

Abraham Loeb and Steven R. Furlanetto

× **THE FIRST  
GALAXIES IN THE  
UNIVERSE**

***NOT APPROVED***

Galaxies fragmented during the plasma epoch when viscous forces first matched gravitational forces at scales less than the scale of causal connection  $ct$ , where  $c$  is the speed of light and  $t$  is the time since the big bang. The time was  $\sim 10^{12}$  seconds (30,000 years). The viscous-gravitational scale depends on the kinematic viscosity of the plasma ( $\nu$ ), the rate-of-strain ( $\gamma$ ), the plasma density ( $\rho$ ), and Newton's gravitational constant ( $G$ ). Condensation was opposed by the rapid expansion of the universe at  $10^{12}$  seconds, so the first structure formation was at density minima by fragmentation to form persistent plasma "proto-galaxies". CHG

## The First Galaxies in the Universe

Density minima were supplied as fossils of turbulent vortex lines of big bang turbulent combustion, Gibson (2004, 2005). These triggered the formation of supervoids that have expanded to 30% of the size of the observable universe ( $\sim 3 \times 10^{25}$  meters), as observed by radio telescopes such as ALMA. Every galaxy has a fossil proto-galaxy at its core, with scale  $\sim 10^{20}$  meters and mass  $\sim 10^{42}$  kg: fossils of the time of first fragmentation  $10^{12}$  s. The proto-galaxies serve as fluid particles for turbulent layers that develop at expanding supervoid boundaries. Superclusters of proto-galaxies terminate the turbulent PG cascade at the plasma-to gas transition at  $\sim 10^{13}$  seconds, as observed. The weak plasma turbulence vortex lines are manifested by chain galaxy clusters (chains of PGs). None of these phenomena are possible according to the standard LCDMHC cosmology, which should be considered obsolete and replaced by hydro-gravitational-dynamics (HGD) cosmology: Schild (1996) and Gibson (1996). See [journalofcosmology.com](http://journalofcosmology.com) for details and illustrations. Unfortunately the present book is completely based on LCDMHC cosmology. From fluid mechanics the concept of "Cold Dark Matter" is dead on arrival. The non-baryonic-dark-matter is weakly collisional and cannot clump or hierarchically cluster. The cosmology is misleading and must be discarded. This book should not be recommended to students except for historical reasons.

## Princeton Series in Astrophysics

Edited by David N. Spergel

Theory of Rotating Stars, *by Jean-Louis Tassoul*Theory of Stellar Pulsation, *by John P. Cox*Galactic Dynamics, Second Edition, *by James Binney and Scott Tremaine*Dynamical Evolution of Globular Clusters, *by Lyman Spitzer, Jr.*Supernovae and Nucleosynthesis: An Investigation of the History of Matter, from the Big Bang to the Present, *by David Arnett*Unsolved Problems in Astrophysics, *edited by John N. Bahcall and Jeremiah P. Ostriker*Galactic Astronomy, *by James Binney and Michael Merrifield*Active Galactic Nuclei: From the Central Black Hole to the Galactic Environment, *by Julian H. Krolik*Plasma Physics for Astrophysics, *by Russell M. Kulsrud*Electromagnetic Processes, *by Robert J. Gould*Conversations on Electric and Magnetic Fields in the Cosmos, *by Eugene N. Parker*High-Energy Astrophysics, *by Fulvio Melia*Stellar Spectral Classification, *by Richard O. Gray and Christopher J. Corbally*Exoplanet Atmospheres: Physical Processes, *by Sara Seager*Physics of the Interstellar and Intergalactic Medium, *by Bruce T. Draine***The First Galaxies in the Universe, *by Abraham Loeb and******Steven R. Furlanetto***

**Fatally flawed by the use of Lambda  
Cold Dark Matter Hierarchical Clustering  
Cosmology. Must use HGD cosmology.**

---

---

# The First Galaxies in the Universe

---

Abraham Loeb and  
Steven R. Furlanetto

Each of the galaxies shown in the book cover (Hubble Space Telescope Ultra-Deep-Field) has a fossil proto-galaxy at its core that was formed during the plasma epoch...a time between  $10^{11}$  seconds when mass exceeded energy and  $10^{13}$  seconds: the time of plasma to gas transition ("recombination"). Proto-Galaxies (PGs) formed at  $\sim 10^{12}$  seconds, and fossilized the density of the plasma ( $\rho_0$ ), size ( $\sim 10^{20}$  meters), and rate-of-strain ( $\gamma_0$ ). All are easily calculated from fluid mechanics, and have been repeatedly confirmed by the flood of excellent observations from the HST and numerous ground based and space telescopes in a variety of frequency bands, as documented in the Journal of Cosmology and elsewhere. The kinematic viscosity of the plasma epoch is estimated by Gibson (1996, 2000, etc.) as  $\nu_0 \sim 10^{26} \text{ m}^2 \text{ s}^{-1}$ , giving a weakly turbulent flow. Such an enormous viscosity cannot be neglected. Chains of protogalaxies fragment along turbulent vortex lines, triggered by the maximum rate-of-strain. These are termed "chain-galaxy-clusters" by HGD cosmology: not "chain galaxies", which is misleading.



PRINCETON UNIVERSITY PRESS  
PRINCETON AND OXFORD

Copyright © 2013 by Princeton University Press

Published by Princeton University Press, 41 William Street,  
Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press, 6 Oxford Street,  
Woodstock, Oxfordshire OX20 1TW

press.princeton.edu

Cover Photograph: *Hubble Extra Deep Field*. Courtesy of NASA, ESA, S. Beckwith (STScI), and the HUDF Team.

All Rights Reserved

Library of Congress Cataloging-in-Publication Data  
Loeb, Abraham.

The first galaxies in the universe / Abraham Loeb and Steven R. Furlanetto.  
p. cm. – (Princeton series in astrophysics)

**X** Summary: "This book provides a comprehensive, self-contained introduction to one of the most exciting frontiers in astrophysics today: the quest to understand how the oldest and most distant galaxies in our universe first formed. Until now, most research on this question has been theoretical, but the next few years will bring about a new generation of large telescopes that promise to supply a flood of data about the infant universe during its first billion years after the big bang. This book bridges the gap between theory and observation. It is an invaluable reference for students and researchers on early galaxies. The *First Galaxies in the Universe* starts from basic physical principles before moving on to more advanced material. Topics include the gravitational growth of structure, the intergalactic medium, the formation and evolution of the first stars and black holes, feedback and galaxy evolution, reionization, 21-cm cosmology, and more. Provides a comprehensive introduction to this exciting frontier in astrophysics. Begins from first principles. Covers advanced topics such as the first stars and 21-cm cosmology. Prepares students for research using the next generation of large telescopes. Discusses many open questions to be explored in the coming decade" – **Provided by publisher.**

Includes bibliographical references and index.

ISBN 978-0-691-14491-7 (hardback) – ISBN 978-0-691-14492-4 (paper)

1. Galaxies–Formation. 2. Stars–Formation. 3. Cosmology.

I. Furlanetto, Steven R. II. Title.

QB857.5.E96L64 2013

523.1'12–dc23 2012018181

British Library Cataloging-in-Publication Data is available

This book has been composed in Scala Lf

Printed on acid-free paper. ∞

Typeset by S R Nova Pvt Ltd, Bangalore, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

**This book should not be recommended to students. The evolution of structure in the universe depends crucially on turbulence, kinematic viscosity, and diffusivity, which are all neglected by this "LCDMHC" treatment based on the Jeans (1902) simplifications of fluid mechanics. CHG**

*To our families*

Fluid mechanical effects on cosmological structure formation can be formulated using dimensional analysis by comparisons of stresses and a definition of turbulence based on the inertial vortex force. The length scale that emerges ( $\sim [\gamma \nu / \rho G]^{1/2}$ ) matches that of Kolmogorov and Obukhov (1941) in their universal similarity hypotheses for turbulent flow at the transition to viscous flow. The Taylor microscale Reynolds number for the big bang was  $\sim 1000$ , but for the Plasma Epoch only  $\sim 100$ : both modest by terrestrial standards, but significantly larger than transitional values  $\sim 10$ . In the plasma this scale is that of proto-galaxies ( $\sim 10^{20}$  m), but in the gas it decreases to that of the Schild (1996) dark matter gas planets ( $\sim 10^{14}$  m). In the Planck scale turbulent combustion fluid of Planck anti-Planck particles, turbulence was damped by gluon viscosity at  $\sim 10^{-27}$  m: Gibson (2004, 2005, 1996 etc.). The huge range of scales is an advantage in establishing the observational facts needed to select HGD cosmology over LCDMHC cosmology using modern space and ground based telescopes. It is an observational fact that neither kinematic viscosity nor turbulence can be neglected in the description of cosmological structure formation. CHG

---



---

## Contents

Preface	xi
PART I. FUNDAMENTALS OF STRUCTURE FORMATION	1
Chapter 1 Introduction and Cosmological Background	3
1.1 Preliminary Remarks	3
1.2 Standard Cosmological Model	5
1.3 Milestones in Cosmic Evolution	15
1.4 Most Matter Is Dark	20
Chapter 2 Linear Growth of Cosmological Perturbations	25
2.1 Growth of Linear Perturbations	25
2.2 The Thermal History during the Dark Ages	35
Chapter 3 Nonlinear Structure and Halo Formation	41
3.1 Spherical Collapse	41
3.2 Cosmological Jeans Mass	45
3.3 Halo Properties	51
3.4 Abundance of Dark Matter Halos	56
3.5 Halo Clustering in Linear Theory	65
3.6 The Nonlinear Power Spectra of Dark Matter and Galaxies	68
3.7 Numerical Simulations of Structure Formation	78
Chapter 4 The Intergalactic Medium	92
4.1 The Cosmic Web	92
4.2 Lyman- $\alpha$ Absorption in the Intergalactic Medium	95
4.3 Theoretical Models of the Lyman- $\alpha$ Forest	100
4.4 The Metagalactic Ionizing Background	114
4.5 The Helium-Ionizing Background	120
4.6 Metal-Line Systems	121
4.7 The Lyman- $\alpha$ Forest at $z > 5$	125



PART II. THE FIRST STRUCTURES	131
Chapter 5 The First Stars	133
5.1 From Virialized Halos to Protostars	136
5.2 From Protostars to Stars	144
5.3 The Second Generation of Stars: "Population III.2"	157
5.4 Properties of the First Stars	163
5.5 The End States of Population III Stars	168
5.6 Gamma-Ray Bursts: The Brightest Explosions	170
Chapter 6 Stellar Feedback and Galaxy Formation	174
6.1 The Ultraviolet Background and H <sub>2</sub> Photodissociation	174
6.2 The X-ray Background: Positive Feedback	184
6.3 Radiative Feedback: Mechanical Effects	186
6.4 Galactic Superwinds and Mechanical Feedback	192
6.5 Metal Enrichment and the Transition to Population II Star Formation	201
6.6 The First Galaxies	211
Chapter 7 Supermassive Black Holes	217
7.1 Quasars and Black Holes: An Overview	217
7.2 Basic Principles of Astrophysical Black Holes	222
7.3 Accretion of Gas onto Black Holes	225
7.4 The First Black Holes and Quasars	232
7.5 Black Holes and Galaxies	237
7.6 Black Hole Binaries	244
7.7 Gravitational Waves from Black Hole Mergers	247
Chapter 8 Physics of Galaxy Evolution	251
8.1 High-Redshift Galaxies	251
8.2 Gas Accretion	253
8.3 Halo Mergers	255
8.4 Disk Formation	256
8.5 Star Formation in Galaxies	258
8.6 Black Hole Growth in Galaxies	263
8.7 Feedback and Galaxy Evolution	264
8.8 From Galaxy Model to Stellar Spectra	266
8.9 Signatures of the Interstellar Medium	269
8.10 Gravitational Lensing	275
Chapter 9 The Reionization of Intergalactic Hydrogen	283
9.1 Propagation of Ionization Fronts	283
9.2 Global Ionization History	288
9.3 The Phases of Hydrogen Reionization	291

CONTENTS	ix
9.4 The Morphology of Reionization	293
9.5 Recombinations inside Ionized Regions	302
9.6 Simulations of Reionization	308
9.7 Statistical Properties of the Ionization Field	315
9.8 Reionization by Quasars and Other Exotic Sources	319
9.9 Feedback from Reionization: Photoheating	326
PART III. OBSERVATIONS OF THE COSMIC DAWN	335
Chapter 10 Surveys of High-Redshift Galaxies	337
10.1 Telescopes for Observing High-Redshift Galaxies	337
10.2 Methods for Identifying High-Redshift Galaxies	340
10.3 Luminosity and Mass Functions	350
10.4 The Statistics of Galaxy Surveys	357
Chapter 11 The Lyman- $\alpha$ Line as a Probe of the Early Universe	367
11.1 Lyman- $\alpha$ Emission from Galaxies	367
11.2 The Gunn-Peterson Trough	375
11.3 IGM Scattering in the Blue Wing of the Lyman- $\alpha$ Line	376
11.4 The Red Damping Wing	382
11.5 The Lyman- $\alpha$ Forest as a Probe of the Reionization Topology	388
11.6 Lyman- $\alpha$ halos around Distant Sources	390
11.7 Lyman- $\alpha$ Emitters during the Reionization Era	396
Chapter 12 The 21-cm Line	408
12.1 Radiative Transfer of the 21-cm Line	410
12.2 The Spin Temperature	412
12.3 The Brightness Temperature of the Spin-Flip Background	420
12.4 The Monopole of the Brightness Temperature	428
12.5 Statistical Fluctuations in the Spin-Flip Background	432
12.6 Spin-Flip Fluctuations during the Cosmic Dawn	439
12.7 Mapping the Spin-Flip Background	446
Chapter 13 Other Probes of the First Galaxies	459
13.1 Secondary Cosmic Microwave Background Anisotropies from the Cosmic Dawn	459
13.2 Diffuse Backgrounds from the Cosmic Dawn	470
13.3 The Cross-Correlation of Different Probes	484
13.4 The Fossil Record of the Local Group	488

x	CONTENTS
Appendix A Useful Numbers	495
Appendix B Cosmological Parameters	497
Notes	499
Further Reading	509
Index	513

---

---

## Preface

The first stars formed  $10^{12}$  s after the PGC clumps of dark matter planets formed at  $10^{13}$  s: within proto-galaxies.

This book captures the latest exciting developments concerning one of the most fascinating unsolved mysteries about our origins: *how did the first stars and galaxies form?* This era, known as the *cosmic dawn* because these sources were the first to illuminate our Universe, assumes central importance in our understanding of the history of the Universe. Most research on this question has been theoretical so far. But the next decade or two will bring about a new generation of large telescopes with unprecedented sensitivity that promise to supply a flood of data about the infant Universe during its first billion years after the Big Bang. Among the new observatories are the James Webb Space Telescope (JWST)—the successor to the Hubble Space Telescope, and three extremely large telescopes on the ground (ranging from 24 to 42 m in diameter), as well as several new arrays of dipole antennas operating at low radio frequencies. The fresh data on the first galaxies and the diffuse gas between them will test existing theoretical ideas about the formation and radiative effects of the first galaxies, and might even reveal new physics that has not yet been anticipated. This emerging interface between theory and observation will constitute an ideal opportunity for students considering a research career in astrophysics or cosmology. Thus the book is intended to provide a self-contained introduction to research on the first galaxies at a technical level appropriate for a graduate student.

The book is organized into three parts that largely build on each other. Part I, *Fundamentals of Structure Formation*, includes chapters on basic cosmology, linear perturbation theory, nonlinear structure formation, and the intergalactic medium. This part provides a broad introduction to studies of cosmological structure and galaxy formation with applications well beyond the first galaxies themselves. The first three chapters provide a crucial introduction to the rest of the book; the fourth (on the intergalactic medium) is not essential for many of the later chapters but is important for understanding the reionization process as well as many of the most important observational probes of the cosmic dawn.

Part II, *The First Structures*, focuses on the physics driving the formation of these objects, as well as the physics that determines their influence on subsequent generations of objects. We review the formation of the first stars and black holes, the importance of stellar feedback, the basic principles of galaxy evolution, and the epoch of reionization. Many of the principles contained here also have wide application to other areas of extragalactic astrophysics, though we focus on their application to the first galaxies. The first two chapters in this part build on each other, but the others can be approached largely independently.

First stars form by mergers of dark matter planets in the gravitational free-fall time of their clumps:  $10^{12}$  seconds.

Part III, *Observations of the Cosmic Dawn*, describes several directions in which we hope to observe the most distant galaxies in the coming decades. This part begins with a discussion of galaxy surveys and then moves on to two unique probes of this era: the Lyman- $\alpha$  and 21-cm lines of neutral hydrogen. It concludes with brief discussions of several other techniques. The chapters in this section are largely independent of each other and may be read in any order.

Throughout the text, we reference seminal papers as well as some recent calculations with endnotes; these are collected in the Notes section. In the Further Reading section, we list useful overviews in the form of books and review papers. We have also included two appendixes. In Appendix A, we include fundamental constants and conversion factors, and, in Appendix B we summarize the cosmological parameters assumed in this book (see also §1.4).

Note that both for the sake of brevity and because the current cosmological measurements are reasonably secure, most of the equations do not explicitly state their dependence on such factors as the baryon density, Hubble constant, or cold dark matter density. Inserting these dependencies is a useful exercise, and we encourage the interested readers to check their understanding in this way.

Various introductory sections of this book are based on an undergraduate-level book, *How Did the First Stars and Galaxies Form?* by one of us (A.L.), which followed a cosmology class he had taught over two decades in the Astronomy and Physics departments at Harvard University. Other parts relate to overviews both of us wrote over the past decade in the form of review articles. Where necessary, selected references are given to advanced papers and other review articles in the scientific literature.

The writing of this book was made possible thanks to the help we received from many individuals. First and foremost, we are grateful to our families for their support and patience during the lengthy writing period of the book. Needless to say, the content of this book echoes many papers and scientific discussions we had over the years with our students, postdocs, and senior collaborators, including Dan Babich, Rennan Barkana, Jon Bittner, Laura Blecha, Judd Bowman, Frank Briggs, Avery Broderick, Volker Bromm, Chris Carilli, Renyue Cen, Benedetta Ciardi, T. J. Cox, Mark Dijkstra, Daniel Eisenstein, Claude-André Faucher-Giguère, Richard Ellis, Idan Ginsburg, Zoltan Haiman, Lars Hernquist, Jackie Hewitt, Loren Hoffman, Bence Kocsis, Girish Kulkarni, Adam Lidz, Andrei Mesinger, Matt McQuinn, Joey Muñoz, Ramesh Narayan, Peng Oh, Ryan O’Leary, Rosalba Perna, Tony Pan, Ue-Li Pen, Jonathan Pritchard, Fred Rasio, Martin Rees, Doug Rubin, George Rybicki, Athena Stacy, Dan Stark, Yue Shen, Nick Stone, Anne Thoul, Hy Trac, Eli Visbal, Stuart Wyithe, and Matias Zaldarriaga. We did not attempt to provide a comprehensive reference list of the related literature, since such a list would be out of date within a few years in this rapidly evolving frontier. Instead we focused pedagogically on the underlying physical principles that will remain relevant in the future, and referred the reader to representative papers, review articles, and books for further reading. We thank Nina Zonnevylle and Uma Mirani for their assistance in obtaining permissions for the figures of the book; Laurie Lites for her

## PREFACE

xiii

assistance with the manuscript; Fred Davies, Lauren Holzbauer, Joey Muñoz, and Ramesh Narayan for their help with several figures; and Natalie Mashian, Doug Rubin, and Anjali Tripathi for their comments on the finished manuscript. Finally, it has been a delightful experience for us to work with our book editor, Ingrid Gnerlich, and the entire production team at Princeton University Press.

—A. L. & S. F.



PART I

Fundamentals of Structure Formation





# Chapter One

---

## Introduction and Cosmological Background

### 1.1 Preliminary Remarks

The initial density was large but not infinite. It was the Planck density  $\sim 10^{97} \text{ kg m}^{-3}$ . See Gibson (2004, 2005) for the big bang mechanism.

On large scales, the Universe is observed to be expanding. As it expands, galaxies separate from one another, and the density of matter (averaged over a large volume of space) decreases. If we imagine playing the cosmic movie in reverse and tracing this evolution backward in time, we would infer that there must have been an instant when the density of matter was infinite. This moment in time is the Big Bang, before which we cannot reliably extrapolate our history. But even before we get all the way back to the Big Bang, there must have been a time when stars like our Sun and galaxies like our Milky Way did not exist, because the Universe was denser than they are. If so, *how and when did the first stars and galaxies form?*

Primitive versions of this question were considered by humans for thousands of years, long before it was realized that the Universe is expanding. Religious and philosophical texts attempted to provide a sketch of the big picture from which people could derive the answer. In retrospect, these attempts appear heroic in view of the scarcity of scientific data about the Universe prior to the twentieth century. To appreciate the progress made over the past century, consider, for example, the biblical story of Genesis. The opening chapter of the Bible asserts the following sequence of events: first, the Universe was created, then light was separated from darkness, water was separated from the sky, continents were separated from water, vegetation appeared spontaneously, stars formed, life emerged, and finally humans appeared on the scene. Instead, the modern scientific order of events begins with the Big Bang, followed by an early period in which light (radiation) dominated and then a longer period in which matter was preeminent and led to the appearance of stars, planets, life on Earth, and eventually humans. Interestingly, the starting and end points of both versions are the same.

Cosmology is by now a mature empirical science. We are privileged to live in a time when the story of genesis (how the Universe started and developed) can be critically explored by direct observations. Because light takes a finite time to travel to us from distant sources, we can see images of the Universe when it was younger by looking deep into space through powerful telescopes.

Existing data sets include an image of the Universe when it was 400,000 years old (in the form of the cosmic microwave background in Figure 1.1), as well as images of individual galaxies when the Universe was older than a billion years. But there is a serious challenge: between these two epochs was a period when

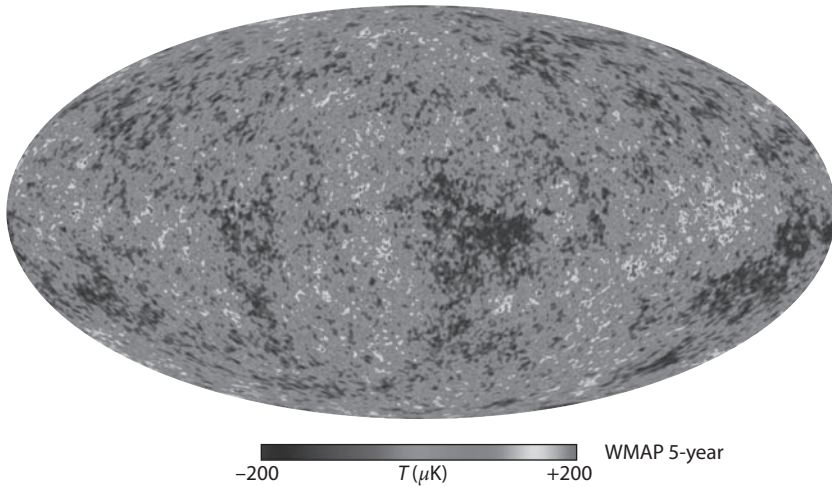


Figure 1.1 Image of the Universe when it first became transparent, 400,000 years after the Big Bang, taken over 5 years by the *Wilkinson Microwave Anisotropy Probe* (WMAP) satellite (see *Color Plate 1* for a color version of this figure). Slight density inhomogeneities at the level of one part in  $\sim 10^5$  in the otherwise uniform early Universe imprinted hot and cold spots in the temperature map of the cosmic microwave background on the sky. The fluctuations are shown in units of microkelvins, and the unperturbed temperature is 2.725 K. The same primordial inhomogeneities seeded the large-scale structure in the present-day Universe. The existence of background anisotropies was predicted in a number of theoretical papers three decades before the technology for taking this image became available. Courtesy of NASA and the WMAP Science Team.

the Universe was dark, stars had not yet formed, and the cosmic microwave background no longer traced the distribution of matter. And this is precisely the most interesting period, when the primordial soup evolved into the rich zoo of objects we now see. *How can astronomers see this dark yet crucial time?*

The situation is similar to having a photo album of a person that begins with the first ultrasound image of him or her as an unborn baby and then skips to some additional photos of his or her years as teenager and adult. The later photos do not simply show a scaled-up version of the first image. We are currently searching for the missing pages of the cosmic photo album that will tell us how the Universe evolved during its infancy to eventually make galaxies like our own Milky Way.

Observers are moving ahead along several fronts. The first involves the construction of large infrared telescopes on the ground and in space that will provide us with new (although rather expensive!) photos of galaxies in the Universe at intermediate ages. Current plans include ground-based telescopes 24–42 m in diameter and NASA's successor to the Hubble Space Telescope, the James Webb Space Telescope (JWST). In addition, several observational groups around the globe are constructing radio arrays that will be capable of

mapping the three-dimensional distribution of cosmic hydrogen left over from the Big Bang in the infant Universe. These arrays are aiming to detect the long-wavelength (redshifted 21-cm) radio emission from hydrogen atoms. Coincidentally, this long wavelength (or low frequency) overlaps the band used for radio and television broadcasting, and so these telescopes include arrays of regular radio antennas that one can find in electronics stores. These antennas will reveal how the clumpy distribution of neutral hydrogen evolved with cosmic time. By the time the Universe was a few hundreds of millions of years old, the hydrogen distribution had been punched with holes and resembled Swiss cheese. These holes were created by the ultraviolet radiation from the first galaxies and black holes, which ionized the cosmic hydrogen in their vicinity.

Theoretical research has focused in recent years on predicting the signals expected from the telescopes described and on providing motivation for these ambitious observational projects.

All these predictions are generated in the context of the modern cosmological paradigm, which turns the Big Bang model into a quantitative tool for understanding our Universe. In the remainder of this chapter, we briefly describe the essential aspects of this paradigm for understanding the formation of the first galaxies in the Universe.


## 1.2 Standard Cosmological Model

### 1.2.1 Cosmic Perspective

In 1915 Einstein formulated the general theory of relativity. He was inspired by the fact that all objects follow the same trajectories under the influence of gravity (the so-called equivalence principle, which by now has been tested to better than one part in a trillion), and realized that this would be a natural result if space–time is curved under the influence of matter. He wrote an equation describing how the distribution of matter (on one side of his equation) determines the curvature of space–time (on the other side of his equation). Einstein then applied his equation to describe the global dynamics of the Universe.

There were no computers available in 1915, and Einstein's equations for the Universe were particularly difficult to solve in the most general case. To get around this obstacle Einstein considered the simplest possible Universe, one that is homogeneous and isotropic. Homogeneity means uniform conditions everywhere (at any given time), and isotropy means the same conditions in all directions seen from one vantage point. The combination of these two simplifying assumptions is known as the *cosmological principle*.

The Universe can be homogeneous but not isotropic: for example, the expansion rate could vary with direction. It can also be isotropic and not homogeneous: for example, we could be at the center of a spherically symmetric mass distribution. But if it is isotropic around *every* point, then it must also be homogeneous.

Under the simplifying assumptions associated with the cosmological principle, Einstein and his contemporaries were able to solve the equations. They were looking for their “lost keys” (solutions) under a convenient “lamppost” (simplifying assumptions), but the real Universe is not bound by any contract to be the simplest that we can imagine. **In fact, it is truly remarkable in the first place that we dare describe the conditions across vast regions of space based on the blueprint of the laws of physics that describe the conditions here on Earth. Our daily life teaches us too often that we fail to appreciate complexity, and that an elegant model for reality is often too idealized for describing the truth (along the lines of approximating a cow as a spherical object).** 

In 1915 Einstein had the wrong notion of the Universe; at the time people associated the Universe with the Milky Way galaxy and regarded all the “spiral nebulae,” which we now know are distant galaxies, as constituents of our own Milky Way galaxy. Because the Milky Way is not expanding, Einstein attempted to reproduce a static universe with his equations. This turned out to be possible only after he added a cosmological constant, whose negative gravity would exactly counteract that of matter. However, Einstein later realized that this solution is unstable: a slight enhancement in density would make the density grow even further. As it turns out, there are no stable static solutions to Einstein’s equations for a homogeneous and isotropic Universe. The Universe must be either expanding or contracting. Less than a decade later, Edwin Hubble discovered that the nebulae previously considered to be constituents of the Milky Way galaxy are receding from us at a speed  $v$  that is proportional to their distance  $r$ , namely,  $v = H_0 r$ , where  $H_0$  is a spatial constant (which can evolve with time), commonly termed the *Hubble constant*.<sup>1</sup> Hubble’s data indicated that the Universe is expanding. (Hubble also resolved individual bright stars in these nebulae, unambiguously determining their nature and their vast distances from the Milky Way.)

Einstein was remarkably successful in asserting the cosmological principle. As it turns out, our latest data indicate that the real Universe is homogeneous and isotropic on the largest observable scales to within one part in  $10^5$ . In particular, isotropy is well established for the distribution of faint radio sources, optical galaxies, the X-ray background, and most important, the cosmic microwave background (CMB). The constraints on homogeneity are less strict, but a cosmological model in which the Universe is isotropic and significantly inhomogeneous in spherical shells around our special location is also excluded based on surveys of galaxies and quasars. Fortunately, Einstein’s simplifying assumptions turned out to be extremely accurate in describing reality: *the keys were indeed lying next to the lamppost*. Our Universe happens to be the simplest we could have imagined, for which Einstein’s equations can easily be solved.

<sup>1</sup>The redshift data examined by Hubble was mostly collected by Vesto Slipher a decade earlier and only partly by Hubble’s assistant, Milton L. Humason. The linear local relation between redshift and distance (based on Hubble and Humason’s data) was first formulated by Georges Lemaître in 1927, 2 years prior to the observational paper written by Hubble and Humason.

Why was the Universe prepared to be in this special state? Cosmologists were able to go one step further and demonstrated that an early phase transition, called *cosmic inflation*—during which the expansion of the Universe accelerated exponentially—could have naturally produced the conditions postulated by the cosmological principle (although other explanations also may create such conditions). One is left to wonder whether the existence of inflation is just a fortunate consequence of the fundamental laws of nature, or whether perhaps the special conditions of the specific region of space–time we inhabit were selected out of many random possibilities elsewhere by the prerequisite that they allow our existence. The opinions of cosmologists on this question are split.

### 1.2.2 Origin of Structure

Planck density  $10^{97} \text{ kg/m}^3$ .

Hubble's discovery of the expansion of the Universe has immediate implications for the past and future of the Universe. If we reverse in our mind the expansion history back in time, we realize that the Universe must have been denser in its past. In fact, there must have been a point in time where the matter density was infinite, at the moment of the so-called Big Bang. Indeed, we do detect relics from a hotter, denser phase of the Universe in the form of light elements (such as deuterium, helium, and lithium) as well as the CMB. At early times, this radiation coupled extremely well to the cosmic gas and produced a spectrum known as a *blackbody*, a form predicted a century ago to characterize matter and radiation in equilibrium. The CMB provides the best example of a blackbody spectrum we have.


To get a rough estimate of when the Big Bang occurred, we may simply divide the distance of all galaxies by their recession velocity. This calculation gives a unique answer,  $\sim r/v \sim 1/H_0$ , that is independent of distance.<sup>ii</sup> The latest measurements of the Hubble constant give a value of  $H_0 \approx 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , which implies a current age for the Universe  $1/H_0$  of 14 billion years (or  $5 \times 10^{17}$  seconds).

The second implication concerns our future. A fortunate feature of a spherically symmetric Universe is that when considering a sphere of matter in it, we are allowed to ignore the gravitational influence of everything outside this sphere. If we empty the sphere and consider a test particle on the boundary of an empty void embedded in a uniform Universe, the particle will experience no net gravitational acceleration. This result, known as *Birkhoff's theorem*, is reminiscent of Newton's "iron sphere theorem." It allows us to solve the equations of motion for matter on the boundary of the sphere through a local analysis without worrying about the rest of the Universe. Therefore, if the sphere has exactly the same conditions as the rest of the Universe, we may deduce the global expansion history of the Universe by examining its behavior. If the sphere is

<sup>ii</sup>Although this is an approximate estimate, it turns out to be within a few percent of the true age of our Universe owing to a coincidence. The cosmic expansion at first decelerated and then accelerated, with the two almost canceling each other at the present time, giving the same age as if the expansion were at a constant speed (as would be strictly true only in an empty Universe).

Not true. The existence of dark matter planets in protoglobular-star-cluster clumps of a trillion that make all the stars causes a systematic dimming error. There was no acceleration of the expansion rate, and no dark energy.

slightly denser than the mean, we will infer how its density contrast will evolve relative to the background Universe.

For the moment, let us ignore the energy density of the vacuum (which is always a good approximation at sufficiently early cosmic times, when matter was denser). Then, the equation describing the motion of a spherical shell of matter is identical with the equation of motion of a rocket launched from the surface of the earth. The rocket will escape to infinity if its kinetic energy exceeds its gravitational binding energy, making its total energy positive. However, if its total energy is negative, the rocket will reach a maximum height and then fall back. To deduce the future evolution of the Universe, we need to examine the energy of a spherical shell of matter relative to the origin. With a uniform density  $\rho$ , a spherical shell of radius  $r$  has a total mass  $M = \rho \times (4\pi r^3/3)$  enclosed within it. Its energy per unit mass is the sum of the kinetic energy due to its expansion speed  $v = Hr$ ,  $(1/2)v^2$ , and its potential gravitational energy,  $-GM/r$  (where  $G$  is Newton's constant), namely,  $E = v^2/2 - GM/r$ . By substituting the preceding relations for  $v$  and  $M$ , we can easily show that  $E = (1/2)v^2(1 - \Omega)$ , where  $\Omega = \rho/\rho_c$ , and  $\rho_c = 3H^2/8\pi G$  is defined as the *critical density*. We therefore find that there are three possible scenarios for the cosmic expansion. The Universe has either (i)  $\Omega > 1$ , making it gravitationally bound with  $E < 0$ —such a “closed Universe” will turn around and end up collapsing toward a “big crunch”; (ii)  $\Omega < 1$ , making it gravitationally unbound with  $E > 0$ —such an “open Universe” will expand forever; or the borderline case, (iii)  $\Omega = 1$ , making the Universe marginally bound or “flat” with  $E = 0$ . 

Einstein's equations relate the geometry of space to its matter content through the value of  $\Omega$ : an open Universe has the geometry of a saddle with a negative spatial curvature, a closed Universe has the geometry of a spherical globe with a positive curvature, and a flat Universe has a flat geometry with no curvature. Our observable section of the Universe appears to be flat.

Now we are in a position to understand how objects like the Milky Way galaxy formed out of small density inhomogeneities that are amplified by gravity.

Let us consider for simplicity the background of a marginally bound (flat) Universe dominated by matter. In such a background, only a slight enhancement in density is required to exceed the critical density  $\rho_c$ . Because of Birkhoff's theorem, a spherical region denser than the mean will behave as if it is part of a closed Universe and will increase its density contrast with time, while an underdense spherical region will behave as if it is part of an open Universe and will appear more vacant with time relative to the background, as illustrated in Figure 1.2. Starting with slight density enhancements that bring them above the critical value,  $\rho_c$ , the overdense regions will initially expand, reach a maximum radius, and then collapse on themselves (like the trajectory of a rocket launched straight up, away from the center of the earth). An initially slightly inhomogeneous Universe will end up clumpy, with collapsed objects forming out of overdense regions. The material to make the objects is drained out of the intervening underdense regions, which end up as voids.

The Universe we live in started with primordial density perturbations of a fractional amplitude  $\sim 10^{-5}$  when the cosmic microwave background last scattered. The overdensities were amplified at late times (once matter dominated

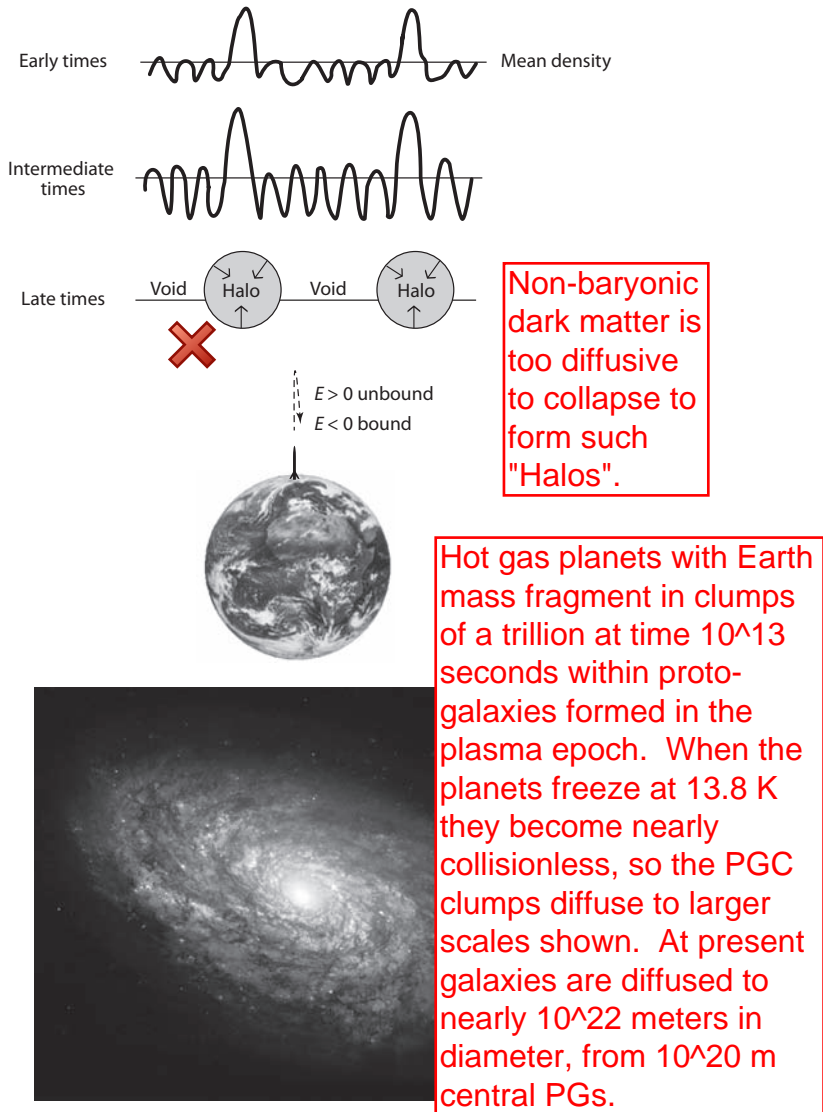


Figure 1.2 *Top*: Schematic illustration of the growth of perturbations to collapsed halos through gravitational instability. The overdense regions initially expand, reach a maximum size, and then turn around and collapse to form gravitationally bound halos if their density exceeds a critical threshold (see §3.1). The material that makes the halos originated in the voids that separate them. *Middle*: A simple model for the collapse of a spherical region. The dynamical fate of a rocket launched from the surface of the earth depends on the sign of its energy per unit mass,  $E = (1/2)v^2 - GM_{\oplus}/r$ . The behavior of a spherical shell of matter on the boundary of an overdense region (embedded in a homogeneous and isotropic Universe) can be analyzed in a similar fashion. *Bottom*: A collapsing region may end up as a galaxy, like NGC 4414, shown here. (Courtesy of NASA and ESA.) The halo gas cools and condenses to a compact disk surrounded by an extended dark matter halo.



the cosmic mass budget) up to values close to unity and collapsed to make objects, first on small scales. We have not yet seen the first small galaxies that started the process that eventually led to the formation of big galaxies like the Milky Way. The search for the first galaxies is a search for our origins and the main subject of this book.

Beyond being uniform, the process of *Big Bang* nucleosynthesis (these were mostly hydrogen in the form of hydrogen). However,

The water molecule includes oxygen, an element that was not made in the Big Bang and did not exist until the first stars had formed. Therefore, our form of life could not have existed in the first hundred million years after the Big Bang, before the first stars had formed. There is also no guarantee that life will persist in the distant future.

All galaxies form as  $10^{20}$  meter proto-galaxies: there are no first small galaxies that lead to the formation of big galaxies like the Milky Way.

Life in its present form began its existence at 2 million years: DNA, RNA, humanoids, on  $10^{80}$  planets.

### 1.2.3 Geometry of Space

The history and fate of our Universe are thus tied inexorably to its contents—be it matter, dark energy, or something even more exotic. However, measuring the average density of the Universe is extraordinarily difficult. Fortunately, Einstein's equations show that the contents of the Universe are also tied to its geometry—so measuring the latter would indirectly constrain its components.

*How can we tell the difference between the flat surface of a book and the curved surface of a balloon?* A simple way is to draw a triangle of straight lines between three points on those surfaces and measure the sum of the three angles of the triangle. The Greek mathematician Euclid demonstrated that the sum of these angles must be  $180^\circ$  (or  $\pi$  radians) on a flat surface. Twenty-one centuries later, the German mathematician Bernhard Riemann extended the field of geometry to curved spaces, which played an important role in the development of Einstein's general theory of relativity. For a triangle drawn on a positively curved surface, like that of a balloon, the sum of the angles is larger than  $180^\circ$ . (This can easily be figured out by examining a globe and noticing that any line connecting one of the poles to the equator opens an angle of  $90^\circ$  relative to the equator. Adding the third angle in any triangle stretched between the pole and the equator would surely result in a total of more than  $180^\circ$ .) According to Einstein's equations, the geometry of the Universe is dictated by its matter content; in particular, the Universe is flat only if the total  $\Omega$  equals unity. *Is it possible to draw a triangle across the entire Universe and measure its geometry?*

Remarkably, the answer is yes. At the end of the twentieth century cosmologists were able to perform this experiment by adopting a simple yardstick provided by the early Universe. The familiar experience of dropping a stone in the middle of a pond results in a circular wave crest that propagates outward. Similarly, perturbing the smooth Universe at a single point at the Big Bang would have resulted in the propagation of a spherical sound wave outward from that point. The wave would have traveled at the speed of sound, which was of the order of the speed of light  $c$  (or, more precisely,  $c/\sqrt{3}$ ) early on

when radiation dominated the cosmic mass budget. At any given time, all the points extending to the distance traveled by the wave are affected by the original pointlike perturbation. The conditions outside this “sound horizon” will remain uncorrelated with the central point, because acoustic information has not yet been able to reach them. The temperature fluctuations of the CMB trace the simple sum of many such pointlike perturbations that were generated in the Big Bang. The patterns they delineate will therefore show a characteristic correlation scale, corresponding to the sound horizon at the time when the CMB was produced, 400,000 years after the Big Bang. By measuring the apparent angular scale of this “standard ruler” on the sky, known as the acoustic peak in the CMB, and comparing it with theory, experimental cosmologists inferred from the simple geometry of triangles that the Universe is flat (or at least very close to it).

The inferred flatness may be a natural consequence of the early period of cosmic inflation during which any initial curvature was flattened. Indeed, a small patch of a fixed size (representing our current observable region in the cosmological context) on the surface of a vastly inflated balloon would appear nearly flat. The sum of the angles on a nonexpanding triangle placed on this patch would get arbitrarily close to  $180^\circ$  as the balloon inflated.

Even though we now know that our Universe is very close to being flat, this flatness only constrains the cumulative energy density in the Universe; it tells us very little about how that energy is distributed among the different components, such as baryons, other forms of matter, and dark energy. We must probe our Universe in other ways to learn about this distribution.

#### 1.2.4 Observing Our Past: Cosmic Archaeology

Our Universe is the simplest possible on two counts: it satisfies the cosmological principle, and it has a flat geometry. The mathematical description of an expanding, homogeneous, and isotropic Universe with a flat geometry is straightforward. We can imagine filling up space with clocks that are all synchronized. At any given snapshot in time the physical conditions (density, temperature) are the same everywhere. But as time goes on, the spatial separation between the clocks will increase. The stretching of space can be described by a time-dependent scale factor,  $a(t)$ . A separation measured at time  $t_1$  as  $r(t_1)$  will appear at time  $t_2$  to have a length  $r(t_2) = r(t_1)[a(t_2)/a(t_1)]$ .

A natural question to ask is whether our human bodies or even the solar system is also expanding as the Universe expands. The answer is no, because these systems are held together by forces whose strength far exceeds the cosmic force. The mean density of the Universe today,  $\bar{\rho}$ , is 29 orders of magnitude smaller than the density of our body. Not only are the electromagnetic forces that keep the atoms in our body together far greater than the force of gravity, but even the gravitational self-force of our body on itself overwhelms the cosmic influence. Only on very large scales does the cosmic gravitational force dominate the scene. This also implies that we cannot observe the cosmic expansion with a local laboratory experiment; to notice the expansion we need to observe sources spread over the vast scales of millions of light-years.

The space–time of an expanding homogeneous and isotropic flat Universe can be described very simply. Because of the cosmological principle, we can establish a unique time coordinate throughout space by distributing clocks that are all synchronized throughout the Universe, so that each clock will measure the same time  $t$  since the Big Bang. The space–time (four–dimensional) line element  $ds$ , commonly defined to vanish for a photon, is described by the Friedmann–Robertson–Walker (FRW) metric,

$$ds^2 = c^2 dt^2 - d\ell^2, \quad (1.1)$$

where  $c$  is the speed of light and  $d\ell$  is the spatial line element. The cosmic expansion can be incorporated through a scale factor  $a(t)$  that multiplies the fixed  $(x, y, z)$  coordinates tagging the clocks, which are themselves “comoving” with the cosmic expansion. For a flat space,

$$d\ell^2 = a(t)^2(dx^2 + dy^2 + dz^2) = a^2(t)(dR^2 + R^2 d\Omega), \quad (1.2)$$

where  $d\Omega = d\theta^2 + \sin^2\theta d\phi^2$ ,  $(R, \theta, \phi)$  are the comoving spherical coordinates centered on the observer, and  $(x, y, z) = (R \cos\theta, R \sin\theta \cos\phi, R \sin\theta \sin\phi)$ . Throughout this book, we quote distances in these comoving units—as opposed to their time-varying *proper* equivalents—unless otherwise specified.

A source located at a separation  $r = a(t)R$  from us will move at a velocity  $v = dr/dt = \dot{a}R = (\dot{a}/a)r$ , where  $\dot{a} = da/dt$ . Here  $r$  is a time-independent tag denoting the present-day distance of the source (when  $a(t) \equiv 1$ ). Defining  $H = \dot{a}/a$ , which is constant in space, we recover the Hubble expansion law  $v = Hr$ .

Edwin Hubble measured the expansion of the Universe using the Doppler effect. We are all familiar with the same effect for sound waves: when a moving car sounds its horn, the pitch (frequency) we hear is different when the car is approaching us than when it is receding from us. Similarly, the wavelength of light depends on the velocity of the source relative to us. As the Universe expands, a light source will move away from us, and its Doppler effect will change with time. The Doppler formula for a nearby source of light (with a recession speed much smaller than the speed of light) gives

$$\frac{\Delta v}{v} \approx -\frac{\Delta v}{c} = -\left(\frac{\dot{a}}{a}\right)\left(\frac{r}{c}\right) = -\frac{(\dot{a}\Delta t)}{a} = -\frac{\Delta a}{a}, \quad (1.3)$$

with the solution  $v \propto a^{-1}$ . Correspondingly, the wavelength scales as  $\lambda = (c/v) \propto a$ . We could have anticipated this outcome, since a wavelength can be used as a measure of distance and should therefore be stretched as the Universe expands. This relation holds also for the de Broglie wavelength,  $\lambda_{\text{dB}} = (h/p) \propto a$ , characterizing the quantum-mechanical wavefunction of a massive particle with momentum  $p$  (where  $h$  is Planck’s constant). Consequently, the kinetic energy of a nonrelativistic particle scales as  $(p^2/2m_p) \propto a^{-2}$ . Thus, in the absence of heat exchange with other systems, the temperature of a gas of nonrelativistic protons and electrons will cool faster ( $\propto a^{-2}$ ) than the CMB temperature ( $h\nu \propto a^{-1}$ ) as the Universe expands and  $a$  increases. The redshift  $z$  is defined through the factor  $(1+z)$  by which the photon wavelength was stretched

(or its frequency reduced) between its emission and observation times. If we define  $a = 1$  today, then  $a = 1/(1+z)$  at earlier times. Higher redshifts correspond to a higher recession speed of the source relative to us (that ultimately approaches the speed of light when the redshift goes to infinity), which in turn implies a larger distance (that ultimately approaches our horizon, which is the distance traveled by light since the Big Bang) and an earlier emission time of the source for the photons to reach us today.

We see high-redshift sources as they looked at early cosmic times. Observational cosmology is like archaeology—the deeper we look into space, the more ancient the clues about our history are (see Figure 1.3).<sup>iii</sup> But there is a limit to how far back we can see: we can image the Universe only if it is transparent. Earlier than 400,000 years after the Big Bang, the cosmic gas was sufficiently hot to be fully ionized, and the Universe was opaque owing to scattering by the dense fog of free electrons that filled it. Thus, telescopes cannot be used to image the infant Universe at earlier times (at redshifts  $> 10^3$ ). The earliest possible image of the Universe can be seen in the cosmic microwave background, the thermal radiation left over from the transition to transparency (Figure 1.1). The first galaxies are believed to have formed long after that.

The expansion history of the Universe is captured by the scale factor  $a(t)$ . We can write a simple equation for the evolution of  $a(t)$  based on the behavior of a small region of space. For that purpose we need to incorporate the fact that in Einstein's theory of gravity, not only does mass density  $\rho$  gravitate but pressure  $p$  does as well. In a homogeneous and isotropic Universe, the quantity  $\rho_{\text{grav}} = (\rho + 3p/c^2)$  plays the role of the gravitating mass density  $\rho$  of Newtonian gravity. There are several examples to consider. For a radiation fluid,<sup>iv</sup>  $p_{\text{rad}}/c^2 = (1/3)\rho_{\text{rad}}$ , which implies that  $\rho_{\text{grav}} = 2\rho_{\text{rad}}$ .

However, if the vacuum has a nonzero energy density that is constant in space and time, the *cosmological constant*, then the pressure of the vacuum is negative, because by opening up a new volume increment  $\Delta V$  one gains an energy  $\rho_{\text{vac}}c^2\Delta V$  instead of losing it, as is the case for normal fluids that expand into more space. In thermodynamics, pressure is derived from the deficit in energy per unit of new volume, which in this case gives  $p_{\text{vac}}/c^2 = -\rho_{\text{vac}}$ . This relation in turn leads to another reversal of signs,  $\rho_{\text{grav}} = (\rho_{\text{vac}} + 3p_{\text{vac}}/c^2) = -2\rho_{\text{vac}}$ , which may be interpreted as repulsive gravity! This surprising result gives rise to the phenomenon of accelerated cosmic expansion, which characterized the early period of cosmic inflation as well as the latest 6 billion years of cosmic history.

**A negative pressure of  $\sim -10^{113}$  Pa was produced by inertial vortex forces of big bang turbulent combustion.**

<sup>iii</sup>Cosmology and archaeology share another similarity: both are *observational*, rather than *experimental*, sciences. Thus, we are forced to interpret the complicated physics of actual systems rather than design elegant experiments that can answer targeted questions. Although simplified models can be built in the laboratory (or even inside computers), the primary challenge of cosmology is figuring out how to extract useful information from real and complex systems that cannot be artificially altered.

<sup>iv</sup>The momentum of each photon is  $1/c$  of its energy. The pressure is defined as the momentum flux along one dimension out of three and is therefore given by  $(1/3)\rho_{\text{rad}}c^2$ , where  $\rho_{\text{rad}}$  is the equivalent mass density of the radiation.

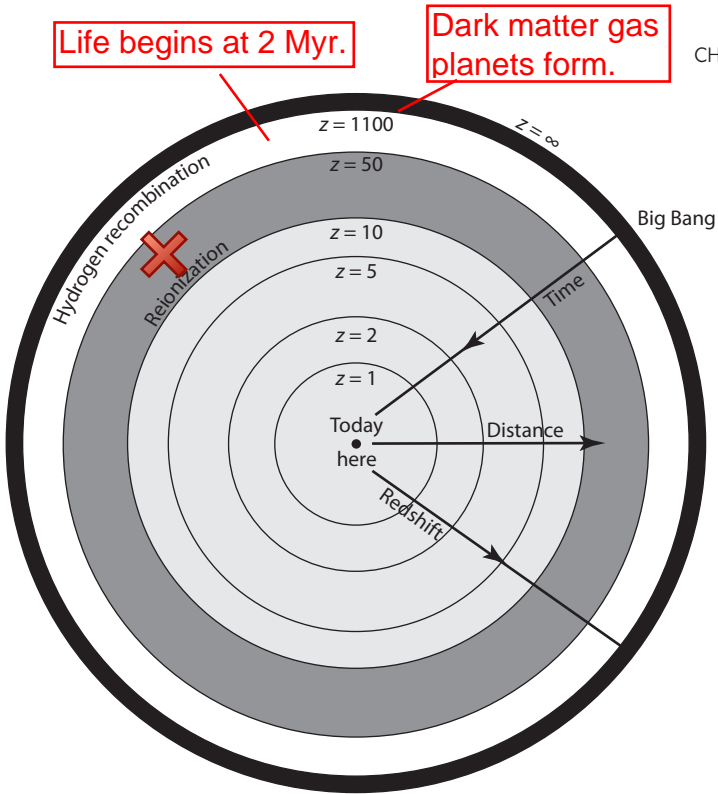


Figure 1.3 Cosmic archaeology of the observable volume of the Universe, in comoving coordinates (which factor out the cosmic expansion). The outermost observable boundary ( $z = \infty$ ) marks the comoving distance that light has traveled since the Big Bang. Future observatories aim to map most of the observable volume of our Universe and to improve dramatically the statistical information we have about the density fluctuations within it. Existing data on the CMB probe mainly a very thin shell at the hydrogen recombination epoch ( $z \sim 10^3$ , beyond which the Universe is opaque), and current large-scale galaxy surveys map only a small region near us at the center of the diagram. The formation epoch of the first galaxies that culminated with hydrogen reionization at a redshift  $z \sim 10$  is shaded dark gray. Note that the comoving volume out to any of these redshifts scales as the distance cubed.

As the Universe expands and the scale factor increases, the matter mass density declines inversely with volume,  $\rho_{\text{matter}} \propto a^{-3}$ , whereas the radiation energy density (which includes the CMB and three species of relativistic neutrinos) decreases as  $\rho_{\text{rad}} c^2 \propto a^{-4}$ , because not only is the density of photons diluted as  $a^{-3}$ , but the energy per photon  $h\nu = hc/\lambda$  (where  $h$  is Planck's constant) declines as  $a^{-1}$ . Today  $\rho_{\text{matter}}$  is larger than  $\rho_{\text{rad}}$  (assuming massless neutrinos) by a factor of  $\sim 3,300$ , but at  $(1+z) \sim 3,300$  the two were equal, and at even higher redshifts the radiation dominated. Since a stable vacuum does not get diluted with cosmic expansion, the present-day  $\rho_{\text{vac}}$  remained a constant and dominated over  $\rho_{\text{matter}}$  and  $\rho_{\text{rad}}$  only at late times (whereas the unstable “false vacuum” that dominated during inflation decayed when inflation ended).

In this book, we will primarily be concerned with the *cosmic dawn*, or the era in which the first galaxies formed at  $z \sim 6\text{--}30$ . At these early times, the cosmological constant was very small compared with the matter densities and can generally be ignored.

### 1.3 Milestones in Cosmic Evolution

The gravitating mass,  $M_{\text{grav}} = \rho_{\text{grav}}V$ , enclosed by a spherical shell of radius  $r(t) = a(t)$  and volume  $V = (4\pi/3)a^3$ , induces an acceleration

$$\frac{d^2a}{dt^2} = -\frac{GM_{\text{grav}}}{a^2}. \quad (1.4)$$

Since  $\rho_{\text{grav}} = \rho + 3p/c^2$ , we need to know how pressure evolves with the expansion factor  $a(t)$ . We obtain this information from the thermodynamic relation mentioned previously between the change in the internal energy  $d(\rho c^2V)$  and the  $p dV$  work done by the pressure,  $d(\rho c^2V) = -p dV$ . This relation implies  $-3pa\dot{a}/c^2 = a^2\dot{\rho} + 3\rho a\dot{a}$ , where an overdot denotes a time derivative. Multiplying equation (1.4) by  $\dot{a}$  and making use of this relation yields our familiar result

$$E = \frac{1}{2}\dot{a}^2 - \frac{GM}{a}, \quad (1.5)$$

where  $E$  is a constant of integration, and  $M \equiv \rho V$ . As discussed before, the spherical shell will expand forever (being gravitationally unbound) if  $E \geq 0$  but will eventually collapse (being gravitationally bound) if  $E < 0$ . Making use of the Hubble parameter,  $H = \dot{a}/a$ , we can rewrite equation (1.5) as

$$\frac{E}{\dot{a}^2/2} = 1 - \Omega, \quad (1.6)$$

where  $\Omega = \rho/\rho_c$ , with

$$\rho_c = \frac{3H^2}{8\pi G} = 9.2 \times 10^{-30} \frac{\text{g}}{\text{cm}^3} \left( \frac{H}{70 \text{ km s}^{-1} \text{ Mpc}^{-1}} \right)^2. \quad (1.7)$$

If we denote the present contributions to  $\Omega$  from *matter* (including cold dark matter as well as a contribution  $\Omega_b$  from ordinary matter of protons and neutrons, or “baryons”), *vacuum density* (cosmological constant), and *radiation*, with  $\Omega_m$ ,  $\Omega_\Lambda$ , and  $\Omega_r$ , respectively, a flat universe with  $E = 0$  satisfies

$$\frac{H(t)}{H_0} = \left[ \frac{\Omega_m}{a^3} + \Omega_\Lambda + \frac{\Omega_r}{a^4} \right]^{1/2}, \quad (1.8)$$

where we define  $H_0$  and  $\Omega_0 = (\Omega_m + \Omega_\Lambda + \Omega_r) = 1$  to be the present-day values of  $H$  and  $\Omega$ , respectively.

In the particularly simple case of a flat Universe, we find that if matter dominates (i.e.,  $\Omega_0 = 1$ ), then  $a \propto t^{2/3}$ ; if radiation dominates, then  $a \propto t^{1/2}$ ;

and if the vacuum density dominates, then  $a \propto \exp\{H_{\text{vac}}t\}$ , where  $H_{\text{vac}} = (8\pi G\rho_{\text{vac}}/3)^{1/2}$  is a constant. After inflation ended, the mass density of our Universe,  $\rho$ , was at first dominated by radiation at redshifts  $z > 3,300$ , by matter at  $0.3 < z < 3,300$ , and finally by the vacuum at  $z < 0.3$ . The vacuum had already started to dominate  $\rho_{\text{grav}}$  at  $z < 0.7$ , or 6 billion years ago. Figure 1.6 illustrates the mass budget in the present-day Universe and during the epoch when the first galaxies formed.

The preceding results for  $a(t)$  have two interesting implications. First, we can calculate the relationship between the time since the Big Bang and redshift, since  $a = (1+z)^{-1}$ . For example, during the matter-dominated era ( $1 < z < 10^3$ , with the low- $z$  end set by the condition  $[1+z] \gg [\Omega_{\Lambda}/\Omega_m]^{1/3}$ ),

$$t \approx \frac{2}{3H_0\Omega_m^{1/2}(1+z)^{3/2}} = \frac{0.95 \times 10^9 \text{ yr}}{[(1+z)/7]^{3/2}}. \quad (1.9)$$


In this same regime, where  $\Omega_m \approx 1$ ,  $H \approx 2/(3t)$ , and  $a = (1+z)^{-1} \approx (3H_0\sqrt{\Omega_m}/2)^{2/3}t^{2/3}$ .

Second, we note the remarkable exponential expansion for a vacuum-dominated phase. This accelerated expansion serves an important purpose in explaining a few puzzling features of our Universe. We have already noticed that our Universe was prepared in a very special initial state: nearly isotropic and homogeneous, with  $\Omega$  close to unity and a flat geometry. In fact, it took the CMB photons nearly the entire age of the Universe to travel toward us. Therefore, it should take them twice as long to bridge their points of origin on opposite sides of the sky. *How is it possible then that the conditions of the Universe (as reflected in the nearly uniform CMB temperature) were prepared to be the same in regions that were never in causal contact before?* Such a degree of organization is highly unlikely to occur at random. If we receive our clothes ironed and folded neatly, we know that there must have been a process that caused this to happen. Cosmologists have identified an analogous “ironing process” in the form of cosmic inflation. This process is associated with an early period during which the Universe was dominated temporarily by the mass density of an elevated vacuum state and experienced exponential expansion by at least  $\sim 60$   $e$ -folds. This vast expansion “ironed out” any initial curvature of our environment and generated a flat geometry and nearly uniform conditions across a region far greater than our current horizon. After the elevated vacuum state decayed, the Universe became dominated by radiation.

The early epoch of inflation was important not just in producing the global properties of the Universe but also in generating the inhomogeneities that seeded the formation of galaxies within it. The vacuum energy density that had driven inflation encountered quantum-mechanical fluctuations. After the perturbations were stretched beyond the horizon of the infant Universe (which today would have occupied a size no bigger than a human hand), they materialized as perturbations in the mass density of radiation and matter. The last perturbations to leave the horizon during inflation eventually reentered after inflation ended (when the scale factor grew more slowly than  $ct$ ). It is tantalizing to contemplate the notion that galaxies, which represent massive classical

objects with  $\sim 10^{67}$  atoms in today's Universe, might have originated from sub-atomic quantum-mechanical fluctuations at early times.

After inflation, an unknown process, called "baryogenesis" or "leptogenesis," generated an excess of particles (baryons and leptons) over antiparticles.<sup>v</sup> As the Universe cooled to a temperature of hundreds of millions of electron-volts (where  $1 \text{ MeV}/k_B = 1.1604 \times 10^{10} \text{ K}$ ), protons and neutrons condensed out of the primordial quark–gluon plasma through the so-called quantum chromodynamics (QCD) phase transition. At about one second after the Big Bang, the temperature declined to  $\sim 1 \text{ MeV}$ , and the weakly interacting neutrinos decoupled. Shortly afterward, the abundance of neutrons relative to protons froze, and electrons and positrons annihilated one another. In the next few minutes, nuclear fusion reactions produced light elements more massive than hydrogen, such as deuterium, helium, and lithium, in abundances that match those observed today in regions where gas has not been processed subsequently through stellar interiors. Although the transition to matter domination occurred at a redshift  $z \sim 3,300$ , the Universe remained hot enough for the gas to be ionized, and electron–photon scattering effectively coupled ordinary matter and radiation. At  $z \sim 1,100$  the temperature dipped below  $\sim 3,000 \text{ K}$ , and free electrons recombined with protons to form neutral hydrogen atoms. As soon as the dense fog of free electrons was depleted, the Universe became transparent to the relic radiation, which is observed at present as the CMB. These milestones of the thermal history are depicted in Figure 1.4.

The Big Bang is the only known event in our past history in which particles interacted with center-of-mass energies approaching the so-called Planck scale<sup>vi</sup> [ $(hc^5/G)^{1/2} \sim 10^{19} \text{ GeV}$ ], at which quantum mechanics and gravity are expected to be unified. Unfortunately, the exponential expansion of the Universe during inflation erased memory of earlier cosmic epochs, such as the Planck time. 

**Many fossils of big bang turbulent combustion persist in the CMB.**

### 1.3.1 Luminosity and Angular-Diameter Distances

When we look at our image reflected off a mirror at a distance of 1 m, we see the way we looked 6 nanoseconds ago, the time it took light to travel to the mirror and back. If the mirror is spaced  $10^{19} \text{ cm} = 3 \text{ pc}$  away, we will see the way we looked 21 years ago. Light propagates at a finite speed, so by observing distant regions, we are able to see what the Universe looked like in the past, a light-travel time ago (see Figure 1.3). The statistical homogeneity of the Universe on large scales guarantees that what we see far away is a fair statistical representation of the conditions that were present in our region of the Universe a long time ago.

This fortunate situation makes cosmology an empirical science. We do not need to guess how the Universe evolved. By using telescopes we can simply see

<sup>v</sup>The origin of the asymmetry in the cosmic abundance of matter over antimatter is still an unresolved puzzle.

<sup>vi</sup>The Planck energy scale is obtained by equating the quantum-mechanical wavelength of a relativistic particle with energy  $E$ , namely,  $hc/E$ , to its "black hole" radius,  $\sim GE/c^4$ , and solving for  $E$ .



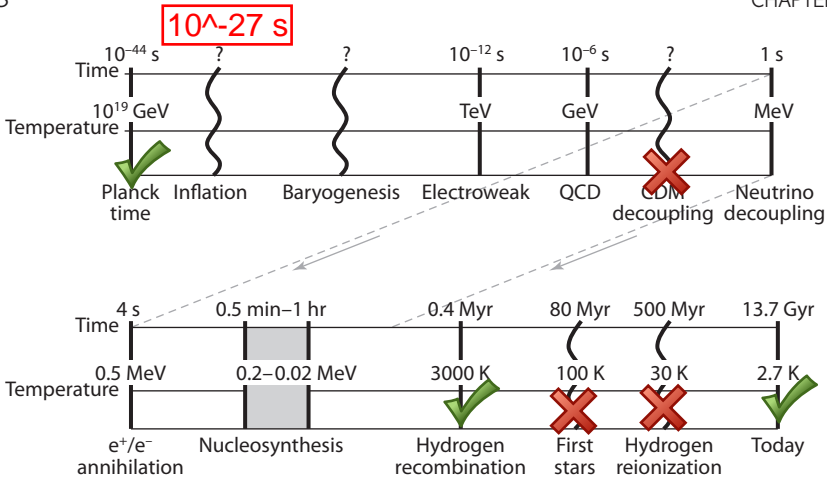


Figure 1.4 Following inflation, the Universe went through several other milestones that left a detectable record. These include baryogenesis (which resulted in the observed asymmetry between matter and antimatter), the electroweak phase transition (during which the symmetry between electromagnetic and weak interactions was broken), the QCD phase transition (during which protons and neutrons nucleated out of a soup of quarks and gluons), the dark matter decoupling epoch (during which the dark matter decoupled thermally from the cosmic plasma), neutrino decoupling, electron–positron annihilation, light-element nucleosynthesis (during which helium, deuterium, and lithium were synthesized), and hydrogen recombination. The cosmic time and CMB temperature of the various milestones are marked. Wavy lines and question marks indicate milestones with uncertain properties. The signatures that the same milestones left in the Universe are used to constrain its parameters.

the way distant regions appeared at earlier cosmic times. Since a greater distance means a fainter flux from a source of a fixed luminosity, the observation of the earliest sources of light requires the development of sensitive instruments and poses technological challenges to observers.

*How faint will the earliest galaxies appear to our telescopes?* In an expanding Universe there is some ambiguity as to which “distance” is most relevant. For example, the framework we described earlier—in which the clocks are synchronized relative to the Big Bang—is not appropriate for observations, because light has a finite speed, so that a signal emitted from one clock at time  $t_A$  will be observed by another clock at a time  $t_B > t_A$ . Which of these times should we use to compute the scale factor in a distance formula? Moreover, the *method* of observation influences the choice of the relevant distance, because the photons themselves evolve as they travel.

To answer these questions, we can easily express the flux observed from a galaxy of luminosity  $L$  at a redshift  $z$ . The observed flux (energy per unit time per unit telescope area) is obtained by spreading the energy emitted from the

source per unit time,  $L$ , over the surface area of a sphere whose radius equals the effective distance of the source,

$$f = \frac{L}{4\pi d_L^2}, \quad (1.10)$$

where  $d_L$  is defined as the *luminosity distance* in cosmology. For a flat Universe, the comoving distance of a galaxy that emitted its photons at a time  $t_{\text{em}}$  and is observed at time  $t_{\text{obs}}$  is obtained by summing over infinitesimal distance elements along the path length of a photon,  $c dt$ , each expanded by a factor  $(1+z)$  to the present time (corresponding to setting  $ds^2 = 0$  in equation 1.1 for a photon trajectory):

$$R_{\text{em}} = \int_{t_{\text{em}}}^{t_{\text{obs}}} \frac{c dt}{a(t)} = \frac{c}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z')^3 + \Omega_\Lambda}}, \quad (1.11)$$

where  $a = (1+z)^{-1}$ . The *angular-diameter distance*  $d_A$ , corresponding to the angular diameter  $\theta = D/d_A$  occupied by a galaxy of size  $D$ , must take into account the fact that we were closer to that galaxy<sup>vii</sup> by a factor  $(1+z)$  when the photons started their journey at a redshift  $z$ , so it is simply given by  $d_A = R_{\text{em}}/(1+z)$ . But to find  $d_L$  we must take account of additional redshift factors.

If a galaxy has an intrinsic luminosity  $L$ , then it will emit an energy  $L dt_{\text{em}}$  over a time interval  $dt_{\text{em}}$ . This energy is redshifted by a factor of  $(1+z)$  and is observed over a longer time interval  $dt_{\text{obs}} = dt_{\text{em}}(1+z)$  after being spread over a sphere of surface area  $4\pi R_{\text{em}}^2$ . Thus, the observed flux will be

$$f = \frac{L dt_{\text{em}}/(1+z)}{4\pi R_{\text{em}}^2 dt_{\text{obs}}} = \frac{L}{4\pi R_{\text{em}}^2 (1+z)^2}, \quad (1.12)$$

which implies that

$$d_L = R_{\text{em}}(1+z) = d_A(1+z)^2. \quad (1.13)$$

Unfortunately, for a flat universe with a cosmological constant, these distance integrals cannot be expressed analytically. However, a convenient numerical approximation, valid to 0.4% relative error in the range  $0.2 \leq \Omega_m \leq 1$  (where  $\Omega_m$  is the total matter density) is<sup>1</sup>

$$d_L = \frac{c}{H_0} a^{-1} [\eta(1, \Omega_m) - \eta(a, \Omega_m)], \quad (1.14)$$

where

$$\eta(a, \Omega_m) = 2\sqrt{s^3+1} \left[ \frac{1}{a^4} - 0.1540 \frac{s}{a^3} + 0.4304 \frac{s^2}{a^2} + 0.19097 \frac{s^3}{a} + 0.066941s^4 \right]^{-1/8}, \quad (1.15)$$

and  $s^3 = 1/\Omega_m - 1$ .

<sup>vii</sup>In a flat Universe, photons travel along straight lines. The angle at which a photon is seen is not modified by the cosmic expansion, since the Universe expands at the same rate both parallel and perpendicular to the line of sight.

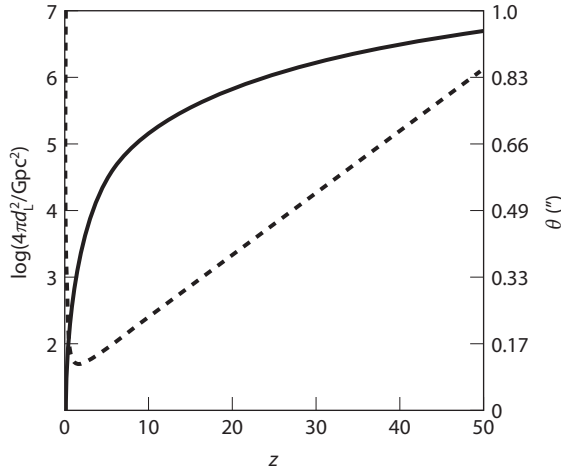


Figure 1.5 The solid line (corresponding to the label on the left-hand side) shows  $\log_{10}$  of the conversion factor between the luminosity of a source and its observed flux,  $4\pi d_L^2$  (in  $\text{Gpc}^2$ ), as a function of redshift,  $z$ . The dashed line (labeled on the right) gives the angle  $\theta$  (in arcseconds) occupied by a galaxy of 1 kpc diameter as a function of redshift.

The area dilution factor  $4\pi d_L^2$  is plotted as a function of redshift in the solid curve of Figure 1.5. If the observed flux is measured over only a narrow band of frequencies, one needs to take account of the additional conversion factor  $(1+z) = (dv_{\text{em}}/dv_{\text{obs}})$  between the emitted frequency interval  $dv_{\text{em}}$  and its observed value  $dv_{\text{obs}}$ . This correction yields the relation  $(df/dv_{\text{obs}}) = (1+z) \times (dL/dv_{\text{em}})/(4\pi d_L^2)$ .

In practice, observed brightnesses are often expressed using the *AB magnitude* system. The conversion from flux density to AB magnitude is

$$\text{AB} = -2.5 \log \left[ \frac{df}{dv_{\text{obs}}} \right] - 48.6, \quad (1.16)$$

where the flux density is expressed in units of  $\text{erg s}^{-1} \text{cm}^{-2} \text{Hz}^{-1}$ .

## 1.4 Most Matter Is Dark

Surprisingly, most of the matter in the Universe is not the same ordinary matter of which we are made (see Figure 1.6). If it were ordinary matter (which also makes stars and diffuse gas), it would have interacted with light, thereby revealing its existence to observations through telescopes. Instead, observations of many different astrophysical environments require the existence of some mysterious dark component of matter that reveals itself only through its gravitational influence and leaves no other clue about its nature. Cosmologists are like detectives who find evidence for some unknown criminal at a crime scene

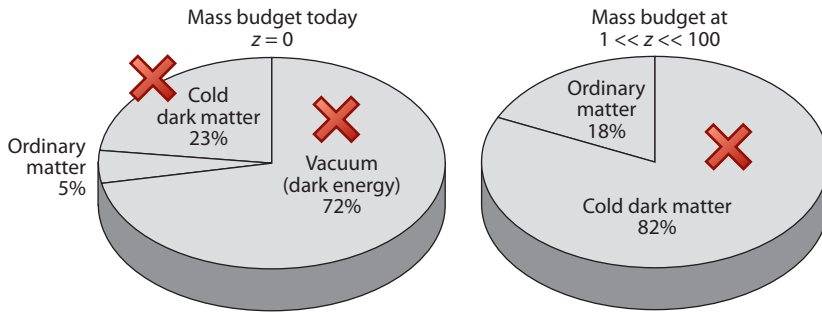


Figure 1.6 Mass budgets of different components in the present-day Universe and in the infant Universe when the first galaxies formed (redshifts  $z = 10\text{--}50$ ). The CMB radiation (not shown) makes up a fraction ( $\sim 0.03\%$ ) of the budget today but was dominant at redshifts  $z > 3,300$ . The cosmological constant (vacuum) contribution was negligible at high redshifts ( $z \gg 1$ ).

and are anxious to find his or her identity. The evidence for dark matter is clear and indisputable, assuming that the laws of gravity are not modified (although a small minority of scientists are exploring this alternative). ✓

Without dark matter we would never have existed by now, because ordinary matter is coupled to the CMB radiation that filled the early Universe. The diffusion of photons on small scales smoothed out perturbations in this primordial radiation fluid. The smoothing length was stretched to a scale as large as hundreds of millions of light-years in the present-day Universe. This is a huge scale by local standards, since galaxies—like the Milky Way—were assembled out of matter in regions a hundred times smaller than that. Because ordinary matter was coupled strongly to the radiation in the early dense phase of the Universe, it also was smoothed on small scales. If there were nothing else in addition to the radiation and ordinary matter, then this smoothing process would have had a devastating effect on the prospects for life in our Universe. Galaxies like the Milky Way would never have formed by the present time, since there would have been no density perturbations on the relevant small scales to seed their formation. The existence of dark matter not coupled to the radiation came to the rescue by remembering the initial seeds of density perturbations on small scales. In our neighborhood, these seed perturbations led eventually to the formation of the Milky Way galaxy inside which the Sun was made as one out of tens of billions of stars, and Earth was born out of the debris left over from the formation process of the Sun. The dark matter of all galaxies is earth mass gas planets: Schild (1996).

We do not know what constitutes the dark matter, but from the good match obtained between observations of large-scale structure and the equations describing a pressureless fluid (see equations 2.3–2.4), we infer that it is likely made of particles with small random velocities. It is therefore called “cold dark matter” (CDM). The popular view is that CDM is composed of particles that weakly interact with ordinary matter, much like the elusive neutrinos we know

to exist. The abundance of such particles would naturally “freeze out” at a temperature  $T > 1$  MeV, at which the Hubble expansion rate is comparable to the annihilation rate of the CDM particles. Interestingly, such a decoupling temperature, together with a weak interaction cross section and particle masses of  $mc^2 > 100$  GeV (as expected for the lightest, and hence stable, supersymmetric particle in simple extensions of the standard model of particle physics), naturally leads through a Boltzmann suppression factor  $\sim \exp(-mc^2/k_B T)$  to  $\Omega_m \sim 1$ . The hope is that CDM particles, owing to their weak but nonvanishing coupling to ordinary matter, will nevertheless be produced in small quantities through collisions of energetic particles in future laboratory experiments such as the Large Hadron Collider (LHC). Other experiments are attempting to detect directly the astrophysical CDM particles in the Milky Way halo. A positive result from any of these experiments will be equivalent to our detective friend’s being successful in finding a DNA sample of the previously unidentified criminal.

The most popular candidate for the CDM particle is a weakly interacting massive particle (WIMP). The lightest supersymmetric particle (LSP) could be a WIMP. The CDM particle mass depends on free parameters in the particle physics model; the LSP hypothesis will be tested at the Large Hadron Collider or in direct detection experiments. The properties of the CDM particles affect their response to the primordial inhomogeneities on small scales. The particle cross section for scattering off standard model particles sets the epoch of their thermal decoupling from the cosmic plasma.

In addition to dark matter, the observed acceleration in the current expansion rate of the Universe implies that the vacuum contributes  $\sim 72\%$  of the cosmic mass density at present. If the vacuum density will behave as a cosmological constant, it will dominate even more in the future (since  $\rho_m/\rho_v \propto a^{-3}$ ). The exponential future expansion will carry all galaxies outside the local group out of our horizon within  $\sim 10^{11}$  years,<sup>2</sup> and will stretch the characteristic wavelength of the cosmic microwave background to be larger than the horizon in  $\sim 10^{12}$  years.<sup>3</sup>

The dark ingredients of the Universe can be probed only indirectly through a variety of luminous tracers. The distribution and nature of the dark matter are constrained by detailed X-ray and optical observations of galaxies and galaxy clusters. The evolution of the dark energy with cosmic time will be constrained over the coming decade by surveys of Type Ia supernovae, as well as surveys of X-ray clusters, up to a redshift of 2.

According to the standard cosmological model, the CDM behaves as a collection of collisionless particles that started out at the epoch of matter domination with negligible thermal velocities and later evolved exclusively under gravitational forces. The model explains how both individual galaxies and the large-scale patterns in their distribution originated from the small initial density fluctuations. On the largest scales, observations of the present galaxy distribution have indeed found the same statistical patterns as seen in the CMB, enhanced as expected by billions of years of gravitational evolution. On smaller scales, the model describes how regions that were denser than average

Table 1.1 Standard Set of Cosmological Parameters (defined and adopted throughout the book). Based on Komatsu, E., et al., *Astrophys. J. Suppl.* **180**, 330 (2009).

$\Omega_\Lambda$	$\Omega_m$	$\Omega_b$	$h$	$n_s$	$\sigma_8$
0.72	0.28	0.05	0.7	1	0.82

collapsed owing to their enhanced gravity and eventually formed gravitationally bound halos, first on small spatial scales and later on larger ones. In this hierarchical model of galaxy formation, the small galaxies formed first and then merged, or accreted gas, to form larger galaxies. At each snapshot of this cosmic evolution, the abundance of collapsed halos, whose masses are dominated by dark matter, can be computed from the initial conditions. The common understanding of galaxy formation is based on the notion that stars formed out of the gas that cooled and subsequently condensed to high densities in the cores of some of these halos.

Gravity thus explains how some gas is pulled into the deep potential wells within dark matter halos and forms galaxies. One might naively expect that the gas outside halos would remain mostly undisturbed. However, observations show that it did not remain neutral (i.e., in atomic form) but was largely ionized by the UV radiation emitted by the galaxies. The diffuse gas pervading the space outside and between galaxies is referred to as the *intergalactic medium* (IGM). For the first hundreds of millions of years after cosmological recombination (when protons and electrons combined to make neutral hydrogen), the so-called cosmic dark ages, the universe was filled with diffuse atomic hydrogen. As soon as galaxies formed, they started to ionize diffuse hydrogen in their vicinity. Within less than a billion years, most of the IGM was reionized. This *reionization epoch* marks a crucial transition in the history of the Universe and is a prime focus of both modern astrophysics research and this book.

The initial conditions of the Universe can be summarized on a single sheet of paper. The small number of parameters that provide an accurate statistical description of these initial conditions are summarized in Table 1.1 (see also Appendix B). However, thousands of books in libraries throughout the world cannot summarize the complexities of galaxies, stars, planets, life, and intelligent life in the present-day Universe. If we feed the simple initial cosmic conditions into a gigantic computer simulation incorporating the known laws of physics, we should be able to reproduce all the complexity that emerged out of the simple early Universe. Hence, all the information associated with this later complexity was encapsulated in those simple initial conditions. We will follow the process through which late-time complexity appeared and established an irreversible arrow to the flow of cosmic time.<sup>viii</sup>

<sup>viii</sup>In previous decades, astronomers used to associate the simplicity of the early Universe with the fact that the data about it were scarce. Although this was true at the infancy of observational cosmology, it is not true any more. With much richer data in our hands, the initial simplicity is now interpreted as an outcome of inflation.

The basic question that cosmology attempts to answer is: *What is the composition of the Universe and what initial conditions generated the observed structures in it?* The first galaxies were shaped, more than any other class of astrophysical objects, by the pristine initial conditions and basic constituents of the Universe. Studying the formation process of the first galaxies could reveal unique evidence for new physics that has so far remained veiled in older galaxies by complex astrophysical processes.

## Chapter Two

---

### Linear Growth of Cosmological Perturbations

After cosmological recombination, the Universe entered the “dark ages,” during which the relic CMB light from the Big Bang gradually faded away. During this “pregnancy” period (which lasted hundreds of millions of years), the seeds of small density fluctuations planted by inflation in the matter distribution grew until they eventually collapsed to make the first galaxies. Here we describe the first stages of that process and introduce the methods conventionally used to describe these fluctuations.

#### 2.1 Growth of Linear Perturbations

As discussed earlier, small perturbations in density grow owing to the unstable nature of gravity. Overdense regions behave as if they reside in a closed Universe. Their evolution ends in a “big crunch,” which results in the formation of gravitationally bound objects like the Milky Way galaxy.

Equation (1.6) explains the formation of galaxies out of seed density fluctuations in the early Universe, at a time when the mean matter density was very close to the critical value  $\Omega_m \approx 1$ . Given that the mean cosmic density was close to the threshold for collapse, a spherical region that was only slightly denser than the mean behaved as if it was part of an  $\Omega > 1$  Universe and therefore eventually collapsed to make a bound object, like a galaxy. The material from which objects are made originated in the underdense regions (voids) that separate these objects (and which behaved as part of an  $\Omega < 1$  Universe), as illustrated in Figure 1.2.

Observations of the CMB show that at the time of hydrogen recombination the Universe was extremely uniform, with spatial fluctuations in the energy density and gravitational potential of roughly one part in  $10^5$ . These small fluctuations grew over time during the matter-dominated era as a result of gravitational instability and eventually led to the formation of galaxies and larger-scale structures, observed today.

In describing the gravitational growth of perturbations in the matter-dominated era ( $z \ll 3,300$ ), we may consider small perturbations of a fractional amplitude  $|\delta| \ll 1$  on top of the uniform background density  $\bar{\rho}$  of cold dark matter. The three fundamental equations describing conservation of mass and momentum along with the gravitational potential can then be expanded to leading order in the perturbation amplitude. We distinguish between physical (or proper) and comoving coordinates (the latter expand with the background



Universe). In vector notation, the fixed coordinate  $\mathbf{r}$  corresponds to a comoving position  $\mathbf{x} = \mathbf{r}/a$ . We describe the cosmological expansion in terms of an ideal pressureless fluid of particles, each of which is at fixed  $\mathbf{x}$ , expanding with the Hubble flow  $\mathbf{v} = H(t)\mathbf{r}$ , where  $\mathbf{v} = d\mathbf{r}/dt$ .

Onto this uniform expansion we impose small fractional density perturbations

$$\delta(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\bar{\rho}} - 1, \quad (2.1)$$

where the mean fluid mass density is  $\bar{\rho}$ , with a corresponding peculiar velocity that describes the deviation from the Hubble flow  $\mathbf{u} \equiv \mathbf{v} - H\mathbf{r}$ . The fluid is then described by the continuity and Euler equations. In comoving coordinates, where the bulk velocity vanishes, we have

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot [(1 + \delta)\mathbf{u}] = 0, \quad (2.2)$$

$$\frac{\partial \mathbf{u}}{\partial t} + H\mathbf{u} + \frac{1}{a}(\mathbf{u} \cdot \nabla)\mathbf{u} = -\frac{1}{a}\nabla\phi - \frac{1}{a\bar{\rho}}\nabla(\delta p). \quad (2.3)$$


The gravitational potential  $\phi$  is given by the Newtonian Poisson equation in terms of the density perturbation:

$$\nabla^2\phi = 4\pi G\bar{\rho}a^2\delta. \quad (2.4)$$

The pressure  $p$  depends on the species under consideration. For cold dark matter, it vanishes; for an ideal gas of baryons at a fixed temperature, the pressure perturbation is  $(\delta p) = c_s^2\delta\bar{\rho}$ . The sound speed for a monatomic gas that obeys the ideal gas equation of state  $p = nkT_e$  and undergoes Hubble expansion is

$$c_s^2 = \frac{dp/da}{d\rho/da} = \frac{k_B T_e}{\mu m_p} \left( 1 - \frac{1}{3} \frac{d \log T_e}{d \log a} \right), \quad (2.5)$$

where  $T_e$  is the gas kinetic temperature, and  $\mu$  is the mean molecular mass in units of  $m_p$ . (For primordial neutral gas including a mass fraction  $Y_p = 0.24$  of helium,  $\mu = 1.22$ .) In this section we adopt this expression for the sound speed, though we note that it assumes that the temperature traces the density field (see §2.2.1 for a more exact treatment).

 This fluid description is valid for describing the evolution of collisionless cold dark matter particles until different particle streams cross. The crossing typically occurs only after perturbations have grown to become nonlinear with  $|\delta| > 1$ , and at that point the individual particle trajectories must in general be followed.

The combination of the preceding equations yields, to leading order in  $\delta$ ,

$$\frac{\partial^2 \delta}{\partial t^2} + 2H \frac{\partial \delta}{\partial t} = 4\pi G\bar{\rho}\delta - \frac{c_s^2 k^2}{a^2} \delta, \quad (2.6)$$

where the last term is the pressure force, which vanishes for cold dark matter. In general, this linear equation has two independent solutions, only one of which grows in time. From random initial conditions this “growing mode”

comes to dominate the density evolution. Thus, until it becomes nonlinear, the density perturbation maintains its shape in comoving coordinates and grows in amplitude in proportion to a growth factor  $D(t)$ . The growth factor in a flat (matter-dominated) Universe at  $z < 10^3$  is given by<sup>1</sup>

$$D(t) \propto \frac{(\Omega_\Lambda a^3 + \Omega_m)^{1/2}}{a^{3/2}} \int_0^a \frac{a'^{3/2} da'}{(\Omega_\Lambda a'^3 + \Omega_m)^{3/2}}. \quad (2.7)$$

In the matter-dominated regime of the redshift range  $1 < z < 10^3$ , the growth factor is simply proportional to the scale factor  $a(t)$ . The normalization is usually chosen to be relative to the perturbation amplitude at the present day; we will discuss how to determine this factor §2.1.3.

In a flat Universe with a cosmological constant, this integral cannot be written in closed form without special functions. However, an approximation accurate to  $\sim 2\%$  in the range  $\Omega_m > 0.1$  is  $D(z) = \mathcal{D}(z)/(1+z)$  with<sup>2</sup>

$$\mathcal{D}(z) = \frac{5\Omega_m(z)}{2} [\Omega_m(z)^{4/7} - \Omega_\Lambda(z) + (1 + \Omega_m(z)/2)(1 + \Omega_\Lambda(z)/70)]^{-1}, \quad (2.8)$$

where (if  $\Omega_m + \Omega_\Lambda = 1$ )

$$\Omega_m(z) = \frac{\Omega_m(1+z)^3}{\Omega_m(1+z)^3 + \Omega_\Lambda}, \quad (2.9)$$

$$\Omega_\Lambda(z) = \frac{\Omega_\Lambda}{\Omega_m(1+z)^3 + \Omega_\Lambda}, \quad (2.10)$$

and  $\Omega_\Lambda$  is the present-day energy density in a cosmological constant scaled to the critical density. Here  $\mathcal{D}(z)$  is normalized to equal unity in a matter-dominated Universe. At the high redshifts of most interest to us, this is a reasonable approximation.

Interestingly, in this matter-dominated regime the gravitational potential  $\phi \propto \delta/a$  does not grow in comoving coordinates, which implies that the potential depth of fluctuations remains frozen in amplitude as fossil relics from the inflationary epoch during which they were generated. Nonlinear collapse changes the potential depth only by a factor of the order of unity, but even inside collapsed objects its rough magnitude remains as testimony to the inflationary conditions. This explains why the characteristic potential depth of collapsed objects such as galaxy clusters ( $\phi/c^2 \sim 10^{-5}$ ) is of the same order as the potential fluctuations probed by the fractional variations in the CMB temperature across the sky. At low redshifts  $z < 1$  and in the future, the cosmological constant dominates ( $\Omega_m \ll \Omega_\Lambda$ ), and the density fluctuations freeze in amplitude [ $D(t) \rightarrow \text{constant}$ ] as their growth is suppressed by the accelerated expansion of space.

It is usually convenient to express the density field as a sum over a complete set of periodic Fourier modes, each with a sinusoidal (wavelike) dependence on space with a comoving wavelength  $\lambda = 2\pi/k$  and wavenumber  $k$ .

Mathematically, we write<sup>i</sup>

$$\delta_{\mathbf{k}} = \int d^3x \delta(x) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (2.11)$$

$$\delta(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} \delta_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}}, \quad (2.12)$$

where  $\mathbf{x}$  is the comoving spatial coordinate. The characteristic amplitude of each  $\mathbf{k}$ -mode defines the typical value of  $\delta$  on the spatial scale  $\lambda$ . It is straightforward to show that equation (2.6) applies to each Fourier mode individually, so the factor  $D(t)$  also describes their growth (in the linear regime), and the evolution of the density field in Fourier space is easy to follow. In particular, note that different spatial scales evolve *independently* in the linear regime.

It is also useful to consider the velocity field  $\mathbf{u}$ . To linear order, the continuity equation (2.2) becomes  $\nabla \cdot \mathbf{u} = -a(d\delta/dt)$ , or in Fourier space

$$-i\mathbf{k} \cdot \mathbf{u}_{\mathbf{k}} = -\frac{a}{D} \frac{dD}{dt} \delta_{\mathbf{k}}, \quad (2.13)$$

where we have assumed that  $\delta_{\mathbf{k}}$  is a pure growing mode. This equation has the solution

$$\mathbf{u}_{\mathbf{k}} = -i \frac{aHf(\Omega)}{k} \delta_{\mathbf{k}} \hat{\mathbf{k}}, \quad (2.14)$$

where  $f(\Omega) = (a/D)(dD/da) \approx \Omega_m^{0.6}$  to a very good approximation (note that it is almost independent of  $\Omega_\Lambda$ ). Interestingly, peculiar velocity perturbations grow proportionally to density fluctuations, and their growing modes are parallel to the wavevector. Note also that  $\mathbf{u}_{\mathbf{k}} \propto \delta_{\mathbf{k}}/k$ , which implies that peculiar velocities on a given scale are sourced by gravitational fluctuations on *larger* scales than those of the density field.

### 2.1.1 The Power Spectrum of Density Fluctuations

The initial perturbation amplitude varies with spatial scale; typically, large-scale regions have a smaller perturbation amplitude than do small-scale regions. The statistical properties of the perturbations as a function of spatial scale can best be captured by their Fourier transforms in comoving wavenumbers. This approach has the convenient property that the spatial scales are *fixed* in time rather than evolving as the perturbation expands or collapses.

Because we cannot observe how particular regions mature and grow over time, we are typically concerned not with the amplitude of individual density perturbations or modes but with the properties of their statistical ensemble. Most often, two complementary statistical measures are used. The first is the *correlation function*,

$$\xi(\mathbf{x}) = \langle \delta(\mathbf{x})\delta(0) \rangle, \quad (2.15)$$

<sup>i</sup>Note that cosmologists typically absorb the volume factors in the Fourier transform into  $\delta_{\mathbf{k}}$ , which has units of volume.

where the angular brackets represent averaging over the entire statistical ensemble of points separated by a comoving distance  $\mathbf{x}$ , and where we made use of the translational invariance of statistical averages in centering our coordinate system on the second point. The correlation function expresses the degree to which a particular overdensity is more likely to be surrounded by other overdense regions. Note that for an isotropic distribution of perturbations,  $\xi$  is a function only of the magnitude of the spatial separation,  $x = |\mathbf{x}|$ .

The second measure is the *power spectrum*,  $P(\mathbf{k})$ , defined by

$$P(\mathbf{k}) = \langle \delta_{\mathbf{k}} \delta_{\mathbf{k}'}^* \rangle = (2\pi)^3 \delta^D(\mathbf{k} - \mathbf{k}') P(\mathbf{k}), \quad (2.16)$$

which has units of volume. This is simply related to the variance of the amplitude of waves on a given scale. Again, it is a function only of  $k = |\mathbf{k}|$  for an isotropic universe.

In fact, the correlation function and power spectrum are intimately related. If we write the former using the Fourier transform of  $\delta(\mathbf{x})$ , we obtain

$$\xi(\mathbf{x}) = \left\langle \int \frac{d^3k}{(2\pi)^3} \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \int \frac{d^3k'}{(2\pi)^3} \delta_{\mathbf{k}'}^* \right\rangle \quad (2.17)$$

$$= \int \frac{d^3k}{(2\pi)^3} \int \frac{d^3k'}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}} \langle \delta_{\mathbf{k}} \delta_{\mathbf{k}'}^* \rangle \quad (2.18)$$

$$= \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}} P(\mathbf{k}), \quad (2.19)$$

where in the first line we have used the fact that  $\delta(0)$  is real. Thus  $\xi(r)$  and  $P(k)$  are simply Fourier transforms of each other. Theoretical calculations are generally simplest using the Fourier representation and power spectrum, but the two approaches have different error properties, so both are used regularly in the literature.

Inflation generates perturbations in which different  $\mathbf{k}$ -modes are statistically independent, and each has a random phase constant in its sinusoid. This makes the density field following inflation a *Gaussian random field*, and its statistical properties are perfectly described by the power spectrum (see §2.1.3). In other words, all higher-order moments and correlations are simply functions of the power spectrum (or correlation function): no additional parameters are needed to understand the distribution, at least until nonlinear evolution becomes important (which does induce higher-order correlations purely through gravitational instability). A very small amount of primordial non-Gaussianity can be accommodated by existing observations; the nonlinear phase of gravitational collapse generates more.

Moreover, in the standard cosmological model, inflation produces a very simple primordial power-law spectrum  $P(k) \propto k^{n_s}$  with  $n_s \approx 1$ . Quantum fluctuations during cosmic inflation naturally result in a nearly scale-invariant spectrum because of the near constancy of the Hubble parameter for a nearly steady vacuum density. This spectrum has the special property that gravitational potential fluctuations of all wavelengths have the same amplitude at the time when they enter the horizon (namely, when their wavelength matches the

distance traveled by light during the age of the Universe), so this spectrum is called *scale invariant*. This is easy to see: the mean square amplitude of mass fluctuations within spheres of comoving radius  $\ell$  is  $(\delta M/M)^2 \propto k^3 P(k)$  for  $k \sim 2\pi/\ell$ . Therefore, the corresponding fluctuation amplitude of the gravitational potential,  $\sim (G\delta M/\ell) \propto \ell^{(1-n_s)/2}$ , is independent of scale if  $n_s = 1$ . This spectrum has the aesthetic appeal that perturbations can always be small on the horizon scale. A different power-law spectrum would lead to an overdensity of the order of unity across the horizon, and result in black hole formation, either in the future or past of the Universe.

However, the power spectrum becomes more complex as perturbations grow at later times in a CDM universe. In particular, the modified final power spectrum is characterized by a turnover at a scale on the order of the horizon  $cH^{-1}$  at matter-radiation equality, and a small-scale asymptotic shape of  $P(k) \propto k^{n_s-4}$ . The turnover results from the fact that density perturbations experience almost no growth during the radiation-dominated era, because the Jeans length at that time ( $\sim ct/\sqrt{3}$ ; see the next chapter) is comparable to the scale of the horizon, inside of which growth is enabled by causality. Therefore, modes on a spatial scale that entered the horizon during the early radiation-dominated era got trapped at their initial small density contrast and so show a smaller amplitude relative to the power-law extrapolation of long-wavelength modes that entered the horizon during the matter-dominated era.

For a scale-invariant index  $n_s \approx 1$ , the small-scale fluctuations have the same amplitude at horizon crossing, and with nearly no growth they have the same amplitude on all subhorizon mass scales at matter-radiation equality. The associated constancy of the fluctuation amplitude on small mass scales (in real space),  $\delta^2 \propto P(k)k^3 \sim \text{constant}$ , implies a small-scale asymptotic slope for  $P(k)$  of  $\approx -3$ . The resulting power spectrum after matter-radiation equality is often parameterized by a *transfer function* that accounts for changes in the shape of the dark matter power spectrum up to this point. The transfer function is defined so that

$$P(k, z) = T^2(k) \frac{D^2(z)}{D^2(z_{\text{eq}})} P_{\text{pri}}(k), \quad (2.20)$$

where  $P_{\text{pri}}(k)$  is the primordial power spectrum. Note that the transfer function is time independent (but scale dependent) because it describes all the evolution from inflation through the era of matter-radiation equality. The growth factor, however is scale independent (but time dependent) because dark matter perturbations do not have any scale dependence during the matter era. The transfer function is crudely described by the fitting function<sup>3</sup>

$$T^2(k) P_{\text{pri}}(k) \propto k^{n_s} / (1 + \alpha_p k + \beta_p k^2)^2, \quad (2.21)$$

with  $\alpha_p = 8(\Omega_m h^2)^{-1}$  Mpc and  $\beta_p = 4.7(\Omega_m h^2)^{-2}$  Mpc<sup>2</sup>. This function provides a reasonable fit to the overall shape of the power spectrum, but small-scale features and subtle modifications not captured by this simple formula are extremely important as well. These include the effects of neutrinos with finite mass (which wash out small-scale structure, thanks to the relativistic motions of

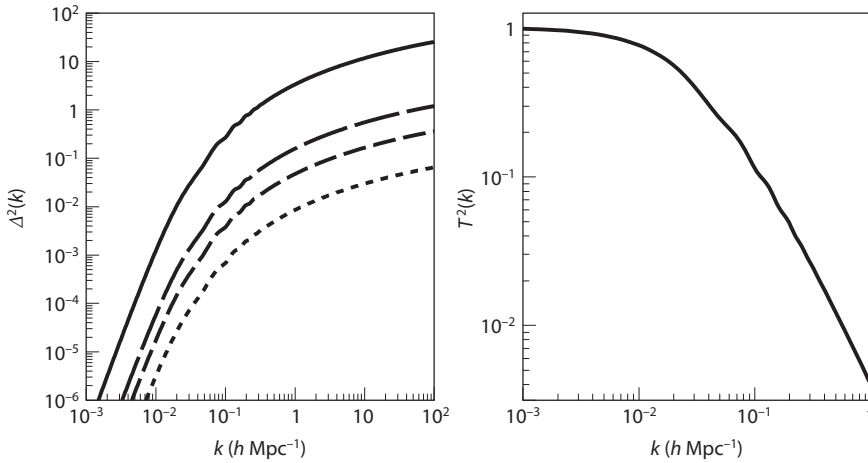


Figure 2.1 *Left*: The matter power spectrum in our fiducial cosmology at  $z = 0, 5, 10,$  and  $25,$  from top to bottom. *Right*: The corresponding transfer function. Computed using the publicly available code CAMB (<http://camb.info>).

these particles) and the influence of baryons, which we discuss in detail next.<sup>4</sup> Figure 2.1 shows the resulting matter power spectra and transfer functions at  $z = 0, 5, 10,$  and  $25,$  using our fiducial cosmological parameters. Note the oscillatory features near  $k \sim 0.1 h \text{ Mpc}^{-1},$  which are called *baryon acoustic oscillations,* whose source we discuss next.

### 2.1.2 Relative Streaming of Baryons and Cold Dark Matter ✘

Species that decouple at a particular time from the cosmic plasma (including the dark matter and the baryons) will show fossil evidence for acoustic oscillations in their power spectrum of inhomogeneities owing to sound waves in the radiation fluid to which they were coupled at early times. This phenomenon can be understood as follows. Imagine a localized point like perturbation from inflation at  $t = 0.$  The small perturbation in density or pressure will send out a sound wave that will reach the sound horizon  $c_s t$  at any later time  $t$  (see also the discussion in §1.2.3); in the radiation fluid, where  $c_s \approx c/\sqrt{3},$  this sound horizon will be near the causal horizon as well. The perturbation will therefore correlate with its surroundings up to the sound horizon, and all  $k$ -modes with wavelengths equal to this scale or its harmonics will be correlated. The result is a series of peaks in the power spectrum corresponding to the harmonics of this physical scale, as shown in Figure 2.1.

These peaks from radiation coupling to the dark matter sector are on very small spatial scales (for weakly interacting particles, they correspond to mass scales of planets or smaller).<sup>5</sup> The mass scales of the perturbations that grew to become the first collapsed objects at  $z < 100$  crossed the horizon in the radiation-dominated era after the dark matter had already decoupled from the cosmic plasma and so were largely unaffected by this streaming.

However, prior to cosmological recombination, the baryons and the cosmic background radiation were tightly coupled and behaved as a single fluid, separate from the dark matter. Because this was relatively late in the history of star formation, the physical scales of these correlations are reasonably large:  $\sim 150$  Mpc today. These large-scale features can be incorporated into the transfer function and, because their locations can be predicted from first principles for a given cosmological model, act as “standard rulers” that are useful in measuring the fundamental parameters of our Universe. The induced correlations occur on such large scales that they do not themselves appreciably affect structure formation at high redshifts.

However, a related effect is potentially very important.<sup>6</sup> When the gas decoupled from the radiation at  $z \approx 10^3$ , it was streaming relative to the dark matter with a root-mean-square (rms) speed of  $v_{bc} \approx 10^{-4}c = 30 \text{ km s}^{-1}$ . This speed is much larger than the sound speed, so it has important implications for the accretion of gas onto dark matter structures (see §3.2.2). Here we will show how these effects can be incorporated into perturbation theory to describe gravitational instability.

Using the continuity equation for the baryons and cold dark matter separately, we can write the Fourier transform of the relative velocity between the two species (to linear order) as

$$\mathbf{u}_{bc}(\mathbf{k}) = \frac{\mathbf{k}}{ik^2} [\theta_b(\mathbf{k}) - \theta_c(\mathbf{k})], \quad (2.22)$$

where  $\theta \equiv a^{-1} \nabla \cdot \mathbf{u}$ . From equation (2.14), the power spectrum of this relative velocity is then

$$\Delta_{vbc}^2(k) = \frac{k^3}{2\pi^2} P_{\text{pri}}(k) \left[ \frac{\theta_b(k) - \theta_c(k)}{k} \right]^2, \quad (2.23)$$

and the total variance is  $\langle u_{bc}^2(\mathbf{x}) \rangle = \int (dk/k) \Delta_{vbc}^2(k)$ . Figure 2.2 shows the variance of the velocity difference perturbations (in units of  $c$ ) per  $\ln k$  as a function of the mode wavenumber  $k$  at  $z = 10^3$ . The power extends to scales as large as the sound horizon at recombination,  $\sim 140$  Mpc, but declines rapidly at  $k > 0.5 \text{ Mpc}^{-1}$ , which indicates that the velocity of the baryons relative to the dark matter was coherent over the photon diffusion (or Silk damping) scale of several comoving megaparsecs. This scale is larger by two orders of magnitude than the size of the regions out of which the first galaxies were assembled at later times. Therefore, in the rest frame of those galaxies, the background intergalactic baryons appeared to be moving coherently as a wind. In the next chapter, we will examine whether this wind had a significant effect on the assembly of baryons onto the earliest galaxies.

In the presence of this relative motion between baryons and cold dark matter, the perturbation analysis becomes somewhat more complex. The simplest approach is to treat the two species as having a spatially constant bulk velocity  $v_{bc}$  that decays with redshift as  $1/a$  as the neutral gas falls into the gravitational potential wells of the dark matter (see equation 2.22). The assumption of a spatially constant background velocity is valid on scales smaller than the coherence

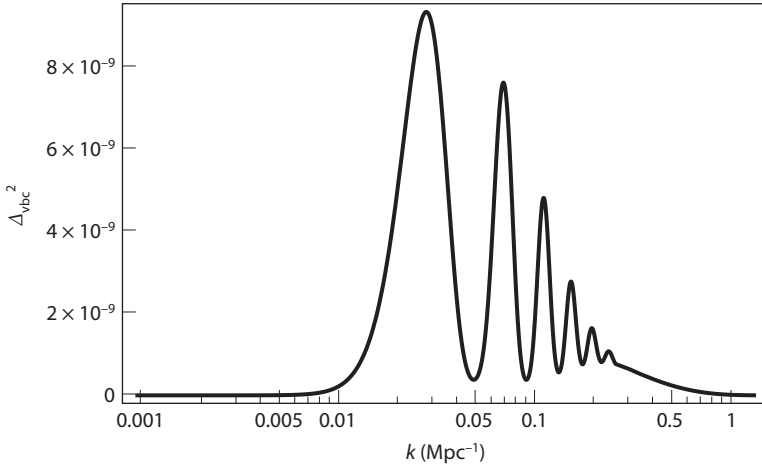


Figure 2.2 The variance of the velocity difference perturbations (in units of  $c$ ) between baryons and dark matter per  $\ln k$  as a function of comoving wavenumber  $k$  at  $z = 10^3$ . Reprinted Figure 1 with permission from Tseliakhovich, D., & Hirata, C., *Phys. Rev. D* **82**, 3520 (2010). Copyright 2010 by the American Physical Society.

length of the velocity field (i.e., several megaparsecs). In the rest frame of the baryons, the analogs to equation (2.2) and (2.3) are (note that we require separate equations for each type of matter)

$$\frac{\partial \delta_c}{\partial t} = \frac{i}{a} \mathbf{u}_{bc} \cdot \mathbf{k} \delta_c - \theta_c, \quad (2.24)$$

$$\frac{\partial \theta_c}{\partial t} = \frac{i}{a} \mathbf{u}_{bc} \cdot \mathbf{k} \theta_c - \frac{3H^2}{2} (\Omega_c \delta_c + \Omega_b \delta_b) - 2H\theta_c, \quad (2.25)$$

$$\frac{\partial \delta_b}{\partial t} = -\theta_b, \quad (2.26)$$

$$\frac{\partial \theta_b}{\partial t} = -\frac{3H^2}{2} (\Omega_c \delta_c + \Omega_b \delta_b) - 2H\theta_c + \frac{c_s^2 k^2}{a^2} \delta_b. \quad (2.27)$$

The first terms on the right-hand side of the cold dark matter equations remain here because the bulk velocity is large and so cannot be ignored during the linearization of the basic fluid equations. When they are large compared with the velocity divergence term, the relative streaming will have a significant effect on structure formation. This occurs at a scale

$$k_{ubc} \sim \frac{aH}{\langle u_{bc}^2 \rangle^{1/2}} \sim 180 \left( \frac{30 \text{ km s}^{-1}}{\langle u_{bc}^2(z_{rec}) \rangle^{1/2}} \right) \left( \frac{1+z}{50} \right)^{-1/2} \text{ Mpc}^{-1}, \quad (2.28)$$

where we have scaled to the typical bulk velocity at recombination and used  $\mathbf{u}_{bc} \propto (1+z)^{-1}$ . The suppression scale is larger at higher redshift, which means that the acoustic feature will affect structure formation to some degree at even



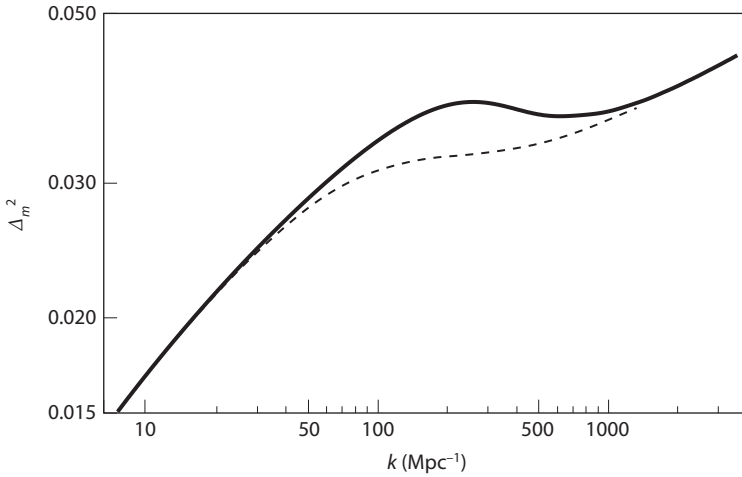


Figure 2.3 The isotropically averaged power spectrum of the matter distribution at  $z = 40$  with and without the relative streaming motions (dashed and solid lines, respectively). Reprinted Figure 2 with permission from Tseliakhovich, D., & Hirata, C., *Phys. Rev. D* **82**, 3520 (2010). Copyright 2010 by the American Physical Society.

larger scales than this estimate shows (see also §3.2). Note as well that the relative velocity term is much, much smaller than the divergence term on scales larger than the coherence length, which has  $k \sim 1 \text{ Mpc}^{-1}$ , so this system of equations is reasonably accurate on large scales as well.

Figure 2.3 shows the effect of these velocities on the total matter power spectrum at high redshifts: because the baryons constitute  $\sim 18\%$  of the matter, the dark matter power spectrum changes significantly on the relevant scales.

### 2.1.3 Normalizing the Power Spectrum

Although the shape of the power spectrum is well determined by linear perturbation theory in an expanding universe, the overall *amplitude* of the power spectrum is not specified by current models of inflation and is usually set by comparison with the observed CMB temperature fluctuations or with measures of large-scale structure based on surveys of galaxies, clusters of galaxies, or the intergalactic gas.

The most popular large-scale structure normalization is through the observed mass fluctuation amplitude (at the present day) on  $8h^{-1} \text{ Mpc}$ , roughly the scale of galaxy clusters. To relate this quantity to the power spectrum, we must consider the statistical distribution of the smoothed density field. We define a window (or filter) function  $W(\mathbf{r})$  normalized so that  $\int d^3r W(\mathbf{r}) = 1$ , where the smoothed density perturbation field is  $\int d^3r \delta(\mathbf{x}) W(\mathbf{r})$ . The simplest observed quantity is a measure of the masses (relative to the mean) inside spheres of radius  $R$ ; in this case we use a “spherical top-hat” window (similar to a

three-dimensional cookie cutter), in which  $W = \text{constant}$  inside a sphere of radius  $R$ , and  $W = 0$  outside.

The normalization of the present day power spectrum at  $z = 0$  is then specified by the variance of this density field when smoothed on the particular scale of  $8h^{-1}\text{Mpc}$ ,  $\sigma_8 \equiv \sigma(R = 8h^{-1}\text{Mpc})$ . For the top-hat filter, the smoothed perturbation field is denoted by  $\delta_R$  or  $\delta_M$ , where the enclosed mass  $M$  is related to the comoving radius  $R$  by  $M = 4\pi\bar{\rho}_m R^3/3$ , in terms of the current mean density of matter  $\bar{\rho}_m$ . We then write the variance  $\langle \delta_M^2 \rangle$  (relative to the mean) as<sup>ii</sup>

$$\sigma^2(M) = \left\langle \frac{1}{V} \int d^3x \delta(\mathbf{x}) W(\mathbf{x}) \frac{1}{V} \int d^3x' \delta(\mathbf{x}') W(\mathbf{x}') \right\rangle \quad (2.29)$$

$$= \frac{1}{V^2} \int d^3x d^3x' W(\mathbf{x}) W(\mathbf{x}') \xi(|\mathbf{x} - \mathbf{x}'|) \quad (2.30)$$

$$= \int \frac{d^3k}{(2\pi)^3} P(k) \frac{|W_{\mathbf{k}}|^2}{V^2}, \quad (2.31)$$

where  $W_{\mathbf{k}}$  is the Fourier transform of the window function. For the usual choice of a spherical top hat, this variance becomes

$$\sigma^2(M) \equiv \sigma^2(R) = \int_0^\infty \frac{dk}{k} \Delta^2(k) \left[ \frac{3j_1(kR)}{kR} \right]^2, \quad (2.32)$$

where  $j_1(x) = (\sin x - x \cos x)/x^2$ , and  $\Delta^2(k) = k^3 P(k)/2\pi^2$  is the so-called dimensionless power spectrum. The term  $\Delta^2$  expresses the contribution, per log wavenumber, of the power spectrum to the net variance.

While the normalization of the power spectrum requires only  $\sigma_8$ , we will see in the next chapter that the function  $\sigma(M)$  plays a major role in fixing the abundance of collapsed objects. We therefore show it in Figure 2.4 as a function of mass and redshift for our standard cosmological model. Note that  $\sigma^2 \propto \delta^2 \propto D(t)^2$ , so the time dependence is trivial (at least in linear theory).

For modes with random phases, the probability that different regions with the same comoving size  $M$  will have a perturbation amplitude between  $\delta$  and  $\delta + d\delta$  is Gaussian with a zero mean and a variance  $\sigma^2(M)$ ,

$$p(\delta)d\delta = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\delta^2/2\sigma^2} d\delta. \quad (2.33)$$

These so-called Gaussian perturbations are a key prediction of inflation; they have the convenient property that the statistical distribution of densities is described entirely by the power spectrum (through  $\sigma^2$ ).

## 2.2 The Thermal History during the Dark Ages

In addition to the density evolution, the second key “initial condition” for galaxy formation is the temperature of the hydrogen and helium gas that will collapse

<sup>ii</sup>Note that  $\sigma^2$  can equally well be considered a function of spatial scale  $R$ .

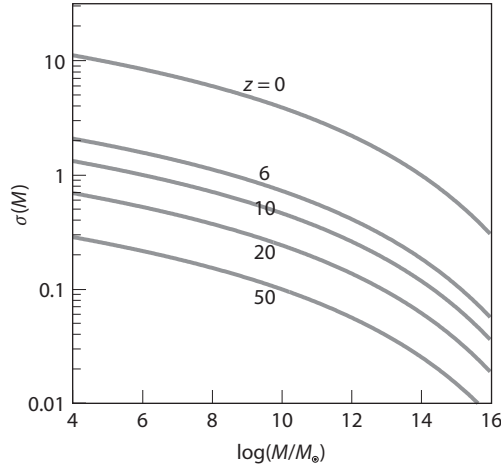


Figure 2.4 The root-mean-square amplitude of linearly extrapolated density fluctuations  $\sigma$  as a function of mass  $M$  (in solar masses  $M_\odot$ , within a spherical top-hat filter) at different redshifts  $z$ . Halos form in regions that exceed the background density by a factor of the order of unity. This threshold is surpassed only by rare (many- $\sigma$ ) peaks for high masses at high redshifts. When discussing the abundance of halos, we will factor out the linear growth of perturbations and use the function  $\sigma(M)$  at  $z = 0$ . The comoving radius of an unperturbed sphere containing a mass  $M$  is  $R = 1.85(M/10^{12} M_\odot)^{1/3}$  Mpc.

into the first galaxies. If it were isolated, the gas would simply cool adiabatically with the overall expansion of the universe. In general, for an ideal gas this cooling rate can be written as  $(\gamma - 1)(\dot{\rho}_b/\rho_b)T_e$ , where  $\rho_b$  is the baryon density, and  $\gamma = 5/3$  is the adiabatic index of a monatomic gas. For gas at the mean density, the factor  $(\dot{\rho}_b/\rho_b) = -3H$  owing to the Hubble expansion.

However, the gas is not thermally isolated: it may exchange energy with the ambient radiation field. Although cosmological recombination at  $z \sim 1100$  results in a nearly neutral universe, a small fraction  $\sim 10^{-4}$  of electrons remain free until the era of the first galaxies. These free electrons scatter off CMB photons and bring the gas closer to equilibrium with the radiation field.

A free electron moving at a speed  $v \ll c$  relative to the cosmic rest frame would probe a Doppler-shifted CMB temperature with a dipole pattern,

$$T(\theta) = T_\gamma \left( 1 + \frac{v}{c} \cos \theta \right), \quad (2.34)$$

where  $\theta$  is the angle relative to its direction of motion, and  $T_\gamma$  is the average CMB temperature. Naturally, the radiation will exert a frictional force on the electron opposite its direction of motion. The CMB energy density within a solid angle  $d\Omega = d \cos \theta d\phi$  (in spherical coordinates) will be  $d\epsilon = a_{\text{rad}} T^4(\theta) d\Omega/4\pi$  (where  $a_{\text{rad}}$  is the radiation constant). Since each photon carries a momentum equal to its energy divided by  $c$ , the electron will be slowed along its direction of motion by a net momentum flux  $c(d\epsilon/c) \times \cos \theta$ . The product of this

momentum flux and the Thomson (Compton) cross section of the electron ( $\sigma_T$ ) yields the net drag force acting on the electron,

$$m_e \frac{dv}{dt} = - \int \sigma_T \cos \theta d\epsilon = - \frac{4}{3c} \sigma_T a_{\text{rad}} T_\gamma^4 v. \quad (2.35)$$

The rate of energy loss by the electron is obtained by multiplying the drag force by  $v$ , which yields

$$\frac{d}{dt} E = - \frac{8\sigma_T}{3m_e c} a_{\text{rad}} T_\gamma^4 E, \quad (2.36)$$

where  $E = (1/2) m_e v^2$ . For a thermal ensemble of electrons at a nonrelativistic temperature  $T$ , the average energy is  $\langle E \rangle = (3/2) k_B T_e$ . If the electrons reach thermal equilibrium with the CMB, then the net rate of energy exchange must vanish. Therefore, there must be a stochastic heating term that balances the cooling term when  $T = T_\gamma$ . The origin of this heating term is obvious. Electrons starting at rest will be pushed around by the fluctuating electric field of the CMB until the ensemble reaches an average kinetic energy per electron of  $\langle E \rangle = (3/2) k_B T_\gamma$ , at which point the ensemble stays in thermal equilibrium with the radiation.

The temperature evolution of gas at the mean cosmic density, which cools only through its coupling to the CMB and its adiabatic Hubble expansion (with no radiative cooling due to atomic transitions or heating by galaxies), is therefore described by the equation

$$\frac{dT_e}{dt} = \frac{x}{(1+x)} \left[ \frac{T_\gamma - T_e}{t_C(z)} \right] - 2HT_e, \quad (2.37)$$

where  $t_C$  is the Compton cooling time,

$$t_C \equiv \left( \frac{8\sigma_T a_{\text{rad}} T_\gamma^4}{3m_e c} \right)^{-1} = 1.2 \times 10^8 \left( \frac{1+z}{10} \right)^{-4} \text{ yr}, \quad (2.38)$$

and  $x$  is the fraction of all electrons that are free. For an electron-proton gas,  $x = n_e/(n_e + n_H)$ , where  $n_e$  and  $n_H$  are the electron and hydrogen densities respectively, and  $T_\gamma \propto (1+z)$ . The second term on the right-hand side of equation (2.37),  $-2HT_e$ , yields the adiabatic scaling  $T_e \propto (1+z)^2$  in the absence of energy exchange with the CMB.

The relative importance of these two heating and cooling mechanisms therefore depends on the residual fraction of free electrons after cosmological recombination. If we ignore helium for simplicity, the rate at which electrons recombine is roughly<sup>iii</sup>

$$\frac{dx}{dt} = -\alpha_B(T_e) x^2 \bar{n}_H, \quad (2.39)$$

<sup>iii</sup>At high redshifts, recombination is delayed by the large photon density and line emission. Detailed calculations at  $z \gg 100$  require tracking the complex network of recombination reactions.

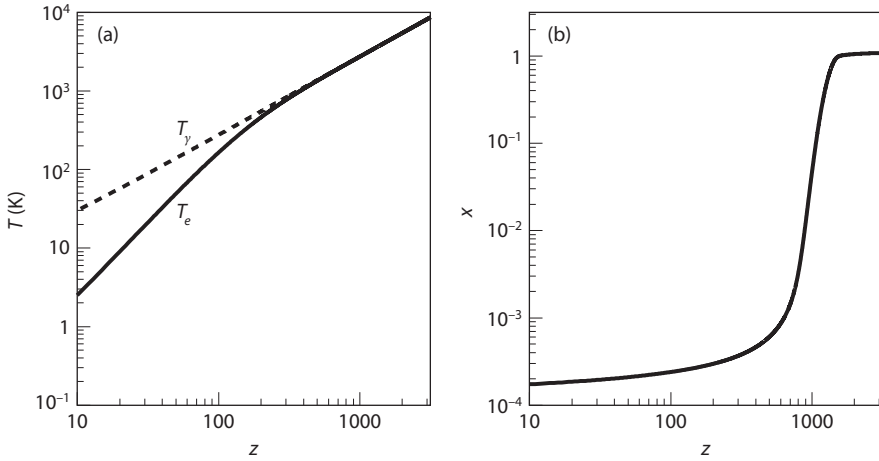


Figure 2.5 Thermal and ionization histories of the Universe before the first stars formed (panels a and b, respectively). In the left panel, the solid and dashed curves show  $T_e$  and  $T_\gamma$ , respectively. Note that the ionized fraction  $x$  decreases rapidly after recombination at  $z \sim 1100$  and then “freezes out” at  $z \sim 300$ . Meanwhile, Compton scattering keeps  $T_e \approx T_\gamma$  until  $z \sim 200$ , after which the declining CMB energy density and small residual ionized fraction are no longer sufficient to maintain thermal contact between the gas and CMB. At later times,  $T_e \propto (1+z)^2$ , as appropriate for an adiabatically expanding non-relativistic gas. These results were produced with the publicly available code RECFAST (<http://www.astro.ubc.ca/people/scott/recfast.html>).

where  $\alpha_B \propto T_e^{-0.7}$  is the case-B recombination coefficient.<sup>iv</sup> With our preferred cosmological parameters, the fractional change in  $x$  per Hubble time is therefore

$$\frac{\dot{n}_e}{Hn_e} \approx 7x(1+z)^{0.8}. \quad (2.40)$$

Electrons “freezeout” and cease to recombine effectively when this factor becomes on the order of unity; after that point, the Hubble expansion time is shorter than the recombination time. More precise numerical calculations give  $x \approx 3 \times 10^{-4}$  at  $z \approx 200$ , as shown in Figure 2.5.

Inserting this value into equation (2.37), we find that the small fraction of residual electrons enforces thermal equilibrium between the gas and CMB down to  $z \approx 200$ , when Compton heating finally becomes inefficient. Figure 2.5 shows a more exact calculation: note how the gas and CMB temperatures begin to depart at  $z \sim 200$ , and the gas begins to follow the expected adiabatic evolution  $T_e \propto (1+z)^2$  at  $z \sim 100$ .

<sup>iv</sup>This ignores recombinations to the ground state, which generate a new ionizing photon and so do not change the net ionized fraction. See §9.2.1 for more discussion of the recombination rate.

Note, however, that Compton cooling can become important again if the Universe is “reionized” by stars or quasars; once  $x \approx 1$ , the Compton cooling time is still shorter than the age of the Universe (and hence significant relative to adiabatic cooling) down to a redshift  $z \sim 6$ .

### 2.2.1 Fluctuations in the IGM Temperature

Equation (2.37) describes the evolution of the mean IGM temperature. However, two factors can induce inhomogeneities in this field. First, the CMB temperature varies slightly across the Universe, so each electron will scatter off a different  $T_\gamma$ . Second, the adiabatic expansion term depends on the local density. In an overdense region, where gravity slows the expansion (or even causes contraction), the cooling is slower (and may turn into heating); in an underdense region, the cooling accelerates. Thus, the IGM will be seeded by small temperature fluctuations that reflect its density structure.

To describe these fluctuations, we write  $\delta_T$  as the fractional temperature fluctuation and  $\delta_\gamma$  as the photon density fluctuation and note that (for a blackbody)  $\delta_\gamma = 4\delta_{T_\gamma}$ , where the latter is the photon temperature fluctuation. Then, the analog of equation (2.37) is

$$\frac{d\delta_T}{dt} = \frac{2}{3} \frac{d\delta_b}{dt} + \frac{x(t)}{t_C(z)} \left[ \delta_\gamma \left( \frac{\bar{T}_\gamma}{\bar{T}_e} - 1 \right) + \frac{\bar{T}_\gamma}{\bar{T}_e} (\delta_{T_\gamma} - \delta_T) \right]. \quad (2.41)$$

Here the first term describes adiabatic cooling due to expansion (allowing for variations in the expansion rate), and the second accounts for variations in the rate of energy exchange through Compton scattering (which can result from variations in either the gas or photon temperatures); overbars denote the mean values for the CMB and electron temperatures.

Meanwhile, the fluctuations in the baryon temperature influence the density evolution as well. If we allow arbitrary fluctuations in the temperature field, rather than forcing them to trace the density fluctuations, equation (2.6) then reads

$$\frac{\partial^2 \delta}{\partial t^2} + 2H \frac{\partial \delta}{\partial t} = \frac{3}{2} H^2 (\Omega_c \delta_c + \Omega_b \delta_b) - \frac{k^2}{a^2} \frac{k_B \bar{T}_e}{\mu m_p} (\delta_b + \delta_T). \quad (2.42)$$

This, result together with equations (2.41), (2.37), and (2.39) for the temperature and ionized fraction evolution, provides a complete set of equations for tracing the density and temperature evolution, in the absence of relative streaming. If streaming is included, the final term in equation (2.27) must be replaced by the final term in equation (2.42).

Figure 2.6 shows the resulting power spectra for  $\delta_c$ ,  $\delta_b$ ,  $\delta_T$ , and  $\delta_{T_\gamma}$  at four different redshifts. Note how the photon perturbations are strongly suppressed on small scales (below the sound horizon) thanks to their large pressure. Near recombination, the baryonic perturbations are also suppressed on these scales, especially in the temperature. After recombination, the baryons fall into the dark matter potential wells, where their perturbations grow rapidly, and temperature fluctuations also grow quickly thanks largely to the variations in the

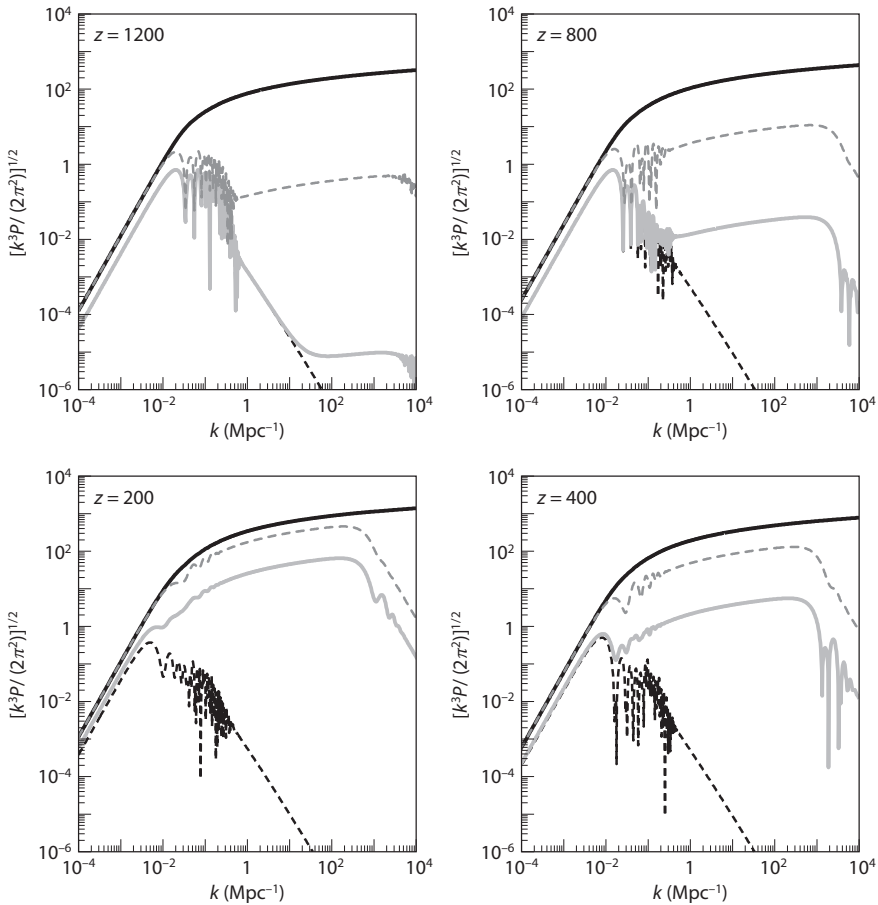


Figure 2.6 Power spectra for density and temperature fluctuations versus comoving wavenumber at four different redshifts. The curves show the CDM density (solid), baryon density (dotted), baryon temperature (short dashes) and photon temperature (long dashes). These curves do not include the relative streaming of the baryons and cold dark matter. Naoz, S., & Barkana, R., *Mon. Not. R. Astron. Soc.* **362**, 1047 (2005). Copyright 2005 by the Royal Astronomical Society.

adiabatic cooling rate. The turnover at very small scales in the baryonic power spectrum is due to the finite pressure of the gas. The baryon acoustic oscillations are also visible near  $k \sim 0.01 \text{ Mpc}^{-1}$ .

Baryon acoustic oscillations do not exist. The pattern of a single peak and two secondary "harmonics" is the pattern of turbulence in the plasma epoch, where the collisional particles of the turbulence are proto-galaxies that fragmented at  $10^{12}$  seconds after the big bang. A similar pattern from big bang turbulence persists at the largest scales. The single peak is the turbulent vortex line. The secondary peaks are from secondary vortices that stretch the primary vortex.

## Chapter Three

---

### Nonlinear Structure and Halo Formation

In the last chapter, we followed the evolution of structure in the linear regime, when the perturbations are small. Of course, most of the objects we study with telescopes are far outside this regime, with typical densities many thousands of times the cosmic mean. In this chapter, we take the next steps toward understanding these objects by studying the evolution of perturbations in the nonlinear regime. We focus for the most part on analytic models that shed light on the physical processes involved.

The advent of computer technology has made numerical studies of nonlinear evolution almost routine, and many of today's theoretical calculations follow this path. The analytic approaches we describe inform these calculations, but the numerical simulations allow us to sharpen our conclusions and predictions. We discuss this synergy and describe "semianalytic" models that can be written analytically but whose ultimate justification lies in their good agreement with numerical simulations. We describe the fundamental aspects of computational methods in the last section of the chapter.

#### 3.1 Spherical Collapse

Existing cosmological data suggest that the dark matter is "cold," that is, its pressure is negligible during the gravitational growth of galaxies. This makes the nonlinear evolution relatively simple, as it depends purely on the gravitational force. We can therefore make some progress in understanding galaxy formation by considering models for this gravitational growth that are sufficiently simple to extend into the nonlinear regime.

For simplicity, let us consider an isolated, spherically symmetric density or velocity perturbation of the smooth cosmological background and examine the dynamics of a test particle at a radius  $r$  relative to the center of symmetry. Birkhoff's theorem (see §1.2.2) implies that we may ignore the mass outside this radius in computing the motion of our particle. The equation of motion describing the system reduces to the usual Friedmann equation for the evolution of the scale factor of a homogeneous Universe, but with a density parameter  $\Omega$  that now takes into account the additional mass interior to the shell and its modified expansion velocity. In particular, despite the arbitrary density and velocity profiles given to the perturbation, only the total mass interior to the particle's radius and the peculiar velocity at the particle's radius contribute to the effective value of  $\Omega$ . We may thus find a solution to the particle's motion that describes



its departure from the background Hubble flow and its subsequent collapse or expansion. This solution holds until our particle crosses paths with one from a different radius, which happens rather late for most initial conditions.

As with the Friedmann equation for a smooth Universe, it is possible to reformulate the problem in a Newtonian form. At some early epoch corresponding to a scale factor  $a_i \ll 1$ , we consider a spherical patch of uniform overdensity  $\delta_i$ , making a so-called top-hat perturbation. If  $\Omega_m$  is essentially unity at this time and if the perturbation is a pure growing mode, then the initial peculiar velocity is radially inward with magnitude  $\delta_i H(t_i)r/3$ , where  $H(t_i)$  is the Hubble constant at the initial time, and  $r$  is the radius from the center of the sphere. This result can easily be derived from mass conservation (the continuity equation) in spherical symmetry. The collapse of a spherical top-hat perturbation beginning at radius  $r_i$  is described by

$$\frac{d^2 r}{dt^2} = H_0^2 \Omega_\Lambda r - \frac{GM}{r^2}, \quad (3.1)$$

where  $r$  is the radius in a fixed (not comoving) coordinate frame,  $H_0$  is the present-day Hubble constant, and the unperturbed Hubble flow velocity (to which the previously mentioned peculiar velocity should be added) is given by  $dr/dt = H(t)r$ . The total mass enclosed within radius  $r$  is  $M = (4\pi/3)r_i^3 \rho_i (1 + \delta_i)$ , where  $\rho_i$  is the background density of the Universe at time  $t_i$ . We next define the dimensionless radius  $x = a_i(r/r_i)$  and rewrite equation (3.1) as

$$\frac{1}{H_0^2} \frac{d^2 x}{dt^2} = -\frac{\Omega_m}{2x^2} (1 + \delta_i) + \Omega_\Lambda x. \quad (3.2)$$

Henceforth we will assume a flat universe with  $\Omega_\Lambda = 1 - \Omega_m$ . Our initial conditions for the integration of this orbit are

$$x(t_i) = a_i, \quad (3.3)$$

$$\frac{dx}{dt}(t_i) = H(t_i)x(t_i) \left(1 - \frac{\delta_i}{3}\right) = H_0 a_i \left(1 - \frac{\delta_i}{3}\right) \sqrt{\frac{\Omega_m}{a_i^3} + \Omega_\Lambda}, \quad (3.4)$$

where  $H(t_i) = H_0[\Omega_m/a_i^3 + (1 - \Omega_m)]^{1/2}$  is the Hubble parameter for a flat Universe at the initial time  $t_i$ . Integrating equation (3.2) we obtain

$$\frac{1}{H_0^2} \left(\frac{dx}{dt}\right)^2 = \frac{\Omega_m}{x} (1 + \delta_i) + \Omega_\Lambda x^2 + K, \quad (3.5)$$

where  $K$  is a constant of integration. Evaluating this expression at the initial time and dropping terms of order  $a_i$  (with  $\delta_i \propto a_i$ ), we find

$$K = -\frac{5\delta_i}{3a_i} \Omega_m. \quad (3.6)$$

If  $K$  is sufficiently negative, the particle will turn around, and the sphere will collapse to zero size at a time

$$H_0 t_{\text{coll}} = 2 \int_0^{a_{\text{max}}} da (\Omega_m/a + K + \Omega_\Lambda a^2)^{-1/2}, \quad (3.7)$$

where  $a_{\max}$  is the value of  $a$  that sets the denominator of the integrand to zero, and we have used the fact that  $\delta_i \ll 1$ . (The integral itself determines the total expansion time; the factor of 2 accounts for the time from maximum expansion to collapse.) The analogy to a test particle escaping a point mass in equation (3.1) is illuminating here: in that case the constant  $K$  is simply proportional to the total energy per unit mass of the system, which determines whether the particle escapes to infinity. Here, a large negative  $K$  (enough to overcome the effective repulsive force from the cosmological constant) implies the same recollapse.

It is easier to solve the equation of motion analytically for the regime in which the cosmological constant is negligible,  $\Omega_\Lambda = 0$  and  $\Omega_m = 1$  (adequate for describing redshifts  $1 < z < 10^3$ ). There are three branches of solutions: one in which the particle turns around and collapses, another in which it reaches an infinite radius with some asymptotically positive velocity, and a third intermediate case in which it reaches an infinite radius but with a velocity that approaches zero. In fact, although we have cast this problem as a test particle in an overdense or underdense region, we could have developed exactly the same equations by carving out a spherical region from a truly uniform medium. Then, the three possibilities would simply correspond to a closed, an open, and a flat Universe (with  $\Omega_\Lambda = 0$ ). The three solutions may be written as

$$\left. \begin{aligned} r &= A(1 - \cos \eta) \\ t &= B(\eta - \sin \eta) \end{aligned} \right\} \quad \text{Closed} \quad \checkmark \quad (0 \leq \eta \leq 2\pi), \quad (3.8)$$

Because the big bang was due to a turbulent combustion instability, a significant amount of entropy was produced that requires a closed universe and eventually a big crunch.

$$\left. \begin{aligned} r &= A(\cosh \eta - 1) \\ t &= B(\sinh \eta - \eta) \end{aligned} \right\} \quad \text{Open} \quad (0 \leq \eta \leq \infty), \quad (3.10)$$

where  $A^3 = GMB^2$  applies in all cases even though the constants have different values in each one. All three solutions have  $r^3 = 9GMt^2/2$  as  $t$  goes to zero, which matches the linear theory expectation that the perturbation amplitude get smaller as one goes back in time. In the closed case, the shell turns around at time  $\pi B$  and radius  $2A$  (when its density contrast relative to the background of an  $\Omega_m = 1$  Universe is  $9\pi^2/16 = 5.6$ ), and collapses to zero radius at time  $2\pi B$ . Interestingly, these collapse times are independent of the initial distance from the origin: perturbations with fixed initial density contrast collapse homogeneously, with all shells turning around and collapsing at the same time. Figure 3.1 illustrates the stages of this collapse process.

This is the fully nonlinear solution for the simplified problem of collapse of a purely spherical top-hat perturbation. Of course, the real density distribution of the Universe is much more complicated. Although we cannot describe analytically the full nonlinear evolution of density perturbations, we *can* fully describe their linear evolution. A compromise is then to use this linear evolution to identify regions (such as galaxies) where spherical nonlinear evolution is not a bad