recalibration of the entire distance scale. All distances outside the solar system increased by about 10% (in fact, the entire Universe suddenly grew by this same factor leading to a decrease in the HST value for the Hubble parameter from about 80 to 72). This meant that the globular clusters were further away, that L_u was 20% larger larger, and the globular clusters were correspondingly younger: $t_{gc} \approx 11.5 \pm 1.3$ Gyr. If we assume that the Universe is about 1 Gyr older than the globular clusters, then the age of the Universe becomes 12.5 ± 2 Gyr [37] which is almost consistent with the Einstein-de Sitter Universe. At least there is no longer any compelling time scale argument for a non zero vacuum energy density, $\Omega_Q > 0$.

A second method for determining the ages of stars is familiar to all physicists, and that is radioactive dating. This has been done recently by observations of a U^{238} line in a metal-poor galactic star (an old star). Although the iron abundance in this star is only 1/800 that of the sun, the abundances of a group of rare earth metals known as r-process elements are enhanced. The r-process is rapid neutron absorption onto iron nuclei (rapid compared to the timescale for subsequent β decay) which contributes to certain abundance peaks in the periodic table and which occurs in explosive events like supernovae. This means that this old star was formed from gas contaminated by an even older supernova event; *i.e.,* the uranium was deposited at a definite time in the past. Now U^{238} is unstable with a half life of 4.5 Gyr which makes it an ideal probe on cosmological times scales. All we have to do is compare the observed abundance of U^{238} to that of a stable r-process element (in this case osmium), with what is expected from the r-process. The answer for the age of this star (or more accurately, the SN which contaminated the gas out of which the star formed) is 12.5 $\pm$3 Gyr, which is completely consistent with the globular cluster ages [38].

If we take $0.6 < h < 0.7$, and $9.5\,\mathrm{Gyr} < t_o < 15.5\,\mathrm{Gyr}$ this implies that $0.59 < H_o t_o < 1.1$. This is consistent with a wide range of FRW cosmologies from Einstein-de Sitter to the concordance model. That is to say, independent measurements of H_o and t_o are not yet precise enough to stand as a confirmation or contradiction to the WMAP result.

4.6 Looking for Discordance: The Classical Tests

4.6.1 The Angular Size Test

The first of the classical cosmological tests we will consider is the angular size test. Here one measures the angular size of a standard meter stick (hopefully) as a function of redshift; different FRW cosmologies make different predictions, but basically, for all FRW models $\theta(z)$ first decreases as $1/z$ (as would be expected in a Euclidean universe) and then increases with z. This is because the angular size distance is given by $D_A = r/(1+z)$ but the radial comoving coordinate approaches a finite value as $z \to \infty$. The angular size distance reaches a maximum at a redshift between 1 and 2 and then decreases again.

When giant radio galaxies at large redshift were discovered in the 1960's there was considerable optimism that these could be used as an angular size cosmological probe. Radio galaxies typically have a double-lobe structure with the radio emitting lobes straddling the visible galaxy; these lobes can extend hundreds of kpc beyond the visible object. Such a linear structure may be oriented at any angle to the observer's line-of-sight, so one needs to measure the angular sizes of a number of radio galaxies in a given redshift bin and only consider the largest ones, *i.e.,* those likely to be nearly perpendicular to the line-of-sight.

The result of all this work was disappointing. It appeared that the angular size of radio sources kept decreasing with redshift just as one would expect for a pure Euclidean universe [39]. The obvious problem, that plagues all classical tests, is that of evolution. Very likely, these radio galaxies are not standard meter sticks at all, but that they were actually smaller at earlier epochs than now. This would be expected, because such objects are thought to result from jets of relativistic particles ejected from the nucleus of the parent galaxy in opposite directions. The jets progress through the surrounding intergalactic medium at a rate determined by the density of that medium, which, of course, was higher at larger redshift.

But there is another class of radio sources that would be less susceptible to such environmental effects: the compact radio sources. These are objects, on a scale of milli-arc-seconds, typically associated with distant quasars, that are observed with radio interferometers having global baselines. The morphology is that of a linear jet with lengths typically less than 30 or 40 pc, so these would presumably be emission from the jets of relativistic particles deep in the galactic nucleus near the central engine producing them. The intergalactic medium, and its cosmological evolution, would be expected to have no effect here [40].

The result of plotting the median angular size of about 150 of these sources as a function of redshift is shown on a log-log plot in Fig. 4.3 [41]. Also shown are the predicted relations for three flat cosmologies ($\Omega_k = 0$) with $\Omega_m = 0.9$, 0.3, 0.1, the remainder being in a cosmological constant (the middle curve

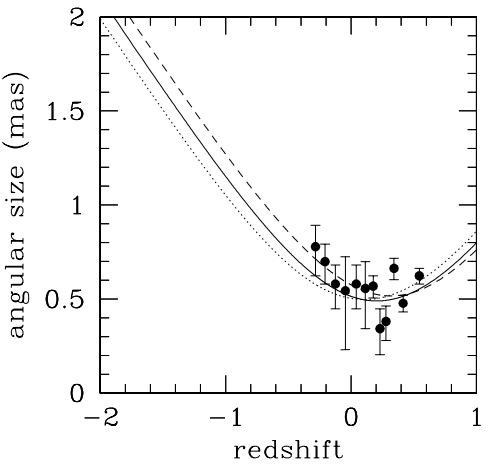

Fig. 4.3. The median angular size vs. redshift (log-log plots) for 145 compact radio sources in 12 redshift bins. The curves are the three flat cosmological models: dashed, $\Omega_\Lambda = 0.9$; solid, $\Omega_\Lambda = 0.7$ (concordance), dotted, $\Omega_\Lambda = 0.1$. The physical size of the sources (20-40 pc) has been chosen for the best fit.

is the concordance model). In each case the linear size of the compact radio sources was chosen to achieve the best fit to the data.

It is evident that the general property of FRW models (that the angular size of a standard meter stick should begin to increase again beyond a redshift of about 1.5) is present in this data. However, no statistical test or maximum likelihood analysis is necessary to see that all three models fit the data equally well. This is basically an imprecise cosmological test and cannot be improved, particularly considering that these objects may also evolve in some unknown way with cosmic time. Looking at the figure, one may notice that measurement of angular sizes for just a few objects at lower redshift might help distinguish between models. However, there are very few such objects at lower redshift, and these have a much lower intrinsic radio power than those near redshift one. It is dangerous to include these objects on such a plot because they are probably of a very different class.

4.6.2 The Modern Angular Size Test: CMB-ology

Although it is not my purpose here to discuss the CMB anisotropies (see A. Challinor's contribution), it is necessary to say a few words on the preferred angular scale of the longest wavelength acoustic oscillations, the "first peak", because this is now the primary evidence for a flat Universe ($\Omega_k = 0$). In Fig. 4.4 we see again the now very familiar plot of the angular power spectrum of anisotropies as observed by WMAP [43] (in my opinion, of all the WMAP papers, this reference provides the clearest discussion of the physics behind the peak amplitudes and positions). The solid line is the concordance model,

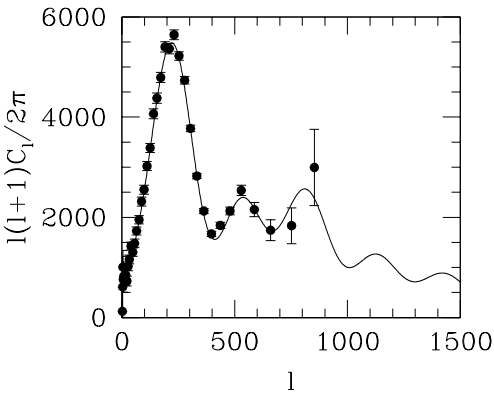

Fig. 4.4. The angular power spectrum of CMB anisotropies observed by WMAP [43]. The solid line is not a fit but the is the concordance model proposed earlier [2].

not a fit, but just the predicted angular power spectrum (via CMBFAST [42]) from the $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ model Universe with an optical depth of $\tau \approx 0.17$ to the surface of last scattering. I must admit that the agreement is impressive.

I remind you that the harmonic index on the horizontal axis is related to angular scale as

$$l \approx \pi/\theta \ , \tag{4.24}$$

so the first peak, at $l \approx 220$, would correspond to an angular scale of about one degree. I also remind you that the first peak corresponds to those density inhomogeneities which entered the horizon sometime before decoupling (at $z = 1000$); enough before so that they have had time to collapse to maximum compression (or expand to maximum rarefaction) just at the moment of hydrogen recombination. Therefore, the linear scale of these inhomogeneities is very nearly given by the sound horizon at decoupling, that is

$$l_h \approx ct_{dec}/\sqrt{3} \ , \tag{4.25}$$

where t_{dec} is the age of the Universe at decoupling.

So one might say, the test is simple: we have a known linear scale l_h which corresponds to an observed angular scale ($\theta \approx 0.014$ rad) so we can determine the geometry of the Universe. It is not quite so simple because the linear scale, l_h depends, via t_{dec} on the matter content of the Universe (Ω_m); basically, the larger Ω_m, the sooner matter dominates the expansion, and the earlier decoupling with a correspondingly smaller l_h. This comoving linear scale is shown in Fig. 4.5 as a function of Ω_m (Ω_Λ hardly matters here, because the vacuum energy density which dominates today has no effect at the epoch of decoupling). Another complication is that the angular size distance to the surface of last scattering not only depends upon the geometry, but also upon the expansion history. This is evident in Fig. 4.6 which shows the comoving

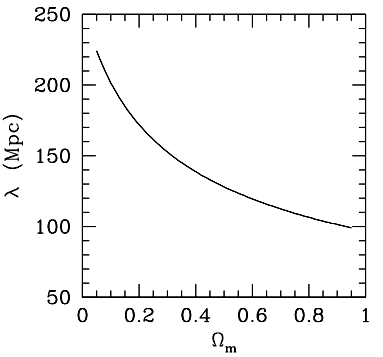

Fig. 4.5. The comoving linear scale of the perturbation corresponding to the first peak as a function of Ω_m

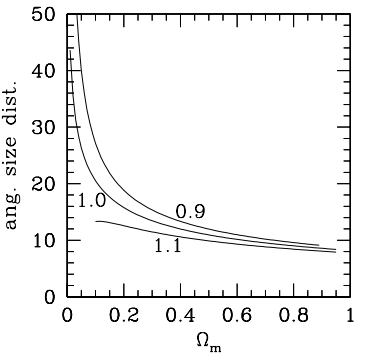

Fig. 4.6. The angular size distance (Gpc) to the last scattering surface ($z = 1000$) as a function of Ω_m for various values of Ω_{tot}.

angular size distance (in Gpc) to the surface of last scattering as a function of Ω_m for three values of $\Omega_{tot} = \Omega_m + \Omega_\Lambda$ (i.e., $\Omega_k = 1 - \Omega_{tot}$). Note that the comoving angular size distance, $D_A(1+z)$, is the same as the radial comoving coordinate r.

We can combine Figs. 4.5 and 4.6 to plot the expected angular size (or harmonic index) of the first peak as a function of Ω_m and Ω_{tot}, and this is shown in Fig. 4.7 with the dashed line giving the observed l of the first peak. We see that a model with $\Omega_{tot} \geq 1.1$ (a closed universe) is clearly ruled out, but it would be possible to have an open model with $\Omega_{tot} = 0.9$ and $\Omega_m = 0.8$ from the position of the first peak alone; the predicted peak amplitude, however, would be about 40% too low. The bottom line of all of this is that the *position* of the first peak does not uniquely define the geometry of the Universe because of a degeneracy with Ω_m (I haven't mentioned the degeneracy with h taken here to be 0.72). To determine whether or not we live in a flat Universe we need an independent handle on Ω_m and that is

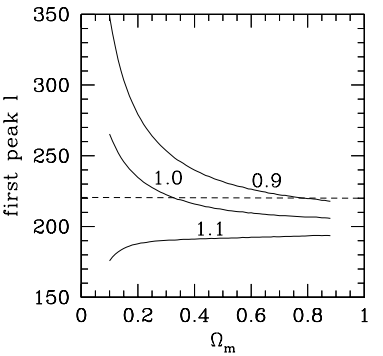

Fig. 4.7. The harmonic index expected for the first peak as a function of Ω_m for various values of Ω_{tot}.

provided, in WMAP data, by the amplitudes of the first two peaks (the more non-baryonic matter, the deeper the forming potential wells, and the lower the amplitudes). From this it is found that $\Omega_m \approx 0.3$, and from Fig. 4.7 we see that the model Universe should be near flat ($\Omega_{tot} \approx 1.0$). Of course if the Universe is near flat with $\Omega_m = 0.3$ then the rest must be in dark energy; this is the indirect evidence from the CMB anisotropies for dark energy.

I just add here that the observed peak amplitudes (given the optical depth to $z = 1000$ determined from WMAP polarization results [44]), is taken now as definitive evidence for CDM. However, alternative physics which affects the amplitude and positions of peaks (e.g. [3]) could weaken this conclusion, as well as affect the derived cosmological parameters. Even taking the peak amplitudes as *prima facie* evidence for the existence of cold dark matter, it is only evidence for CDM at the epoch of recombination ($z = 1000$) and not in the present Universe. To address the cosmic coincidence problem, models have been suggested in which dark matter transmutes into dark energy (e.g. [45]).

Now I turn to the direct evidence for dark energy.

4.6.3 The Flux-Redshift Test: Supernovae Ia

Type I supernovae are thought to be nuclear explosions of carbon/oxygen white dwarfs in binary systems. The white dwarf (a stellar remnant supported by the degenerate pressure of electrons) accretes matter from an evolving companion and its mass increases toward the Chandrasekhar limit of about 1.4 $M_\odot$ (this is the mass above which the degenerate electrons become relativistic and the white dwarf unstable). Near this limit there is a nuclear detonation in the core in which carbon (or oxygen) is converted to iron. A nuclear flame propagates to the exterior and blows the white dwarf apart (there are alternative models but this is the favored scenario [46]).

These events are seen in both young and old stellar populations; for example, they are observed in the spiral arms of spiral galaxies where there is

active star formation at present, as well as in elliptical galaxies where vigorous star formation apparently ceased many Gyr ago. Locally, there appears to be no difference in the properties of SNIa arising in these two different populations, which is important because at large redshift the stellar population is certainly younger.

The peak luminosity of SNIa is about 10^{10} $L_\odot$ which is comparable to that of a galaxy. The characteristic decay time is about one month which, in the more distant objects, is seen to be stretched by 1+z as expected. The light curve has a characteristic form and the spectra contain no hydrogen lines, so given reasonable photometric and spectroscopic observations, they are easy to identify as SNIa as opposed to type II supernovae; these are thought to be explosions of young massive stars and have a much larger dispersion in peak luminosity [47].

The value of SNIa as cosmological probes arises from the high peak luminosity as well as the observational evidence (locally) that this peak luminosity is the sought-after standard candle. In fact, the absolute magnitude, at peak, varies by about 0.5 magnitudes which corresponds to a 50%-60% variation in luminosity; this, on the face of it, would make them fairly useless as standard candles. However, the peak luminosity appears to be well-correlated with decay time: the larger L_{peak}, the slower the decay. There are various ways of quantifying this effect [47], such as

$$M_B \approx 0.8(\Delta m_{15} - 1.1) - 19.5 \ , \tag{4.26}$$

where M_B is the peak absolute magnitude and $\Delta\ m_{15}$ is the observed change in apparent magnitude 15 days after the peak [48]. This is an empirical relationship, and there is no consensus about the theoretical explanation, but, when this correction is applied it appears that $\Delta L_{peak} < 20\%$. If true, this means that SNIa are candles that are standard enough to distinguish between cosmological models at $z \approx 0.5$.

In a given galaxy, supernovae are rare events (on a human time scale, that is), with one or two such explosions per century. But if thousands of galaxies can be surveyed on a regular and frequent basis, then it is possible to observe several events per year over a range of redshift. About 10 years ago two groups began such ambitious programs [49, 50]; the results have been fantastically fruitful and have led to a major paradigm shift.

The most recent results are summarized in [51]: at present, about 230 SNIa have been observed out to $z = 1.2$. The bottom line is that SNIa are 10% to 20% fainter at $z \approx 0.5$ than would be expected in an empty ($\Omega_{tot} = 0$) non-accelerating Universe. But, significantly, at $z \geq 1$ the supernovae appear to become brighter again relative to the non-accelerating case; this should happen in the concordance model at about this redshift because it is here that the cosmological constant term in the Friedmann equation (4.10) first begins to dominate over the matter term. This result is shown in Fig. 4.8 which is a plot of the median Δm, the observed deviation from the non-accelerating case, in various redshift bins as a function of redshift (*i.e.,* the

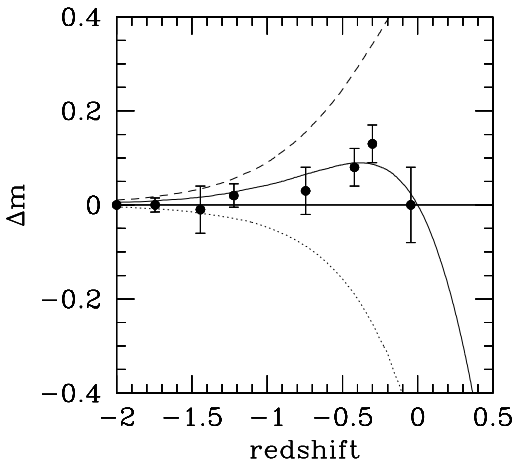

Fig. 4.8. The Hubble diagram for SNIa normalized to an empty non-accelerating Universe. The points are binned median values for 230 supernovae [51] The curves show the predictions for three flat ($\Omega_{tot} = 0$) cosmological models: The dashed line is the model dominated by a cosmological constant ($\Omega_\Lambda = 0.9$), the solid curve is the concordance model ($\Omega_\Lambda = 0.7$), and the dotted curve is the matter dominated model ($\Omega_\Lambda = 0.1$).

horizontal line at $\Delta m = 0$ corresponds to the empty universe). The solid curves show the prediction for various flat ($\Omega_{tot} = 1$) models with the value of the cosmological term indicated. It is evident that models dominated by a cosmological term or by matter are inconsistent with the observations at extremely high levels of significance, while the concordance model agrees quite well with the observations.

It is also evident from the figure that the significance of the effect is not large, perhaps 3 or 4 σ (quite a low level of significance on which to base a paradigm shift). When all the observed supernovae are included on this plot, it is quite a messy looking scatter with a minimum χ^2 per degree of freedom (for flat models) which is greater than one. Moreover the positive result depends entirely upon the empirical peak luminosity-decay rate relationship and, of course, upon the assumption that this relation does not evolve. So, before we become too enthusiastic we must think about possible systematic effects and how these might affect the conclusions. These effects include:

1) Dust: It might be that supernovae in distant galaxies are more (or less) dimmed by dust than local supernovae. But normal dust, with particle sizes comparable to the wavelength of light, not only dims but also reddens (for the same reason, Rayleigh scattering, that sunsets are red). This is quantified by the so-called color excess. Remember I said that astronomers measure the color of an object by its B-V color index (the logarithm of a flux ratio). The color excess is defined as

$$E(B-V) = (B-V)_{obs} - (B-V)_{int} \ , \tag{4.27}$$

where *obs* means the observed color index and *int* means the intrinsic color index (the color the object would have with no reddening). In our own galaxy it is empirically the case that the magnitudes of absorption is proportional to this color excess, *i.e.,*

$$A_V = R_V E(B-V) \ , \tag{4.28}$$

where R_V is roughly constant and depends upon average grain properties. So assuming that the dust in distant galaxies is similar to the dust in our own, it should be possible to estimate and correct for the dust obscuration. Significantly [49], it appears that there is no difference between E(B-V) for local and distant supernovae. This implies that the distant events are not more or less obscured than the local ones.

2) Grey dust: It is conceivable (but unlikely) that intergalactic space contains dust particles which are significantly larger than the wavelength of light. Such particles would dim but not redden the distant supernovae and so would be undetectable by the method described above [52]. It is here that the very high redshift supernovae ($z > 1$) play an important role. If this is the cause of the apparent dimming we might expect that the supernovae would not become brighter again at higher redshift.

3) Evolution: It is possible that the properties of these events may have evolved with cosmic time. As I mentioned above, the SN exploding at high redshift come from a systematically younger stellar population than the objects observed locally. Moreover, the abundance of metals was smaller in the earlier Universe than now; this evolving composition, by changing the opacity in the outer layers or the composition of the fuel itself could lead to a systematic evolution in peak luminosity. Here it is important to look for observational differences between local and distant supernovae, and there seem to be no significant differences in most respects, the spectrum or the light curve. There is, however, a suggestion that distant supernovae are intrinsically bluer than nearby objects [47]. If this effect is verified, then it could not only point to a systematic difference in the objects themselves, but could also have lead to an underestimate of the degree of reddening in the distant SN. It is difficult, in general, to eliminate the possibility that the events themselves were different in the past and that this could mimic the effect of a cosmological constant [53]; a deeper theoretical understanding of the SNIa process is required in order to realistically access this possibility.

4) Sample evolution: The sample of SN selected at large redshift may differ from the nearby sample that is used, for example, to calibrate the peak luminosity-decline rate correlation. There does appear to be an absence, at large redshift, of SN with very slowly declining light curves– which is to say, very luminous SN that are seen locally. Perhaps a class of more luminous objects is missing in the more distant Universe due to the fact that these SN emerge from a systematically younger stellar population. One would hope

that the luminosity-decline rate correlation would correct for this effect, assuming, of course, that this relation itself does not evolve.

5) Selection biases: There is a dispersion in the luminosity-decline rate relationship, and in a flux-limited sample, one tends to select the higher luminosity objects. Astronomers call this sort of bias the "Malmquist effect" and it is always present in such observational data. Naively, one would expect such a bias to lead to an underestimate of the true luminosity, and, therefore an underestimate luminosity distance; the bias actually diminishes the apparent acceleration. But there is another effect which is more difficult to access: The most distant supernovae are being observed in the UV of their own rest frame. SNIa are highly non-uniform in the UV, and K-corrections are uncertain. This could introduce systematic errors at the level of a few hundredths of a magnitudes [51].

We see that there are a number of systematic effects that could bias these results. A maximum likelihood analysis over the entire sample [51], confirms earlier results that the confidence contours in Ω_m-Ω_Λ space are stretched along a line $\Omega_\Lambda = 1.4\Omega_m + 0.35$ and that the actual best fit is provided by a model with $\Omega_m \approx 0.7$ and $\Omega_\Lambda \approx 1.3$, not the concordance model. Of course, if we add the condition that $\Omega_{tot} = 1$ (a flat Universe) then the preferred model becomes the concordance model. In [51] it is suggested that this apparent deviation is due to the appearance of one or more of the systematic effects discussed above near $z = 1$ at the level of 0.04 magnitudes.

The result that SNIa are systematically dimmer near $z = 0.5$ than expected in a non-accelerating Universe is robust. At the very least it can be claimed with reasonable certainty that the Universe is not decelerating at present. However, given the probable presence of systematic uncertainties at the level of a few hundredths of a magnitude, it is difficult to constrain the equation of state (w) of the dark energy or its evolution (dw/dt) until these effects are better understood. I will just mention that lines of constant age, $t_o H_o$, are almost parallel to the best fit line in the Ω_m-Ω_Λ plane mentioned above. This then gives a fairly tight constraint on the age in Hubble times [51]; i.e. $t_o H_o = 0.96 \pm 0.4$, which is consistent with the WMAP result. In a near flat Universe this rules out the dominance of matter and requires a dark energy term.

4.6.4 Number Counts of Faint Galaxies

The final classical test I will discuss is that of number counts of distant objects, what radio astronomers call the log(N)-log(S) test. Basically one counts the number of galaxies N brighter than a certain flux limit S. If we lived in a static Euclidean universe, then the number of galaxies out to distance R would be $N \propto R^3$ but the flux is related to R as $S \propto R^{-2}$. This implies that $N \propto S^{-3/2}$ or $log(N) = -3/2\, log(S) + const = 0.6m + const.$ where m is the magnitude corresponding to the flux S.

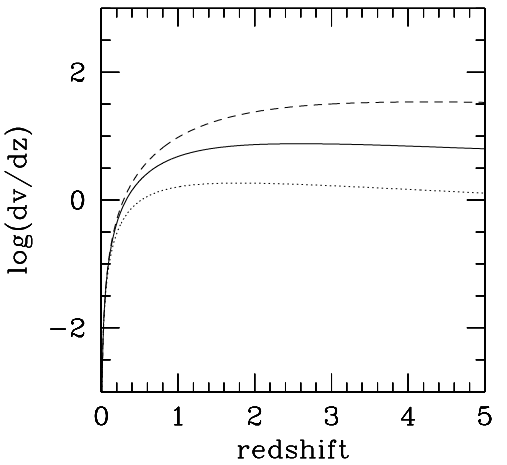

Fig. 4.9. The log of the incremental volume per incremental redshift (in units of the Hubble volume) as a function of redshift for the three flat cosmological models.

But we do not live in a static Euclidean universe; we live in an evolving universe with a non-Euclidean geometry where the differential number counts probe dV(z), the comoving volume as a function of redshift. In Fig. 4.9 we see log(dV/dz) as a function of redshift for three different($\Omega_{tot} = 1$) cosmological models: the matter dominated Universe, the cosmological constant dominated Universe, and the concordance model. For small z, dV/dz increases as z^2 for all models as would be expected in a Euclidean Universe, but by redshift one, the models are obviously diverging, with the models dominated by a cosmological constant having a larger comoving incremental volume. Therefore if we can observe faint galaxies extending out to a redshift of one or two, we might expect number counts to provide a cosmological probe.

There is a long history of counting objects as a function of flux or redshift. Although cosmological conclusions have been drawn (see, e.g. [54]), the overall consensus is that this is not a very good test because the galaxy population evolves strongly with redshift. Galaxies evolve because stars evolve. In the past, the stellar populations were younger and contained relatively more massive, luminous stars. Therefore we expect galaxies to be more luminous at higher redshift. It is also possible that the density of galaxies evolves because of merging, as would be consistent with the preferred model of hierarchical structure formation in the Universe.

The distribution of galaxies by redshift can be used, to some extent, to break this degeneracy between evolution and cosmology. If we can measure the redshifts of galaxies with infrared magnitudes between 23 and 26, for example, that distribution will be skewed toward higher redshift if there is more luminosity evolution.

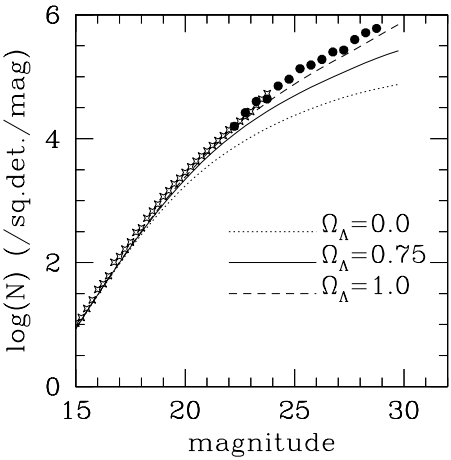

Fig. 4.10. The solid points are the faint galaxy number counts from the Hubble Deep Fields (north and south [56, 57]) and the star shaped points are the number counts from ground based data. The curves are the no-evolution predictions from three flat cosmological models.

I have recently reconsidered the number counts of the faint galaxies in the Hubble Deep Fields, north and south [56, 57]. These are two separate small patches of empty sky observed with the Hubble Space Telescope down to a very low flux limit– about $m_I = 30$ (the I band is a far red filter centered around 8000 angstroms). The differential number counts are shown by the solid round points in Fig. 4.10 where ground based number counts at fainter magnitudes are also shown by the starred points.

For this same sample of galaxies, there are also estimates of the redshifts based upon the galaxy colors, the so called photometric redshifts [58]. In order to calculate the expected number counts and redshift distribution one must have some idea of the form of the luminosity function, the distribution of galaxies by redshift. Here, like everyone else, I have have assumed that this form is given by the Schechter function [59]:

$$N(L)dL = N_o(L/L_*)^{-\alpha}exp(-L/L_*)dL \ , \tag{4.29}$$

which is characterized by three parameters: α, a power law at low luminosities, L_* a break-point above which the number of galaxies rapidly decreases, and N_o a normalization. I take this form because the overall galaxy distribution by luminosity at low redshifts is well fit by such a law [60], so I am assuming that at least the form of the luminosity function does not evolve with redshift.

But when I consider faint galaxies at high redshift in a particular band I have to be careful to apply the K-correction mentioned above; that is, I must correct the observed flux in that band to the rest frame. Making this

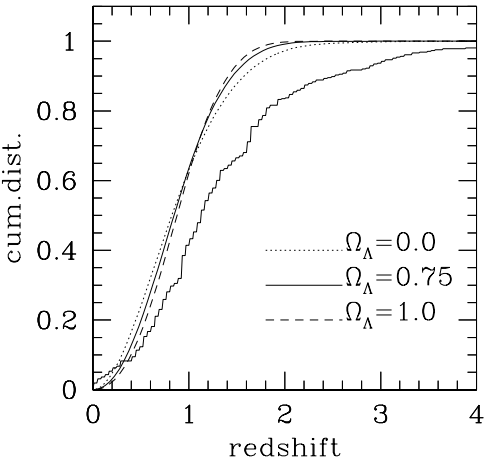

Fig. 4.11. The cumulative redshift distribution for galaxies between apparent I-band magnitudes of 23 and 26 (photometric redshifts from [58]). The curves are the predicted no-evolution distributions for the three cosmological models.

correction [61], but assuming no luminosity or density evolution, I find the differential number counts appropriate to our three flat cosmological models shown by the indicated curves in Fig. 4.10. We see that the predicted number counts all fall short of the observed counts, but that the cosmological constant dominated model comes closest to matching the observations. However, the distribution by redshift of HDF galaxies between I-band magnitudes of 22 and 26 is shown in Fig. 4.11 (this is obviously the cumulative distribution). Here we see that all three models seriously fail to match the observed distribution, in the sense that the predicted mean redshift is much too small.

This problem could obviously be solved by evolution. If galaxies are brighter in the past, as expected, then we would expect to shift this distribution toward higher redshifts. One can conceive of very complicated evolution schemes, involving initial bursts of star formation with or without continuing star formation, but it would seem desirable to keep the model as simple as possible; let's take a "minimalist" model for galaxy evolution. A simple one parameter scheme with the luminosity brightening proportional to the look-back time squared, *i.e.,* every galaxy brightens as

$$\Delta M_I = q\,(H_o t_{lb})^2\ , \tag{4.30}$$

where q is the free parameter, can give a reasonable match to evolution models for galaxies [61]. (we also assume that all galaxies are the same, they are not divided into separate morphological classes). I choose the value of q such that the predicted redshift distribution most closely matches the observed distribution for all three models, and the results are shown in Fig. 4.12.

The required values of q (in magnitudes per ${t_H}^2$) for the three cosmological models are: $q = 2.0$ ($\Omega_\Lambda = 1.0$), $q = 3.0$ ($\Omega_\Lambda = 0.7$), and $q = 11.0$ ($\Omega_\Lambda = 0.0$).

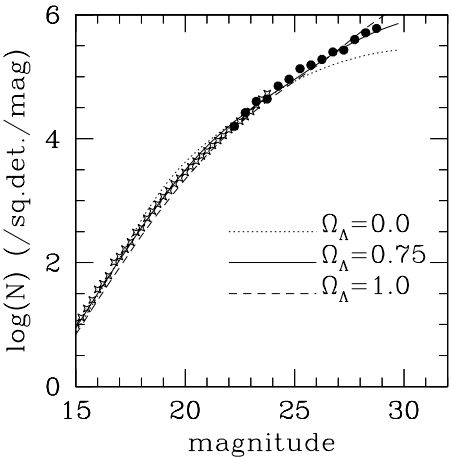

Fig. 4.12. As in Fig. 4.10 above the observed galaxy number counts and the predictions for the cosmological models with luminosity evolution sufficient to explain the number counts.

Obviously, the matter-dominated model requires the most evolution, and with this simple evolution scheme, cannot be made to perfectly match the observed distribution by redshift (this in itself is not definitive because one could always devise more complicated schemes which would work). For the concordance model, the required evolution would be about two magnitudes out to $z = 3$.

For these same evolutionary models, that is, with evolution sufficient to match the number counts, the predicted redshift distributions are shown in Fig. 4.13. Here we see that the model dominated by a cosmological constant predicts too many low redshift galaxies, the matter dominated model predicts too few, and the model that works perfectly is very close to the concordance model! Preforming this operation for a number of flat models with variable Ω_Λ, I find that $0.59 < \Omega_\Lambda < 0.71$ to 90% confidence.

Now there are too many assumptions and simplifications to make this definitive. The only point I want to make is that faint galaxy number counts and redshift distributions are completely consistent with the concordance model when one considers the simplest minimalist model for pure luminosity evolution. One may certainly conclude that number counts provide no contradiction to the generally accepted cosmological model of the Universe (to my disappointment).

4.7 Conclusions

In this contribution I have been looking for discord, but have not found it. The classical tests return results for cosmological parameters that are

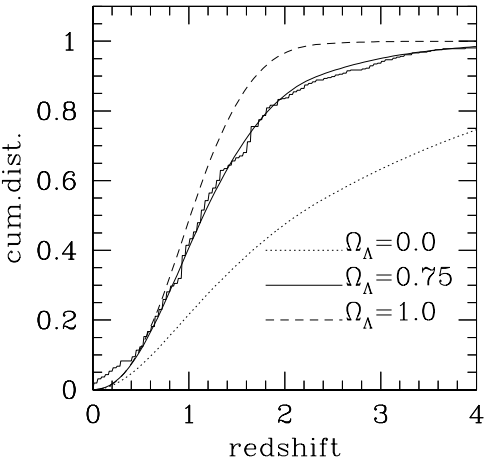

Fig. 4.13. The cumulative redshift distribution for galaxies between apparent i-band magnitudes of 22 and 26 (photometric redshifts from [58]). The curves are the predicted distributions for the three cosmological models with evolution sufficient to explain the number counts.

consistent with but considerably less precise than those implied by the CMB anisotropies, given the usual assumptions. It is fair to say that the numbers characterizing the concordance model, $\Omega_m \approx 0.3$, $\Omega_\Lambda \approx 0.7$ are robust *in the context of the framework of FRW cosmology.* It is, in fact, the peculiar composition of the Universe embodied by these numbers which calls that framework into question.

Rather small changes in the assumptions underlying pure FRW cosmology (with only an evolving vacuum energy density in addition to more familiar fluids) can make a difference. For example, allowing $w = -0.6$ brings the number counts and z-distribution of faint galaxies into agreement with a Universe strongly dominated by dark energy ($\Omega_Q = 0.9$). The same also true of the high-z supernovae observations [51]. Allowing a small component of correlated iso-curvature initial perturbations, as expected in braneworld cosmologies, can affect the amplitudes and positions of the peaks in the angular power spectrum of the CMB anisotropies [3], and therefore the derived cosmological parameters.

But even more drastic changes have been suggested. Certain braneworld scenarios, for example, in which 4D gravity is induced on the brane [62] imply that gravity is modified at large scale where gravitons begin to leak into the bulk [63]. It is possible that the observed acceleration is due to such modifications and not to dark energy. More ad hoc modifications of General Relativity [5] have also been proposed because of a general unease with dark energy, proposals whereby gravity is modified in the limit of small curvature scalar. My own opinion is that we should also feel uneasy with the mysterious

non-baryonic cold dark matter, because the only evidence for its existence, at present, is its gravitational influence; when the theory of gravity is modified to eliminate dark energy, it might also be found that the need for dark matter vanishes.

In general, more attention is being given to so-called infrared modifications of gravity (e.g. [64]), and this is a positive development. High energy modifications, that affect the evolution of the early Universe, are, as we have seen, strongly constrained by considerations of primordial nucleosynthesis (now, in combination with the CMB results). It is more likely that modifications play a role in the late, post-recombination evolution of the Universe, where the peculiarities of the concordance model suggest that they are needed. The fact that the same rather un-natural values for the comparable densities of dark energy and matter keep emerging in different observational contexts may be calling attention to erroneous underlying assumptions rather than to the actual existence of these "ethers".

Convergence toward a parameterized cosmology is not, without deeper understanding, sufficient reason for triumphalism. Rather, it should be a motivation to look more carefully at the possible systematic effects in the observations and to question more critically the underlying assumptions of the models.

Acknowledgements

I thank Rien van de Weygaert, Ole Möller, Moti Milgrom, Art Wolfe, Jacob Bekenstein, and Scott Trager for useful comments on the manuscript. I also thank Gary Steigman, Wendy Freedman, and Luis Ho for permission to use Figs. 4.1 and 4.2. I am very grateful to the organizers of the Second Aegean Summer School on the Early Universe, and especially, Lefteris Papantonopoulos, for all their work and for inviting me to the very pleasant island of Syros.

References

1. D.N. Spergel, et al., Astrophys. J. Suppl., **148**, 175 (2003), astro-ph/0302209.
2. J.P. Ostriker, P.J. Steinhardt, Nature **377**, 600 (1995).
3. R. Maartens, gr-qc/0312059.
4. Deffayet C, Phys. Lett. **B 502**, 199 (2001).
5. S.M. Carroll, V. Duvvuri, M. Trodden, M.S. Turner, astro-ph/0306438 (2003).
6. J.F. Navarro, C.S. Frenk, S.D.M. White, Astrophys. J. **462** 563 (1996).
7. W.J.G. de Blok, S.S. McGaugh, V.C. Rubin, Astron. J., **122**, 2396 (2001).
8. T. Treu, L.V.E. Koopmans, D.J. Sand, G.P. Smith, R.S. Ellis, To appear in the proceedings of IAU Symposium 220 "Dark matter in galaxies", S. Ryder, D.J. Pisano, M. Walker, and K. Freeman eds, astro-ph/0311052.
9. S.M. Carroll, Living Rev.Rel. **4**, 1 (2001), astro-ph/0004075.

10. P.J.E. Peebles, B. Ratra', Rev. Mod. Phys., **75**, 559 (2003).
11. T. Padmanabhan, hep-th/0212290.
12. J.A. Peacock, Phil. Trans. Roy. Soc. Lond., **A361** 2479 (2003), astro-ph/0309238.
13. H. Hoekstra, H. Yee, M. Gladders, Astrophys. J., **577**, 595 (2002).
14. M. Milgrom, Astrophys. J., **270**, 365 (1983).
15. W. Rindler, Essential Relativity, (Heidelberg, Springer-Verlag).
16. R.R. Caldwell, Phys. Lett. **B545**, 23 (2002).
17. J.C. Mather et al., Astrophys. J., **512**, 511 (1999).
18. P. Molaro, S.A. Levshakov, M. Dessauges-Zavadsky, S. D'Odorico, Astron.Astrophys., **381**, 64 (2002), astro-ph/0111589.
19. G. Steigman, Carnegie Observatories Astrophysics Series Vol. 2, "Measuring and Modeling the Universe", W. Freedman ed. (Cambridge, Cambridge Univ. Press) (2003), astro-ph/ 0307244.
20. S. Burles, D. Tytler, Astrophys. J., **499**, 699 (1998).
21. J.M. O'Meara et al., Astrophys. J., **552**, 718 (2001).
22. M. Pettini, D.V. Bowen, Astrophys. J., **560**, 41 (2001).
23. D. Kirkman, et al., Astrophys. J. Suppl, in press (2003), astro-ph/0302006.
24. P.Jakobsen et al., Astrophys. J., **417**, 528 (1993).
25. K. Olive, G. Steigman, Astrophys. J. Supp., **97**, 49 (1995).
26. Y.I. Izotov, T.X. Thuan, Astrophys. J., **500**, 188 (1998).
27. V. Barger, J.P. Kneller, P. Langacker, D. Marfatia, G. Steigman, Phys.Lett. **B569** 123 (2003), hep-ph/0306061.
28. C. Barcelo, M. Visser, Phys.Lett. **B482**, 183 (2000), hep-th/0004056.
29. W.L. Freedman et al., Astrophys. J., **553**, 47 (2001).
30. R.A. Sunyaev, Ya.B. Zeldovich, Ann. Rev. Astron. Astrophys., **18**, 537 (1980).
31. J.E. Carlstrom, G.P. Holder, E.D. Reese, Ann.Rev.Astron.Astrophys., **40**, 643 (2002).
32. S. Refsdal, Mon. Not. RAS, **128**, 307 (1964).
33. L.L.R. Williams, P. Saha, Astron. J., **119**, 439 (2000).
34. L.V.E. Koopmans, C.D. Fassnacht, Astrophys. J., **527**, 513 (1999).
35. C.S. Kochanek, P.L. Schechter, to appear in Measuring and Modeling the Universe (Carnegie Observatories Astrophysics Series, Vol. 2), ed. W.L. Freedman (2003), astro-ph/0306040.
36. L.V.E. Koopmans et al., Astrophys. J., **599**, 70 (2003).
37. B. Chaboyer, P. Demarque, P.J. Kernan, L.M. Krauss, Astrophys. J., **494**, 96 (1998).
38. R. Cayrel et al., Nature, **409**, 691 (2001), astro-ph/0104357.
39. G.K. Miley, Mon. Not. RAS, **152**, 477 (1971).
40. K.I. Kellerman, Nature, **361**, 134 (1993).
41. L.I. Gurvits, K.I. Kellerman, S. Frey, Astron. Astrophys, **342**, 378 (1999).
42. U. Seljak, M. Zeldarriaga, Astrophys. J., **469**, 437 (1996).
43. L. Page et al., Astrophys. J. Suppl., **148**, 233 (2003).
44. A. Kogut et al., Astrophys. J. Suppl., **148**, 161 (2003).
45. B.A. Bassett, M. Kunz, D. Parkinson, C. Ungarelli, Phys. Rev., **D68**, 043504 (2003).
46. D. Branch, Ann. Rev. Astron. Astrophys., **36**, 17 (1998).
47. B. Leibundgut, Ann. Rev. Astron. Astrophys, **39**, 67 (2001).
48. M. Hamuy, M.M. Phillips, R.A. Schommer, N.B. Suntzeff, Astron. J., **112**, 2391 (1996).

49. Perlmutter S, et al., Astrophys. J., **517**, 565 (1999).
50. P.M. Garnevitch et al. , Astrophys. J., **493**, 53 (1998).
51. J.L. Tonry et al., Astrophys. J., **594**, 1 (2003), astro-ph/0305008.
52. A. Aguirre, Z. Haiman, Astrophys. J., **532**, 28 (2000).
53. P.S. Drell, T.J. Loredo, I. Wasserman, Astrophys. J., **530**, 593 (2000).
54. M. Fukugita, F. Takahara, K. Yamashita, Y. Yoshi, Astrophys. J., **361**, L1 (1990).
55. Y. Yoshii, F. Takahara, Astrophys. J., **346**, 28 (1989).
56. R.E. Williams et al., Astron. J., **112**, 1335 (1996).
57. S. Casertano et al., Astron. J., **120**, 247 (2000).
58. A. Fernández-Soto, K.M. Lanzetta, A. Yahil, Astrophys. J., **513**, 34 (1999).
59. P.L. Schechter, Astrophys. J., **203**, 297 (1976).
60. M.R. Blanton et al., Astron. J., **121**, 2358 (2001).
61. B.M. Poggianti, Astron. Astrophys. Suppl, **122**, 399 (1997).
62. G. Dvali, G. Gabadadaze, M. Porrati, Phys. Lett., **B485**, 208 (2000), hep-th/0005016.
63. C. Deffayet, G. Dvali, G. Gabadadze, Phys. Rev. **D65** 044023 (2002), astro-ph/0105068.
64. N. Arkani-Hamed, H-C. Cheng, M.A. Luty, S. Mukohyama, hep-th/0312099.

5 Dark Matter and Dark Energy

Varun Sahni

Inter-University Center for Astronomy and Astrophysics, Puné 411 007, India

Abstract. I briefly review our current understanding of dark matter and dark energy. The first part of this review focusses on issues pertaining to dark matter including observational evidence for its existence, current constraints and the 'abundance of substructure' and 'cuspy core' issues which arise in CDM. I also briefly describe MOND. The second part of this review focusses on dark energy. In this part I discuss the significance of the cosmological constant problem which leads to a predicted value of the cosmological constant which is almost 10^{123} times larger than the observed value $\lambda/8\pi G \simeq 10^{-47}\mathrm{GeV}^4$. Setting λ to this small value ensures that the acceleration of the universe is a fairly recent phenomenon giving rise to the 'cosmic coincidence' conundrum according to which we live during a special epoch when the density in matter and λ are almost equal. Anthropic arguments are briefly discussed but more emphasis is placed upon dynamical dark energy models in which the equation of state is time dependent. These include Quintessence, Braneworld models, Chaplygin gas and Phantom energy. Model independent methods to determine the cosmic equation of state and the Statefinder diagnostic are also discussed. The Statefinder has the attractive property $\dddot{a}\ /aH^3 = 1$ for LCDM, which is helpful for differentiating between LCDM and rival dark energy models. The review ends with a brief discussion of the fate of the universe in dark energy models.

5.1 Dark Matter

Observations of the cosmic microwave background (CMB) and the deuterium abundance in the Universe suggest that $\omega_{\rm baryon}h^2 \simeq 0.02$, or $\omega_{\rm baryon} \simeq 0.04$ if the current Hubble expansion rate is $h = H_0/100\mathrm{km/sec/Mpc} = 0.7$. Although $\omega_{\rm baryon}$ is much larger than the observed mass in stars, $\omega_{\rm stars} \simeq 0.005$ [1], it is nevertheless very much smaller than the total energy density in the universe inferred from the observed anisotropy in the cosmic microwave background [193]

$$\Omega_{\rm total} \equiv \frac{8\pi G\rho_{\rm total}}{3H^2} = 1.02 \pm 0.02\ . \tag{5.1}$$

Both dark matter and dark energy are considered essential missing pieces in the cosmic jigsaw puzzle

[1] This suggests that most of the baryonic matter at $z = 0$ is not contained in stars but might be contained in hot gas [30].

V. Sahni, Dark Matter and Dark Energy, Lect. Notes Phys. **653**, 141–179 (2005)
http://www.springerlink.com/

$$\Omega_{\rm total} - \omega_{\rm baryons} = \ ? \tag{5.2}$$

Although the nature of neither dark matter (DM) nor dark energy (DE) is currently known, it is felt that both DM and DE are non-baryonic in origin, and that DM is distinguished from DE by the fact that the former clusters on sub-Megaparsec scales (in order to explain galactic rotation curves) whereas the latter has a large negative pressure (and can make the universe accelerate). In addition there is strong evidence to suggest that

$$\Omega_m \simeq 1/3 \ , \quad \Omega_{\rm DE} \simeq 2/3 \ . \tag{5.3}$$

In this contribution I will briefly review some properties of both dark matter and dark energy.

Though the observational evidence favouring a flat Universe with $\Omega_{\rm total} \simeq 1$ is fairly recent, the nature of the 'unseen' component of the universe (which dominates its mass density), is a long-standing issue in modern cosmology. Indeed, the need for dark matter was originally pointed out by Zwicky (1933) who realized that the velocities of individual galaxies located within the Coma cluster were quite large, and that this cluster would be gravitationally bound only if its total mass substantially exceeded the sum of the masses of its component galaxies. For clusters which have relaxed to dynamical equilibrium the mean kinetic and potential energies are related by the virial theorem [50]

$$K + \frac{U}{2} = 0 \ , \tag{5.4}$$

where $U \simeq -GM^2/R$ is the potential energy of a cluster of radius R, $K \simeq 3M\langle v_r^2\rangle/2$ is the kinetic energy and $\langle v_r^2\rangle^{1/2}$ is the dispersion in the line-of-sight velocity of cluster galaxies. (Clusters in the Abell catalogue typically have $R \simeq 1.5h^{-1}$ Mpc.) This relation allows us to infer the mean gravitational potential energy if the kinetic energy is accurately known. The mass-to-light ratio in clusters can be as large as $M/L \simeq 300 M_\odot/L_\odot$. However since most of the mass in clusters is in the form of hot, x-ray emitting intracluster gas, the extent of dark matter in these objects is estimated to be $M/M_{\rm lum} \simeq 20$, where $M_{\rm lum}$ is the total mass in luminous matter including stars and gas.

In individual galaxies the presence of dark matter has been convincingly established through the use of Kepler's third law

$$v(r) = \sqrt{\frac{GM(r)}{r}} \tag{5.5}$$

to determine the 'rotation curve' $v(r)$ at a given radial distance from the galactic center. Observations of galaxies taken at distances large enough for there to be no luminous galactic component indicate that, instead of declining at the expected rate $v \propto r^{-1/2}$ true if $M \simeq$ constant, the velocity curves flattened out to $v \simeq$ constant implying $M(r) \propto r$ (see Fig. 5.1). This observation suggests that the mass of galaxies continues to grow even when

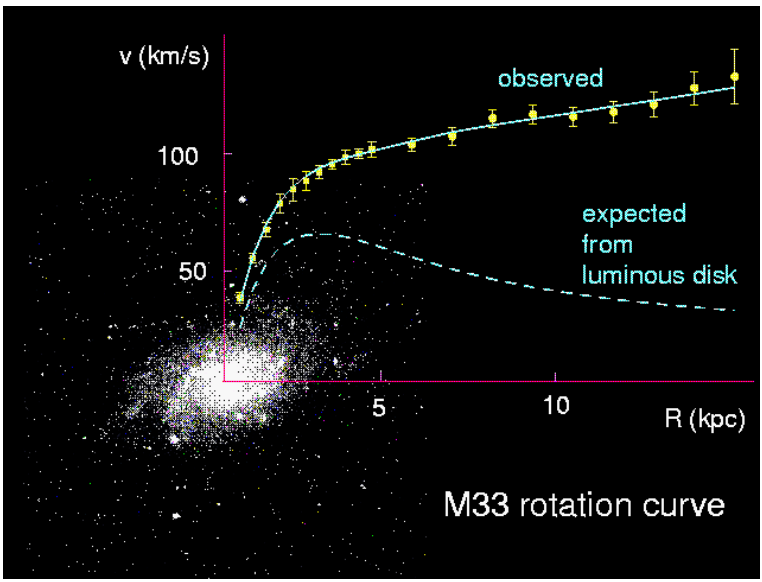

Fig. 5.1. The observed rotation curve of the dwarf spiral galaxy M33 extends considerably beyond its optical image (shown superimposed); From Roy [164].

there is no luminous component to account for this increase. Velocity curves have been compiled for over 1000 spiral galaxies usually by measuring the 21 cm emission line from neutral hydrogen (HI) [148, 191]. The results indicate that $M/L = (10-20)M_\odot/L_\odot$ in spiral galaxies and in ellipticals, while this ratio can increase to $M/L \simeq (200-600)M_\odot/L_\odot$ in low surface brightness galaxies (LSB's) and in dwarfs. For instance, a recent measurement of the Draco dwarf spheroidal galaxy located at a distance of only 79 kpc from the Milky Way shows the presence of a considerable amount of dark matter $M/L|_{\rm Draco} = (440 \pm 240)M_\odot/L_\odot$ [97] ! It is interesting that the total mass of an individual galaxy is still somewhat of an unknown quantity since a turn around to the $v \propto r^{-1/2}$ law at large radii has not been convincingly observed.

An important difference between the distribution of dark matter in galaxies and clusters needs to be emphasised: whereas dark matter appears to *increase* with distance in galaxies, in clusters exactly the reverse is true, the dark matter distribution actually *decreases* with distance. Indeed, for certain dwarfs (such as DD0154) the rotation curve has been measured to almost 15 optical length scales indicating that the dark matter surrounding this object is extremely spread out (see also Fig. 5.1). A foreground cluster, on the other hand, acts as a gravitational lens which focuses the light from background objects such as galaxies and QSO's thereby allowing us to determine the depth of the cluster potential well. Observations of strong lensing by clusters indicate that dark matter is strongly concentrated in central regions with a

projected mass of $(10^{13} - 10^{14})M_{\odot}$ being contained within (0.2 - 0.3) Mpc of the central region. As we shall see later, this observation may prove to be problematic for alternatives to the dark matter hypothesis such as the Modified Newtonian Dynamics (MOND) approach of Milgrom [122].

As discussed earlier, the fact that only 4% of the cosmic density is baryonic suggests that the dark matter which we are observing could well be non-baryonic in origin. The need for non-baryonic forms of dark matter gets indirect support from the fact that baryonic models find it difficult to grow structure from small initial conditions and hence to reconcile the existence of a well developed cosmic web of filaments, sheets and clusters at the present epoch with the exceedingly small amplitude of density perturbations ($\delta\rho/\rho \sim 10^{-5}$ at $z \simeq 1,100$) inferred from COBE measurements and more recent CMB experiments. Indeed, it is well known that, if the effects of pressure are ignored, linearized density perturbations in a spatially flat matter dominated universe grow at the rate $\delta \propto t^{2/3} \propto (1+z)^{-1}$, where $1 + z = a_0/a(t)$ is the cosmological redshift. (Contrast this relatively slow growth rate with the exponential 'Jeans instability' of a static matter distribution $\delta \propto \exp\sqrt{4\pi G\rho}t$.) In a baryonic universe, the large radiation pressure (caused by Thompson scattering of CMB photons off electrons) ensures that density perturbations in the baryonic component can begin growing only after hydrogen recombines at $z \simeq 1,100$ at which point of time baryons and radiation decouple. Requiring $\delta > 1$ today implies $\delta > 10^{-3}$ at recombination, which contradicts CMB observations by over an order of magnitude ! In non-baryonic models on the other hand, the absence of any significant coupling between dark matter and radiation allows structure to grow much earlier, significantly before hydrogen in the universe has recombined. After recombination baryons simply fall into the potential wells created for them by the dominant non-baryonic component. As a result a universe with a substantial non-baryonic component can give rise to the structure which we see today from smaller initial fluctuations.

The growth of structure via gravitational instability depends both upon the nature of primordial perturbations (adiabatic/isocurvature) and upon whether the dark matter species is hot or cold. The issue of density perturbations has been discussed in considerable detail by R. Durrer in her contribution to this volume and I shall not touch upon this important topic any further. Let me instead say a few words about hot and cold dark matter. Non-baryonic *Hot Dark Matter* (HDM) particles are assumed to have decoupled from the rest of matter/radiation when they were relativistic and so have a very large velocity dispersion (hence called 'hot'). *Cold Dark Matter* (CDM) particles, on the other hand, have a very small velocity dispersion and decouple from the rest of matter/radiation when they are non-relativistic. The free-streaming (collisionless phase mixing) of non-baryonic particles as they travel from high density to low density regions (and vice versa) introduces an important length scale called the 'free-streaming distance' $\lambda_{\rm fs}$ – which is the mean distance travelled by a relativistic particle species until its momentum becomes non-relativistic. In both HDM and CDM the processed *final*

spectrum of density perturbations differs from its initial form. In the case of HDM this difference arises because fluctuations on scales smaller than $\lambda_{\rm fs}$ are wiped out due to free streaming with the result that the processed final spectrum has a well defined cutoff on scales smaller than $\lambda \sim \lambda_{\rm fs}$. Perhaps the best example of HDM is provided by a light neutrino of mass about 30 eV. In this case $\lambda_{\rm fs} \simeq 41(30{\rm eV}/m_\nu)$ Mpc with the result that large proto-pancakes having masses comparable to those of rich clusters of galaxies $M \sim 10^{15} M_\odot$ are the first objects to form in HDM. Smaller objects (galaxies) are formed by the fragmentation of the proto-pancake. This *top-down* scenario for structure formation was originally suggested by Zeldovich and coworkers in connection with adiabatic baryonic models and subsequently applied to HDM. It has since fallen out of favour mainly due to the strong observational constraints on the mass of the neutrino $\sum_{\nu_i} m_{\nu_i} < 0.7$ eV and on the relic neutrino density $10^{-3} \lesssim \Omega_\nu h^2 \lesssim 10^{-1}$ [61, 193, 63, 123]. It also faces considerable difficulty in forming structure sufficiently early to explain the existence of galaxies and QSO's at high redshifts.

In contrast to HDM, constituents of CDM have a much smaller free-streaming distance. Because of this small scales are the first to go non-linear and gravitational clustering proceeds in a *bottom up* fashion in this scenario. A key quantity defining gravitational clustering is the power spectrum of density perturbations $P(k) \equiv |\delta_k|^2$, which is related to the mean square density fluctuation via

$$\left\langle \left(\frac{\delta\rho}{\rho} \right)^2 \right\rangle = 4\pi \int_0^\infty P(k) k^2 dk \ . \tag{5.6}$$

Inflationary models predict $P_i(k) \propto k^n$, $n \simeq 1$, at an early epoch. As the universe expands the power spectrum gets modified. The 'processed' final spectrum depends upon the nature of dark matter, the epoch of matter-radiation equality and other cosmological quantities. The final and initial spectra are related through a transfer function

$$P_f(k) = P_i(k) \times T^2(k) \ . \tag{5.7}$$

CDM-type spectra have the following approximate form of the transfer function [165, 194, 166]

$$T(k) = \left(1 + \frac{Ak^2}{\log(1 + Bk)} \right)^{-1} \ . \tag{5.8}$$

Equations (5.7) and (5.8) illustrate the 'turn around' of the power spectrum from its primordial scale invariant form $P(k) \propto k$ on the largest scales to $P(k) \propto k^{-3} \log^2 k$ on small scales. (The precise location of the turn-around and the amplitude of $P(k)$ depend upon specific details of the cosmological model, see for instance [16].)

The 'standard' cold dark matter (SCDM) paradigm, which assumed that $\Omega_{\rm CDM} = 1$, was introduced during the early 1980's at roughly the same time

when HDM was perceived to be in trouble (see [101, 96, 50, 166] for references to earlier work on this subject). Although SCDM was very successful in explaining a host of observational details, it was clear already a decade ago, that the processed power spectrum of SCDM lacked sufficient power on large scales and so fell short of explaining the angular two point correlation function for galaxies on scales ~ 50 Mpc [60]. The relevant cosmological quantity in this case is the shape of the power spectrum of density perturbations, which for CDM-like models, can be characterised by the 'shape parameter' $\Gamma = \Omega_m h$. SCDM models with $\Omega_m = 1$ and the HST-determined value $h \simeq 0.7$ predict $\Gamma \simeq 0.5$ which is much larger than the observed value $\Gamma = 0.207 \pm 0.030$ inferred from observations of galaxy clustering in the sloan digital sky survey (SDSS) [154]. A modification of SCDM called LCDM assumes that, in addition to CDM the universe consists of a smoothly distributed component called a cosmological constant or a Lambda-term. LCDM models with $h \simeq 0.7$ and $\Omega_m = 0.3$ predict a smaller value for the shape parameter, $\Gamma \simeq 0.2$, and the resulting amplitude and shape of the power spectrum is in excellent agreement with several different sets of observations as demonstrated in Fig. 5.2.

From (5.6), (5.7) and (5.8) we find that on small scales, the contribution to the *rms* density fluctuation from a given logarithmic interval in k is

$$\left(\frac{\delta\rho}{\rho}\right)_k^2 \sim k^3 P_f(k) \propto \log^2 k \ , \tag{5.9}$$

which illustrates the fact that, although the smallest scales are the first to go non-linear, there is significant power to drive gravitational instability rapidly to larger scales in this model. Indeed, detailed N-body simulations of large scale structure show that filaments defining the cosmic web first form on the smallest scales. The scale-length characterizing the cosmic web grows as the universe expands, until at the present epoch the cosmic web consists of a fully developed supercluster-void network with a scale-length of several tens of Megaparsec [181, 183, 119, 206].

Promising candidates for cold dark matter include a $(100 - 1000)$ GeV particle called a neutralino. The neutralino is a weakly interacting massive particle (WIMP). As its name suggests it is neutral and is a fermionic partner to the gauge and Higgs bosons (usually called the 'bino, wino and higgsino'). It is believed that the lightest supersymmetric particle will be stable due to R-parity which makes the neutralino an excellent candidate for cold dark matter (see [163, 89] for reviews of particle dark matter). A radically different particle candidate for cold dark matter is an ultra-light pseudo-Goldstone boson called an axion with a mass of only $m_a \sim 10^{-5\pm1}$ eV. Although ultralight, the axion is 'cold' because it was created as a zero-momentum condensate. Its existence is a by-product of an attempt to resolve QCD of what is commonly called the 'strong CP problem' which arises because non-perturbative effects in QCD give rise to an electric dipole moment for the neutron, in marked contrast with observations [101]. Other candidates for non-baryonic

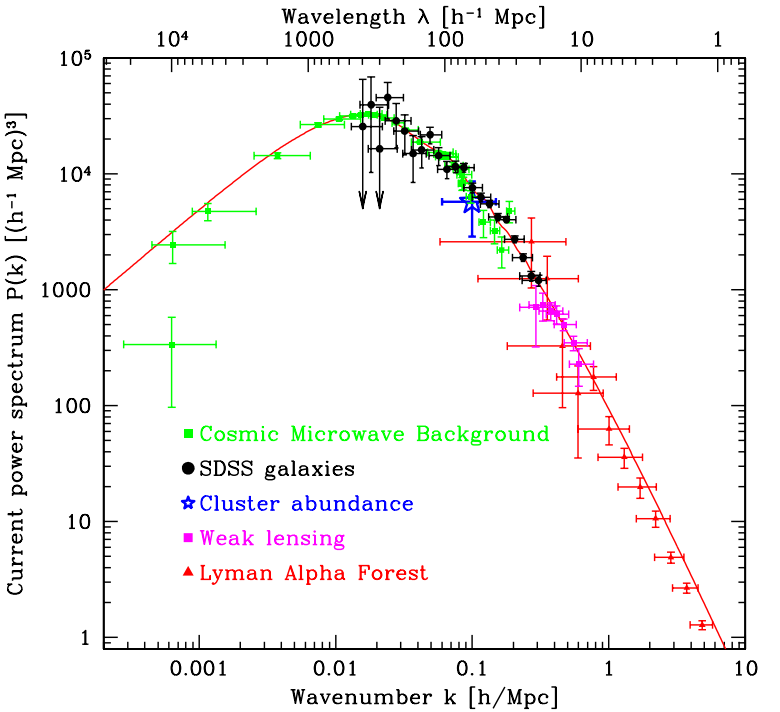

Fig. 5.2. The power spectrum inferred from observations of large scale structure, the Lymanα forest, gravitational lensing and the CMB. The solid line shows the power spectrum prediction for a flat scale-invariant LCDM model with $\Omega_m = 0.28$, $\Omega_b/\Omega_m = 0.16$, $h = 0.72$. From Tegmark et al. [200].

cold dark matter include string theory motivated modulii fields [32]; non-thermally produced super-heavy particles having a mass $\sim 10^{14}$ GeV and dubbed Wimpzillas [100]; as well as axino's and gravitino's, superpartners of the axion and graviton respectively [163].

Since WIMP's cluster gravitationally, one should expect to find a flux of these particles in our own solar system and attempts are being made to determine dark matter particles by measuring the scattering of WIMP's on target nucleii through nuclear recoils. Now the earth orbits the sun (see Fig. 5.3) with a velocity $\simeq 30$ km/sec, even as the sun orbits the galaxy with $v_{M_\odot} \simeq 220$ km/sec. Furthermore the plane of the Earth's orbit is inclined at an angle of 60° to the glactic plane, because of which the dark matter flux on Earth is expected to be larger in June (when the Earth's velocity and the Sun's velocity add together) than in December (when these two velocities subtract). The resulting rate variation is about 7% between the flux measured during summer and winter. Precisely such a signal was reported by the DAMA experiment whose data (collected since 1996) appears to show

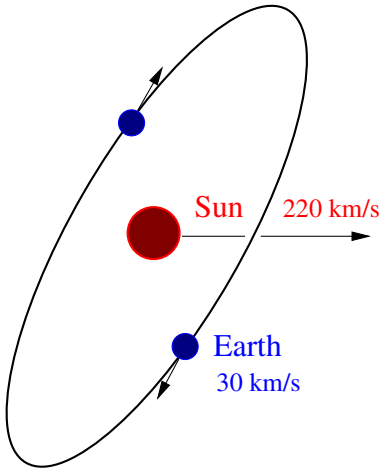

Fig. 5.3. The Earth's motion around the Sun; From Khalil and Munoz (2001).

a yearly modulation with greater events reported in June than in December [17]. However results obtained by the DAMA group remain controversial since they have not been substantiated by other groups which report negative results for similar searches (see [129, 95] for recent reviews on this subject).

Despite the excellent agreement of LCDM with large scale observations, some concerns have recently been expressed about the ability of this model to account for a number of smaller scale observations which can be summarized as follows:

– *The substructure problem*

  It is used to describe the fact that the cold dark matter model (with or without a cosmological constant) predicts an excessive number of dark matter subhaloes (or substructure) within a larger halo. If one (perhaps naively) associates each halo with a gravitationally bound baryonic object then the predicted number of dwarf-galaxy satellites within the local group exceeds the observed number by over an order of magnitude. Indeed, detailed N-body simulations as well as theoretical estimates predict around 1000 dark matter satellites in our local group which is much larger than the 40 or so observed at present [98, 126, 94, 35, 192, 24, 120, 199, 64].

– *The cuspy core problem*

  CDM predicts a *universal density profile* for dark matter halos in the wide range $(10^7 - 10^{15})M_\odot$ which applies both to galaxy clusters as well as individual galaxies including dwarfs and LSB's. [2] The density profile

[2] Low Surface Brightness Galaxies (LSB's) are dominated by their dark matter content and therefore provide particularly good astrophysical objects with which to test dark matter models.

originally suggested by Navarro, Frenk and White [133] is

$$\rho(r) = \rho_0 \, (r_s/r) \left[1 + \left(\frac{r}{r_s} \right) \right]^{-2} , \tag{5.10}$$

which gives $\rho \propto r^{-1}$ for $r \ll r_s$ and $\rho \propto r^{-3}$ for $r \gg r_s$, where r_s is the scale radius and ρ_0 is the characteristic halo density. (Other groups using higher resolution computations found somewhat steeper density profiles at small radii, such as $\rho \propto r^{-1.5}$ [127, 87].)

The cuspy core problem refers to the apparent contradiction between N-body experiments, which show that the density profile in CDM halos has a $1/r$ (or steeper) density cusp at the center, and observations, which appear to favour significantly shallower density cores in galaxy clusters as well as in individual dwarf and LSB galaxies (see [69, 36, 28, 37, 155, 199, 103, 162, 179, 109] for detailed discussions of this issue).

Although disconcerting, given the very considerable success of LCDM in explaining gravitational clustering on large scales, it may at this point be premature to condemn this model on the basis of small scale observations alone. It could be that the difficulties alluded to above are a result of an oversimplification of the complex physical processes involved and that a more careful analysis of the baryonic physics on small scales including the hydrodynamical effects of star formation and supernova feedback needs to be undertaken. For instance both dwarfs and LSB's have very shallow potential wells, a strong burst of star formation and supernova activity may therefore empty dark matter halos of their baryonic content resulting in a large number of 'failed galaxies' and providing a possible resolution to the 'satellite catastrophe'. (The failed galaxies will act as gravitational lenses and should therefore be detectable through careful observations.) Other explanations include the effects of tidal stripping recently discussed in [103]. Likewise issues involving beam smearing, the influence of bars and the interaction of baryons and dark matter in the central regions of galaxies and clusters could be intricately linked with the central cusp issue and must be better understood if one wishes to seriously test the CDM hypothesis on small scales.

In concluding this discussion on dark matter I would like to briefly mention Modified Newtonian Dynamics (MOND) which, in some circles, is regarded as an alternative to the dark matter hypothesis. As the name suggests, MOND is a modification of Newtonian physics which proposes to explain the flat rotation curves of galaxies without invoking any assumptions about dark matter. Briefly, MOND assumes that Newtons law of inertia ($F = ma$) is modified at sufficiently low accelerations ($a < a_0$) to

$$\mathbf{F} = m\mathbf{a}\mu(a/a_0) \, , \tag{5.11}$$

where $\mu(x) = x$ when $x \ll 1$ and $\mu(x) = 1$ when $x \gg 1$ [122, 180]. It is easy to see that this results in the modification of the conventional formula for

gravitational acceleration $\mathbf{F} = m\mathbf{g_N}$, resulting in the following relation between the true acceleration and the Newtonian value: $a = \sqrt{g_N a_0}$. For a body orbiting a point mass M, $g_N = GM/r^2$. Since the centripetal acceleration $a = v^2/r$ now equals the *true* acceleration a, one gets

$$v^4 = GMa_0 \ , \tag{5.12}$$

i.e. for sufficiently low values of the acceleration the rotation curve of an isolated body of mass M does not depend upon the radial distance r at which the velocity is measured, in other words not only does this theory predict flat rotation curves it also suggests that the individual halo associated with a galaxy is infinite in extent ! (This latter prediction may be a problem for MOND since recent galaxy-galaxy lensing results [82] suggest that galaxy halo's may have a maximum extent of about 0.5 Mpc.) The value of a_0 needed to explain observations is $a_0 \sim 10^{-8}\mathrm{cm/s^2}$ which is of the same order as cH_0 ! This has led supporters of this hypothesis to conjecture that MOND may reflect "the effect of cosmology on local particle dynamics" [180]. Although MOND gives results which are in good agreement with observations of individual galaxies, it is not clear whether it is as successful for explaining clusters for which strong gravitational lensing indicates a larger mass concentration at cluster centers than accounted for by MOND [180, 52]. Another difficulty with MOND is that it is problematic to embed this theory within a more comprehensive relativistic theory of gravity and hence, at present, it is not clear what predictions a MOND-type theory may make for gravitational lensing and other curved space-time effects. For some recent developments in this direction see [23].

To summarise, current observations make a strong case for clustered, non-baryonic dark matter to account for as much as a third of the total matter density in the Universe $\Omega_m \simeq 1/3$. The remaining two-thirds is thought to reside in a relative smooth component having large negative pressure and called Dark Energy.

5.2 Dark Energy

5.2.1 The Cosmological Constant and Vacuum Energy

Type Ia supernovae, when treated as standardized candles, suggest that the expansion of the universe is speeding up rather than slowing down. The case for an accelerating universe also receives independent support from CMB and large scale structure studies. All three data sets can be simultaneously satified if one postulates that the dominant component of the universe is relatively smooth, has a large negative pressure and $\Omega_{\rm DE} \simeq 2/3$.

The simplest example of dark energy is a cosmological constant, introduced by Einstein in 1917. The Einstein equations, in the presence of the cosmological constant, aquire the form

$$R_{ik} - \frac{1}{2} g_{ik} R = \frac{8\pi G}{c^4} T_{ik} + \lambda g_{ik} \ . \tag{5.13}$$

Although Einstein originally introduced the cosmological constant (λ) into the left hand side of his field equations, it has now become conventional to move the λ-term to the RHS, treating it as an effective form of matter. In a homogeneous and isotropic Friedmann-Robertson-Walker (FRW) universe consisting of pressureless dust (dark matter) and λ, the Raychaudhury equation, which follows from (5.13), takes the form

$$\ddot{a} = -\frac{4\pi G}{3} a \rho_m + \frac{\lambda}{3} \ . \tag{5.14}$$

Equation (5.14) can be rewritten in the form of a force law:

$$\mathcal{F} = -\frac{GM}{R^2} + \frac{\lambda}{3} R, \quad (R \equiv a) \tag{5.15}$$

which demonstrates that the cosmological constant gives rise to a *repulsive force* whose value increases with distance. The repulsive nature of λ could be responsible for the acceleration of the universe as demonstrated in (5.14).

Although introduced into physics in 1917, the physical basis for a cosmological constant remained a bit of a mystery until the 1960's, when it was realised that zero-point vacuum fluctuations must respect Lorenz invariance and therefore have the form $\langle T_{ik} \rangle = \lambda g_{ik}$ [214]. As it turns out, the vacuum expectation value of the energy momentum is divergent both for bosonic and fermionic fields, and this gives rise to what is known as 'the cosmological constant problem'. Indeed the effective cosmological constant generated by vacuum fluctuations is

$$\frac{\lambda}{8\pi G} = \langle T_{00} \rangle_{\rm vac} \propto \int_0^\infty \sqrt{k^2 + m^2} k^2 dk \ , \tag{5.16}$$

since the integral diverges as k^4 one gets an infinite value for the vacuum energy. Even if one chooses to 'regularize' $\langle T_{ik} \rangle$ by imposing an ultraviolet cutoff at the Planck scale, one is still left with an enormously large value for the vacuum energy $\langle T_{00} \rangle_{\rm vac} \simeq c^5/G^2\hbar \sim\sim 10^{76} {\rm GeV}^4$ which is 123 orders of magnitude larger than the currently observed $\rho_\lambda \simeq 10^{-47} {\rm GeV}^4$. A smaller ultraviolet cut-off does not fare much better since a cutoff at the QCD scale results in $\Lambda_{QCD}^4 \sim 10^{-3} {\rm GeV}^4$, which is still forty orders of magnitude larger than observed.

In the 1970's the discovery of supersymmetry led to the hope that, since bosons and fermions (of identical mass) contribute equally but with opposite sign to the vacuum expectation value of physical quantities, the cosmological constant problem may be resolved by a judicious balance between bosons and fermions in nature. However supersymmetry (if it exists) is broken at the low temperatures prevailing in the universe today and on this account one should expect the cosmological constant to vanish in the early universe,

but to reappear during late times when the temperature has dropped below $T_{\rm SUSY}$. This is clearly an undesirable scenario and almost the very opposite of what one is looking for, since, a large value of λ at an early time is useful from the viewpoint of inflation, whereas a very small current value of λ is in agreement with observations [172, 171].

In the absence of a resolution to the cosmological constant problem the following possibility connecting the vacuum energy with the SUSY and Planck scales may be worth exploring [172, 171]. The mass scale associated with the scale of supersymmetry breaking in some models, $M_{\rm SUSY} \sim 1$ TeV, lies midway between the Planck scale and 10^{-3} eV. One could conjecture that the small observed value of the cosmological constant $\rho_\lambda \simeq (10^{-3} eV)^4$ is associated with the vacuum in a theory which had a fundamental mass scale $M_X \simeq M_{\rm SUSY}^2/M_{\rm Pl}$, such that $\rho_{\rm vac} \sim M_X^4 \sim (10^{-3} eV)^4$.

The cosmological constant is also relevant from the perspective of models with spontaneous symmetry breaking [209]. Indeed, if one examines the Lagrangian

$$\mathcal{L} = \frac{1}{2}\, g^{ij}\partial_i\phi\partial_j\phi - V(\phi)$$
$$V(\phi) = V_0 - \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4, \qquad (5.17)$$

one notices that the symmetric state at $\phi = 0$ is unstable and the system settles in the ground state $\phi = +\sigma$ or $\phi = -\sigma$, where $\sigma = \sqrt{\mu^2/\lambda}$, thereby breaking the reflection symmetry $\phi \leftrightarrow -\phi$ present in the Lagrangian. For $V_0 = 0$ this potential gives rise to a large negative cosmological constant $\lambda_{eff} = V(\phi = \sigma) = -\mu^4/4\lambda$ in the broken symmetry state. This embarrassing situation can be avoided only if one chooses a value for V_0 which almost exactly cancels λ_{eff}, namely $V_0 \simeq +\mu^4/4\lambda$ so that $\lambda_{eff}/8\pi G = V_0 - \mu^4/4\lambda \simeq 10^{-47}{\rm GeV}^{\ 4}$.

The cosmological consequences of this rather ad-hoc 'regularization' exercise are instructive. Unless the value of λ_{eff} lies in a very small window, the universe will be a very different place from the one we are used to. For instance a negative value of the λ-term $\lambda_{eff}/8\pi G < -10^{-43}{\rm GeV}^{\ 4}$ will cause the universe to recollapse (the effect of λ is attractive now instead of being repulsive) less than a billion years after the big bang – a period which is much too short for galaxies to form and for life (as we know it) to emerge. On the other hand a large positive $\lambda_{eff}/8\pi G > 10^{-43}{\rm GeV}^4$ makes the universe accelerate much before the present epoch, thereby inhibiting structure formation and precluding the emergence of life.

The very small window in λ which allows life to emerge has led some cosmologists to propose anthropic arguments for the existence of a small cosmological constant [20, 118, 76, 210]. One such possibility is the following "if our big bang is just one of many big bangs, with a wide range of vacuum energies, then it is natural that some of these big bangs should have a vacuum energy in the narrow range where galaxies can form, and of course it is just

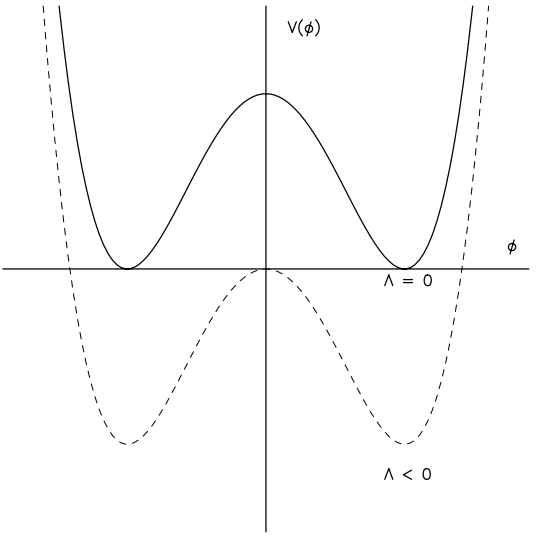

Fig. 5.4. Spontaneous symetry breaking in many field theory models takes the form of the Mexican top hat potential shown above. The dashed line shows the potential before the cosmological constant has been 'renormalized' and the solid line after. (From Sahni and Starobinsky 2000.)

these big bangs in which there could be astronomers and physicists wondering about the vacuum energy" [210].

I will not discuss the anthropic argument any further in this review but will point the interested reader to [118, 76, 210] for further discussion of this issue.

It is important to note that there is no known fundamental symmetry in nature which will set the value of λ to zero. In its absence, the small observed value of the dark energy remains somewhat of a dilemma which remains to be fully understood and resolved. [3]

5.2.2 Dynamical Models of Dark Energy

The cosmological constant is but one example of a form of matter (dark energy) which could drive an accelerated phase in the history of our universe. Indeed, (5.14) is easily generalised to

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\sum_i(\rho_i + 3p_i) = -\frac{4\pi G}{3}\sum_i \rho_i(1+3w_i)\ , \qquad (5.18)$$

[3] The important role played by symmetries is illustrated by the U(1) gauge symmetry of electrodynamics whose presence implies a zero rest mass for the photon. No analogous symmetry exists for the neutrino and recent experiments do indicate that neutrino's could have a small mass.

where the summation is over all forms of matter present in the universe with equation of state $w_i = p_i/\rho_i$. Equation (5.18) together with its companion equation

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\sum_i \rho_i - \frac{k}{a^2} \tag{5.19}$$

completely describes the dynamics of a FRW universe (k/a^2 is the Gaussian curvature of space).

Clearly a universe consisting of only a single component will accelerate if $w < -1/3$. Fluids satisfying $\rho + 3p \geq 0$ or $w \geq -1/3$ are said to satisfy the 'strong energy condition' (SEC). We therefore find that, in order to accelerate, 'dark energy' must violate the SEC. Another condition which is usually assumed to be sacrosanct, but has recently been called into question is the 'weak energy condition' (WEC) $\rho + p \geq 0$ or $w \geq -1$. Failure to satisfy the WEC can result in faster-than-exponential expansion for the universe and in a cosmic 'Big Rip', which we shall come to in a moment.

It is often more convenient to rewrite (5.18) in terms of the 'deceleration parameter'

$$q = -\frac{\ddot{a}}{aH^2} = \sum_i (\frac{4\pi G\rho_i}{3H^2})(1+3w_i) = \frac{(1+3w_{\rm X}\Omega_{\rm X})}{2}\ , \tag{5.20}$$

where $\Omega_i = 8\pi G\rho_i/3H^2$ and we have assumed a flat universe with $\Omega_{\rm m}+\Omega_{\rm X} = 1$ ($\Omega_{\rm X} \equiv \Omega_{\rm DE}$). The condition for accelerated expansion ($q < 0$) is equivalent to

$$w_X < -\frac{1}{3(1-\Omega_{\rm m})}\ , \tag{5.21}$$

which leads to

$$w < -\frac{1}{3} \quad \text{for } \Omega_m = 0 \tag{5.22}$$

$$w < -\frac{1}{2} \quad \text{for } \Omega_m = 1/3\ . \tag{5.23}$$

Equation (5.19) can be used to develop an expression for the Hubble parameter $H \equiv \dot{a}/a$ in terms of the cosmological redshift $z = a_0/a(t) - 1$:

$$H(z) = H_0\left[\Omega_m(1+z)^3 + \Omega_X(1+z)^{3(1+w)}\right]^{1/2}, \tag{5.24}$$

where $H_0 = H(z=0)$ is the present value of the Hubble parameter, $\Omega_m = 8\pi G\rho_{0m}/3H_0^2$, $\Omega_X = 8\pi G\rho_{0\rm DE}/3H_0^2$, describe the dimensionless density of matter and dark energy respectively, ($w \equiv w_{\rm DE}$), and we have made the assumption of a flat universe so that $\Omega_m + \Omega_X = 1$.

In LCDM cosmology $w = -1$, $\Omega_\Lambda = \lambda/3H_0^2$, and the expansion factor has the elegant form [172]

$$a(t) \propto \left(\sinh \frac{3}{2} \sqrt{\frac{\Lambda}{3}} ct \right)^{2/3} , \tag{5.25}$$

which smoothly interpolates between a matter dominated universe in the past ($a \propto t^{2/3}$) and accelerated expansion in the future ($a \propto \exp \sqrt{\frac{\Lambda}{3}} t$).

We are now in a position to appreciate the evidence for an accelerating universe which originates in observations of the light flux from high redshift type Ia supernovae. Type Ia supernovae are extremely bright objects, ($M_B \simeq -19.5$) which makes them ideally suited for studying the properties of the universe at large distances.

The light flux received from a distant supernova is related to its absolute luminosity $\mathcal{L}$ and its 'luminosity distance' d_L through the relation

$$F = \frac{\mathcal{L}}{4\pi d_L^2}. \tag{5.26}$$

If one views this problem from within the Newtonian perspective then, since the geometry of space is Euclidean, $d_L = \sqrt{x^2 + y^2 + z^2}$. In general relativity, on the other hand, the geometry of space can be non-Euclidean, and the luminosity distance to an object located at redshift z will, in general, depend both upon the geometry of space as well as the expansion history of the universe. Indeed, it can be shown that in a spatially flat and expanding FRW universe, the luminosity distance has the form

$$d_L(z) = (1+z) \int_0^z \frac{dz'}{H(z')}. \tag{5.27}$$

The luminosity distance is shown in Fig. 5.5 for a number of cosmological models with varying amounts of Ω_m and Ω_Λ. The limiting case $\Omega_m = 1$, $\Omega_\Lambda = 0$ corresponds to standard cold dark matter (SCDM) in which the universe decelerates as a weak power law $a(t) \propto t^{2/3}$. The other extreme example $\Omega_\Lambda = 1$, $\Omega_m = 0$ describes the de Sitter universe (also known as steady state cosmology) which accelerates at the steady rate $a(t) \propto \exp \sqrt{\frac{\Lambda}{3}} t$. From Fig. 5.5 we see that a supernova at redshift $z = 3$ will appear 9 times brighter in SCDM than it will in de Sitter space !

Systematic studies of type Ia supernovae have revealed that:

– Type Ia Sn are excellent stadardized candles. The dispersion in peak supernova luminosity is small: $\Delta m \simeq 0.3$, and the corresponding change in intensity is about 25%. In addition the light curve of a type Ia supernova is correlated with its peak luminosity [149] to a precision of $\sim 7\%$, so that *brighter supernovae take longer to fade.* (Type Ia Sn take roughly 20 days to rise from relative obscurity to maximum light.) This allows us to 'standardize' supernova light curves thereby reducing the scatter in their luminosities to $\sim 12\%$ which turns type Ia supernovae into very good standard candles.

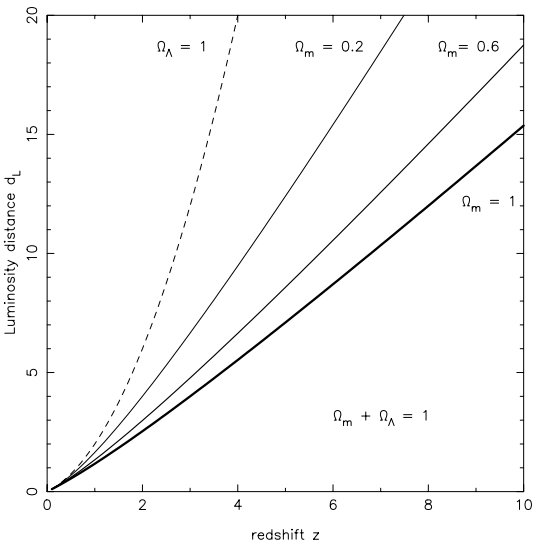

Fig. 5.5. The luminosity distance d_L (in units of H_0^{-1}) is shown as a function of cosmological redshift z for spatially flat cosmological models with $\Omega_m + \Omega_\Lambda = 1$. Heavier lines correspond to larger values of Ω_m. The dashed line shows the luminosity distance in the spatially flat de Sitter universe ($\Omega_\Lambda = 1$). From Sahni and Starobinsky [172].

– Type Ia supernovae at higher redshifts are consistently dimmer than their counterparts at lower redshifts relative to what might be expected in SCDM cosmology. If type Ia supernovae are treated as standard candles then, assuming systematic effects such as cosmological evolution and dimming by intergalactic dust are either not vitally important or have been corrected for, the systematic dimming of high$-z$ Sn can be interpreted as evidence for an accelerated expansion of the universe caused by a form of 'dark energy' having large negative pressure.

The evidence for an accelerating universe from high redshift type Ia supernovae has now received independent support from an analysis of CMB fluctuations (see A. Challinor's contribution) together with the HST key project determination of the Hubble parameter. Interestingly, the degeneracy in parameter space $\{\Omega_m, \Omega_\lambda\}$ arising from Sn observations is almost orthogonal to the degeneracy which arises from CMB measurements. This principle of 'cosmic complementarity' serves to significantly reduce the errors on Ω_m and Ω_λ when the two sets of observations are combined, as shown in Fig. 5.6.

If dark energy is described by an unevolving equation of state $w = p_X/\rho_X$, then the transition between deceleration and acceleration ($\ddot{a} = 0$) occurs at the redshift

$$(1 + z_{\rm a})^{-3w} = -(1 + 3w)\frac{\Omega_X}{\Omega_m} \;\; w < 0. \tag{5.28}$$

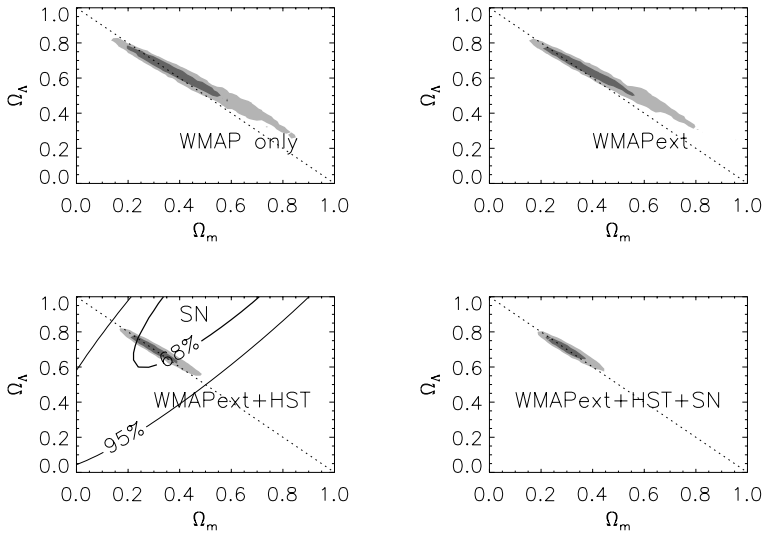

Fig. 5.6. Constraints on the density of dark matter Ω_m and dark energy in the form of a cosmological constant Ω_λ, determined using WMAP (upper left), WMAP + other CMB experiments (WMAPext; upper right), WMAPext + HST key project data (lower left) and WMAPext + HST + supernova data (lower right); from Spergel et al (2003).

Another important redshift describes the epoch when the densities in dark matter and dark energy are equal

$$(1 + z_{\rm eq})^{3w} = \left(\frac{\Omega_m}{\Omega_X}\right) \ . \tag{5.29}$$

Substituting $\Omega_\Lambda = 0.7,\ \Omega_m = 0.3$ we find $z_{\rm a} \simeq 0.73$, $z_{\rm eq} \simeq 0.37$ for LCDM. The fact that the acceleration of the universe is a fairly recent phenomenon illustrates the 'cosmic coincidence' puzzle according to which we appear to live during a special epoch when the densities in dark energy and in dark matter are almost equal. A recent origin for the acceleration epoch is supported by supernova observations which suggest a decelerating universe at $z \gtrsim 0.5$ [160]. It is important to note that dark energy models with an unevolving equation of state need to have their initial conditions properly 'tuned' in order to dominate the universe at precisely the present epoch. This problem is most acute for the cosmological constant. Since the cosmological constant does not evolve while both matter and radiation evolve rapidly ($\rho_m \propto a^{-3}$, $\rho_r \propto a^{-4}$), it follows that the small current value $\rho_\Lambda = \lambda/8\pi G \simeq 10^{-47}$ GeV^4 implies $\rho_\Lambda/\rho_r \simeq 10^{-123}$ at the Planck time (when the temperature of the universe was $T \sim 10^{19}$ GeV), or $\rho_\Lambda/\rho_r \simeq 10^{-55}$ at the time of the electroweak phase transition ($T \sim 100$ GeV). Thus an extreme fine-tuning of initial conditions is required in order to ensure that $\rho_\Lambda/\rho_m \sim 1$ today !

The fine tuning problem which plagues λ also affects DE models in which $w = \text{constant} \neq -1$. A combined analaysis of CMB, galaxy clustering and supernovae data indicates that a constant equation of state for dark energy must satisfy $w < -0.82$ at the 95% confidence level [193, 201], and it is easy to show that for these models the fine tuning (and cosmic coincidence) problems are almost as acute as they are for the cosmological constant. This constraint on w also virtually rules out two interesting DE candidates based on topological defect models: a tangled network of cosmic strings $w \simeq -1/3$ and domain walls $w \simeq -2/3$.

5.2.3 Quintessence

It is interesting that the fine tuning problem facing dark energy models with a constant equation of state can be alleviated if we assume that the equation of state is time dependent. An important class of models having this property are scalar fields (quintessence)[4] which couple minimally to gravity so that their Lagrangian density and energy momentum tensor is

$$\mathcal{L} = \frac{1}{2}\dot{\phi}^2 - V(\phi) \tag{5.30}$$

$$\rho \equiv T^0_0 = \frac{1}{2}\dot{\phi}^2 + V(\phi),\ p \equiv -T^\alpha_\alpha = \frac{1}{2}\dot{\phi}^2 - V(\phi)\ , \tag{5.31}$$

where we have assumed, for simplicity, that the field is homogeneous. Potentials which are sufficiently steep to satisfy $\Gamma \equiv V''V/(V')^2 \geq 1$ have the interesting property that scalar fields rolling down such a potential approach a common evolutionary path from a wide range of initial conditions [217] (see Fig. 5.7). In these so-called 'tracker' models the scalar field density (and its equation of state) remains close to that of the dominant background matter during most of cosmological evolution. An excellent example of a tracker potential is provided by $V(\phi) = V_0/\phi^\alpha$ [157]. During tracking the ratio of the energy density of the scalar field (quintessence) to that of radiation/matter gradually increases $\rho_\phi/\rho_B \propto t^{4/(2+\alpha)}$ while its equation of state remains marginally smaller than the background value $w_\phi = (\alpha w_B - 2)/(\alpha + 2)$. For large values of ϕ this potential becomes flat ensuring that the scalar field rolls sufficiently slowly ($\dot{\phi}^2 \ll V(\phi)$) to allow the universe to accelerate. Note that for quintessence fields the condition's (5.22) and (5.23) translate into

$$\begin{aligned} w_\phi < -\frac{1}{3} &\ \Rightarrow\ \dot{\phi}^2 < V(\phi) \\ w_\phi < -\frac{1}{2} &\ \Rightarrow\ \dot{\phi}^2 < \frac{2}{3}V(\phi)\ . \end{aligned} \tag{5.32}$$

(Current observations imply $\alpha < 2$.)

[4] Quintessence is named after the all pervasive fifth element of ancient philosophical thought. Note that the quintessence Lagrangian is the same as that used for Inflationary model building.

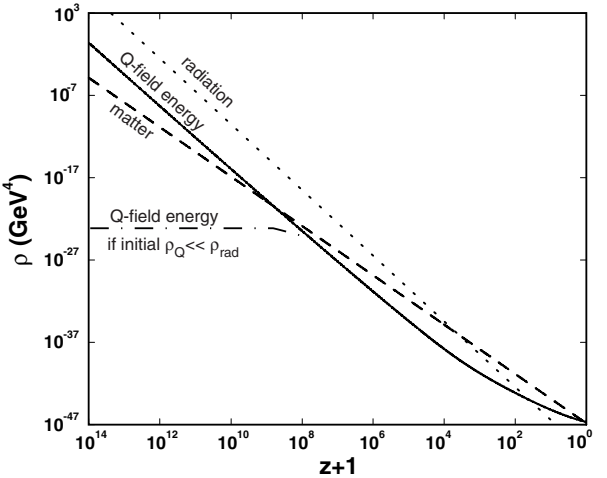

Fig. 5.7. The quintessence Q-field while rolling an inverse power law potential tracks first radiation then matter, before coming to dominate the energy density of the universe at present. If the initial value of the Q-field density is small then $\rho_{\rm Q}$ remains constant until $\rho_{\rm Q} \sim \rho_{\rm rad}$, and then follows the tracker trajectory. From Zlatev, Wang and Steinhardt [217].

An extreme example of quintessence is provided by the exponential potential $V(\phi) = V_0 \exp(-\sqrt{8\pi}\lambda\phi/M_p)$ [157, 213], where $M_p = 1/\sqrt{G}$ is the Planck mass. In this case

$$\frac{\rho_\phi}{\rho_B + \rho_\phi} = \frac{3(1+w_B)}{\lambda^2} = \text{constant} < 0.2, \tag{5.33}$$

ρ_B is the background energy density while w_B is the associated background equation of state. The lower limit $\rho_\phi/\rho_{\rm total} < 0.2$ arises because of nucleosynthesis constraints which prevent the energy density in quintessence from being large initially (at $t \sim few$ sec.). Equation (5.33) suggests that the exponential potential will remain subdominant if it was so initially. An interesting potential which interpolates between an exponential and a power law can however give rise to late time acceleration from tracker-like initial conditions [168]

$$V(\phi) = V_0[\cosh \lambda\phi - 1]^p, \tag{5.34}$$

has the property that $w_\phi \simeq w_B$ at early times whereas $\langle w_\phi \rangle = (p-1)/(p+1)$ at late times. Consequently (5.34) describes *quintessence* for $p \leq 1/2$ and pressureless 'cold' dark matter (CDM) for $p = 1$. Thus the cosine hyperbolic potential (5.34) is able to describe both dark matter and dark energy within a tracker framework (also see [204, 12]).

Remarkably, quintessence can even accommodate a constant equation of state ($w =$ constant) by means of the potential [172, 173, 203]

Table 5.1.

Quintessence Potential	Reference
$V_0 \exp(-\lambda\phi)$	Ratra and Peebles (1988), Wetterich (1988), Ferreira and Joyce (1998)
$m^2\phi^2, \lambda\phi^4$	Frieman et al (1995)
$V_0/\phi^\alpha, \alpha > 0$	Ratra and Peebles (1988)
$V_0 \exp(\lambda\phi^2)/\phi^\alpha$	Brax and Martin (1999, 2000)
$V_0(\cosh\lambda\phi - 1)^p$	Sahni and Wang (2000)
$V_0 \sinh^{-\alpha}(\lambda\phi)$	Sahni and Starobinsky (2000), Ureña-López & Matos (2000)
$V_0(e^{\alpha\kappa\phi} + e^{\beta\kappa\phi})$	Barreiro, Copeland and Nunes (2000)
$V_0(\exp M_p/\phi - 1)$	Zlatev, Wang and Steinhardt (1999)
$V_0[(\phi - B)^\alpha + A]e^{-\lambda\phi}$	Albrecht and Skordis (2000)

$$V(\phi) \propto \sinh^{\frac{2(1+w)}{w}}(C\phi + D)\ , \tag{5.35}$$

with suitably chosen values of C, D.

Quintessence models can be divided into two categories: models which roll to large values of $\phi/m_P \gtrsim 1$ and models for which $\phi/m_P \ll 1$ at the present epoch. An important concern for the former is the effect of quantum corrections which, if large, could alter the shape of the quintessence potential [102, 33, 59, 182]. An important related issue is that the coupling between standard model fields and quintessence must be small in order to have evaded detection. Moreover even small couplings between quintessence and standard model fields can give rise to interesting changes in cosmology as shown in [6, 114].

I would like to end this section by mentioning that, due to the shortage of time I have not been able to cover all of the DE models suggested in the literature (a number that is growing rapidly !) For this reason this review will not discuss DE due to vacuum polarization[167, 141], k-essence [13], Cardassian expansion [73], Quasi-Steady State Cosmology [132], scalar-tensor models [5, 26, 45, 145, 205, 146, 161, 29, 147]. For other interesting approaches see [15, 83, 84, 62, 137, 80, 104, 105, 125, 187, 215]. A partial list of some popular quintessence models is given in Table 1, and the reader is also referred to the dark energy reviews in [172, 41, 143, 171, 140].

5.2.4 Dark Energy in Braneworld Models

Inspired by the Randall-Sundrum [156] scenario, braneworld cosmology suggests that we could be living on a three dimensional 'brane' which is embedded in a higher (usually four) dimensional bulk. According to such a scheme, all matter fields are confined to the brane whereas the graviton if free to propagate in the brane as well as in the bulk (see R. Maartens's contribution to this volume and [110] for a comprehensive discussion of Braneworld cosmology.) Within the RS setting the equation of motion of a scalar field propagating on the brane is

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 \ , \tag{5.36}$$

where [184]

$$\begin{aligned} H^2 &= \frac{8\pi}{3m^2}\rho(1+\frac{\rho}{2\sigma}) + \frac{\Lambda_4}{3} + \frac{\mathcal{E}}{a^4} \\ \rho &= \frac{1}{2}\dot{\phi}^2 + V(\phi) \ . \end{aligned} \tag{5.37}$$

$\mathcal{E}$ is an integration constant which transmits bulk graviton influence onto the brane. The brane tension σ provides a relationship between the four dimensional Planck mass (m) and the five-dimensional Planck mass (M)

$$m = \sqrt{\frac{3}{4\pi}}(\frac{M^3}{\sqrt{\sigma}}) \ . \tag{5.38}$$

σ also relates the four-dimensional cosmological constant Λ_4 on the brane to the five-dimensional (bulk) cosmological constant $\Lambda_{\rm b}$ through

$$\Lambda_4 = \frac{4\pi}{M^3}\left(\Lambda_{\rm b} + \frac{4\pi}{3M^3}\sigma^2\right). \tag{5.39}$$

Note that (5.37) contains an additional term ρ^2/σ whose presence can be attributed to junction conditions imposed at the bulk-brane boundary. Because of this term the damping experienced by the scalar field as it rolls down its potential *dramatically increases* so that inflation can be sourced by potentials which are normally too steep to produce slow-roll. Indeed the slow-roll parameters in braneworld models (for $V/\sigma \gg 1$) are [111]

$$\epsilon \simeq 4\epsilon_{\rm FRW}(V/\sigma)^{-1}, \ \eta \simeq 2\eta_{\rm FRW}(V/\sigma)^{-1}, \tag{5.40}$$

illustrating that slow-roll ($\epsilon, \eta \ll 1$) is easier to achieve when $V/\sigma \gg 1$. Inflation can therefore arise for the very steep potentials associated with quintessence such as $V \propto e^{-\lambda\phi}$, $V \propto \phi^{-\alpha}$ etc. This gives rise to the intriguing possibility that both inflation and quintessence may be sourced by one and the same scalar field. Termed 'quintessential inflation', these models have

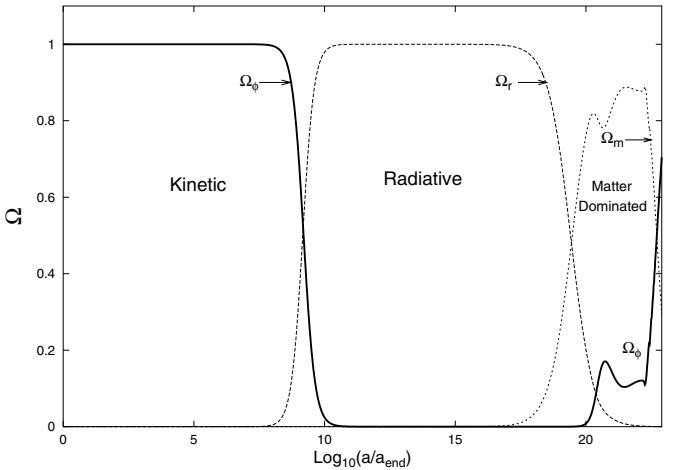

Fig. 5.8. The post-inflationary density parameter Ω is plotted for the scalar field (solid line) radiation (dashed line) and cold dark matter (dotted line) in the quintessential-inflationary model decribed by (5.34) with $p = 0.2$. Late time oscillations of the scalar field ensure that the mean equation of state turns negative $\langle w_\phi \rangle \simeq -2/3$, giving rise to the current epoch of cosmic acceleration with $a(t) \propto t^2$ and present day values $\Omega_{0\phi} \simeq 0.7, \Omega_{0m} \simeq 0.3$. From Sahni, Sami and Souradeep [169].

been examined in [142, 53, 85, 169, 116, 66, 106, 57, 185, 178, 177]. An example of quintessential inflation is shown in Fig. 5.8.

A radically different way of making the Universe accelerate was suggested in [55, 170]. The braneworld model developed by Deffayet, Dvali and Gabadadze (DDG) was radically different from the RS model in that both the bulk cosmological constant and the brane tension were set to zero, while a curvature term was introduced in the brane action so that the theory was described by

$$S = M^3 \int_{\rm bulk} \mathcal{R} + m^2 \int_{\rm brane} R + \int_{\rm brane} L_{\rm matter}\,. \tag{5.41}$$

The rationale for the $\int_{\rm brane} R$ term is that quantum effects associated with matter fields are likely to give rise to such a term in the Einstein action as discussed by Sakharov in his development of induced gravity [176].

The resulting Hubble parameter in the DDG braneworld is

$$H = \sqrt{\frac{8\pi G \rho_{\rm m}}{3} + \frac{1}{l_c^2}} + \frac{1}{l_c}\,, \tag{5.42}$$

where $l_c = m^2/M^3$ is a new length scale determined by the four dimensional Planck mass m and and the five dimensional Planck mass M respectively.

An important property of this model is that the acceleration of the universe is not caused by the presence of any 'dark energy'. Instead, since gravity becomes five dimensional on length scales $R > l_c = 2H_0^{-1}(1-\Omega_{\rm m})^{-1}$, one finds that the expansion of the universe is modified during *late times* instead of early times as in the RS model.

A more general class of braneworld models which includes RS cosmology and the DDG brane as subclasses was developed in [51, 189] and is described by the action

$$S = M^3 \int_{\rm bulk} (\mathcal{R} - 2\Lambda_{\rm b}) + \int_{\rm brane} \left(m^2 R - 2\sigma\right) + \int_{\rm brane} L_{\rm matter}\,. \qquad (5.43)$$

For $\sigma = \lambda_{\rm b} = 0$ (5.43) reduces to the action describing the DDG model, whereas for $m = 0$ it describes the Randall-Sundrum model.

As demonstrated by Sahni and Shtanov [170] the braneworld which follows from the action (5.43) describes an accelerating universe at late times with the Hubble parameter

$$\frac{H^2(z)}{H_0^2} = \Omega_{\rm m}(1+z)^3 + \Omega_\sigma + 2\Omega_l \mp 2\sqrt{\Omega_l}\,\sqrt{\Omega_{\rm m}(1+z)^3 + \Omega_\sigma + \Omega_l + \Omega_{\lambda_{\rm b}}}\,, \qquad (5.44)$$

where

$$\Omega_l = \frac{1}{l_c^2 H_0^2}\ ,\ \Omega_{\rm m} = \frac{\rho_{0m}}{3m^2H_0^2}\ ,\ \Omega_\sigma = \frac{\sigma}{3m^2H_0^2}\ ,\ \Omega_{\lambda_{\rm b}} = -\frac{\lambda_{\rm b}}{6H_0^2}\ . \qquad (5.45)$$

(The $\mp$ signs refer to the two different ways in which the brane can be embedded in the bulk, both signs give rise to interesting cosmology [170].) As in the DDG model $l_c \sim H_0^{-1}$ if $M \sim 100$ MeV. On short length scales $r \ll l_c$ and at early times, one recovers general relativity, whereas on large length scales $r \gg l_c$ and at late times brane-related effects begin to play an important role. Indeed by setting $M = 0$ ($\Omega_l = 0$) (5.44) reduces to the LCDM model

$$\frac{H^2(z)}{H_0^2} = \Omega_{\rm m}(1+z)^3 + \Omega_\sigma\ , \qquad (5.46)$$

whereas for $\sigma = \lambda_{\rm b} = 0$ (5.44) reduces to the DDG braneworld. An important feature of the braneworld (5.44) is that it can lead to an effective equation of state of dark energy $w_{\rm eff} \leq -1$. This is easy to see from the expression for the current value of the effective equation of state [170]

$$w_0 = \frac{2q_0 - 1}{3\,(1-\Omega_{\rm m})} = -1 \pm \frac{\Omega_{\rm m}}{1-\Omega_{\rm m}}\sqrt{\frac{\Omega_\ell}{\Omega_{\rm m} + \Omega_\sigma + \Omega_\ell + \Omega_{\Lambda_{\rm b}}}}\,, \qquad (5.47)$$

we find that $w_0 < -1$ when we take the lower sign in (5.47), which corresponds to choosing one of two possible embeddings of this braneworld in the higher dimensional bulk. (The second choice of embedding gives $w_0 > -1$.)

It is also possible, in this model, for the acceleration of the universe to be a transient phenomenon which ends once the universe returns to matter dominated expansion after the current accelerating phase. As discussed in [170] such a braneworld will not have an event horizon and may therefore help in reconciling an accelerating universe with the demands of string/M-theory. Other possibilities of obtaining dark energy from extra dimensions have been discussed in [7, 144, 44, 150, 151, 117, 34, 138]. The possibility that DE could arise due to modifications of gravitational physics has also been examined in [108, 40, 43, 58, 124, 135, 136].

5.2.5 Chaplygin Gas

A completely different route to dark energy is provided by the Chaplygin gas [92] which obeys the equation of state

$$p_c = -A/\rho_c \ . \tag{5.48}$$

The conservation equation $dE = -pdV \Rightarrow d(\rho a^3) = -pd(a^3)$ immediately gives

$$\rho_c = \sqrt{A + \frac{B}{a^6}} = \sqrt{A + B(1+z)^6} \ , \tag{5.49}$$

where B is a constant of integration. Thus the Chaplygin gas behaves like pressureless dust at early times and like a cosmological constant during very late times.

The Hubble parameter for a universe containing cold dark matter and the Chaplygin gas is given by

$$H(z) = H_0 \left[\Omega_m (1+z)^3 + \frac{\Omega_m}{\kappa} \sqrt{\frac{A}{B} + (1+z)^6} \right]^{1/2} \ , \tag{5.50}$$

where $\kappa = \rho_{0m}/\sqrt{B}$ and it is easy to see from (5.50) that

$$\kappa = \frac{\rho_{0m}}{\rho_c}(z \to \infty) \ . \tag{5.51}$$

Thus, κ defines the ratio between CDM and the Chaplygin gas energy densities at the commencement of the matter-dominated stage. It is easy to show that

$$A = B \left\{ \kappa^2 \left(\frac{1-\Omega_m}{\Omega_m} \right)^2 - 1 \right\} \ . \tag{5.52}$$

It is interesting that the Chaplygin gas can be derived from an underlying Lagrangian in two distinct ways:

– One can derive it from a quintessence Lagrangian (5.30) with the potential [92]

$$V(\phi) = \frac{\sqrt{A}}{2} \left(\cosh 3\phi + \frac{1}{\cosh 3\phi} \right) \ . \tag{5.53}$$

– The Chaplygin gas can also be derived from the Born-Infeld form of the Lagrangian density

$$\mathcal{L} = -V_0\sqrt{1-\phi_{,\mu}\phi^{,\mu}}\ , \tag{5.54}$$

where $\phi_{,\mu} \equiv \partial\phi/\partial x^\mu$. For time-like $\phi_{,\mu}$ one can define a four velocity

$$u^\mu = \frac{\phi^{,\mu}}{\sqrt{\phi_{,\alpha}\phi^{,\alpha}}}\ , \tag{5.55}$$

this leads to the standard form for the hydrodynamical energy-momentum tensor

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu - pg_{\mu\nu}\ , \tag{5.56}$$

where [75]

$$\rho = \frac{V_0}{\sqrt{1-\phi_{,\mu}\phi^{,\mu}}}\ ,\quad p = -V_0\sqrt{1-\phi_{,\mu}\phi^{,\mu}}\ , \tag{5.57}$$

i.e. we have recovered (5.48) with $A = V_0^2$.

The fact that the properties of the Chaplygin gas interpolate between those of CDM and a λ-term led to the hope that the CG might provide a conceptual framework for a unified model of dark matter and dark energy. It should however be noted that in contrast to CDM and baryons, the sound velocity in the Chaplygin gas $v_c = \sqrt{dp_c/d\rho_c} = \sqrt{A}/\rho_c$ quickly grows $\propto t^2$ during the matter-dominated regime and becomes of the order of the velocity of light at present (it approaches light velocity asymptotically in the distant future). Thus, when one examines classical inhomogeneities, the properties of the Chaplygin gas during the matter-dominated epoch appear to be rather unusual and resemble those of hot dark matter rather than CDM, despite the fact that the Chaplygin gas formally carries negative pressure [2].

A 'generalized Chaplygin gas' has also been proposed for which $p \propto -1/\rho^\alpha$. The equation of state in this case is

$$w(a) = -\frac{|w_0|}{\left[|w_0| + \frac{1-|w_0|}{a^{3(1+\alpha)}}\right]}\ , \tag{5.58}$$

which interpolates between $w = 0$ at early times ($a \ll 1$) and $w = -1$ at late times ($a \gg 1$); w_0 is the current equation of state at $a = 1$. (The constant α regulates the transition time in the equation of state.) WMAP, supernovae and large scale sructure data have all been used to test Chaplygin gas models; see [27, 65, 78, 9, 14, 22, 115, 128, 56, 25].

5.2.6 Is Dark Energy a Phantom?

In an influential paper Caldwell [38] noticed that a very good fit to the supernova-derived luminosity distance was provided by dark energy which

violated the weak energy condition so that $w < -1$. He called this Phantom dark energy.[5] Indeed, a study of high-z Sn [99] finds that the DE equation of state has a 99% probability of being < -1 if no priors are placed on Ω_m ! When these Sn results are combined with CMB and 2dFGRS the 95% confidence limits on an unevolving equation of state are $-1.61 < w < -0.78$ [99], which is consistent with estimates made by other groups [193, 201].

A universe filled with Phantom energy has some interesting but bizarre properties.

- If $t_{\rm eq}$ marks the epoch when the densities in matter and phantom energy are equal then the expansion factor of a universe dominated by phantom energy grows as

$$a(t) \simeq a\left(t_{\rm eq}\right)\left[(1+w)\frac{t}{t_{\rm eq}} - w\right]^{2/3(1+w)}, \quad w < -1\ , \tag{5.59}$$

  and therefore *diverges in a finite* amount of cosmic time

$$a(t) \to \infty \ \ \text{as} \ \ t \to t_{\rm BR} = \left(\frac{w}{1+w}\right) t_{\rm eq}\,. \tag{5.60}$$

  By substituting $w < -1$ into (5.24) we immediately find that the Hubble parameter also diverges as $t \to t_{\rm BR}$, implying that an infinitely rapid expansion rate for the universe has been reached in a *finite* time. The divergence of the Hubble parameter is associated with the divergence of phantom density which grows without bound

$$\rho(t) \propto \left[(1+w)\frac{t}{t_{\rm eq}} - w\right]^{-2}, \tag{5.61}$$

  and reaches a singular value in a finite interval of time $\rho(t) \to \infty, t \to t_{\rm BR}$. Thus a universe dominated by Phantom energy culminates in a future curvature singularity ('Big Rip') at which the notion of a classical space-time breaks down. (See also [196, 47, 38, 121, 39, 42, 71, 72, 186, 88, 10, 93].)
- The ultra-negative phantom equation of state suggests that the effective velocity of sound in the medium $v = \sqrt{|dp/d\rho|}$ can become larger than the velocity of light in this model.
- Although a dynamical model of phantom energy can be constructed with the 'wrong' sign of the kinetic term, see (5.31), such models are plagued with instabilities at the quantum level [49] which makes their existence suspected.

  It should be pointed out that phantom is not the only way to get $w < -1$. A model with similar properties but sharing none of phantom's pathologies is the braneworld model of [170, 1], which has $w_{\rm eff} < -1$ today but does not run into a 'Big Rip' in the future.

[5] Phantom takes its name from Part I of the Star Wars movie series – the Phantom Menace.

5.2.7 Reconstructing Dark Energy and the Statefinder Diagnostic

In view of the considerable number of dark energy models suggested in the literature, it becomes meaningful to ask whether we can reconstruct the properties of DE from observations in a model independent manner. This indeed may be possible if one notices that the Hubble parameter is related to the luminosity distance [195, 175]

$$H(z) = \left[\frac{d}{dz}\left(\frac{d_L(z)}{1+z}\right)\right]^{-1}, \tag{5.62}$$

and that, in the case of quintessence, the scalar field potential as well as its equation of state can be directly expressed in terms of the Hubble parameter and its derivative [195, 175]

$$\frac{8\pi G}{3H_0^2}V(x) = \frac{H^2}{H_0^2} - \frac{x}{6H_0^2}\frac{dH^2}{dx} - \frac{1}{2}\Omega_m\, x^3 \tag{5.63}$$

$$\frac{8\pi G}{3H_0^2}\left(\frac{d\phi}{dx}\right)^2 = \frac{2}{3H_0^2 x}\frac{d\ln H}{dx} - \frac{\Omega_m x}{H^2},\ x = 1+z \tag{5.64}$$

$$\tag{5.65}$$

$$w_\phi(x) \equiv \frac{p}{\varepsilon} = \frac{(2x/3)d\ln H/dx - 1}{1-(H_0^2/H^2)\,\Omega_m x^3}. \tag{5.66}$$

Both the quintessence potential $V(\phi)$ as well as the equation of state $w_\phi(z)$ may therefore be reconstructed provided the luminosity distance $d_L(z)$ is known to reasonable accuracy from observations.

In practice it is useful to have an ansatz for either one of three cosmological quantities: $d_L(z)$, $H(z)$ or $w(z)$, which can then be used for cosmological reconstruction [175, 131, 112, 211]. Popular fitting functions discussed in the literature include:
(i) An ansatz for the dark energy [173]

$$\rho_{\rm DE}(x) = \sum_{i=0}^{N} A_i x^i, \quad x = 1+z\ . \tag{5.67}$$

(ii) Fitting functions to the dark energy equation of state [212, 107]:

$$w(z) = \sum_{i=0}^{N} w_i z^i$$
$$w(z) = w_0 + \frac{w_1 z}{1+z}\ . \tag{5.68}$$

The fitting parameters w_i, A_i are obtained by matching to observations. In practice the first few terms in either series (5.67), (5.68) is sufficient since the current Sn data is quite noisy; see [46, 212, 54, 77, 113, 2] for a discussion

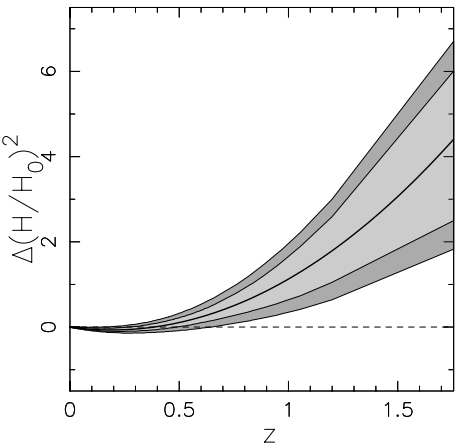

Fig. 5.9. The relative difference between the Hubble parameter reconstructed from Sn data and the LCDM value is shown as a function of redshift. Sn data from Tonry et al (2003) were used for the reconstruction. The best-fit is represented by the thick solid line assuming $\Omega_m = 0.3$. The light (dark) grey contours represents the 1σ (2σ) confidence levels around the best-fit. The dashed horizontal line shows LCDM. From Alam, Sahni, Saini and Starobinsky [4].

of these issues. An example of cosmological reconstruction of the Hubble parameter from Sn data is shown in Fig. 5.9; see also [208, 134].

The Sn inventory is increasing dramatically every year and so are increasingly precise measurements of galaxy clustering and the CMB. To keep pace with the better quality observational data which will soon become available and the increasing sophistication of theoretical modelling, a new diagnostic of DE called 'Statefinder' was introduced in [173].

The statefinder probes the expansion dynamics of the universe through higher derivatives of the expansion factor $\dddot{a}$ and is a natural companion to the deceleration parameter which depends upon $\ddot{a}$ (5.20). The statefinder pair $\{r, s\}$ is defined as follows:

$$r \equiv \frac{\dddot{a}}{aH^3} = 1 + \frac{9w}{2}\Omega_X(1+w) - \frac{3}{2}\Omega_X\frac{\dot{w}}{H} \ , \tag{5.69}$$

$$s \equiv \frac{r-1}{3(q-1/2)} = 1 + w - \frac{1}{3}\frac{\dot{w}}{wH} \ . \tag{5.70}$$

Inclusion of the statefinder pair $\{r, s\}$, increases the number of cosmological parameters to four[6]: H, q r, s. The Statefinder is a 'geometrical' diagnostic in the sense that it depends upon the expansion factor and hence upon the metric describing space-time. An important property of the Statefinder is that spatially flat LCDM corresponds to the fixed point

[6] r has also been called 'cosmic jerk' in [207].

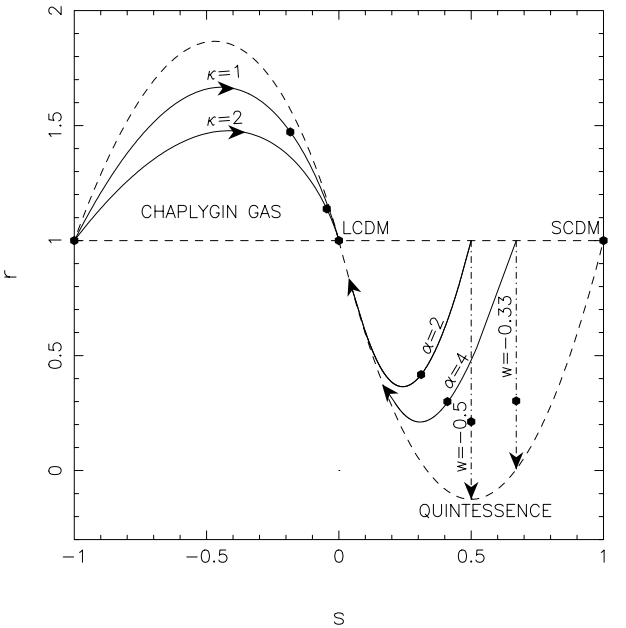

Fig. 5.10. The time evolution of the statefinder pair $\{r, s\}$ for quintessence models and the Chaplygin gas. Solid lines to the right of LCDM represent tracker potentials $V = V_0/\phi^\alpha$, while those to the left correspond to the Chaplygin gas. Dot-dashed lines represent DE with a constant equation of state w. Tracker models tend to approach the LCDM fixed point $(r = 1, s = 0)$ from the right at $t \to \infty$, whereas the Chaplygin gas approaches LCDM from the left. For Chaplygin gas κ is the ratio between matter density and the density of the Chaplygin gas at early times. The dashed curve in the lower right is the envelope of all quintessence models, while the dashed curve in the upper left is the envelope of Chaplygin gas models (the latter is described by $\kappa = \Omega_m/1-\Omega_m$). The region outside the dashed curves is forbidden for both classes of dark energy models. The ability of the Statefinder to differentiate between dark energy models is clearly demonstrated. From Alam, Sahni, Saini and Starobinsky [2].

$$\{r, s\}\Big|_{\rm LCDM} = \{1, 0\} \ . \tag{5.71}$$

Departure of a given DE model from this fixed point provides a good way of establishing the 'distance' of this model from LCDM [2]. As demonstrated in [173, 2, 78, 216] the Statefinder can successfully differentiate between a wide variety of DE models including the cosmological constant, quintessence, the Chaplygin gas, braneworld models and interacting DE models; an example is provided in Fig. 5.10.

5.2.8 Big Rip, Big Crunch or Big Horizon? – The Fate of the Universe in Dark Energy Models

The nature of dark energy affects the future of our Universe in a very significant way. If DE is simply the cosmological constant, then the universe will accelerate for ever. Of great importance is the fact that an accelerating LCDM universe developes an event horizon similar to the one surrounding a black hole [196]. Consider an event (r_1, t_1) which we wish to observe at our location at $r = 0$. Setting $ds^2 = 0$ we get

$$\int_0^{r_1} \frac{dr}{\sqrt{1-\kappa r^2}} = \int_{t_1}^{t} \frac{cdt'}{a(t')}. \tag{5.72}$$

Any event in the universe will one day be observed by us if the integral in the RHS of (5.72) *diverges* as $t \to \infty$. For power law expansion this clearly implies $a \propto t^p$, $p < 1$, *i.e.* a *decelerating universe*. In an accelerating universe exactly the opposite is true, the integral in the RHS converges signalling the presence of an event horizon. In this case our civilization will receive signals only from those events which satisfy [172]

$$\int_0^{r_1} \frac{dr}{\sqrt{1-\kappa r^2}} \leq \int_{t_1}^{\infty} \frac{cdt'}{a(t')}. \tag{5.73}$$

For de Sitter-like expansion $a = a_1 \exp H(t - t_1), H = \sqrt{\lambda/3}$, we get $r_1 = c/a_1 H$, so that the *proper distance* to the event horizon is $R_H = a_1 r_1 = c/H$. In LCDM cosmology,

$$H \equiv H(t \to \infty) = \sqrt{\lambda/3} = H_0\sqrt{1-\Omega_m}\ , \tag{5.74}$$

and the proper distance to the horizon is

$$R_H = \frac{c}{H_0\sqrt{1-\Omega_m}} \simeq 3.67 h^{-1}\ \text{Gpc}\ , \tag{5.75}$$

if $\Omega_m \simeq 1/3$. Thus our observable universe will progressively shrink as astrophysical bodies which are not gravitationally bound to the local group get pushed to distances beyond R_H. (More generally, horizons exist in a universe which begins to perpetually accelerate after a given point of time [81, 68, 171]. To this category belong models of dark energy with equation of state $-1 < w < -1/3$, as well as 'runaway scalar fields' [198] which satisfy $V, V', V'' \to 0$ and $V'/V, V''/V \to 0$ as $\phi \to \infty$.)

The presence of an event horizon implies that, at any given moment of time t_0, there is a 'sphere of influence' around our civilization. This sphere has an associated redshift z_H, and a celestial body having $z > z_H$ will be unreachable by any signal emitted by our civilization now or in the future; $z_H \simeq 1.8$ in LCDM cosmology with $\Omega_\lambda \simeq 2\Omega_m \simeq 2/3$. Thus all celestial bodies with $z > 1.8$ lie beyond our event horizon and there is no possibility of causal contact with any of them.

Interestingly, an N-body simulation tracking the future of an LCDM universe has shown that ~ 100 billion years from now the observable universe will consist of only a single massive galaxy within our event horizon, the merger product of the Milky Way and Andromeda galaxies [130]. Furthermore, since the growth of large scale structure freezes in an accelerating universe, the mass distribution of bound objects will cease to evolve after about 30 billion years.

This somewhat gloomy future scenario is not absolutely essential and can be avoided if the currently observed acceleration of the universe is a transient phenomenon.[7] Just such a possibility exists in a class of braneworld models [170] in which the current accelerating phase is succeeded by a decelerating matter dominated regime. Quintessence potentials can also have this property, as discussed in [21]. An interesting class of transiently accelerating DE models is constructed around a scalar field potential which decays with time and becomes negative at late times [74, 48, 139, 90, 91, 3]. An example is $V = V_0 \cos \phi/f$ which describes axionic quintessence [74, 48, 139, 3]. Such a universe recollapses in the future when $H(t_0 + \Delta T) = 0$, and contracts thereafter towards a 'Big Crunch' singularity. Supernova observations indicate that, for typical decaying potentials, the universe will not collapse for at least $\Delta T \simeq 20$ Gyrs [3].

DE models have also been proposed which encounter a 'quiescent singularity' while expanding. At the 'quiescent singularity' the second derivative of the expansion factor diverges while its first derivative remains finite [190, 79] (*i.e.* $\ddot{a} \to -\infty$, $\dot{a} \simeq$ constant). In such models the expansion of the universe 'brakes' to a virtual standstill as the universe approaches the singular regime at which invariants of the space-time metric diverge ($R_{iklm}R^{iklm} \to \infty$) while, curiously, the Hubble parameter and the energy density remain finite. Cosmological consequences of models which encounter a future quiescent singularity (or a 'Big Break' [79]) have been briefly discussed in [1, 190, 79] but need to be examined in more detail.

Finally, as discussed in Sect. 5.2.6, Phantom models with $w < -1$ expand towards a Big Rip, at which the density and all curvature invariants become infinite. As in the case of the Big Crunch singularity, the Big Rip will occur only in the very distant future (if it occurs at all). For instance, if $w =$ constant ≥ -1.5, $H_0 = 70$km/sec/Mpc and $\Omega_m = 0.3$, the time to the Big Rip exceeds 22 Gyr [196].

[7] Accelerating cosmologies without event horizons are important in a different context. Since the conventional S-matrix approach may not work in a universe with an event horizon, models with horizons may pose a serious challenge to a fundamental theory of interactions such as string/M-theory.

5.3 Conclusions and Future Directions

From the theoretical standpoint the single most important question to be asked of dark energy is

$$\text{Is } w = -1 \text{ ?}$$

Rephrased in terms of the Statefinder diagnostic the question is:

$$\text{Is } \dddot{a}\,/aH^3 = 1 \text{ ?}$$

If future observations do answer this question in the affirmative[8] then, in all likelyhood the cosmological constant is the vacuum energy, and one will need to review the cosmological constant problem again, in order to fathom why the formally infinite quantity $\langle T_{ik} \rangle$ is in fact so very small.

If on the other hand, either $w \neq -1$ or if the DE density is shown to be time dependent, then the cosmological constant problem may need to be decoupled from the DE conundrum and searches for evolving DE models which produce $\rho_{\rm DE} \simeq 10^{-47}{\rm GeV}^4$ without exacerbating 'cosmic coincidence' will need to be examined deeply in the light of developments both in high energy physics and in gravitation theory (superstring/M-theory, extra dimensions etc.). In either case the key to determining the properties of DE to great precision clearly lies with ongoing and future astrophysical experiments and observations.

Since the original discovery of an accelerating universe [152, 153, 159] the Sn data base has grown considerably and data pertaining to ~ 200 type Ia supernovae are avaliable in the literature [202, 99, 19, 160]. Although systematic effects such as luminosity evolution, dimming by intervening extragalactic material (alternatively brightening due to gravitational lensing) continue to be a cause of some concern, recall that a luminosity evolution of $\sim 25\%$ over a lookback time of ~ 5 Gyr is sufficient to nullify the DE hypothesis [158], it is reassuring that recent observations of CMB anisotropies and estimates of galaxy clustering in the 2dF and SDSS surveys, make a strong and independent case for dark energy [193, 200, 201]. Indeed, a joint analysis of CMB data from WMAP + HST Key Project determination of H_0 imply $w_{\rm DE} < -0.5$ at the 95% confidence level [193].

It is of paramount importance that Sn observations continue to be supplemented by other investigations which are sensitive to the geometry of space and can be used as independent tests of the DE hypothesis. The volume-redshift test, Sunyaev-Zeldovich surveys, the Alcock-Paczynski test, the angular size-redshift test and gravitational lensing have all been suggested as possible probes of dark energy, and will doubtless enrich the theory vs observations debate in the near future. In addition, the proposed SNAP satellite which aims to measure light curves of ~ 2000 supernovae within a single year [219], should provide a big step forward in our understanding of type Ia supernovae and help determine the cosmological parameters to great precision, as shown in Fig. 5.11.

[8] *i.e.* if $w = -1$ is measured to satisfyingly high accuracy.

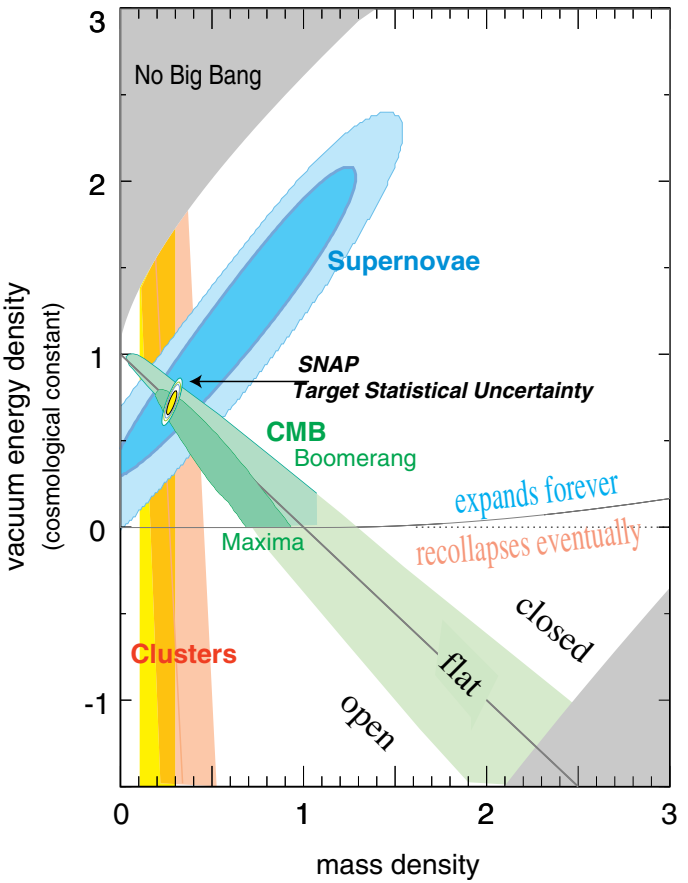

Fig. 5.11. Target statistical uncertainty of the SNAP experiment is shown overlayed with current results from CMB and LSS observations. From Aldering [11].

Acknowledgements

I thank E. Papantonopoulos and the other organisers of this school for their hospitality, and for organising such a friendly school in such a beautiful environment. I also thank Yuri Shtanov and Alexei Starobinsky for their helpful comments on the first draft of this article.

References

1. Alam, U. and Sahni, V., astro-ph/0209443.
2. Alam, U., Sahni, V., Saini, T.D. and Starobinsky, A.A., Mon. Not. Roy. Ast. Soc. **344**, 1057 (2003), astro-ph/0303009.
3. Alam, U. Sahni, V. and Starobinsky, A.A., JCAP **0304**, 002 (2003), astro-ph/0302302.
4. Alam, U., Sahni, V., Saini, T.D. and Starobinsky, A.A., astro-ph/0311364.
5. Amendola, L. (1999) Phys. Rev. D **60**, 043501.
6. Amendola, L. and Tocchini-Valentini, D. Phys. Rev. D **64**, 043509 (2001); Tocchini-Valentini, D. and Amendola, L. Phys. Rev. D **65**, 063508 (2002).
7. Albrecht, A., Burgess, C.P., Ravndal, F. and Skordis, C., Phys. Rev. D **65**, 123507 (2002).
8. Albrecht, A. and Skordis, C. Phys. Rev. Lett. **84**, 2076 (2000).
9. Alcaniz, J.S., Jain, D. and Dev, A., Phys. Rev. D **67**, 043514 (2003), astro-ph/0210476.
10. Alcaniz, J.S., astro-ph/0312424.
11. Aldering, G., astro-ph/0209550.
12. Arbey, A., Lesgourgues, J and Salati, P., Phys. Rev. D **68** 023511 (2003).
13. Armendariz-Picon, C., Mukhanov, V. and Steinhardt, P.J., Phys. Rev. Lett. **85**, 4438 (2000).
14. Avelino, C., Beca, L.M.G., de Carvalho, J.P.M., Martins, C.J.A.P. and Pinto, P., Phys. Rev. D **67**, 023511 (2003), astro-ph/0208528.
15. Axenides, M. and Dimopoulos, K., hep-th/0401238.
16. Bardeen, J.M., Bond, J.R., Kaiser, N. and Szalay, A.S., Astrophys. J.**304**, 15 (1986).
17. Barnebei, R. et al, Phys. Lett. B **480**, 23 (2000).
18. Barreiro,T., Copeland, E.J. and Nunes, N.J., Phys. Rev. D **61**, 127301 (2000).
19. Barris, B. J. *et al.*, astro-ph/0310843.
20. Barrow, J.D. and Tipler, F.J. (1986) *The Anthropic Cosmological Principle*, New York: Oxford University Press.
21. Barrow, J.D., Bean, R. and Magueijo, J., MNRAS **316**, L41 (2000).
22. Bean, R. and Dore, O., Phys. Rev. D **68** 023515 (2003), astro-ph/0301308.
23. Bekenstein, J.D., astro-ph/0403694.
24. Benson, A.J. et al. MNRAS **333**, 177 (2002).
25. Bento, M., Bertolami and Sen, A.A., Phys. Lett. B **575** 172, (2003), astro-ph/0303538.
26. Bertolami, O. and Martins, P.J. (1999) gr-qc/9910056.
27. Bilic, N., Tupper, G.B. and Viollier, R., Phys. Lett. B **535**, 17 (2002).
28. Bolatto, A,D, *et al.* Astrophys. J. **565**, 238 (2002).
29. Boisseau, B., Esposito-Farese, G., Polarski, D. and Starobinsky, A.A., Phys. Rev. Lett. **85**, 2236 (2000).
30. Bosma, A., astro-ph/0312154.
31. Brax, P. and Martin, J. Phys. Rev. D **61**, 103502 (2000); Phys. Lett. B **468**, 40 (1999).
32. Brustein, R., hep-ph/9810526.
33. Burgess, C.P., Grenier, P. and Hoover, D., hep-ph/0308252.
34. Burgess, C.P., hep-th/0402200.

35. Bullock, J.S., Kravtsov, A.V. and Weinberg, D.H., Astrophys. J.**539**, 517 (2000).
36. Burkert, A., Astrophys. J.**447**, L25 (1995).
37. de Blok, W.J.G. and McGaugh, S.S., MNRAS **290**, 533 (1997).
38. Caldwell, R.R., Phys. Lett. B **545**, 23 (2002), astro-ph/9908168.
39. Caldwell, R.R., Kamionkowski, M. and Weinberg, N.N., Phys. Rev. Lett. **91** 071301 (2003), astro-ph/0302506.
40. Capozziello, S., Carloni, S. and Troisi, A., astro-ph/0303041.
41. Carroll, S.M., Living Rev. Rel. **4** 1 (2001), astro-ph/0004075.
42. Carroll, S.M., Hoffman, M. and Trodden, M., Phys. Rev. D **68**, 023509 (2003).
43. Carroll, S.M., Duvvuri, V., Trodden, M. and Turner, M.S., astro-ph/0306438.
44. Chen, J.-W., Luty, M.A. and Ponton, E., JHEP **0009**, 012 (2000).
45. Chiba, T. Phys. Rev. D **D60**, 083508 (1999).
46. Chiba, T. and Nakamura, T., Phys. Rev. D **62**, 121301(R) (2000).
47. Chiba, T., Okabe, T. and Yamaguchi, M, 2000, Phys. Rev. D **62**, 023511.
48. Choi, K., Phys.Rev. D **62** 043509 (2000), hep-ph/9902292.
49. Cline, J.M., Jeon, S. and Moore, G.D., hep-ph/0311312.
50. Coles, P. and Lucchin, F. *Cosmology, The origin and evolution of cosmic structure*, Wiley.
51. Collins,H. and and Holdom, B., Phys. Rev. D **62**, 105009 (2000), hep-ph/0003173.
52. Combes, F., astro-ph/0206126.
53. Copeland, E.J., Liddle, A.R. and Lidsey, J.E. Phys. Rev. D **64** 023509 (2001).
54. Corasaniti, P.S. and Copeland, E.J., Phys. Rev. D **67** 063521 (2003), astro-ph/0205544.
55. Deffayet, C., Dvali, G. and Gabadadze, G., Phys. Rev. D **65**, 044023 (2002), astro-ph/0105068; Deffayet, C., Landau, S.J., Raux, J., Zaldarriaga, M. and Astier, P., Phys. Rev. D **66**, 024019 (2002), astro-ph/0201164.
56. Dev, A., Jain, D. and Alcaniz, J. astro-ph/0311056.
57. Dimopoulos, K., Phys. Rev. D **68**, 123506 (2003), astro-ph/0212264.
58. Dolgov, A.D. and Kawasaki, M, Phys. Lett. B **573**, 1 (2003), astro-ph/0307285.
59. Doran M. and Jaeckel, J., Phys. Rev. D **66**, 043519 (2003), astro-ph/0203018.
60. Efstathiou, G., Sutherland, W. and Maddox, S.J., Nature **348**, 705 (1990).
61. Elgaroy, O. et al., Phys. Rev. Lett. **90**, 021802, hep-ph/0204152.
62. Elizalde, E., Lidsey, J., Nojiri, S. and Odintsev, S.D., Phys. Lett. **B574**, 1 (2003), hep-th/0307177.
63. Ellis, J., astro-ph/0304183.
64. Evans, N.W. astro-ph/0102082.
65. Fabris, J.S., Goncalves, S.V. and de Souza, P.E., astro-ph/0207430.
66. Feng, B. and Li, M., Phys. Lett., **B** 564, 169 (2003), hep-ph/0212213.
67. Ferreira, P.G. and Joyce, M. Phys. Rev. Lett. **79**, 4740 (1997); Ferreira, P.G. and Joyce, M. Phys. Rev. D **58**, 023503 (1998).
68. Fischler, W., Kashani-Poor, A., McNees, R. and Paban, S., JHEP 0107 003 (2001), hep-th/0104181.
69. Flores, R.A. and Primack, J.R. Astrophys. J.**427**, L1 (1994).
70. Frampton, P.H. and Takahashi, T., Phys. Lett. B **557**, 135 (2003), astro-ph/0211544.
71. Frampton, P., Phys. Lett. B **555**, 139 (2003).

72. Frampton, P. and Takahashi, T., Phys. Lett. B **557**, 135 (2003).
73. Freese, K. and Lewis, M., Phys. Lett. B **540** 1 (2002), astro-ph/0201229.
74. Frieman, J., Hill, C.T., Stebbins,A. and Waga, I., Phys. Rev. Lett. **75**, 2077 (1995).
75. Frolov, A., Kofman, L. and Starobinsky, A.A., Phys. Lett. B **545**, 8, (2002), hep-th/0204187.
76. Garriga, J. and Vilenkin, A., Phys. Rev. D **61** 083502 (2000).
77. Gerke, B, & Efstathiou, G., *Mon. Not. Roy. Ast. Soc.* **335** 33 (2002), astro-ph/0201336.
78. Gorini, V., Kamenshchik, A. and Moschella, U., astro-ph/0209395.
79. Gorini, V., Kamenshchik, A., Moschella, U., Pasquier, V. hep-th/0311111.
80. Gu, P., Wang, X. and Zhang, X., Phys. Rev. D **68**, 087301 (2003), hep-ph/0307148.
81. Hellerman, S., Kaloper, N. and Susskind, L., JHEP 0106 003 (2001), hep-th/0104180.
82. Hoekstra, H., yee, H.K.C. and Gladders, M.D., astro-ph/0109514.
83. Hsu, S. and Murrey, B., astro-ph/0402541.
84. Hsu, S., hep-th/0403052.
85. Huey, G. Huey Lidsey, J. Phys. Lett. B **514**, 217 (2001).
86. Huterer, D. and Turner, M. S., Phys. Rev. D , **60** 81301 (1999).
87. Jing, Y.P. and Suto, Y., apj **529**, L69 (2000).
88. Johri, V.B., 2003, astro-ph/0311293.
89. Jungman, J., Kamionkowski, M. and Griest, K., Phys. Rep. **267**, 195 (1996).
90. Kallosh,R., Linde, A., Prokushkin, S. and Shmakova, M., Phys. Rev. D **66** 123503 (2002), hep-th/0208156.
91. Kallosh, R. and Linde,A. JCAP **02** 02 (2003), astro-ph/0301087.
92. Kamenshchik, A., Moschella, U. and Pasquier, V., Phys. Lett. B **511** 265 (2001), gr-qc/0103004.
93. Kaplinghat, M. and Bridle, S., astro-ph/0312430.
94. Kauffmann, G., White, S.D.M. and Guiderdoni, B., MNRAS **264**, 201 (1993).
95. Khalil, S. and Munoz, C., Contemp.Phys. **43**, 51 (2002), hep-ph/0110122.
96. Khlopov, M. Yu., Cosmoparticle physics, World Scientific, 1999.
97. Klenya, J., et al, Astrophys. J.**563**, L115 (2001).
98. Klypin, A. et al. Astrophys. J.**522**, 82 (1999).
99. Knop, R.A., et al., 2003, astro-ph/0309368.
100. Kolb, E., Phys. Rev. Lett. **81**, 4048 (1998).
101. Kolb, E.W. and Turner, M.S. (1990) *The Early Universe*, Addison Wesley.
102. Kolda, C. and Lyth, D.H., Phys. Lett. **B458** 197 (1999).
103. Kravtsov, A.V., Gnedin, O. and Klypin, A., astro-ph/0401088.
104. Li, M., Wang, X., Feng, B. and Zhang, X., Phys. Rev. D **65**, 103511 (2002), hep-ph/0112069.
105. Li, M., and Zhang, X., Phys. Lett. **B573**, 20 (2003), hep-ph/0209093.
106. Liddle, A. R. and Urena-Lopez, L.A., astro-ph/0302054.
107. Linder, E.V., Phys. Rev. Lett. **90** 091301, (2003), astro-ph/0208512.
108. Lue, A., Scoccimaro, R. and Starkman, G.D., Phys. Rev. D **69**, 044005 (2004), astro-ph/0307034.
109. Ma, C.P. and Boylan-Kolchin, M., astro-ph/0403102.
110. Maartens, R., gr-qc/0312059.
111. Maartens, R., Wands, D., Bassett, B.A. and Heard, I.P.C., Phys. Rev. D **62**, 041301 (2000).

112. Maor, I., Brustein, R. and Steinhardt, P.J., Phys. Rev. Lett. **86**, 6 (2001).
113. Maor, I. *et al.*, Phys. Rev. D **65** 123003 (2002), astro-ph/0112526.
114. Maccio, A.V. et al., astro-ph/0309671.
115. Mackler, M., de Oliviera, Q. S. and Waga, I., Phys. Rev. D **68** 123521 (2003), astro-ph/0306507.
116. Majumdar, A. S. Phys. Rev. D **64**, 083503 (2001).
117. Maroto, A.L., hep-ph/0402278.
118. Martel, H., Shapiro, P., and Weinberg, S., Astrophys. J.**492**, 29 (1998).
119. Martinez, V.J. and Saar, E. "Statistics of the galaxy distribution", Chapman and Hall, 2002.
120. Mateo, M.L., A.R.A.A. **36**, 435 (1998).
121. McInnes, B., JHEP **0208**, 029 (2002), hep-th/0112066.
122. Milgrom, M., Astrophys. J.**270**, 365; **270**, 371; **270**, 384 (1983).
123. Minakata, H. and Sugiyama, H., hep-ph/0212240.
124. Mofatt, J.W., astro-ph/0403266.
125. Mongan, T.R., Gen.Rel.Grav. **33**, 1415 (2001), gr-qc/0103021.
126. Moore, B. *et al.*, Astrophys. J. **524**, L19 (1999).
127. Moore, B. *et al.*, Astrophys. J. **310**, 1147 (1999).
128. Multimaki, T., Manera. M. and Gaztanaga, E., Phys. Rev. D **69** 023004, (2004), astro-ph/0307533.
129. Munoz, C., IJMPA (in press) hep-ph/0309346.
130. Nagamine, K. and Loeb, A., New Astron. **8**, 439 (2003), astro-ph/0204249.
131. Nakamura, T. and Chiba, T., *Mon. Not. Roy. Ast. Soc.* , **306**, 696 (1999).
132. Narlikar, J.V. *et al.*, Astrophys. J.**585** 1, (2003) astro-ph/0211036, and references therein.
133. Navarro, J.F., Frenk, C.S. and White, S.D.M. MNRAS **275**, 720 (1995); apj **462**, 562 (1996); Astrophys. J. **490**, 493 (1997).
134. Nesseris, S. and Perivolaroupolos, L., astro-ph/0401556.
135. Nojiri, S. and Odintsev, S.D., Phys. Rev. D **68**, 123512 (2003), hep-th/0307288.
136. Nojiri, S. and Odintsev, S.D., Phys. Lett. **B576**, 5 (2003), hep-th/0307071.
137. Nojiri, S. and Odintsev, S.D., Phys. Lett. **B562**, 147 (2003), hep-th/0303117.
138. Nojiri, S. and Odintsev, S.D., Phys. Lett. **B565**, 1, (2003), hep-th/0304131.
139. Ng, S.C. and Wiltshire, D.L., Phys.Rev. D **64** 123519 (2001), astro-ph/0107142.
140. Padmanabhan, T., Phys. Rep. **380**, 235 (2003), hep-th/0212290.
141. Parker, L. and Raval, A., Phys. Rev. D **60**, 063512, 123502 (1999).
142. Peebles, P.J.E. and Vilenkin, A., Phys. Rev. D **59** 063505 (1999).
143. Peebles, P.J.E. and Ratra, B., Rev.Mod.Phys. **75**, 559 (2002), astro-ph/0207347.
144. Perivolaropoulos, L. and Sourdis, C., Phys. Rev. D **66**, 084018 (2002).
145. Perrotta, F., Baccigalupi, C. and Matarrese, S. Phys. Rev. D **61**, 023507 (1999).
146. Perrotta, F. and Baccigalupi, C. Phys. Rev. D **59**, 123508 (1999).
147. Perrotta, F., Matarrese, S., Pietroni, M. and Schimd, C., astro-ph/0310359.
148. Persic, M., Salucci, P. and Stel, F., astro-ph/9506004.
149. Phillips, M.M. Astrophys. J.**413**, L105 (1993).
150. Peitroni, M., Phys. Rev. D **67**, 103523 (2003).
151. Peloso, M. and Poppitz, E., Phys. Rev. D **68**, 125009 (2003).

152. Perlmutter, S.J. et al., Nature **391**, 51 (1998).
153. Perlmutter, S.J. et al., Astrophys. J.**517**, 565 (1999).
154. Pope, A.C., et al, astro-ph/0401249.
155. Primack, J., astro-ph/0112255.
156. Randall, L. and Sundrum, R., Phys. Rev. Lett. **83**, 4690 (1999).
157. Ratra, B. and Peebels, P.J.E., Phys. Rev. D **37**, 3406 (1988).
158. Riess, A.G., Filipenko, A.V., Li, W. and Schmidt, B.P. Astron.J. **118** 2668 (1999), astro-ph/9907038.
159. Riess, A.G. et al., Astron. J. **116**, 1009 (1998).
160. Riess, A.G. et al., astro-ph/0402512.
161. de Ritis, R., Marino, A.A., Rubano, C. and Scudellaro, P. Phys.Rev. D **62** 043506 (2000).
162. Rhee, G., Klypin, A. and Valenzuela, O., astro-ph/0311020.
163. Roszkowski, L., hep-ph/9903467.
164. Roy, D.P. physics/0007025; also see E. Corbelli and P. Salucci, astro-ph/9909252.
165. Sahni, V., PhD thesis, Moscow State University, Moscow, 1984.
166. Sahni, V. and Coles, P., Phys. Rept., **262**, 1 (1995).
167. Sahni, V. and Habib, S., Phys. Rev. Lett. **81**, 1766, (1998), hep-ph/9808204.
168. Sahni, V. and Wang, L., Phys. Rev. D **62**, 103517 (2000).
169. Sahni, V., Sami, M. and Souradeep, T., Phys. Rev. D **65** 023518 (2002).
170. Sahni, V. and Shtanov, Yu.V., JCAP 0311,014, (2003), astro-ph/0202346.
171. Sahni, V., Class.Quant.Grav. **19** 3435 (2002), astro-ph/0202076.
172. Sahni, V. and Starobinsky, A.A. IJMP D **9**, 373 (2000).
173. Sahni, V., Saini, T.D., Starobinsky, A.A. and Alam, U., JETP Lett. **77** 201 (2003), astro-ph/0201498.
174. Sahni, V., Chaos, Solitons and Fractals **16**, 527 (2003).
175. Saini, T.D., Raychaudhury, S., Sahni, V. and Starobinsky, A.A., Phys. Rev. Lett. **85**, 1162 (2000).
176. Sakharov, A.D., Dokl. Akad. Nauk SSSR. Ser. Fiz. **177**, 70 (1967) [Sov. Phys. Dokl. **12**, 1040 (1968)]; reprinted in: Usp. Fiz. Nauk **161**, 64 (1991) [Sov. Phys. Usp. **34**, 394 (1991)]; Gen. Rel. Grav. **32**, 365 (2000).
177. Sami, M., Dadhich, N. and Shiromizu, T., Phys. Lett. B **568** 118 (2003), hep-th/0304187.
178. Sami, M. and Sahni, V., hep-th/0402086.
179. Sand, D.J., Treu, T, Smith, G.P. and Ellis, R.E., astro-ph/0309465.
180. Sanders, R.H. and McGaugh, S.S., astro-ph/0204521.
181. Sathyaprakash, B.S., Sahni, V. and Shandarin, S.F., ApJL, **462**, L5 (1996).
182. Seery, D. and Bassett, B.A., astro-ph/0310208.
183. Shandarin, S.F., Sheth, J.V. and Sahni, V., astro-ph/0312110.
184. Shiromizu, T., Maeda, K. and Sasaki, M., Phys. Rev. D **62**, 024012 (2000).
185. Shiromizu,, T., Torii, T. and Uesugi, T., hep-th/0302223.
186. Singh, P., Sami, M. and Dadhich, N.K., Phys. Rev. D **68**, 023522 (2003), hep-th/0305110.
187. Shapiro, I. and Sola, J., Phys. Lett. **B475**, 236 (2000), hep-ph/9910462.
188. Shapiro, I., Sola, J., Espana-Bonet, C. and Ruiz-Lapuente, P., Phys. Lett. **B574**, 149 (2003), astro-ph/0303306.
189. Shtanov, Yu., hep-th/0005193.
190. Shtanov, Yu. and Sahni, V., Class. Quant. Grav. **19**, L101 (2003), gr-qc/0204040.

191. Sofue, Y. and Rubin. V., astro-ph/0010594.
192. Somerville, R., Astrophys. J.**572** L23 (2002).
193. Spergel, D.N., et al, Astrophys.J.Suppl. **148**, 175 (2003), astro-ph/0302209.
194. Starobinsky, A. A. and Sahni, V., in *Modern Theoretical and Experimental Problems of General relativity* MGPI Press, Moscow, 1984, p. 77.
195. Starobinsky, A.A., JETP Lett. **68**, 757 (1998).
196. Starobinsky, A.A. Grav. Cosmol. **6**, 157 (2000).
197. Steinhardt,P.J., Wang, L., and Zlatev, I., Phys. Rev. D **59**, 123504 (1999).
198. Steinhardt, P.J., "Quintessential Cosmology and Cosmic Acceleration", http://feynman.princeton.edu/ steinh.
199. Tasitsiomi, A. astro-ph/0205464.
200. Tegmark, M. et al, astro-ph/0310725.
201. Tegmark, M. et al, astro-ph/0310723.
202. Tonry, J.L., et al., 2003, Astrophys. J.**594**, 1, astro-ph/0305008.
203. Ureña-López L.A. and Matos, T. , Phys. Rev. D **62**, 081302 (2000).
204. Ureña-López, L.A. and Liddle, A., Phys. Rev. D **66**, 083005, (2002), astro-ph/0207493.
205. Uzan, J. Phys. Rev. D **59**, 123510 (1999).
206. van de Weygaert, R., 2002, "Froth across the Universe, Dynamics and Stochastic Geometry of the Cosmic Foam", astro-ph/0206427.
207. Visser, M. gr-qc/0309109.
208. Wang. Y. and Mukherjee, P., astro-ph/0312192.
209. Weinberg, S., Rev. Mod. Phys. **61**, 1 (1989).
210. Weinberg, S., astro-ph/0104482.
211. Weller, J. and Albrecht, A. Phys. Rev. Lett. **86**, 1939 (2001).
212. Weller, J. and Albrecht, A., Phys. Rev. D **65**, 103512 (2002), astro-ph/0106079.
213. Wetterich, C., Nuclear Physics B **302**, 668 (1988).
214. Zel'dovich, Ya.B., Sov. Phys. – Uspekhi **11**, 381 (1968).
215. Zhu, Z.-H., Fujimoto, M.-K. and Tatsumi, D. Astron. Astrophys. **372**, 377 (2001), astro-ph/0107234.
216. Zimdahl, W. and Pavon, D. gr-qc/0311067.
217. Zlatev, I., Wang, L. and Steinhardt, P.J., Phys. Rev. Lett. **82**, 896 (1999).
218. Zwicky, F., *Helv. Phys. Acta* **6**, 110 (1933).
219. http://snap.lbl.gov/.

6 String Cosmology

André Lukas

Department of Physics and Astronomy, University of Sussex,
Brighton BN1 9QH, UK

Abstract. This review provides a non-technical introduction into time dependent phenomena in string theory and their possible applications to cosmology, mainly within the context of string low energy effective theories. Particular emphasis is placed on the relation between string theory and inflation. We also discuss certain topology changing processes in string theory in the context of cosmology.

6.1 Introduction

At present, M-theory represents the most promising candidate for a fundamental theory. It includes both gauge theories and gravity and is capable of reproducing at least the generic features of the standard model of particle physics. It is a natural question, therefore, if M-theory cosmology may provide a coherent picture of the evolution of the early universe and possibly lead to prediction which can be tested observationally. Despite suggestions of a low string scale of order TeV [1, 2], it seems likely to the author that the string scale is actually of the order of the Planck scale or at least not much smaller. Hence cosmology may be the only way in which M-theory can eventually be tested.

The theoretical development of M-theory is by no means a closed subject and a complete, satisfying formulation of the theory is still outstanding. String theory became popular as a candidate fundamental theory around 1985 after it had been realized that it is free of quantum anomalies and always contains among its states a spin-2 particle which can be interpreted as the graviton. For some time after that, string theory seemed to be a fairly coherent and well-founded subject dealing mostly with the perturbative formulation of the five consistent ten-dimensional string theories. However, it was realized around 1995 [3, 4, 5] that all five string theories plus eleven-dimensional supergravity were related and had to be viewed as different limits of a unique underlying theory. This hypothetical theory was named M-theory. Currently, M-theory is accessible mainly through these six limiting theories and the relations between them. It was also realized [6] that, in addition to strings, M-theory contains extended objects, so-called p-branes, of all spatial dimensions p. This state of affairs has to be kept in mind when discussing M-theory cosmology.

A. Lukas, String Cosmology, Lect. Notes Phys. **653**, 181–211 (2005)
http://www.springerlink.com/

In practise, M-theory cosmology is nowhere near being able to provide a coherent picture of the early universe or make reliable predictions, at present, although this situation may well change in the future. Consequently, this review will not present specific detailed scenarios for M-theory cosmology but rather discuss time-dependent solutions of M-theory and their possible applications in some generality. For applications, we will focus on inflation which seems to be the most likely link between particle physics and cosmology. We will also attempt to work in a context which takes requirements from particle physics into consideration. The discussion will be non-technical and mostly in the language of low-energy effective supergravity theories which is the conceptually simplest approach and probably also the most coherent one, at present.

In the next section, we start with a brief informal overview explaining the structure of M-theory and its compactifications to lower dimensions. In Sect. 6.3 we discuss a number of time-dependent M-theory solutions and their application to inflation is the topic of Sect. 6.4. Section 6.5 deals with the phenomenon of topology change in M-theory and its possible cosmological significance and we conclude in Sect. 6.6.

6.2 M-Theory Basics

This section introduces the basic structure of M-theory and a number of concepts which will be relevant in the context of cosmology. The level of the discussion is kept as informal and non-technical as possible. For more extensive reviews of string- and M-theory see, for example [7, 8, 9, 5, 10].

6.2.1 The Main Players

Currently, no fundamental formulation of M-theory which would be completely satisfying from a theoretical point of view is known. While one would hope such a formulation is found eventually, the best one can do at present is to "define" M-theory through a patchwork of different theories, all thought to be limiting cases of the complete theory, which are related by a web of so-called duality transformations. We begin by describing this structure in some detail.

M-theory can be described in terms of six constituent theories, five of them ten-dimensional theories and one eleven-dimensional. These six theories can be uniformly understood as supergravity theories, a viewpoint which we will adopt for now. Generically, the field content of these supergravity theories is given by a metric (in the appropriate space-time dimension), possibly scalar fields, anti-symmetric tensor fields of various degrees and corresponding fermions as required by supersymmetry. The associated supergravity actions can be thought of as higher-dimensional generalizations of the Einstein-Hilbert action to include the other fields and respect supersymmetry. Let us

now describe thesse theories in more detail, focusing on the bosonic fields and beginning with the single theory in eleven dimensions, eleven-dimensional supergravity. This theory is the unique supergravity theory in eleven dimensions and its field content is given by a metric g_{IJ}, a three-form C_{IJK} and associated fermions (the gravitino, in this case), where $I, J, \cdots = 0, \ldots, 10, 11$. The bosonic part of the action is given by

$$S_{11} = -\frac{1}{2}\kappa_{11}^2 \int \left[\sqrt{-g}R + G \wedge \star G + \frac{2}{3} C \wedge G \wedge G\right] + (\text{fermions and higher order terms}) \,, \tag{6.1}$$

where $G = dC$ is the four-form field strength of C and κ_{11} is the eleven-dimensional Newton constant. The theory has thirty two supersymmetries corresponding the the minimal, $N = 1$ supersymmetry in eleven dimensions.

The five ten-dimensional constituent theories of M-theory can be described as the five consistent (that is, anomaly-free) supergravity theories in this dimension. They split into two theories with thirty two supersymmetries, corresponding to extended $N = 2$ supersymmetry in ten dimensions, referred to as type IIA and IIB, and three theories with sixteen supersymmetries, that is $N = 1$ supersymmetry in ten dimensions, called $E_8 \times E_8$ heterotic , $SO(32)$ heterotic and type I. All five theories have a common sector of fields called the NSNS (for (Neveu-Schwarz Neveu-Schwarz)^2) sector which consists of the ten-dimensional metric g_{AB}, a two-form B_{AB} (with indices $A, B, \cdots = 0, \ldots, 9$) and a scalar Φ, the dilaton. The associated part of the action, also common to all five theories, reads

$$S_{\text{NS}} = -\frac{1}{2\kappa_{10}} \int \sqrt{-g} e^{-2\Phi} \left[R + 4\partial_A \Phi \partial^A \Phi + \frac{1}{12} H_{ABC} H^{ABC}\right] + (\text{fermions and higher order terms}) \,. \tag{6.2}$$

In addition to the NSNS sector, type II theories have anti-symmetric tensor fields in the so-called RR (for Ramond Ramond) sector. For type IIA these are odd-degree forms, specifically a one-form $A_{(1)}$ and a three form $A_{(3)}$ while for type IIB we have even-degree forms, specifically a zero-form $A_{(0)}$ (a scalar), a two-form $A_{(2)}$ and a (self-dual) four-form $A_{(4)}$. In addition to the terms in (6.2) the actions for type II theories contain kinetic terms $F_{(p)} \wedge \star F_{(p)}$ for the RR forms $A_{(p)}$, where $F_{(p)} = dA_{(p)}$ is the field strength and certain topological (that is, metric-independent) Chern-Simons terms involving the RR forms and the NS two-form B.

The three ten-dimensional theories with $N = 1$ supersymmetry contain non-Abelian gauge fields in addition to the NSNS sector. For the two heterotic theories the associated gauge groups are precisely as indicated by their names, while the gauge group for type I is $SO(32)$. The NS action (6.2) has to be supplemented by the kinetic terms for these non-Abelian gauge fields and a few other modifications.

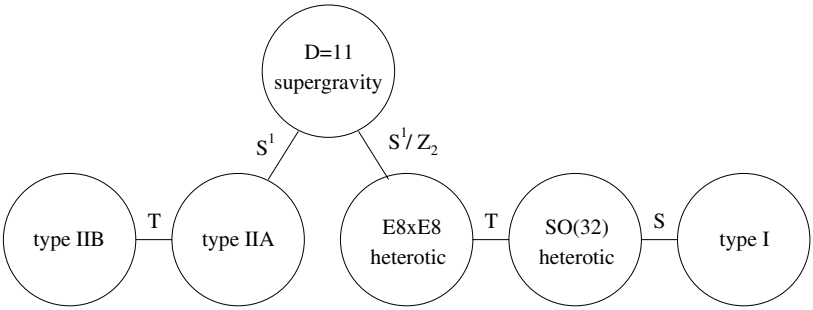

Fig. 6.1. Relation between the six constituent theories of M-theory, following [5].

These six supergravity theories should be thought of as low-energy effective theories which describe certain limits of the full M-theory in a low-energy approximation where all energies are small compared to the fundamental scale $M_{\rm F}$ set by the ten- or eleven-dimensional Newton constants. The various relations between those six theories (or, rather, their M-theory counterparts) via duality maps are indicated in Fig. 6.1. For example, Type IIA theory is (believed to be) equivalent to eleven-dimensional part of M-theory by compactifying one of the ten spatial dimensions on a small circle S^1. A simple indication why this may be true follows from a Kaluza-Klein reduction of eleven-dimensional supergravity on S^1. Splitting up the eleven-dimensional indices into the S^1 direction (identified with $I = 11$) and the remaining non-compact ten directions, labelled by $I = A$, we find the metric components g_{AB}, g_{A11} and $g_{11,11}$ while the three-form splits into C_{ABC} and C_{AB11}. These fields match the particle content of IIA, roughly identifying g_{AB} with the ten-dimensional metric, $g_{A11} \to A_{(1)A}$, $R_{11} \to g_{11,11} \to \Phi$, $C_{ABC} \to A_{(3)ABC}$ and $C_{AB11} \to B_{AB}$. Indeed, the full IIA supergravity can be shown to follow from a Kalaza-Klein reduction of 11-dimensional supergravity on S^1. It is instructive to look at the more precise relation

$$g_S \equiv e^{\Phi} \sim R_{11}^{3/2} \tag{6.3}$$

between the radius R_{11} of S^1 and the dilaton. Here g_S is the string coupling and it represents the coupling constant of IIA, as we will discuss below. Relation (6.3) tells us we should think of the eleven-dimensional theory on a large circle S^1 as the strong-coupling limit of IIA. Similarly, $D = 11$ supergravity on the orbifold $S^1/\mathbf{Z}_2$ constitutes the strong coupling limit of the $E_8 \times E_8$ heterotic theory with the relation between the heterotic string coupling g_S and the radius of the orbi-circle similar to (6.3).

The other relations in Fig. 6.1 can be described in a similar way. Type IIA and IIB are related via a T-duality transformation, which amounts to compactifying both theories on circles S^1 with respective radii $R_{\rm IIA}$ and $R_{\rm IIB}$ and subsequently identifying $R_{\rm IIA} = 1/R_{\rm IIB}$. The T-duality relation between the two heterotic theories has an analogous structure. Finally, the $SO(32)$-

heterotic theory is related to type I by an S-duality transformation which inverts the coupling constant $g_S = e^{\Phi}$. Hence, type I can be interpreted as strong coupling limit of $SO(32)$ heterotic and vice versa.

6.2.2 Branes

The supergravity fields listed in the previous subsection are not the only degrees of freedom within M-theory. There exist a variety of extended object of diverse dimensions [6], called p-branes, where p denotes the number of spatial dimension of these objects. Such p-dimensional "sheets" are parameterized by $p+1$ worldvolume coordinates σ^a, where $a, b, \cdots = 0, \ldots, p$ and σ^0 corresponds to world-sheet time, and their embedding into D-dimensional space-time can be described by D functions $X^I = X^I(\sigma^a)$ (see Fig. 6.2). Most directly, the p-branes can be obtained as specific solutions of the supergravity theories of the previous subsection. As a general rule, a D-dimensional supergravity theory with a $(p+1)$-form anti-symmetric tensor field $C_{(p+1)}$ has an "elementary" p-brane solution charged under $C_{(p+1)}$ and a "solitonic" $(D-p-4)$-brane solution charged under the Hodge-dual of $C_{(p+1)}$. The fluctuations of a p-brane can be described by a world-volume theory which (for elementary p-branes) takes the typical form

$$S_{(p)} = -T_p \int \left[d^{p+1}\sigma \sqrt{-\gamma} + C_{(p+1)} \right] + \ldots \ . \tag{6.4}$$

Here the dots stand for terms involving other possible fields, such as world-volume gauge fields or fermions. The constant T_p is called the p-brane tension and is typically of the order M_{F}^{p+1}. Further, γ_{ab} is the induced metric on the brane world-volume defined as

$$\gamma_{ab} = \partial_a X^I \partial_b X^J g_{IJ} \ . \tag{6.5}$$

Hence, the first term in the action (6.4) measures the volume of the p-brane with respect to the induced metric. The second term, called the Wess-Zumino

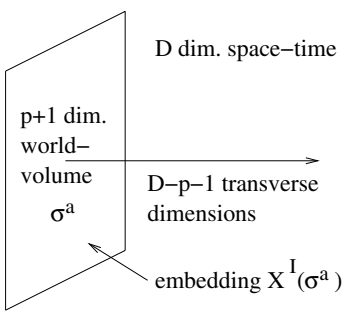

Fig. 6.2. A p-brane embedded into D-dimensional space-time.

term, indicates that the p-brane is charged under the $(p+1)$-form $C_{(p+1)}$ and it can be seen as the generalization of the term $\int A$ in the world-line action of a particle which carries charge under the $U(1)$ gauge field A.

Let us discuss the types of p-branes which arise in the various supergravity limits of M-theory, starting with eleven-dimensional supergravity. The only form of eleven-dimensional supergravity is the three-form C which couples to an elementary two-brane, also called M2-brane or M-theory membrane, and a solitonic five-brane, also called M5-brane. The five ten-dimensional supergravities all contain the NSNS two-form B which leads to an elementary one-brane or string and a solitonic five-brane. In addition, in a type II theory with a RR form $A_{(p+1)}$ we have an elementary p-brane and a solitonic $6-p$ branes. These p branes charged under RR forms are called Dp-branes (where D is for Dirichlet). Type IIA has odd-degree RR forms and, hence, Dp-branes for even p while type IIB contains even-degree forms and, hence Dp-branes for odd p.

The elementary strings which couple to the NSNS two-forms B play a particularly important role for the five 10-dimensional theories. Their two-dimensional world-volume theories can be consistenly quantized leading to a spectrum with a finite number of zero-mass states and an infinite tower of massive modes with masses of the order $n/\sqrt{\alpha'}$, where n is an integer and α' is the Regge slope, related to the string tension $T_{\rm S}$ by $T_{\rm S} = 1/(2\pi\alpha')$. It turns out that the spectrum of zero modes for each of these five string theories coincides with the spectrum of the five 10-dimensional supergravities. Moreover, the 10-dimensional supergravities can be shown to be the correct low-energy theories for the string zero modes. More precisely, they can be derived in string perturbation theory and represent the lowest-order terms arising in a double expansion in the string coupling $g_S = e^{\Phi}$, controlling loop corrections and the Regge-slope α', controlling higher-derivative corrections. Within this framework of perturbative string theory Dp-branes can also be characterized as locations where open strings can end [9].

Analogously, eleven-dimensional supergravity can be viewed as an effective theory for the zero modes of the quantized worldvolume theory of the M2-brane which corresponds to the theory of a collection of D0-branes and is known under the name of M-theory matrix model [11]. However, the M-theory matrix model and the relation to its effective theory is less well understood than its string counterparts.

In this review we will mostly focus on the low-energy effective supergravity approach since it provides the most direct and intuitive way to make contact with cosmology. However, some basic questions in string cosmology, such as the one about the fate of the initial singularity of standard cosmology, can probably not be resolved within the context of effective field theories but require the more fundamental viewpoint of quantized world-volume theories or even a proper understanding of M-theory at a fundamental level. We will return to this point later.

6.2.3 Compactification

While there may be a number of cosmological problems which can be studied in the context of a ten- or eleven-dimensional toy model it is quite clear that a successful model of string cosmology must evolve towards an effectively four-dimensional state at late time. Moreover, this effective four-dimensional theory should reproduce the known features of particle physics, that is, the particle physics standard model or some suitable (perhaps supersymmetric) extension thereof. In this review, I will discuss string cosmology in this spirit, aiming at combing the desired properties both from cosmology and particle physics.

Let us begin by introducing some of the basic concepts of compactification. For more detailed accounts see, for example, [7]. Starting point is a particular class of solutions of one of the six supergravity theories which must respect four-dimensional Poincaré invariance. Accordingly, the metric for such solutions has the general structure

$$ds^2 = e^{2A(\mathbf{y})} dx^\mu dx^\nu \eta_{\mu\nu} + g_{mn}(\mathbf{y}) dy^m dy^n \ , \tag{6.6}$$

where $\mu, \nu, \dots = 0, 1, 2, 3$ are four-dimensional indices and $\mathbf{y} = (y^m)$ are the coordinates of a six- or seven-dimensional internal compact space X with metric g_{mn}. Note that four-dimensional Poincaré invariance is compatible with the so called warp factor $\exp(2A(\mathbf{y}))$ which multiplies the four-dimensional part of the metric, as long as this factor only depends on the internal coordinates $\mathbf{y}$.

If all other background fields vanish the above metric must, to lowest order, be a solution of the free Einstein equations, that is, it must be Ricci-flat. This implies that the warp factor must be trivial, $A = 0$, and that the internal metric g_{mn} must be Ricci-flat. The simplest example for an internal space is then a torus T^6 or T^7 with a flat metric. While such torus compactifications are interesting to study on theoretical grounds they do not break any of the supersymmetries of the higher-dimensional theory and, therefore, lead to unrealistic four-dimensional models with sixteen or thirty two supercharges, corresponding to $N = 4$ or $N = 8$, respectively. However, there exist more complicated Ricci-flat spaces, namely Calabi-Yau three-folds [12] (six-dimensional manifolds with holonomy $SU(3)$) or G_2 spaces [13] (seven-dimensional spaces with holonomy G_2) which reduce the amount of supersymmetry to 1/4 or 1/8, respectively. Hence, such manifolds or suitable singular (orbifold) limits thereof, reduce the four-dimensional supersymmetry to a more suitable $N = 1$ or $N = 2$.

In more complicated cases, when fields other than the metric are non-trivial in the background, the warp-factor A can be non-vanishing and the internal metric will be no longer Ricci-flat. Particularly interesting examples are provided by flux compactifications [14, 16, 17, 18, 19, 20], where the field strengths $G_{(p)} = dC_{(p)}$ of anti-symmetric tensor fields are non-trivial on the internal space.

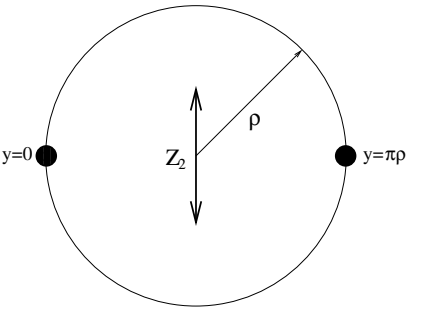

Fig. 6.3. The orbifold $S^1/\mathbf{Z}_2$.

Branes can well be part of such a compactification background. In order to preserve four-dimensional Poicaré invariance they must span the four-dimensional space and, hence, only p-branes with $p \geq 3$ are suitable. In addition, for $p > 3$, these p-branes must wrap a $(p-3)$-dimensional cycle inside the internal manifold X. Such compactifications with branes can be described by combining one of the supergravity "bulk" actions (6.1), (6.2) with p-branes actions (6.4). Anti-symmetric tensor field charges on a compact manifold must add up to zero. This cannot be achieved with p-branes alone since their charge has to be always positive for supersymmetric compactifications [1]. The internal charge can be balanced by allowing the space X to develop orbifold (or orientifold) singularities which can carry negative charge without breaking supersymmetry. The simplest example of such an orbifold is given by $S^1/\mathbf{Z}_2$ (see Fig. 6.3) where the circle coordinate y is restricted to the range $[-\pi\rho, +\pi\rho]$ (with endpoints identified) and the $\mathbf{Z}_2$ action is defined by $y \to -y$. Orbifolding by this $\mathbf{Z}_2$ identifies points y on the upper half of the circle with corresponding points $-y$ on the lower half while $y = 0$ and $y = \pi\rho$ are fixed points. Alternatively, this space can also be seen as line $[0, \pi\rho]$ with the endpoints corresponding to the fix points. Despite the singular structure of orbifolds at their fixed points string/M-theory can be frequently consistently defined on such spaces. Moreover, as mentioned before, negative form-field charges can arise at the singularities which can counterbalance the positive brane charges while preserving some supersymmetry. Formally, such singularities can be included by adding to the bulk action a term $-T\int\sqrt{-g_{\text{ind}}}$ (plus appropriate couplings to forms to account for their charge) where g_{ind} is the induced metric on the fixed location and T is the tension which, unlike brane tensions, can be negative.

Backgrounds suitable for compactifications as described above typically come in continuous families. More precisely, the metric (6.6) should then be written as

[1] Non-supersymmetric compactifications, for example, of type II theories with branes have been considered [21, 22], however, there is serious doubts about the stability of such models.

$$ds^2 = e^{2A(\mathbf{y},b^i)} dx^\mu dx^\nu g_{\mu\nu} + g_{mn}(\mathbf{y}, b^i) dy^m dy^n \ , \tag{6.7}$$

where $g_{\mu\nu}$ is the four-dimensional metric and b^i are a number of continuous parameters or moduli. Geometrically, the moduli b^i can usually be interpreted as the radii or cycle volumes of the internal manifold X. Further such moduli can arise from zero modes of the form-fields and we will generically denote these by ν^i. Zero modes can also arise from branes which are part of the background. They can be, for example, brane positions in the internal space or zero modes of other brane-world volume fields. All these zero modes become fields in the four-dimensional effective theory and we now turn to discuss its generic structure.

6.2.4 The Four-Dimensional Effective Theory

The four-dimensional effective theory can be viewed as a theory for the zero modes (or collective modes) of a class of background geometries, In this review, I will focus on backgrounds which preserve four-dimensional $N = 1$ supersymmetry which is the right amount to guarantee some degree of stability of the backgrounds and allow for contact with particle physics. Hence, the effective theory must be a four-dimensional $N = 1$ supergravity theory. In the following I will assume familiarity with such supergravity theories. For a pedagogical introduction see, for example, [23, 15].

The four-dimensional fields can be split into two groups according to their higher-dimensional origin. First, there are field which originate from the bulk and they account for gravity and all gravity-like fields, such as the bulk moduli b^i and ν^i. Secondly, there are four-dimensional fields which descend from branes and they should account for all the gauge and gauge matter fields. Accordingly, we split the chiral multiplets into two types [2], the moduli multiplets T^i, typically containing a geometrical "radius" modulus b^i and an axion ν^i

$$T^i = b^i + i\nu^i \ , \tag{6.8}$$

and the brane gauge matter fields C^a. For a successful particle physics model, those latter fields should contain the spectrum of the (supersymmetric) standard model. At the perturbative level and without flux, the Kähler potential K, the superpotential W and the gauge-kinetic function f take the generic form

$$K = K_{\rm mod}(T^i, \bar{T}^i) + Z_{ab}(T^i, \bar{T}^i) C^a \bar{C}^b \tag{6.9}$$

$$W = d_{abc}(T^i) C^a C^b C^c \tag{6.10}$$

$$f = f(T^i) \ , \tag{6.11}$$

[2] However, not quite respecting our previous distinction we will also denote brane moduli, such as internal brane positions, as T^i.

where $K_{\rm mod}$ is the moduli Kähler potential, Z_{ab} is the (moduli-dependent) matter field Kähler metric and d_{abc} are (moduli-dependent) Yukawa couplings. Higher order terms in C from both K and W have been dropped. When flux or non-perturbative effects are included the most important addition to the above structure is the introduction of an additional moduli superpotential $W_{\rm mod}(T^i)$. The validity of the effective theory is restricted by whatever restricted the validity of the original higher-dimensional theory it descended from, by possible constraints from the validity of the background solutions plus the requirement that all energies be smaller than $1/R_c$, where R_c is the typical compactification radius.

Before going into more detail, it is worth noting that the above theories provide plenty of scope for time-varying coupling constants, a topic of some interest in cosmology. In this review, we will not discuss this subject in any detail.

A central issue for cosmology as well as particle physics is the dynamics of the moduli fields T^i. Let us, therefore, consider some typical expressions for the moduli Kähler potential and superpotential. For a generic modulus $T = b + i\nu$ the Kähler potential frequently takes the form

$$K_{\rm mod} = -k \ln(T + \bar{T}) \,, \tag{6.12}$$

where k is a constant. The associated kinetic terms for b and ν are then

$$\frac{k}{b^2}(\partial_\mu b \partial^\mu b + \partial_\mu \nu \partial^\mu \nu) \,. \tag{6.13}$$

This implies the canonically normalized field is β defined by

$$b = \exp\left(\frac{\beta}{\sqrt{k}}\right) \,. \tag{6.14}$$

Typical moduli superpotentials contain terms

$$W_{\rm mod} \sim \begin{cases} e^{-cT} \\ T \end{cases} \,, \tag{6.15}$$

where the upper contribution originates from non-perturbative sources such as gaugino-condensation [14, 15] or brane instantons [24, 25] and the lower contribution originates from flux[3]. Note that in terms of the canonically normalized field β non-perturbative terms are double-exponential while flux terms are single-exponential.

Recall that the scalar potential in four-dimensional $N = 1$ supergravity for a collection $\{\phi_I\}$ of chiral multiplets is given by

$$V = e^K \left[K^{IJ} F_I \bar{F}_J - 3|W|^2\right] \,, \tag{6.16}$$

[3] Flux superpotentials can be more complicated functions of the moduli but, unlike non-perturbative contributions, they are not exponential in T.

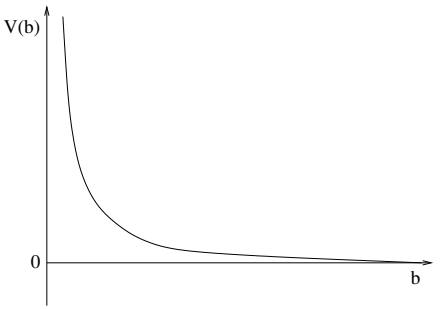

Fig. 6.4. Typical shape of the potential in the direction of a radius modulus b.

where

$$F_I = \frac{\partial W}{\partial \phi_I} + \frac{\partial K}{\partial \phi_I} W \ , \tag{6.17}$$

and K^{IJ} is the inverse of the Kähler metric

$$K_{IJ} = \frac{\partial^2 K}{\partial \phi_I \partial \bar{\phi}_J} \ . \tag{6.18}$$

This can be used to calculate the potential for moduli Kähler potentials of the type (6.12) and moduli superpotentials which are a sum of the possible contributions in (6.15). In the axion direction ν, such potentials usually have a number of substantial minima (in the absence of flux the potential is, in fact, periodic in ν). A typical shape in the b direction has been sketched in Fig. 6.4. Such a runaway shape without a minimum arises for a single non-perturbative term [14] a sum of such terms unless carefully tuned [26] or simple flux-only superpotentials (the tail falling off double-exponential or exponential in β depending on the case). A more interesting situation with a minimum at finite b and a negative cosmological constant can be achieved for a carefully tuned superposition of non-perturbative contributions [26] or, much more generically, by the superposition of a non-perturbative and a flux term entering W with different signs [14, 27]. Such a shape has been sketched in Fig. 6.5. Note that the value of b at the minimum normally corresponds to compactification radii not too far away from the fundamental length scale $1/M_F$ of the theory. From our current understanding it seems unlikely, therefore, that scenarios with very large additional dimensions [1, 2] can be embedded into M-theory. There are indications [27] that the minimum in Fig. 6.5 can be lifted to a positive cosmological constant by inclusion of certain supersymmetry breaking effects, such as anti-branes and that all moduli can be stabilized for certain constructions. This would result in a shape as shown in Fig. 6.6. A major problem in extracting concrete predictions from string theory is its large vacuum degeneracy. One manifestation of this degeneracy is the presence of the moduli which, we recall, parameterize classes of background solutions. As we have seen above, through the inclusion of moduli potentials one can at

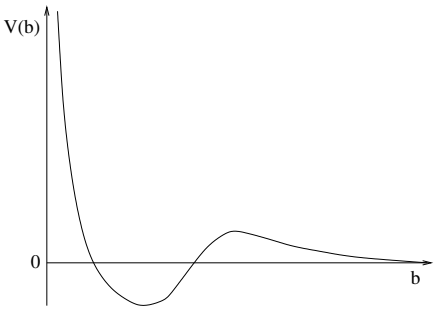

Fig. 6.5. Potential with negative cosmological constant minimum in the direction of a radius modulus b.

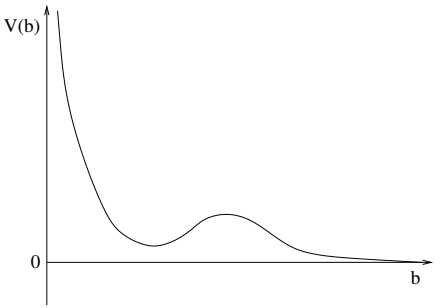

Fig. 6.6. Potential with positive cosmological constant minimum in the direction of a radius modulus b.

least begin to address the dynamics of these fields and, perhaps, single out a background geometry by fixing the moduli. We will return to this point in Sect. 6.4 when we discuss inflation. However, a different type of degeneracy arises from the existence of a large number of topologically different classes of backgrounds (see Fig. 6.7). They arise, for example, from topogically different internal manifold X or topologically different brane configurations. Low-energy theories associated to different topologies can be vastly different, such as in their particle spectrum or gauge group. The problem how to single out a specific topology is largely unsolved even in principle. We will return to this issue in Sect. 6.5 when we discuss topology change.

6.2.5 A Specific Example: Heterotic M-Theory

Let us now illustrate the previous general discussion with a concrete class of examples within heterotic M-theory [28]. The construction starts with eleven-dimensional supergravity (6.1) on the orbifold $S^1/\mathbf{Z}_2$ which has been described in Sect. 6.2.3. In addition to the fields of eleven-dimensional supergravity, this introduces two ten-dimensional E_8 super-Yang-Mills multiplets, one an each of the ten-dimensional orbifold fixed planes (boundaries). This

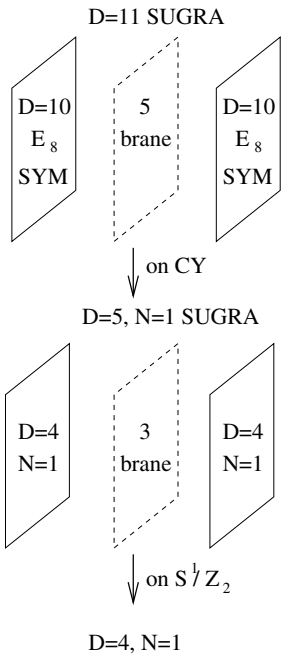

Fig. 6.7. Reduction of heterotic M-theory to five and four dimensions.

theory constitutes the strong coupling limit of the $E_8 \times E_8$ heterotic string (see Fig. 6.1) in much the same way eleven-dimensional supergravity is the strong coupling limit of IIA.

Upon reduction of this theory on Calabi-Yau three-folds [29, 30] one obtains a class of five-dimensional brane-world theories with a bulk described by five-dimensional $N = 1$ supergravity and coupled to two four-dimensional $N = 1$ theories localized on the two, now four-dimensional boundaries M_1 and M_2. There exist examples where the theory on one of the boundaries is very close to the (supersymmetric) standard model of particle physics [31, 32] and one expects a large number of such examples to exist. Further, one can introduce M5-branes wrapping two-cycles within the Calabi-Yau three-fold into the compactification. In the five-dimensional brane-world theory, they become 3-branes which also carry a four-dimensional $N = 1$ theory. For simplicity, we consider a single such 3-brane which (together with its $\mathbf{Z}_2$ mirror) we denote by M_3. Taking the radius of the orbifold sufficiently small one can further derive an effective four-dimensional $N = 1$ theory. The two steps of this reduction are summarized in Fig. 6.7. A minimal version of five-dimensional heterotic M-theory is given by [30, 33]

$$S = S_{\text{bulk}} + \sum_{n=1}^{3} S_n \ , \tag{6.19}$$

where

$$S_{\text{bulk}} = -\frac{1}{2\kappa_5^2}\int d^5x\sqrt{-g}\left[\frac{1}{2}R + \frac{1}{4}\partial\varphi^2 + \frac{1}{3}\alpha^2 e^{-2\varphi}\right] \tag{6.20}$$

$$S_n = -\frac{1}{\kappa_5^2}\int_{M_n} d^4x\sqrt{-g_4}\alpha_n e^{-\varphi} \text{ for } n = 1,2 \tag{6.21}$$

$$S_3 = -\frac{1}{2\kappa_5^2}\int_{M_3}\sqrt{-\gamma}|\alpha_3|e^{-\varphi}\ . \tag{6.22}$$

Here, we have only kept the minimal field content and have, in particular, omitted gauge and matter fields from the boundary actions S_1 and S_2. More complete actions can be found in [29, 34]. The parameter α which appears in the bulk potential is defined by

$$\alpha = \sum_{n=1}^{3}\alpha_n\theta(M_n)\ , \tag{6.23}$$

where θ is the theta function. The boundary charges α_1, α_2 and the 3-brane charge α_3 have to satisfy

$$\sum_{n=1}^{3}\alpha_n = 0\ , \tag{6.24}$$

so that the total charge is zero. We note that the boundary tensions can be negative (since α_1 and α_2 can be) while the three-brane tension is always positive, as it should. For $\alpha_3 > 0$ we have a three-brane and the theory is supersymmetric as described above. We can also consider the case of an anti three-brane, taking $\alpha_3 < 0$. In this case supersymmetry is completely broken. The scalar field φ is the five-dimensional dilaton and e^{φ} is proportional to the volume of the internal Calabi-Yau three-fold.

The associated four-dimensional effective $N = 1$ theory (for $\alpha_3 > 0$) can be obtained as the theory for the collective modes of a class of domain wall solutions of the above five-dimensional action [30]. It contains three geometrical moduli, namely $\beta \sim \ln < g_{55} >$ measuring the radius of the orbifold, $\phi \sim < \varphi >$ which measures the volume of the Calabi-Yau space and $z \in [0,1]$ which specifies the position of the 3-brane (where $z = 0$ and $z = 1$ correspond to the boundaries M_1 and M_2, respectively). Along with these fields go three chiral superfields

$$\begin{aligned} S &= e^{\phi} + \ldots \\ T &= e^{\beta} + \ldots \\ Z &= e^{\beta} z + \ldots \end{aligned} \tag{6.25}$$

where the dots indicate axion and higher-order terms. The Kähler potential for these moduli is given by [35, 34]

$$K_{\rm mod} = -\ln\left[S+\bar{S}-q_3\frac{(Z+\bar{Z})^2}{T+\bar{T}}\right]-3\ln(T+\bar{T})\ , \tag{6.26}$$

where q_3 is a constant proportional to α_3. Non-perturbative potentials take the general form

$$W_{\rm mod}\sim\exp(-c_1S+c_2T+c_3Z)\ , \tag{6.27}$$

with positive constants c_n while a flux superpotential is a constant [4] in S, T and Z.

We stress that the above four-dimensional effective theory is in general different from the four-dimensional "on-the-brane" theory obtained by projecting the five-dimensional brane-world theory onto one of the boundaries. The four-dimensional effective theory provides a valid description of all five-dimensional dynamics sufficiently close to the domain wall vacua of the five-dimensional theory. The "on-the-brane" theory, on the other hand, provides an incomplete description of all solutions of the five-dimensional theory. Therefore, working with the "on-the-brane" theory can be misleading in that it suggests four-dimensional physics while, in fact, it may describe situations for away from the effective four-dimensional limit required at late time.

We have argued above, that compact dimensions in M-theory are unlikely to be much larger than the fundamental length scale of the theory. Accordingly, most of the evolution can be safely described in terms of the effective four-dimensional theory. Only in the very early universe, possibly during inflation, and if the orbifold size is relatively large may working with the five-dimensional theory be required.

6.3 Classes of Simple Time-Dependent Solutions

At the heart of cosmological models is a background solution with a four-dimensional space-time metric of Friedmann-Robertson-Walker (FRW) type. Finding FRW solutions in the context of M-theory, therefore, seems to be a natural starting point for M-theory cosmology. In this section, we present simple classes of such solutions and discuss their properties. For an extensive list of references to the vast literature on the subject see, for example, [37].

6.3.1 Rolling Radii Solutions

We would now like to consider the free evolution (that is, the potential is set to zero) of a number of four-dimensional moduli fields $b^i=\exp(\beta_i)$ which measure radii of the internal manifold X. Such solutions are also referred to as rolling radii solutions [38] and are the analogue of Kasner solutions in

[4] However, it depends on another class of moduli, the complex structure moduli which we have not discussed explicitly. Their values are usually fixed by the flux potential [36].

M-theory. As we have explained in Sect. 6.2.4, moduli potentials in M-theory rapidly decrease for large values of the radius moduli. Hence, rolling radii solutions should provide a good approximation to the dynamics at sufficiently large values of the radii.

Let us start with a class of examples [39] which arises in the context of compactifying eleven-dimensional supergravity on a manifold of G_2 holonomy. The G_2 manifolds under consideration have seven radius moduli β^i with associated chiral multiplet $T^i = \exp(\beta^i) + \text{axion}$ plus a number of additional moduli which we suppress. The full model has been analyzed in [40]. The T^i part of the Kähler potential is simply

$$K_{\text{mod}} = -\sum_{i=1}^{7} \ln(T^i + \bar{T}^i) \,, \tag{6.28}$$

One can consistently set the axions in T^i to constants which results in the simple Lagrangian

$$\mathcal{L} = -\sqrt{-g} \left[R + \sum_{i=1}^{7} \partial\beta^i \partial\beta^i \right] \,. \tag{6.29}$$

Consider an Ansatz with time-dependent moduli $\beta^i = \beta^i(t)$ and a FRW metric

$$ds^2 = -dt^2 + e^{2\alpha(t)} d\mathbf{x}^2 \tag{6.30}$$

with scale factor α and, for simplicity, flat spatial sections. The solutions are then given by

$$a \equiv e^{\alpha} \sim |t|^{\frac{1}{3}} \tag{6.31}$$

$$\beta_i = p_i \ln|t| + c_i \,. \tag{6.32}$$

The expansion powers $\mathbf{p} = (p_i)$ are constrained by

$$\mathbf{p}^2 = \frac{4}{3} \,, \tag{6.33}$$

while c_i are arbitrary constants. Note that the expansion power of the scale factor a is $1/3$, as expected for a kinetic energy driven expansion. Every one of the above solutions exists on two branches, the positive-time branch, $t > 0$, which begins in a curvature singularity and corresponds to sub-luminal expansion of the universe and the negative-time branch, $t < 0$, which ends in a curvature singularity and describes super-luminal contraction. Of course, close to the singularity in either branch the four-dimensional effective field theory breaks down as higher-derivative corrections become important. Whether and how this singularity is regulated in the full M-theory must still be viewed as an open question to which we return in Sect. 6.4.3. We also note that the power-law evolution of the scalar fields β^i describes the expansion or contraction of the internal G_2 space X.

Another, similar class of rolling radii solutions can be obtained in the context of four-dimensional heterotic M-theory without five-branes. We start with the Kähler potential (6.26), drop the five-brane modulus Z and write $S = e^{\phi} + \text{axion}$, $T = e^{\beta} + \text{axion}$. This leads to the Lagrangian

$$\mathcal{L} = -\sqrt{-g}\left[R + \frac{1}{2}(\partial\phi)^2 + \frac{3}{2}(\partial\beta)^2\right] . \tag{6.34}$$

For a FRW metric (6.30) and time-dependent fields $\phi = \phi(t)$, $\beta = \beta(t)$ one finds the solutions

$$\begin{aligned} a &\sim |t|^{\frac{1}{3}} \\ \phi &= p_\phi \ln|t| + c_\phi \\ \beta &= p_\beta \ln|t| + c_\beta \, . \end{aligned} \tag{6.35}$$

The expansion powers c_ϕ and c_β satisfy

$$3p_\beta^2 + p_\phi^2 = \frac{4}{3} , \tag{6.36}$$

while c_ϕ and c_β are arbitrary constants. Recall that ϕ and β describe the evolution of the Calabi-Yau volume and the radius of the orbifold $S^1/\mathbf{Z}_2$, respectively.

6.3.2 Including Axions

Recall that axions ν^i usually arise as zero modes of the anti-symmetric tensor fields of M-theory and, at low energy, become the imaginary parts of the moduli multiplets $T^i = \exp(\beta^i) + i\nu^i$. Consider a single such axion ν coupled to a number of radius moduli β^i. The Lagrangian (6.29) is then modified to

$$\mathcal{L} = -\sqrt{-g}\left[R + \sum_i \partial\beta^i\partial\beta^i + \exp\left(\sum_i q_i\beta^i\right)\partial\nu\partial\nu\right] . \tag{6.37}$$

Here $\mathbf{q} = (q_i)$ is a constant vector which depends on the specific case under consideration. The general solution for this Lagrangian with time-dependent scalar fields and a FRW metric (6.30) can be found [41, 42] analytically. Here it will be sufficient to describe the qualitative behaviour of this solution. In each branch a solution to (6.37) has an early and a late asymptotic region where ν is approximately constant and the evolution of β^i is described by a rolling radii solution (6.32) with expansion power $\mathbf{p}_{(i)}$ for the initial asymptotic state and a generally different expansion power $\mathbf{p}_{(f)}$ for the final asymptotic state. The two asymptotic regions are connected by an intermediate phase where the axion ν moves by a finite distance in field space. It can be shown that the initial and final expansion powers are related by

$$\mathbf{p}_{(f)} = \mathbf{p}_{(i)} - \frac{2\mathbf{q}\cdot\mathbf{p}_{(i)}}{|\mathbf{q}|^2}\mathbf{q}\,. \tag{6.38}$$

Hence the axion "mediates" between two rolling radii solutions. The structure of curvature singularities is the same as for simple rolling radii solutions. Further, it can be seen that the non-trivial kinetic function $\exp(\sum_i q_i\beta^i)$ for ν in the Lagrangian (6.37) always increases asymptotically early and late.

6.3.3 Moving Branes

As we have mentioned before, in addition to bulk moduli there also generally exist moduli which descend from brane world volume theories, for example, scalar fields which indicate the position of a brane in the internal transverse space. In the context of heterotic M-theory we have already encountered such a modulus, namely the modulus Z which measures the position of a three-brane in the orbifold direction. We will now study the cosmological evolution of such brane moduli for the case of heterotic M-theory. However, many of our conclusions are quite general and apply to other models as well.

Recall, that in heterotic M-theory we have the generic chiral moduli multiplets S, T and Z with associated Kähler potential (6.26). In terms of the component fields ϕ, β and z defined in (6.25) and after truncating off the axions this leads to the Lagrangian [34]

$$\mathcal{L} = -\sqrt{-g}\left[R + \frac{1}{2}(\partial\phi)^2 + \frac{3}{2}(\partial\beta)^2 + |q_3|e^{\beta-\phi}(\partial z)^2\right]\,, \tag{6.39}$$

which is a generalization of (6.34) to include the position modulus z. We recall that the three scalar fields in this Lagrangian have a well-defined interpretation in terms of the underlying higher-dimensional geometry, as sketched in Fig. 6.8. Specifically, e^{ϕ} and e^{β} are proportional to the Calabi-Yau volume and the orbifold radius, respectively, while z measures the position of

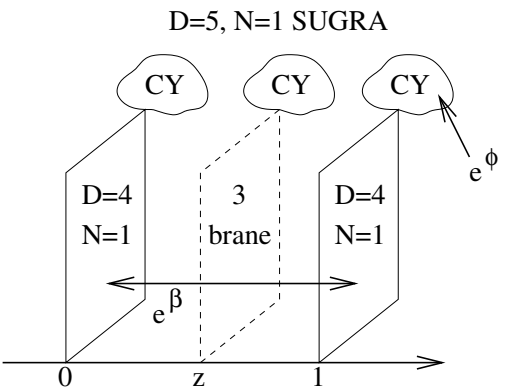

Fig. 6.8. Interpretation of moduli ϕ, β and z in heterotic M-theory.

the three-brane with $z = 0$ and $z = 1$ corresponding to the locations of the boundaries.

An elementary observation about the Lagrangian (6.39) is the existence of the non-trivial kinetic term for the brane position modulus z. This term implies that the bulk moduli ϕ and β necessarily vary in time whenever z does. As we will see this variation of the bulk moduli is actually important and has important implications for the qualitative features of the solutions. It would, therefore, be quite inappropriate, in this model as well as in other brane models, to study the evolution of the brane in a static background.

We note that the Lagrangian (6.39) is a special case of the Lagrangian (6.37) which couples radius moduli to an axion, with the position modulus z formally playing the role of the axion. Hence, we conclude that the cosmological solutions [43] to (6.39) have two asymptotic regions where the brane is approximately in rest and ϕ and β evolve as rolling radii with expansion coefficients satisfying (6.36). These two asymptotic early and late regions are connected by a period where the brane moves by a finite coordinate distance. In short, the brane evolution interpolates between two rolling radii solutions. Since the brane moves by a finite distance it may or may not collide with one of the boundaries depending on initial conditions. In the case of collision, the system may undergo a topology-changing small instanton transition [44, 45, 46] where the three-brane is "absorbed". Some dynamical aspects related to this transition have been studied in [47]. We also note that the heterotic strong coupling expansion parameter defined by $\epsilon_S \sim \exp(\beta - \phi)$ always increases asymptotically for those solutions. Hence the effective theory is always invalidated asymptotically, a fact one would have missed had one set the bulk moduli to constants.

6.3.4 Duality Symmetries and Cosmological Solutions

Dualities are global, discrete transformations on the M-theory moduli space which are thought to be exact symmetries of the theory. They can usually be found as symmetries of low-energy effective theories and, at perturbative level, they enlarge to continuous version of the symmetry. A typical example for a Kähler simple potential

$$K_{\text{mod}} = -k \ln(T + \bar{T}) \tag{6.40}$$

with a constant k would be given by the transformations

$$T \to \frac{aT + ib}{cT + id} , \tag{6.41}$$

where $ad - bc = 1$ which form the group $SL(2, R)$. The invariance of the action can be seen by noting that these transformation can be generated by $T \to aT + ib$ and $T \to 1/T$ where the former obviously leaves the above Kähler potential invariant (up to an irrelevant constant) and the latter corresponds to a Kähler transformation of K_{mod} which leaves the action unchanged.

Such continuous versions of duality symmetries can be quite useful to generate new cosmological solutions from old ones [37]. For example, the above transformations mix the real and imaginary part of T and can, hence, be used to generate solutions with a non-trivial axion (the imaginary part of T) from pure rolling radii solutions.

6.4 M-Theory and Inflation

The dynamics of scalar fields and inflation in particular constitutes probably the most important area in which one expects a close relation between cosmology and particle physics (or fundamental theories in general). Let us, therefore, discuss inflation in the context of M-theory in some detail.

6.4.1 Reminder Inflation

In this review, I will understand "inflation" as an era in the history of the universe where modes leave the Hubble horizon. The ratio of the size of a particular mode and the horizon size is proportional to $|H|a$ where $a = \exp\alpha$ is the scale factor in (6.30) and $H = \dot{a}/a$. Modes leave the Hubble horizon if the time-derivative of $|H|a$ is positive. For a power-law evolution with

$$a \sim |t|^p \tag{6.42}$$

and some constant power p this implies that

$$\frac{d}{dt}\left(|H|a\right) = \mathrm{sgn}(t)\ddot{a} \sim \mathrm{sgn}(t)p(p-1)|t|^{p-2} \tag{6.43}$$

must be positive for inflation. This leads to two different possibilities. First, in the positive-time branch, $t > 0$, we should have $p > 1$. This case is realized by potential-driven inflation [48] which can arise if the potential[5] $V = V(\phi)$ satisfies the slow-roll conditions

$$\epsilon \sim \left(\frac{V'}{V}\right)^2 \ll 1\ , \qquad \eta \sim \frac{V''}{V} \ll 1\ , \tag{6.44}$$

in any part of fields space. Here the prime denotes the derivative with respect to the (canonically normalized) field ϕ.

The second case, for which (6.43) is positive arises in the negative-time branch, $t < 0$, if $p \in [0, 1]$. This possibility is realized by many solutions in the negative-time branch and is also referred to as pre-big-bang inflation [49, 50]. Note, that in this case inflation evolves towards an apparent curvature singularity which arises within the low-energy effective theory.

[5] Here, ϕ is a generic scalar field.

All proposed mechanisms (known to me) to solve the problems of standard cosmology fall into one of these two categories. This includes more recent proposals which involve moving branes such as the one in [51].

We will now discuss these two different types of inflation separately in the context of M-theory.

6.4.2 Potential-Driven Inflation

M-theory low-energy effective theories typically contain a large number of scalar fields and inflation, if realized at all, is unlikely to arise in the form of one of the simple single-scalar field models which is commonly studied in inflationary model-building. The possibilities can be broadly classified into "bulk field inflation" where the inflaton is a combination of bulk scalar fields and "brane field inflation" where inflation arises from the evolution of scalar fields which originate from brane world-volume theories.

We start discussing bulk field inflation. The obvious problem in this case is to find a suitable M-theory potential for bulk moduli which satisfies the slow-roll conditions (6.44) at least in some part of moduli space. For multi-field potentials these conditions have to be imposed on all directions in field space which are not stabilized. In Sect. 6.2.4 we have discussed typical shapes for such potentials in the direction of radius moduli. The generic problem is that the potentials originating from superpotentials (6.15) are normally to steep to be compatible with slow-roll. Hence, for potentials with a shape as in Fig. 6.4 there is little hope for any relevant amount of inflationary expansion [52, 53], even though other directions in field space, such as some of the axion directions, may well be compatible with slow-roll. The situation is somewhat more promising with potentials such as in Fig. 6.5 which have a minimum with negative cosmological constant in the direction of the radius moduli. Some inflation will be possible starting from the maximum (although a sufficient amount is likely to require initial conditions fine-tuned to the maximum) but the final state with negative cosmological constant is unattractive given the observational evidence for a positive cosmological constant in the late universe. There are indications for the existence of minima with positive cosmological constant [27] such as in Fig. 6.6. In this case, the most promising scenario may be to interpret the minimum as the true late-time vacuum. Inflation may then be generated by fixing the radius moduli at the minimum and use an axion direction or a brane field [54] as the inflaton.

With these remarks on bulk field inflation in mind we now turn to brane field inflation. A recently much discussed example of this type is anti-brane inflation [55], where the inflaton is the position modulus of an anti-brane. Let us discuss this mechanism in the context of heterotic M-theory with an anti 5-brane. The effective four-dimensional theory for this situation is given by (6.39) with an additional potential

$$V \sim (|q_3| - q_3) e^{-\phi - 2\beta} \left[1 - \frac{2}{3} q_1 e^{\beta - \phi} z \right]^{-1} . \tag{6.45}$$

Recall that the anti-brane case corresponds to $q_3 < 0$ where this potential is indeed non-vanishing. Here, q_1 is proportional to the charge α_1 on the boundary at $y = 0$. This potential satisfies the slow-roll conditions in the z-direction as long as the strong coupling expansion parameter $\epsilon_S \sim e^{\beta-\phi}$ is much smaller than one, a condition that is anyway required for the validity of the effective theory. On would, therefore, hope for an inflationary period while the anti-brane slowly moves along the orbifold direction which ends when it collides with one of the boundaries. An obvious problem (which is quite generic for such anti-brane models) is that the potential (6.45) is by no means slow-roll in the direction of the radius moduli ϕ and β. Without any additional mechanism to stabilize the radius moduli no inflation will be generated, therefore. The obvious solution seems to be to combine a potential as in Fig. 6.6 which stabilizes the radius moduli with (6.45). Unfortunately, it turns out, the amount of inflation generated before the brane collides with the boundary is too small. Roughly, this happens because the kinetic function for z in (6.39) is identical to the parameter ϵ_S which controls the slow-roll properties of the potential (6.45). Again this problem arises in many types of anti-brane inflation models, although it can be avoided in some constructions [56].

There is no need to insist on the anti-brane position modulus as the inflaton. There will generically be many brane scalar fields, such as the scalar partners of matter fields, with polynomial superpotentials as in (6.10), which may be well-suited. These possibilities will now have to be examined for string vacua with moduli potentials as in Fig. 6.6 and all moduli fixed.

So far we have exclusively focused on inflation in the context of an effective four-dimensional theory. As mentioned before, the reason is that the existence of very large additional dimensions seem unlikely given the known mechanism of stabilizing moduli in M-theory. However, the energy scale during inflation can be quite large (up to $\sim 10^{16}$GeV) and, hence, inflation may be able to probe dimensions one or two orders of magnitude larger than the fundamental length scale $1/M_F$. Such a possiblity may exist within heterotic M-theory [33] where coupling unification suggest a "largish" orbfield size [57]. In this case, the analysis should be carried out in the context of five-dimensional heterotic M-theory reviewed in Sect. 6.2.5. For further discussions in this direction see [58].

6.4.3 Pre-Big-Bang Inflation

The basic idea of pre-big-bang cosmology [49, 50, 37] is to use solutions in the negative-time branch to generate inflation. We note that this mechanism by itself is not particularly "stringy". It simply relies on the existence of a time-reflection symmetry of General Relativity and can, hence, be used in any type of gravitational theory.

However, implementing the idea in a string context is rather straightforward. Consider the heterotic Lagrangian (6.34) for the moduli ϕ and β with

its cosmological solutions (6.35). Since the expansion power of the scale factor is always $1/3$ (and, hence, in the range $[0, 1]$) any of these solutions in the negative time branch can be used to generate pre-big-bang inflation. The original model [49] focused on the special case of constant β, that is, $p_\beta = 0$ which, from (6.36) implies a dilaton expansion power

$$p_\phi = \pm \frac{2}{\sqrt{3}} \, . \tag{6.46}$$

Of course, more complicated versions of the model with evolving β or with an additional moduli potential added to (6.34) are possible.

The obvious problem of this proposal is the future curvature singularity of negative-time branch solutions together with the necessity to generate a transition from the negative-time branch with contracting universe to standard cosmology in the positive-time branch. Clearly, positive- and negative-time branch solutions cannot be simply continuously matched together since they correspond to opposite signs of the Hubble parameter. Hence, additional physics not contained in simple effective Lagrangian such as (6.34) must be responsible for this hypothetical transition which has also been called "graceful exit" [59]. Let us briefly discuss why it is not easy to find a mechanism for a graceful exit. Consider, the second order equation for the scale factor

$$\dot{H} = -(\rho + p) \, , \tag{6.47}$$

where ρ and p are the total energy density and pressure for the model including contributions from whatever is supposed to generate the exit. For the required transition from a state with negative H to a state with positive H one needs stress energy which satisfies $\rho + p < 0$, at least for some period of time around the transition. However, most "decent" sources of stress energy such as scalar fields with potentials, gauge fields, forms etc. satisfy the null energy condition $\rho + p \geq 0$, simply from the algebraic form of their stress energy tensor.

A form of stress energy which potentially does violate the null energy condition are higher curvature terms, such as R^2 terms. It is well-known that such terms arise in string theory as higher-order α' corrections. Clearly, such terms will become important when the pre-big-bang phase approaches the curvature singularity. The hope has, therefore, been that string theory will somehow regulate the curvature singularity, thereby generating a graceful exit.

The question whether and how string theory resolves cosmological singularities has been addressed in different frameworks. Effective actions including higher-order α' corrections and loop corrections have been studied [60, 61]. Although, there are some indications that an exit can be achieved within this framework there is an obvious problem with keeping only the lowest order corrections of an infinite series when these lowest order corrections need to be large to generate the exit. It is hard to see how this problem can be overcome

in the context of an effective field theory approach. On a more fundamental level, there have been studies of exact string backgrounds in the context of Wess-Zumino-Witten models [62] and, more recently, a perturbative string analysis of time-dependent orbifolds [63]. It is fair to say that none of these approaches has led, so far, to conclusive results about the fate of cosmological singularities in M-theory.

There has been some renewed interest [51] in pre-big-bang cosmology in the context of negative-time branch solutions with a moving three-brane within heterotic M-theory, based on the Lagrangian (6.39) (possibly with an additional moduli potential). For example, one could use the negative-time branch of the solutions described in Sect. 6.3.3 as the basis for a moving-brane pre-big-bang cosmology. While this proposal is nothing new in terms of the mechanism for inflation it contains an interesting new aspect related to the exit problem. In the pre-big-bang phase, the three-brane can be arranged to collide with one of the boundaries well before the system runs into large curvature. The hope is that the brane-collision and the possible associated topology change ends the pre-big-bang phase and generates a graceful exit. This idea seems quite appealing in that it avoids the difficult problem of dealing with the large-curvature region. However, it is far from clear at present that such a brane-boundary collision can indeed lead to an exit from the pre-big-bang phase.

A final remark on the calculation of scalar density perturbations in pre-big-bang models is in order. As a general rule, while potential-driven inflation generically leads to a scale-invariant spectrum with spectral index n close to one, spectral indices in the context of pre-big-bang cosmology vary substantially and depend on the particular solution. For example, the spectral index for an axion field in a background of one of the solutions (6.35) depends on p_ϕ and p_β [37]. Frequently, n is substantially larger than one. Moreover, it is not completely clear how to match perturbations across the transition from pre- to post-big-bang cosmology, a fact which has been highlighted in the recent debate [64]. Presumably, the question of how this should be done can only be conclusively answered once a physical exit mechanism has been found.

6.5 Topology Change in Cosmology

As we have mentioned before, M-theory has a large number of topologically different compactifications. At present, it is unclear what, if anything, will choose one particular topology over the others. It may well be that the cosmological evolution has some role to play in this context. Fortunately, within M-theory the topology of a compactification can change in a number of well-defined processes. Hence, one can think of the M-theory moduli space as one entity (or, at least few entities) which consists of a web of different topologies connected by topology-changing processes. Indeed, such a picture seems to be a pre-requisite for any mechanism which would single out a particular topol-

ogy. One may hope that the closer analysis of topology-changing processes and their cosmological implications leads to some clues as to the nature of this mechanism. In this spirit we will now discuss the simplest topology-changing process, the M-theory flop transition [65], and its cosmology [66, 67].

6.5.1 M-Theory Flops

The subsequent analysis will be in the context of eleven-dimensional supergravity (M-theory) on Calabi-Yau three folds, leading to an effective $N = 1$ supergravity in five dimensions. We consider transitions between two topologically different Calabi-Yau spaces X and $\tilde{X}$. The Kähler moduli spaces of these Calabi-Yau three-folds are parametrized by moduli b^i and $\tilde{b}^i$, respectively. Their number is identical for the two topologies. One can think of these moduli as measuring the volumes of two-cycles within the Calabi-Yau spaces. Let us focus on a particular direction $b = \gamma_i b^i$ in moduli space, where γ_i are constant coefficient. We think of b as measuring the size of a particular two-sphere S^2 with the Calabi-Yau space. When this two-sphere collapses to zero size, $b \to 0$, the Calabi-Yau space X becomes singular. The two-sphere can then be blown up in a topologically different way to produce another Calabi-Yau space $\tilde{X}$ with the modulus b now extending to negative values (see Fig. 6.9). The important point is that this mathematical process, called a flop, can be smoothly realized as a physical transition within M-theory. The key for a smooth description of the transition is the inclusion of certain additional states q which have mass $m \sim b$ and, hence, become massless at the transition point. These states, which we will call "transition states", originate from a membrane which wraps the collapsing two-sphere.

Once these states are included the transition can, in fact, be described in terms of the low-energy effective five-dimensional supergravity. The relevant terms in the supergravity action are [66]

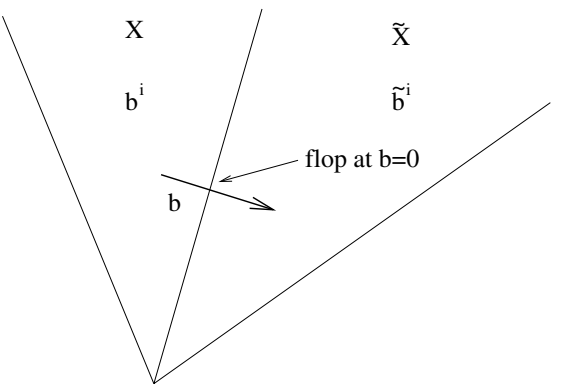

Fig. 6.9. Adjacent Kähler moduli spaces of Calabi-Yau spaces X and $\tilde{X}$ and flop transition.

$$S = -\frac{1}{2\kappa_5^2}\int d^5x \left[\frac{1}{2}R + \frac{1}{4}G_{ij}\partial b^i \partial b^j + (\partial q)^2 + V\right] , \qquad (6.48)$$

where the metric G_{ij} is defined by

$$G_{ij} = -\partial_i \partial_j \ln K , \qquad K = d_{ijk} b^i b^j b^k , \qquad (6.49)$$

where d_{ijk} are certain topological numbers (intersection numbers) of the Calabi-Yau space X [6]. They change under a flop transition to a new set of numbers $\tilde{d}_{ijk}$ for $\tilde{X}$ and replacing d_{ijk} by $\tilde{d}_{ijk}$ (and b^i with $\tilde{b}^i$) is, in fact, all that has to be done to the above action across the flop transition. The moduli fields b^i are subject to the constraint

$$K = 6 . \qquad (6.50)$$

Finally, the potential V is given by

$$V = \frac{g^2}{4}\left[b^2 q^2 + (G^{ij}\gamma_i\gamma_j - b^2)q^4\right] . \qquad (6.51)$$

Note that without transition states (setting $q = 0$ effectively) there would be no potential and the moduli b^i would be completely flat directions. The first term in the above potential precisely represents the mass term for the transition states.

6.5.2 Flops in Cosmology

We are now ready to consider a flop-transition in cosmology by studying time-dependent solutions of the action (6.48). We focus on a specific pair X and $\tilde{X}$ of Calabi-Yau three-folds related by a flop. The details of this example are described in [66]. It contains three moduli fields which we call (T, U, W) and the Kähler potential on X is given by

$$K = \frac{9}{4}U^3 + 3T^2U - W^3 , \qquad (6.52)$$

while on $\tilde{X}$ it changes (as a result of the change in intersection numbers) to

$$\tilde{K} = \frac{5}{4}U^3 + 3U^2W - 3UW^2 + 3T^2U . \qquad (6.53)$$

The moduli space of X is characterized by $U > W$ and the one for $\tilde{X}$ by $U < W$. The flop occurs precisely at $U = W$. Hence the field b which measures the size of the collapsing two-sphere is given by $b = U - W$, in this particular

[6] We are here using a canonical kinetic term for the transition states which we can think off as the lowest order expansion of a non-trivial sigma-model metric. Such a non-trivial metric has been included in [67].

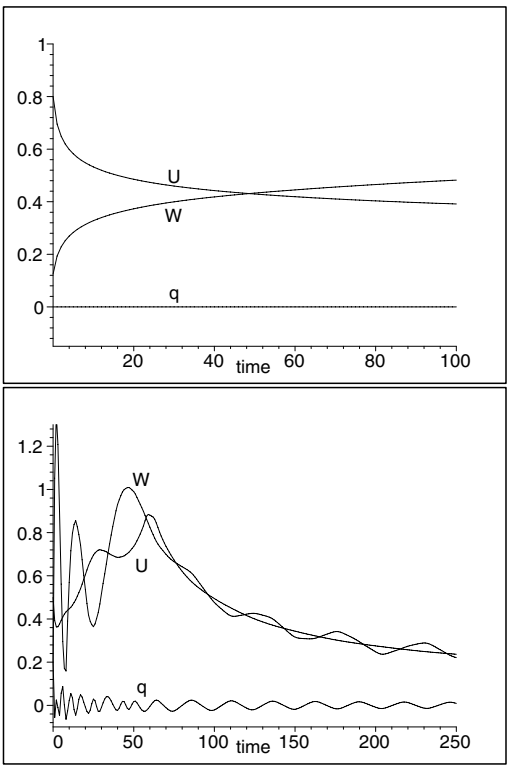

Fig. 6.10. Evolution for moduli U and W across a flop at $U = W$ for vanishing transition state q (upper plot) and non-vanishing transition state (lower plot).

example. By virtue of the constraint (6.50) only two of the three fields are independent and we take these to be U and W. The flop then occurs at the line $U = W$ in this two-dimensional moduli space.

It is easy to set up initial conditions for the system so it evolves towards the flop (basically by choosing the initial velocity $\dot{b}$ negative) and perform a numerical integration. The qualitative features of the result depend on whether the transition state q is set to zero (which can be done consistently in the action (6.48) by setting q and $\dot{q}$ to zero initially) or whether it is non-zero. Typical results for these two cases are shown in Fig. 6.10. These results are easy to interpret. For a vanishing transition state q the potential (6.51) is switched off and the moduli evolve freely. Starting out in X ($U > W$) they cross the transition region and evolve into the moduli space of $\tilde{X}$ at $U < W$. Hence, the topology indeed changes dynamically. However, when the transition state is non-zero the potential (6.51) becomes important and effects the evolution. From the lower plot we see that the system approaches the flop and then oscillates around the transition region. It does not evolve into the moduli space $\tilde{X}$ and, hence, the topology-changing transition is not

completed. This behaviour which is quite generic and has been confirmed in the analysis of [67] can be understood from the first term proportional to b^2q^2 in the potential (6.51). Once q is non-zero it effectively corresponds to a mass term for b and the system oscillates around the flop-line $b = 0$. It is unlikely to "find" the flat direction in b which arises for $q = 0$ and the topological transition remains incomplete. It is interesting that the dynamics of flops singles out a particular region in moduli space close to the transition region as the preferred state of the system. We also remark that the evolution of fields is always smooth even in the transition region and at no time is there a violation of the null energy condition. Hence, a transition from a negative- to a positive-time branch cannot be generated by a flop. It remains to be seen whether more severe topology-changing processes, such as the small-instanton transition of heterotic M-theory, are similar in this respect.

6.6 Conclusions

We have seen that obtaining "successful" early universe cosmology from M-theory is by no means straightforward. We have discussed the problems in implementing potential driven-inflation which arise from the typical properties of moduli potentials in string theory. For pre-big-bang cosmology the exit problem essentially remains an unresolved issue. Three key tasks in M-theory cosmology are suggested by the current state of affairs. First, obtaining a realistic and predictive model of potential-driven inflation within M-theory would be a major step forward. Some progress in this direction has been made through the systematic study of flux compactifications, as we have discussed. Secondly, the question of whether and how M-theory resolves cosmological curvature singularities should be answered. Thirdly, the relation between cosmological evolution, topology change and the choice of topology should be investigated. A solution of the latter two problems may well require a more satisfying, fundamental formulation of M-theory.

Acknowledgements

The author would like to thank the organizers of the 2nd Aegean Summer School for a very interesting and pleasant conference and acknowledges support by a PPARC Advanced Fellowship.

References

1. I. Antoniadis, "A Possible New Dimension At A Few TeV", Phys. Lett. B **246** (1990) 377.
2. I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, "New dimensions at a millimeter to a Fermi and superstrings at a TeV", Phys. Lett. B **436** (1998) 257, hep-ph/9804398.
3. C. M. Hull and P. K. Townsend, "Unity of superstring dualities", Nucl. Phys. B **438** (1995) 109, hep-th/9410167.
4. E. Witten, "String theory dynamics in various dimensions", Nucl. Phys. B **443** (1995) 85, hep-th/9503124.
5. P. K. Townsend, "Four lectures on M-theory", hep-th/9612121.
6. M. J. Duff, R. R. Khuri and J. X. Lu, "String solitons", Phys. Rept. **259** (1995) 213, hep-th/9412184.
7. M. B. Green, J. H. Schwarz and E. Witten, "Superstring Theory", vols. 1 and 2, Cambridge University Press, Cambridge (1986).
8. J. Polchinski, "String Theory", vols. 1 and 2, Cambridge University Press, Cambridge (1998).
9. J. Polchinski, "Lectures on D-branes", hep-th/9611050.
10. A. Sen, "An introduction to non-perturbative string theory", hep-th/9802051.
11. For a review see: W. Taylor, "M(atrix) theory: Matrix quantum mechanics as a fundamental theory", Rev. Mod. Phys. **73** (2001) 419, hep-th/0101126.
12. For a review see: B. R. Greene, "String theory on Calabi-Yau manifolds", hep-th/9702155.
13. G. Papadopoulos and P. K. Townsend, "Compactification of D = 11 supergravity on spaces of exceptional holonomy", Phys. Lett. B **357** (1995) 300, hep-th/9506150.
14. M. Dine, R. Rohm, N. Seiberg and E. Witten, "Gluino Condensation In Superstring Models", Phys. Lett. B **156** (1985) 55.
15. H. P. Nilles, "Supersymmetry, Supergravity And Particle Physics", Phys. Rept. **110** (1984) 1.
16. K. Dasgupta, G. Rajesh and S. Sethi, "M theory, orientifolds and G-flux", JHEP **9908** (1999) 023, hep-th/9908088.
17. S. B. Giddings, S. Kachru and J. Polchinski, "Hierarchies from fluxes in string compactifications", Phys. Rev. D **66** (2002) 106006, hep-th/0105097.
18. S. Kachru, M. B. Schulz and S. Trivedi, "Moduli stabilization from fluxes in a simple IIB orientifold", JHEP **0310** (2003) 007, hep-th/0201028.
19. R. Blumenhagen, L. Goerlich, B. Kors and D. Lust, "Noncommutative compactifications of type I strings on tori with magnetic background flux", JHEP **0010** (2000) 006, hep-th/0007024.
20. R. Blumenhagen, B. Kors and D. Lust, "Type I strings with F- and B-flux", JHEP **0102** (2001) 030, hep-th/0012156.
21. R. Blumenhagen, B. Kors, D. Lust and T. Ott, "The standard model from stable intersecting brane world orbifolds", Nucl. Phys. B **616** (2001) 3, hep-th/0107138.
22. L. E. Ibanez, F. Marchesano and R. Rabadan, "Getting just the standard model at intersecting branes", JHEP **0111** (2001) 002, hep-th/0105155.
23. J. Wess and J. Bagger, "Supersymmetry and Supergravity", Princeton University Press, Princeton (1982).

24. K. Becker, M. Becker and A. Strominger, "Five-branes, membranes and non-perturbative string theory", Nucl. Phys. B **456** (1995) 130, hep-th/9507158.
25. J. A. Harvey and G. W. Moore, "Superpotentials and membrane instantons", hep-th/9907026.
26. L. J. Dixon, "Supersymmetry Breaking In String Theory", SLAC-PUB-5229 *Invited talk given at 15th APS Div. of Particles and Fields General Mtg., Houston,TX, Jan 3-6, 1990.*
27. S. Kachru, R. Kallosh, A. Linde and S. P. Trivedi, "De Sitter vacua in string theory", Phys. Rev. D **68** (2003) 046005, hep-th/0301240.
28. P. Horava and E. Witten, "Eleven-Dimensional Supergravity on a Manifold with Boundary", Nucl. Phys. B **475** (1996) 94, hep-th/9603142.
29. A. Lukas, B. A. Ovrut, K. S. Stelle and D. Waldram, "Heterotic M-theory in five dimensions", Nucl. Phys. B **552** (1999) 246, hep-th/9806051.
30. A. Lukas, B. A. Ovrut, K. S. Stelle and D. Waldram, "The universe as a domain wall", Phys. Rev. D **59** (1999) 086001, hep-th/9803235.
31. R. Donagi, A. Lukas, B. A. Ovrut and D. Waldram, "Non-perturbative vacua and particle physics in M-theory", JHEP **9905** (1999) 018, hep-th/9811168.
32. R. Donagi, B. A. Ovrut, T. Pantev and D. Waldram, "Standard models from heterotic M-theory", Adv. Theor. Math. Phys. **5** (2002) 93, hep-th/9912208.
33. A. Lukas, B. A. Ovrut and D. Waldram, "Boundary inflation", Phys. Rev. D **61** (2000) 023506, hep-th/9902071.
34. M. Brandle and A. Lukas, "Five-branes in heterotic brane-world theories", Phys. Rev. D **65** (2002) 064024, hep-th/0109173.
35. J. P. Derendinger and R. Sauser, "A five-brane modulus in the effective N = 1 supergravity of M-theory", Nucl. Phys. B **598** (2001) 87, hep-th/0009054.
36. R. Rohm and E. Witten, "The Antisymmetric Tensor Field In Superstring Theory", Annals Phys. **170** (1986) 454.
37. J. E. Lidsey, D. Wands and E. J. Copeland, "Superstring cosmology," Phys. Rept. **337** (2000) 343, hep-th/9909061.
38. M. Mueller, "Rolling Radii And A Time Dependent Dilaton", Nucl. Phys. B **337** (1990) 37.
39. A. Lukas and S. Morris, "Rolling G(2) moduli", hep-th/0308195.
40. A. Lukas and S. Morris, "Moduli Kaehler potential for M-theory on a G(2) manifold", hep-th/0305078.
41. A. Lukas, B. A. Ovrut and D. Waldram, "Cosmological solutions of type II string theory", Phys. Lett. B **393** (1997) 65, hep-th/9608195.
42. A. Lukas, B. A. Ovrut and D. Waldram, "String and M-theory cosmological solutions with Ramond forms", Nucl. Phys. B **495** (1997) 365, hep-th/9610238.
43. E. J. Copeland, J. Gray and A. Lukas, "Moving five-branes in low-energy heterotic M-theory", Phys. Rev. D **64** (2001) 126003, hep-th/0106285.
44. E. Witten, "Small Instantons in String Theory", Nucl. Phys. B **460** (1996) 541, hep-th/9511030.
45. O. J. Ganor and A. Hanany, "Small E_8 Instantons and Tensionless Non-critical Strings", Nucl. Phys. B **474** (1996) 122, hep-th/9602120.
46. B. A. Ovrut, T. Pantev and J. Park, "Small instanton transitions in heterotic M-theory", JHEP **0005** (2000) 045, hep-th/0001133.
47. J. Gray, A. Lukas and G. I. Probert, "Gauge five brane dynamics and small instanton transitions in heterotic models", hep-th/0312111.

48. For a review see: D. H. Lyth and A. Riotto, "Particle physics models of inflation and the cosmological density perturbation", Phys. Rept. **314** (1999) 1, hep-ph/9807278.
49. M. Gasperini and G. Veneziano, "Pre - big bang in string cosmology", Astropart. Phys. **1** (1993) 317, hep-th/9211021.
50. M. Gasperini and G. Veneziano, "The pre-big bang scenario in string cosmology", Phys. Rept. **373** (2003) 1, hep-th/0207130.
51. J. Khoury, B. A. Ovrut, P. J. Steinhardt and N. Turok, "The ekpyrotic universe: Colliding branes and the origin of the hot big bang", Phys. Rev. D **64** (2001) 123522, hep-th/0103239.
52. R. Brustein and P. J. Steinhardt, "Challenges For Superstring Cosmology", Phys. Lett. B **302** (1993) 196, hep-th/9212049.
53. T. Barreiro, B. de Carlos and E. J. Copeland, "Stabilizing the dilaton in superstring cosmology", Phys. Rev. D **58** (1998) 083513, hep-th/9805005.
54. S. Kachru, R. Kallosh, A. Linde, J. Maldacena, L. McAllister and S. P. Trivedi, "Towards inflation in string theory", JCAP **0310** (2003) 013, hep-th/0308055.
55. G. R. Dvali and S. H. H. Tye, "Brane inflation", Phys. Lett. B **450** (1999) 72, hep-ph/9812483.
56. S. Buchan, B. Shlaer, H. Stoica and S. H. H. Tye, "Inter-brane interactions in compact spaces and brane inflation", hep-th/0311207.
57. E. Witten, "Strong Coupling Expansion Of Calabi-Yau Compactification", Nucl. Phys. B **471** (1996) 135, hep-th/9602070.
58. A. Lukas and D. Skinner, "Brane-world inflation and the transition to standard cosmology", JHEP **0109** (2001) 020, hep-th/0106190.
59. R. Brustein and G. Veneziano, "The Graceful exit problem in string cosmology", Phys. Lett. B **329** (1994) 429, hep-th/9403060.
60. R. Brustein and R. Madden, "A model of graceful exit in string cosmology", Phys. Rev. D **57** (1998) 712, hep-th/9708046.
61. C. Cartier, E. J. Copeland and R. Madden, "The graceful exit in string cosmology", JHEP **0001** (2000) 035, hep-th/9910169.
62. S. Elitzur, A. Giveon, D. Kutasov and E. Rabinovici, "From big bang to big crunch and beyond", JHEP **0206** (2002) 017, hep-th/0204189.
63. H. Liu, G. Moore and N. Seiberg, "Strings in a time-dependent orbifold", JHEP **0206** (2002) 045, hep-th/0204168.
64. C. Cartier, R. Durrer and E. J. Copeland, "Cosmological perturbations and the transition from contraction to expansion", Phys. Rev. D **67** (2003) 103517, hep-th/0301198.
65. E. Witten, "Phase Transitions In M-Theory And F-Theory", Nucl. Phys. B **471** (1996) 195, hep-th/9603150.
66. M. Brandle and A. Lukas, "Flop transitions in M-theory cosmology", Phys. Rev. D **68** (2003) 024030, hep-th/0212263.
67. L. Jarv, T. Mohaupt and F. Saueressig, "M-theory cosmologies from singular Calabi-Yau compactifications", hep-th/0310174.

7 Brane-World Cosmology

Roy Maartens

Institute of Cosmology and Gravitation, University of Portsmouth,
Portsmouth PO1 2EG, UK

Abstract. According to recent ideas from particle physics, the universe could be a higher-dimensional spacetime, with our observable part of the universe being a four-dimensional "brane" surface. In this picture, Standard Model particles and fields are confined to the brane while gravity propagates freely in all dimensions. If one of the extra spatial dimensions is very large relative to the Planck scale, this lowers the fundamental gravity scale, possibly even down to electroweak scales ($\sim$TeV). At low energies, gravity is localized at the brane and general relativity is recovered, but at high energies gravity "leaks" off the brane, leading to significant corrections to general relativistic dynamics and perturbations. Here I review the geometry, dynamics and perturbations of simple brane-world models for cosmology.

7.1 Introduction

Einstein's theory of general relativity breaks down at high enough energies, where quantum gravity theory takes over. The fully quantum regime entails the breakdown of the space-time continuum, but even when spacetime can be modelled as a continuum, significant corrections to general relativity will arise at energies below, but near, the fundamental scale. Traditionally, the fundamental scale has been thought to be the Planck scale, $M_{\rm p} \sim 10^{19}$ GeV. However, recent developments in M theory, a leading candidate quantum gravity theory [1], indicate that the $M_{\rm p}$ may be an effective energy scale, with the true fundamental scale being lower.

A fundamental aspect of string theory and M theory is the need for extra spatial dimensions. If there are d extra (spatial) dimensions, then the gravitational potential is $V(r) \propto 1/r^{1+d}$. If the length scale of the extra dimensions is L, then on scales $r \lesssim L$, the potential is $4+d$-dimensional, while on scales large relative to L, where the extra dimensions do not contribute to variations in the potential, V behaves like a 4-dimensional potential, *i.e.* $r \sim L$ in the d extra dimensions, and $V \sim L^{-d} r^{-1}$. This means that the usual Planck scale becomes an effective coupling constant, describing gravity on scales much larger than the extra dimensions, and related to the fundamental scale via the volume of the extra dimensions:

$$M_{\rm p}^2 \sim M_{4+d}^{2+d}\, L^d\,. \tag{7.1}$$

R. Maartens, Brane-World Cosmology, Lect. Notes Phys. **653**, 213–252 (2005)
http://www.springerlink.com/

If the extra-dimensional volume is Planck scale, *i.e.* $L \sim M_{\rm p}^{-1}$, then $M_{4+d} \sim M_{\rm p}$. But if the extra-dimensional volume is significantly above Planck scale, then the true fundamental scale M_{4+d} can be much less. In this case, we understand the weakness of gravity as due to the fact that it "spreads" into extra dimensions and only a part of it is felt in four dimensions. Experiments in colliders and table-top tests of gravitational force [2, 3] imply the bounds $L \lesssim 0.1$ mm and $M_{4+d} \gtrsim 1$ TeV.

String theory thus incorporates the possibility that the fundamental scale is much less than the Planck scale felt in four dimensions. There are five distinct (1+9)-dimensional superstring theories, all giving quantum theories of gravity. Discoveries in the mid-90's of duality transformations that relate these superstring theories and the (1+10)-dimensional supergravity theory, led to the conjecture that all of these theories arise as different limits of a single theory, which has come to be known as M theory. The eleventh dimension in M theory is related to the string coupling strength; the size of this dimension grows as the coupling becomes strong. At low energies, M theory can be approximated by (1+10)-dimensional supergravity.

It was also discovered that p-branes, which are extended objects of higher dimension than strings (one-branes), play a fundamental role in the theory. In the weak coupling limit, p-branes ($p > 1$) become infinitely heavy, so that they do not appear in the perturbative theory. Of particular importance among p-branes are the D-branes, on which open strings can end. Roughly speaking, open strings, which describe the non-gravitational sector, are attached at their endpoints to branes, while the closed strings of the gravitational sector can move freely in the full spacetime (the "bulk"). Classically, this is realised via the localization of matter and radiation fields on the brane, with gravity propagating in the bulk (see Fig. 7.1).

In the Horava-Witten solution [4], gauge fields of the standard model are confined on two (1+9)-branes (or domain walls) located at the end points of an S^1/Z_2 orbifold, *i.e.* a circle folded on itself across a diameter. The six extra dimensions on the branes are compactified on a very small scale, close to the fundamental scale, and their effect on the dynamics is felt through "moduli" fields, i.e. 5D scalar fields. A 5D realization of the Horava-Witten theory and the corresponding brane-world cosmology is given in [5].

These solutions can be thought of as effectively five-dimensional, with an extra dimension that can be large relative to the fundamental scale. They provide the basis for the Randall-Sundrum type I models of five-dimensional gravity [6]. The single-brane Randall-Sundrum type II models [7] with infinite extra dimension arise when the orbifold radius tends to infinity. The RS models are not the only phenomenological realizations of M theory ideas. They were preceded by the Arkani-Hamed-Dimopoulos-Dvali (ADD) [8] brane-world models, which put forward the idea that a large volume for the compact extra dimensions would lower the fundamental Planck scale, and thus address the long-standing "hierarchy" problem, *i.e.* why there is such a large gap between the electroweak and Planck scales.

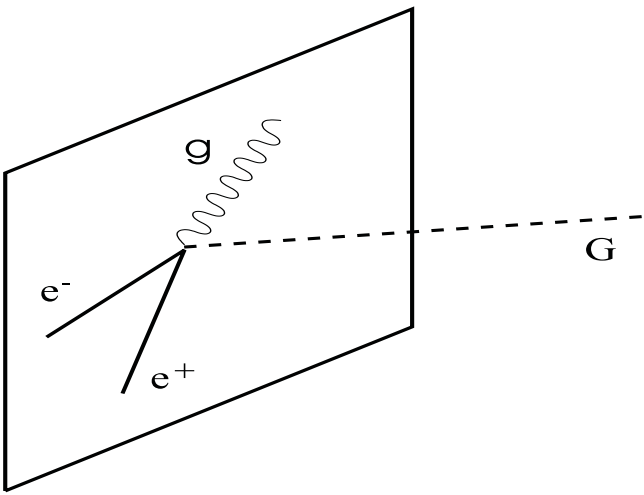

Fig. 7.1. Schematic of confinement of matter to the brane, while gravity propagates in the bulk (From [2]).

In the ADD models, more than one extra dimension is required for agreement with experiments, and there is "democracy" amongst the equivalent extra dimensions, which are furthermore flat. By contrast, the RS models have a "preferred" extra dimension, with other extra dimensions treated as ignorable (i.e., stabilized except at energies near the fundamental scale). Furthermore, this extra dimension is curved or "warped" rather than flat: the bulk is a portion of anti de Sitter (AdS_5) spacetime. As in the Horava-Witten solutions, the RS branes are Z_2-symmetric (mirror symmetry), and have a tension, which serves to counter the influence of the negative bulk cosmological constant on the brane. This also means that the self-gravity of the branes is incorporated in the RS models. The novel feature of the RS models compared to previous higher-dimensional models is that the observable three dimensions are protected from the large extra dimension (at low energies) by curvature rather than straightforward compactification.

The RS brane-worlds and their generalizations (to include matter on the brane, scalar fields in the bulk, etc.) provide phenomenological models that reflect at least some of the features of M theory, and that bring exciting new geometric and particle physics ideas into play. The RSII models also provide a framework for exploring holographic ideas that have emerged in M theory. This review focuses mainly on RSII-type brane-worlds (see also [9, 10, 11]).

The dilution of gravity via extra dimensions weakens gravity, and it also extends the range of graviton modes felt on the brane, beyond the massless mode of four-dimensional gravity. For simplicity, consider a flat brane with one flat extra dimension, compactified through the identification $y \leftrightarrow y +$

$2\pi nL$, where $n = 0, 1, 2, \cdots$. The perturbative 5D graviton amplitude can be Fourier expanded as

$$f(x^a, y) = \sum_n e^{iny/L} f_n(x^a)\,, \tag{7.2}$$

where f_n are the amplitudes of the KK modes, i.e. the effective 4D modes of the the 5D graviton. To see that these KK modes are massive from the brane viewpoint, we start from the 5D wave equation that the massless 5D field f satisfies (in a suitable gauge):

$$^{(5)}\Box f = 0 \;\Rightarrow\; \Box f + \partial_y^2 f = 0\,. \tag{7.3}$$

It follows that the KK modes satisfy a 4D Klein-Gordon equation with an effective 4D mass, m_n,

$$\Box f_n = m_n^2 f_n\,, \quad m_n = \frac{n}{L}\,. \tag{7.4}$$

The massless mode, f_0, is the usual 4D graviton mode. But there is a tower of massive modes, $L^{-1}, 2L^{-1}, \cdots$, which imprint the effect of the 5D gravitational field on the 4D brane. Compactness of the extra dimension leads to discreteness of the spectrum. For an infinite extra dimension, $L \to \infty$, the separation between the modes disappears and the tower forms a continuous spectrum.

Extra dimensions lead to new scalar and vector degrees of freedom on the brane. In 5D, the spin-2 graviton is represented by a metric perturbation $^{(5)}h_{AB}$ that is transverse traceless:

$$^{(5)}h^A{}_A = 0 = \partial_B\, {}^{(5)}h_A{}^B\,. \tag{7.5}$$

In a suitable gauge, $^{(5)}h_{AB}$ contains a 3D transverse traceless perturbation h_{ij}, a 3D transverse vector perturbation Σ_i and a scalar perturbation β, each of which satisfies the 5D wave equation (7.3). The 5 degrees of freedom (polarizations) in the 5D graviton are thus split into:
a 4D spin-2 graviton h_{ij} (2 polarizations); a 4D spin-1 gravi-vector (gravi-photon) Σ_i (2 polarizations); a 4D spin-0 gravi-scalar β. The standard 4D graviton corresponds to the massless zero-mode of h_{ij}.

In the general case of d extra dimensions, the number of degrees of freedom in the graviton follows from the irreducible tensor representations of the isometry group as $\frac{1}{2}(d+1)(d+4)$.

7.2 Randall-Sundrum Brane-Worlds

RS brane-worlds do not use compactification to localize gravity at the brane, but the curvature of the bulk. Gravity at low energies is prevented from

leaking into the extra dimension by the gravitational effect of a negative bulk cosmological constant,

$$\Lambda_5 = -\frac{6}{\ell^2} = -6\mu^2, \tag{7.6}$$

where ℓ is the curvature radius of AdS_5:

$${}^{(5)}R_{ABCD} = -\frac{1}{\ell^2}\left[{}^{(5)}g_{AC}\,{}^{(5)}g_{BD} - {}^{(5)}g_{AD}\,{}^{(5)}g_{BC}\right], \tag{7.7}$$

and μ is the corresponding energy scale. In Gaussian normal coordinates $X^A = (x^\mu, y)$, based on the brane at $y = 0$, the AdS_5 metric takes the form

$${}^{(5)}ds^2 = e^{-2|y|/\ell}\eta_{\mu\nu}dx^\mu dx^\nu + dy^2\,, \tag{7.8}$$

with $\eta_{\mu\nu}$ the Minkowski metric. The exponential warp factor reflects the confining role of the bulk cosmological constant. The Z_2-symmetry about the brane at $y = 0$ is reflected via the $|y|$ term. In the bulk, this metric is a solution of the 5D Einstein equations,

$${}^{(5)}G_{AB} = -\Lambda_5\,{}^{(5)}g_{AB}\,. \tag{7.9}$$

The brane is a flat Minkowski spacetime, with self-gravity in the form of brane tension. One can also use Poincare coordinates, which bring the metric into manifestly conformally flat form:

$${}^{(5)}ds^2 = \frac{\ell^2}{z^2}\left[\eta_{\mu\nu}dx^\mu dx^\nu + dz^2\right]\,,\ z = \ell e^{y/\ell}\,. \tag{7.10}$$

The two RS models are distinguished as follows:

– **RSI:**
There are two branes in RSI [6], at $y = 0$ and $y = L$, with Z_2-symmetry about each. The branes have equal and opposite tensions, $\pm\lambda$, where

$$\lambda = \frac{3M_\mathrm{p}^2}{4\pi\ell^2}\,. \tag{7.11}$$

The positive-tension "TeV" (or "hidden") brane has fundamental scale $M_5 \sim 1$ TeV. Standard Model fields are confined on the negative tension "Planck" (or "visible") brane. Because of the exponential warping factor, the effective scale on the visible brane at $y = L$ is M_p, where

$$M_\mathrm{p}^2 = M_5^3\,\ell\left[1 - e^{-2L/\ell}\right]\,. \tag{7.12}$$

Thus RSI gives a new solution to the hierarchy problem. Because of the finite separation between the branes, the KK spectrum is discrete. In order to recover 4D general relativity at low energies, a mechanism is required to stabilize the inter-brane distance – a scalar field degree of freedom known as the radion [12].

– **RSII:**
In RSII [7], there is only one, positive tension, brane; the negative tension brane is removed to infinity, $L \to \infty$. Then the energy scales are related via

$$M_5^3 = \frac{M_{\rm p}^2}{\ell}\,. \tag{7.13}$$

The infinite extra dimension makes a finite contribution to the 5D volume because of the warp factor:

$$\int d^5X\sqrt{-{}^5g} = 2\int d^4x \int_0^\infty dy e^{-4y/\ell} = \frac{\ell}{2}\int d^4x\,. \tag{7.14}$$

Thus the effective size of the extra dimension probed by the 5D graviton is ℓ. The RSII models are the most simple and geometrically appealing form of brane-world model, while at the same time providing a framework for cosmology and for the AdS/CFT correspondence.

On the RSII brane, the negative Λ_5 is balanced by the positive brane tension λ. The fine-tuning in (7.11) ensures that there is zero effective cosmological constant on the brane, so that the brane has the induced geometry of Minkowski spacetime. To see how gravity is localized at low energies, consider the 5D graviton perturbations of the metric in RS gauge [7, 13]:

$${}^{(5)}g_{AB} \to {}^{(5)}g_{AB} + e^{-2|y|/\ell}\,{}^{(5)}h_{AB}\,, \quad {}^{(5)}h_{Ay} = 0 = {}^{(5)}h^{\mu}{}_{\mu} = {}^{(5)}h^{\mu\nu}{}_{,\nu}\,. \tag{7.15}$$

(See Fig. 7.2.) The five polarizations of the 5D graviton are contained in the five independent components of $h_{\mu\nu}$ in the RS gauge, which has no remaining gauge freedom.

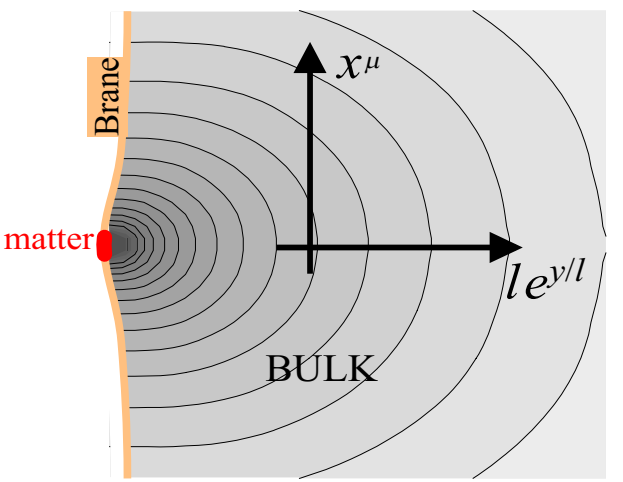

Fig. 7.2. The gravitational field of a small point mass on the brane in RS gauge (From [13]).

We split the amplitude f of ${}^{(5)}h_{AB}$ into 3D Fourier modes, and the linearized 5D Einstein equations lead to the wave equation ($y > 0$)

$$e^{2y/\ell}\left[\ddot{f} + k^2 f\right] = f'' - \frac{4}{\ell} f' \,. \tag{7.16}$$

Writing

$$f(t,y) = \sum_m \varphi_m(t)\, f_m(y)\,, \tag{7.17}$$

the wave equation reduces to

$$\ddot{\varphi}_m + (m^2 + k^2)\varphi_m = 0 \tag{7.18}$$

$$f_m'' - \frac{4}{\ell} f_m' + e^{2y/\ell} f_m = 0\,. \tag{7.19}$$

The zero mode solution is

$$\varphi_0(t) = A_{0+} e^{+ikt} + A_{0-} e^{-ikt} \tag{7.20}$$

$$f_0(y) = B_0 + C_0 e^{4y/\ell}\,, \tag{7.21}$$

and the $m > 0$ solutions are

$$\varphi_m(t) = A_{m+} \exp(+i\sqrt{m^2 + k^2}\, t) + A_{m-} \exp(-i\sqrt{m^2 + k^2}\, t) \tag{7.22}$$

$$f_m(y) = e^{2y/\ell}\left[B_m J_2\left(m\ell e^{y/\ell}\right) + C_m Y_2\left(m\ell e^{y/\ell}\right)\right]\,. \tag{7.23}$$

The boundary condition follows from the junction conditions, see (7.41) below, and leads to $f'(t,0) = 0$ (since the transverse traceless part of the perturbed energy-momentum tensor on the brane vanishes). Thus

$$C_0 = 0\,, \quad C_m = -\frac{J_1(m\ell)}{Y_1(m\ell)}\, B_m\,. \tag{7.24}$$

The zero mode contribution to the gravitational potential $V = \frac{1}{2}\,{}^{(5)}h_{00}$ gives the 4D result, $V \propto r^{-1}$. The contribution of the massive KK modes sums to a correction of the 4D potential. For $r \ll \ell$, one obtains

$$V(r) \approx \frac{GM\ell}{r^2}\,, \tag{7.25}$$

which simply reflects the fact that the potential becomes truly 5D on small scales. For $r \gg \ell$,

$$V(r) \approx \frac{GM}{r}\left(1 + \frac{2\ell^2}{3r^2}\right)\,, \tag{7.26}$$

which gives the small correction to 4D gravity at low energies from extra-dimensional effects.

Table-top tests of Newton's laws give $\ell \lesssim 0.1$ mm in (7.26). Then by (7.11) and (7.13), this leads to lower limits on the RS brane tension and the RS2 fundamental scale:

$$\lambda > (1\ \text{TeV})^4\,,\ M_5 > 10^5\ \text{TeV}\,. \tag{7.27}$$

7.3 Covariant Generalization of RS Brane-Worlds

The RS models and the subsequent generalization from a Minkowski brane to a Friedmann-Robertson-Walker (FRW) brane [14, 15, 16], were derived as solutions in particular coordinates of the 5D Einstein equations, together with the junction conditions at the Z_2-symmetric brane. The covariant Shiromizu-Maeda-Sasaki approach [17], is independent of coordinates and applies for general brane and bulk metrics. The basic idea is to use the Gauss-Codazzi equations to project the 5D curvature along the brane.

The 5D field equations determine the 5D curvature tensor; in the bulk, they are

$$^{(5)}G_{AB} = -\Lambda_5\, {}^{(5)}g_{AB} + \kappa_5^2\, {}^{(5)}T_{AB}\,, \tag{7.28}$$

where $^{(5)}T_{AB}$ represents any 5D energy-momentum of the gravitational sector (e.g., dilaton-type scalar fields, form fields).

If y is a Gaussian normal coordinate orthogonal to the brane (at $y=0$), then $n_A dX^A = dy$, with n^A the unit normal. The 5D metric in terms of the induced metric on $\{y = \text{ const}\}$ surfaces is

$$^{(5)}g_{AB} = g_{AB} + n_A n_B\,,\ {}^{(5)}ds^2 = g_{\mu\nu}(x^\alpha, y)dx^\mu dx^\nu + dy^2\,. \tag{7.29}$$

The extrinsic curvature of $\{y = \text{ const.}\}$ surfaces is

$$K_{AB} = g_A{}^C\ {}^{(5)}\nabla_C n_B\,,\ K_{[AB]} = 0 = K_{AB}n^B\,, \tag{7.30}$$

where square brackets denote anti-symmetrization. The Gauss equation gives the 4D curvature tensor as

$$R_{ABCD} = {}^{(5)}R_{EFGH}\, g_A{}^E g_B{}^F g_C{}^G g_D{}^H + 2K_{A[C}K_{D]B}\,, \tag{7.31}$$

and the Codazzi equation determines the change of K_{AB} along $\{y = \text{ const}\}$:

$$\nabla_B K^B{}_A - \nabla_A K = {}^{(5)}R_{BC}\, g_A{}^B n^C\,, \tag{7.32}$$

where $K = K^A{}_A$.

7.3.1 Field Equations on the Brane

Using (7.28) and (7.31), it follows that

$$\begin{aligned} G_{\mu\nu} = -\frac{1}{2}\Lambda_5 g_{\mu\nu} + \frac{2}{3}\kappa_5^2\left[{}^{(5)}T_{AB}g_\mu{}^A g_\nu{}^B + \left\{{}^{(5)}T_{AB}n^A n^B - \frac{1}{4}\,{}^{(5)}T\right\}g_{\mu\nu}\right] \\ + KK_{\mu\nu} - K_\mu{}^\alpha K_{\alpha\nu} + \frac{1}{2}\left[K^{\alpha\beta}K_{\alpha\beta} - K^2\right]g_{\mu\nu} - \mathcal{E}_{\mu\nu}\,, \end{aligned} \tag{7.33}$$

where $^{(5)}T = {}^{(5)}T^A{}_A$ and

$$\mathcal{E}_{\mu\nu} = {}^{(5)}C_{ACBD}\, n^C n^D g_\mu{}^A g_\nu{}^B\,, \tag{7.34}$$

is the projection of the bulk Weyl tensor, with

$$\mathcal{E}_{AB}n^B = 0 = \mathcal{E}_{[AB]} = \mathcal{E}_A{}^A \,. \tag{7.35}$$

Evaluating (7.33) as $y \to \pm 0$, gives the field equations on the brane.

First, we need to determine $K_{\mu\nu}$ at the brane from the junction conditions. The total energy-momentum tensor on the brane is

$$T^{\rm brane}_{\mu\nu} = T_{\mu\nu} - \lambda g_{\mu\nu} \,, \tag{7.36}$$

where $T_{\mu\nu}$ is the energy-momentum tensor of particles and fields confined to the brane (so that $T_{AB}n^B = 0$). The 5D field equations, including explicitly the contribution of the brane, are then

$${}^{(5)}G_{AB} = -\Lambda_5 \, {}^{(5)}g_{AB} + \kappa_5^2 \left[{}^{(5)}T_{AB} + T^{\rm brane}_{AB}\delta(y) \right] . \tag{7.37}$$

Integrating along the extra dimension and taking the limit, this leads to the junction conditions at the brane,

$$g^+_{\mu\nu} - g^-_{\mu\nu} = 0 \tag{7.38}$$

$$K^+_{\mu\nu} - K^-_{\mu\nu} = -\kappa_5^2 \left[T^{\rm brane}_{\mu\nu} - \frac{1}{3} T^{\rm brane} g_{\mu\nu} \right] , \tag{7.39}$$

where $T^{\rm brane} = g^{\mu\nu}T^{\rm brane}_{\mu\nu}$. The Z_2 symmetry means that when you approach the brane from one side and go through it, you emerge into a bulk that looks the same, but with the normal reversed, $n^A \to -n^A$. Then by (7.30)

$$K^-_{\mu\nu} = -K^+_{\mu\nu} \,, \tag{7.40}$$

so that we can use the junction condition (7.39) to determine the extrinsic curvature on the brane:

$$K_{\mu\nu} = -\frac{1}{2}\kappa_5^2 \left[T_{\mu\nu} + \frac{1}{3}\left(\lambda - T\right) g_{\mu\nu} \right] , \tag{7.41}$$

where $T = T^\mu{}_\mu$, we have dropped the $(+)$ and we evaluate quantities on the brane by taking the limit $y \to +0$.

Finally we arrive at the induced field equations on the brane:

$$G_{\mu\nu} = -\Lambda g_{\mu\nu} + \kappa^2 T_{\mu\nu} + 6\frac{\kappa^2}{\lambda}\mathcal{S}_{\mu\nu} - \mathcal{E}_{\mu\nu} + 4\frac{\kappa^2}{\lambda}\mathcal{F}_{\mu\nu} \,, \tag{7.42}$$

where

$$\kappa^2 \equiv \kappa_4^2 = \frac{1}{6}\lambda\kappa_5^4 \tag{7.43}$$

$$\Lambda \equiv \Lambda_4 = \frac{1}{2}\left[\Lambda_5 + \kappa^2\lambda\right] \tag{7.44}$$

$$\mathcal{S}_{\mu\nu} = \frac{1}{12}TT_{\mu\nu} - \frac{1}{4}T_{\mu\alpha}T^\alpha{}_\nu + \frac{1}{24}g_{\mu\nu}\left[3T_{\alpha\beta}T^{\alpha\beta} - T^2\right] \tag{7.45}$$

$$\mathcal{F}_{\mu\nu} = {}^{(5)}T_{AB}g_\mu{}^A g_\nu{}^B + \left[{}^{(5)}T_{AB}n^A n^B - \frac{1}{4}\,{}^{(5)}T \right] g_{\mu\nu} \,. \tag{7.46}$$

What about the conservation equations? Using (7.28), (7.32) and (7.41), one obtains

$$\nabla^\nu T_{\mu\nu} = -2\,{}^{(5)}T_{AB}n^A g^B{}_\mu\,. \tag{7.47}$$

Thus in general there is exchange of energy-momentum between the bulk and the brane. From now on, we will assume that ${}^{(5)}T_{AB} = 0 \;\Rightarrow\; \mathcal{F}_{\mu\nu} = 0$, so that

$${}^{(5)}G_{AB} = -\Lambda_5\,{}^{(5)}g_{AB}\,, \text{ (bulk)} \tag{7.48}$$

$$G_{\mu\nu} = -\Lambda g_{\mu\nu} + \kappa^2 T_{\mu\nu} + 6\frac{\kappa^2}{\lambda}\mathcal{S}_{\mu\nu} - \mathcal{E}_{\mu\nu}\,, \text{ (brane)} \tag{7.49}$$

and one then recovers from (7.47) the standard 4D conservation equations,

$$\nabla^\nu T_{\mu\nu} = 0\,. \tag{7.50}$$

Then there is no exchange of energy-momentum between the bulk and the brane; their interaction is purely gravitational. The 4D contracted Bianchi identities ($\nabla^\nu G_{\mu\nu} = 0$), applied to (7.42), lead to

$$\nabla^\mu \mathcal{E}_{\mu\nu} = \frac{6\kappa^2}{\lambda}\,\nabla^\mu \mathcal{S}_{\mu\nu}\,, \tag{7.51}$$

which shows qualitatively how (1+3) spacetime variations in the matter-radiation on the brane can source KK modes.

The induced field equations (7.42) show two key modifications to the standard 4D Einstein field equations arising from extra-dimensional effects:

– $\mathcal{S}_{\mu\nu} \sim (T_{\mu\nu})^2$ is the high energy correction term, which is negligible for $\rho \ll \lambda$, but dominant for $\rho \gg \lambda$:

$$\frac{|\kappa^2\mathcal{S}_{\mu\nu}/\lambda|}{|\kappa^2 T_{\mu\nu}|} \sim \frac{|T_{\mu\nu}|}{\lambda} \sim \frac{\rho}{\lambda}\,. \tag{7.52}$$

– $\mathcal{E}_{\mu\nu}$, the projection of the bulk Weyl tensor on the brane, encodes corrections from 5D graviton effects (the KK modes in the linearized case).

From the brane-observer viewpoint, the energy-momentum corrections in $\mathcal{S}_{\mu\nu}$ are local, whereas the KK corrections in $\mathcal{E}_{\mu\nu}$ are nonlocal, since they incorporate 5D gravity wave modes. These nonlocal corrections cannot be determined purely from data on the brane. In the perturbative analysis of RSII which leads to the corrections in the gravitational potential, (7.26), the KK modes that generate this correction are responsible for a nonzero $\mathcal{E}_{\mu\nu}$; this term is what carries the modification to the weak-field field equations. The nine independent components in the trace-free $\mathcal{E}_{\mu\nu}$ are reduced to five degrees of freedom by (7.51); these arise from the five polarizations of the 5D graviton.

Note that the covariant formalism applies also to the two-brane case. In that case, the gravitational influence of the second brane is felt via its contribution to $\mathcal{E}_{\mu\nu}$.

7.3.2 The Brane Observer's Viewpoint

The effects of bulk gravity are conveyed, from a brane observer viewpoint, via the local ($\mathcal{S}_{\mu\nu}$) and nonlocal ($\mathcal{E}_{\mu\nu}$) corrections to Einstein's equations. (In the more general case, bulk effects on the brane are also carried by $\mathcal{F}_{\mu\nu}$, which describes any 5D fields.) The $\mathcal{E}_{\mu\nu}$ term cannot in general be determined from data on the brane, and the 5D equations above (or their equivalent) need to be solved in order to find $\mathcal{E}_{\mu\nu}$.

The general form of the brane energy-momentum tensor for any matter fields (scalar fields, perfect fluids, kinetic gases, dissipative fluids, etc.), including a combination of different fields, can be covariantly given in terms of a chosen four-velocity u^μ as

$$T_{\mu\nu} = \rho u_\mu u_\nu + p h_{\mu\nu} + \pi_{\mu\nu} + q_\mu u_\nu + q_\nu u_\mu \,. \tag{7.53}$$

Here ρ and p are the energy density and isotropic pressure, and

$$h_{\mu\nu} = g_{\mu\nu} + u_\mu u_\nu \,, \tag{7.54}$$

projects into the comoving rest space orthogonal to u^μ on the brane. The momentum density obeys $q_\mu = q_{\langle\mu\rangle}$, and the anisotropic stress obeys $\pi_{\mu\nu} = \pi_{\langle\mu\nu\rangle}$, where angled brackets denote the spatially projected, symmetric and trace-free part:

$$V_{\langle\mu\rangle} = h_\mu{}^\nu V_\nu \,, \quad W_{\langle\mu\nu\rangle} = \left[h_{(\mu}{}^\alpha h_{\nu)}{}^\beta - \frac{1}{3} h^{\alpha\beta} h_{\mu\nu} \right] W_{\alpha\beta} \,. \tag{7.55}$$

In an inertial frame at any point on the brane, we have

$$u^\mu = (1, \mathbf{0}) \,, \; h_{\mu\nu} = \mathrm{diag}(0,1,1,1) \,, \; q_\mu = (0, q_i) \,, \; \pi_{\mu 0} = 0 = \sum \pi_{ii} \,. \tag{7.56}$$

The tensor $\mathcal{S}_{\mu\nu}$, which carries local bulk effects onto the brane, may then be irreducibly decomposed as

$$\begin{aligned} \mathcal{S}_{\mu\nu} = {} & \frac{1}{24} \left[2\rho^2 - 3\pi_{\alpha\beta}\pi^{\alpha\beta} \right] u_\mu u_\nu + \frac{1}{24} \left[2\rho^2 + 4\rho p + \pi_{\alpha\beta}\pi^{\alpha\beta} - 4 q_\alpha q^\alpha \right] h_{\mu\nu} \\ & - \frac{1}{12}(\rho + 2p)\pi_{\mu\nu} + \pi_{\alpha\langle\mu}\pi_{\nu\rangle}{}^\alpha \\ & + q_{\langle\mu} q_{\nu\rangle} + \frac{1}{3}\rho q_{(\mu} u_{\nu)} - \frac{1}{12} q^\alpha \pi_{\alpha(\mu} u_{\nu)} \,. \end{aligned} \tag{7.57}$$

This simplifies for a perfect fluid or minimally-coupled scalar field:

$$\mathcal{S}_{\mu\nu} = \frac{1}{12}\rho \left[\rho u_\mu u_\nu + (\rho + 2p)\, h_{\mu\nu} \right] \,. \tag{7.58}$$

The trace free $\mathcal{E}_{\mu\nu}$ carries nonlocal bulk effects onto the brane, and contributes an effective "dark" radiative energy-momentum on the brane, with

energy density ρ^*, pressure $\rho^*/3$, momentum density q^*_μ and anisotropic stress $\pi^*_{\mu\nu}$:

$$-\frac{1}{\kappa^2}\mathcal{E}_{\mu\nu} = \rho^*\left(u_\mu u_\nu + \frac{1}{3}h_{\mu\nu}\right) + q^*_\mu u_\nu + q^*_\nu u_\mu + \pi^*_{\mu\nu}\,. \tag{7.59}$$

We can think of this as a KK or Weyl "fluid". The brane "feels" the bulk gravitational field through this effective fluid. More specifically:

- The KK (or Weyl) anisotropic stress $\pi^*_{\mu\nu}$ incorporates the scalar or spin-0 ("Coulomb"), the vector (transverse) or spin-1 (gravimagnetic) and the tensor (transverse traceless) or spin-2 (gravitational wave) 4D modes of the spin-2 5D graviton.
- The KK momentum density q^*_μ incorporates spin-0 and spin-1 modes, and defines a velocity v^*_μ of the Weyl fluid relative to u^μ via $q^*_\mu = \rho^* v^*_\mu$.
- The KK energy density ρ^*, often called the "dark radiation", incorporates the spin-0 mode.

The RS models have a Minkowski brane in an AdS_5 bulk. This bulk is also compatible with an FRW brane. However, the most general vacuum bulk with a Friedmann brane is Schwarzschild-anti de Sitter spacetime [18]. It follows from the FRW symmetries that $q^*_\mu = 0 = \pi^*_{\mu\nu}$, where $\rho^* = 0$ only if the mass of the black hole in the bulk is zero. The presence of the bulk black hole generates via Coulomb effects the dark radiation on the brane.

The brane-world corrections can conveniently be consolidated into an effective total energy density, pressure, momentum density and anisotropic stress.

$$\rho^{\rm tot} = \rho + \frac{1}{4\lambda}\left(2\rho^2 - 3\pi_{\mu\nu}\pi^{\mu\nu}\right) + \rho^* \tag{7.60}$$

$$p^{\rm tot} = p + \frac{1}{4\lambda}\left(2\rho^2 + 4\rho p + \pi_{\mu\nu}\pi^{\mu\nu} - 4q_\mu q^\mu\right) + \frac{\rho^*}{3} \tag{7.61}$$

$$q^{\rm tot}_\mu = q_\mu + \frac{1}{2\lambda}\left(2\rho q_\mu - 3\pi_{\mu\nu}q^\nu\right) + q^*_\mu \tag{7.62}$$

$$\pi^{\rm tot}_{\mu\nu} = \pi_{\mu\nu} + \frac{1}{2\lambda}\left[-(\rho+3p)\pi_{\mu\nu} + 3\pi_{\alpha\langle\mu}\pi_{\nu\rangle}{}^\alpha + 3q_{\langle\mu}q_{\nu\rangle}\right] + \pi^*_{\mu\nu}\,. \tag{7.63}$$

These general expressions simplify in the case of a perfect fluid (or minimally coupled scalar field, or isotropic one-particle distribution function), *i.e.* for $q_\mu = 0 = \pi_{\mu\nu}$:

$$\rho^{\rm tot} = \rho\left(1 + \frac{\rho}{2\lambda} + \frac{\rho^*}{\rho}\right) \tag{7.64}$$

$$p^{\rm tot} = p + \frac{\rho}{2\lambda}(2p+\rho) + \frac{\rho^*}{3} \tag{7.65}$$

$$q^{\rm tot}_\mu = q^*_\mu \tag{7.66}$$

$$\pi^{\rm tot}_{\mu\nu} = \pi^*_{\mu\nu}\,. \tag{7.67}$$

Note that nonlocal bulk effects can contribute to effective imperfect fluid terms even when the matter on the brane has perfect fluid form: there is in general an effective momentum density and anisotropic stress induced on the brane by the 5D graviton.

The effective total equation of state and sound speed follow from (7.64) and (7.65) as

$$w^{\rm tot} \equiv \frac{p^{\rm tot}}{\rho^{\rm tot}} = \frac{w + (1+2w)\rho/2\lambda + \rho^*/3\rho}{1 + \rho/2\lambda + \rho^*/\rho} \tag{7.68}$$

$$c^2_{\rm tot} \equiv \frac{\dot{p}^{\rm tot}}{\dot{\rho}^{\rm tot}} = \left[c^2_{\rm s} + \frac{\rho+p}{\rho+\lambda} + \frac{4\rho^*}{9(\rho+p)(1+\rho/\lambda)}\right] \left[1 + \frac{4\rho^*}{3(\rho+p)(1+\rho/\lambda)}\right]^{-1}, \tag{7.69}$$

where $w = p/\rho$ and $c^2_{\rm s} = \dot{p}/\dot{\rho}$. At very high energies, i.e., $\rho \gg \lambda$, we can generally neglect ρ^* (e.g., in an inflating cosmology), and the effective equation of state and sound speed are stiffened:

$$w^{\rm tot} \approx 2w + 1\,, \quad c^2_{\rm tot} \approx c^2_{\rm s} + w + 1\,. \tag{7.70}$$

This can have important consequences in the early universe and during gravitational collapse. For example, in a very high energy radiation era, $w = \frac{1}{3}$, the effective cosmological equation of state is ultra-stiff: $w^{\rm tot} \approx \frac{5}{3}$. In late-stage gravitational collapse of pressureless matter, $w = 0$, the effective equation of state is stiff, $w^{\rm tot} \approx 1$, and the effective pressure is nonzero and dynamically important.

7.3.3 Conservation Equations: Ordinary and "Weyl" Fluids

Conservation of $T_{\mu\nu}$ gives the standard general relativity energy and momentum conservation equations:

$$\dot{\rho} + \Theta(\rho+p) + {\rm D}^\mu q_\mu + 2A^\mu q_\mu + \sigma^{\mu\nu}\pi_{\mu\nu} = 0 \tag{7.71}$$

$$\dot{q}_{\langle\mu\rangle} + \frac{4}{3}\Theta q_\mu + {\rm D}_\mu p + (\rho+p)A_\mu + {\rm D}^\nu \pi_{\mu\nu} + A^\nu \pi_{\mu\nu} + \sigma_{\mu\nu}q^\nu - \varepsilon_{\mu\nu\alpha}\omega^\nu q^\alpha = 0\,. \tag{7.72}$$

In these equations, an overdot denotes $u^\nu\nabla_\nu$, $\Theta = \nabla^\mu u_\mu$ is the volume expansion rate of the u^μ worldlines, $A_\mu = \dot{u}_\mu = A_{\langle\mu\rangle}$ is their four-acceleration, $\sigma_{\mu\nu} = {\rm D}_{\langle\mu}u_{\nu\rangle}$ is their shear rate, and $\omega_\mu = -\frac{1}{2}\varepsilon_{\mu\alpha\beta}{\rm D}^\alpha u^\beta = \omega_{\langle\mu\rangle}$ is their vorticity rate. On a Friedmann brane,

$$A_\mu = \omega_\mu = \sigma_{\mu\nu} = 0\,, \quad \Theta = 3H\,, \tag{7.73}$$

where $H = \dot{a}/a$ is the Hubble rate. Here ${\rm D}_\mu$ is the spatially projected part of the brane covariant derivative, defined by

$$\mathrm{D}_\mu F^{\alpha\cdots}{}_{\cdots\beta} = (\nabla_\mu F^{\alpha\cdots}{}_{\cdots\beta})_\perp = h_\mu{}^\nu h^\alpha{}_\gamma \cdots h_\beta{}^\delta \nabla_\nu F^{\gamma\cdots}{}_{\cdots\delta}\,. \tag{7.74}$$

In a local inertial frame at a point on the brane, with $u^\mu = \delta^\mu{}_0$, we have: $0 = A_0 = \omega_0 = \sigma_{0\mu} = \varepsilon_{0\alpha\beta}$ and $\mathrm{D}_\mu F^{\alpha\cdots}{}_{\cdots\beta} = \delta_\mu{}^i \delta^\alpha{}_j \cdots \delta_\beta{}^k \nabla_i F^{j\cdots}{}_{\cdots k}$, where $i, j, k = 1, 2, 3$.

The absence of bulk source terms in the matter conservation equations is a consequence of having Λ_5 as the only 5D source in the bulk. For example, if there is a bulk scalar field, then there is energy-momentum exchange between the brane and bulk (in addition to the gravitational interaction) [19].

Equation (7.51) may be called the "nonlocal conservation equation for the Weyl fluid". Projecting along u^μ gives the nonlocal energy conservation equation [20], which is a propagation equation for ρ^*:

$$\begin{aligned}
&\dot\rho^* + \frac{4}{3}\Theta\rho^* + \mathrm{D}^\mu q^*_\mu + 2A^\mu q^*_\mu + \sigma^{\mu\nu}\pi^*_{\mu\nu} \\
&\quad = \frac{1}{4\lambda}\left[6\pi^{\mu\nu}\dot\pi_{\mu\nu} + 6(\rho+p)\sigma^{\mu\nu}\pi_{\mu\nu} + 2\Theta\left(2q^\mu q_\mu + \pi^{\mu\nu}\pi_{\mu\nu}\right) + 2A^\mu q^\nu \pi_{\mu\nu}\right. \\
&\qquad \left. - 4q^\mu \mathrm{D}_\mu\rho + q^\mu \mathrm{D}^\nu \pi_{\mu\nu} + \pi^{\mu\nu}\mathrm{D}_\mu q_\nu - 2\sigma^{\mu\nu}\pi_{\alpha\mu}\pi_\nu{}^\alpha - 2\sigma^{\mu\nu}q_\mu q_\nu\right]\,.
\end{aligned} \tag{7.75}$$

Projecting into the comoving rest space gives the nonlocal momentum conservation equation, which is a propagation equation for q^*_μ:

$$\begin{aligned}
&\dot q^*_{\langle\mu\rangle} + \frac{4}{3}\Theta q^*_\mu + \frac{1}{3}\mathrm{D}_\mu\rho^* + \frac{4}{3}\rho^* A_\mu + \mathrm{D}^\nu \pi^*_{\mu\nu} + A^\nu \pi^*_{\mu\nu} + \sigma_\mu{}^\nu q^*_\nu - \varepsilon_\mu{}^{\nu\alpha}\omega_\nu q^*_\alpha \\
&\quad = \frac{1}{4\lambda}\left[-4(\rho+p)\mathrm{D}_\mu\rho + 6(\rho+p)\mathrm{D}^\nu\pi_{\mu\nu} + q^\nu\dot\pi_{\langle\mu\nu\rangle} + \pi_\mu{}^\nu \mathrm{D}_\nu(2\rho+5p)\right. \\
&\qquad - \frac{2}{3}\pi^{\alpha\beta}\left(\mathrm{D}_\mu\pi_{\alpha\beta} + 3\mathrm{D}_\alpha\pi_{\beta\mu}\right) - 3\pi_{\mu\alpha}\mathrm{D}_\beta\pi^{\alpha\beta} + \frac{28}{3}q^\nu\mathrm{D}_\mu q_\nu \\
&\qquad + 4\rho A^\nu\pi_{\mu\nu} - 3\pi_{\mu\alpha}A_\beta\pi^{\alpha\beta} + \frac{8}{3}A_\mu\pi^{\alpha\beta}\pi_{\alpha\beta} - \pi_{\mu\alpha}\sigma^{\alpha\beta}q_\beta \\
&\qquad + \sigma_{\mu\alpha}\pi^{\alpha\beta}q_\beta + \pi_{\mu\nu}\varepsilon^{\nu\alpha\beta}\omega_\alpha q_\beta - \varepsilon_{\mu\alpha\beta}\omega^\alpha\pi^{\beta\nu}q_\nu + 4(\rho+p)\Theta q_\mu \\
&\qquad \left. + 6q_\mu A^\nu q_\nu + \frac{14}{3}A_\mu q^\nu q_\nu + 4q_\mu\sigma^{\alpha\beta}\pi_{\alpha\beta}\right]\,.
\end{aligned} \tag{7.76}$$

The (1+3)-covariant decomposition shows two key features:

- inhomogeneous and anisotropic effects from the 4D matter-radiation distribution on the brane are a source for the 5D Weyl tensor, which nonlocally "backreacts" on the brane via its projection $\mathcal{E}_{\mu\nu}$;
- there are evolution equations for the dark radiative (nonlocal, Weyl) energy (ρ^*) and momentum (q^*_μ) densities (carrying scalar and vector modes from bulk gravitons), but there is no evolution equation for the dark radiative anisotropic stress ($\pi^*_{\mu\nu}$) (carrying tensor, as well as scalar and vector, modes), which arises in both evolution equations.

In particular cases, the Weyl anisotropic stress $\pi^*_{\mu\nu}$ may drop out of the nonlocal conservation equations, *i.e.* when we can neglect $\sigma^{\mu\nu}\pi^*_{\mu\nu}$, $\mathrm{D}^\nu\pi^*_{\mu\nu}$

and $A^\nu \pi^*_{\mu\nu}$. This is the case when we consider linearized perturbations about an FRW background (which remove the first and last of these terms) and further when we can neglect gradient terms on large scales (which removes the second term) [20]. But in general, on small scales in cosmology, and especially in astrophysical contexts, the $\pi^*_{\mu\nu}$ terms cannot be neglected.

All of the matter source terms on the right of these two equations, except for the first term on the right of (7.76), are imperfect fluid terms, and most of these terms are quadratic in the imperfect quantities q_μ and $\pi_{\mu\nu}$. For a single perfect fluid or scalar field, only the $\mathrm{D}_\mu\rho$ term on the right of (7.76) survives, but in realistic cosmological and astrophysical models, further terms will survive. For example, terms linear in $\pi_{\mu\nu}$ will carry the photon quadrupole in cosmology or the shear viscous stress in stellar models. If there are two fluids (even if both fluids are perfect), then there will be a relative velocity v_μ generating a momentum density $q_\mu = \rho v_\mu$, which will serve to source nonlocal effects.

In general, the four independent equations in (7.75) and (7.76) constrain four of the nine independent components of $\mathcal{E}_{\mu\nu}$ on the brane. What is missing, is an evolution equation for $\pi^*_{\mu\nu}$ (which has up to five independent components). Thus in general, the projection of the five-dimensional field equations onto the brane does not lead to a closed system, as expected, since there are bulk degrees of freedom whose impact on the brane cannot be predicted by brane observers. The Weyl or KK anisotropic stress $\pi^*_{\mu\nu}$ encodes the nonlocality.

In special cases the missing equation does not matter. For example, if $\pi^*_{\mu\nu} = 0$ by symmetry, as in the case of an FRW brane, then the evolution of $\mathcal{E}_{\mu\nu}$ is determined by (7.75) and (7.76). However, small perturbations of this special case will immediately restore the problem of missing information.

If the matter on the brane has a perfect-fluid energy-momentum tensor, the local conservation equations (7.71) and (7.72) reduce to

$$\dot\rho + \Theta(\rho + p) = 0 \tag{7.77}$$

$$\mathrm{D}_\mu p + (\rho + p)A_\mu = 0\,, \tag{7.78}$$

while the nonlocal conservation equations (7.75) and (7.76) reduce to

$$\dot\rho^* + \frac{4}{3}\Theta\rho^* + \mathrm{D}^\mu q^*_\mu = 0 \tag{7.79}$$

$$\dot q^*_{\langle\mu\rangle} + \frac{4}{3}\Theta q^*_\mu + \frac{1}{3}\mathrm{D}_\mu\rho^* + \frac{4}{3}\rho^* A_\mu + \mathrm{D}^\nu \pi^*_{\mu\nu} = -\frac{(\rho + p)}{\lambda}\mathrm{D}_\mu\rho\,, \tag{7.80}$$

where we have also linearized about an FRW background.

7.4 Brane-World Cosmology: Dynamics

In the FRW case, (7.80) is trivially satisfied, while (7.79) becomes

$$\dot{\rho}^* + 4H\rho^* = 0\,. \tag{7.81}$$

This equation has the dark radiation solution

$$\rho^* = \rho_0^* \left(\frac{a_0}{a}\right)^4 . \tag{7.82}$$

In natural static coordinates, the Schwarzschild-AdS$_5$ metric for an FRW brane-world is

$${}^{(5)}ds^2 = -F(R)dT^2 + \frac{dR^2}{F(R)} + R^2\left(\frac{dr^2}{1-Kr^2} + r^2 d\Omega^2\right) \tag{7.83}$$

$$F(R) = K + \frac{R^2}{\ell^2} - \frac{m}{R^2}\,, \tag{7.84}$$

where $K = 0, \pm 1$ is the FRW curvature index and m is the mass parameter of the black hole at $R = 0$ (note that the 5D gravitational potential has R^{-2} behaviour). The bulk black hole gives rise to dark radiation on the brane via its Coulomb effect. The FRW brane moves radially along the fifth dimension, with $R = a(T)$, where a is the FRW scale factor, and the junction conditions determine the velocity via the Friedmann equation for a [18]. Thus one can interpret the expansion of the universe as motion of the brane through the static bulk. In the special case $m = 0$ and $da/dT = 0$, the brane is fixed and has Minkowski geometry, *i.e.* the original RSII brane-world is recovered, in different coordinates.

The velocity of the brane is coordinate-dependent, and can be set to zero. We can use Gaussian normal coordinates, in which the brane is fixed but the bulk metric is not manifestly static [14]:

$${}^{(5)}ds^2 = -N^2(t,y)dt^2 + A^2(t,y)\left[\frac{dr^2}{1-Kr^2} + r^2 d\Omega^2\right] + dy^2\,. \tag{7.85}$$

Here $a(t) = A(t,0)$ is the scale factor on the FRW brane at $y = 0$, and t may be chosen as proper time on the brane, so that $N(t,0) = 1$. In the case where there is no bulk black hole ($m = 0$), the metric functions are

$$N = \frac{\dot{A}(t,y)}{\dot{a}(t)} \tag{7.86}$$

$$A = a(t)\left[\cosh\left(\frac{y}{\ell}\right) - \left\{1 + \frac{\rho(t)}{\lambda}\right\}\sinh\left(\frac{|y|}{\ell}\right)\right]. \tag{7.87}$$

Again, the junction conditions determine the Friedmann equation. The extrinsic curvature at the brane is

$$K^{\mu}{}_{\nu} = \mathrm{diag}\left(\frac{N'}{N}, \frac{A'}{A}, \frac{A'}{A}, \frac{A'}{A}\right)_{\mathrm{brane}} . \tag{7.88}$$

Then, by (7.41),

$$\frac{N'}{N}\Big|_{\mathrm{brane}} = \frac{\kappa_5^2}{6}(2\rho + 3p - \lambda) \tag{7.89}$$

$$\frac{A'}{A}\Big|_{\mathrm{brane}} = -\frac{\kappa_5^2}{6}(\rho + \lambda) . \tag{7.90}$$

The field equations yield the first integral [14]

$$(AA')^2 - \frac{A^2}{N^2}\dot{A}^2 + \frac{\Lambda_5}{6}A^4 + m = 0 , \tag{7.91}$$

where m is constant. Evaluating this at the brane, using (7.90), gives the modified Friedmann equation

$$H^2 = \frac{\kappa^2}{3}\rho\left(1 + \frac{\rho}{2\lambda}\right) + \frac{m}{a^4} + \frac{1}{3}\Lambda - \frac{K}{a^2} . \tag{7.92}$$

By (7.82),

$$m = \frac{\kappa^2}{3}\rho_0^* a_0^4 . \tag{7.93}$$

The Friedmann and matter energy conservation equations yield

$$\dot{H} = -\frac{\kappa^2}{2}(\rho + p)\left(1 + \frac{\rho}{\lambda}\right) - 2\frac{m}{a^4} + \frac{K}{a^2} . \tag{7.94}$$

When the bulk black hole mass vanishes, the bulk geometry reduces to AdS_5 and $\rho^* = 0$. In order to avoid a naked singularity, we assume that the black hole mass is non-negative, so that $\rho_0^* \geq 0$. [By (7.84), it is possible to avoid a naked singularity with negative m when $K = -1$, provided $|m| \leq \ell^2/4$.] This additional effective relativistic degree of freedom is constrained by nucleosynthesis and CMB observations to be no more than $\sim$5% of the radiation energy density [21, 22]:

$$\frac{\rho^*}{\rho_{\mathrm{rad}}}\Big|_{\mathrm{nuc}} . \lesssim 0.05 \tag{7.95}$$

The other modification to the Hubble rate is via the high energy correction ρ/λ. In order to recover the observational successes of general relativity, the high energy regime where significant deviations occur must take place before nucleosynthesis, i.e., cosmological observations impose the lower limit

$$\lambda > (1\ \mathrm{MeV})^4 \ \Rightarrow\ M_5 > 10^4\ \mathrm{GeV} . \tag{7.96}$$

This is much weaker than the limit imposed by table-top experiments, (7.27). Since ρ^2/λ decays as a^{-8} during the radiation era, it will rapidly become negligible after the end of the high energy regime, $\rho = \lambda$.

If $\rho^* = 0$ and $K = 0 = \Lambda$, then the exact solution of the Friedmann equations for $w = p/\rho =$ const. is [14]

$$a = \text{ const}\,[t(t+t_\lambda)]^{1/3(w+1)}\,, \quad t_\lambda = \frac{M_{\rm p}}{\sqrt{3\pi\lambda}(1+w)} \lesssim (1+w)^{-1}10^{-9}\,\text{sec}\,, \tag{7.97}$$

where $w > -1$. If $\rho^* \neq 0$ (but $K = 0 = \Lambda$), then the solution for the radiation era ($w = \frac{1}{3}$) is

$$a = \text{ const}\,[t(t+t_\lambda)]^{1/4}\,, \quad t_\lambda = \frac{\sqrt{3}\,M_{\rm p}}{4\sqrt{\pi\lambda}\,(1+\rho^*/\rho)}\,. \tag{7.98}$$

For $t \gg t_\lambda$ we recover from (7.97) and (7.98) the standard behaviour, $a \propto t^{2/3(w+1)}$, whereas for $t \ll t_\lambda$, we have the very different behaviour of the high energy regime,

$$\rho \gg \lambda \;\Rightarrow\; \mu \propto t^{1/3(w+1)}\,. \tag{7.99}$$

When $w = -1$ we have $\rho = \rho_0$ from the conservation equation. If $K = 0 = \Lambda$, we recover the de Sitter solution for $\rho^* = 0$ and an asymptotically de Sitter solution for $\rho^* > 0$:

$$\rho^* = 0: \quad a = a_0 \exp[H_0(t-t_0)]\,, \quad H_0 = \kappa\sqrt{\frac{\rho_0}{3}\left(1+\frac{\rho_0}{2\lambda}\right)} \tag{7.100}$$

$$\rho^* > 0: \quad a^2 = \sqrt{\frac{m}{H_0}}\,\sinh[2H_0(t-t_0)]\,. \tag{7.101}$$

7.5 Brane-World Inflation

In RSII-type brane-worlds, where the bulk has only a vacuum energy, inflation on the brane must be driven by a 4D scalar field trapped on the brane. In more general brane-worlds, where the bulk contains a 5D scalar field, it is possible that the 5D field induces inflation on the brane via its effective projection [23].

More exotic possibilities arise from the interaction between two branes, including possible collision, which is mediated by a 5D scalar field and which can induce either inflation [24] or a hot big-bang radiation era, as in the "ekpyrotic" or cyclic scenario [25]. Here we discuss the simplest case of a 4D scalar field ϕ with potential $V(\phi)$.

High energy brane-world modifications to the dynamics of inflation on the brane have been investigated [26, 27]. Essentially, the high energy corrections provide increased Hubble damping, since $\rho \gg \lambda$ implies H is larger for a given energy than in 4D general relativity. This makes slow-roll inflation possible even for potentials that would be too steep in standard cosmology [26, 28, 29].

The field satisfies the Klein-Gordon equation

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0\,. \tag{7.102}$$

In 4D general relativity, the condition for inflation, $\ddot{a} > 0$, is $\dot{\phi}^2 < V(\phi)$, i.e., $p < -\frac{1}{3}\rho$, where $\rho = \frac{1}{2}\dot{\phi}^2 + V$ and $p = \frac{1}{2}\dot{\phi}^2 - V$. The modified Friedmann equation leads to a stronger condition for inflation: using (7.92), with $m = 0 = \Lambda = K$, and (7.102), we find that

$$\ddot{a} > 0 \;\Rightarrow\; w < -\frac{1}{3}\left[\frac{1+2\rho/\lambda}{1+\rho/\lambda}\right], \tag{7.103}$$

where the square brackets enclose the brane correction to the general relativity result. As $\rho/\lambda/ \to 0$, the 4D result $w < -\frac{1}{3}$ is recovered, but for $\rho > \lambda$, w must be more negative for inflation. In the very high energy limit $\rho/\lambda \to \infty$, we have $w < -\frac{2}{3}$. When the only matter in the universe is a self-interacting scalar field, the condition for inflation becomes

$$\dot{\phi}^2 - V + \left[\frac{\frac{1}{2}\dot{\phi}^2 + V}{\lambda}\left(\frac{5}{4}\dot{\phi}^2 - \frac{1}{2}V\right)\right] < 0\,, \tag{7.104}$$

which reduces to $\dot{\phi}^2 < V$ when $\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V \ll \lambda$.

In the the slow-roll approximation,

$$H^2 \approx \frac{\kappa^2}{3}V\left[1 + \frac{V}{2\lambda}\right] \tag{7.105}$$

$$\dot{\phi} \approx -\frac{V'}{3H}\,. \tag{7.106}$$

The brane-world correction term V/λ in (7.105) serves to enhance the Hubble rate for a given potential energy, relative to general relativity. Thus there is enhanced Hubble 'friction' in (7.106), and brane-world effects will reinforce slow-roll at the same potential energy. We can see this by defining slow-roll parameters that reduce to the standard parameters in the low-energy limit:

$$\epsilon \equiv -\frac{\dot{H}}{H^2} = \frac{M_{\rm p}^2}{16\pi}\left(\frac{V'}{V}\right)^2\left[\frac{1+V/\lambda}{(1+V/2\lambda)^2}\right] \tag{7.107}$$

$$\eta \equiv -\frac{\ddot{\phi}}{H\dot{\phi}} = \frac{M_{\rm p}^2}{8\pi}\left(\frac{V''}{V}\right)\left[\frac{1}{1+V/2\lambda}\right]. \tag{7.108}$$

Self-consistency of the slow-roll approximation then requires $\epsilon, |\eta| \ll 1$. At low energies, $V \ll \lambda$, the slow-roll parameters reduce to the standard form. However at high energies, $V \gg \lambda$, the extra contribution to the Hubble expansion helps damp the rolling of the scalar field and the new factors in square brackets become $\approx \lambda/V$:

$$\epsilon \approx \epsilon_{\rm gr}\left[\frac{4\lambda}{V}\right],\; \eta \approx \eta_{\rm gr}\left[\frac{2\lambda}{V}\right], \tag{7.109}$$

where $\epsilon_{\rm gr}, \eta_{\rm gr}$ are the standard general relativity slow-roll parameters. In particular, this means that steep potentials which do not give inflation in general relativity, can inflate the brane-world at high energy and then naturally stop inflating when V drops below λ. These models can be constrained because they typically end inflation in a kinetic-dominated regime and thus generate a blue spectrum of gravitational waves, which can disturb nucleosynthesis [28]. They also allow for the novel possibility that the inflaton could act as dark matter or quintessence at low energies [28, 30].

The number of e-folds during inflation, $N = \int H dt$, is, in the slow-roll approximation,

$$N \approx -\frac{8\pi}{M_{\rm p}^2} \int_{\phi_{\rm i}}^{\phi_{\rm f}} \frac{V}{V'} \left[1 + \frac{V}{2\lambda}\right] d\phi \,. \tag{7.110}$$

Brane-world effects at high energies increase the Hubble rate by a factor $V/2\lambda$, yielding more inflation between any two values of ϕ for a given potential. Thus we can obtain a given number of e-folds for a smaller initial inflaton value $\phi_{\rm i}$. For $V \gg \lambda$, (7.110) becomes

$$N \approx -\frac{128\pi^3}{3M_5^6} \int_{\phi_{\rm i}}^{\phi_{\rm f}} \frac{V^2}{V'} \, d\phi \,. \tag{7.111}$$

The key test of any modified gravity theory during inflation, will be the spectrum of perturbations produced due to quantum fluctuations of the fields about their homogeneous background values. We will discuss brane-world cosmological perturbations in the next section. In general, perturbations on the brane are coupled to bulk metric perturbations, and the problem is very complicated. However on large scales on the brane, the density and curvature perturbations decouple from the bulk metric perturbations [20, 21]. Thus we are justified in neglecting the nonlocal effects carried by $\mathcal{E}_{\mu\nu}$ when computing the density perturbations.

To quantify the amplitude of scalar (density) perturbations we evaluate the usual gauge invariant quantity

$$\zeta \equiv \mathcal{R} - \frac{H}{\dot\rho} \delta\rho \,, \tag{7.112}$$

which reduces to the curvature perturbation, $\mathcal{R}$, on uniform density hypersurfaces ($\delta\rho = 0$). This is conserved on large scales for purely adiabatic perturbations, as a consequence of energy conservation (independently of the field equations) [31]. The curvature perturbation on uniform density hypersurfaces is given in terms of the scalar field fluctuations on spatially flat hypersurfaces, $\delta\phi$, by

$$\zeta = H \frac{\delta\phi}{\dot\phi} \,. \tag{7.113}$$

The field fluctuations at Hubble crossing ($k = aH$) in the slow-roll limit are given by $\langle \delta\phi^2 \rangle \approx (H/2\pi)^2$, a result for a massless field in de Sitter space that

is also independent of the gravity theory [31]. For a single scalar field the perturbations are adiabatic and hence the curvature perturbation ζ can be related to the density perturbations when modes re-enter the Hubble scale during the matter dominated era which is given by $A_{\rm s}^2 = 4\langle\zeta^2\rangle/25$. Using the slow-roll equations and (7.113), this gives

$$A_{\rm s}^2 \approx \left(\frac{512\pi}{75M_{\rm p}^6}\,\frac{V^3}{V'^2}\right)\left[\frac{2\lambda+V}{2\lambda}\right]^3\Bigg|_{k=aH} . \tag{7.114}$$

Thus the amplitude of scalar perturbations is *increased* relative to the standard result at a fixed value of ϕ for a given potential.

The scale-dependence of the perturbations is described by the spectral tilt

$$n_{\rm s} - 1 \equiv \frac{d\ln A_{\rm s}^2}{d\ln k} \approx -6\epsilon + 2\eta\,, \tag{7.115}$$

where the slow-roll parameters are given in (7.107) and (7.108). Because these slow-roll parameters are both suppressed by an extra factor λ/V at high energies, we see that the spectral index is driven towards the Harrison-Zel'dovich spectrum, $n_{\rm s} \to 1$, as $V/\lambda \to \infty$; however, this does not necessarily mean that the brane-world case is closer to scale-invariance than the general relativity case.

In comparing the high energy brane-world case to the standard 4D case, we implicitly require the same potential energy. However, precisely because of the high energy effects, large scale perturbations will be generated at different values of V than in the standard case, specifically at lower values of V, closer to the reheating minimum. Thus there are two competing effects, and it turns out that the shape of the potential determines which is the dominant effect [32]. For the quadratic potential, the lower location on V dominates, and the spectral tilt is slightly further from scale invariance than in the standard case. The same holds for the quartic potential. Data from WMAP and 2dF can be used to constrain inflationary models via their deviation from scale invariance, and the high energy brane-world versions of the quadratic and quartic potentials are thus under more pressure from data than their standard counterparts [32].

Other perturbation modes are also generated in inflation:

– High energy inflation on the brane also generates a zero mode (4D graviton mode) of tensor perturbations, and stretches it to super-Hubble scales, as will be discussed below. This zero mode has the same qualitative features as in general relativity, remaining frozen at constant amplitude while beyond the Hubble horizon. Its amplitude is enhanced at high energies, although the enhancement is much less than for scalar perturbations [33]:

$$A_{\rm t}^2 \approx \left(\frac{32V}{75M_{\rm p}^2}\right)\left[\frac{3V^2}{4\lambda^2}\right] \tag{7.116}$$

$$\frac{A_{\rm t}^2}{A_{\rm s}^2} \approx \left(\frac{M_{\rm p}^2}{16\pi}\,\frac{V'^2}{V^2}\right)\left[\frac{6\lambda}{V}\right]. \tag{7.117}$$

Equation (7.117) means that brane-world effects suppress the large scale tensor contribution to CMB anisotropies. The tensor spectral index at high energy has a smaller magnitude than in general relativity,

$$n_{\rm t} = -3\epsilon\,, \tag{7.118}$$

but remarkably the same consistency relation as in general relativity holds [29]:

$$n_{\rm t} = -2\frac{A_{\rm t}^2}{A_{\rm s}^2}\,. \tag{7.119}$$

The massive KK modes of tensor perturbations remain in the vacuum state during slow-roll inflation [33, 34]. The evolution of the super-Hubble zero mode is the same as in general relativity, so that high energy brane-world effects in the early universe serve only to rescale the amplitude. However, when the zero mode re-enters the Hubble horizon, massive KK modes can be excited.

– Vector perturbations in the bulk metric can support vector metric perturbations on the brane, even in the absence of matter perturbations. However, there is no normalizable zero mode, and the massive KK modes stay in the vacuum state during brane-world inflation [35]. Therefore, as in general relativity, we can neglect vector perturbations in inflationary cosmology.

7.6 Brane-World Cosmology: Perturbations

The background dynamics of brane-world cosmology are simple because the FRW symmetries simplify the bulk and rule out nonlocal effects. But perturbations on the brane in general release the nonlocal KK modes. Then the 5D bulk perturbation equations must be solved in order to solve for perturbations on the brane. These 5D equations are partial differential equations for the three-Fourier modes, with complicated initial and boundary conditions.

The theory of gauge invariant perturbations in brane-world cosmology has been extensively investigated and developed (see references given in the reviews [9, 10, 11]) and is qualitatively well understood (for a general review on cosmological perturbations see the review article by R. Durrer). The key remaining task is integration of the coupled brane-bulk perturbation equations. Special cases have been solved, where these equations effectively decouple, as in the previous section, and approximation schemes have recently been developed [36, 37, 38, 39, 40] for the more general cases where the coupled system

must be solved. From the brane viewpoint, the bulk effects, *i.e.* the high energy corrections and the KK modes, act as source terms for the brane perturbation equations. At the same time, perturbations of matter on the brane can generate KK modes (*i.e.* emit 5D gravitons into the bulk) which propagate in the bulk and can subsequently interact with the brane. This nonlocal interaction amongst the perturbations is at the core of the complexity of the problem. It can be elegantly expressed via integro-differential equations [41], which take the form (assuming no incoming 5D gravitational waves)

$$A_k(t) = \int dt'\, \mathcal{G}(t,t')B_k(t')\,, \tag{7.120}$$

where $\mathcal{G}$ is the bulk retarded Green's function evaluated on the brane, and A_k, B_k are made up of brane metric and matter perturbations and their (brane) derivatives, and include high energy corrections to the background dynamics. Solving for the bulk Green's function which then determines $\mathcal{G}$, is the core of the 5D problem.

We can isolate the KK anisotropic stress $\pi^*_{\mu\nu}$ as the term that must be determined from 5D equations. Once $\pi^*_{\mu\nu}$ is determined in this way, the perturbation equations on the brane form a closed system. The solution will be of the form, expressed in Fourier modes:

$$\pi^*_k(t) \propto \int dt'\ \mathcal{G}(t,t')F_k(t')\,, \tag{7.121}$$

where the functional F_k will be determined by the brane perturbation quantities and their derivatives. It is known in the case of a Minkowski background, but not in the cosmological case.

7.6.1 Metric-Based Perturbations

A review of this approach is given in [42]. In an arbitrary gauge, and for a flat FRW background, the perturbed metric has the form

$$\delta\,{}^{(5)}g_{AB} = \left[\begin{array}{cc|c} -2N^2\psi & A^2(\partial_i\mathcal{B} - S_i) & N\alpha \\ A^2(\partial_j\mathcal{B} - S_j) & A^2\left\{2\mathcal{R}\delta_{ij} + 2\partial_i\partial_j\mathcal{C} \right. & A^2(\partial_i\beta - \chi_i) \\ & \left. +2\partial_{(i}F_{j)} + f_{ij}\right\} & \\ \hline N\alpha & A^2(\partial_j\beta - \chi_j) & 2\nu \end{array}\right], \tag{7.122}$$

where the background metric functions A, N are given by (7.86) and (7.87). The scalars $\psi, \mathcal{R}, \mathcal{C}, \alpha, \beta, \nu$ represent scalar perturbations. The vectors S_i, F_i and χ_i are transverse, so that they represent 3D vector perturbations, and the tensor f_{ij} is transverse traceless, representing 3D tensor perturbations.

In the Gaussian normal gauge, the brane coordinate-position remains fixed under perturbation,

$$^{(5)}ds^2 = \left[g^{(0)}_{\mu\nu}(x,y) + \delta g_{\mu\nu}(x,y)\right] dx^\mu dx^\nu + dy^2\,, \tag{7.123}$$

where $g^{(0)}_{\mu\nu}$ is the background metric, (7.85), so that

$$\alpha = \beta = \nu = \chi_i = 0\,. \tag{7.124}$$

In the 5D longitudinal gauge,

$$-\mathcal{B} + \dot{\mathcal{C}} = 0 = -\beta + \mathcal{C}'\,. \tag{7.125}$$

In this gauge, and for an AdS_5 background, the metric perturbation quantities can all be expressed in terms of a "master variable" [41] Ω which obeys a wave equation. In the case of scalar perturbations, we have for example,

$$\mathcal{R} = \frac{1}{6A}\left(\Omega'' - \frac{1}{N^2}\,\ddot{\Omega} - \frac{\Lambda_5}{3}\,\Omega\right)\,, \tag{7.126}$$

with similar expressions for the other quantities. All of the metric perturbation quantities are determined once solution is found for the wave equation

$$\left(\frac{1}{NA^3}\,\dot{\Omega}\right)^{\cdot} + \left(\frac{\Lambda_5}{6} + \frac{k^2}{A^2}\right)\frac{N}{A^3}\,\Omega = \left(\frac{N}{A^3}\,\Omega'\right)'\,. \tag{7.127}$$

The junction conditions (7.41) relate the off-brane derivatives of metric perturbations to the matter perturbations:

$$\partial_y\,\delta g_{\mu\nu} = -\kappa_5^2\left[\delta T_{\mu\nu} + \frac{1}{3}\left\{\lambda - T^{(0)}\right\}\delta g_{\mu\nu} - \frac{1}{3}g^{(0)}_{\mu\nu}\delta T\right]\,, \tag{7.128}$$

where

$$\delta T^0{}_0 = -\delta\rho\,,\ \delta T^0{}_i = a^2 q_i \tag{7.129}$$

$$\delta T^i{}_j = \delta p\,\delta^i{}_j + \delta\pi^i{}_j\,. \tag{7.130}$$

For scalar perturbations in the Gaussian normal gauge, this gives

$$\partial_y\psi(x,0) = \frac{\kappa_5^2}{6}(2\delta\rho + 3\delta p) \tag{7.131}$$

$$\partial_y\mathcal{B}(x,0) = \kappa_5^2\delta p \tag{7.132}$$

$$\partial_y\mathcal{C}(x,0) = -\frac{\kappa_5^2}{2}\delta\pi \tag{7.133}$$

$$\partial_y\mathcal{R}(x,0) = -\frac{\kappa_5^2}{6}\delta\rho - \partial_i\partial^i\,\mathcal{C}(x,0)\,, \tag{7.134}$$

where $\delta\pi$ is the scalar potential for the matter anisotropic stress,

$$\delta\pi_{ij} = \partial_i\partial_j\delta\pi - \frac{1}{3}\delta_{ij}\,\partial_k\partial^k\delta\pi\,. \tag{7.135}$$

The perturbed nonlocal (dark radiation) stress tensor on the brane is given by

$$\delta\mathcal{E}^0{}_0 = \kappa^2\delta\rho^*\,,\ \delta\mathcal{E}^0{}_i = -\kappa^2 a^2 q_i^* \tag{7.136}$$

$$\delta\mathcal{E}^i{}_j = -\frac{\kappa^2}{3}\delta\rho^*\,\delta^i{}_j + \delta\pi^{*i}{}_j\,. \tag{7.137}$$

The evolution of the bulk metric perturbations is determined by the perturbed 5D field equations in the vacuum bulk,

$$\delta\,{}^{(5)}G^A{}_B = 0\,. \tag{7.138}$$

Then the matter perturbations on the brane enter via the perturbed junction conditions, (7.128).

For example, for scalar perturbations in Gaussian normal gauge,

$$\delta\,{}^{(5)}G^y{}_i = \partial_i\left\{-\psi' + \left(\frac{A'}{A} - \frac{N'}{N}\right)\psi - 2\mathcal{R}' - \frac{A^2}{2N^2}\left[\dot{\mathcal{B}}' + \left(5\frac{\dot{A}}{A} - \frac{\dot{N}}{N}\right)\mathcal{B}'\right]\right\}\,. \tag{7.139}$$

For tensor perturbations (in any gauge), the only nonzero components of the perturbed Einstein tensor are

$$\delta\,{}^{(5)}G^i{}_j = -\frac{1}{2}\left\{-\frac{1}{N^2}\ddot{f}^i{}_j + f''^i{}_j - \frac{k^2}{A^2}f^i{}_j + \frac{1}{N^2}\left(\frac{\dot{N}}{N} - 3\frac{\dot{A}}{A}\right)\dot{f}^i{}_j + \left(\frac{N'}{N} + 3\frac{A'}{A}\right)f'^i{}_j\right\}\,. \tag{7.140}$$

7.6.2 Curvature Perturbations and the Sachs–Wolfe Effect

The curvature perturbation $\mathcal{R}$ on uniform density surfaces is defined in (7.122). The associated gauge-invariant quantity

$$\zeta = \mathcal{R} + \frac{\delta\rho}{3(\rho+p)}\,, \tag{7.141}$$

may be defined for matter on the brane. Similarly, for the Weyl "fluid" if $\rho^* \neq 0$ in the background, the curvature perturbation on hypersurfaces of uniform dark energy density is

$$\zeta^* = \mathcal{R} + \frac{\delta\rho^*}{4\rho^*}\,. \tag{7.142}$$

On large scales, the perturbed dark energy conservation equation is [21]

$$(\delta\rho^*)^{\cdot} + 4H\delta\rho^* + 4\rho^*\dot{\mathcal{R}} = 0\,, \tag{7.143}$$

which leads to

$$\dot{\zeta}^* = 0\,. \tag{7.144}$$

For adiabatic matter perturbations, by the perturbed matter energy conservation equation,

$$(\delta\rho)^{\cdot} + 3H(\delta\rho + \delta p) + 3(\rho + p)\dot{\mathcal{R}} = 0 \tag{7.145}$$

we find

$$\dot{\zeta} = 0\,. \tag{7.146}$$

This is independent of brane-world modifications to the field equations, since it depends on energy conservation only. For the total, effective fluid, the curvature perturbation is defined as follows [21]: if $\rho^* \neq 0$ in the background,

$$\zeta^{\rm tot} = \zeta + \left[\frac{4\rho^*}{3(\rho + p)(1 + \rho/\lambda) + 4\rho^*}\right](\zeta^* - \zeta)\,, \tag{7.147}$$

and if $\rho^* = 0$ in the background,

$$\zeta^{\rm tot} = \zeta + \frac{\delta\rho^*}{3(\rho + p)(1 + \rho/\lambda)} \tag{7.148}$$

$$\delta\rho^* = \frac{\delta C^*}{a^4}\,, \tag{7.149}$$

where δC^* is a constant. It follows that the curvature perturbations on large scales, like the density perturbations, can be found on the brane without solving for the bulk metric perturbations.

Note that $\dot{\zeta}^{\rm tot} \neq 0$ even for adiabatic matter perturbations; for example, if $\rho^* = 0$ in the background,

$$\dot{\zeta}^{\rm tot} = H\left(c_{\rm tot}^2 - \frac{1}{3}\right)\frac{\delta\rho^*}{(\rho + p)(1 + \rho/\lambda)}\,. \tag{7.150}$$

The KK effects on the brane contribute a non-adiabatic mode, although $\dot{\zeta}^{\rm tot} \to 0$ at low energies.

Although the density and curvature perturbations can be found on super-Hubble scales, the Sachs-Wolfe effect requires $\pi^*_{\mu\nu}$ in order to translate from density/ curvature to metric perturbations. In the 4D longitudinal gauge of the metric perturbation formalism, the gauge invariant curvature and metric perturbations on large scales are related by

$$\zeta^{\rm tot} = \mathcal{R} - \frac{H}{\dot{H}}\left(\frac{\dot{\mathcal{R}}}{H} - \psi\right) \tag{7.151}$$

$$\mathcal{R} + \psi = -\kappa^2 a^2 \delta\pi^*\,, \tag{7.152}$$

where the radiation anisotropic stress on large scales is neglected, as in general relativity, and $\delta\pi^*$ is the scalar potential for $\pi^*_{\mu\nu}$ (equivalent to the covariant quantity Π). In 4D general relativity, the right hand side of (7.152) is zero. The (non-integrated) Sachs-Wolfe formula has the same form as in general relativity:

$$\left.\frac{\delta T}{T}\right|_{\rm now} = (\zeta_{\rm rad} + \psi - \mathcal{R})|_{\rm dec}\,. \tag{7.153}$$

The brane-world corrections to the general relativistic Sachs-Wolfe effect are then given by [21]

$$\frac{\delta T}{T} = \left(\frac{\delta T}{T}\right)_{\rm gr} - \frac{8}{3}\left(\frac{\rho_{\rm rad}}{\rho_{\rm cdm}}\right) S^* - \kappa^2 a^2 \delta\pi^* + \frac{2\kappa^2}{a^{5/2}}\int da\; a^{7/2}\,\delta\pi^*\,, \tag{7.154}$$

where S^* is the KK entropy perturbation (determined by $\delta\rho^*$). The KK term $\delta\pi^*$ cannot be determined by the 4D brane equations, so that $\delta T/T$ cannot be evaluated on large scales without solving the 5D equations. Equation (7.154) has been generalized to the two-brane case, in which the radion makes a contribution to the Sachs-Wolfe effect [43].

The presence of the KK (Weyl, dark) component has essentially two possible effects.

- A contribution from the KK entropy perturbation S^* that is similar to an extra isocurvature contribution.
- The KK anisotropic stress $\delta\pi^*$ also contributes to the CMB anisotropies. In the absence of anisotropic stresses, the curvature perturbation $\zeta^{\rm tot}$ is sufficient to determine the metric perturbation $\mathcal{R}$ and hence the large angle CMB anisotropies, via (7.151), (7.152) and (7.153). However bulk gravitons can also generate anisotropic stresses which, although they do not affect the large-scale curvature perturbation $\zeta^{\rm tot}$, can affect the relation between $\zeta^{\rm tot}$, $\mathcal{R}$ and ψ, and hence can affect the CMB anisotropies at large angles.

7.7 Gravitational Wave Perturbations

The tensor perturbations are given by (7.122), *i.e.* (for a flat background brane),

$${}^{(5)}ds^2 = -N^2(t,y)dt^2 + A^2(t,y)\left[\delta_{ij} + f_{ij}\right]dx^i dx^j + dy^2\,. \tag{7.155}$$

The transverse traceless f_{ij} satisfies (7.140), which implies, on splitting f_{ij} into Fourier modes with amplitude $f(t,y)$,

$$\frac{1}{N^2}\left[\ddot{f} + \left(3\frac{\dot{A}}{A} - \frac{\dot{N}}{N}\right)\dot{f}\right] + \frac{k^2}{A^2}f = f'' + \left(3\frac{A'}{A} + \frac{N'}{N}\right)f'\,. \tag{7.156}$$

By the transverse traceless part of (7.128), the boundary condition is

$$f'_{ij}\big|_{\rm brane} = \bar{\pi}_{ij}\,, \tag{7.157}$$

where $\bar{\pi}_{ij}$ is the tensor part of the anisotropic stress of matter-radiation on the brane.

The wave equation (7.156) cannot be solved analytically except if the background metric functions are separable, and this only happens for maximally symmetric branes, *i.e.* branes with constant Hubble rate H_0. This includes the RS case $H_0 = 0$, already treated above. The cosmologically relevant case is the de Sitter brane, $H_0 > 0$. We can calculate the spectrum of gravitational waves generated during brane inflation [33, 34, 44, 45], if we approximate slow-roll inflation by a succession of de Sitter phases. The metric for a de Sitter brane, dS_4, in AdS_5 is given by (7.85)–(7.87) with

$$N(t,y) = n(y)\,, \quad A(t,y) = a(t)n(y) \tag{7.158}$$

$$n(y) = \cosh\mu y - \left(1+\frac{\rho_0}{\lambda}\right)\sinh\mu|y| \tag{7.159}$$

$$a(t) = a_0 \exp H_0(t-t_0)\,, \quad H_0^2 = \frac{\kappa^2}{3}\rho_0\left(1+\frac{\rho_0}{2\lambda}\right)\,, \tag{7.160}$$

where $\mu = \ell^{-1}$.

The linearized wave equation (7.156) is separable. As before, we separate the amplitude as $f = \sum \varphi_m(t) f_m(y)$ where m is the 4D mass, and this leads to:

$$\ddot{\varphi}_m + 3H_0\dot{\varphi}_m + \left[m^2 + \frac{k^2}{a^2}\right]\varphi_m = 0 \tag{7.161}$$

$$f''_m + 4\frac{n'}{n}f'_m + \frac{m^2}{n^2}f_m = 0\,. \tag{7.162}$$

The general solutions for $m > 0$ are

$$\varphi_m(t) = \exp\left(-\frac{3}{2}H_0 t\right) B_\nu\left(\frac{k}{H_0}e^{-H_0 t}\right) \tag{7.163}$$

$$f_m(y) = n(y)^{-3/2} L^\nu_{3/2}\left(\sqrt{1+\frac{\mu^2}{H_0^2}n(y)^2}\right)\,, \tag{7.164}$$

where B_ν is a linear combination of Bessel functions, $L^\nu_{3/2}$ is a linear combination of associated Legendre functions, and

$$\nu = i\sqrt{\frac{m^2}{H_0^2} - \frac{9}{4}}\,. \tag{7.165}$$

It is more useful to reformulate (7.162) as a Schrödinger-type equation,

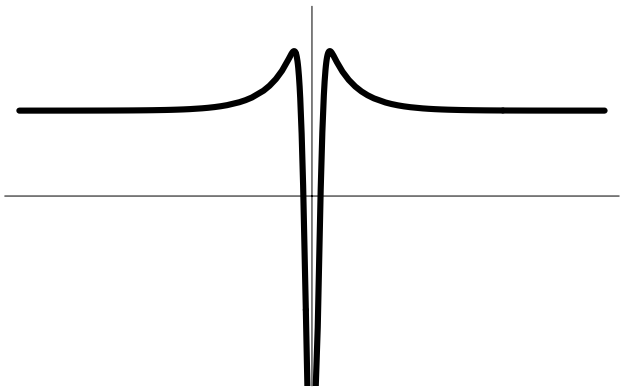

Fig. 7.3. The graviton "volcano" potential around the dS_4 brane, showing the mass gap (From [11]).

$$\frac{d^2\Psi_m}{dz^2} - V(z)\Psi_m = -m^2\Psi_m\,, \tag{7.166}$$

using the conformal coordinate

$$z = z_{\rm b} + \int_0^y \frac{d\tilde{y}}{n(\tilde{y})}\,, \quad z_{\rm b} = \frac{1}{H_0}\sinh^{-1}\left(\frac{H_0}{\mu}\right)\,, \tag{7.167}$$

and defining $\Psi_m \equiv n^{3/2} f_m$. The potential is given by (see Fig. 7.3)

$$V(z) = \frac{15H_0^2}{4\sinh^2(H_0 z)} + \frac{9}{4}H_0^2 - 3\mu\left[1 + \frac{\rho_0}{\lambda}\right]\delta(z - z_{\rm b})\,. \tag{7.168}$$

The "volcano" shape of the potential shows how the 5D graviton is localized at the brane at low energies.

The non-zero value of the Hubble parameter implies the existence of a mass gap [16],

$$\Delta m = \frac{3}{2}H_0\,, \tag{7.169}$$

between the zero mode and the continuum of massive KK modes. This result has been generalized: for dS_4 brane(s) with bulk scalar field, a universal lower bound on the mass gap of the KK tower is [44] $\Delta m \geq \sqrt{3/2}\,H_0$.

The massive modes decay during inflation, according to (7.163), leaving only the zero mode, which is effectively a 4D gravitational wave. The zero mode, satisfying the boundary condition, $f_0'(x,0) = 0$, is given by

$$f_0 = \sqrt{\mu}\; F(H_0/\mu)\;, \tag{7.170}$$

where the normalization condition

$$2\int_{z_{\rm b}}^{\infty} |\Psi_0^2| dz = 1\,, \tag{7.171}$$

implies that the function F is given by

$$F(x) = \left\{ \sqrt{1+x^2} - x^2 \ln \left[\frac{1}{x} + \sqrt{1+\frac{1}{x^2}} \right] \right\}^{-1/2} . \tag{7.172}$$

At low energies ($H_0 \ll \mu$), we recover the general relativity amplitude: $F \to 1$. At high energies, the amplitude is considerably enhanced:

$$H_0 \gg \mu \;\Rightarrow\; F \approx \sqrt{\frac{3H_0}{2\mu}} . \tag{7.173}$$

The factor F determines the modification of the gravitational wave amplitude relative to the standard 4D result:

$$A_{\rm t}^2 = \left[\frac{8}{M_{\rm p}^2} \left(\frac{H_0}{2\pi} \right)^2 \right] F^2(H_0/\mu). \tag{7.174}$$

This enhanced zero mode produced by brane inflation remains frozen outside the Hubble radius, as in general relativity, but when it re-enters the Hubble radius during radiation or matter domination, it will no longer be separated from the massive modes, since H will not be constant. Instead, massive modes will be excited during re-entry. In other words, energy will be lost from the zero mode as 5D gravitons are emitted into the bulk, *i.e.* as massive modes are produced on the brane. Self-consistent low energy approximations to compute this effect are developed in [39, 40].

At zero order, the low-energy approximation is based on the following. In the radiation era, at low energy, the background metric functions obey $A(t,y) \to a(t)e^{-\mu y}$ and $N(t,y) \to e^{-\mu y}$. To lowest order, the wave equation therefore separates, and the mode functions can be found analytically. The massive modes in the bulk, $f_m(y)$, are the same as for a Minkowski brane. The massive modes decay on super-Hubble scales, unlike the zero-mode. Expanding the wave equation in ρ_0/λ, one arrives at the first order, where mode-mixing arises. The numerical integration of the equations [39] confirms the effect of massive mode generation and consequent damping of the zero-mode, as shown in Fig. 7.4.

7.8 Brane-World CMB Anisotropies

Recently, the anisotropies in the CMB for RS-type brane-world cosmologies have been calculated using a low-energy approximation [37] (for a review on CMB anisotropies see A. Challinor's contribution). The basic idea of the low-energy approximation [36] is to use a gradient expansion to exploit the fact that, during most of the history of the universe, the curvature scale on

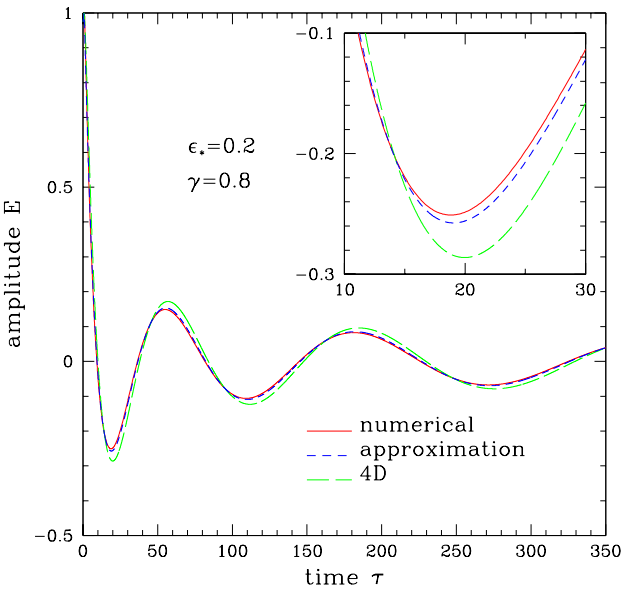

Fig. 7.4. The damping of cosmological gravity waves on horizon re-entry due to massive mode generation. The solid curve is the numerical solution, the short-dashed curve the low-energy approximation, and the long-dashed curve the standard general relativity solution. $\epsilon_* = \rho_0/\lambda$ and γ is a parameter giving the location of the regulator brane. (From [39]).

the observable brane is much greater than the curvature scale of the bulk ($\ell < 1$ mm):

$$L \sim |R_{\mu\nu\alpha\beta}|^{-1/2} \gg \ell \sim |\,{}^{(5)}R_{ABCD}|^{-1/2} \tag{7.175}$$

$$\Rightarrow |\nabla_\mu| \sim L^{-1} \ll |\partial_y| \sim \ell^{-1}\,. \tag{7.176}$$

These conditions are equivalent to the low energy regime, since $\ell^2 \propto \lambda^{-1}$ and $|R_{\mu\nu\alpha\beta}| \sim |T_{\mu\nu}|$:

$$\frac{\ell^2}{L^2} \sim \frac{\rho}{\lambda} \ll 1\,. \tag{7.177}$$

Using (7.176) to neglect appropriate gradient terms in an expansion in ℓ^2/L^2, the low-energy equation

$$\nabla^\nu \mathcal{E}_{\mu\nu} = 0\,, \tag{7.178}$$

can be solved. However, two boundary conditions are needed to determine all functions of integration. This is achieved by introducing a second brane, as in the RSI scenario. This brane is to be thought of either as a regulator brane, whose backreaction on the observable brane is neglected (which will only be true for a limited time), or as a shadow brane with physical fields, which have a gravitational effect on the observable brane.

The background is given by low energy FRW branes with tensions $\pm\lambda$, proper times $t_\pm$, scale factors $a_\pm$, energy densities $\rho_\pm$ and pressures $p_\pm$, and dark radiation densities $\rho^*_\pm$. The physical distance between the branes is $\ell\bar{d}(t)$, and

$$\frac{d}{dt_-} = e^{\bar{d}}\,\frac{d}{dt_+}\,,\ a_- = a_+e^{-\bar{d}}\,,\ H_- = e^{\bar{d}}\left(H_+ - \dot{\bar{d}}\right)\,,\ \rho^*_- = e^{4\bar{d}}\rho^*_+\,. \quad (7.179)$$

Then the background dynamics are given by

$$H^2_\pm = \pm\frac{\kappa^2}{3}\left(\rho_\pm \pm \rho^*_\pm\right) \quad (7.180)$$

$$\ddot{\bar{d}} + 3H_+\dot{\bar{d}} - \dot{\bar{d}}^2 = \frac{\kappa^2}{6}\left[\rho_+ - 3p_+ + e^{2\bar{d}}(\rho_- - 3p_-)\right]\,. \quad (7.181)$$

The dark energy obeys $\rho^*_+ = C/a^4_+$, where C is a constant. From now on, we drop the $(+)$ subscripts which refer to the physical, observed quantities.

The perturbed metric on the observable (positive tension) brane is described, in longitudinal gauge, by the metric perturbations ψ and $\mathcal{R}$, and the perturbed radion is $d = \bar{d} + N$. The approximation for the KK (Weyl) energy-momentum tensor on the observable brane is

$$\mathcal{E}^\mu{}_\nu = \frac{2}{e^{2d}-1}\left[-\frac{\kappa^2}{2}\left(T^\mu_\nu + e^{-2d}T^\mu_-{}_\nu\right)\right.$$
$$\left. - \nabla^\mu\nabla_\nu d + \delta^\mu_\nu\nabla^2 d - \left\{\nabla^\mu d\nabla_\nu d + \frac{1}{2}\delta^\mu_\nu(\nabla d)^2\right\}\right], \quad (7.182)$$

and the field equations on the observable brane can be written in scalar-tensor form as

$$G^\mu{}_\nu = \frac{\kappa^2}{\chi}T^\mu_\nu + \frac{\kappa^2(1-\chi)^2}{\chi}T^\mu_-{}_\nu + \frac{1}{\chi}\left(\nabla^\mu\nabla_\nu\chi - \delta^\mu_\nu\nabla^2\chi\right)$$
$$+ \frac{\omega(\chi)}{\chi^2}\left[\nabla^\mu\chi\nabla_\nu\chi - \frac{1}{2}\delta^\mu_\nu(\nabla\chi)^2\right]\,, \quad (7.183)$$

where

$$\chi = 1 - e^{-2d}\,,\ \omega(\chi) = \frac{3}{2}\frac{\chi}{1-\chi}\,. \quad (7.184)$$

The perturbation equations can then be derived as generalizations of the standard equations. For example, the $\delta G^0{}_0$ equation is [46]

$$H^2\psi - H\dot{\mathcal{R}} - \frac{1}{3}\frac{k^2}{a^2}\mathcal{R}$$
$$= -\frac{1}{6}\kappa^2\frac{e^{2\bar{d}}}{e^{2\bar{d}}-1}\left(\delta\rho + e^{-4\bar{d}}\delta\rho_-\right) + \frac{2}{3}\kappa^2\frac{e^{2\bar{d}}}{e^{2\bar{d}}-1}\rho^*\,N$$
$$- \frac{1}{e^{2\bar{d}}-1}\left[\left(\dot{\bar{d}} - H\right)\dot{N} + \left(\dot{\bar{d}} - H\right)^2 N\right.$$
$$\left. - \dot{\bar{d}}^2\psi + 2H\dot{\bar{d}}\psi - \dot{\bar{d}}\dot{\mathcal{R}} - \frac{1}{3}\frac{k^2}{a^2}N\right]. \quad (7.185)$$

The trace part of the perturbed field equation shows that the radion perturbation determines the crucial quantity, $\delta\pi^*$:

$$\mathcal{R}+\psi=-\frac{2}{e^{2\bar{d}}-1}N=-\kappa^2a^2\delta\pi^*\,, \tag{7.186}$$

where the last equality follows from (7.152). The radion perturbation itself satisfies the wave equation

$$\begin{aligned}\ddot{N}+\left(3H-2\dot{\bar{d}}\right)\dot{N}-\left(2\dot{H}+4H^2+2\dot{\bar{d}}^2-6H\dot{\bar{d}}-2\ddot{\bar{d}}\right)N+\frac{k^2}{a^2}N\\ -\dot{\bar{d}}\dot{\psi}+3\dot{\bar{d}}\dot{\mathcal{R}}+\left(-2\ddot{\bar{d}}-6H\dot{\bar{d}}+2\dot{\bar{d}}^2\right)\psi\\ =\frac{\kappa^2}{6}\left[\delta\rho-3\delta p+e^{-2\bar{d}}(\delta\rho_--3\delta p_-)\right].\end{aligned} \tag{7.187}$$

A new set of variables $\varphi_\pm,\ E$ turns out be very useful [43, 37]:

$$\begin{aligned}\mathcal{R}&=-\varphi_+-\frac{a^2}{k^2}H\dot{E}+\frac{1}{3}E\\ \psi&=-\varphi_+-\frac{a^2}{k^2}(\ddot{E}+2H\dot{E})\\ N&=\varphi_--\varphi_+-\frac{a^2}{k^2}\dot{\bar{d}}\dot{E}\,.\end{aligned} \tag{7.188}$$

Equation (7.186) gives

$$\ddot{E}+\left(3H+\frac{2\dot{\bar{d}}}{e^{2\bar{d}}-1}\right)\dot{E}-\frac{1}{3}\frac{k^2}{a^2}E=-\frac{2e^{2\bar{d}}}{e^{2\bar{d}}-1}\frac{k^2}{a^2}\left(\varphi_+-e^{-2\bar{d}}\varphi_-\right)\,. \tag{7.189}$$

The variable E determines the metric shear in the bulk, whereas $\varphi_\pm$ give the brane displacements, in transverse traceless gauge. The latter variables have a simple relation to the curvature perturbations on large scales [43, 37] (restoring the (+) subscripts):

$$\zeta_\pm^{\rm tot}=-\varphi_\pm+\frac{H_\pm^2}{\dot{H}_\pm}\left(\frac{\dot{\varphi}_\pm}{H_\pm}+\varphi_\pm\right)\,, \tag{7.190}$$

where $\dot{f}_\pm\equiv df_\pm/dt_\pm$.

The simplest model is the one in which

$$\rho^*=0=\dot{\bar{d}} \tag{7.191}$$

in the background, with $p_-/\rho_-=p/\rho$. By (7.181), it follows that

$$\rho_-=-\rho e^{2\bar{d}}\,, \tag{7.192}$$

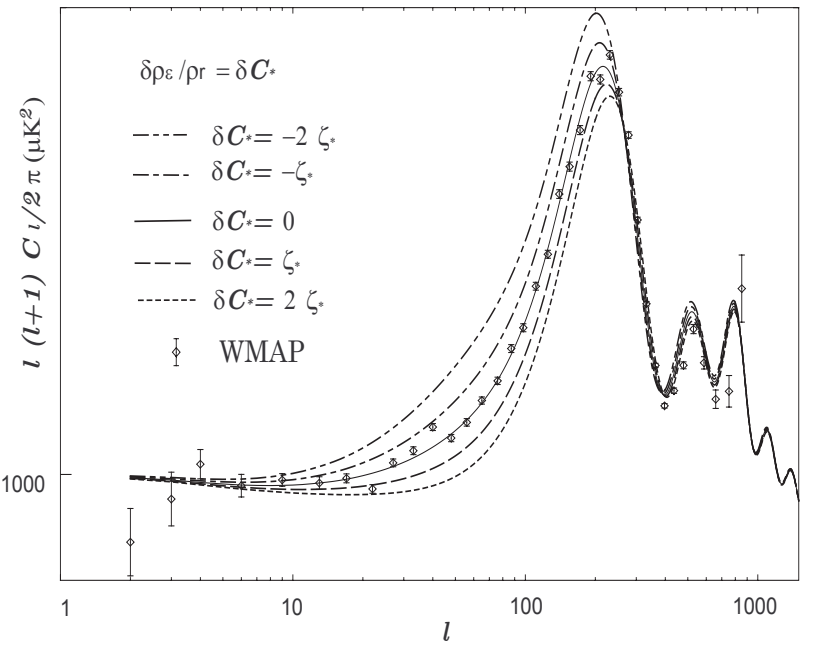

Fig. 7.5. The CMB power spectrum with brane-world effects, encoded in the dark radiation fluctuation parameter δC_* as a proportion of the large-scale curvature perturbation for matter (denoted ζ_* in the plot). The dark radiation perturbation $\delta\rho^*$ is denoted $\delta\rho_\varepsilon$. (From [37]).

i.e. the shadow matter must have fine-tuned and negative energy density to prevent the shadow brane from moving in the background. With these assumptions, and further assuming adiabatic perturbations for the matter, there is only one independent brane-world parameter, *i.e.* the parameter measuring dark radiation fluctuations:

$$\delta C_* = \frac{\delta\rho^*}{\rho_{\rm rad}}\,. \tag{7.193}$$

This assumption has a remarkable consequence on large scales: the Weyl anisotropic stress $\delta\pi^*$ terms in the Sachs-Wolfe formula (7.154) cancel the entropy perturbation from dark radiation fluctuations, so that there is no difference on the largest scales from the standard general relativity power spectrum. On small scales, beyond the first acoustic peak, the brane-world corrections are negligible. On scales up to the first acoustic peak, brane-world effects can be significant, changing the height and the location of the first peak. These features are apparent in Fig. 7.5. However, it is not clear to what extent these features are general brane-world features (within the low energy approximation), and to what extent they are consequences of

the simple assumptions imposed on the background. Further work remains to be done. (A related low energy approximation, using the moduli space approximation, has been developed for certain two-brane models with bulk scalar field [38].)

7.9 Conclusions

Simple brane-world models of RS type provide a rich phenomenology for exploring some of the ideas that are emerging from M theory. These models are the simplest brane-worlds with curved extra dimension that allow for a meaningful approach to cosmology. There are generalizations that attempt to make these models more realistic, or that explore other aspects of higher-dimensional gravity which are not probed by these simple models. Important examples are:

- The inclusion of dynamical interaction between the brane(s) and a bulk scalar field, described by the action [19, 23]

$$S = \int d^5x\sqrt{-{}^{(5)}g}\left[\frac{{}^{(5)}R}{2\kappa_5^2} - \frac{1}{2}\partial_A\Phi\partial^A\Phi - \Lambda_5(\Phi)\right] + \int_{\text{brane(s)}} d^4x\sqrt{-g}\left[-\lambda(\Phi) + \frac{K}{\kappa_5^2} + L_{\text{matter}}\right]. \quad (7.194)$$

  In two-brane models, the brane separation is described by the radion. For general potentials of the scalar field which provide radion stabilization, 4D Einstein gravity is recovered at low energies on either brane. (By contrast, in the absence of a bulk scalar, low energy gravity is Brans-Dicke type [13].)
  In particular, such models will allow some fundamental problems to be addressed: the hierarchy problem of particle physics; an extra-dimensional mechanism for initiating inflation (or the hot radiation era with super-Hubble correlations) via brane interaction (building on the initial work in [24, 25]); an extra-dimensional explanation for the dark energy (and possibly also dark matter) puzzles: could dark energy or late time acceleration of the universe be a result of gravitational effects on the visible brane of the shadow brane, mediated by the bulk dilaton?
- Moduli fields that encode the influence on the brane of further extra dimensions (for example, the six extra dimensions in Horava-Witten type solutions, in addition to the large extra dimension) [38, 47].
- Higher-order curvature invariants, which arise in the AdS/CFT correspondence as next-to-leading order corrections in the CFT. These may be thought of as stringy corrections to the Eisntein-Hilbert action. The Gauss-Bonnet combination in particular has unique properties in 5D, giving field equations which are second-order in the bulk metric (and linear in the second derivatives), and being ghost-free. The action is

$$S = \int d^5x\sqrt{-{}^{(5)}g}\left[\frac{{}^{(5)}R}{2\kappa_5^2} - \Lambda_5 \right.$$
$$\left. + \alpha\left\{{}^{(5)}R^2 - 4{}^{(5)}R_{AB}{}^{(5)}R^{AB} + {}^{(5)}R_{ABCD}{}^{(5)}R^{ABCD}\right\}\right]$$
$$+ \int_{\rm brane} d^4x\sqrt{-g}\left[-\lambda + \frac{K}{\kappa_5^2} + L_{\rm matter}\right], \qquad (7.195)$$

where α is the Gauss-Bonnet coupling constant, related to the string scale. The cosmological dynamics of these brane-worlds is investigated in [48]. In the early universe, the Gauss-Bonnet corrections to the Friedmann equation have the dominant form

$$H^2 \propto \rho^{2/3}, \qquad (7.196)$$

so that the high energy behaviour $H^2 \propto \rho^2$ of the RS brane-world is radically modified (although the big bang singularity is not removed).

Quantum field theory corrections to the action, arising from the coupling between brane matter and bulk gravitons, and leading to an induced 4D Ricci term in the brane action. The original induced-gravity brane-world [49] was put forward as an alternative to the RS mechanism: the bulk was flat Minkowski 5D spacetime (and as a consequence there is no normalizable zero-mode of the bulk graviton), and there was no brane tension. Another viewpoint is to see the induced-gravity term in the action as a correction to the RS action:

$$S = \int d^5x\sqrt{-{}^{(5)}g}\left[\frac{{}^{(5)}R}{2\kappa_5^2} - \Lambda_5\right]$$
$$+ \int_{\rm brane} d^4x\sqrt{-g}\left[\frac{R}{2\kappa^2} - \lambda + \frac{K}{\kappa_5^2} + L_{\rm matter}\right]. \qquad (7.197)$$

Unlike the other brane-worlds discussed, these models lead to 5D behaviour on large scales rather than small scales. The cosmological models have been analysed in [50]. The late-universe 5D behaviour of gravity can naturally produce a late-time acceleration, even without dark energy, although the fine-tuning problem is not evaded.

The effect of the induced-gravity correction at early times is to remove the high energy ρ^2 term in the early universe Friedmann equation; instead, the Friedmann equation is close the standard general relativity form.

Thus we have a striking result that both forms of correction to the gravitational action, *i.e.* Gauss-Bonnet and induced gravity, suppress the Randall-Sundrum type high energy modifications to the Friedmann equation. (Cosmologies with both induced-gravity and Gauss-Bonnet corrections to the RS action are considered in [51].)

Brane-world gravity opens up exciting prospects for subjecting M theory ideas to the increasingly stringent tests provided by high-precision astronomical observations. In addition, brane-world models provide a rich arena for

probing the geometry and dynamics of the gravitational field and its interaction with matter.

Acknowledgments

I thank the organisers for a very enjoyable and stimulating summer school. I am supported by PPARC.

References

1. For recent reviews, see, e.g., J.H. Schwarz, astro-ph/0304507; J. Polchinski, hep-th/0209105; R. Kallosh, hep-th/0205315.
2. M. Cavaglia, Int. J. Mod. Phys. **A18**, 1843 (2003), hep-ph/0210296.
3. J. Hewett, M. Spiropulu, Ann. Rev. Nucl. Part. Sci. **52**, 397 (2002), hep-ph/0205106.
4. P. Horava, E. Witten, Nucl. Phys. **B460**, 506 (1996), hep-th/9510209.
5. A. Lukas, B.A. Ovrut, K.S. Stelle, D. Waldram, Phys. Rev. D**59**, 086001 (1999), hep-th/9803235; A. Lukas, B.A. Ovrut, D. Waldram, Phys. Rev. D**60**, 086001 (1999), hep-th/9806022; ibid., **61**, 023506 (2000), hep-th/9902071.
6. L. Randall, R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999), hep-ph/9905221.
7. L. Randall, R. Sundrum, Phys. Rev. Lett. **83**, 4690 (1999), hep-th/9906064.
8. N. Arkani-Hamed, S. Dimopoulos, G. Dvali, Phys. Lett. **B429**, 263 (1998), hep-ph/9803315; I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos, G. Dvali, Phys. Lett. **B436**, 257 (1998), hep-ph/9804398.
9. R. Maartens, gr-qc/0312059.
10. P. Brax, C. van de Bruck, Class. Quantum Grav. **20**, R201 (2003), hep-th/0303095; N. Deruelle, Astrophys. Space Sci. **283**, 619 (2003), gr-qc/0301035; S. Räsänen, astro-ph/0208282; D. Langlois, Prog. Theor. Phys. Suppl. **148**, 181 (2002), hep-th/0209261; E. Papantonopoulos, Lect. Notes Phys. **592**, 458 (2002), hep-th/0202044; D. Langlois, gr-qc/0207047; D. Wands, Class. Quantum Grav. **19**, 3403 (2002), hep-th/0203107; D. Langlois, Int. J. Mod. Phys. A**17**, 2701 (2002), gr-qc/0205004; J. March-Russell, hep-ph/0012151.
11. D. Langlois, astro-ph/0301021.
12. W.D. Goldberger, M.B. Wise, Phys. Rev. Lett. **83**, 4922 (1999), hep-ph/9907447.
13. J. Garriga, T. Tanaka, Phys. Rev. Lett. **84**, 2778 (2000), hep-th/9911055.
14. P. Binetruy, C. Deffayet, U. Ellwanger, D. Langlois, Phys. Lett. **B477**, 285 (2000), hep-th/9910219.
15. P. Kraus, JHEP **12**, 011 (1999), hep-th/9910149; N. Kaloper, Phys. Rev. D**60**, 123506 (1999), hep-th/9905210; A. Kehagias, E. Kiritsis, JHEP **11**, 022 (1999), hep-th/9910174; S.S. Gubser, Phys. Rev. D**63** 084017 (2001), hep-th/9912001; S. Mukohyama, Phys. Lett. **B473**, 241 (2000), hep-th/9911165; D. Ida, JHEP **09**, 014 (2000), gr-qc/9912002; E.E. Flanagan, S.-H. Henry Tye, I. Wasserman, Phys. Rev. D **62**, 044039 (2000), hep-ph/9910498.
16. J. Garriga, M. Sasaki, Phys. Rev. D**62**, 043523 (2000), hep-th/9912118.

17. T. Shiromizu, K. Maeda, M. Sasaki, Phys. Rev. D**62**, 024012 (2000), gr-qc/9910076.
18. S. Mukohyama, T. Shiromizu, K. Maeda, Phys. Rev. D**62**, 024028 (2000), hep-th/9912287; P. Bowcock, C. Charmousis, R. Gregory, Class. Quantum Grav. **17**, 4745 (2000), hep-th/0007177.
19. K. Maeda, D. Wands, Phys. Rev. D**62**, 124009 (2000), hep-th/0008188.
20. R. Maartens, Phys. Rev. D**62**, 084023 (2000), hep-th/0004166.
21. D. Langlois, R. Maartens, M. Sasaki, D. Wands, Phys. Rev. D**63**, 084009 (2001), hep-th/0012044.
22. J.D. Barrow, R. Maartens, Phys. Lett. **B532**, 153 (2002), gr-qc/0108073; K. Ichiki, M. Yahiro, T. Kajino, M. Orito, G.J. Mathews, Phys. Rev. D**66**, 043521 (2002), astro-ph/0203272; J.D. Bratt, A.C. Gault, R.J. Scherrer, T.P. Walker, Phys. Lett. **B546**, 19 (2002), astro-ph/0208133.
23. S. Kobayashi, K. Koyama, J. Soda, Phys. Lett. **B501**, 157 (2001), hep-th/0009160; Y. Himemoto, M. Sasaki, Phys. Rev. D**63**, 044015 (2001), gr-qc/0010035; E.E. Flanagan, S.-H. Henry Tye, I. Wasserman, Phys. Lett. **B522**, 155 (2001), hep-th/0110070; N. Sago, Y. Himemoto, M. Sasaki, Phys. Rev. D**65**, 024014 (2002), gr-qc/0104033; Y. Himemoto, T. Tanaka, M. Sasaki, Phys. Rev. D**65**, 104020 (2002), gr-qc/0112027; Y. Himemoto, T. Tanaka, Phys. Rev. D**67**, 084014 (2003), gr-qc/0212114; T. Tanaka, Y. Himemoto, Phys. Rev. D**67**, 104007 (2003), gr-qc/0301010; B. Wang, L-H. Xue, X. Zhang, W-Y.P. Hwang, Phys. Rev. D**67**, 123519 (2003), hep-th/0301072; K. Koyama, K. Takahashi, Phys. Rev. D**67**, 103503 (2003), hep-th/0301165; D. Langlois, M. Sasaki, Phys. Rev. D**68**, 064012 (2003), hep-th/0302069; Y. Himemoto, M. Sasaki, Prog. Theor. Phys. Suppl. **148**, 235 (2002), gr-qc/0302054; R.H. Brandenberger, G. Geshnizjani, S. Watson, Phys. Rev. D**67**, 123510 (2003), hep-th/0302222; M. Minamitsuji, Y. Himemoto, M. Sasaki, Phys. Rev. D**68**, 024016 (2003), gr-qc/0303108; S. Kanno, J. Soda, hep-th/0303203; J. Martin, G.N. Felder, A.V. Frolov, M. Peloso, L. Kofman, hep-th/0309001; A.V. Frolov, L. Kofman, hep-th/0309002; P.R. Ashcroft, C. van de Bruck, A.-C. Davis, astro-ph/0310643.
24. G. Dvali, S.-H.H. Tye, Phys. Lett. **B450**, 72 (1999), hep-ph/9812483; S. Kanno, M. Sasaki, J. Soda, Prog. Theor. Phys. **109**, 357 (2003), hep-th/0210250.
25. J. Khoury, B.A. Ovrut, P.J. Steinhardt, N. Turok, Phys. Rev. D**64**, 123522 (2001), hep-th/0103239; R. Kallosh, L. Kofman, A. Linde, Phys. Rev. D**64**, 123523 (2001), hep-th/0104073; A. Neronov, JHEP **11**, 007 (2001), hep-th/0109090; P.J. Steinhardt, N. Turok, Phys. Rev. D**65**, 126003 (2002), hep-th/0111098; D. Langlois, K. Maeda, D. Wands, Phys. Rev. Lett. **88**, 181301 (2002), gr-qc/0111013; N.E. Mavromatos, hep-th/0210008; A.J. Tolley, N. Turok, P.J. Steinhardt, hep-th/0306109.
26. R. Maartens, D. Wands, B.A. Bassett, I.P.C. Heard, Phys. Rev. D**62**, 041301 (2000), astro-ph/0009131.
27. N. Kaloper, Phys. Rev. D**60**, 123506 (1999), hep-th/9905210; J.M. Cline, C. Grojean, G. Servant, Phys. Rev. Lett. **83**, 4245 (1999), hep-ph/9906523; H. Stoica, S.-H. Henry Tye, I. Wasserman, Phys. Lett. **B482**, 205 (2000), hep-th/0004126; L. Mendes, A.R. Liddle, Phys. Rev. D**62**, 103511 (2000), astro-ph/0006020; A. Mazumdar, Phys. Rev. D**64**, 027304 (2001), hep-ph/0007269; S.C. Davis, W.B. Perkins, A.-C. Davis, I.R. Vernon, Phys. Rev. D**63**, 083518 (2001), hep-ph/0012223; A.R. Liddle, A.N. Taylor, Phys. Rev. D**65**, 041301 (2002), astro-ph/0109412; M.C. Bento, O. Bertolami, Phys. Rev. D**65**, 063513

(2002), astro-ph/0111273; M.C. Bento, O. Bertolami, A.A. Sen, Phys. Rev. D**67**, 023504 (2003), gr-qc/0204046; ibid., 063511 (2003), hep-th/0208124; S. Mizuno, K. Maeda, K. Yamamoto, Phys. Rev. D**67**, 024033 (2003), hep-ph/0205292; R. Hawkins, J.E. Lidsey, Phys. Rev. D**68**, 083505 (2003), astro-ph/0306311; K.E. Kunze, hep-th/0310200.
28. E.J. Copeland, A.R. Liddle, J.E. Lidsey, Phys. Rev. D**64**, 023509 (2001), astro-ph/0006421; A. S. Majumdar, Phys. Rev. D**64**, 083503 (2001), astro-ph/0105518; V. Sahni, M. Sami, T. Souradeep, Phys. Rev. D**65**, 023518 (2002), gr-qc/0105121; N.J. Nunes, E.J. Copeland, Phys. Rev. D**66**, 043524 (2002), astro-ph/0204115; A.R. Liddle, L.A. Urena-Lopez, Phys. Rev. D**68**, 043517 (2003), astro-ph/0302054.
29. G. Huey, J.E. Lidsey, Phys. Lett. **B514**, 217 (2001), astro-ph/0104006.
30. A. Albrecht, C.P. Burgess, F. Ravndal, C. Skordis, Phys. Rev. D**65**, 123507 (2002), astro-ph/0107573; S. Mizuno, K. Maeda, Phys. Rev. D**64**, 123521 (2001), hep-ph/0108012; J.E. Lidsey, T. Matos, L.A. Urena-Lopez, Phys. Rev. D**66**, 023514 (2002), astro-ph/0111292; C.P. Burgess, astro-ph/0207174; D. Seery, B.A. Bassett, astro-ph/0310208.
31. D. Wands, K. A. Malik, D. H. Lyth, A. R. Liddle, Phys. Rev. D**62**, 043527 (2000), astro-ph/0003278.
32. A.R. Liddle, A.J. Smith, Phys. Rev. D**68**, 061301 (2003), astro-ph/0307017.
33. D. Langlois, R. Maartens, D. Wands, Phys. Lett. **B489**, 259 (2000), hep-th/0006007.
34. D.S. Gorbunov, V.A. Rubakov, S.M. Sibiryakov, JHEP **10**, 15 (2001), hep-th/0108017.
35. H.A. Bridgman, K.A. Malik, D. Wands, Phys. Rev. D**63**, 084012 (2001), hep-th/0010133.
36. J. Soda, S. Kanno, Phys. Rev. D**66**, 083506 (2002), hep-th/0207029; T. Wiseman, Class. Quantum Grav. **19**, 3083 (2002), hep-th/0201127; T. Shiromizu, K. Koyama, Phys. Rev. D**67**, 084022 (2003), hep-th/0210066; J. Soda, S. Kanno, Astrophys. Space Sci. **283**, 639 (2003), gr-qc/0209086; ibid., hep-th/0303203.
37. K. Koyama, Phys. Rev. Lett., in press (2003), astro-ph/0303108.
38. C.S. Rhodes, C. van de Bruck, Ph. Brax, A.C. Davis, Phys. Rev. D**68**, 083511 (2003), astro-ph/0306343; P. Brax, C. van de Bruck, A.-C. Davis, C.S. Rhodes, hep-ph/0309181.
39. T. Hiramatsu, K. Koyama, A. Taruya, hep-th/0308072.
40. R. Easther, D. Langlois, R. Maartens, D. Wands, JCAP **10**, 014 (2003), hep-th/0308078.
41. S. Mukohyama, Phys. Rev. D**62**, 084015 (2000), hep-th/0004067; S. Mukohyama, Phys. Rev. D**64**, 064006 (2001), hep-th/0104185.
42. H.A. Bridgman, K.A. Malik, D. Wands, Phys. Rev. D**65**, 043502 (2002), astro-ph/0107245.
43. K. Koyama, Phys. Rev. D**66**, 084003 (2002), gr-qc/0204047.
44. A. Frolov, L. Kofman, hep-th/0209133.
45. T. Kobayashi, H. Kudoh, T. Tanaka, Phys. Rev. D**68**, 044025 (2003), gr-qc/0305006.
46. K. Koyama, unpublished notes.
47. R. Brustein, S.P. de Alwis, E.G. Novak, Phys. Rev. D**68**, 043507 (2003), hep-th/0212344; L. Kofman, astro-ph/0303614; S. Kachru, R. Kallosh, A. Linde, J. Maldacena, L. McAllister, S.P. Trivedi, JCAP **10**, 013 (2003), hep-th/0308055; M. Peloso, E. Poppitz, hep-ph/0307379; J. Gray, A. Lukas, hep-th/0309096.

48. N. Deruelle, T. Dolezel, Phys. Rev. D**62**, 103502 (2000), gr-qc/0004021; C. Germani, C.F. Sopuerta, Phys. Rev. Lett. **88**, 231101 (2002), hep-th/0202060; C. Charmousis, J-F. Dufaux, Class. Quantum Grav. **19**, 4671 (2002), hep-th/0202107; P. Binetruy, C. Charmousis, S.C. Davis, J.-F. Dufaux, Phys. Lett. **B544**, 183 (2002), hep-th/0206089; N.E. Mavromatos, J. Rizos, Int. J. Mod. Phys. **A18**, 57 (2003), hep-th/0205299; J.E. Lidsey, N.J. Nunes, Phys. Rev. D**67**, 103510 (2003), astro-ph/0303168; N. Deruelle, J. Madore, gr-qc/0305004; C. Barcelo, C. Germani, C.F. Sopuerta, Phys. Rev. D**68**, 104007 (2003), gr-qc/0306072; N. Deruelle, C. Germani, gr-qc/0306116; C. Charmousis, S.C. Davis, J.-F. Dufaux, hep-th/0309083; K. Maeda, T. Torii, hep-th/0309152.
49. G. Dvali, G. Gabadadze, M. Porrati, Phys. Lett. **B485**, 208 (2000), hep-th/0005016. See also, H. Collins, B. Holdom, Phys. Rev. D**62**, 105009 (2000), hep-ph/0003173; Yu.V. Shtanov, hep-th/0005193.
50. C. Deffayet, G. Dvali, G. Gabadadze, Phys. Rev. D**65**, 044023 (2002), astro-ph/0105068; G. Kofinas, JHEP **08**, 034 (2001), hep-th/0108013; G.A. Diamandis, B.C. Georgalas, N.E. Mavromatos, E. Papantonopoulos, I. Pappa, Int. J. Mod. Phys. **A17**, 2241 (2002), hep-th/0107124; E. Kiritsis, N. Tetradis, T.N. Tomaras, JHEP **03**, 019 (2002), hep-th/0202037; C. Deffayet, S.J. Landau, J. Raux, M. Zaldarriaga, P. Astier, Phys. Rev. D**66**, 024019 (2002), astro-ph/0201164; V. Sahni, Y. Shtanov, astro-ph/0202346; Y. Shtanov, V. Sahni, Class. Quantum Grav. **19**, L101 (2002), gr-qc/0204040; V. Sahni, Y. Shtanov, Int. J. Mod. Phys. **D11**, 1 (2002), gr-qc/0205111; P. Singh, R.G. Vishwakarma, N.K. Dadhich, hep-th/0206193; U. Alam, V. Sahni, astro-ph/0209443; K. Maeda, S. Mizuno, T. Torii, Phys. Rev. D**68**, 024033 (2003), gr-qc/0303039; A. Lue, G. Starkman, Phys. Rev. D**67**, 064002 (2003), astro-ph/0212083; T. Multamäki, E. Gaztañaga, M. Manera, Mon. Not. Roy. Astron. Soc. **344**, 761 (2003), astro-ph/0303526; B. Gumjudpai, gr-qc/0308046.
51. G. Kofinas, R. Maartens, E. Papantonopoulos, JHEP **10**, 066 (2003), hep-th/0307138.

8 Gravitational Wave Astronomy: The High Frequency Window

Nils Andersson[1] and Kostas D. Kokkotas[2,3]

[1] School of Mathematics, University of Southampton, Southampton SO17 1BJ, UK
[2] Department of Physics, Aristotle University of Thessaloniki, Thessaloniki 541 24 Greece
[3] Center for Gravitational Wave Physics, 104 Davey Laboratory, University Park, PA 16802, USA

Abstract. As several large scale interferometers are beginning to take data at sensitivities where astrophysical sources are predicted, the direct detection of gravitational waves may well be imminent. This would (finally) open the long anticipated gravitational-wave window to our Universe, and should lead to a much improved understanding of the most violent processes imaginable; the formation of black holes and neutron stars following core collapse supernovae and the merger of compact objects at the end of binary inspiral. Over the next decade we can hope to learn much about the extreme physics associated with, in particular, neutron stars.

This contribution is divided in two parts. The first part provides a text-book level introduction to gravitational radiation. The key concepts required for a discussion of gravitational-wave physics are introduced. In particular, the quadrupole formula is applied to the anticipated "bread-and-butter" source for detectors like LIGO, GEO600, EGO and TAMA300: inspiralling compact binaries. The second part provides a brief review of high frequency gravitational waves. In the frequency range above (say) 100 Hz, gravitational collapse, rotational instabilities and oscillations of the remnant compact objects are potentially important sources of gravitational waves. Significant and unique information concerning the various stages of collapse, the evolution of protoneutron stars and the details of the supranuclear equation of state of such objects can be drawn from careful study of the gravitational-wave signal. As the amount of exciting physics one may be able to study via the detections of gravitational waves from these sources is truly inspiring, there is strong motivation for the development of future generations of ground based detectors sensitive in the range from hundreds of Hz to several kHz.

8.1 Introduction

One of the central predictions of Einsteins' general theory of relativity is that gravitational waves will be generated as masses are accelerated. Despite decades of effort these ripples in spacetime have still not been observed directly. Yet we have strong indirect evidence for their existence from the excellent agreement between the observed inspiral rate of the binary pulsar PSR1913+16 and the theoretical prediction (better than 1% in the phase evolution). This provides confidence in the theory and suggests that

N. Andersson and K.D. Kokkotas, Gravitational Wave Astronomy: The High Frequency Window, Lect. Notes Phys. **653**, 255–276 (2005)
http://www.springerlink.com/

"gravitational-wave astronomy" should be viewed as a serious proposition. Provided that i) detectors with the required sensitivity can be constructed, and ii) the significant data analysis challenge can be dealt with, this new window to the Universe promises to bring unprecedented insights into the most violent events imaginable; supernova explosions, binary mergers and the big bang itself. A key reason for this expectation follows from a comparison between gravitational and electromagnetic waves:

- While electromagnetic waves are radiated by individual particles, gravitational waves are due to non-spherical bulk motion of matter. In essence, this means that the information carried by electromagnetic waves is stochastic in nature, while the gravitational waves provide insights into coherent mass currents.
- The electromagnetic waves will have been scattered many times. In contrast, gravitational waves interact weakly with matter and arrive at the Earth in pristine condition. This means that gravitational waves can be used to probe regions of space that are opaque to electromagnetic waves, It is, of course, a blessing in disguise since the weak interaction with matter also makes the gravitational waves fiendishly hard to detect.
- Standard astronomy is based on deep imaging of small fields of view, while gravitational-wave detectors cover virtually the entire sky.
- Electromagnetic radiation has a wavelength smaller than the size of the emitter, while the wavelength of a gravitational wave is usually larger than the size of the source. This means that we cannot use gravitational-wave data to create an image of the source. In fact, gravitational-wave observations are more like audio than visual.

Morale: Gravitational waves carry information which would be very difficult to glean by other means. By analysing gravitational-wave data we can expect to learn a lot about the extreme physics governing compact objects. This should lead to answers to many outstanding questions in astrophysics, for example,

- What is the black-hole population of the Universe? Observations of the nonlinear spacetime dynamics associated with binary merger, as well as the quasinormal mode (QNM) ringing which is likely to dominate the radiation at late times, should provide direct proof of the presence of a black hole, as well as a measure of it's mass and spin.
- Are astrophysical black holes, indeed, described by the Kerr metric? By studying the inspiral of a low-mass object into a supramassive black hole (as anticipated at the center of most galaxies) we can hope to construct a detailed map of the exterior black-hole spacetime.
- What is the supranuclear neutron star equation of state? This is a very difficult question, which potentially involves an understanding of the role of exotic phases of matter, like deconfined quarks, superfluidity/conductivity, extreme magnetic fields etcetera. At an immediate level,

gravitational-wave observations of the various oscillation modes of a compact star may be used to solve the "inverse problem" for parameters like the mass and the radius of the star. At a more subtle level, the spectrum of the star depends on the internal physics. For example, strong composition gradients which affect the gravity g-modes significantly, and superfluidity which leads to the existence of new classes of oscillations associated with the more or less decoupled additional degrees of freedom of such a system. Finally, the internal physics also plays a crucial role in determining the extent to which various rotational instabilities will be able to grow to an interesting amplitude. In the case of the Coriolis driven r-modes, their instability may be severely suppressed by the presence of hyperons in the stars core. A key role is also played by the internal magnetic fields.

At the time of writing, the new generation of interferometric gravitational-wave detectors (in particular LIGO and GEO600) is already collecting data at a sensitivity at least one order of magnitude better than that of the operating resonant detectors. In the first instance, the broadband detectors will be sensitive in a range of frequencies between 50 and a few hundred Hz. This frequency window is of great interest since an inspiraling compact binary will move through it during the last few minutes before merger. Such sources are the natural "bread and butter" source for the detectors. The next generation of interferometers will broaden the bandwidth somewhat but will still not be very sensitive to frequencies above 500-600 Hz, unless they are operated in a narrow-band configuration [1, 2]. There are also interesting suggestions for wide band resonant detectors in the kHz band [3]. The move towards higher frequencies is driven by the wealth of exciting sources that radiate in the range from a few hundred Hz up to several kHz.

Our contribution to these proceedings is divided in two parts. The first part describes gravitational-wave physics at an introductory level. The second part provides a brief review of the main sources that radiate in the frequency band above a few hundred Hz. We believe that these sources are the natural targets for a third generation of ground based detectors. As we will discuss, there are a variety of sources associated with very interesting physics in this high-frequency window. These sources clearly deserve special attention, and if either resonant or narrow-band interferometers can achieve the required sensitivity, a plethora of unique information can be gathered.

8.2 Einstein's Elusive Waves

The aim of the first part of our contribution is to provide a condensed textbook level introduction to gravitational waves. Although in no sense complete this description should prepare the reader for the discussion of high-frequency sources which follows.

8.2.1 The Nature of the Waves

The first aspect of gravitational waves that we need to appreciate is their *tidal* nature. This is important because it implies that they can only be measured through the relative motion of bodies. That this should be the case is easy to understand. In general relativity we can always construct a local inertial frame associated with a given observer. In this local frame, spacetime will by construction be flat which means that we cannot hope to observe the local deformations which would correspond to a gravitational wave. Consider two test particles, A and B, that are initially at rest, as shown in Fig. 8.1. Assume that they are separated by a purely spatial vector ξ^j (hereafter latin indices are spatial and run from 1–3, while greek indices are spacetime and run from 0–4), and use the local inertial frame in which particle A remains at the origin for the calculation. In this case the equation of geodesic deviation can be written

$$\frac{\partial^2 \xi^j}{\partial t^2} = -R^j{}_{0k0}\xi^k \ , \tag{8.1}$$

where $R_{\mu\nu\delta\gamma}$ is the Riemann tensor. Here it represents the curvature induced by the gravitational wave. Letting $\xi^j = x_0^j + \delta x^j$, with δx^j the small displacement away from the original position, we get

$$\frac{\partial^2 \delta x^j}{\partial t^2} \approx -R^j{}_{0k0}x_0^k = -R_{j0k0}x_0^k \ . \tag{8.2}$$

Now it is natural to define the *gravitational wave-field* h_{jk} through

$$R_{j0k0} \equiv -\frac{1}{2}\frac{\partial^2 h_{jk}}{\partial t^2} \longrightarrow \frac{\partial^2 \delta x_j}{\partial t^2} \approx \frac{1}{2}\frac{\partial^2 h_{jk}}{\partial t^2}x_0^k \tag{8.3}$$

which integrates to

$$\delta x_j = \frac{1}{2}h_{jk}x_0^k \quad \text{or} \quad h \approx \frac{\Delta L}{L} \ , \tag{8.4}$$

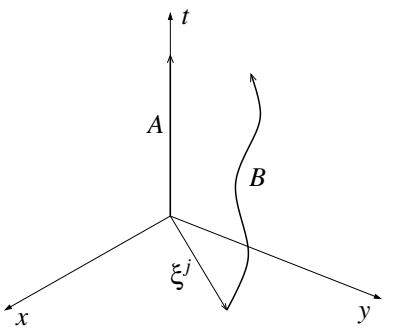

Fig. 8.1. An illustration of the two test particles discussed in the text. If a gravitational wave passes through, particle A will observe particle B oscillating back and forth.

where h is the dimensionless gravitational-wave strain. From this exercise we learn that, in order to detect gravitational waves we need to monitor (with extreme precision) the relative motion of test masses. Let us now assume that the waves propagate in the z-direction, i.e. that we have $h_{jk} = h_{jk}(t - z)$. Then one can show that we have only two independent components;

$$h_+ = h^{\rm TT}_{xx} = -h^{\rm TT}_{yy} \ , \qquad h_\times = h^{\rm TT}_{xy} = h^{\rm TT}_{yx} \ . \tag{8.5}$$

What effect does h_+ have on matter? Consider a particle initially located at (x_0, y_0) and let $h_\times = 0$ to find that

$$\delta x = \frac{1}{2} h_+ x_0 \ , \tag{8.6}$$

$$\delta y = -\frac{1}{2} h_+ y_0 \ . \tag{8.7}$$

That is, if h_+ is oscillatory (a wave!) then an object will first experience a stretch in the x-direction accompanied by a squeeze in the y-direction. One half-cycle later, the squeeze is in the x-direction and the stretch in the y-direction. It is straightforward to show that the effect of $h_\times$ is the same, but rotated by 45 degrees. This is illustrated in Fig. 8.2. A general wave will be a linear combination of the two polarisations.

But wait a second! We are discussing the effect of gravitational *waves* without actually having proven that Einstein's theory predicts their existence. To remedy this, consider small perturbations away from a flat spacetime. That is, use $g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta}$ to get the (linearised) Ricci tensor:

$$R_{\mu\nu} = \frac{1}{2}(h_{\alpha\nu,\mu}{}^{\alpha} + h_{\mu\alpha}{}^{,\alpha}{}_{\nu} - h^{\alpha}{}_{\alpha,\mu\nu} - h_{\mu\nu,\alpha}{}^{\alpha}) \ . \tag{8.8}$$

In vacuum, Einstein's equations are equivalent to requiring that

$$R_{\mu\nu} = 0 \ . \tag{8.9}$$

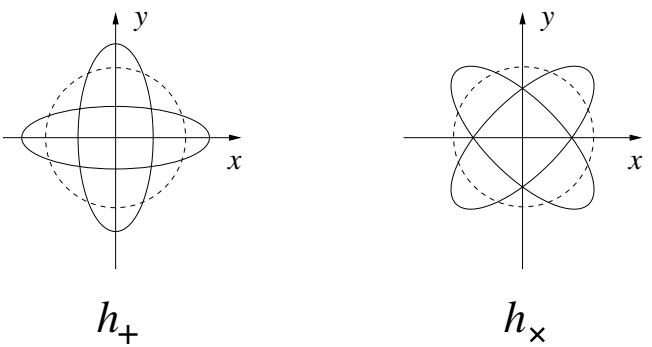

Fig. 8.2. The two gravitational-wave polarisations h_+ and $h_\times$.

Now consider what happens if we make a coordinate transformation. For

$$x^\alpha \to x^\alpha + \xi^\alpha \tag{8.10}$$

we get

$$h_{\mu\nu} \to h_{\mu\nu} - \xi_{\mu,\nu} - \xi_{\nu,\mu} \ . \tag{8.11}$$

Use this freedom to impose the harmonic gauge condition ${h_{\mu\alpha}}^{,\alpha} = 0$. Since the gauge remains unchanged for any transformation such that $\Box\xi_\mu = 0$, we can impose further conditions. We take one of these to be ${h_\alpha}^\alpha = 0$, to get

$${h_{\mu\nu,\alpha}}^\alpha = \Box h_{\mu\nu} = 0 \ . \tag{8.12}$$

That is, the metric variations are governed by a standard wave equation. Finally, use $h_{00} = h_{j0} = h_{0j} = 0$ to get $h_{jk} = h^{\rm TT}_{jk}$ as before. The set of coordinates we have introduced is known as TT-gauge.

Before we move on to discuss the modelling of various gravitational-wave sources, it is worth elucidating an issue that caused serious debate until the late 1960s. We need to demonstrate that gravitational waves carry energy. This is a tricky problem because, as we have already pointed out, one can only deduce the presence of a wave from the relative effect on two (or more) test particles. This means that one cannot localize the wave to individual points in space, and hence cannot directly "measure" its energy. In order to construct a meaningful energy expression we need to average over one (or more) wavelengths. Defining perturbations with respect to an averaged spacetime metric, i.e. using

$$g^{\rm (B)}_{\mu\nu} = \langle g_{\mu\nu} \rangle \longrightarrow g_{\mu\nu} = g^{\rm (B)}_{\mu\nu} + h_{\mu\nu} \ . \tag{8.13}$$

and expanding in powers of h (which is presumed small), we have the (schematic) Einstein equations,

$$\langle G_{\mu\nu} \rangle = G^{\rm B}_{\mu\nu} + \underbrace{\left\langle G^{(1)}_{\mu\nu} \right\rangle}_{O(h_{\mu\nu})} + \underbrace{\left\langle G^{(2)}_{\mu\nu} \right\rangle}_{O(h_{..}h_{..})} = 0 \ . \tag{8.14}$$

Here, the term that is linear in h will vanish if we average over a wavelength. This means that we can deduce an expression for the stress energy-tensor for gravitational waves:

$$G^{\rm B}_{\mu\nu} = -8\pi \left\langle \frac{G^{(2)}_{\mu\nu}}{8\pi} \right\rangle \equiv 8\pi T^{\rm GW}_{\mu\nu} \ . \tag{8.15}$$

Working out the algebra one can show that, in TT-gauge we will have

$$T^{\rm GW}_{\mu\nu} = \frac{1}{32\pi} \left\langle h^{\rm TT}_{ij,\mu} {h^{{\rm TT}ij}}_{,\nu} \right\rangle \ . \tag{8.16}$$

In particular, the energy propagating in the z-direction then follows from

$$T_{0z}^{\rm GW} = -T_{00}^{\rm GW} = -\frac{1}{16\pi}\left\langle \dot{h}_+^2 + \dot{h}_\times^2 \right\rangle \longrightarrow \dot{E} = -T_{0z}^{\rm GW} = \frac{\omega^2}{16\pi}\left\langle h_+^2 + h_\times^2 \right\rangle \ . \tag{8.17}$$

where the frequency of the wave is ω. Finally assuming that $h_+ \sim h_\times \sim h\sin\omega(t-z)$, and integrating over a sphere with radius r, we get

$$\dot{E} = \frac{\omega^2 r^2}{4} h^2 \rightarrow |\dot{h}|^2 = \frac{4G}{c^3 r^2}\dot{E} \ . \tag{8.18}$$

As we will now demonstrate, this is a very useful relation.

8.2.2 Estimating the Gravitational-Wave Amplitude

We can use the expression (8.18) to infer the gravitational-wave strain associated with typical gravitational-wave sources. Let us characterise a given event by a timescale τ and assume that the signal is monochromatic (with frequency f). Then we can use $dE/dt \approx E/\tau$ and $\dot{h} \approx 2\pi f h$ to deduce that

$$h \approx 5\times 10^{-22}\left(\frac{E}{10^{-3}M_\odot c^2}\right)^{1/2}\left(\frac{\tau}{1\ {\rm ms}}\right)^{-1/2}\left(\frac{f}{1\ {\rm kHz}}\right)^{-1}\left(\frac{r}{15\ {\rm Mpc}}\right)^{-1} . \tag{8.19}$$

If the signal analysis is based on matched filtering, the *effective amplitude* improves roughly as the square root of the number of observed cycles n. Using $n \approx f\tau$ we get

$$h_{\rm c} \approx 5\times 10^{-22}\left(\frac{E}{10^{-3}M_\odot c^2}\right)^{1/2}\left(\frac{f}{1\ {\rm kHz}}\right)^{-1/2}\left(\frac{r}{15\ {\rm Mpc}}\right)^{-1} . \tag{8.20}$$

This is a crucial expression. We see that the "detector sensitivity" essentially depends only on the radiated energy, the characteristic frequency and the distance to the source. That is, in order to obtain a rough estimate of the relevance of a given gravitational-wave source at a given distance we only need to estimate the frequency and the radiated energy. Alternatively, if we know the energy released can work out the distance at which these sources can be detected.

It is quite easy to obtain a rough idea of the frequencies involved. The dynamical frequency of any self-bound system with mass M and radius R is

$$f \approx \frac{1}{2\pi}\sqrt{\frac{GM}{R^3}} \ . \tag{8.21}$$

Thus, the natural frequency of a (non-rotating) black hole should be

$$f_{\rm BH} \approx 10^4\left(\frac{M_\odot}{M}\right)\ {\rm Hz} \ . \tag{8.22}$$

Medium sized black holes, with masses in the range $10 - 100M_\odot$, will be prime sources for ground-based interferometers, while black holes with masses

$\sim 10^6 M_\odot$ should radiate in the LISA bandwidth. Meanwhile, neutron stars, with a mass of $1.4 M_\odot$ compressed inside a radius of 10 km, will radiate at

$$f_{\rm NS} \approx 2 \text{ kHz} \ . \tag{8.23}$$

This means that one would expect neutron physics to be in the range for ground based detectors. In fact, given the likely need to detect signals with frequencies above 1 kHz, neutron star signals provide a strong motivation for the development of (third generation) high-frequency detectors.

Given that the weak signals are going to be buried in detector noise, we need to obtain as accurate theoretical models as possible. The rough order of magnitude estimates we just derived will certainly not be sufficient, even though they provide an indication as to whether it is worth spending the time and effort required to build a detailed model. Such source models are typically obtained using either

- approximate perturbation techniques, eg. expansions in small perturbations away from a known solution to the Einstein equations, the archetypal case being black-hole and neutron star oscillations.
- post-Newtonian approximations, essentially an expansion in the ratio between a characteristic velocity of the system and the speed of light, most often used to model the inspiral phase of a compact binary system.
- numerical relativity, where the Einstein equations are formulated as an initial-value problem and solved on the computer. This is the only way to make progress in situations where the full nonlinearities of the theory must be included, eg. in the merger of black holes and neutron stars or a supernova core collapse.

Here we will only describe the first step beyond a Newtonian description, where the gravitational radiation is described by the so-called *quadrupole formula*. For a source with weak internal gravity we have (in TT-gauge)

$$\Box h_{\mu\nu} = -16\pi T_{\mu\nu} \ . \tag{8.24}$$

We can solve this equation using the standard retarded Green's function to get

$$h_{\mu\nu}(t, \boldsymbol{x}) = 4 \int \frac{T_{\mu\nu}(\boldsymbol{x}', t' = t - |\boldsymbol{x} - \boldsymbol{x}'|)}{|\boldsymbol{x} - \boldsymbol{x}'|} d^3x \ . \tag{8.25}$$

Matching of the near-zone solution to an outgoing wave solution far away leads to the expression

$$h_{jk}^{\rm TT} = \frac{2G}{rc^4} \ddot{\mathcal{I}}_{jk}^{\rm TT}(t - r) \ , \tag{8.26}$$

where

$$\mathcal{I}_{jk} \equiv \int \rho \left(x_j x_k - \frac{1}{3} r^2 \delta_{jk} \right) d^3x \ . \tag{8.27}$$

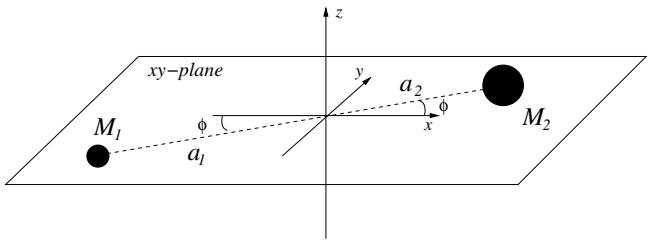

Fig. 8.3. A schematic illustration of a compact binary system.

is the reduced quadrupole moment of the source. Consider a system of mass M with typical internal velocity v. Then we see that

$$h \approx \left(\frac{2GM}{c^2}\right)\left(\frac{v}{c}\right)^2 \frac{1}{r} \tag{8.28}$$

which shows (no surprise!) that in order to generate strong gravitational waves we need large masses moving at high speeds.

From the formulas we derived earlier, we find that energy is radiated at a rate

$$\frac{dE}{dt} = -\frac{G}{5c^5}\left\langle \dddot{\mathcal{I}}_{jk}\dddot{\mathcal{I}}^{jk}\right\rangle \ . \tag{8.29}$$

The radiated angular momentum follows from the (usually) weaker current multipole radiation, which is governed by a similar expression. Let us now apply the above results to the potentially most important gravitational-wave source, a compact binary system. Gravitational waves are emitted as the stars (or black holes) orbit each other and as a result the binary separation decreases. Consider a binary system, as illustrated in Fig. 8.3, with individual masses M_1 and M_2 and separation $2R$. Introduce the total and reduced masses

$$M = M_1 + M_2 \qquad \text{and} \qquad \mu = \frac{M_1 M_2}{M} \longrightarrow M_1 a_1 = M_2 a_2 = \mu R \tag{8.30}$$

and work in the coordinate system illustrated in Fig. 8.3. Working out the required (time-varying) components of the quadrupole moment, we have

$$\mathcal{I}_{xx} = -\mathcal{I}_{yy} = \frac{\mu R^2}{2}\cos 2\phi \quad \text{and} \quad \mathcal{I}_{xy} = \mathcal{I}_{yx} = \frac{\mu R^2}{2}\sin 2\phi \tag{8.31}$$

and we find that

$$\frac{dE}{dt} = \frac{G}{5c^5}\left\langle \dddot{\mathcal{I}}_{jk}\dddot{\mathcal{I}}^{jk}\right\rangle = \frac{32}{5}\frac{G}{c^5}\mu^2 R^4 \Omega^6 \ . \tag{8.32}$$

Next, determining the orbital rotation frequency from Kepler's law $\Omega^2 = M/R^3$, and introducing the so-called "chirp mass" $\mathcal{M} = \mu^{3/5} M^{2/5}$ we have the final result

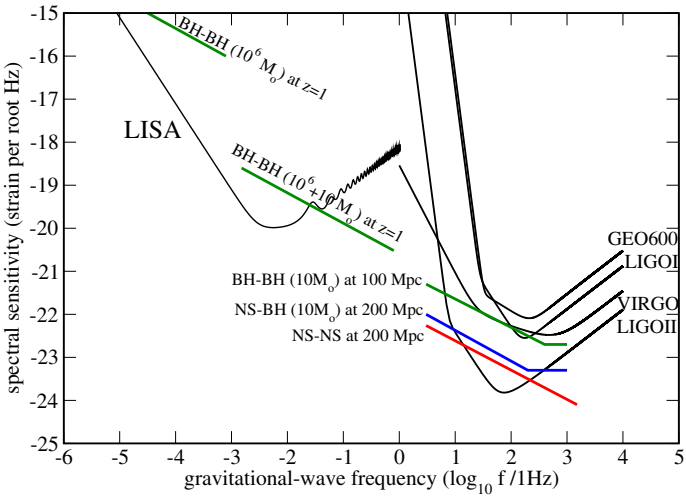

Fig. 8.4. Estimated signal strengths for various inspiralling binaries relevant for ground- and space-based detectors.

$$\frac{dE}{dt} = \frac{32}{5}\frac{G^4}{c^5}(\mathcal{M}\Omega)^{10/3} \ . \tag{8.33}$$

Moreover, from

$$h_{xx}^{\rm TT} \approx -\frac{2\mathcal{M}^{5/3}\Omega^{2/3}}{r}\cos 2\Omega(t-r) \ , \tag{8.34}$$

we can estimate the effective amplitude of the binary signal:

$$h_c \approx \sqrt{ft}\, h \sim \frac{M}{r}\left(\frac{R}{M}\right)^{1/4} \ . \tag{8.35}$$

This shows that, even though the actual signal gets stronger, its detectability decreases as the orbit shrinks. Figure 8.4 compares our estimated gravitational-wave strain to the predicted noise-curves for various gravitational-wave detectors. Essentially, one would expect that

- Advanced LIGO may observe several binary systems per year.
- The space-based LISA detector may suffer an "embarrassment of riches", with a large number of known galactic binaries leading to detectable signals (most likely generating to a "binary noise" which will be difficult to filter out).

To predict the rate at which the binary orbit shrinks as a result of gravitational-wave emission we need to estimate to total energy of the system:

$$E = \frac{M_1 v_1^2}{2} + \frac{M_2 v_2^2}{2} - \frac{GM_1M_2}{R} = -\frac{G\mu M}{2R} = -G\frac{\mathcal{M}^{5/3}\Omega^{2/3}}{2} \ . \tag{8.36}$$

From this we find that the period P of the system changes as

$$\frac{\dot{P}}{P} = -\frac{96}{5}\frac{G^3}{c^5}\mathcal{M}^{5/3}\Omega^{8/3} \ . \tag{8.37}$$

For the *binary pulsar* 1913+16, the predicted change in orbital period agrees with the theoretical prediction to within 1%. This indirect proof that gravitational waves exist led to Hulse and Taylor being awarded the 1993 Nobel prize in physics.

Finally, note that that the chirp-mass $\mathcal{M}$ plays a key role in our various expressions. From

$$\frac{\dot{P}}{P} \sim \mathcal{M}^{5/3}\Omega^{8/3} \quad \text{and} \quad h \sim \frac{\mathcal{M}^{5/3}\Omega^{2/3}}{r} \ , \tag{8.38}$$

we see that this is the only combination of the two masses that can be infered from the signal (at this level of approximation: higher order post-Newtonian corrections depend on the individual masses, the spins etcetera). Suppose that we detect both the change in the amplitude h and the shift in the frequency. Then one can infer both the chirp mass and the distance to the source, and in effect coalescing binaries are *standard candles* which can be used to constrain cosmological parameters.

One often classifies gravitational-wave sources by the nature of the waves. This is convenient because the different classes require different approaches to the data-analysis problem;

- *Chirps*. As a binary system radiates gravitational waves and loses energy the two constituents spiral closer together. As the separation decreases the gravitational-wave amplitude increases, leading to a characteristic "chirp" signal.
- *Bursts*. Many scenarios lead to burst-like gravitational waves. A typical example would be black-hole oscillations excited during binary merger.
- *Periodic*. Systems where the gravitational-wave backreaction leads to a slow evolution (compared to the observation time) may radiate persistent waves with a virtually constant frequency. This would be the gravitational-wave analogue of the radio pulsars.
- *Stochastic*. A stochastic (non-thermal) background of gravitational waves is expected to have been generated following the Big Bang. One may also have to deal with stochastic gravitational-wave signals when the sources are too abundant for us to distinguish them as individuals.

8.3 High-Frequency Gravitational Wave Sources

Having introduced the key concepts required for a discussion of gravitational-wave physics we will now focus our attention on sources that radiate above a

hundred Hz or so, i.e. which may at least in principle be detectable from the ground. As we will see there are strong motivations for constructing detectors which are sensitive up to (ideally) several kHz.

8.3.1 Radiation from Binary Systems

It is easy to understand why binary systems are considered the "best" sources of gravitational waves. They emit copious amounts or gravitational radiation, and for a given system we know quite accurately the amplitude and frequency of the gravitational waves in terms of the masses of the two bodies and their separation (see Sect. 8.2.2).

The gravitational-wave signal from inspiraling binaries is approximately sinusoidal, see equation (8.34), with a frequency which is twice the orbital frequency of the binary. As the binary system evolves the orbit shrinks and the frequency increases in the characteristic chirp. Eventually, depending on the masses of the binaries, the frequency of the emitted gravitational waves will enter the bandwidth of the detector at the low-frequency end and will evolve quite fast towards higher frequencies. A system consisting of two neutron stars will be detectable by LIGO when the frequency of the gravitational waves is $\sim$10Hz until the final coalescence around 1 kHz. This process will last for about 15 min and the total number of observed cycles will be of the order of 10^4, which leads to an enhancement of the detectability by a factor 100 (remember $h_c \sim \sqrt{n}h$). Binary neutron star systems and binary black hole systems with masses of the order of $50M_\odot$ are the primary sources for LIGO. Given the anticipated sensitivity of LIGO, binary black hole systems are the most promising sources and could be detected as far as 200 Mpc away. For the present estimated sensitivity of LIGO the event rate is probably a few per year, but future improvements of detector sensitivity (the LIGO II phase) could lead to the detection of at least one event per month. Supermassive black hole systems of a few million solar masses are the primary source for LISA. These binary systems are rare, but due to the huge amount of energy released, they should be detectable from as far away as the boundaries of the observable universe. Finally, the recent discovery of the highly relativistic binary pulsar J0737-3039 [4] enchanced considerably the expected coalescence event rate of NS-NS binaries [5]. The event rate for initial LIGO is in the best case 0.2 per year while advanced LIGO might be able to detect 20-1000 events per year.

8.3.2 Gravitational Collapse

One of the most spectacular events in the Universe is the supernova (SN) collapse to create a neutron star (NS) or a black hole (BH). Core collapse is a very complicated event and a proper study demands a deep understanding of neutrino emission, amplification of the magnetic fields, angular momentum distribution, pulsar kicks, etc. There are many viable models for each of

the above issues but it is still not possible to combine all of them together into a consistent explanation. Gravitational waves emanating from the very first moments of the core collapse might shed light on all the above problems and help us understand the details of this dramatic event. Gravitational collapse compresses matter to nuclear densities, and is responsible for the core bounce and the shock generation. The event proceeds extremely fast, lasting less than a second, and the dense fluid undergoes motions with relativistic speeds ($v/c \sim 0.2 - 0.4$). Even small deviations from spherical symmetry during this phase can generate significant amounts of gravitational waves. However, the size of these asymmetries is not known. From observations in the electromagnetic spectrum we know that stars more massive than $\sim 8M_\odot$ end their evolution in core collapse and that $\sim 90\%$ of them are stars with masses $\sim 8 - 20M_\odot$. During the collapse most of the material is ejected and if the progenitor star has a mass $M \lesssim 20M_\odot$ it leaves behind a neutron star. If $M \gtrsim 20M_\odot$ more than 10% falls back and pushes the proto-neutron-star (PNS) above the maximum NS mass leading to the formation of a black hole (*type II collapsars*). Finally, if the progenitor star has a mass $M \gtrsim 40M_\odot$ no supernova takes place. Instead, the star collapses directly to a BH (*type I collapsars*).

A significant amount of the ejected material can fall back, subsequently spinning up and reheating the nascent NS. Instabilities can be excited during such a process. If a BH was formed, its quasi-normal modes (QNM) can be excited for as long as the process lasts. "Collapsars" accrete material during the very first few seconds, at rates $\sim 1 - 2M_\odot$/sec. Later the accretion rate is reduced by an order of magnitude but still material is accreted for a few tenths of seconds. Typical frequencies of the emitted gravitational waves are in the range 1-3kHz for $\sim 3 - 10M_\odot$ BHs. If the disk around the central object has a mass $\sim 1M_\odot$ self-gravity becomes important and gravitational instabilities (spiral arms, bars) might develop and radiate gravitational waves. Toroidal configurations can be also formed around the collapsed object. Their instabilities and oscillations might be an interesting source of gravitational waves[7]. There is also the possibility that the collapsed material might fragment into clumps, which orbit for some cycles like a binary system (*fragmentation instability* [8]).

The supernova event rate is 1-2 per century per galaxy [6] and about 5-40% of them produce BHs through the fallback material [9]. Conservation of angular momentum suggests that the final objects should rotate close to the mass shedding limit, but this is still an open question, since there is limited knowledge of the initial rotation rate of the final compact object. Pulsar statistics suggest that the initial periods are probably considerably shorter than 20 ms. This strong increase of rotation during the collapse has been observed in many numerical simulations (see e.g. [10, 11]).

Core collapse as a potential source of gravitational waves has been studied for more than three decades (some of the most recent calculations can be found in [12, 13, 14, 15, 11, 16, 17, 18, 19]). All these numerical calculations

show that signals from Galactic supernova ($d \sim 10$kpc) are detectable even with the initial LIGO/EGO sensitivity at frequencies $\lesssim$1kHz. Advanced interferometers can detect signals from distances of 1 Mpc but it will be difficult with the designed broadband sensitivity to resolve signals from the Virgo cluster ($\sim$15Mpc). The typical gravitational wave amplitude from the 2D numerical simulations [11, 16] for an observer located on the equatorial plane of the source is

$$h \approx 9 \times 10^{-21} \varepsilon \left(\frac{10\text{kpc}}{d} \right) , \tag{8.39}$$

where $\varepsilon \sim 1$ is the normalized gravitational wave amplitude. The total energy radiated in gravitational waves during the collapse is $\lesssim 10^{-6} - 10^{-8} M_{\odot} c^2$. However, these numerical estimates are not yet conclusive, as important aspects such as 3D hydrodynamics combined with proper spacetime evolution have been neglected. The influence of the magnetic fields have been ignored in most calculations. The proper treatment of these issues might not change the above estimates by orders of magnitude but it will provide a conclusive answer. There are also issues that need to be understood such as the pulsar kicks (velocities even higher than 1000 km/s) which suggest that in a fraction of newly-born NSs (and BHs) the process may be strongly asymmetric [20, 21, 22, 23, 24]. Also, the polarization of the light spectra in supernovae indicates significant asymmetries [25]. Better treatment of the microphysics and construction of accurate progenitor models for the angular momentum distributions are needed. All these issues are under investigation by many groups.

Accretion Incuded Collapse (AIC) is also a possible source of high frequency gravitational waves. AIC takes place when a white dwarf (WD) exceeds the Chandrasekhar limit due to accretion of material and begins to collapse. The cooling via neutrino emission does not reduce the heating significantly and the collapsing WDs reach appropriate temperatures for ignition of nuclear burning (Type Ia supernova). Estimates suggest that about $0.1 M_{\odot}$ material is ejected. Since the WD is pushed over the Chandrasekhar limit due to accretion, it will rotate fast enough to allow various types of instabilities [26]. The galactic rate of accretion induced collapse is about 10^{-5}/yr which means that AIC are about 1000 times rarer than core collapse SN.

8.3.3 Rotational Instabilities

Newly born neutron stars are expected to rotate rapidly enough to be subject to rotation induced instabilities. These instabilities arise from non-axisymmetric perturbations having angular dependence $e^{im\varphi}$. Early Newtonian estimates have shown that a *dynamical bar-mode ($m = 2$) instability* is excited if the ratio $\beta = T/W$ of the rotational kinetic energy T to the gravitational binding energy W is larger than $\beta_{\rm dyn} = 0.27$. The instability develops on a dynamical time scale (the time that a sound wave needs to travel across the star) which is about one rotation period, and may last from

1 to 100 rotations depending on the degree of differential rotation in the PNS. Another class of instabilities are those driven by dissipative effects such as fluid viscosity or gravitational radiation. Their growth time is much longer (many rotational periods) but they can be excited for significantly lower rotational rates, $\beta \gtrsim 0.14$ in the case of the fundamental modes of oscillation of the star.

8.3.4 Bar-Mode Instability

The dynamical bar-mode instability can be excited in a hot PNS, a few milliseconds after the core-bounce, given a sufficiently large β. It might also be excited a few tenths of seconds later, when the NS cools enough due to neutrino emission and contracts still further ($\beta \sim 1/R$). The amplitude of the emitted gravitational waves can be estimated as $h \sim MR^2\Omega^2/d$, where M is the mass of the body, R its size, Ω the rotational rate and d the distance from Earth. This leads to an estimate of the gravitational wave amplitude

$$h \approx 9 \times 10^{-23} \left(\frac{\epsilon}{0.2}\right) \left(\frac{f}{3\text{kHz}}\right)^2 \left(\frac{15\text{Mpc}}{d}\right) M_{1.4} R_{10}^2 \ . \tag{8.40}$$

where ϵ measures the ellipticity of the bar. Note that the gravitational wave frequency f is twice the rotational frequency Ω. Such a signal is detectable only from sources in our galaxy or the nearby ones (our Local Group). If the sensitivity of the detectors is improved in the kHz region, then signals from the Virgo cluster may be detectable. If the bar persists for many ($\sim$ 10-100) rotation periods, then even signals from distances considerably larger than the Virgo cluster could be detectable, cf. Fig. 8.5. The event rate is of the same order as the SN rate (a few events per century per galaxy): this means that given the appropriate sensitivity at frequencies between 1-3kHz we might be able to observe a few events per year. The bar-mode instability may also be excited during the merger of NS-NS, BH-NS, BH-WD and even in type II collapsars (see discussion in [27]).

In general, the above estimates rely on Newtonian hydrodynamics calculations; GR enhances the onset of the instability slightly, $\beta_{\text{dyn}} \sim 0.24$ [28] and β_{dyn} may be even lower for large values of the compactness (larger M/R). The bar-mode instability may be excited for significantly smaller β if centrifugal forces produce a peak in the density off the sources rotational center[29]. Rotating stars with a high degree of differential rotation are also dynamically unstable for significantly lower $\beta_{\text{dyn}} \gtrsim 0.01$ [30, 31]. In this scenario the unstable neutron star settles down to a non-axisymmetric quasistationary state which is a strong emitter of quasi-periodic gravitational waves

$$h_{\text{eff}} \approx 3 \times 10^{-22} \left(\frac{R_{\text{eq}}}{30\text{km}}\right) \left(\frac{f}{800\text{Hz}}\right)^{1/2} \left(\frac{100\text{Mpc}}{d}\right) M_{1.4}^{1/2} \ . \tag{8.41}$$

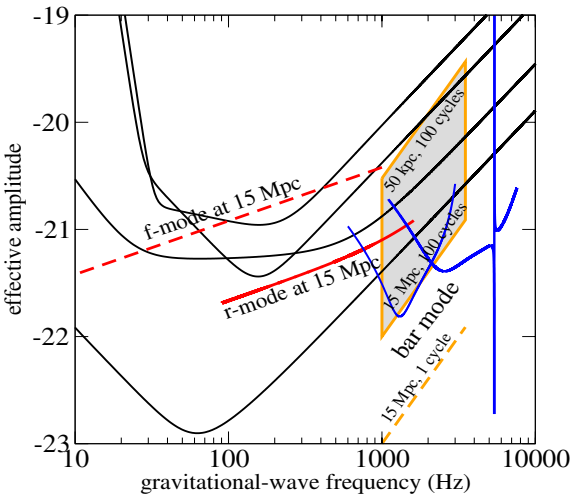

Fig. 8.5. The estimated strength of gravitational waves from the dynamical bar-mode instability and the CFS instability of the f- and r-modes. The estimates are compared to the predicted noise of the various interferometers and also the possible noise curve for a dual cylinder detector.

The bar-mode instability of differentially rotating neutron stars could be an excellent source of gravitational waves provided that the dissipation of non-axisymmetric perturbations by viscosity and magnetic fields is negligible. That this is the case is far from clear. Magnetic fields might actually enforce the uniform rotation of the star on a dynamical timescale and the persistent non-axisymmetric structure might not have time to develop at all.

Numerical simulations have shown that the $m = 1$ one-armed spiral mode might become dynamically unstable for considerably lower rotational rates [29, 32]. This $m = 1$ instability depends critically on the softness of the equation of state (EoS) and the degree of differential rotation.

8.3.5 CFS Instability, f- and r-Modes

After the initial bounce, neutron stars may maintain a considerable amount of deformation. They settle down to an axisymmetic configuration mainly due to emission of gravitational waves, viscosity and magnetic fields. During this phase QNMs are excited. Technically speaking, an oscillating non-rotating star has equal values $\pm|\sigma|$ (the frequency of a mode) for the forward and backward propagating modes (corresponding to $m = \pm|m|$). Rotation changes the mode frequency by an amount $\delta\sigma \sim m\Omega$ and both the prograde and retrograde modes will be dragged forward by the stellar rotation. If the star spins sufficiently fast, the originally retrograde mode will appear to be moving forwards in the inertial frame (according to an observer at infinity),

but still backwards in the rotating frame (for an observer rotating with the star). Thus, an inertial observer sees gravitational waves with positive angular momentum emitted by the retrograde mode, but since the perturbed fluid rotates slower than it would in absence of the perturbation, the angular momentum of the mode itself is negative. The emission of gravitational waves consequently makes the angular momentum of the mode increasingly negative leading to an instability. From the above, one can easily conclude that a mode will be unstable if it is retrograde in the rotating frame and prograde for a distant observer measuring a mode frequency $\sigma - m\Omega$ i.e. the criterion will be $\sigma(\sigma - m\Omega) < 0$.

This class of *frame-dragging instabilities* is usually referred to as Chandrasekhar–Friedman–Schutz (CFS) [34, 35] instabilities. For the high frequency (f and p) modes the instability is possible only for large values of Ω or for quite large m. In general, for every mode there will always be a specific value of Ω for which the mode will become unstable, although only modes with $|m| < 5$ have an astrophysically significant growth time. The CFS mechanism is not only active for fluid modes but also for the *spacetime* or the so-called w-modes[33]. It is easy to see that the CFS mechanism is not unique to gravitational radiation: any radiative mechanism will have the same effect.

In GR, the *f-mode* ($l = m = 2$) becomes unstable for $\beta \approx 0.06 - 0.08$ [36]. If the star has significant differential rotation the instability is excited for somewhat higher values of β (see e.g. [37, 38]). The f-mode instability is an excellent source of gravitational waves. After the brief dynamical phase, the PNS becomes unstable and the instability deforms the star into a non-axisymmetric configuration via the $l = 2$ bar mode. Since the star loses angular momentum it spins down, and the gravitational wave frequency sweeps from 1 kHz down to about 100 Hz [39]. If properly modelled such a signal can be detected from a distance of 100 Mpc (if the mode grows to a large nonlinear amplitude).

Rotation not only shifts the frequencies of the various modes; it also gives rise to the Coriolis force, and an associated new family of *rotational* or *inertial* modes. Inertial modes are primarily velocity perturbations. Of special interest is the quadrupole inertial mode (r-mode) with $l = m = 2$. The frequency of the r-mode in the rotating frame of reference is $\sigma = 2\Omega/3$. Using the CFS criterion for stability we can easily show that the r-mode is unstable for any rotation rate of the star. For temperatures between $10^7 - 10^9$K and rotation rates larger than 5-10% of the Kepler limit, the growth time of the unstable mode is much shorter than the damping times due to bulk and shear viscosity. The mode grows until it saturates due to non-linear effects [40, 41, 42]. The strength of the emitted gravitational waves depends on the saturation amplitude α. Mode coupling might not allow the growth of the instability to amplitudes larger than $\alpha \approx 10^{-2} - 10^{-3}$ [45]. The existence of a crust [43, 44] or of hyperons in the core [47] and strong magnetic fields [46], affect the efficiency of the instability (for extended reviews see [48, 49]). For newly-born neutron stars the amplitude of gravitational waves might not be

such that the signals will be detectable only from the local group of galaxies ($d < 1$Mpc)

$$h(t) \approx 10^{-21}\alpha \left(\frac{\Omega}{1\text{kHz}}\right)\left(\frac{100\text{kpc}}{d}\right) \ , \tag{8.42}$$

(see Fig. 8.5).

If the compact object is a strange star, then the r-mode instability will not reach high amplitudes ($\alpha \sim 10^{-3} - 10^{-4}$) but it will persist for a few hundred years and in this case there might be up to ten unstable stars per galaxy radiating gravitational waves at any time [54]. Integrating data for a few weeks can then lead to an effective amplitude $h_{\rm eff} \sim 10^{-21}$ for galactic signals at frequencies $\sim 700 - 1000$Hz. The frequency of the signal changes only slightly on a timescale of a few months, so the radiation is practically monochromatic.

Old accreting neutron stars, radiating gravitational waves due to the r-mode instability, at frequencies 400-700Hz, are probably a better source [50, 51, 52, 53]. Still, the efficiency and the actual duration of the process depends on the saturation amplitude α. If the accreting compact object is a strange star or has a hyperon core then it might be a persistent source which radiates gravitational waves for as long as accretion lasts [54, 55].

8.3.6 Oscillations of Black Holes and Neutron Stars

Black-hole ringing. The merger of two neutron stars or black holes or the collapse of a supermassive star (collapsar of type I or II) will produces a black hole. The newly formed black hole will ring, emitting a characteristic signal until it settles down to the stationary Kerr state. This characteristic signal, the so-called quasi-normal mode oscillation, will be a unique probe of the black hole's existence. Although the ringing phase does not last very long (a few tenths of a ms), the ringing due to the excitation by the fallback material might last for secs[56, 57]. The frequency and damping time of the black-hole ringing for the $l = m = 2$ oscillation mode can be estimated via the relations [58]

$$\sigma \approx 3.2\text{kHz}\ M_{10}^{-1}\left[1 - 0.63(1 - a/M)^{3/10}\right] \ , \tag{8.43}$$

$$Q = \pi\sigma\tau \approx 2\,(1-a)^{-9/20} \ . \tag{8.44}$$

These relations together with similar ones either for the 2nd QNM or the $l = 2, m = 0, \pm 1$ can uniquely determine the mass M and angular momentum a of the BH if the frequency and the damping time of the signal have been accurately extracted [59, 60]. The amplitude of the ring-down waves depends on the BH's initial distortion. If the excitation of the BH is due to infalling material then the energy is roughly $\Delta E \gtrsim \epsilon\mu c^2(\mu/M)$ where $\epsilon \gtrsim 0.01$ [61]. This leads to an effective gravitational wave amplitude

$$h_{\rm eff} \approx 2 \times 10^{-21} \left(\frac{\epsilon}{0.01}\right) \left(\frac{10{\rm Mpc}}{d}\right) \left(\frac{\mu}{M_\odot}\right) \ . \tag{8.45}$$

This approximate result has been verified by more detailed full non-linear simulations [62]

Neutron star ringing. If the collapse leaves behind a compact star, various types of oscillation modes might be excited which can help us estimate parameters of the star such as radius, mass, rotation rate and EoS[63, 64, 65]. This *gravitational wave asteroseismology* is a unique way to find the radius and the EoS of compact stars. One can derive approximate formulas in order to connect the observable frequencies and damping times of the various stellar modes to the stellar parameters. For example, for the fundamental oscillation ($l = 2$) mode (f-mode) of non-rotating stars we get [63]

$$\sigma({\rm kHz}) \approx 0.8 + 1.6 M_{1.4}^{1/2} R_{10}^{-3/2} + \delta_1 m \bar{\Omega} \ , \tag{8.46}$$

$$\tau^{-1}({\rm secs}^{-1}) \approx M_{1.4}^3 R_{10}^{-4} \left(22.9 - 14.7 M_{1.4} R_{10}^{-1}\right) + \delta_2 m \bar{\Omega} \ , \tag{8.47}$$

where $\bar{\Omega}$ is the normalized rotation frequency of the star, and δ_1 and δ_2 are constants estimated by sampling data from various EoS. The typical frequencies of the NS modes are higher than 1 kHz. On the other hand, 2D simulations of rotating core-collapse have shown that if a rapidly rotating NS is created, then the dominant mode is the quasi-radial mode ("$l = 0$"), radiating through its $l = 2$ piece at frequencies $\sim$ 800Hz-1kHz [11]. Since each type of mode is sensitive to the physical conditions where the amplitude of the mode eigenfunction is greatest, the more information we get from the various classes of modes the better we will understand the details of the star.

Concluding, we should mention that the tidal disruption of a NS by a BH [66] or the merging of two NSs [67] may give valuable information for the radius and the EoS if we can detect the signal at frequencies higher than 1 kHz.

8.4 Gravitational Waves of Cosmological Origin

The monumental discovery of the cosmological microwave background radiation in the mid 1960s had catalytic influence on our understanding of the early Universe. More recent data by COBE and WMAP provide a detailed snapshot of the Universe about 400,000 years after the hot Big Bang, i.e. of the time when the Universe became transparent to electromagnetic radiation. Although this early picture of the Universe provides us with unprecedented information, fundamental processes would have taken place at much earlier times than the photon decoupling. In contrast to electromagnetic signals, gravitational radiation can travel virtually unaffected even if it was produced during the first seconds of the Universe's existence. This is due to the weak

interaction of gravitational waves with matter which is, of course, also a fundamental reason for our difficulty to detect them. A number of processes during the very early stages of the Universe, including quantum fluctuations during inflation, bubble collisions in a first-order phase transition, the decays of cosmic strings and the processes that acted as seeds for galaxy formation, can generate gravitational waves.

Gravitational waves produced in the early Universe will form a stochastic background which can span a very wide range of frequencies. The intensity of a stochastic gravitational-wave background is characterized by the dimensionless quantity

$$\Omega(f) = \frac{1}{\rho_c}\frac{d\rho_{\rm gw}}{d\log f} \ , \tag{8.48}$$

where $\rho_{\rm gw}$ is the energy density of the gravitational waves, f is the frequency and ρ_c is the critical energy density for closing the Universe. The critical density is given as a function of the present value of the Hubble constant H_0 i.e. $\rho_c = 3H_0^2/8\pi G$. A particular gravitational-wave detector will measure the quantity

$$h_c(f) \approx 1.3 \times 10^{-20}\sqrt{\Omega(f)h_0^2}\left(\frac{100{\rm Hz}}{f}\right) \ , \tag{8.49}$$

where h_0 is the rescaled Hubble constant, expected to lie in the range $0.4 \lesssim h_0 \lesssim 0.9$. For example, for $\Omega(f) \sim 10^{-8}$ at a frequency of 100 Hz the strain in a gravitational-wave detector is $h_c \sim 10^{-24}$. It is obvious from equation (8.49) that gravitational radiation with higher frequencies will be more difficult to detect.

There are four frequency bands in which one can search for a stochastic background. First, the detailed analysis of microwave background anisotropies might reveal the existence of inflationary gravitational waves with frequencies $f \lesssim 10^{-16}$Hz. The data from COBE and WMAP were not detailed enough for such an analysis, but future data from PLANCK might reveal the existence of this gravitational-wave background. Analysis of pulsar data can provide hints of the existence of a stochastic background at frequencies $f \sim 10^{-9}$Hz. It will be detected as a timing noise in the electromagnetic signal and one may be able to resolve stochastic noise of order $h_0^2\Omega \sim 10^{-8}$. The space-based gravitational wave detector LISA will be sensitive to stochastic gravitational waves at frequencies of about $10^{-4} - 10^{-3}$Hz. By collecting data for about one year LISA could detect $h_0^2\Omega \sim 10^{-10}$. Finally, combination of data by ground based detectors can reveal the stochastic background at frequencies $10 - 10^3$Hz with sensitivity $h_0^2\Omega \sim 5 \times 10^{-9}$. Detailed reviews on the possible sources of cosmological gravitational waves and the probability of their detection can be found in [68, 69].

Acknowledgements

This work has been supported by the EU Programme 'Improving the Human Research Potential and the Socio-Economic Knowledge Base' (Research Training Network Contract HPRN-CT-2000-00137). KK acknowledges support throught the Center of Gravitational Wave Physics, which is funded by the NSF number cooperative agreement PHY 01-14375.

References

1. C. Cutler, K.S. Thorne, *in proceedings of GR16 (Durban South Africa, 2001)*, gr-qc/0204090.
2. R. Schnabel, J. Harms, K.A Strain, K. Danzmann, Class. Quantum Grav. **21**, S1155 (2004).
3. M. Bonaldi, M. Cerdonio, L. Conti, G.A. Prodi, L. Taffarello, J.P. Zendri, Class. Quantum Grav. **21**, S1155 (2004).
4. M. Burgay, N. D'Amico, A. Possenti, R.N. Manchester, A.G. Lyne, B.C. Joshi, M.A. McLaughlin, M. Kramer, J.M. Sarkissian, F. Camilo, V. Kalogera, C. Kim, D.R. Lorimer, Nature, **426**, 531 (2003).
5. V. Kalogera, C. Kim, D.R. Lorimer, M. Burgay, N. D'Amico, A. Possenti, R.N. Manchester, A.G. Lyne, B.C. Joshi, M.A. McLaughlin, M. Kramer, J.M. Sarkissian, F. Camilo, Astrophys.J. **601**, L179 (2004).
6. E. Cappellaro, M. Turatto, D.Yu. Tsvetov, O.S. Bartunov, C. Pollas, R. Evans and M. Hamuy, A&A **351**, 459 (1999).
7. O. Zanotti, L. Rezzolla, J. A. Font, M.N.R.A.S., **341** 832 (2003).
8. C.L. Fryer, S.E. Woosley, A. Heger, Astrophys. J. **550**, 372 (2001).
9. C.L. Fryer, V. Kalogera, Astrophys. J. **554**, 548 (2001).
10. C.L. Fryer, A. Heger, Astrophys. J. **541** 1033 (2000).
11. H. Dimmelmeier, J.A. Font, E. Muller, A& A **393**, 523 (2002).
12. S.L. Finn, C.R. Evans, Astrophys. J, **351**, 588 (1990).
13. T. Zwerger, E. Müller, A & A **320**, 209 (1997).
14. M. Rampp, E. Müller, M. Ruffert, A& A **332**, 969 (1998).
15. C.L. Fryer, D.E. Holz, S.A. Hudges, Astrophys. J. **565**, 430 (2002).
16. C.D. Ott, A. Burrows, E. Livne, R. Walder, Astrophys. J. **600**, 834 (2004).
17. K. Kotake, S. Yamada, K. Sato K, Phys. Rev. D **68**, 044023 (2003).
18. M. Shibata, Astrophys. J. **595**, 992 (2003).
19. L. Baiotti, I. Hawke, P. . Montero, F. Löffler, L. Rezzolla, N. Stergioulas, J. A. Font, E. Seidel, gr-qc/0403029.
20. P.A. Caraveo, Astrophys. J. **415**, L111 (1993).
21. A. Burrows, J. Hayes, Phys. Rev. Lett. **76**, 1037 (1996).
22. E. Müller, H.-Th. Janka, A& A **317**, 140 (1997).
23. H. C. Spruit, E.S. Phinney, Nature **393**, 139 (1998).
24. L. C. Loveridge, Phys. Rev. D **69**,024008 (2004).
25. P. Hoeflich, A. Khokhlov, L. Wang, J. C. Wheeler, D. Baade, IAU Symposium 212 on Massive Stars, D. Reidel Conf. Series, ed. E. van den Hucht, astro-ph/0207272.
26. Y. T Liu, L. Lindblom, M.N.R.A.S. **324**, 1063 (2001).

27. S. Kobayasi, P. Meszaros, Astrophys. J. **589**,861 (2003).
28. M. Shibata, T. W. Baumgarte, S. L. Shapiro, Astrophys. J. **542**, 453 (2000).
29. J. M. Centrella, K. C. B. New, L.L. Lowe, J.D. Brown, Astrophys. J. Lett. **550** 193 (2001).
30. M. Shibata, S. Karino, Y. Eriguchi, M.N.R.A.S. **334**, L27 (2002).
31. M. Shibata, S. Karino, Y. Eriguchi, M.N.R.A.S. **343**, 619 (2003).
32. M. Saijo, T. W. Baumgarte, S. L. Shapiro, Astrophys. J. **595**, 352 (2002).
33. K.D. Kokkotas, J. Ruoff, N. Andersson: *gr-qc/0212419*
34. S. Chandrasekhar, Phys. Rev. Lett. **24**, 611 (1970).
35. J.L. Friedman, B.F. Schutz, Astroph. J. **222**, 281 (1978).
36. N. Stergioulas, J.L. Friedman, Astrophys. J. **492**, 301 (1998).
37. S. Yoshida, L. Rezzolla, S. Karino, Y. Eriguchi, Astrophys. J. Lett. **568**, 41 (2002).
38. N. Stergioulas, Living Reviews in Relativity **6** 3 (2003).
39. D. Lai, S.L. Shapiro, Astrophys. J. **442**, 259 (1995).
40. L. Lindblom, B. J. Owen, S. M. Morsink, Phys. Rev. Lett., **80** 4843 (1998).
41. B. Owen, L. Lindblom, C. Cutler, B.F. Schutz, A. Vecchio, N. Andersson, Phys. Rev. D **58**, 084020 (1998).
42. N. Andersson, K. D. Kokkotas, B. F. Schutz, Astrophys. J. **510**, 846 (1999).
43. L. Bildsten, G. Ushomirsky, Astrophys. J. **529**, L33 (2000).
44. L. Lindblom, B. J. Owen, G. Ushomirsky, Phys. Rev. D **62**, 084030 (2000).
45. A. K. Schenk, P. Arras, E.E. Flanagan, S.A. Teukolsky, I. Wasserman, Phys. Rev. D **65**, 024001 (2002).
46. L. Rezzolla, F. K. Lamb, S. L. Shapiro, Astrophys. J. **531**, L139 (2000).
47. L. Lindblom, B. Owen, Phys. Rev. D **65**, 063006 (2002).
48. N. Andersson and K. D. Kokkotas, I.J.M.P. D**10**,381 (2001).
49. N. Andersson, Class. Quantum Grav., **20** R105 (2003).
50. N. Andersson, K. D. Kokkotas, N. Stergioulas, Astrophys. J. **307**, 314 (1999).
51. N. Andersson, D.I. Jones, K.D. Kokkotas, N. Stergioulas, Astrophys. J. Lett. **534** L75 (2000).
52. J. Heyl, Astrophys. J. Lett. **574** L57 (2002).
53. R.V. Wagoner, Astrophys. J. Lett. **578** L63 (2002).
54. N. Andersson, D.I. Jones, K. D. Kokkotas, M.N.R.A.S. **337** 1224 (2002).
55. A. Reisenegger, A. Bonacic, Phys. Rev. Lett. **91**, 201103 (2003).
56. R.A. Araya-Góchez *astro-ph/0311001*
57. A. Nagar, G. Diaz, J. A. Pons, J. A. Font, gr-qc/0403077.
58. E. Echeverria, Phys. Rev. D **40** 3194 (1988).
59. L.S. Finn, Phys. Rev. D **8** 3308 (1992).
60. O. Dryer, B. Kelly, B. Krishnan, L. S. Finn, D. Garrison, R. Lopez-Aleman, Class. Quantum Grav. **21**, 787 (2004).
61. M. Davis, R. Ruffini, W.H. Press, R.H. Price, Phys. Rev. Lett.**27**, 1466 (1971).
62. P. Anninos, R. H. Price, J. Pullin, E. Seidel, W.-M. Suen, Phys. Rev. D **52**, 4462 (1995).
63. N. Andersson, K.D. Kokkotas, M.N.R.A.S. **299**, 1059 (1998).
64. K.D. Kokkotas, T. Apostolatos, N. Andersson, M.N.R.A.S. **320** 307 (2001).
65. N. Andersson, G.L. Comer, Phys. Rev. Lett.**87**, 241101 (2001).
66. M. Vallisneri, Phys. Rev. Lett. **84** 3519 (2000).
67. J.A. Faber, P. Grandclement, F.A. Rasio, K. Taniguchi, Phys. Rev. Lett. **89** 231102 (2002).
68. B. Allen in Les Houches School on Astrophysicsl Sources of Gravitational Waves, eds J.A. Marcka and J.P. Lasota, Cambridge University Press (1996).
69. M. Maggiore, Phys. Reports **331**, 283 (2000).

9 Computational Black Hole Dynamics

Pablo Laguna[1] and Deirdre M. Shoemaker[2]

[1] Department of Astronomy and Astrophysics, Institute for Gravitational Physics and Geometry, Center for Gravitational Wave Physics, Penn State University, University Park, PA 16802, USA
[2] Center for Radiophysics and Space Research, Cornell University, Ithaca, NY 14853, USA

Abstract. Over the last decade, advances in computer hardware and numerical algorithms have opened the door to the possibility that simulations of sources of gravitational radiation can produce valuable information of direct relevance to gravitational wave astronomy. One source in particular is believed to be of extreme importance: the inspiral and merger of a binary black hole system. Simulations of binary black hole systems involve solving the Einstein equation in full generality. Such a daunting task has been one of the primary goals of the numerical relativity community. This review article focuses on the computational modelling of binary black holes. It provides a basic introduction to the subject and is intended for non-experts in the area of numerical relativity.

9.1 Introduction

A new era in astronomy will begin once gravitational wave interferometers such as LIGO, GEO, VIRGO, TAMA and, in the future, LISA detect *first light*. It is expected that these detectors will provide a revolutionary view of the Universe, complementary to the electromagnetic perspective. In this new astronomy, the messengers are gravitational waves, ripples in the fabric of spacetime. These waves will have encoded detailed knowledge of the coherent, bulk motions of matter and the vibrations in the curvature of spacetime produced by a vast class of astrophysical sources; compact object binaries, supernovae, spinning neutron stars, gamma ray bursts and stochastic backgrounds are just a few examples of these sources.

Detection of the stochastic background of gravitational waves will revolutionize early universe cosmology and high energy physics by tapping into information unaccessible by any other means. Because gravitational waves interact so weakly, they penetrate much deeper into the universe and its history. If these relic gravitational waves are observed, they will have imprinted on them a snapshot of the very early universe at the time when they were decoupled from the primordial plasma. This will provide the means to investigate the origins of the universe and to measure cosmological parameters from a new perspective. A summary of how ground based gravitational wave detectors do cosmology can be found in [1]. This topic is of such importance that NASA has developed a program to investigate it called "Beyond Einstein", details at http://universe.gsfc.nasa.gov/.

P. Laguna and D.M. Shoemaker, Computational Black Hole Dynamics, Lect. Notes Phys. **653**, 277–298 (2005)
http://www.springerlink.com/

The same reason that gravitational waves have such a long-reach in time and space is the reason they are so difficult to detect. The detection is a formidable undertaking, requiring innovative engineering, powerful data analysis tools and careful theoretical modelling. Our article focuses on the computational modelling of binary black holes. It provides a basic introduction to the subject and is intended for non-experts in the area of numerical relativity. The article is an expanded version of the lecture notes given by one of us at the Second Aegean Summer School. For a comprehensive review, we recommend the one by Baumgarte and Shapiro [2].

Among the sources of gravitational radiation, binary systems consisting of black holes and/or neutron stars are expected to play a dominant role [3]. Over the last couple of decades, advances in numerical algorithms and computer hardware have increased the impact that constructing numerical solutions of the Einstein equation has on astrophysical systems such as black-hole binaries. The ultimate goal is to develop *generic* numerical codes that, given initial data and boundary conditions, will produce a spacetime representing multiple black-hole singularities. The hope is that these codes will be capable of modeling the dynamics of a binary black-hole collision, including inspiral, merger, and ringdown phases.

This article is divided as follows: In Sect. 9.2 we presents the *traditional* introduction to the (3+1) decomposition of the Einstein equation. Since the primary focus is on black hole spacetimes, we will concentrate on the vacuum case. Also in this section, we will briefly discuss the two most popular (3+1) formulations currently used by the numerical relativity community. Section 9.3 is dedicated to black-hole horizons and the infrastructure developed around them that is used in numerical simulations. In particular, we will describe the *excision* method to handle black hole singularities. In Sect. 9.4, we review the Kerr-Schild form of a single black-hole spacetime and its use as a building block for numerical evolutions. In Sect. 9.5, we review the current status of black-hole simulations. Conclusions and a look at what lies ahead in the near future are given in Sect. 9.6.

9.2 Einstein Equation and Numerical Relativity

The essence of Einstein equation is the notion that geometry is related to matter-energy. Specifically,

$$G_{ab} = 8\pi T_{ab} \ , \tag{9.1}$$

where G_{ab} is the Einstein tensor (geometry) and T_{ab} is the stress-energy tensor (matter-energy). The Einstein tensor G_{ab} is an operator acting on the spacetime metric ${}^{(4)}g_{ab}$. This operator is nonlinear, involving first and second derivatives of the metric. Since this review focuses on black holes, we will concentrate on vacuum spacetimes and thus set $T_{ab} = 0$. Also, Latin

indices from the first part of the alphabet will denote spacetime indices, and Latin indices from the middle of the alphabet will denote spatial indices. Geometrical units are assumed.

When one expands the Einstein equation into each of its components, one immediately finds that of the ten equations only six of them contain second time derivatives. Consequently, if one views general relativity as an initial value problem, namely initial data with subsequent unique dynamical evolution, it appears that there are only enough equations for six of the metric components or six combinations of metric components. What happens is that the four components of the Einstein equation involving at the most first time derivatives are constraint conditions that the initial data must satisfy. There are then four metric components or combinations of metric components for which the Einstein equation does not provide evolution equations. As a consequence, one is free to introduce four conditions to fix those quantities. As we shall see, this freedom encapsulates the property in general relativity of performing coordinate transformations without affecting the physical content of the solution.

In numerical relativity, the most popular approach to construct spacetimes is based on a (3+1) decomposition of the Einstein equation. This is the view called *Geometrodynamics* by J.A. Wheeler, referring to the history of the geometry of space-like hypersurfaces. There are other ways to decompose the Einstein equation including conformal and characteristic treatments. The characteristic formulation has been tremendously successful in single black hole and black hole plus matter calculations; however, there are serious complications in applying this method to binary black hole computations. For details, we suggest a review of conformal and characteristic work by Lehner [4].

Under the (3+1) approach, the spacetime manifold $\mathcal{M}$ is sliced or foliated into a sequence of space-like hypersurfaces Σ_τ, parameterized by a scalar function τ as illustrated in Fig. 9.1. The foliation is characterized by a closed one-form $\Omega_a \equiv \nabla_a \tau$ such that

$$\alpha^{-2} \equiv - {}^{(4)}g^{ab} \Omega_a \, \Omega_b \, . \tag{9.2}$$

The function α is called the lapse function. Its geometrical interpretation will become clear later. Associated with Ω_a, here is a *dual* vector N^a such that $\Omega_a \, N^a = 1$. It is straightforward to show that

$$N^a = -\alpha^2 \, {}^{(4)}g^{ab} \, \Omega_b \, . \tag{9.3}$$

Notice also that N^a points in the direction of increasing τ and is a time-like vector,

$${}^{(4)}g_{ab} \, N^a \, N^b = -\alpha^2 < 0 \, . \tag{9.4}$$

The unit-norm vector n^a associated with N^a is $n^a = \alpha^{-1} \, N^a$.

The vector N^a is not the only dual vector to Ω_a such that $\Omega_a \, N^a = 1$. In general, any vector t^a of the form

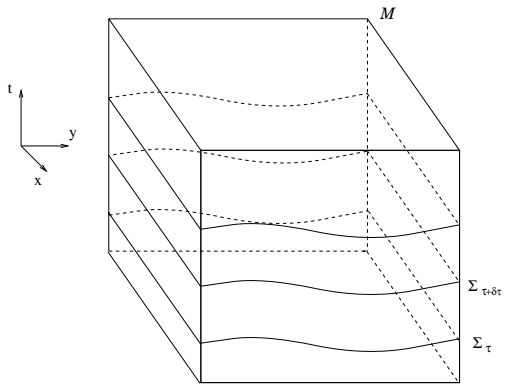

Fig. 9.1. Sketch of the foliation of a three-dimensional spacetime manifold $\mathcal{M}$ into space-like hypersurfaces, Σ_τ, sliced in time.

$$t^a = N^a + \beta^a = \alpha n^a + \beta^a \ , \tag{9.5}$$

with $\beta^a \, \Omega_a = 0$ will yield $\Omega_a \, t^a = 1$. In particular, one can choose the vector t^a to be the tangent to the world lines of coordinates threading the family of hypersurfaces Σ_τ. With this choice, the scalar function α and vector β^a have the following interpretation. Given two hypersurfaces Σ_τ and $\Sigma_{\tau+\delta\tau}$, the proper time between these two hypersurfaces along the normal n^a is $\alpha \, \delta\tau$, here the reason for the name *lapse function*. On the other hand, in a general situation, a coordinate world-line intersecting Σ_τ at a point P will not intersect $\Sigma_{\tau+\delta\tau}$ at a point Q along the normal direction but instead at a point R *shifted* a coordinate distance $\beta^a \, \delta\tau$ from Q (see Fig. 9.2). That is why β^a is called the *shift vector*. It represents the freedom of relabeling coordinate points in subsequent hypersurfaces.

Decomposing the Einstein equations into a (3+1) form reduces to projections into and perpendicular to Σ_τ. This is accomplished with a projection tensor

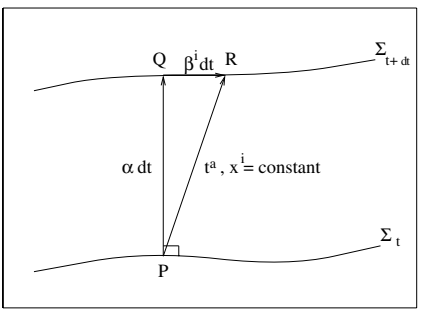

Fig. 9.2. Sketch of two adjacent hypersurfaces (Σ_t and Σ_{t+dt}).

$$\perp^a{}_b \equiv \delta^a{}_b + n^a n_b \ . \tag{9.6}$$

Notice that by construction $\perp^a{}_b n^b = 0$. For instance, given an arbitrary tensor $T^{\cdots}{}_{..a..}$, the projection of the covariant index a is accomplished by $\perp^b{}_a T^{\cdots}{}_{..b..}$, and similarly for other indices. The notation $\perp$ preceding a tensor implies projecting every free index in the tensor. In particular, the projection of the spacetime metric is

$$g_{ab} \equiv \perp\ {}^{(4)}g_{ab} = \perp^c{}_a \perp^d{}_b\ {}^{(4)}g_{cd} = {}^{(4)}g_{ab} + n_a n_b \ . \tag{9.7}$$

The tensor g_{ab} is the intrinsic metric to Σ_τ. Because g_{ab} and β^a are tensors in Σ_τ, one can also write $g_{ij} = g_{ab}$ as well as $\beta^i = \beta^a$. The spacetime metric given as a line element can then be decomposed as follows:

$$ds^2 = -\alpha^2 dt^2 + g_{ij}(dx^i + \beta^i dt)(dx^j + \beta^j dt) \ . \tag{9.8}$$

The spatial metric g_{ab} induces a covariant differentiation operator D_a in Σ_τ. That is, $D_a g_{ab} = 0$ with $D_a = \perp \nabla_a$. For instance

$$D_a T^{b_1 \ldots b_l}{}_{c_1 \ldots c_m} = \perp^{b_1}{}_{d_1} \ldots \perp^{b_l}{}_{d_l} \perp^{e_1}{}_{c_1} \ldots \perp^{e_m}{}_{c_m} \perp^f{}_a \nabla_f T^{d_1 \ldots d_l}{}_{e_1 \ldots e_m} \ . \tag{9.9}$$

In order to completely characterize the spacetime from the point of view of the (3+1) decomposition, it is not enough to have the intrinsic metric g_{ab}. One needs in addition information on how the hypersurfaces are embedded in the spacetime, namely the extrinsic curvature K_{ab} of Σ_τ. The extrinsic curvature is defined as

$$K_{ab} = - \perp \nabla_{(a} n_{b)} = -\frac{1}{2} \perp \mathcal{L}_n g_{ab} = -\frac{1}{2} \mathcal{L}_n\ {}^{(4)}g_{ab} \, , \tag{9.10}$$

where $\mathcal{L}_n$ is the Lie derivative along the normal n^a. It is not difficult to show that

$$K_{ab} = -\nabla_a n_b - n_a\, a_b \ , \tag{9.11}$$

where $a^b = n^a \nabla_a n^b$ is the 4-acceleration of coordinate observers.

Finally, given an arbitrary vector W^a in Σ_τ, the curvature Riemann tensor $R_{abc}{}^d$ is defined via the action of D_a on W_a as:

$$(D_a D_b - D_b D_a) W_c = R_{abc}{}^d W_d \ . \tag{9.12}$$

Similarly, the spatial Ricci tensor R_{ab} and the spatial Ricci scalar R are obtained from $R_{ab} = R_{acb}{}^c$ and $R = R_a{}^a$, respectively.

The relation between the curvature tensor $\mathcal{R}_{abcd}$ of the spacetime manifold $\mathcal{M}$ and the curvature tensor R_{abcd} of hypersurface manifold Σ_τ is given by the Gauss, Codazzi and Ricci equations:

$$\perp \mathcal{R}_{abcd} = R_{abcd} + K_{ac} K_{bd} - K_{ad} K_{bc} \tag{9.13}$$

$$\perp \mathcal{R}_{abc\hat{n}} = D_a K_{ac} - D_a K_{bc} \tag{9.14}$$

$$\perp \mathcal{R}_{a\hat{n}b\hat{n}} = \mathcal{L}_n K_{ab} + \frac{1}{\alpha} D_a\, D_b\, \alpha + K_a{}^c\, K_{ab} \, , \tag{9.15}$$

where we have used the following notation

$$n^a \, T^{\cdots}_{\cdot\cdot a \cdot\cdot} \equiv T^{\cdots}_{\cdot\cdot \hat{n} \cdot\cdot} \, . \tag{9.16}$$

We are now in the position of considering the decomposition of the Einstein tensor $G_{ab} = \mathcal{R}_{ab} - g_{ab}\,\mathcal{R}/2$. Without loss of generality, this tensor can be rewritten as:

$$G_{ab} = \perp G_{ab} - 2 n_{(a} \perp G_{b)\hat{n}} + n_a n_b G_{\hat{n}\hat{n}} \, . \tag{9.17}$$

With the help of (9.13), it becomes

$$2\, G_{\hat{n}\hat{n}} = R + K^2 - K^a{}_b K^b{}_a \, . \tag{9.18}$$

Similarly, making use of (9.14), one obtains

$$\perp G^{a\hat{n}} = -D_b K^{ab} + D^a K \, . \tag{9.19}$$

The next step would be to write $\perp G_{ab}$ in terms of (3+1) quantities. However, as originally pointed out by York [5], it is more convenient to use $\perp \mathcal{R}_{ab}$ instead. A straightforward but tedious calculation yields

$$\perp \mathcal{R}_{ab} = \mathcal{L}_n K_{ab} + \frac{1}{\alpha} D_a \, D_b \, \alpha + 2 \, K_a{}^c \, K_{cb} - K \, K_{ab} - R_{ab} \, . \tag{9.20}$$

Note that (9.18–9.20) involve only spatial tensors; therefore, they can equally be written with spatial indices.

Since we are interested in the vacuum case, the Einstein equation simply reads $G_{ab} = 0$. Equations (9.18–9.20) reduce to

$$R + K^2 - K^i{}_j K^j{}_i = 0 \tag{9.21}$$

$$D_j K^{ij} - D^i K = 0 \tag{9.22}$$

$$\partial_o K_{ij} = -D_i D_j \alpha + \alpha \, (R_{ij} + K K_{ij} - 2 \, K_{il} K^l{}_j) \tag{9.23}$$

$$\partial_o g_{ij} = -2 \, \alpha \, K_{ij} \, , \tag{9.24}$$

where we have introduced the following notation: $\partial_o \equiv \partial_t - \mathcal{L}_\beta$. Notice also that we have added (9.24), which is basically the definition of the extrinsic curvature (see (9.10)).

Equations (9.21) and (9.22) are known as the *Hamiltonian* and *momentum* constraints. Equations (9.23) and (9.24) are also known as the *K-dot* and *g-dot* equations or the ADM evolution equations. The ADM name was originated because of the work on (3+1) decompositions of the Einstein equation by Arnowitt, Deser, and Misner [6]. Strictly speaking calling (9.23) and (9.24) the ADM equations is incorrect. The equations derived by Arnowitt, Deser, and Misner use a extrinsic curvature density and the evolution equation for the densitized extrinsic curvature is derived from $\perp G_{ab}$. To our knowledge, the system (9.21-9.24) as it stands was first introduced by York [5].

The traditional approach in numerical relativity to solve the system of equations (9.21-9.24) for spacetimes containing black hole singularities requires the following action items:

- Specify spatial coordinates $\{x^i\}$ (e.g. Cartesian, spherical, etc.) and topology of the initial slice.
- Construct initial data $\{g_{ij}, K_{ij}\}$ such that the constraints (9.21) and (9.22) are satisfied.
- Prescribe a recipe to fix α and β^i.
- Impose boundary conditions.
- Handling black hole singularities.
- Evolve or update in time the data $\{g_{ij}, K_{ij}\}$ via (9.23) and (9.24).

There are of course a series of important and complicated issues in handling each of the items above. For instance regarding initial data, there are twelve quantities in $\{g_{ij}, K_{ij}\}$ but only four constraint equations. How do we select among these twelve quantities those that are derived from the constraints? York and collaborators [5], based on earlier work by Lichnerowicz, developed an extremely powerful methodology that singles out those pieces in $\{g_{ij}, K_{ij}\}$ that are to be obtained from the constraints (9.21) and (9.22). This method is commonly known as the York *conformal* approach to the initial data problem in general relativity. For an excellent review and modern variations, we encourage the reader to look at the review by Cook [7].

Similarly, there are important aspects to be considered in the choice of the gauge variables α and β^i as well as boundary conditions. Throughout the various disciplines of physics, boundary conditions are often one of the most demanding aspects. Recently in numerical relativity, a great deal of attention has been devoted to boundary conditions that are constraint preserving [8]. This work is attempting to relieve the prevalence of constraint violation during evolutions by applying explicit constraint preserving boundary conditions. Regarding the choice of the lapse and the shift, the numerical relativity community seems to be converging on gauge condition of the driver type [9]. The idea behind these gauge condition is to design evolution equations for *alpha* and β^i that drive certain quantities to a *steady state*. These conditions have been successfully used in both black hole and neutron star calculations.

The most common procedure to evolve initial data in numerical relativity is the so called *free evolution*. In a free evolution, the constraint equations are only used to monitor the quality of the evolved data, namely the departures of the numerically data from satisfying the constraints. Because of truncation or discretization errors, the evolved data will satisfy at best the discrete version of the constraints. More often, these errors drive the solutions rapidly away from the constraint surface, rendering the simulation unstable. It is fair to say that one of the primary tasks in numerical relativity has been the control of these constraint violating modes [10, 11]. Recently there have been attempts to explicitly solve the constraints as a way of projecting the evolved data back into the constraint surface [12].

The ADM formulation is not unique. There are in principle an infinite number of ways of formulating the Einstein equation with a (3+1) structure. One can simply add terms to the evolution equations involving the constraints

to create a different set of evolution equations. For instance,

$$\partial_o g_{ij} = \text{r.h.s. Eq. (9.24)} + P_{ij}\,\rho + U^k{}_{ij}\,J_k \tag{9.25}$$
$$\partial_o K_{ij} = \text{r.h.s. Eq. (9.23)} + Q_{ij}\,\rho + V^k{}_{ij}\,J_k\,, \tag{9.26}$$

where

$$\rho \equiv R + K^2 - K^i{}_j K^j{}_i \tag{9.27}$$
$$J^i \equiv D_j K^{ij} - D^i K\,. \tag{9.28}$$

Above, P_{ij}, Q_{ij}, $U^k{}_{ij}$ and $V^k{}_{ij}$ are tensors, symmetric in the ij indices involving parameters as well as metric functions [13]. These types of adjustment have a deep impact on the well-posedness properties of the system. Among all the formulations that have been proposed, there are two that are currently enjoying widespread popularity. We will quickly summarize both. For an extensive review on hyperbolic formulations of the Einstein equation the reader should consult Reula's review [14].

The first of the two formulations is the so called BSSN system. This is a formulation re-introduced by Baumgarte and Shapiro [15] and originally developed by Shibata and Nakamura [16]. The first step in obtaining the BSSN formulation is to abandon g_{ij} and K_{ij} as primary variables and work instead with Φ, $\hat{g}_{ij}$, K and $\hat{A}_{ij}$. The relationships between these and the ADM variables are

$$\Phi = \frac{1}{6}\,\ln g^{1/2} \tag{9.29}$$
$$\hat{g}_{ij} = e^{-4\Phi}\,g_{ij} \tag{9.30}$$
$$\hat{A}_{ij} = e^{-4\Phi}\,A_{ij}\,, \tag{9.31}$$

where $A_{ij} = K_{ij} - g_{ij}\,K/3$. Given (9.29)-(9.31), it is not difficult to show that the ADM system of evolution equations take the form

$$\partial_o \Phi = -\frac{1}{6}\alpha\,K \tag{9.32}$$
$$\partial_o \hat{g}_{ij} = -2\,\alpha\,\hat{A}_{ij} \tag{9.33}$$
$$\partial_o K = -\nabla_i\nabla^i\alpha + \alpha\,(\hat{A}_{ij}\hat{A}^{ij} + K^2/3) \tag{9.34}$$
$$\partial_o \hat{A}_{ij} = e^{-4\Phi}\,(-\nabla_i\nabla_j\alpha + \alpha\,R_{ij})^{TF} + \alpha\,(K\hat{A}_{ij} - 2\hat{A}_{il}\hat{A}^l{}_j)\,. \tag{9.35}$$

where in (9.35) the superscript TF denotes the trace free part of the tensor between brackets, e.g. $T^{TF}_{ij} \equiv T_{ij} - g_{ij}g^{kl}T_{kl}/3$ for any tensor T_{ij}. The key ingredient of the BSSN system is the introduction of a conformal connection

$$\widehat{\Gamma}^i \equiv \hat{g}^{jk}\widehat{\Gamma}^i_{jk} = -\partial_j \hat{g}^{ij}, \tag{9.36}$$

where the $\widehat{\Gamma}^i_{jk}$ are the connection coefficients associated with $\hat{g}_{ij}$. The motivation for introducing these connections is to substitute in the Ricci tensor $\hat{R}_{ij}$ all the derivatives of $\hat{g}^{jk}\widehat{\Gamma}^i_{jk}$ in favor of derivatives of $\widehat{\Gamma}^i$. In doing so, the Ricci tensor yields a system of evolution equations with a *hyperbolic flavor*.

Expanding the number of primary variables in the system to include the conformal connections implies the need of an additional evolution equation. From definition (9.36), one can show that the evolution equation for $\widehat{\Gamma}^i$ is given by

$$\begin{aligned}\partial_o \widehat{\Gamma}^i = {} & \hat{g}^{jk}\partial_{jk}\beta^i + \frac{1}{3}\hat{g}^{ij}\partial_{jk}\beta^k - 2\,\hat{A}^{ij}\partial_j\alpha \\ & + 2\,\alpha\widehat{\Gamma}^i_{jk}\hat{A}^{jk} + 12\,\alpha\hat{A}^{ij}\partial_j\Phi - \frac{4}{3}\alpha\widehat{\nabla}^i K\,. \end{aligned} \tag{9.37}$$

In summary, the BSSN system involves conformal transformations, a trace-free decomposition of the extrinsic curvature, the introduction of a conformal connection and the use of constraints to eliminate the Ricci scalar and derivatives of the trace-free extrinsic curvature.

The other currently most popular (3+1) formulation of the Einstein equation is the KST formulation developed by Kidder, Scheel and Teukolsky [17]. There are a number of ways to construct hyperbolic formulations [14], but we only include the KST formulation in any detail. Strictly speaking this is a family of hyperbolic formulations foliated by twelve parameters. This formulation has been demonstrated to be successful in the evolution of single black holes. It is expected that in the near future this success will be translated to evolutions of binary black holes.

The first step in the derivation of the KST system consists of introducing a new variable to eliminate second derivatives of the spatial metric. The new variable (symmetric on its last two indices) is

$$d_{kij} \equiv \partial_k g_{ij}\ , \tag{9.38}$$

and its traces $d_k \equiv g^{ij}d_{kij}$ and $b_k \equiv g^{ij}d_{ijk}$. An evolution equation for d_{kij} is obtained by taking a spatial derivative of (9.24). This yields

$$\partial_o d_{kij} = -2\alpha\partial_k K_{ij} - 2K_{ij}\partial_k\alpha\,. \tag{9.39}$$

In terms of the new variables, the evolution equation for the extrinsic curvature becomes

$$\begin{aligned}\partial_o K_{ij} = {} & \alpha\Big[\frac{1}{2}g^{kl}\left(\partial_{(i}d_{klj)} + \partial_k d_{(ij)l} - \partial_k d_{lij} - \partial_{(i}d_{j)kl}\right) \\ & + \frac{1}{2}b^k d_{kij} - \frac{1}{4}d^k d_{kij} - b^k d_{(ij)k} - \frac{1}{2}d_{kj}{}^l d_{li}{}^k \\ & + \frac{1}{2}d^k d_{(ij)k} + \frac{1}{4}d_i{}^{kl}d_{jkl} + \frac{1}{2}d^{kl}{}_i d_{klj} - 2K_{ik}K_j{}^k \\ & + KK_{ij}\Big] - \partial_i\partial_j\alpha - \frac{1}{2}d^k{}_{ij}\partial_k\alpha + d_{(ij)}{}^k\partial_k\alpha\,. \end{aligned}$$

By adding terms proportional to the constraints, the evolution equations for K_{ij} and d_{kij} can be rewritten as

$$\partial_o K_{ij} = (\ldots) + \gamma\alpha g_{ij}\mathcal{C} + \zeta\alpha g^{kl}\mathcal{C}_{k(ij)l} \tag{9.40}$$

$$\partial_o d_{kij} = (\ldots) + \eta\alpha g_{k(i}\mathcal{C}_{j)} + \chi\alpha g_{ij}\mathcal{C}_k \ , \tag{9.41}$$

where $(\ldots)$ represents the right-hand side of either (9.39) or (9.40). The parameters $\{\gamma, \zeta, \eta, \chi\}$ are arbitrary constants. The evolution equations are now given by

$$\partial_o g_{ij} \simeq 0, \tag{9.42}$$

$$\begin{aligned}\partial_o K_{ij} \simeq & -\frac{1}{2}\alpha g^{kl}\left[\partial_k d_{lij} - (1+\zeta)\partial_k d_{(ij)l}\right. \\ & - (1-\zeta)\partial_{(i}d_{klj)} + (1+2\sigma)\partial_{(i}d_{j)kl} \\ & \left. -\gamma g_{ij}g^{mn}\partial_k d_{mnl} + \gamma g_{ij}g^{mn}\partial_k d_{lmn}\right],\end{aligned} \tag{9.43}$$

$$\begin{aligned}\partial_o d_{kij} \simeq & -2\alpha\partial_k K_{ij} + \alpha g^{lm}\left(\eta g_{k(i}\partial_l K_{mj)} + \chi g_{ij}\partial_l K_{mk}\right. \\ & \left. -\eta g_{k(i}\partial_{j)}K_{lm} - \chi g_{ij}\partial_k K_{lm}\right),\end{aligned} \tag{9.44}$$

where $\simeq$ denotes equal to the principal part. In [17] it is shown that one can find values of the parameters such that the system is weakly hyperbolic. It is also found that densitizing the lapse is a necessary condition for strong hyperbolicity, namely:

$$\alpha = g^{\sigma}\, e^{Q}\,, \tag{9.45}$$

with Q an arbitrary function and σ a parameter.

Finally, in order to arrive at the KST system, one introduces a generalized extrinsic curvature P_{ij} using the relation

$$P_{ij} \equiv K_{ij} + \hat{z} g_{ij} K \ , \tag{9.46}$$

where $\hat{z}$ is an arbitrary parameter. Similarly one introduces a generalized derivative of the metric, M_{kij}, using the relation

$$\begin{aligned}M_{kij} = & \frac{1}{2}\left\{\hat{k}d_{kij} + \hat{e}d_{(ij)k} + g_{ij}\left[\hat{a}d_k + \hat{b}b_k\right]\right. \\ & \left. + g_{k(i}\left[\hat{c}d_{j)} + \hat{d}b_{j)}\right]\right\}\ ,\end{aligned} \tag{9.47}$$

with $\hat{k}$, $\hat{e}$, $\hat{a}$, $\hat{b}$, $\hat{c}$ and $\hat{d}$ additional parameters. With these definitions, the principal parts of the evolution equations for P_{ij} and M_{kij} are

$$\partial_o g_{ij} \simeq 0 \tag{9.48}$$

$$\begin{aligned}\partial_o P_{ij} \simeq & -\alpha g^{kl}\left(\mu_1\partial_k M_{lij} + \mu_2\partial_k M_{(ij)l} + \mu_3\partial_{(i}M_{klj)}\right. \\ & + \mu_4\partial_{(i}M_{j)kl} + \mu_5 g_{ij}g^{mn}\partial_k M_{mnl} \\ & \left. +\mu_6 g_{ij}g^{mn}\partial_k M_{lmn}\right)\end{aligned} \tag{9.49}$$

$$\begin{aligned}\partial_o M_{kij} \simeq & -\alpha\left(\nu_1\partial_k P_{ij} + \nu_2\partial_{(i}P_{j)k} + \nu_3 g^{mn}g_{k(i}\partial_m P_{nj)}\right. \\ & + \nu_4 g_{ij}g^{mn}\partial_m P_{nk} + \nu_5 g^{mn}g_{k(i}\partial_{j)}P_{mn} \\ & \left. +\nu_6 g_{ij}g^{mn}\partial_k P_{mn}\right)\ ,\end{aligned} \tag{9.50}$$

where μ_A and ν_A with $A = 1...6$ are parameters function of the 12 parameters above.

9.3 Black Hole Horizons and Excision

Back hole horizons are of crucial importance in numerical simulations that may either contain or lead to the formation of black holes. Specifically in numerical relativity, horizons are used to identify the formation of a black hole, to locate a black hole or to characterize properties of a black hole. To successfully utilize horizons in computations, they must be formulated in terms appropriate to numerical approximations. We will briefly review the types of black hole horizons that have been used in connection with numerical simulations.

A black hole is most often defined in terms of its *event horizon*, *i.e.* the future boundary of the causal past of future null infinity. The event horizon is a mathematically elegant and powerful definition of the black hole surface; however, for the purposes of locating in a numerical simulation the position of the black hole in the computational domain this definition is not always useful. By definition, the event horizon is global in nature; meaning, the entire spacetime must be known *a priori* in order to determine its location. This is demanding the end product of the evolution before we even begin. In practice, an alternative surface called *apparent horizon* is used to localize or track the position of a black hole.

An apparent horizon is a closed two-sphere on Σ_τ. Therefore, it is well suited for numerical relativity since it only requires information available at an instant of time. There is a price to pay however. Because they are defined from quantities in Σ_τ, they dependent on the way one chooses to foliate the spacetime. One can in principle find foliations of a black hole spacetime in which a hypersurface Σ_τ does not contain an apparent horizon. Fortunately, these type of slicings seem to be rare. Furthermore, the world tube connecting apparent horizons from one time to the next could be discontinuous. This is in contrast with an event horizon which is a continuous worldtube.

The apparent horizon is defined as the outermost marginally trapped surface in Σ_τ. As mentioned before, the definition of an apparent horizon requires only knowledge of quantities on Σ_τ. Following [18], let S be a surface with S^2 topology with k^a and l^a respectively the outgoing and ingoing null vectors to S, see Fig. 9.3. That is, $k^a s_a > 0$ and $l^a s_a < 0$, where s^a is a spacelike unit vector to S in Σ_τ, namely $s^a n_a = 0$ with n^a the unit time-like normal to Σ_τ.

A trapped surface is defined as

$$\Theta \equiv \nabla_a k^a \leq 0 \ . \tag{9.51}$$

By making use of $k^a = n^a + s^a$ and the projector $\perp$, (9.51) can be rewritten in terms of quantities on Σ_τ as follows:

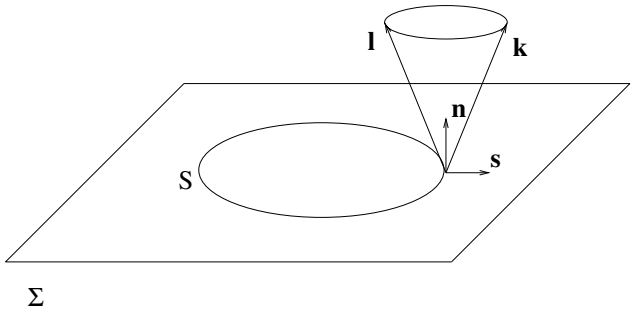

Fig. 9.3. Representation of a two-Sphere embedded in a hypersurface, Σ_τ.

$$\Theta = D_i s^i - K + s^i s^j K_{ij}\,, \tag{9.52}$$

where as before D_i denotes the covariant derivative associated with the 3-metric g_{ij} in Σ_τ. For a marginally trapped surface, $\Theta = 0$. An apparent horizon is the *outermost* marginally trapped surface. Thus, finding marginally trapped surfaces involves finding solutions to

$$D_i s^i - K + s^i s^j K_{ij} = 0\,. \tag{9.53}$$

There are many ways to solve this equation, see reference [19] and those therein.

The motivation for finding apparent horizons during numerical evolutions of black hole spacetimes is to localize the singularity intrinsic to the black hole on each space-like hypersurface. Numerically, the singularity must be treated specially since infinite gradients are impossible to handle in calculating derivatives of the fields. Once the apparent horizon is located by solving (9.53), a numerical code can use this information to avoid computing near the singularity contained within the horizon. One approach to deal with the singularity is known as *singularity avoidance*. In this technique, one takes advantage of the freedom in foliating the spacetime to construct coordinates that avoid the singularity. Originally, this method encountered problems caused by the increase in proper separation between neighboring points, also known as *grid stretching*. However, the problem has been alleviated to some extent through clever choice of shift [20].

An alternative and perhaps more robust method is to physically remove or *excise* the singularity from the computational domain. Physically, this procedure requires no boundary conditions as it respects the causality of the spacetime, *i.e.* events inside the horizon are not in causal contact with external events. This means that all physics information within the boundary cannot escape and consequently can be ignored. However, this does not apply to non-physical information such as gauge modes. Unless all the characteristic speeds of the evolved fields are within the light code, these modes

can propagate out and affect the numerical stability. Excision has been implemented in many different formulations [21, 22, 23, 24, 25, 26, 27]. With the recent work on formulations, excision algorithms have been stably coded in hyperbolic and BSSN-type codes in stationary and dynamic back hole systems [28, 29, 30, 31, 32].

Invariant physical information contained in the source simulations must be both extracted and interpreted if we are to construct a complete picture of gravitational physics in the strong field regime. This is a daunting challenge in general relativity where there is freedom in the choice of gauge and the form of the equations.

Of particular interest is the quantification of the mass M and angular momentum J of a black hole. One way to attribute a mass and an angular momentum to a black hole is to calculate the corresponding ADM quantities at infinity. The main difficulty is that the ADM mass and angular momentum refer to the whole spacetime. In a dynamical situation, such a spacetime will contain gravitational radiation and it is not clear how much of the mass or angular momentum should be attributed to the black hole itself and, if there is more than one black hole, to each individual black hole. It is desirable to have a framework that combines the properties of apparent horizons with the powerful tools available at infinity. In the regime when the black hole is isolated in an otherwise dynamical spacetime, such a framework now exists in the form of *isolated horizons* [33, 34]. Isolated horizons provide a way to identify a black hole quasi-locally and allow for the calculation of M and J. It has been shown recently that the formula for angular momentum and mass arising from the isolated horizon formalism are valid even in dynamical situations [35, 36].

The theory of dynamical and isolated horizons gives rise to definitions for M and J that are similar in form to the ADM definitions, but are calculated at the dynamical horizon Δ:

$$J_\Delta = \frac{1}{8\pi} \oint_S (\varphi^a s^b K_{ab}) \, \mathrm{d}^2 V \,, \tag{9.54}$$

where K_{ab} is the extrinsic curvature on S and φ^a is a Killing vector related to the fact that S must by axisymmetric in order for angular momentum to be defined. In [37] a method for calculating φ^a based on the Killing transport equation is detailed.

Given J_Δ, the horizon mass M_Δ is obtained from [38, 39]

$$M_\Delta = \frac{1}{2R_\Delta} \sqrt{R_\Delta^4 + 4J_\Delta^2} \,, \tag{9.55}$$

where R_Δ is the area radius of the horizon: $R_\Delta = (A_\Delta/4\pi)^{1/2}$. This formula depends on R_Δ and J_Δ in the same way as in the Kerr solution. However, this is a result of the calculation and not an assumption. Furthermore, under some physically reasonable assumptions on fields near future time-like infinity

(i^+), one can show that $M_\Delta - M_{ADM}$ is equal to the energy radiated across future null-infinity if the isolated horizon extends all the way to i^+. Thus, M_Δ is the mass left over after all the gravitational radiation has left the system. This lends further support for identifying M_Δ with the mass of the black hole.

9.4 Initial Data and the Kerr-Schild Metric

There are two general approaches to represent black holes for the construction of initial data. One is based on *punctures* [40] and the other on using the Kerr-Schild form of the single back hole solution to Einstein's equations [41]. They both have advantages and disadvantages. There are two positive reasons for using the Kerr-Schild form: (1) the metric is regular at the horizon and (2) the metric is Lorentz form-invariant under boosts. We will concentrate the discussion on the Kerr-Schild approach. For puncture data, see the review by Cook [7].

The Kerr-Schild metric is given by

$$ {}^{(4)}g_{ab} = \eta_{ab} + 2\,H\,l_a\,l_b\,, \tag{9.56}$$

with l^a a null vector with respect to both ${}^{(4)}g_{ab}$ and the flat metric η_{ab}. In a (3+1) form, this metric takes the form

$$ g_{ij} = \eta_{ij} + 2Hl_il_j \tag{9.57}$$

$$ \alpha = \frac{1}{\sqrt{1+2Hl_t^2}} \tag{9.58}$$

$$ \beta_i = 2Hl_tl_i\ . \tag{9.59}$$

The relation between the lapse and shift dictates that the horizon stays at a constant coordinate location in a non-boosted Kerr-Schild solution.

The Kerr-Schild metric is form-invariant under a Lorentz transformation. Consider the transformation matrix Λ

$$ \Lambda^{\bar{t}}_{t} = \gamma \tag{9.60}$$

$$ \Lambda^{\bar{t}}_{i} = -v\gamma\hat{v}_i \tag{9.61}$$

$$ \Lambda^{\bar{i}}_{j} = (\gamma-1)\hat{v}^i\hat{v}_j + \eta^i_j\ , \tag{9.62}$$

with $\eta_{ij}\hat{v}^i\hat{v}^j = 1$ and $\gamma = 1/\sqrt{1-v^2}$. With this transformation, a new null vector l_μ and a new function H are determined

$$ l_\mu = \Lambda^{\bar{\nu}}_{\mu}\bar{l}_{\bar{\nu}} \tag{9.63}$$

$$ H = \bar{H}(\Lambda^{\bar{\nu}}_{\mu}\bar{x}_\nu)\ . \tag{9.64}$$

The form of the spacetime metric and its (3+1) decomposition remains unchanged in terms of l_μ and H. To give a specific example, consider the case of a non-rotating black hole boosted with a velocity v in the x-direction. Then

$$\bar{t} = \gamma(t - vx) \tag{9.65}$$
$$\bar{x} = \gamma(x - vt) \tag{9.66}$$
$$\bar{y} = y \tag{9.67}$$
$$\bar{z} = z \ , \tag{9.68}$$

where v is the boost velocity and $\gamma = 1/\sqrt{1-v^2}$ with $c = 1$. After the boost, l_i becomes

$$l_i = \partial_i(M/H) - \gamma v_i \ . \tag{9.69}$$

Since H is a scalar, it is invariant under this transformation. The null vector and scalar function become explicitly

$$r^2 = \gamma^2(x - vt)^2 + y^2 + z^2 \tag{9.70}$$
$$l_t = \gamma(1 - v\gamma(x - vt)/r) \tag{9.71}$$
$$l_x = \gamma(\gamma(x - vt)/r - v) \tag{9.72}$$
$$l_y = y/r \tag{9.73}$$
$$l_z = z/r \tag{9.74}$$
$$H = M/r \ . \tag{9.75}$$

The new metric is given as before, namely (9.57–9.59) with l_μ and H above.

Given the boosted solution of a single black hole in Kerr-Schild form, one can construct data representing binary black holes by superposing two Kerr-Schild solutions [42]. If the initial data is obtained following York's conformal approach [5], the freely specifiable data is the conformal metric $\hat{g}_{ij}$, trace of the extrinsic curvature K and the conformal, transverse, traceless part of the extrinsic curvature $\hat{A}^*_{ij}$. One can then set the conformal metric to be

$$\hat{g}_{ij} = {}_{(1)}g_{ij} + {}_{(2)}g_{ij} - \eta_{ij} \ , \tag{9.76}$$

with

$${}_{(1)}g_{ij} = \eta_{ij} + H\, l_i\, l_j|_{(1)} \tag{9.77}$$
$${}_{(2)}g_{ij} = \eta_{ij} + H\, l_i\, l_j|_{(2)} \ , \tag{9.78}$$

being the the isolated Kerr-Schild metric forms with l_i and H corresponding to the single black holes. The arguments of H and l_j are

$${r_1}^2 = (x - x_1)^i (x - x_1)^j \eta_{ij} \tag{9.79}$$
$${r_2}^2 = (x - x_2)^i (x - x_2)^j \eta_{ij} \ , \tag{9.80}$$

with ${x_1}^i$ and ${x_2}^j$ the coordinate positions of the holes on the initial slice.

Similarly the trace of the extrinsic curvature K can be obtained from

$$\hat{K}^i_j = {}_{(1)}\hat{K}^i_j + {}_{(2)}\hat{K}^i_j \ , \tag{9.81}$$

where ${}_{(1)}\hat{K}^i_j$ and ${}_{(2)}\hat{K}^i_j$ are the individual extrinsic curvatures associated with the individual Kerr-Schild metrics. That is, one sets $K = \hat{K}(1) + \hat{K}(2)$. Given $\hat{g}_{ij}$ and K, one is in the position to apply York's conformal approach.

While both the puncture and superposed Kerr-Schild methods have been used to generate initial data for binary black hole evolutions, it is important to note that posing astrophysically relevant initial data in numerical relativity is very much an open question. Correct initial data would be supplied by Post-Newtonian calculations with numerical evolutions taking over where the Post-Newtonian approximations break down. This is not yet possible with current codes and computer technology. For this reason, it is important to develop physical insight about how much spurious radiation is present in the initial data. Some preliminary work in this area has been done [43].

9.5 Black Hole Evolutions

The first milestone to be achieved for the successful orbit and merger of two black holes is the stable, long term evolution of a single, static black hole. Despite the fact that the solution for a charge free single black hole is known to be either Schwarzschild or Kerr, the numerical relativity community has been struggling to obtain a generic, three-dimensional stable evolution for years. Early on in the effort, two types of evolutions were pursued, those using the standard ADM formulation [44, 21, 22] and those using the characteristic formulation [45, 23]. Characteristic formulations achieved astounding success at evolving single black holes for arbitrary amounts of time while the (3+1) suffered severe stability problems when applied in three spatial dimensions.

Over the last several years, the community has come to understand that the problems were not solely numerical but originated in part from the structure of the equations. The result is that (3+1) codes can now evolve single black holes stably in three dimensions. With a code based on the KST hyperbolic formulation, the Cornell/Caltech group has been able to achieve evolution of a single black hole for $(600 - 8000)M$ depending on the coordinates chosen [17]. Here constraint violations were tracked and determined to be the possible culprit in the failure to evolve a single black hole in (3+1) for so many years. Similarly, several groups with codes based on the BSSN form of equations carried out black hole evolutions for hundreds to thousands of M [46, 47, 31]. A crucial aspect in these simulations has the use of dynamical (driver) gauge conditions, as well as a densitized lapse [31].

In order to successfully compute the last few orbits and merger of two black holes, it is highly likely that one would have to develop a code capable of moving excised black holes through the computational domain. While much of the current effort on orbits and head-on collisions has been accomplished by holding the black holes fixed to the grid, it is our believe that adding the flexibility for black holes to move or drift will greatly facilitate coalescence simulations. This implies the development of excision algorithms

that populate points on the numerical grid with field values where there were none at the previous time step [30].

Characteristic formulations had early success at moving black holes through the grid [23]. Although there was an early attempt at moving black holes in (3+1) formulations [21], success depended on recent improvements to the formulation of the (3+1) equations that has allowed the stable evolution of a stationary black hole. As a test of moving excision, evolutions of a scalar field in a static or boosted Kerr-Schild black hole background [29] have been accomplished. Further, in [30, 31], full three-dimension evolutions were performed of a single black hole in a coordinates systems in which the coordinate location of the black hole did not remained fixed.

A truly dynamical test of codes designed to evolve black holes is the evolution of distorted black hole spacetimes. Distorted black holes are dynamical systems demanding many of the same technical and analytic developments that binary black hole systems do. In addition there are interesting physics regarding gravitational radiation and the dynamics of horizons that can be investigated with these systems. Many three dimensional simulations of distorted black holes have been carried out [48, 46, 49, 50, 51]. Typically, the black hole is distorted by a wave, often a Brill wave [52]. The simulations were primarily intended at testing gravitational wave extraction techniques and the stability properties of the codes. Recently, distorted black holes [46] were used to test gauge conditions. These results showed a match to the lowest two quasi normal mode frequencies.

Highly distorted black holes provide a mechanism for a detailed characterization of the transition from a highly, nonlinear distorted black hole to a ringing black hole and to the onset of quasi normal mode frequencies. Allen and collaborators [51] probed the nonlinear generation of harmonics for small amplitudes of the ingoing wave. More recently, Papadopoulos [53] and Zlochower and his collaborators [54] have evolved distorted black holes using the Characteristic framework. Both report interesting nonlinear effects such as mode mixing, larger phase shifts and amplitudes that may lend insight into the observations of gravitational waves.

The effort on simulating head-on collisions provides a good picture of the early history of vacuum numerical relativity. The first attempt to solve this problem took place as early as 1964 [55]. This and subsequent efforts on head-on collision were carried out in axisymmetry, namely as a (2+1) problem [56, 25]. It was not until the late 90's that computer speed and memory was such that 3D simulations were possible [20].

Despite the early difficulties in evolving single black holes stably, some groups were capable of carrying out evolutions of black hole grazing collisions. Grazing collisions refer to collisions of slightly off-center head-on black hole collisions. The initial separation of the black holes was limited by the onset of instabilities. Separation of the black holes were such that a joint apparent horizon formed early in the simulation. These tests, however, demonstrated the ability to do non-axisymmetric dynamic binary black hole simulations.

Grazing collisions continue to be a step toward the orbit and merger of two black holes. Brügmann [57] completed the first grazing collision using the Brandt and Brügmann puncture method [40] to avoid the singularities during the computation. Although this evolution ended too prematurely to be of astrophysical interest, it was the first three-dimensional binary black hole evolution. The first attempt to use dynamic singularity excision in grazing collisions was carried out by the Penn State/Pittsburgh/Texas collaboration [22] using a code based on the standard ADM formulation. The initial data for these simulations was the superposed Kerr-Schild data [42, 58]. These simulations were successful in demonstrating the use of dynamic excision to follow two back holes as they merge; however, the results were too short-lived and affected by boundary effects to allow for gravitational wave extraction. Soon after, a second grazing collision was completed by the AEI group at Germany [59]. This grazing collision did not used excision but punctures to handle the black hole singularities. A combination of large amount of supercomputing power and coordinates to push outer boundaries sufficiently far allowed wave extraction to be obtained in these simulations.

Given the limitations of performing long-lived binary evolutions, the Lazarus group [60] developed a framework to extend the life of the simulations by connecting at the end of fully nonlinear calculations a perturbative calculation. Using an ADM code, the Lazarus effort was successful in constructing waveforms from head-on [61] and merger from ISCO [62]. What appears to be the first fully nonlinear evolution of an orbit of two black holes was performed only recently [63]. The simulation uses corotating coordinates and dynamical shift conditions that force the black holes to stay fixed on the grid. The simulation lasts about one orbital period before crashing. While this works does not yet provide detailed waveforms useful to data analysis effort, they however demonstrate the advances made over the last few years.

9.6 Conclusions and Future Work

Over the last decade or more, the numerical relativity community has focused most of its attention to the binary black hole problem. Although early efforts were seriously plagued by numerical instabilities, simulations in which these instabilities are been tamed is gaining momentum. There are still many open problems before simulations of orbits and mergers, such as the ones reported above, can be enhanced to the point in which useful astrophysical predictions are possible. One of the major obstacles is specifying the astrophysically relevant initial data; this requires a framework for incorporating post-Newtonian information into numerical source simulations. A second obstacle is accurate wave extraction, namely the translation of numerically evolved quantities into physical invariants. In addition, we are lacking good outer boundary conditions although it is encouraging to see that there has been work toward the design of constraint-preserving boundary conditions. There are more difficul-

ties ahead, but the difference between today and five years ago is that today the challenges are met with codes that can successfully evolve single black hole, as well as modest binary black hole evolutions.

Of great significance has been the progress made in formulating the (3+1) decomposition of the Einstein equation and dynamical gauge conditions. With the advent of adaptive mesh refinements (AMR), the outlook for numerical relativity is positive and fast paced. AMR is essential if one has any hope of simulating astrophysical black hole orbits. Currently the effort has been focused on fixed mesh refinement [63, 64, 65]. Adaptivity can be also achieved by other means beside AMR. For instance, the groups at Meudon and Caltech/Cornell have pioneered the use of pseudo-spectral methods having this in mind [66, 67, 68]. Preliminary work is currently taking place also on the use of finite element techniques, a method with a high degree of intrinsic adaptivity.

The detection of gravitational radiation will supply us with means to probe the fundamental nature of gravity. The information obtained from source simulations will have a profound impact to the data analysis community now searching for evidence of gravitational waves in the ground-based detector's data stream. With the possibility of high quality source simulations, the community must begin to interface with data analysis. While it is likely that the outcome from numerical simulations to be handed to the data analysis community will not be an exhaustive collection of waveforms, source simulations could nonetheless provide robust information (e.g. frequency evolutions, mode content, etc.) of extremely high value to the observational effort.

Acknowledgments

This work was supported by NSF grants PHY-9900672 and PHY-0312072 at Cornell and PHY-9800973 and PHY-0114375 at Penn State. Work supported in part by the Center for Gravitational Wave Physics funded by the National Science Foundation under Cooperative Agreement PHY-0114375. The authors would like to thank the organizers of this conference for the opportunity to be part of the Second Aegean school and to visit such a beautiful place as Syros.

References

1. M. Maggiore, Gravitational Wave Experiments and Early Universe Cosmology, *Phys. Reports*, **331**, 283 (2000).
2. T.W. Baumgarte and S.L. Shapiro. Numerical relativity and compact binaries. *Phys. Reports*, **376**, 41 2003.
3. C. Cluter and K.S. Thorne. An overview of gravitational-wave sources. In *Proceeding of GR16*, 2001.

4. L. Lehner, Numerical relativity: A review, Class. Quant. Grav. **18**, R25 (2001).
5. J. W. York Jr, Kinematics and dynamics of general relativity, In L. L. Smarr, editor, *Sources of gravitational radiation*, pages 83–126. Cambridge University Press, Cambridge, (1979).
6. R. Arnowitt, S. Deser, and C.W. Misner, The dynamics of general relativity, In L. Witten, editor, *Gravitation an introduction to current research*, pages 227–265. John Wiley, New York, (1962).
7. G.B. Cook, Living Rev. Rel. **5**, 1 (2000).
8. G. Calabrese and O. Sarbach, J. Math. Phys. **44**, 3888 (2003).
9. M. Alcubierre, Class. Quant. Grav. **20**, 607 (2003).
10. M. Tiglio, L. Lehner, and D. Nilsen, gr-qc/0312001 (2003).
11. L. Lindblom, M. A. Scheel, L. E. Kidder, H. P. Pfeiffer, D. Shoemaker, and S. A. Teukolsky, gr-qc/0402027, (2004).
12. M. Anderson amd R.A. Matzner, gr-qc/0307055 (2003).
13. H. Shinkai and G. Yoneda, gr-qc/0209111, (2002).
14. O. Reula, Living Rev. Rel. **3**, 1 (1998).
15. T.W. Baumgarte and S.L. Shapiro, Phys. Rev. D **59**, 024007 (1999).
16. M. Shibata and T. Nakamura, Phys. Rev. D **52**, 5428 (1995).
17. L.E. Kidder, M.A. Scheel, and S.A. Teukolsky, Extending the lifetime of 3d black hole computations with a new hyperbolic system of evolution equations, Phys. Rev. **D64**, 064017 (2001).
18. J. York, In C.R. Evans and L.S. Finn and D.W. Hobill, editors, *Frontiers in Numerical Relativity*, Cambridge University Press, Cambridge, (1989).
19. J. Thornburg, A fast apparent-horizon finder for 3-dimensional cartesian grids in numerical relativity, Class. Quant. Grav. **21**, 743 (2004).
20. M. Alcubierre, B. Brügmann, P. Diener, M. Koppitz, D. Pollney, E. Seidel, and R. Takahashi, Gauge conditions for long-term numerical black hole evolutions without excision, Phys. Rev. **D67**, 084023 (2003).
21. G.B. Cook and BBH Alliance, Boosted three-dimensional black hole evolutions with singularity excision, Phys. Rev. Lett. **80**, 2512 (1998).
22. S. Brandt, R. Correll, R. Gómez, M. Huq, P. Laguna L. Lehner, D. Neilsen, R. Matzner, J. Pullin, E. Schnetter, D. Shoemaker, and J. Winicour, Grazing collisions of black holes via the excision of singularities, Phys. Rev. Lett. **85**, 5496 (2000).
23. R. Gómez, L. Lehner, R. Marsa, and J. Winicour, Moving black holes in 3d, Phys. Rev. **D57**, 4778 (1998).
24. E. Seidel and W. Suen, Phys. Rev. Lett. **69**, 1845 (1992).
25. P. Anninos, G. Gaues, J. Masso, E. Seidel, and L. Smarr, Horizon boundary condition for black hole spacetimes, Phys. Rev. **D51**, 5562 (1995).
26. R. Marsa and M. Choptuik, Black hole–scalar field interactions in spherical symmetry, Phys. Rev. **D54**, 4929 (1996).
27. M. Scheel, T. Baumgarte, G. Cook, S. Shapiro, and S. Teukolsky, Numerical evolution of black holes with a hyperbolic formulation of general relativity, Phys. Rev. **D56**, 6320 (1997).
28. M. Alcubierre and B. Brügmann, Simple excision of a black hole in 3+1 numerical relativity, Phys. Rev. **D63**, 104006 (2001).
29. H. Yo, T. Baumgarte, and S. Shapiro, A numerical testbed for singularity excision in moving black hole spacetimes, Phys. Rev. **D64**, 124011 (2001).
30. D. Shoemaker, K. Smith, U. Sperhake, P. Laguna, E. Schnetter, and D. Fiske, Class. Quant. Grav. **20**, 3729 (2003).

31. U. Sperhake, K. Smith, B. Kelly, P. Laguna, and D. Shoemaker, Phys. Rev. **D69**, 024012 (2004).
32. G. Calabrese, L. Lehner, D. Neilsen, J. Pullin, O. Reula, O. Sarbach, and M. Tiglio, Novel finite-differencing techniques for numerical relativity: application to black hole excision, Class. Quant. Grav. **20**, L245 (2003).
33. A. Ashtekar, C. Beetle, and S. Fairhurst, Mechanics of isolated horizons, Class. Quant. Grav. **17**, 253 (2000).
34. A. Ashtekar, C. Beetle, O. Dreyer, S. Fairhurst, B. Krishnan, J. Lewandowski, and J. Wisniewski, Generic isolated horizons and their applications, Phys. Rev. Lett. **85**, 3564 (2000).
35. A. Ashtekar and B. Krishnan, Dynamical horizons: Energy, angular momentum, fluxes and balance laws, Phys. Rev. Lett. **89**, 261101 (2002).
36. A. Ashtekar and B. Krishnan, Dynamical horizons and their properties, Phys. Rev. **D68**, 104030 (2003).
37. O. Dreyer, B. Krishnan, E. Schnetter, and D. Shoemaker, Introduction to isolated horizons in numerical relativity, Phys. Rev. **D67**, 024018 (2003).
38. A. Ashtekar, S. Fairhurst, and B. Krishnan, Isolated horizons: Hamiltonian evolution and the first law, Phys. Rev. **D62**, 104025 (2000).
39. A. Ashtekar, C. Beetle, and J. Lewandowski, Mechanics of rotating isolated horizons, Phys. Rev. **D64**, 044016 (2001).
40. S. Brandt and B. Brügmann, A simple construction of initial data for multiple black holes, Phys. Rev. Lett. **78**, 3606 (1997).
41. C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation.* Freeman: San Francisco, (1973).
42. R.A. Matzner, M.F. Huq, and D. Shoemaker, Initial data and coordinates for multiple black hole systems, Phys. Rev. **D59**, 024015 (1999).
43. Harald P. Pfeiffer and Gregory B. Cook and Saul A. Teukolsky, Comparing initial-data sets for binary black holes, Phys. Rev. **D66**, 024047 (2002).
44. P. Anninos, K. Camarda J. Massó, E. Seidel, W. Suen, and J. Towns, Three dimensional numerical relativity: the evolution of black holes, Phys. Rev. **D52**, 2059 (1995).
45. R. Gómez *et al,* Stable characteristic evolution of generic three-dimensional single black hole space-times, Phys. Rev. Lett. **80**, 3915 (1998).
46. M. Alcubierre, B. Brügmann, D. Pollney, E. Seidel, and R. Takahashi, Black hole excision for dynamic black holes, Phys. Rev. **D64**, 061501 2001.
47. H.-J. Yo, T. W. Baumgarte, and S. L. Shapiro, Phys. Rev. **D66**, 084026 (2002).
48. S. Brandt, K. Camarda, E. Seidel, and R. Takahashi, Three dimensional distorted black holes, Class. Quant. Grav. **20**, 1 (2003).
49. R. Gómez, Gravitational waveforms with controlled accuracy, Phys. Rev. **D64**, 024007 (2001).
50. J. Baker, S. Brandt, M. Campanelli, C. Lousto, E. Seidel, and R. Takahashi, Nonlinear and perturbative evolution of distorted black holes. ii. odd-parity modes, Phys. Rev. **D62**, 127701 (2000).
51. G. Allen, K. Camarda, and E. Seidel, Black hole spectroscopy: Determining waveforms from 3d excited black holes, gr-qc/9806036 (1998).
52. D.R. Briland and R.W. Lindquist, Phys. Rev. **131**, 471 (1963).
53. P. Papadopoulos, Nonlinear harmonic generation in finite amplitude black hole oscillations, Phys. Rev. **D65**, 084016 2002.
54. Y. Zlochower, R Gómez, S. Husa, L. Lehner, and J. Winicour, Mode coupling in the nonlinear response of black holes, Phys. Rev. **D68**, 084014 (2003).

55. S. Hahn and R. Lindquist, The two body problem in geometrodynamics, Annals of Physics **29**, 304 (1964).
56. L. L. Smarr, Gauge conditions, radiation formulae and the two black hole collisions, In L. L. Smarr, editor, *Sources of gravitational radiation*, page 275. Cambridge University Press, Cambridge, (1979).
57. B. Brügmann, Binary black hole mergers in 3d numerical relativity, Int. J. Mod. Phys. **D8**, 85 (1999).
58. P. Marronetti, M. Huq, P. Laguna, L. Lehner, R. Matzner, and D. Shoemaker, Approximate analytical solutions to the initial data problem of black hole binary systems, Phys. Rev. **D62**, 024017 (2000).
59. M. Alcubierre, W. Benger, B. Brügmann, G. Lanfermann, L. Nerger, E. Seidel, and R. Takahashi, The 3d grazing collision of two black holes, Phys. Rev. Lett. **87**, 271103 (2001).
60. J. Baker, M. Campanelli, C.O. Lousto, and R. Takahashi, The lazarus project: A pragmatic approach to binary black hole evolutions, Phys. Rev. D **65**, 124012 (2002).
61. J. Baker, B. Brügmann, M. Campanelli, and C.O. Lousto, Gravitational waves from black hole collisions via an eclectic approach, Class. Quant. Grav. **17**, L149 (2000).
62. J. Baker, , B. Brügmann, M. Campanelli, C.O. Lousto, and R. Takahashi, Plunge waveforms from inspiralling binary black holes, Phys. Rev. Lett. **87**, 121103 (2001).
63. B. Brügmann, W. Tichy, and N. Jansen, Numerical simulation of orbiting black holes, gr-qc/0312112 (2003).
64. E. Schnetter, S. Hawley, and I. Hawke, Evolutions in 3d numerical relativity using fixed mesh refinement, Class. Quant. Grav **21**, 1465 (2004).
65. B. Imbiriba, J. Baker, D. Choi, J. Centrella, D. Fiske, J. Brown, J. van Meter, and K. Olson, Evolving a puncture black hole with fixed mesh refinement, gr-qc/0403048 (2004).
66. L. Kidder, M. Scheel, S. Teukolsky, E. Carlson, and G. Cook, Black hole evolution by spectral methods, Phys. Rev. **D62**, 084032 (2000).
67. S. Bonazzola, E. Gourgoulhon, and J. Marck, Spectral methods in general relativistic astrophysics, J. Comput. Appl. Math. **109**, 892 (1999).
68. P. Grandclément, S. Bonazzola, E. Gourgoulhon, and J. Marck, A multi-domain spectral method for scalar and vectorial poisson equations with non-compact sources, J. Comput. Phys. **170**, 231 (2001).

Index