THEORETICAL ASTROPHYSICS

Volume III: Galaxies and Cosmology

T. PADMANABHAN

Inter-University Centre for Astronomy and Astrophysics
Pune, India

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

© Cambridge University Press 2002

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2002

Printed in the United States of America

Typeface Times Roman 11/13 pt. *System* LaTeX 2ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

ISBN 0 521 56242 2 hardback
ISBN 0 521 56630 4 paperback

Contents

Preface xiii

1 Overview: Galaxies and Cosmology

1.1 Introduction 1

1.2 Evolution of the Universe 3

1.3 Formation of Dark-Matter Halos 8

1.4 Galaxy Formation 12

1.5 Morphological Classification of Galaxies 18

1.6 The Evolution of Galaxies 26

1.7 Properties of Disk Galaxies 34

1.8 Properties of Elliptical Galaxies 41

1.9 Milky Way Galaxy 46

- 1.9.1 Components of the Milky Way 47
- 1.9.2 Metallicity 50
- 1.9.3 Kinematics 51
- 1.9.4 The Galactic Centre 54

1.10 Features of Active Galactic Nuclei 55

- 1.10.1 Compact Sizes, Variability, and Continuum Emission 56
- 1.10.2 Radio Emission and Jets 62
- 1.10.3 Emission Lines 63
- 1.10.4 Absorption Systems 64

1.11 Taxonomy of Active Galactic Nuclei 65

- 1.11.1 Radio Galaxies 65
- 1.11.2 Quasars 66
- 1.11.3 BL Lac Objects 67
- 1.11.4 Seyfert Type I Galaxies 67
- 1.11.5 Seyfert Type II Galaxies 67
- 1.11.6 Low-Ionisation Nuclear-Emission Regions 67

1.12 Luminosity Function of Galaxies and Quasars 70
  1.12.1 Galaxy Counts and Luminosity Function 71
  1.12.2 Quasar Counts and Luminosity Function 76
1.13 Distribution of Matter 79
1.14 Extragalactic Background Radiation 82

2 Galactic Structure and Dynamics

2.1 Introduction 85
2.2 Models for Galaxies in Steady State 86
  2.2.1 Polytropes 88
  2.2.2 Isothermal Spheres 89
  2.2.3 King Model 90
  2.2.4 Axisymmetric Systems 92
2.3 Aspects of Stellar Orbits 94
  2.3.1 Spherically Symmetric Potentials 94
  2.3.2 Rotation Curves of Disk Galaxies 96
  2.3.3 Epicyclic Approximation in Axisymmetric Potentials 101
  2.3.4 Planar Nonaxisymmetric Potentials 104
  2.3.5 Potentials in the Rotating Frame 108
2.4 Application of the Jeans Equations 112
  2.4.1 Asymmetric Drift 113
  2.4.2 Mass and Velocity Dispersion 115
  2.4.3 Rotation of Elliptical Galaxies 117
2.5 Stellar Dynamics at Galactic Cores 121
2.6 Spiral Structure 125
2.7 Warps 137
2.8 Chemical Evolution of Galaxies 139
2.9 Galaxy Interactions and Mergers 149
  2.9.1 Galactic Cannibalism 151
  2.9.2 Galaxy Collisions 154
  2.9.3 Numerical Simulations 156

3 Friedmann Model of the Universe

3.1 Introduction 161
3.2 The Friedmann Model 161
3.3 Kinematics of the Friedmann Model 167
3.4 Dynamics of the Friedmann Model 176
3.5 Observational Tools in Friedmann Models 187
3.6 Gravitational Lensing 196
  3.6.1 Constant Surface Density 202
  3.6.2 Point Mass 203
  3.6.3 Isothermal Sphere 206

4 **Thermal History of the Universe**

4.1 Introduction 210
4.2 Distribution Functions in the Early Universe 210
4.3 Relic Background of Relativistic Particles 218
4.4 Relic Background of Wimps 226
4.5 Synthesis of Light Nuclei 230
4.6 A Simplified Model for Primordial Nucleosynthesis 243
4.7 Decoupling of Matter and Radiation 248
4.8 Very Early Universe and Cosmological Scalar Fields 263

5 **Structure Formation**

5.1 Introduction 272
5.2 Growth of Inhomogeneities 273
5.3 Linear Growth in the General Relativistic Regime 276
5.4 Gauge Dependence of Perturbations: An Illustration 279
  5.4.1 Synchronous Gauge 281
  5.4.2 Poisson Gauge 282
5.5 Gravitational Clustering in the Newtonian Limit 289
5.6 Linear Perturbations in the Newtonian Limit 292
5.7 Origin of Density Perturbations 304
5.8 Transfer Functions and Statistical Indicators 315
5.9 Zeldovich Approximation 326
5.10 Spherical Approximation 329
5.11 Scaling Laws 335
5.12 Nonlinear Scaling Relations 338

6 **Cosmic Microwave Background Radiation**

6.1 Introduction 349
6.2 Processes Leading to Distortions in CMBR 349
6.3 Angular Pattern of CMBR Anisotropies 352
6.4 CMBR Anisotropies: Simplified Derivation 360
6.5 CMBR Anisotropies: A More Rigorous Derivation 366
6.6 Comparison with Observations 379
  6.6.1 Dipolar Anisotropy 379
  6.6.2 Anisotropies at Large Angular Scales 382
  6.6.3 Anisotropies at Small Angular Scales 385
6.7 Spectral Distortions of CMBR 388
  6.7.1 Distortions Due to Global-Energy Injection 388
  6.7.2 Sunyaev–Zeldovich Effect 395

7 **Formation of Baryonic Structures**

7.1 Introduction 397
7.2 Linear Perturbations in Baryons 398

7.3 Nonlinear Collapse of Baryons 404
7.4 Mass Functions and Abundances 415
7.5 Angular Momentum of Galaxies 424
7.6 Galaxy Formation and Evolution 427
7.7 Galaxy Distributions in Projection 439
7.8 Magnetic Fields in the Universe 443

8 Active Galactic Nuclei

8.1 Introduction 447
8.2 The Black Hole Paradigm 447
8.3 Optical and UV Continua from AGN 451
8.4 High-Energy Spectra: X Rays and Gamma Rays 462
  8.4.1 Comptonisation 462
  8.4.2 Pair Production 468
  8.4.3 Line Emission from Iron 473
8.5 Radio Emission from Quasars 475
8.6 Radio Jets 481
8.7 Effects of Bulk Relativistic Motion 490
8.8 The Broad-Line and Narrow-Line Regions 495
  8.8.1 Broad-Line Regions 495
  8.8.2 Narrow-Line Regions 505
8.9 Intrinsic Absorbers in AGN 508
8.10 Quasar Luminosity Function 510

9 Intergalactic Medium and Absorption Systems

9.1 Introduction 518
9.2 Gunn–Peterson Effect 518
9.3 Ionisation of the IGM 523
  9.3.1 Photoionisation Equilibrium of the IGM 523
  9.3.2 Photoionisation of the IGM by Discrete Sources 527
  9.3.3 Collisional Ionisation 532
9.4 Background Radiation from High-Redshift Sources 534
9.5 Lyman-α Absorption by a Diffuse IGM 541
  9.5.1 Classification of Lyman-α Absorption Lines 541
  9.5.2 Lyman-α Forest and a Diffuse IGM 544
9.6 Damped Lyman-α Clouds 547

10 Cosmological Observations

10.1 Introduction 552
10.2 Cosmic Distance Scale 552
  10.2.1 Examples of Direct Distance Estimates 553
  10.2.2 Development of a Cosmic Distance Ladder 555
10.3 Age of the Universe 562

10.4 Observational Evidence for Dark Matter 564
  10.4.1 Solar Neighbourhood 566
  10.4.2 Rotation Curves of Other Disk Galaxies 571
  10.4.3 Cores of Spiral Galaxies and Dwarf Spheroidals 571
  10.4.4 Dark-Matter Estimates from the Dynamics of the Local Group 573
  10.4.5 Groups of Galaxies 575
  10.4.6 Clusters of Galaxies 577
  10.4.7 Virgo-Centric Flow and Velocity Fields 579
10.5 Nature of Dark Matter 581
  10.5.1 Baryonic Dark Matter 582
  10.5.2 Nonbaryonic Dark Matter 586
10.6 Axions 592
10.7 Cosmological Constant 595

Notes and References 599

Index 609

1

Overview: Galaxies and Cosmology

1.1 Introduction

Attempts to understand extragalactic objects and the universe by using the laws of physics lead to difficulties that have no parallel in the application of the laws of physics to systems of a more moderate scale. The key difficulty arises from the fact that our universe exhibits temporal evolution and is not in steady state. Thus different epochs in the past evolutionary history of the universe are unique (and have occurred only once), and the current state of the universe is a direct consequence of the conditions that were prevalent in the past. For example, most of the galaxies in the universe have formed sometime in the past during a particular phase in the evolution of the universe. This is in contrast to star formation within a galaxy that we can observe directly and study by using standard statistical methods.

In principle, we should be able to see the events that took place in the universe in the past because of the finite light travel time. By observing sufficiently faraway regions of the universe, we will be able to observe the universe as it was in the past. Although technological innovation will eventually allow us to directly observe and understand all the past events in the history of the universe (especially when neutrino astronomy and gravitational wave astronomy start complementing photon-based observations), we are far from such a satisfactory state of affairs at present. Direct observational evidence today spans only a tiny fraction in the past history of the universe and is not available for sufficiently early epochs. Hence the straightforward approach of starting with known initial conditions for the laws of physics (expressed as a differential equation, say) and integrating them forward in time cannot be adopted to the study of the universe.

An alternative procedure is to start with the current state of the universe and integrate the same equations backwards in time in order to understand its past history. Even in this attempt, progress is not easy because data available at the present epoch are insufficient. The primary problem is what was stressed in Vol. II, Chap. 1: Observational data of adequate quality and quantity become scarce as we probe larger and larger scales. Further, we have no direct laboratory evidence regarding nearly 90% of the matter that is present in the universe; there is also some indirect evidence to suggest that nearly 60% of the matter present in the universe today obeys a fairly exotic equation of state.

These difficulties – which are unique when we attempt to apply the laws of physics to an evolving universe – require us to proceed in a multifaceted manner. Our approach will be to develop a broad paradigm describing the evolution of the universe and the formation of structures in it and iterate the details by constantly comparing the theoretical predictions with observational data. This paradigm is based on the idea that the universe was reasonably homogeneous, isotropic, and fairly featureless – except for small fluctuations in the energy density – at sufficiently early times. It is then possible to set up the equations that describe a model for the universe and integrate them forward in time. The results will depend on the composition of the universe, its current expansion rate, and the initial spectrum of density perturbations. Varying these parameters allows us to construct a library of evolutionary models for the universe that can then be compared with observations in order to restrict the parameter space.

Our approach in many of the chapters in this volume are based on the preceding paradigm of *parameterised cosmology*. The aim will be to deduce as many features of the observed universe as possible from a small set of parameters. Such an approach has proved to be extremely successful in the past two decades, mainly because of the advances in technology that allow good-quality observations. Some of the observations planned during the next two decades hold the hope of determining fairly accurately the parameters that characterize the universe, thereby reducing the problem to one of integration of the relevant equations.

It is possible to consider the study of extragalactic astronomy and cosmology from a broader perspective and ask why the parameters describing the universe have the values that are attributed to them. In other words, why does the observed universe follow one template out of a class of models that can be constructed based on the known laws of physics? Such a question – although intuitively appealing – has no mathematically rigorous and unique formulation and hence will be ignored in our discussion.

A completely different issue will be whether the laws of physics can be used to reduce the number of independent parameters and assumptions in any cosmological model. This is certainly possible once our knowledge of high-energy interactions of particles gets better. At present direct laboratory evidence exists for particle interactions only at energies less than about 100 GeV, and particle-physics models describing higher energies do not have the level of certainty required for making definite *predictions* about the evolution of the universe. Eventually, when our understanding of high-energy particle physics improves to an adequate level, it can be applied to the early phases of the universe. We stress the fact that the procedure of applying laws of physics to understand the behaviour of the universe is hindered *only* because we are ignorant about the relevant laws of physics at sufficiently high energies.[1] (The superscripted numbers throughout the book refer to items in the Notes and References chapter at the end of the book.)

1.2 Evolution of the Universe

Observations suggest that the universe at large scales is homogeneous and isotropic. The fractional fluctuations $(\delta\rho/\rho)_R$ in the mass (and energy) density ρ (which is due to the existence of structures like galaxies, clusters etc.), within a randomly placed sphere of radius R, decrease with R as a power law. This suggests that we can model the universe as being made up of a smooth background with an average density, superposed with fluctuations in the density that are large at small scales but decrease with scale. At sufficiently large scales, the universe may be treated as being homogeneous and isotropic with a uniform density.

It was shown in Vol. 1, Chap. 1, that the only large-scale motion compatible with homogeneity and isotropy is the one with the velocity field of the form $\dot{\mathbf{r}}(t) = \mathbf{v}(t) = f(t)\mathbf{r}$. This allows us to describe the position $\mathbf{r}$ of any material body in the universe in the form $\mathbf{r} = a(t)\mathbf{x}$, where $a(t)$ is another arbitrary function related to $f(t)$ by $f(t) = (\dot{a}/a)$ and $\mathbf{x}$ is a constant for any given material body in the universe. It is conventional to call $\mathbf{x}$ and $\mathbf{r}$ the *comoving* and the *proper* coordinates of the body and $a(t)$ the *expansion factor*. (Even though, if $\dot{a} < 0$, it acts as a contraction factor.)

The dynamics of the universe is entirely determined by the function $a(t)$. The simplest choice will be $a(t) =$ constant, in which case there will be no motion in the universe and all matter will be distributed uniformly in a static configuration. It is, however, clear that such a configuration will be violently unstable when the mutual gravitational forces of the bodies are taken into account. Any such instability will eventually lead to the random motion of particles in localized regions, thereby destroying the initial homogeneity. Observations, however, indicate that this is not true and that the relation $\mathbf{v} = (\dot{a}/a)\mathbf{r}$ does hold in the observed universe with $\dot{a} > 0$. In that case, the dynamics of $a(t)$ can be qualitatively understood along the following lines. Consider a particle of *unit* mass at the location r with respect to some coordinate system. Equating the sum of its kinetic energy $v^2/2$ and gravitational potential energy $[-GM(r)/r]$ that is due to the attraction of matter inside a sphere of radius r, to a constant, we find that $a(t)$ should satisfy the condition

$$\frac{1}{2}\dot{a}^2 - \frac{4\pi G\rho(t)}{3}a^2 = \text{constant}, \tag{1.1}$$

where ρ is the mean density of the universe; that is,

$$\frac{\dot{a}^2}{a^2} + \frac{k}{a^2} = \frac{8\pi G}{3}\rho(t), \tag{1.2}$$

where k is a constant. Although the preceding argument to determine this equation is fallacious, Eq. (1.2) happens to be exact and arises from the proper application of Einstein's theory of relativity to a homogeneous and isotropic distribution of matter with ρ interpreted as the energy density. We shall now describe some simple aspects of such an evolution that will be taken up for detailed study in the later chapters.

Observations suggest that our universe today (at $t = t_0$) is governed by Eq. (1.2) with $(\dot{a}/a)_0 \equiv H_0 = 0.3 \times 10^{-17} h \text{ s}^{-1}$, where $h \approx (0.5\text{–}1)$. This is equivalent to $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$, where $1 \text{ Mpc} \approx 3 \times 10^{24}$ cm is a convenient unit for cosmological distances. (We will also use the units $1 \text{ kpc} = 10^{-3}$ Mpc and $1 \text{ pc} = 10^{-6}$ Mpc in our discussion.) From H_0 we can form the time scale $t_{\text{univ}} \equiv H_0^{-1} \approx 10^{10} h^{-1}$ yr and the length scale $cH_0^{-1} \approx 3000h^{-1}$ Mpc; t_{univ} characterizes the evolutionary time scale of the universe and cH_0^{-1} is of the order of the largest length scales currently accessible in cosmological observations. The relation $\mathbf{v} = f(t)\mathbf{r} = (\dot{a}/a)\mathbf{r} = H_0\mathbf{r}$ is called *Hubble's law*, and H_0 is called *Hubble's constant*. From H_0 we can also construct a quantity with the dimensions of density, called the *critical density*:

$$\begin{aligned} \rho_c = \frac{3H_0^2}{8\pi G} &= 1.88h^2 \times 10^{-29} \text{ gm cm}^{-3} = 2.8 \times 10^{11} h^2 \, M_\odot \text{ Mpc}^{-3} \\ &= 1.1 \times 10^4 h^2 \text{ eV cm}^{-3} = 1.1 \times 10^{-5} h^2 \text{ protons cm}^{-3}. \end{aligned} \tag{1.3}$$

(The last two "equalities" should be interpreted in terms of conversion of mass into energy by a factor c^2 and the conversion of mass into number of baryons by a factor m_p^{-1}, where m_p is the proton mass.) It is conventional to measure all other mass and energy densities in the universe in terms of the critical density. If ρ_i is the mass or the energy density associated with a particular species, then we define a density parameter Ω_i through the ratio $\Omega_i \equiv (\rho_i/\rho_c)$. In general, both ρ_i and ρ_c can be defined at any given epoch in the universe and not necessarily at the present moment $t = t_0$; by convention, ρ_c is always defined in terms of the present value of the Hubble constant, although ρ_i could, in general, be a function of time: $\rho_i = \rho_i(t)$. In this case, Ω_i will also depend on time and we define $\Omega_i(t) \equiv \rho_i(t)/\rho_c$.

It is obvious from Eq. (1.2) that the numerical value of k can be absorbed into the definition of $a(t)$ by rescaling it so that we can treat k as having one of the three values $(0, -1, +1)$. The choice among these three values for k is decided by Eq. (1.2) depending on the value of Ω; we see that $k = 1, 0$ or -1, depending on whether Ω is greater than, equal to, or less than unity. The fact that k is proportional to the total energy of the dynamical system described by Eq. (1.2) shows that $a(t)$ will have a maximum value followed by a contracting phase to the universe if $k = 1$ and $\Omega > 1$.

To determine the nature of the cosmological model we need to determine the value of Ω for the universe, taking into account all forms of energy densities that exist at present. Further, to determine the form of $a(t)$ we need to determine how the energy density of any given species varies with time. We now briefly describe the issues involved in this task.

If a particular kind of energy density is described by an equation of state of the form $p = w\rho$, where p is the pressure and w is a constant, then the equation for energy conservation in an expanding background, $d(\rho a^3) = -pd(a^3)$, can be integrated to give $\rho \propto a^{-3(1+w)}$. Equation (1.2) can be now written in the form

$$\frac{\dot{a}^2}{a^2} = H_0^2 \sum_i \Omega_i \left(\frac{a_0}{a}\right)^{3(1+w_i)} - \frac{k}{a^2}, \tag{1.4}$$

where each of these species is identified by density parameter Ω_i and the equation of state is characterized by w_i. The most familiar forms of energy densities are those due to pressureless matter with $w_i = 0$ (that is, nonrelativistic matter with rest-mass-energy density ρc^2 dominating over the kinetic-energy density, $\rho v^2/2$) and radiation with $w_i = (1/3)$. The density parameter contributed today by visible, nonrelativistic, baryonic matter in the universe is $\Omega_B \approx (0.01\text{–}0.2)$ and the density parameter that is due to radiation is $\Omega_R \approx 2 \times 10^{-5}$. Unfortunately, models for the universe with just these two constituents for the energy density are in violent disagreement with observations. As we shall see in later chapters, it appears to be necessary to postulate (1) the existence of pressureless ($w = 0$) *non*baryonic dark matter that does not couple with radiation and has a density of at least $\Omega_{\rm DM} \approx 0.3$; because it does not emit light, it is called *dark matter*; (2) an exotic form of matter (called either *cosmological constant* or *vacuum-energy density*) with an equation of state $p = -\rho$ (that is, $w = -1$) that has a density parameter of $\Omega_V \approx 0.7$. The evidence for the existence of nonbaryonic dark matter seems to be fairly definitive whereas the evidence for the existence of cosmological constant is somewhat less definitive. Keeping this in mind, we will concentrate on two typical cosmological models throughout this volume. The first one will have $\Omega_V = 0$ and $0 \leq \Omega_{\rm DM} \leq 1$; the second one will have $\Omega_V + \Omega_{\rm DM} = 1$.

Figure 1.1 provides an inventory of the density contributed by different forms of matter in the universe, and these entries will be discussed in different sections of this chapter. The x axis is actually a combination of Ω and the Hubble parameter h because different components are measured by different techniques. (Usually $n = 1$ or 2; numerical values are for $h = 0.7$.) The top two positions in the contribution to Ω are from a cosmological constant and nonbaryonic dark matter. It is unfortunate that we do not have laboratory evidence for the existence of the first two dominant contributions to the energy density in the universe. This feature alone could make most of the cosmological paradigm described in this book irrelevant at a future date. Alternatively, laboratory detection of a nonbaryonic dark-matter candidate will be an important discovery in establishing the standard paradigm of structure formation.

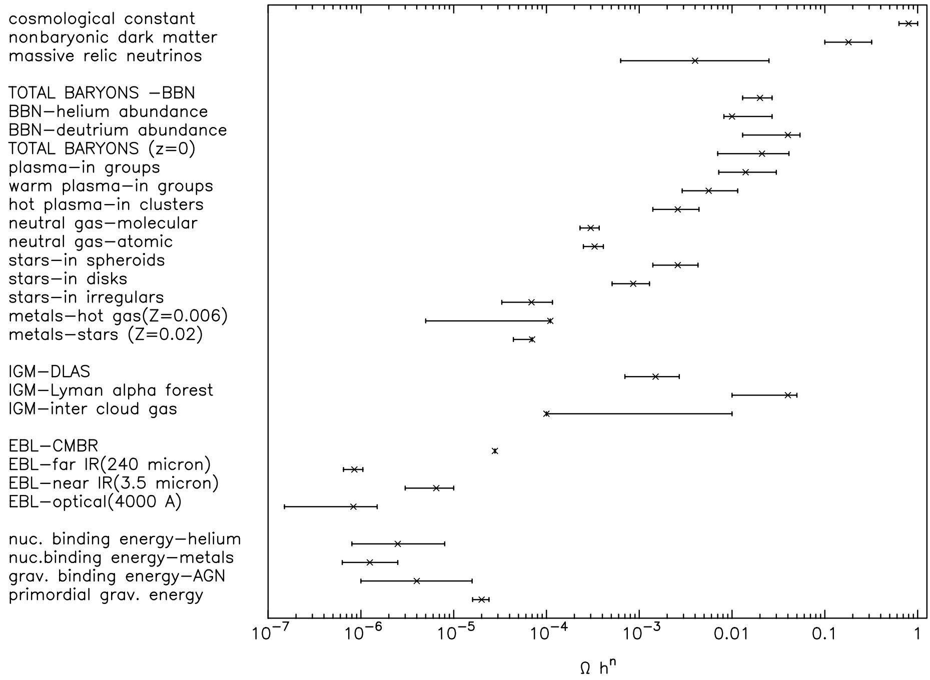

Fig. 1.1. Cosmic inventory of energy densities. See text for description.

Exercise 1.1

Determining the matter content: Let us assume that the universe contains material with several different equations of state, each characterized by a constant value $w = p/\rho$. Introduce the parameter $\alpha \equiv 3(1 + w)$ and the function $\Omega(\alpha)$ that describes the amount of energy density contributed by a species with a given value of α. Explain how the knowledge of the function $a(t)$ can be used to determine $\Omega(\alpha)$. [Answer: We first note that the term (k/a^2) can be thought of as contributed by a hypothetical species of matter with $w = -(1/3)$. Hence Eq. (1.4) can be written in the form

$$\frac{\dot{a}^2}{a^2} = H_0^2 \sum_i \Omega_i \left(\frac{a_0}{a}\right)^{3(1+w_i)}, \tag{1.5}$$

with a term having $w_i = -(1/3)$ added to the sum. In the continuum limit, this equation can be rewritten as

$$\left(\frac{dq}{d\tau}\right)^2 = \int_{-\infty}^{\infty} d\alpha\ \Omega(\alpha) e^{-\alpha q}, \tag{1.6}$$

where $(a/a_0) = \exp(q)$ and $\tau = H_0 t$. The function $\Omega(\alpha)$ is assumed to have finite support or to decrease fast enough for the expression on the right-hand side to converge. If the observations determine the function $a(t)$, then the left-hand side can be expressed as a function of q. An inverse Laplace transform of this equation will then determine the form of $\Omega(\alpha)$, thereby determining the composition of the universe, as long as all matter can be described by an equation of state of the form $p = w\rho$.]

The evolution of the universe, with the energy content as described above, is straightforward to determine and we shall illustrate it for a simple model with $\Omega_{\mathrm{DM}} + \Omega_B + \Omega_R \approx 1; \Omega_V = 0$. If neither particles nor photons are created or destroyed during the expansion, then the number density of particles or photons will decrease as a^{-3} as a increases. In the case of photons, the wavelength will also get stretched during expansion with $\lambda \propto a$; because the energy density of material particles is nmc^2 whereas that of photons of frequency ν is $nh\nu = (nhc/\lambda)$, it follows that the energy densities of radiation and matter vary as $\rho_{\mathrm{rad}} \propto a^{-4}$ and $\rho_{\mathrm{matter}} \propto a^{-3}$. Combining with the result $\rho_{\mathrm{rad}} \propto T^4$ for thermal radiation, it follows that any thermal spectrum of photons in the universe will have its temperature varying as $T \propto a^{-1}$. In the past, when the universe was smaller, it would also have been (1) denser, (2) hotter, and – at sufficiently early epochs – (3) dominated by radiation-energy density.

The light emitted at an earlier epoch by an object will reach us today with the wavelengths stretched because of the expansion. If the light was emitted at $a = a_e$ and received today (when $a = a_0$), the wavelength will change by the factor $(1 + z_e) = (a_0/a_e)$, where z_e is called the *redshift*, which corresponds to the epoch of emission a_e. Because the observed luminosity L of a source is proportional to $(p_\gamma c)\, d^3 p_\gamma \propto \nu^3\, d\nu \propto (1+z)^{-4}$, where $p_\gamma = (\epsilon/c) = (h\nu/c)$ is the photon momentum, it will decrease as $(1+z)^{-4}$.

When the temperature of the universe is higher than the temperatures corresponding to the atomic ionisation energy, the matter content in the universe will be a high-temperature plasma. Further, when the temperature of the universe is higher than the binding energy of the nuclei ($\sim$MeV), none of the heavy elements (helium and the metals) could have existed in the universe. Starting from such a hot initial plasma stage, the universe cools as it expands and nucleosynthesis of some amount of deuterium, helium and lithium takes place when $k_B T \lesssim$ MeV. This process does not proceed to form any other heavier elements in significant quantities. This is because – for the observed range of matter and radiation-energy densities – the universe expands too fast to allow the synthesis of heavier metals. The primordial abundance of helium and deuterium is therefore a sensitive test of the different parameters of the universe and will be explored in detail in Chap. 4. The three terms in Fig. 1.1 marked BBN give the constraints arising from *big bang nucleosynthesis*.

In the early hot phase, the radiation will be in thermal equilibrium with matter; as the universe cools below $k_B T \simeq (\epsilon_a/10)$, where ϵ_a is the binding energy of atoms, the electrons and ions will combine to form neutral atoms and radiation will decouple from matter. This occurs at $T_{\mathrm{dec}} \simeq 3 \times 10^3$ K. As the universe expands further, these photons will exist in the form of thermal background radiation with a temperature that scales as $T \propto (1/a)$. It turns out that the major component of the extragalactic background light (EBL) that exists today is in the microwave band and can be fitted very accurately by a thermal spectrum at a temperature of ~2.7 K. It seems reasonable to interpret this radiation as a relic arising from the early hot phase of the evolving universe. The intensity per logarithmic band of frequency, νB_ν, for this radiation peaks at a wavelength of 1 mm and the maximum intensity is 5.3×10^{-7} W m^{-2} rad^{-2} over the entire sky. The intensity per square arcsecond of the sky is approximately 1.33×10^{-17} W m^{-2} arcsec^{-2}. The energy density that is due to this radiation today will be $\rho_\gamma \simeq (k_B T)^4/(\hbar c)^3 \simeq 5.7 \times 10^{-13}$ ergs cm^{-3}, which corresponds to a *mass* density of $(\rho_\gamma/c^2) = 5.7 \times 10^{-34}$ gm cm^{-3} (this is marked as the entry EBL-CMBR in Fig. 1.1; CMBR stands for *cosmic microwave background radiation*). Taking the matter density today as $\rho_0 = 10^{-30}$ gm cm^{-3}, we find that $\rho_\gamma \simeq 5.7 \times 10^{-4}\rho_0$; radiation (with $\rho_\gamma \propto a^{-4}$) would have dominated over matter (with $\rho \propto a^{-3}$) when the redshift was larger than $z_{\rm eq} \equiv (\rho/\rho_\gamma) \approx 1.7 \times 10^3$.

1.3 Formation of Dark-Matter Halos

The considerations of the last section were independent of the explicit form of $a(t)$. We now turn to the solutions of Eq. (1.2) that determine $a(t)$ and the issue of the formation of structures. The simplest solution to Eq. (1.2) will occur for $k = 0$ if we take the matter density in the universe to decrease as a^{-3} with expansion. Then we get $a(t) = (t/t_0)^{2/3}$ with $t_0^{-2} = (6\pi G\rho_0)$, and $a(t)$ is normalised to $a = 1$ at the present epoch $t = t_0$.

Such a totally uniform universe, of course, will never lead to any of the inhomogeneous structures seen today. However, if the universe has even the slightest inhomogeneity in the past, then gravitational instability can amplify the density perturbations. To see how this comes about in the simplest context, consider Eq. (1.2) written in the equivalent form as

$$\ddot{a} = -\frac{4\pi G\rho_0}{3a^2} = -\left(\frac{2}{9t_0^2}\right)\frac{1}{a^2}, \tag{1.7}$$

where we have put $\rho = (\rho_0 a_0^3/a^3)$ and differentiated Eq. (1.2) once with respect to t. If we perturb $a(t)$ slightly to $a(t) + \delta a(t)$ such that the corresponding fractional density perturbation is $\delta \equiv (\delta\rho/\rho) = -3(\delta a/a)$, we find that δa satisfies the equation

$$\frac{d^2}{dt^2}\delta a = \left(\frac{4}{9t_0^2}\right)\frac{\delta a}{a^3} = \frac{4}{9}\frac{\delta a}{t^2}. \tag{1.8}$$

This equation has the growing solution $\delta a \propto t^{4/3} \propto a^2$. Hence the density perturbation $\delta = -3(\delta a/a)$ grows as $\delta \propto a$. When the perturbations have grown sufficiently, their self-gravity will start dominating and the matter can collapse to form a gravitationally bound system. The dark matter will form virialised, gravitationally bound structures with different masses and radii. The baryonic matter will cool by radiating energy, sink to the centres of the dark-matter halos, and form galaxies. We now discuss some of the features of such a case for structure formation, starting with the formation of dark-matter haloes. The formation of galaxies will be discussed in the next section.

To describe the growth of structures in the universe, it is convenient to use the spatial Fourier transform $\delta_{\mathbf{k}}(t)$ of the density contrast $\delta(t, \mathbf{x}) \equiv [\rho(t, \mathbf{x}) - \rho_{\mathrm{bg}}]/\rho_{\mathrm{bg}}$, where $\rho_{\mathrm{bg}}(t)$ is the smooth background density. We treat the density fluctuation $\delta_{\mathbf{k}}(t)$ as a realisation of a random processes. Then we can define the power spectrum of fluctuations at a given wave number k by $P(k, t) \equiv \langle |\delta_{\mathbf{k}}(t)|^2 \rangle$, where the averaging symbol denotes that we are treating $P(k, t)$ as a statistical quantity averaged over an ensemble of possibilities; statistical isotropy of the universe implies that the power spectrum can depend on only the magnitude $|\mathbf{k}|$ of the wave number. The power per logarithmic band in k is given by

$$\Delta_k^2(t) = \frac{k^3\, |\delta_k(t)|^2}{2\pi^2} = \frac{k^3 P(k, t)}{2\pi^2}. \tag{1.9}$$

For a smoothly varying power spectrum, this quantity is related to the mean-square fluctuation in density (or mass) at the scale $R \approx k^{-1}$ in the universe by

$$\Delta_k^2 = \left(\frac{\delta\rho}{\rho}\right)^2_{R \simeq k^{-1}} = \left(\frac{\delta M}{M}\right)^2_{R \simeq k^{-1}} \cong \sigma^2(R, t). \tag{1.10}$$

Since we can associate a mass scale $M = (4\pi/3)\rho_{\mathrm{bg}}(t_0)R^3$ with a length scale R, one can also treat σ^2 as a function of mass scale: $\sigma^2 = \sigma^2(M, t)$. We shall see in Chap. 5 that the power spectrum of fluctuations in the universe is fairly smooth and hence can be approximated by a power law in k locally at any given time so that $P(k) \propto k^n$. From the result derived above, $\delta \propto a$, it follows that

$$\Delta_k^2(t) \propto a^2 k^{n+3}, \qquad \sigma^2(R, t) \propto a^2 R^{-(n+3)} \tag{1.11}$$

as long as $\sigma \ll 1$, with n being a slowly varying function of scale k or R.

The pattern of density fluctuations is thus characterised by the power spectrum $P(k, t)$ at any given time. The gravitational potential that is due to a density perturbation $\delta\rho = \bar{\rho}\delta$ in a region of size R will be $\phi \propto (\delta M/R) \propto \bar{\rho}\delta R^2$. In an expanding universe $\bar{\rho} \propto a^{-3}$ and $R \propto a$, and the perturbation δ grows as $\delta \propto a$ [see the discussion following Eq. (1.8)], making ϕ constant in time. In particular, the fluctuations that existed in the universe at the time when radiation decoupled from matter would have left their imprint on the radiation field. Because photons climbing out of a potential well of size ϕ will lose energy and undergo a redshift $(\Delta\nu/\nu) \approx (\phi/c^2)$, we would expect to see a temperature anisotropy in the microwave radiation of the order of $(\Delta T/T) \approx (\Delta\nu/\nu) \approx (\phi/c^2)$. The largest potential wells would have left their imprint on the cosmic background radiation at the time of decoupling of radiation and matter. We shall see later that the galaxy clusters constitute the deepest gravitational potential wells in the universe from which the escape velocities are $v_{\text{clus}} \approx (GM/R)^{1/2} \approx 10^3\ \text{km s}^{-1}$. This will lead to a temperature anisotropy of $\Delta T/T \approx (v_{\text{clus}}/c)^2 = 10^{-5}$. Such a temperature perturbation has indeed been observed in the microwave background radiation, vindicating the case for structure formation.

The entry marked gravitational binding energy in Fig. 1.1 is essentially a measure of $(v/c)^2$ for the largest scales that are gravitationally bound. Equivalently, it can be thought of as the amount of power in the gravitational potential per logarithmic band in Fourier space. Its value can be determined from the temperature anisotropies in CMBR and will be discussed in Chap. 6.

When $\sigma(R, t) \to 1$, that particular scale characterized by R will go nonlinear and matter at that scale will collapse and form a bound structure. Because this occurs when the density contrast σ reaches some critical value $\sigma_c \approx 1$, it follows from relations (1.11) that the scale that goes nonlinear at any given time t in the past (corresponding to a redshift z) obeys the relation

$$R_{\text{NL}}(t) \propto a(t)^{2/(n+3)} = R_{\text{NL}}(t_0)(1+z)^{-2/(n+3)}. \tag{1.12}$$

Equivalently, structures with mass $M \propto R_{\text{NL}}^3$ will form at a redshift z where

$$M_{\text{NL}}(z) = M_{\text{NL}}(t_0)(1+z)^{-6/(n+3)}. \tag{1.13}$$

Such virialised, gravitationally bound structures – once formed – will remain frozen at a mean density $\bar{\rho}$, which is approximately $f_c \simeq 200$ times the background density of the universe at the redshift of formation, z (see Chap. 5). Taking the background density of the universe at redshift z to be $\rho_{\text{bg}}(z) = \rho_c \Omega (1+z)^3$, we find that the mean density $\bar{\rho}$ of an object that would have collapsed at redshift z is given by $\bar{\rho} \simeq \Omega \rho_c f_c (1+z)^3$. We define the *circular velocity* v_c for such a collapsed body as

$$v_c^2 \equiv \frac{GM}{r} \equiv \frac{4\pi G}{3}\bar{\rho} r^2. \tag{1.14}$$

If $\bar{\rho}$ is eliminated in terms of v_c, the redshift of formation of an object can be expressed in the form

$$(1+z) \cong 5.8 \left(\frac{200}{\Omega f_c}\right)^{1/3} \frac{(v_c/200\ \text{km s}^{-1})^{2/3}}{(r/h^{-1}\ \text{Mpc})^{2/3}}. \tag{1.15}$$

It is interesting that such a fairly elementary calculation leads to an acceptable result regarding the redshift for the formation of first structures. If we consider small-scale halos (approximately a few kiloparsecs), the formation redshift can go up to, say, 20. This calculation also introduces the notion of *hierarchical clustering* in which smaller scales go nonlinear and virialise earlier on and the merging of these smaller structures leads to hierarchically bigger and bigger structures. Of course, the process is supplemented by the larger scales going nonlinear by themselves but the importance of mergers cannot be ignored in galaxy-formation scenarios.

It is in fact possible to work out all the properties of collapsed structures from the preceding formalism. To do this, we begin by noting that $\sigma(M, t)$ denotes the *typical* value of density fluctuations that exist in the universe at a given time t. In a statistical description, we could think of regions having a density contrast of, say, ν times larger than the typical value $\sigma(M, t)$ occurring with a probability $\mathcal{P}(\nu)$ that could be approximated as a Gaussian of unit variance in most models. Such a $\nu\sigma$ fluctuation will collapse at a redshift z, which is determined by the condition $\nu\sigma(M, t) = \sigma_c = \mathcal{O}(1)$. Because $\sigma(M, t) \propto a(t)$, it follows that $\sigma(M, t) = \sigma_0(M)(1 + z)^{-1}$, where $\sigma_0(M)$ is the fiducial value of the density fluctuation today. This leads to the first result, namely the redshift $z_{\text{coll}}(M)$ at which any given mass scale collapses in terms of $\sigma_0(M)$:

$$(1 + z_{\text{coll}}) = \frac{\nu\sigma_0(M)}{\sigma_c}. \tag{1.16}$$

Next we use the fact that the density of structures collapsing at redshift z is f_c times larger than the background density at z; hence $\rho = \rho_c f_c(1 + z)^3$ in an $\Omega = 1$ universe. Given M and ρ, we can compute the radius by $R^3 = (3M/4\pi\rho)$, the circular velocity by $v_c = (GM/R)^{1/2}$, and the gravitational potential energy by $U = -(3/5)(GM^2/R)$. The only input needed to compute these quantities is $\sigma_0(M)$, and we obtain by straightforward algebra the following results:

$$\rho = 5.6 \times 10^{13} h^2 \left(\frac{f_c}{200}\right) \left(\frac{\nu}{\sigma_c}\right)^3 \sigma_0^3(M)\, M_\odot\ \text{Mpc}^{-3}, \tag{1.17}$$

$$R = 2.57 h^{-2/3} \left(\frac{M}{10^{11}\, M_\odot}\right)^{1/3} \left(\frac{f_c}{200}\right)^{-1/3} \left(\frac{\sigma_c}{\nu}\right) \sigma_0^{-1}(M)\ \text{Mpc}, \tag{1.18}$$

$$v_c = 75.5 h^{1/3} \left(\frac{M}{10^{11}\, M_\odot}\right)^{1/3} \left(\frac{f_c}{200}\right)^{1/6} \left(\frac{\sigma_c}{\nu}\right)^{-1/2} \sigma_0^{1/2}(M)\ \text{km s}^{-1}, \tag{1.19}$$

$$U = 6.72 \times 10^{57} h^{2/3} \left(\frac{M}{10^{11}\, M_\odot}\right)^{5/3} \left(\frac{f_c}{200}\right)^{1/3} \left(\frac{\sigma_c}{\nu}\right)^{-1} \sigma_0(M)\ \text{ergs}. \tag{1.20}$$

The free-fall time scale for a constant density sphere of radius R and mass M is $t_{\text{ff}} = (\pi/2\sqrt{2})(R^3/GM)^{1/2}$ (see Vol. II, Chap. 3, Eq. 3.13) and is approximately

$$t_{\text{ff}} = 0.77 h^{-1} \left(\frac{f_c}{200}\right)^{-1/2} \left(\frac{\sigma_c}{\nu}\right)^{3/2} \sigma_0^{-3/2}(M)\ \text{Gyr}. \tag{1.21}$$

From virial theorem, we know that the collapse of a protogalaxy from a marginally bound initial state ($E \simeq 0; T_i \simeq |U|$) to a virialised final state ($T_f \simeq |U|/2$) will lead to the release of a comparable amount of energy. If this is released over a free-fall time scale of $t_{\mathrm{ff}} \simeq 10^8$ yr, then the resulting luminosity is approximately $L = U/t_{\mathrm{ff}}$, which scales as

$$L = 7 \times 10^7 h^{5/3} \left(\frac{M}{10^{11}\, M_\odot}\right)^{5/3} \left(\frac{f_c}{200}\right)^{5/6} \left(\frac{\sigma_c}{\nu}\right)^{-5/2} \sigma_0^{5/2}(M)\, L_\odot. \tag{1.22}$$

Figure 1.2 shows M, v_c, and L in terms of the collapsed redshift for a cosmological model in which $\sigma_0(M)$ is adequately approximated by the fitting function

$$\sigma_0(M) = \frac{7.2 m^{-0.035}}{1 + 0.82 m^{0.23}}, \qquad m = \frac{M}{10^{11}\, M_\odot}, \tag{1.23}$$

in the range $10^7 \lesssim (M/M_\odot) \lesssim 10^{14}$. We have also taken $\sigma_c = 1.68$ and $f_c = 170$; these numerical values as well as the nature of $\sigma_0(M)$ will be justified in Chap. 5, but their orders of magnitude should be obvious from physical considerations. The three curves (from bottom to top) in each of the frames are for $\nu = 1, 2, 3$, which roughly correspond to collapsed objects with fractional abundances of $(2\pi)^{-1/2} \exp(-\nu^2/2) = (0.24, 0.054, 4.4 \times 10^{-3})$. Figure 1.2 shows that the mass scales from 10^7 to 10^{12} $M_\odot$ collapse and virialise, forming nonlinear self-gravitating structures in the redshift range $z = (10\text{–}1)$. Further, it is clear that, although a 1σ fluctuation at the mass scale of, say, 10^{11} $M_\odot$ will collapse at $z \approx 1$, a 3σ fluctuation at the same scale can collapse as early as $z \approx 6$. Because the probability for a 3σ fluctuation relative to a 1σ fluctuation is down by a factor of $(0.0044/0.24) = 0.018$, we would expect ~2% of the structures having a mass of 10^{11} $M_\odot$ to form as early as $z = 6$, although copious production of such structures will occur only at $z \approx (1\text{–}2)$. This analysis shows that galaxy formation is an extended process rather than a single event in the models we will consider.

Exercise 1.2

Putting asunder what gravity has put together: We saw in Vol. II, Chap. 4, that a supernova explosion releases $\sim 10^{51}$ ergs of energy. Show that this hydrodynamic energy can, in principle, unbind the baryonic gas in the dark-matter halos at $z > 5$. Take the binding energy of baryons to be a factor $\Omega_B/\Omega_{\mathrm{DM}} \approx 0.15$ less than the binding energy U computed above.

1.4 Galaxy Formation

The description in the previous section applies only to dark matter that is not directly coupled to radiation. The baryons inside the dark-matter halos can radiate energy, cool, and sink to the centres of the dark-matter halos. As the halos merge, the baryonic structures can survive provided they can cool and condense sufficiently within these halos. We now estimate the conditions for this.

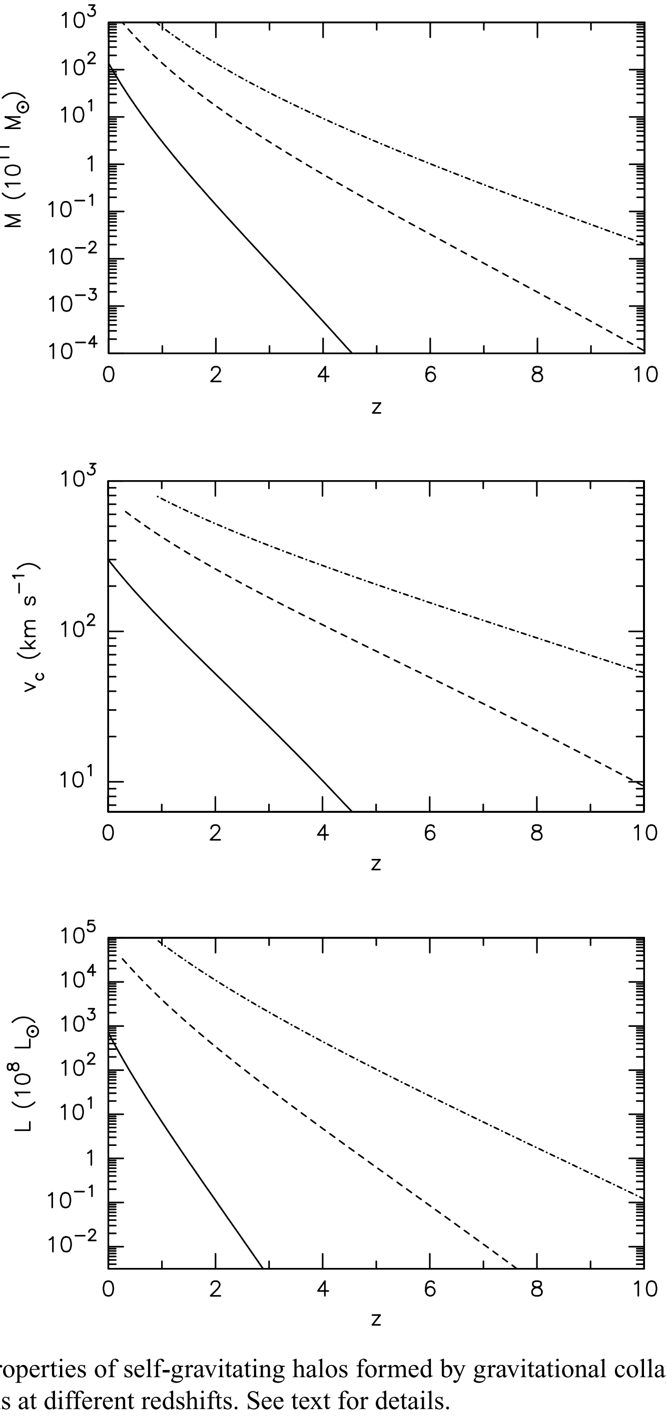

Fig. 1.2. Properties of self-gravitating halos formed by gravitational collapse of overdense regions at different redshifts. See text for details.