

Malcolm S. Longair

Galaxy Formation

Second Edition

With 202 Figures and 20 Tables

 Springer

Preface

This is the second edition of my book *Galaxy Formation*. Many people liked the first edition which appeared in 1998, just before the explosion of magnificent new data which have completely changed the face of astrophysical cosmology. Many of the agonies which had to be gone through in the first edition have disappeared and, to many people's amazement, including mine, there is now a *concordance model* for cosmology, the cosmologist's equivalent of the particle physicist's *standard model*. Just like the standard model, however, the concordance model creates as many problems as it solves. This is not a cause for concern, but rather one for celebration because we are now able to ask much better and deeper questions than in the past. These questions indicate clearly the need for physics and astrophysics 'Beyond the Concordance Model'.

The object of this new edition is to bring this amazing story up-to-date, very much in the spirit of the first edition. To recapitulate some of the points made in the previous preface about the origin of the book, I was asked by Springer-Verlag to expand the set of lecture notes that I prepared in 1988 for the First Astrophysics School organised by the European Astrophysics Doctoral Network into a full-length book. The set of notes was entitled *Galaxy Formation* and was published as a chapter of the volume *Evolution of Galaxies: Astronomical Observations* (eds. I. Appenzeller, H.J. Habing and P. Lena, pages 1 to 93, Springer-Verlag Berlin, Heidelberg, 1989). In that chapter, I attempted to bridge the gap between elementary cosmology and the technical papers appearing in the literature which can seem quite daunting on first encounter. The objective was to present the physical ideas and key results as clearly as possible as an introduction and guide to the technical literature.

In 1993, more lecture notes on *The Physics of Background Radiation* were prepared for the 23rd Advanced Course of the Swiss Society of Astrophysics and Astronomy, the topic being *The Deep Universe* (A.R. Sandage, R.G. Kron and M.S. Longair, Springer-Verlag Berlin, Heidelberg, 1995). Then, also in 1993, I completed a history of twentieth century astrophysics and cosmology, which was published as Chap. 23 of a three-volume work entitled *Twentieth Century Physics* (eds. L.M. Brown, A. Pais and A.B. Pippard, IOP Publications, AIP Press Bristol, and New York 1995). A much enlarged full-length book on this topic entitled *The Cosmic Century: A History of Astrophysics and Cosmology* was published by Cambridge University Press in 2006. That book brought the story of the origin of

VIII Preface

galaxies and the large-scale structure of the Universe up-to-date as of October 2005 and it has been further updated and expanded in the present book. Just as in the first edition, the present volume is much more than a recycled and concatenated version of previously published works. I have rewritten and rethought the original versions, expanded some parts, brought everything up-to-date and included new material.

I often find that I understand things best, and present them most clearly, when I have to prepare them for students, at either the undergraduate or the post-graduate level, and so I have adopted the same form of presentation here. I have intentionally presented the material in an informal, pedagogical manner, and attempted to avoid getting bogged down in formalities and technicalities. If the material becomes too difficult, I simply summarise the key points, give some appropriate references and pass on. My approach is to reduce the problems to their simplest form and rationalise from these examples the results of more complete analyses. Wherever it is feasible without excessive effort, we will attempt to derive exact results. The level of presentation is intended to be appropriate for a final-year undergraduate or first-year post-graduate course of lectures. In other words, it is assumed that the reader has a good grasp of basic physics but does not necessarily have the appropriate background in astronomy, astrophysics or cosmology. My aim has been to write a user-friendly book, taking particular care to expound carefully areas where I have found students have difficulty.

When I wrote the original set of lecture notes on galaxy formation, my objective was to tell the story of modern astrophysical cosmology from the perspective of one of its most important and fundamental problems of cosmology – how did the galaxies come about? I enjoy this approach to the exposition of modern cosmology because, to do the problem justice, it is essential to introduce the whole of what I call *classical cosmology*, as the framework for the discussion. This approach has, for me, the great advantage of concentrating upon a crucial problem of astrophysical cosmology rather than regarding the objective of cosmology as being simply the delineation of a preferred cosmological model, however interesting that is in its own right. As we will show, the origin of galaxies and larger-scale structures in the Universe is one of the great cosmological problems and has provided us with unique and direct information about the physics of the very early Universe.

This new understanding brings with it the question of whether or not the old structure of the book is really appropriate – do we really need to grind through all the old story in order to understand the problems raised by the concordance model? My decision has been to maintain much of the original structure of the book, largely because the approach was very strongly physics-motivated and the old story reveals much of the essential physics of the concordance model.

One final warning is in order. I make no claim that this presentation is complete, unbiased or objective. You should regard the book as my own impressions and opinions of what I consider to be the important issues of modern astrophysical cosmology. Others would tell the story in a completely different way and put

emphasis
I will end
but the te
encourage
biases.

Good I

Venice and
July 2007

galaxies and the large-scale structure of the Universe up-to-date as of October 2005 and it has been further updated and expanded in the present book. Just as in the first edition, the present volume is much more than a recycled and concatenated version of previously published works. I have rewritten and rethought the original versions, expanded some parts, brought everything up-to-date and included new material.

I often find that I understand things best, and present them most clearly, when I have to prepare them for students, at either the undergraduate or the post-graduate level, and so I have adopted the same form of presentation here. I have intentionally presented the material in an informal, pedagogical manner, and attempted to avoid getting bogged down in formalities and technicalities. If the material becomes too difficult, I simply summarise the key points, give some appropriate references and pass on. My approach is to reduce the problems to their simplest form and rationalise from these examples the results of more complete analyses. Wherever it is feasible without excessive effort, we will attempt to derive exact results. The level of presentation is intended to be appropriate for a final-year undergraduate or first-year post-graduate course of lectures. In other words, it is assumed that the reader has a good grasp of basic physics but does not necessarily have the appropriate background in astronomy, astrophysics or cosmology. My aim has been to write a user-friendly book, taking particular care to expound carefully areas where I have found students have difficulty.

When I wrote the original set of lecture notes on galaxy formation, my objective was to tell the story of modern astrophysical cosmology from the perspective of one of its most important and fundamental problems of cosmology – how did the galaxies come about? I enjoy this approach to the exposition of modern cosmology because, to do the problem justice, it is essential to introduce the whole of what I call *classical cosmology*, as the framework for the discussion. This approach has, for me, the great advantage of concentrating upon a crucial problem of astrophysical cosmology rather than regarding the objective of cosmology as being simply the delineation of a preferred cosmological model, however interesting that is in its own right. As we will show, the origin of galaxies and larger-scale structures in the Universe is one of the great cosmological problems and has provided us with unique and direct information about the physics of the very early Universe.

This new understanding brings with it the question of whether or not the old structure of the book is really appropriate – do we really need to grind through all the old story in order to understand the problems raised by the concordance model? My decision has been to maintain much of the original structure of the book, largely because the approach was very strongly physics-motivated and the old story reveals much of the essential physics of the concordance model.

One final warning is in order. I make no claim that this presentation is complete, unbiased or objective. You should regard the book as my own impressions and opinions of what I consider to be the important issues of modern astrophysical cosmology. Others would tell the story in a completely different way and put

emphasis upon different parts of what is unquestionably a multi-dimensional story. I will endeavour to include as wide a spectrum of ideas and opinions as possible, but the text will inevitably be incomplete. I do not worry about this – it should encourage you to read as widely as possible in order to neutralise my prejudices and biases.

Good Luck!

Venice and Cambridge,
July 2007

Malcolm Longair

Since the first edition was published, many colleagues have pointed out minor errors and suggested further topics for discussion. The preparation of the present edition began in December 2005, at the beginning of a two-year period of sabbatical leave kindly granted by the University of Cambridge. So much had changed in the matter of just a few years that essentially everything had to be rewritten in the light of this new understanding.

From February to May 2006, I was Adjunct Professor of Astronomy at the Astronomy Department of the University of Massachusetts at Amherst. I am most grateful to Ronald Snell and his colleagues for their warm welcome and the many helpful discussions we had about galaxies, Martin Weinberg, Neal Katz and Hou-Jun Mo were very stimulating colleagues. A special thank you goes to Dan McIntosh who introduced me to the red and blue sequences of galaxies and showed how our perception of what galaxies are has changed.

From October to December 2006, I was Visiting Professor at the Max Planck Institute for Astrophysics at Garching at the invitation of Rashid Sunyaev. The discussions with Rashid, Simon White and their colleagues were very helpful indeed in sharpening my understanding of many of the issues of astrophysical cosmology. I was back in Garching from February to April 2007 at the invitation of Catherine Cesarsky as a Visiting Scientist at the European Southern Observatory. This also proved to be a very stimulating visit. The wealth of expertise at ESO and the nearby Institutes made Garching a ideal haven in which to complete much of the final chapters of the new edition. I am most grateful to Bruno Leibundgut and all the staff members, visitors, post-docs and graduate students on whom I tried out some of the new sections of the book. The very final sections of the book were completed while I was a guest at the Osservatorio Astrofisico di Arcetri in Florence. I am most grateful to Francesco Palla and his colleagues for their kindness and hospitality.

Four colleagues are worthy of special thanks. Paul Schechter used the first edition of the book as a text for his courses on cosmology and kindly set up a web-site with corrections to it. He also made helpful suggestions for the present edition. Megan Donahue also used the first edition as a class text and provided a very helpful list of corrections and suggested improvements. She then ‘battle-tested’ Parts 2 and 3 of the new edition on her students and made helpful comments. I am most grateful to Megan and her students for this invaluable help which will undoubtedly have improved the book. Luigi Guzzo kindly read a number of sections of the book very carefully and provided helpful suggestions and corrections. Peter Schneider reviewed the draft text at the invitation of Springer-Verlag and suggested a number of very helpful improvements.

Finally, the book is dedicated to my family, Deborah, Mark and Sarah, whose constant love, support and patience have made it possible.

Contents

Part I Preliminaries	
1	A Brief History of Cosmology and Galaxy Formation
1.1	Pre-History
1.2	The Galaxies and the Structure of our Galaxy
1.3	The Theory of the Expanding Universe
1.4	The Big Bang
1.5	Galaxy and Structure Formation
1.6	Hot and Cold Dark Matter
1.7	The Very Early Universe
2	The Large-Scale Structure of the Universe
2.1	The Spectrum and Isotropy of the Cosmic Microwave Background Radiation
2.1.1	The Spectrum of the Cosmic Microwave Background Radiation
2.1.2	The Isotropy of the Cosmic Microwave Background Radiation
2.2	The Large-Scale Distribution of Galaxies
2.2.1	Two-Point Correlation Functions
2.2.2	Walls and Voids in the Distribution of Galaxies on Large Scales
2.3	Hubble’s Law and the Expansion of the Universe
2.4	Conclusion
3	Galaxies
3.1	Introduction
3.2	The Revised Hubble Sequence for Galaxies
3.3	Peculiar and Interacting Galaxies
3.4	The Light Distribution in Galaxies
3.4.1	Elliptical Galaxies
3.4.2	Spiral and Lenticular Galaxies
3.4.3	Putting the Light Distributions Together
3.5	The Masses of Galaxies

XIV Contents

3.5.1	The Virial Theorem for Clusters of Stars, Galaxies and Clusters of Galaxies	63	
3.5.2	The Rotation Curves of Spiral Galaxies	66	
3.5.3	The Velocity Dispersions of Elliptical Galaxies	69	
3.6	The Properties of Spiral and Elliptical Galaxies	70	
3.6.1	The Faber-Jackson Relation and the Fundamental Plane ..	70	
3.6.2	Ellipticals Galaxies as Triaxial Systems	71	
3.6.3	The Tully-Fisher Relation for Spiral Galaxies	73	
3.6.4	Luminosity-Metallicity Relations	74	
3.7	The Luminosity Function of Galaxies	77	
3.7.1	Aspects of the Luminosity Function of Galaxies	79	
3.7.2	The Integrated Luminosity and the Mean Mass-to-Luminosity Ratio for Visible Matter in the Universe		
3.8	The Properties of Galaxies: Correlations Along the Hubble Sequence	83	
3.9	The Red and Blue Sequences	84	
3.9.1	Colour Versus Absolute Magnitude	88	
3.9.2	Sérsic Index and Colour	89	
3.9.3	Mean Stellar Age and Concentration Index C	91	
3.9.4	The Effect of the Galaxy Environment	92	
3.9.5	The New Perspective	93	
3.10	Concluding Remark	94	
4	Clusters of Galaxies	95	
4.1	The Large-Scale Distribution of Clusters of Galaxies	95	
4.1.1	The Abell Catalogues of Rich Clusters of Galaxies	96	
4.1.2	Comparison with Clusters Selected from the Sloan Digital Sky Survey	98	
4.1.3	Abell Clusters and the Large-Scale Distribution of Galaxies ..	99	
4.2	The Distribution of Galaxies in Clusters of Galaxies	101	
4.2.1	The Galaxy Content and Spatial Distribution of Galaxies in Clusters	101	
4.2.2	Clusters of Galaxies and Isothermal Gas Spheres	103	
4.2.3	The Luminosity Function for Cluster Galaxies	108	
4.2.4	Summary of the Properties of Rich Clusters of Galaxies ..	109	
4.3	Dynamical Estimates of the Masses of Clusters of Galaxies ..	110	
4.4	X-Ray Observations of Hot Gas in Clusters of Galaxies	114	
4.5	The Sunyaev-Zeldovich Effect in Hot Intraccluster Gas	125	
4.6	Gravitational Lensing by Galaxies and Clusters of Galaxies ..	128	
4.6.1	Basic Theory of Gravitational Deflections	128	
4.6.2	Magnification of Images by Gravitational Lensing	130	
4.6.3	Extended Deflectors	133	
4.6.4	Gravitational Lensing and the Astrophysics of Galaxies ..	136	
4.7	Forms of Dark Matter	139	

P
—
—
5

6

7

7

<p>3.5.1 The Virial Theorem for Clusters of Stars, Galaxies and Clusters of Galaxies 63</p> <p>3.5.2 The Rotation Curves of Spiral Galaxies 66</p> <p>3.5.3 The Velocity Dispersions of Elliptical Galaxies 69</p> <p>3.6 THE Properties of Spiral and Elliptical Galaxies 70</p> <p>3.6.1 The Faber-Jackson Relation and the Fundamental Plane 70</p> <p>3.6.2 Ellipticals as Triaxial Systems 71</p> <p>3.6.3 The Tully-Fisher Relation for Spiral Galaxies 73</p> <p>3.6.4 Luminosity-Metallicity Relations 74</p> <p>3.7 THE Luminosity Function of Galaxies 77</p> <p>3.7.1 Aspects of the Luminosity Function of Galaxies 77</p> <p>3.7.2 The Integrated Luminosity and the Mean Mass-to-Luminosity Ratio for Visible Matter in the Universe 83</p> <p>3.8 The Properties of Galaxies: Correlations Along the Hubble Sequence 84</p> <p>3.9 THE Red and Blue Sequences 88</p> <p>3.9.1 Colour Versus Absolute Magnitude 89</p> <p>3.9.2 Sérsic Index and Colour 91</p> <p>3.9.3 Mean Stellar Age and Concentration Index C 91</p> <p>3.9.4 The Effect of the Galaxy Environment 92</p> <p>3.9.5 The New Perspective 93</p> <p>3.10 Concluding Remark 94</p>	<p>4.7.1 Baryonic Dark Matter 139</p> <p>4.7.2 Non-Baryonic Dark Matter 142</p> <p>4.7.3 Astrophysical and Experimental Limits 144</p> <p>5 THE Theoretical Framework 149</p> <p>5.1 The Cosmological Principle 149</p> <p>5.2 Isotropic Curved Spaces 150</p> <p>5.3 The Space-Time Metric for Isotropic Curved Spaces 155</p> <p>5.4 The Robertson-Walker Metric 158</p> <p>5.5 Observations in Cosmology 162</p> <p>5.5.1 The Cosmological Redshift 162</p> <p>5.5.2 Hubble's Law 165</p> <p>5.5.3 Angular Diameters 167</p> <p>5.5.4 Apparent Intensities 168</p> <p>5.5.5 Number Densities 170</p> <p>5.5.6 The Age of the Universe 171</p> <p>5.6 Summary 171</p> <p>6 An Introduction to Relativistic Gravity 173</p> <p>6.1 The Principle of Equivalence 173</p> <p>6.2 The Gravitational Redshift 176</p> <p>6.3 The Bending of Light Rays 179</p> <p>6.4 Further Complications 179</p> <p>6.5 The Route to General Relativity 181</p> <p>6.5.1 Four-Tensors in Relativity 184</p> <p>6.5.2 What Einstein Did 184</p> <p>6.6 Experimental and Observational Tests of General Relativity 189</p> <p>6.6.1 The Four Tests of General Relativity 189</p> <p>6.6.2 Pulsars and General Relativity 192</p> <p>6.6.3 Parameterised Post-Newtonian Models 195</p> <p>6.6.4 Variation of the Gravitational Constant with Cosmic Epoch 197</p> <p>6.7 Summary 198</p> <p>7 The Friedman World Models 199</p> <p>7.1 Einstein's Field Equations 199</p> <p>7.2 The Standard Friedman World Models with $\Lambda = 0$ 202</p> <p>7.2.1 The Newtonian Analogue of the Friedman World Models 202</p> <p>7.2.2 The Critical Density and the Density Parameter 204</p> <p>7.2.3 The Dynamics of the Friedman Models with $\Lambda = 0$ 205</p> <p>7.3 Friedman Models with Non-Zero Cosmological Constant 207</p> <p>7.3.1 The Cosmological Constant and the Vacuum Energy Density 208</p>
---	---

7.3.2	Varying the Equation of State of the Vacuum Energy	210	
7.3.3	The Dynamics of World Models with $\Lambda \neq 0$:		
7.4	General Considerations	211	
7.4.1	Observations in Cosmology	215	10.5
7.4.2	The Deceleration Parameter	216	10.6
7.4.3	The Cosmic Time–Redshift Relation	217	
7.4.4	Distance Measures as a Function of Redshift	218	
7.4.5	Angular Diameter–Redshift Relations	221	
7.4.6	Flux Density–Redshift Relations	223	
7.5	The Comoving Volume Within Redshift z	226	
7.6	Angular Diameter Distances Between Any Two Redshifts	228	11.1
7.7	The Flatness Problem	230	
7.7	Inhomogeneous World Models	231	
A7	The Robertson–Walker Metric for an Empty Universe	237	11.2
8	The Determination of Cosmological Parameters	241	11.3
8.1	The Cosmological Parameters	241	11.4
8.2	Testing the Friedman Models	242	
8.3	Hubble's Constant H_0	246	
8.4	The Age of the Universe T_0	250	
8.5	The Deceleration Parameter q_0	252	
8.5.1	The Redshift–Magnitude Relation for the Brightest Galaxies in Clusters	252	11.5
8.5.2	The Redshift–Magnitude Relation for Radio Galaxies	254	11.6
8.5.3	The Redshift–Magnitude Relation for Type 1a Supernovae	256	11.7
8.5.4	The Number Counts of Galaxies	259	
8.5.5	The Angular Diameter–Redshift Test	261	12.1
8.6	Ω_A and the Statistics of Gravitational Lenses	263	12.2
8.7	The Density Parameter Ω_0	267	
8.8	Summary	270	
9	The Thermal History of the Universe	271	1
9.1	Radiation-Dominated Universes	271	1
9.2	The Matter and Radiation Content of the Universe	273	12.3 S
9.3	The Epoch of Recombination	277	12.4 T
9.4	The Radiation-Dominated Era	281	11
9.5	The Speed of Sound as a Function of Cosmic Epoch	285	11
9.6	Early Epochs	286	12.5 D
10	Nucleosynthesis in the Early Universe	289	12.6 Is
10.1	Equilibrium Abundances in the Early Universe	289	12.7 B
10.2	The Decoupling of Neutrinos and the Neutrino Barrier	290	12
10.3	The Synthesis of the Light Elements	292	12.8 W
10.4	The Abundances of the Light Elements	295	

7.3.2 Varying the Equation of State of the Vacuum Energy	210
7.3.3 The Dynamics of World Models with $\Lambda \neq 0$:	
General Considerations	211
7.4 Observations in Cosmology	215
7.4.1 The Deceleration Parameter	216
7.4.2 The Cosmic Time–Redshift Relation	217
7.4.3 Distance Measures as a Function of Redshift	218
7.4.4 Angular Diameter–Redshift Relations	221
7.4.5 Flux Density–Redshift Relations	223
7.4.6 The Comoving Volume Within Redshift z	226
7.5 Angular Diameter Distances Between Any Two Redshifts	228
7.6 The Flatness Problem	230
7.7 Inhomogeneous World Models	231
A7 The Robertson–Walker Metric for an Empty Universe	237
8 The Determination of Cosmological Parameters	241
8.1 The Cosmological Parameters	241
8.2 Testing the Friedman Models	242
8.3 Hubble's Constant H_0	246
8.4 The Age of the Universe T_0	250
8.5 The Deceleration Parameter q_0	252
8.5.1 The Redshift–Magnitude Relation for the Brightest Galaxies in Clusters	252
8.5.2 The Redshift–Magnitude Relation for Radio Galaxies	254
8.5.3 The Redshift–Magnitude Relation for Type Ia Supernovae	256
8.5.4 The Number Counts of Galaxies	259
8.5.5 The Angular Diameter–Redshift Test	261
8.6 Ω_A and the Statistics of Gravitational Lenses	263
8.7 The Density Parameter Ω_0	267
8.8 Summary	270
9 The Thermal History of the Universe	271
9.1 Radiation-Dominated Universes	271
9.2 The Matter and Radiation Content of the Universe	273
9.3 The Epoch of Recombination	277
9.4 The Radiation-Dominated Era	281
9.5 The Speed of Sound as a Function of Cosmic Epoch	285
9.6 Early Epochs	286
10 Nucleosynthesis in the Early Universe	289
10.1 Equilibrium Abundances in the Early Universe	289
10.2 The Decoupling of Neutrinos and the Neutrino Barrier	290
10.3 The Synthesis of the Light Elements	292
10.4 The Abundances of the Light Elements	295
11 The Evolution of Perturbations in the Standard Big Bang	311
11.1 What the Theorists Are Trying to Do	311
11.1.1 Why this Programme Is Feasible	312
11.1.2 A Warning	312
11.2 The Non-Relativistic Wave Equation for the Growth of Small Perturbations in the Expanding Universe	313
11.3 The Jeans' Instability	317
11.4 The Jeans' Instability in an Expanding Medium	319
11.4.1 Small Perturbation Analysis	319
11.4.2 Perturbing the Friedman Solutions	321
11.4.3 Falling Poles	322
11.4.4 The General Solution	324
11.5 The Evolution of Peculiar Velocities in the Expanding Universe	327
11.6 The Relativistic Case	331
11.7 The Basic Problem	332
12 More Tools and Problems	335
12.1 Horizons and the Horizon Problem	335
12.2 Pedagogical Interlude—Space-Time Diagrams for the Standard World Models	335
12.2.1 Distance and Times	339
12.2.2 The Past Light Cone	339
12.2.3 The Critical World Model $\Omega_0 = 1, \Omega_A = 0$	342
12.2.4 The Reference World Model $\Omega_0 = 0.3, \Omega_A = 0.7$	344
12.3 Superhorizon Scales	347
12.4 The Adiabatic Baryonic Fluctuations in the Standard Big Bang	350
12.4.1 The Radiation-Dominated Era	352
12.4.2 The Matter-Dominated Era	353
12.5 Dissipation Processes in the Pre-Recombination Era	355
12.6 Isothermal Perturbations	357
12.7 Baryonic Theories of Galaxy Formation	360
12.7.1 The Adiabatic Scenario	361
12.7.2 The Isothermal Scenario	363
12.8 What Went Wrong?	364

XVIII Contents

13 Dark Matter and Galaxy Formation	367	I
13.1 Introduction	367	I
13.2 Forms of Non-Baryonic Dark Matter	369	I
13.3 WIMPs as Dark Matter Particles	370	I
13.4 Metric Perturbations and Hot and Cold Dark Matter	374	15.4 I
13.5 Free Streaming and the Damping of Hot Dark Matter Perturbations	375	I
13.6 Instabilities in the Presence of Dark Matter	377	I
13.7 The Evolution of Hot and Cold Dark Matter Perturbations	380	I
13.7.1 Hot Dark Matter Scenario	380	15.5 Ir
13.7.2 Cold Dark Matter Scenario	381	15.6 S
13.8 Conclusion	384	I
14 Correlation Functions and the Spectrum of the Initial Fluctuations	385	I
14.1 The Two-Point Correlation Function for Galaxies	385	15.7 Tl
14.2 The Perturbation Spectrum	388	15.8 Tl
14.2.1 The Relation Between $\xi(r)$ and the Power Spectrum of the Fluctuations	388	15
14.2.2 The Initial Power Spectrum	390	15
14.2.3 The Harrison-Zeldovich Power Spectrum	391	15
14.3 Evolution of the Initial Perturbation Spectrum – Transfer Functions	393	15
14.3.1 Adiabatic Cold Dark Matter	393	15
14.3.2 Adiabatic Hot Dark Matter	396	15.9 De
14.3.3 Isocurvature Cold Dark Matter	396	15.10 Otl
14.3.4 The Subsequent Evolution	399	15.11 Re
14.4 Biasing	401	
14.5 Reconstructing the Processed Initial Power Spectrum	405	
14.5.1 Redshift Biases	406	
14.5.2 Non-Linear Development of Density Perturbations	407	
14.5.3 The Role of Baryon Perturbations	409	
14.6 The Acoustic Peaks in the Power Spectrum of Galaxies	411	
14.6.1 The 2dF Galaxy Redshift Survey	412	
14.6.2 Sloan Digital Sky Survey	412	
14.7 Variations on a Theme of Cold Dark Matter	415	
Part IV The Epoch of Recombination		
16 The Post-Recombinational Epoch		
16.1 The Epoch		
16.2 The Epoch		
16.3 The Epoch		
16.3.1 The Epoch		
16.3.2 The Epoch		
17 The Evolution of the Universe		
17.1 The Epoch		
17.2 The Epoch		
17.2.1 The Epoch		
17.2.2 The Epoch		
17.2.3 The Epoch		
17.2.4 The Epoch		
17.2.5 The Epoch		
17.2.6 The Epoch		
17.2.7 The Epoch		
17.2.8 The Epoch		
17.2.9 The Epoch		
17.2.10 The Epoch		
17.2.11 The Epoch		
17.2.12 The Epoch		
17.2.13 The Epoch		
17.2.14 The Epoch		
17.2.15 The Epoch		
17.2.16 The Epoch		
17.2.17 The Epoch		
17.2.18 The Epoch		
17.2.19 The Epoch		
17.2.20 The Epoch		
17.2.21 The Epoch		
17.2.22 The Epoch		
17.2.23 The Epoch		
17.2.24 The Epoch		
17.2.25 The Epoch		
17.2.26 The Epoch		
17.2.27 The Epoch		
17.2.28 The Epoch		
17.2.29 The Epoch		
17.2.30 The Epoch		
17.2.31 The Epoch		
17.2.32 The Epoch		
17.2.33 The Epoch		
17.2.34 The Epoch		
17.2.35 The Epoch		
17.2.36 The Epoch		
17.2.37 The Epoch		
17.2.38 The Epoch		
17.2.39 The Epoch		
17.2.40 The Epoch		
17.2.41 The Epoch		
17.2.42 The Epoch		
17.2.43 The Epoch		
17.2.44 The Epoch		
17.2.45 The Epoch		
17.2.46 The Epoch		
17.2.47 The Epoch		
17.2.48 The Epoch		
17.2.49 The Epoch		
17.2.50 The Epoch		
17.2.51 The Epoch		
17.2.52 The Epoch		
17.2.53 The Epoch		
17.2.54 The Epoch		
17.2.55 The Epoch		
17.2.56 The Epoch		
17.2.57 The Epoch		
17.2.58 The Epoch		
17.2.59 The Epoch		
17.2.60 The Epoch		
17.2.61 The Epoch		
17.2.62 The Epoch		
17.2.63 The Epoch		
17.2.64 The Epoch		
17.2.65 The Epoch		
17.2.66 The Epoch		
17.2.67 The Epoch		
17.2.68 The Epoch		
17.2.69 The Epoch		
17.2.70 The Epoch		
17.2.71 The Epoch		
17.2.72 The Epoch		
17.2.73 The Epoch		
17.2.74 The Epoch		
17.2.75 The Epoch		
17.2.76 The Epoch		
17.2.77 The Epoch		
17.2.78 The Epoch		
17.2.79 The Epoch		
17.2.80 The Epoch		
17.2.81 The Epoch		
17.2.82 The Epoch		
17.2.83 The Epoch		
17.2.84 The Epoch		
17.2.85 The Epoch		
17.2.86 The Epoch		
17.2.87 The Epoch		
17.2.88 The Epoch		
17.2.89 The Epoch		
17.2.90 The Epoch		
17.2.91 The Epoch		
17.2.92 The Epoch		
17.2.93 The Epoch		
17.2.94 The Epoch		
17.2.95 The Epoch		
17.2.96 The Epoch		
17.2.97 The Epoch		
17.2.98 The Epoch		
17.2.99 The Epoch		
17.2.100 The Epoch		
17.2.101 The Epoch		
17.2.102 The Epoch		
17.2.103 The Epoch		
17.2.104 The Epoch		
17.2.105 The Epoch		
17.2.106 The Epoch		
17.2.107 The Epoch		
17.2.108 The Epoch		
17.2.109 The Epoch		
17.2.110 The Epoch		
17.2.111 The Epoch		
17.2.112 The Epoch		
17.2.113 The Epoch		
17.2.114 The Epoch		
17.2.115 The Epoch		
17.2.116 The Epoch		
17.2.117 The Epoch		
17.2.118 The Epoch		
17.2.119 The Epoch		
17.2.120 The Epoch		
17.2.121 The Epoch		
17.2.122 The Epoch		
17.2.123 The Epoch		
17.2.124 The Epoch		
17.2.125 The Epoch		
17.2.126 The Epoch		
17.2.127 The Epoch		
17.2.128 The Epoch		
17.2.129 The Epoch		
17.2.130 The Epoch		
17.2.131 The Epoch		
17.2.132 The Epoch		
17.2.133 The Epoch		
17.2.134 The Epoch		
17.2.135 The Epoch		
17.2.136 The Epoch		
17.2.137 The Epoch		
17.2.138 The Epoch		
17.2.139 The Epoch		
17.2.140 The Epoch		
17.2.141 The Epoch		
17.2.142 The Epoch		
17.2.143 The Epoch		
17.2.144 The Epoch		
17.2.145 The Epoch		
17.2.146 The Epoch		
17.2.147 The Epoch		
17.2.148 The Epoch		
17.2.149 The Epoch		
17.2.150 The Epoch		
17.2.151 The Epoch		
17.2.152 The Epoch		
17.2.153 The Epoch		
17.2.154 The Epoch		
17.2.155 The Epoch		
17.2.156 The Epoch		
17.2.157 The Epoch		
17.2.158 The Epoch		
17.2.159 The Epoch		
17.2.160 The Epoch		
17.2.161 The Epoch		
17.2.162 The Epoch		
17.2.163 The Epoch		
17.2.164 The Epoch		
17.2.165 The Epoch		
17.2.166 The Epoch		
17.2.167 The Epoch		
17.2.168 The Epoch		
17.2.169 The Epoch		
17.2.170 The Epoch		
17.2.171 The Epoch		
17.2.172 The Epoch		
17.2.173 The Epoch		
17.2.174 The Epoch		
17.2.175 The Epoch		
17.2.176 The Epoch		
17.2.177 The Epoch		
17.2.178 The Epoch		
17.2.179 The Epoch		
17.2.180 The Epoch		
17.2.181 The Epoch		
17.2.182 The Epoch		
17.2.183 The Epoch		
17.2.184 The Epoch		
17.2.185 The Epoch		
17.2.186 The Epoch		
17.2.187 The Epoch		
17.2.188 The Epoch		
17.2.189 The Epoch		
17.2.190 The Epoch		
17.2.191 The Epoch		
17.2.192 The Epoch		
17.2.193 The Epoch		
17.2.194 The Epoch		
17.2.195 The Epoch		
17.2.196 The Epoch		
17.2.197 The Epoch		
17.2.198 The Epoch		
17.2.199 The Epoch		
17.2.200 The Epoch		
17.2.201 The Epoch		
17.2.202 The Epoch		
17.2.203 The Epoch		
17.2.204 The Epoch		
17.2.205 The Epoch		
17.2.206 The Epoch		
17.2.207 The Epoch		
17.2.208 The Epoch		
17.2.209 The Epoch		
17.2.210 The Epoch		
17.2.211 The Epoch		
17.2.212 The Epoch		
17.2.213 The Epoch		
17.2.214 The Epoch		
17.2.215 The Epoch		
17.2.216 The Epoch		
17.2.217 The Epoch		
17.2.218 The Epoch		
17.2.219 The Epoch		
17.2.220 The Epoch		
17.2.221 The Epoch		
17.2.222 The Epoch		
17.2.223 The Epoch		
17.2.224 The Epoch		
17.2.225 The Epoch		
17.2.226 The Epoch		
17.2.227 The Epoch		
17.2.228 The Epoch		
17.2.229 The Epoch		
17.2.230 The Epoch		
17.2.231 The Epoch		
17.2.232 The Epoch		
17.2.233 The Epoch		
17.2.234 The Epoch		
17.2.235 The Epoch		
17.2.236 The Epoch		
17.2.237 The Epoch		
17.2.238 The Epoch		
17.2.239 The Epoch		
17.2.240 The Epoch		
17.2.241 The Epoch		
17.2.242 The Epoch		
17.2.243 The Epoch		
17.2.244 The Epoch		
17.2.245 The Epoch		
17.2.246 The Epoch		
17.2.247 The Epoch		
17.2.248 The Epoch		
17.2.249 The Epoch		
17.2.250 The Epoch		
17.2.251 The Epoch		
17.2.252 The Epoch		
17.2.253 The Epoch		
17.2.254 The Epoch		
17.2.255 The Epoch		
17.2.256 The Epoch		
17.2.257 The Epoch		
17.2.258 The Epoch		
17.2.259 The Epoch		
17.2.260 The Epoch		
17.2.261 The Epoch		
17.2.262 The Epoch		
17.2.263 The Epoch		
17.2.264 The Epoch		
17.2.265 The Epoch		
17.2.266 The Epoch		
17.2.267 The Epoch		
17.2.268 The Epoch		
17.2.269 The Epoch		
17.2.270 The Epoch		
17.2.271 The Epoch		
17.2.272 The Epoch		
17.2.273 The Epoch		
17.2.274 The Epoch		
17.2.275 The Epoch		
17.2.276 The Epoch		
17.2.277 The Epoch		
17.2.278 The Epoch		
17.2.279 The Epoch		
17.2.280 The Epoch		
17.2.281 The Epoch		
17.2.282 The Epoch		
17.2.283 The Epoch		
17.2.284 The Epoch		
17.2.285 The Epoch		
17.2.286 The Epoch		
17.2.287 The Epoch		
17.2.288 The Epoch		
17.2.289 The Epoch		
17.2.290 The Epoch		
17.2.291 The Epoch		
17.2.292 The Epoch		
17.2.293 The Epoch		
17.2.294 The Epoch		
17.2.295 The Epoch		
17.2.296 The Epoch		
17.2.297 The Epoch		
17.2.298 The Epoch		
17.2.299 The Epoch		
17.2.300 The Epoch		
17.2.301 The Epoch		
17.2.302 The Epoch		
17.2.303 The Epoch		
17.2.304 The Epoch		
17.2.305 The Epoch		
17.2.306 The Epoch		
17.2.307 The Epoch		
17.2.308 The Epoch		
17.2.309 The Epoch		
17.2.310 The Epoch		
17.2.311 The Epoch		
17.2.312 The Epoch		
17.2.313 The Epoch		
17.2.314 The Epoch		
17.2.315 The Epoch		
17.2.316 The Epoch		
17.2.317 The Epoch		
17.2.318 The Epoch		
17.2.319 The Epoch		
17.2.320 The Epoch		
17.2.321 The Epoch		
17.2.322 The Epoch		
17.2.323 The Epoch		
17.2.324 The Epoch		
17.2.325 The Epoch		
17.2.326 The Epoch		
17.2.327 The Epoch		
17.2.328 The Epoch		
17.2.329 The Epoch		
17.2.330 The Epoch		
17.2.331 The Epoch		
17.2.		

13 Dark Matter and Galaxy Formation	367	Contents	XIX
13.1 Introduction	367		
13.2 Forms of Non-Baryonic Dark Matter	369		
13.3 WIMPs as Dark Matter Particles	370		
13.4 Metric Perturbations and Hot and Cold Dark Matter	374		
13.5 Free Streaming and the Damping of Hot Dark Matter Perturbations	375		
13.6 Instabilities in the Presence of Dark Matter	377		
13.7 The Evolution of Hot and Cold Dark Matter Perturbations	380		
13.7.1 Hot Dark Matter Scenario	380		
13.7.2 Cold Dark Matter Scenario	381		
13.8 Conclusion	384		
14 Correlation Functions and the Spectrum of the Initial Fluctuations	385		
14.1 The Two-Point Correlation Function for Galaxies	385		
14.2 The Perturbation Spectrum	388		
14.2.1 The Relation Between $\xi(r)$ and the Power Spectrum of the Fluctuations	388		
14.2.2 The Initial Power Spectrum	390		
14.2.3 The Harrison-Zeldovich Power Spectrum	391		
14.3 Evolution of the Initial Perturbation Spectrum – Transfer Functions	393		
14.3.1 Adiabatic Cold Dark Matter	393		
14.3.2 Adiabatic Hot Dark Matter	396		
14.3.3 Isocurvature Cold Dark Matter	396		
14.3.4 The Subsequent Evolution	399		
14.4 Biasing	401		
14.5 Reconstructing the Processed Initial Power Spectrum	405		
14.5.1 Redshift Biases	406		
14.5.2 Non-Linear Development of Density Perturbations	407		
14.5.3 The Role of Baryon Perturbations	409		
14.6 The Acoustic Peaks in the Power Spectrum of Galaxies	411		
14.6.1 The 2dF Galaxy Redshift Survey	412		
14.6.2 Sloan Digital Sky Survey	412		
14.7 Variations on a Theme of Cold Dark Matter	415		
15 Fluctuations in the Cosmic Microwave Background Radiation	421		
15.1 The Ionisation of the Intergalactic Gas Through the Epoch of Recombination	422		
15.2 The Physical and Angular Scales of the Fluctuations	424		
15.2.1 The Last Scattering Layer	425		
15.2.2 The Silk Damping Scale	426		
15.2.3 The Sound Horizon at the Last Scattering Layer	427		
15.2.4 The Particle Horizon Scales	429		
15.2.5 Summary	430		
15.3 The Power Spectrum of Fluctuations in the Cosmic Microwave Background Radiation	431		
Part IV The Post-Recombination Universe			
16 The Post-Recombination Era	471		
16.1 The Non-Linear Collapse of Density Perturbations	471		
16.1.1 Isotropic Top-Hat Collapse	472		
16.1.2 The Zeldovich Approximation	475		
16.2 The Role of Dissipation	477		
16.3 The Press-Schechter Mass Function	482		
16.3.1 Exposition – Elementary Theory	482		
16.3.2 Development and Recapitulation	485		
17 The Evolution of Galaxies and Active Galaxies with Cosmic Epoch	491		
17.1 Introduction	491		
17.2 Counts of Galaxies and Active Galaxies	492		
17.2.1 Euclidean Source Counts	493		
17.2.2 Source Counts for the Standard World Models	494		
17.2.3 Submillimetre Counts of Dusty Galaxies	500		
17.2.4 Number Counts in Models with Finite S_A	501		
17.2.5 Fluctuations in the Background Radiation due to Discrete Sources	504		

XX Contents

17.3	The V/V_{\max} or Luminosity-Volume Test	507	
17.4	Background Radiation	510	
17.4.1	Background Radiation and Source Counts	510	19
17.4.2	Evaluating the Background due to Discrete Sources	510	19
17.4.3	The Effects of Evolution – the Case of the Radio Background Emission	512	19
17.5	The Evolution of Active Galaxies with Cosmic Epoch	514	19
17.5.1	Number Counts and V/V_{\max} Tests for Extragalactic Radio Sources	514	
17.5.2	Radio Quiet Quasars	518	
17.5.3	X-Ray Source Counts	524	
17.5.4	X-Ray Clusters of Galaxies	529	
17.6	Infrared and Submillimetre Number Counts	532	
17.7	Counts of Galaxies	537	
17.8	Clusters of Galaxies	543	
18	The Intergalactic Medium	547	
18.1	The Background Emission of and Absorption by the Intergalactic Gas	548	20.
18.2	The Gunn–Peterson Test	549	20.
18.3	The Lyman- α Absorption Clouds	552	20.
18.3.1	The Properties of the Lyman- α Absorption Clouds	552	
18.3.2	The Nature of the Clouds in the Lyman- α Forest	554	
18.3.3	The Evolution of Lyman- α Absorption Clouds with Cosmic Epoch	556	
18.3.4	The Power Spectrum of the Lyman- α Forest	557	
18.4	The Lukewarm Intergalactic Gas	560	
18.4.1	The X-Ray Background and a Cosmic Conspiracy	561	20.
18.4.2	The Collisional Excitation of the Intergalactic Gas	562	20.
18.4.3	The Emission and Absorption of Diffuse Lukewarm Intergalactic Gas	564	
18.4.4	The Proximity Effect and the Diffuse Ultraviolet Background Radiation at Large Redshifts	567	
18.5	The Lyman Continuum Opacity of the Intergalactic Gas	569	
18.6	Modelling the Evolution of the Intergalactic Medium	571	
18.7	The Epoch of Reionisation	574	
18.8	The Origin of Magnetic Fields	577	
19	Making Real Galaxies	583	
19.1	Star and Element Formation in Galaxies	583	
19.1.1	The Background Radiation and Element Formation	584	
19.1.2	The Global Star Formation Rate from Optical and Ultraviolet Observations of Star-Forming Galaxies	587	
19.1.3	The Lyman-Break Galaxies	590	
19.1.4	The Hubble Deep and Ultra Deep Fields	590	
			References ..
			Name Index ..
			Index ..

	Contents	xxi
17.3	The V/V_{\max} or Luminosity-Volume Test	507
17.4	Background Radiation	510
17.4.1	Background Radiation and Source Counts	510
17.4.2	Evaluating the Background due to Discrete Sources	510
17.4.3	The Effects of Evolution – the Case of the Radio Background Emission	512
17.5	The Evolution of Active Galaxies with Cosmic Epoch	514
17.5.1	Number Counts and V/V_{\max} Tests for Extragalactic Radio Sources	514
17.5.2	Radio Quiet Quasars	518
17.5.3	X-Ray Source Counts	524
17.5.4	X-Ray Clusters of Galaxies	529
17.6	Infrared and Submillimetre Number Counts	532
17.7	Counts of Galaxies	537
17.8	Clusters of Galaxies	543
18	The Intergalactic Medium	547
18.1	The Background Emission and Absorption by the Intergalactic Gas	548
18.2	The Gunn-Peterson Test	549
18.3	The Lyman- α Absorption Clouds	552
18.3.1	The Properties of the Lyman- α Absorption Clouds	552
18.3.2	The Nature of the Clouds in the Lyman- α Forest	554
18.3.3	The Evolution of Lyman- α Absorption Clouds with Cosmic Epoch	556
18.3.4	The Power Spectrum of the Lyman- α Forest	557
18.4	The Lukewarm Intergalactic Gas	560
18.4.1	The X-Ray Background and a Cosmic Conspiracy	561
18.4.2	The Collisional Excitation of the Intergalactic Gas	562
18.4.3	The Emission and Absorption of Diffuse Lukewarm Intergalactic Gas	564
18.4.4	The Proximity Effect and the Diffuse Ultraviolet Background Radiation at Large Redshifts	567
18.5	The Lyman Continuum Opacity of the Intergalactic Gas	569
18.6	Modelling the Evolution of the Intergalactic Medium	571
18.7	The Epoch of Reionisation	574
18.8	The Origin of Magnetic Fields	577
19	Making Real Galaxies	583
19.1	Star and Element Formation in Galaxies	583
19.1.1	The Background Radiation and Element Formation	583
19.1.2	The Global Star Formation Rate from Optical and Ultraviolet Observations of Star-Forming Galaxies	587
19.1.3	The Lyman-Break Galaxies	590
19.1.4	The Hubble Deep and Ultra Deep Fields	590
	Contents	xxi
	19.1.5 Submillimetre Determinations of the Cosmic Star-Formation Rate	596
	19.2 The Abundances of Elements in Lyman- α Absorption Systems	598
	19.3 The Equations of Cosmic Chemical Evolution	604
	19.4 The Old Red Galaxies	607
	19.5 The Origin of Rotation	610
	19.6 Putting It All Together – Semi-Analytic Models of Galaxy Formation	613
20	The Very Early Universe	621
20.1	The Big Problems	621
20.1.1	The Horizon Problem	621
20.1.2	The Flatness Problem	621
20.1.3	The Baryon-Asymmetry Problem	622
20.1.4	The Primordial Fluctuation Problem	622
20.1.5	The Values of the Cosmological Parameters	623
20.1.6	The Way Ahead	624
20.2	The Limits of Observation	624
20.3	The Anthropic Cosmological Principle	625
20.4	The Inflationary Universe – Historical Background	626
20.5	The Origin of the Spectrum of Primordial Perturbations	629
20.5.1	The Equation of State	630
20.5.2	The Duration of the Inflationary Phase	630
20.5.3	The Shrinking Hubble Sphere	631
20.5.4	Scalar Fields	634
20.5.5	The Quantised Harmonic Oscillator	635
20.5.6	The Spectrum of Fluctuations in the Scalar Field	641
20.6	Baryogenesis	642
20.7	The Planck Era	642
	References	645
	Name Index	695
	Index	701

Part I

Preliminaries

1 A Brief History of Cosmology and Galaxy Formation

We begin with a broad-brush historical overview of the development of ideas and concepts concerning galaxies, cosmology and galaxy formation from the time of Tycho Brahe and Newton to the present day. This chapter therefore summarises qualitatively many of the key topics to be dealt with in quantitative detail in the rest of this book.¹ If you do not need this gentle introduction, or misguidedly think that history is boring, you may pass straightaway to Chap. 2.

1.1 Pre-History

It always comes as a surprise to me to realise how recent our understanding of galaxies, cosmology and galaxy formation really is. The motions of the Sun, Moon and planets against the background of the ‘fixed stars’ had been studied from ancient times, but the scientific study of their motions in the modern sense only began in the sixteenth century.

The developments which led to the Newtonian revolution can be traced to the technological and observational achievements of Tycho Brahe in the final decades of the sixteenth century. I have told this remarkable story elsewhere (Longair, 2003). Tycho Brahe measured the positions of the Sun, Moon, planets and 777 stars over a period of 20 years, resulting in an order of magnitude improvement in the accuracy with which their orbits were determined over all previous measurements. In the year before his death, he employed Johannes Kepler as his assistant and assigned him the task of working out the orbits of the planets from his magnificent data sets. In the period 1601 to 1619, Kepler succeeded in interpreting the mass of Tycho’s data in terms of elliptical planetary orbits about the Sun, which was located in one of the foci of each ellipse. Kepler’s discovery of his three laws of planetary motion was a miracle of geometrical analysis. The three laws embody not only the elliptical orbits of the planets (the first law), but also the areal law – that equal areas are swept out by the radius vector from the Sun to the planet in equal times (the second law) – and the dependence of the period T of the planet’s orbit about the Sun upon the three-halves power of its mean distance r from the Sun, $T \propto r^{3/2}$, (the third law).

¹ Comprehensive references to the original papers discussed in this chapter can be found in my book *The Cosmic Century: A History of Astrophysics and Cosmology* (Longair, 2006).

In 1664, at the age of only 21, Isaac Newton first derived his law of gravity from Kepler's third law. Writing *Newton's law of gravity* in modern vector notation,

$$\mathbf{f} = -G \frac{M_1 M_2}{r^2} \mathbf{i}_r, \quad (1.1)$$

where \mathbf{f} is the gravitational force acting between two point masses M_1 and M_2 separated by a distance r and \mathbf{i}_r is the unit vector in the direction from one mass to the other. Newton's achievement was to unify the terrestrial law of gravity with celestial dynamics by showing that (1.1) could explain the acceleration of falling apples on Earth and the orbits of the planets. It is no exaggeration to say that astronomy, astrophysics and cosmology are the *sciences of gravity* – all the systems we study in astronomy and cosmology are attempting to counteract the omnipresent attractive force of gravity by one means or another.

In 1692 Richard Bentley gave the first series of Boyle Lectures which Robert Boyle had founded 'to combat atheism'. Bentley took as his theme Newton's 'sublime discoveries' and entered into a short but profound correspondence with Newton about the nature of our physical Universe. The question at issue was the stability of a finite or infinite Universe filled with stars under the attractive force of gravity. The conclusion of the correspondence was that the Universe must be infinite because, if it were not, it would collapse to its centre under gravity. With remarkable insight, they recognised, however, that an infinite Universe filled with stars is gravitationally unstable. If a star is displaced from its equilibrium position, it continues to accelerate in that direction. To quote Harrison:

(Newton) agreed with Bentley that providence had designed a universe of infinite extent in which uniformly distributed stars stand poised in unstable equilibrium like needles on their points (Harrison, 1987).

It was only in the twentieth century that the nature of this instability was fully appreciated. For a static medium, the instability criterion and the growth rate of the instability were derived by James Jeans in 1902 and the corresponding results for an expanding medium by Georges Lemaître, Howard Robertson and Evgenii Lifshitz in the 1930s and 1940s. Their results are central to the understanding of the problems of the formation of structure in the Universe and the modern working out of their basic insights will dominate much of the discussion throughout this book.

As part of the dialogue with Bentley, Newton proposed that the stars are objects like the Sun and he made star counts in an attempt to show that the stars are indeed uniformly distributed in space. From the seventeenth century onwards, most astronomers assumed that the stars are objects similar to the Sun, but at vastly greater distances. The problem was to find means of measuring their distances. If they were assumed to have the same intrinsic luminosities as the Sun, the inverse square law could be used to estimate distances by comparing the relative brightnesses of the Sun and the distant stars. The technical problem was that the Sun is so much brighter than the brightest stars that it was difficult to make good estimates of the ratio of their observed flux densities, or apparent magnitudes. An ingenious solution was discovered in 1668 by James Gregory, who used Jupiter as an intermediate

lumin
disc
magn
of Si
same
estim.
Th
the st
stellar
meas
measu
and b
Sun.

1.2 T

In his
Galile
observ
cosmo
publis
system
Hypot
the 'Di
in 176
models
of the
cosmol

To v
astrono
detail c
of the i
assumi
picture
with di
(Fig. 1.

John
a fixed l
1767, M
binary s
of stars
absolute
bright st
produce

In 1664, at the age of only 21, Isaac Newton first derived his law of gravity from Kepler's third law. Writing *Newton's law of gravity* in modern vector notation,

$$\mathbf{f} = -G \frac{M_1 M_2}{r^2} \mathbf{i}_r, \quad (1.1)$$

where \mathbf{f} is the gravitational force acting between two point masses M_1 and M_2 separated by a distance r and \mathbf{i}_r is the unit vector in the direction from one mass to the other. Newton's achievement was to unify the terrestrial law of gravity with celestial dynamics by showing that (1.1) could explain the acceleration of falling apples on Earth and the orbits of the planets. It is no exaggeration to say that astronomy, astrophysics and cosmology are the *sciences of gravity* – all the systems we study in astronomy and cosmology are attempting to counteract the omnipresent attractive force of gravity by one means or another.

In 1692 Richard Bentley gave the first series of Boyle Lectures which Robert Boyle had founded 'to combat atheism'. Bentley took as his theme Newton's 'sublime discoveries' and entered into a short but profound correspondence with Newton

about the nature of our physical Universe. The question at issue was the stability of a finite or infinite Universe filled with stars under the attractive force of gravity. The conclusion of the correspondence was that the Universes must be infinite because, if it were not, it would collapse to its centre under gravity. With remarkable insight, they recognised, however, that an infinite Universe filled with stars is gravitationally unstable. If a star is displaced from its equilibrium position, it continues to accelerate in that direction. To quote Harrison:

(Newton) agreed with Bentley that providence had designed a universe of infinite extent in which uniformly distributed stars stand poised in unstable equilibrium like needles on their points (Harrison, 1987).

It was only in the twentieth century that the nature of this instability was fully appreciated. For a static medium, the instability criterion and the growth rate of the instability were derived by James Jeans in 1902 and the corresponding results for an expanding medium by Georges Lemaître, Howard Robertson and Evgenii Lifshitz in the 1930s and 1940s. Their results are central to the understanding of the problems of the formation of structure in the Universe and the modern working out of their basic insights will dominate much of the discussion throughout this book.

As part of the dialogue with Bentley, Newton proposed that the stars are objects like the Sun and he made star counts in an attempt to show that the stars are indeed uniformly distributed in space. From the seventeenth century onwards, most astronomers assumed that the stars are objects similar to the Sun, but at vastly greater distances. The problem was to find means of measuring their distances. If they were assumed to have the same intrinsic luminosities as the Sun, the inverse square law could be used to estimate distances by comparing the relative brightnesses of the Sun and the distant stars. The technical problem was that the Sun is so much brighter than the brightest stars that it was difficult to make good estimates of the ratio of their observed flux densities, or apparent magnitudes. An ingenious solution was discovered in 1668 by James Gregory, who used Jupiter as an intermediate

luminosity calibrator, assuming that its light was entirely sunlight reflected from the disc of the planet and that its surface was a perfect reflector. Then, the apparent magnitudes of Jupiter and the bright star Sirius could be compared and the distance of Sirius was found to be about 83,190 astronomical units (Gregory, 1668). The same method was used by John Michell in 1767 using Saturn as an intermediary to estimate a distance of 460,000 astronomical units for Vega, or α Lyrae.

The method of Gregory and Michell depended upon the assumption that all the stars have the same absolute luminosities. The first direct measurements of stellar distances were made by the technique of stellar parallaxes, the first successful measurement being announced in 1838 by Friedrich Bessel for the star 61 Cygni. The measurement of stellar parallaxes was however difficult and demanding technically and by 1900 only about 100 parallaxes were known for stars in the vicinity of the Sun.

1.2 The Galaxies and the Structure of our Galaxy

In his extraordinary text of 1610, the *Sidereus Nuncius* or *The Sidereal Messenger*, Galileo Galilei demonstrated that the Milky Way can be resolved into stars when observed through the telescope. These observations led to the earliest speculative cosmologies of the modern era. The 'island universe' model of René Descartes, published in *The World* of 1636, involved an interlocking jigsaw puzzle of solar systems. In 1750, Thomas Wright of Durham published *An Original Theory or New Hypothesis of the Universe*, in which the Sun was one of many stars which orbit the 'Divine Centre' of the star system. Immanuel Kant in 1755 and Johann Lambert in 1761 took these ideas further and developed the first hierarchical, or fractal, models of the Universe. Kant also made the prescient suggestion that the flattening of these 'island universes' was due to their rotation. The problem with these early cosmologies was that they lacked observational validation.

Towards the end of the eighteenth century, William Herschel was one of the first astronomers to attempt to define the distribution of stars in the Universe in some detail on the basis of careful astronomical observation. To determine the structure of the Milky Way, he counted the numbers of stars in different directions. Then, assuming that they all have the same intrinsic luminosities, he derived his famous picture for the structure of our Galaxy which consisting of a flattened disc of stars with diameter about five times its thickness, the Sun being located close to its centre (Fig. 1.1) (Herschel, 1785).

John Michell had already warned Herschel that the assumption that the stars have a fixed luminosity was a poor approximation. In his remarkable pioneering paper of 1767, Michell introduced statistical methods into astronomy in order to show that binary and star clusters must be real physical systems and not random associations of stars on the sky (Michell, 1767). Consequently, there must be a dispersion in the absolute luminosities of the stars from their observed range of apparent magnitudes in bright star clusters, such as the Pleiades. Despite this warning, Herschel proceeded to produce a number of different versions of his model for the structure of our Galaxy,

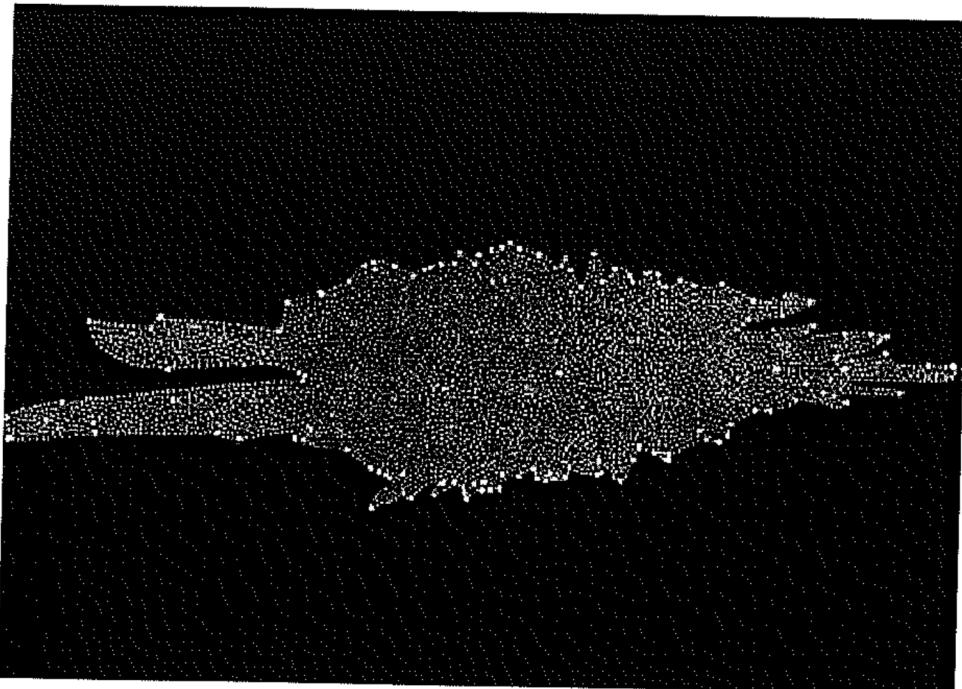


Fig. 1.1. William Herschel's model of the Galaxy based upon star counts in different directions. The Sun is located close to the centre of the disc of stars (Herschel, 1785)

adding appendages to account for various features of the star counts in different directions.

In 1802, Herschel measured the magnitudes of visual binary stars and was forced to agree with Michell's conclusion about the wide dispersion in the luminosities of the stars (Herschel, 1802). Equally troubling was the fact that observations with his magnificent 40-foot telescope showed that, the fainter he looked, the more stars he continued to find. There seemed to be no edge to the Galaxy and Herschel gradually lost faith in his model. In addition, the importance of interstellar extinction by dust was not appreciated – it was only in the 1930s that its central importance for studies of our own and other galaxies was fully appreciated.

Even before the discovery of the telescope, it had been realised that there exist 'nebulous' objects which differ from the stars in having a diffuse or fuzzy appearance. Kant, Lambert, Swedenborg and Wright argued that these objects were 'island universes' similar to the Milky Way, but too distant to be resolved into stars. There was, however, no observational basis for this hypothesis. Herschel also inferred that the nebulae were island universes similar to our Galaxy. A test of this picture was to show that the nebulae could be resolved into stars and he believed that this had been achieved in a number of cases. In others, he assumed that the nebulae were too distant to be resolved into individual stars. This picture came into question, however, when he discovered that, among the nebulae were the planetary nebulae, which consist of a central star surrounded by a shell of gas. Herschel recognised

that these
'luminous'

The catalogues
interest
to enable
contains
galactic
numbers.

The...
and his...
century...
publicatio...
*Clusters o...
observatio...
1888, Joh...
*General C...
suppleme...
Objects i...
numbers.**

While...
a mystery.
Huggins'
(Huggins a...
were objec...
were beyon...
could be...
Debate' an...
and, secon...
'island uni...
the history...
theorists (S...
To illus...
the lumino...
directions,...
flattened wi...
size in the C...
Meanwi...
termination...
remarkable...
Clouds (Fig...
nomical dis...
and their di...
Cepheid var...
of the Galax...

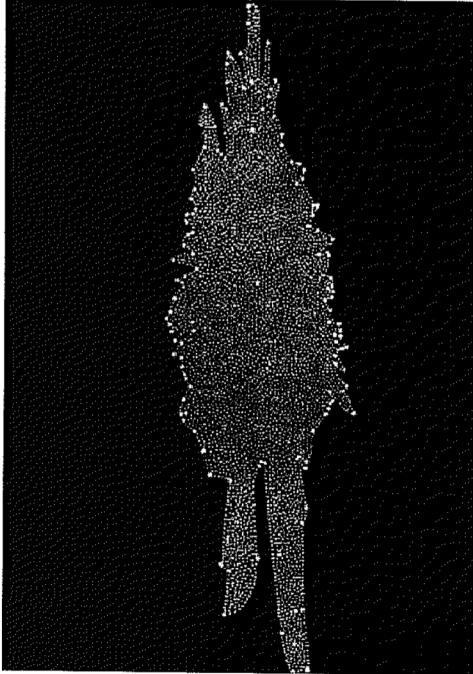


Fig. 1.1. William Herschel's model of the Galaxy based upon star counts in different directions. The Sun is located close to the centre of the disc of stars (Herschel, 1785)

adding appendages to account for various features of the star counts in different directions.

In 1802, Herschel measured the magnitudes of visual binary stars and was forced to agree with Mitchell's conclusion about the wide dispersion in the luminosities of the stars (Herschel, 1802). Equally troubling was the fact that observations with his magnificent 40-foot telescope showed that, the fainter he looked, the more stars he continued to find. There seemed to be no edge to the Galaxy and Herschel gradually lost faith in his model. In addition, the importance of interstellar extinction by dust was not appreciated – it was only in the 1930s that its central importance for studies of our own and other galaxies was fully appreciated.

Even before the discovery of the telescope, it had been realised that there exist 'nebulos' objects which differ from the stars in having a diffuse or fuzzy appearance. Kant, Lambert, Swedenborg and Wright argued that these objects were 'island universes' similar to the Milky Way, but too distant to be resolved into stars. There was, however, no observational basis for this hypothesis. Herschel also inferred that the nebulae were island universes similar to our Galaxy. A test of this picture was to show that the nebulae could be resolved into stars and he believed that this had been achieved in a number of cases. In others, he assumed that the nebulae were too distant to be resolved into individual stars. This picture came into question, however, when he discovered that, among the nebulae were the planetary nebulae, which consist of a central star surrounded by a shell of gas. Herschel recognised

that these nebulae were unlikely to be resolved into stars but rather consisted of 'luminous fluid' surrounding the central star.

The cataloguing of the bright nebulae was begun by Charles Messier whose catalogue of 109 objects was compiled during the years 1771 to 1784. Messier's interest was primarily in comets and his objective in compiling the catalogue was to enable him to distinguish between diffuse nebulae and comets. The catalogue contains a mixture of what we now know are the brightest Galactic and extra-galactic nebulae and they are still commonly referred to by their Messier, or M, numbers.

The systematic cataloguing of the nebulae was begun by William Herschel and his sister Caroline and was continued through the first half of the nineteenth century by his son John Herschel. The results of these huge endeavours was the publication by John Herschel in 1864 of the *General Catalogue of Nebulae and Clusters of Stars* containing 5079 objects. These catalogues were based upon visual observations long before photography became a standard tool of the astronomer. In 1888, John Dreyer published an expanded catalogue which was known as the *New General Catalogue of Nebulae and Clusters of Stars* which, together with the two supplementary *Index Catalogues* of 1895 and 1908, contain some 15,000 objects. Objects in these catalogues are still commonly referred to by their NGC or IC numbers.

While the cataloguing of the nebulae proceeded apace, their nature remained a mystery. Undoubtedly, some of them were gas clouds, as demonstrated by William Huggins' pioneering spectroscopic observations of diffuse nebulae in the 1860s (Huggins and Miller, 1864). The big question was whether or not the spiral nebulae were objects within our own Galaxy or were more distant systems. These nebulae were beyond the distances at which conventional techniques of distance measurement could be used. This problem culminated in what became known as 'The Great Debate' and concerned two related issues. Firstly, what is the size of our own Galaxy and, secondly, are the spiral nebulae members of our Galaxy or are they separate 'island universes', well beyond the confines of our Galaxy? This key episode in the history of modern astronomy should be required reading for all observers and theorists (Sandage, 1961b; Hoskin, 1976; Smith, 1982; Trimble, 1995).

To illustrate the nature of the problem, by 1920, Jacobus Kapteyn had determined the luminosity function of stars near the Sun and so, from star counts in different directions, determined the structure of the Galaxy which he found to be highly flattened with dimensions 1500 pc perpendicular to the plane and about 8 times that size in the Galactic plane (Fig. 1.2) (Kapteyn, 1922).

Meanwhile, Harlow Shapley had adopted a quite different approach to the determination of Galactic structure. In 1912, Henrietta Leavitt had discovered the remarkable period-luminosity relation for Cepheid variable stars in the Magellanic Clouds (Fig. 1.3). This discovery provided a powerful means of measuring astronomical distances because the Cepheid variables are intrinsically luminous stars and their distinctive light curves can be recognised in stars in distant systems. The Cepheid variables were the tools used by Harlow Shapley to determine the structure of the Galaxy through his studies of globular clusters. He found the scale of the

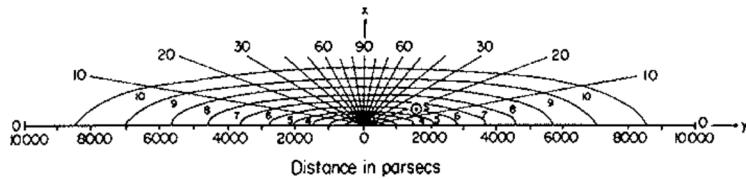


Fig. 1.2. Kapteyn's model for the distribution of stars in the Galaxy (Kapteyn, 1922). The diagram shows the distribution of stars in a plane perpendicular to the Galactic plane. The curves are lines of constant number density of stars and are in equal logarithmic steps. The Sun S is slightly displaced from the centre of the system

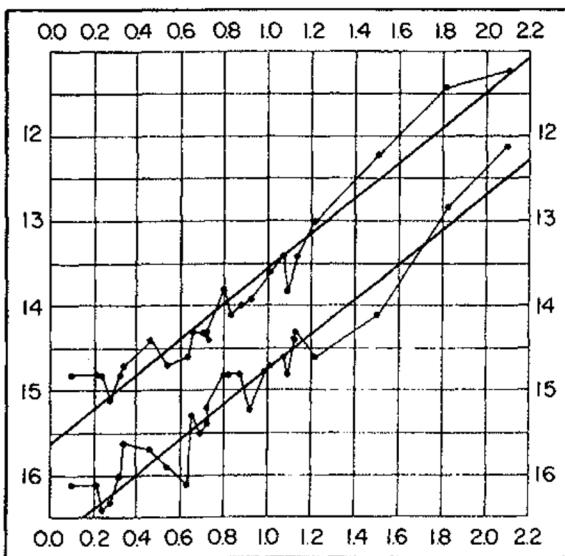


Fig. 1.3. A plot of the period-luminosity relation for the 25 Cepheid variables discovered by Leavitt in the Small Magellanic Cloud (Leavitt, 1912). The *upper locus* is found for the maximum light of the Cepheid variables and the *lower line* for their minimum brightnesses

globular cluster system to be enormous, the most distant globular cluster having a distance of 67 kpc. Furthermore, the globular cluster system was not centred upon the Solar System, but rather most of the globular clusters were found in a direction centred upon the constellation of Sagittarius (Fig. 1.4) (Shapley, 1918).

The course of the debate was complex, but the issues were resolved finally and conclusively in 1925 by Edwin Hubble's observations of Cepheid variables in the Andromeda Nebula. Using the period-luminosity relation for Cepheid variables, he established to everyone's satisfaction that the spiral nebulae are distant extragalactic systems.

Within a year, Hubble had published the first major survey of the properties of galaxies as extragalactic systems. In his remarkable paper (Hubble, 1926), he introduced an early version of his classification of galaxies into ellipticals, spirals and irregulars, estimated mass-to-light ratios for these different types of galaxies, used number counts of galaxies to show that they are uniformly distributed in space and hence estimated the mean density of matter in the Universe in the form of galaxies. Adopting Einstein's static model of the Universe, he found that the radius of curvature of its spherical geometry was 27,000 Mpc. He estimated that, with the

Fig. 1.4. A measure of the size and concentration of the globular cluster system located at the center of the cluster system

100-inch
radius of

... it ma
unive

This paper
to learn th
200-inch
\$6 million
which wa

In 1925
showed th
that their r
(Fig. 1.5a)
from such
had exten
apparent m
indicators
to as Hubi
was that,
that the w

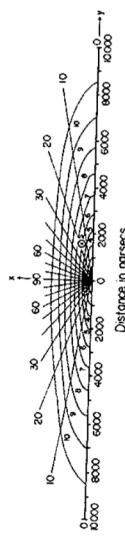


Fig. 1.2. Kapteyn's model for the distribution of stars in the Galaxy (Kapteyn, 1922). The diagram shows the distribution of stars in a plane perpendicular to the Galactic plane. The curves are lines of constant number density of stars and are in equal logarithmic steps. The Sun S is slightly displaced from the centre of the system

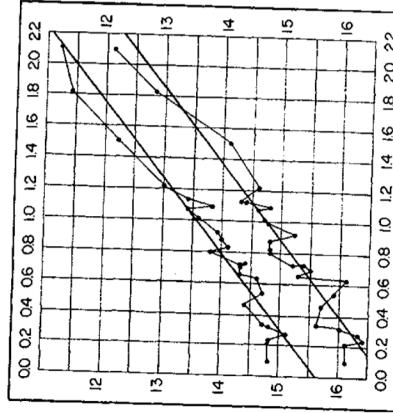


Fig. 1.3. A plot of the period-luminosity relation for the 25 Cepheid variables discovered by Leavitt in the Small Magellanic Cloud (Leavitt, 1912). The upper locus is found for the maximum light of the Cepheid variables and the lower line for their minimum brightnesses

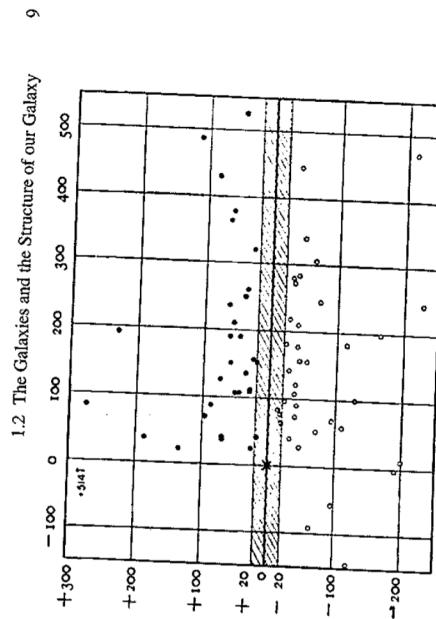


Fig. 1.4. The distribution of globular clusters in the Galaxy according to Shapley's distance measurements (Shapley, 1918). The scales on the abscissa and ordinate are in units of 100 pc and correspond to distances in and perpendicular to the Galactic plane respectively. The Sun, located at zero coordinates on the abscissa and ordinate, lies towards one edge of the globular cluster system

globular cluster system to be enormous, the most distant globular cluster having a distance of 67 kpc. Furthermore, the globular cluster system was not centred upon the Solar System, but rather most of the globular clusters were found in a direction centred upon the constellation of Sagittarius (Fig. 1.4) (Shapley, 1918).

The course of the debate was complex, but the issues were resolved finally and conclusively in 1925 by Edwin Hubble's observations of Cepheid variables in the Andromeda Nebula. Using the period-luminosity relation for Cepheid variables, he established to everyone's satisfaction that the spiral nebulae are distant extragalactic systems.

Within a year, Hubble had published the first major survey of the properties of galaxies as extragalactic systems. In his remarkable paper (Hubble, 1926), he introduced an early version of his classification of galaxies into ellipticals, spirals and irregulars, estimated mass-to-light ratios for these different types of galaxies, used number counts of galaxies to show that they are uniformly distributed in space and hence estimated the mean density of matter in the Universe in the form of galaxies. Adopting Einstein's static model of the Universe, he found that the radius of curvature of its spherical geometry was 27,000 Mpc. He estimated that, with the

100-inch Hooker telescope, he could observe typical galaxies to about 1/600 of the radius of the Universe. He concluded with the remark that

... with reasonable increases in the speed of plates and sizes of telescopes, it may become possible to observe an appreciable fraction of the Einstein universe.

This paper marked the beginning of extragalactic astronomy. It comes as no surprise to learn that George Ellery Hale began his campaign to raise funds for the Palomar 200-inch telescope in 1928 – before the year was out, he had secured a grant of \$6 million from the Rockefeller Foundation for the telescope, the construction of which was completed in 1949.

In 1929, Hubble made his second fundamental contribution to cosmology. He showed that the extragalactic nebulae are all moving away from our own Galaxy and that their recessional velocities v are proportional to their distances r from our Galaxy (Fig. 1.5a) (Hubble, 1929). It is remarkable that he was able to deduce this key result from such a small sample of nearby galaxies but, within five years, he and Humason had extended the relation to very much greater velocities and distances using the apparent magnitudes of the fifth brightest members of clusters of galaxies as distance indicators (Fig. 1.5b). The velocity-distance relation $v = H_0 r$ is commonly referred to as *Hubble's law* and H_0 as *Hubble's constant*. The significance of this discovery was that, combined with the isotropy of the Universe, Hubble's law demonstrates that the whole system of galaxies is partaking in a uniform expansion.

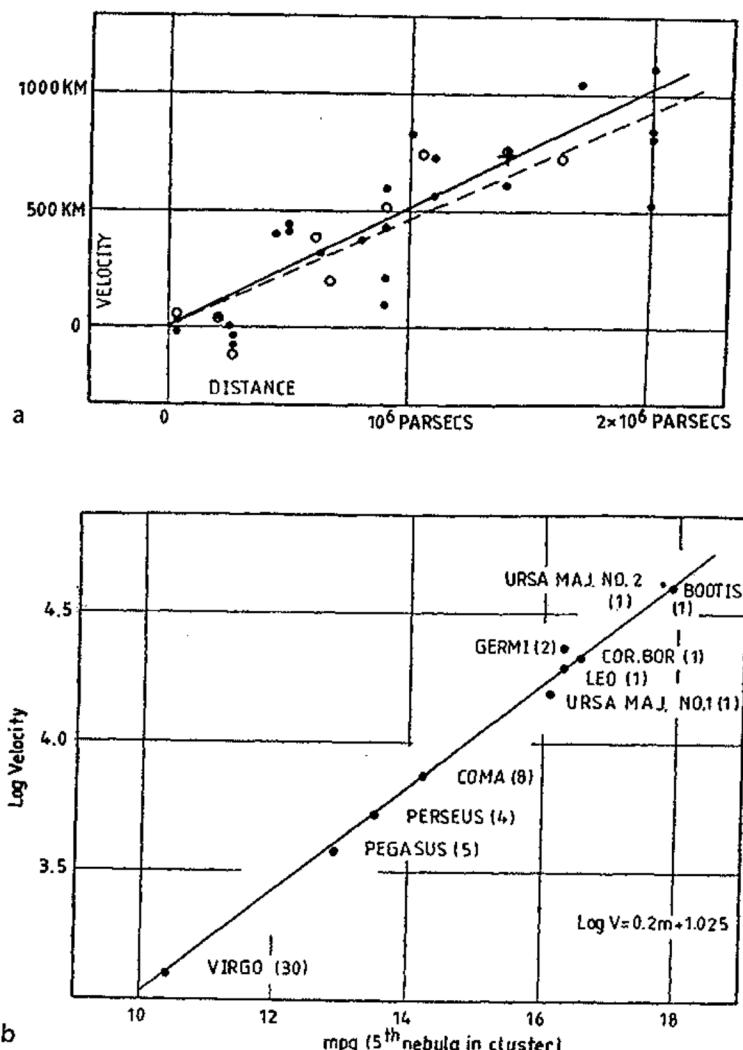


Fig. 1.5. a Hubble's first velocity-distance relation for nearby galaxies (Hubble, 1929). The filled circles and the full line represent a solution for the solar motion using the nebulae individually; the open circles and the dashed line represent a solution combining the nebulae into groups. The cross is an estimate of the mean distance of the other 20 galaxies for which radial velocities were available. b The velocity-apparent magnitude relation for the fifth brightest member of clusters of galaxies, corrected for galactic obscuration (Hubble and Humason, 1934). Each cluster velocity is the mean of the various individual velocities observed in the cluster, the number being indicated by the figure in brackets.

1.3 The Theory of the Expanding Universe

In Newton's *Principia Mathematica*, he emphatically took the position that all motion takes place with respect to a system of absolute space and time. He fiercely rejected the idea that the motion of a body could only be described relative to those of

other bodies, and others by. The issue was relative to one of reference parlance. This which is fixed principle to t

During th seriously by parallel lines a self-consist be Euclidean Carl Friedrich aware of the measuring the Hoherhagen a of the state of the data himsel triangle was 1

The father rector of Kaz then part of H existence of no not be deduced In his papers e an astronomer minimum para

where a is the geometry. He using an obser a prescient stat remarked:

There is n exactness

Non-Euclidean mann, who also was introduced Until Albert Ein of the geometr of the Universe

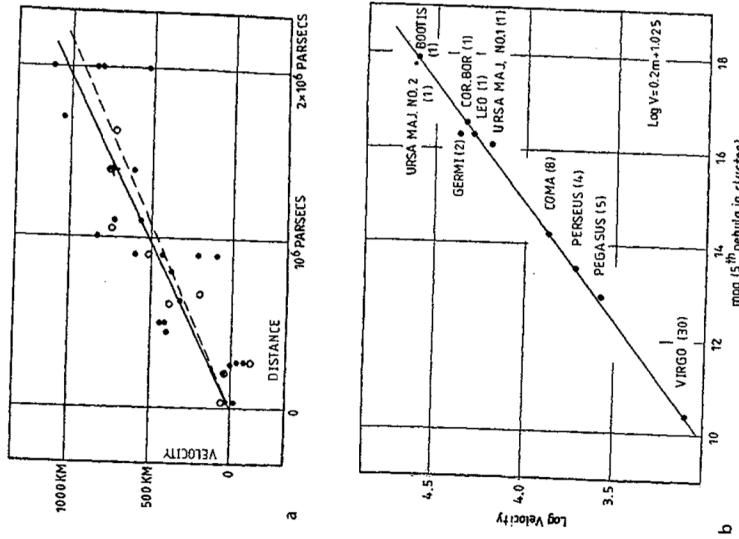


Fig. 1.5. a Hubble's first velocity-distance relation for nearby galaxies (Hubble, 1929). The *filled circles* and the *full line* represent a solution for the solar motion using the nebulae individually; the *open circles* and the *dashed line* represent a solution combining the nebulae into groups. The cross is an estimate of the mean distance of the other 20 galaxies for which radial velocities were available. b The velocity-apparent magnitude relation for the fifth brightest member of clusters of galaxies, corrected for galactic obscuration (Hubble and Humason, 1934). Each cluster velocity is the mean of the various individual velocities observed in the cluster, the number being indicated by the figure in brackets.

1.3 The Theory of the Expanding Universe

In Newton's *Principia Mathematica*, he emphatically took the position that all motion takes place with respect to a system of absolute space and time. He fiercely rejected the idea that the motion of a body could only be described relative to those of

other bodies. This position was challenged by Bishop Berkeley, Christiaan Huygens and others but, at least until the late nineteenth century, Newton's view prevailed. The issue was revived by Ernst Mach who argued that motion can only be defined relative to other bodies. Specifically, he took the view that the local inertial frame of reference is determined by the frame of the distant stars, or galaxies in modern parlance. Thus, a freely swinging Foucault pendulum swings in a reference frame which is fixed relative to the distant galaxies. Albert Einstein gave the name *Mach's principle* to this idea.

During the late eighteenth century, non-Euclidean geometries began to be taken seriously by mathematicians who realised that the fifth postulate of Euclid, that parallel lines meet only at infinity, might not be essential for the construction of a self-consistent geometry. Proposals that the global geometry of space might not be Euclidean were discussed by Girolamo Saccheri and Johann Lambert. In 1816, Carl Friedrich Gauss repeated this proposal in a letter to Christian Gerling and was aware of the fact that a test of the local geometry of space could be carried out by measuring the sum of the angles of a triangle between three high peaks, the Brocken, Hoherhagen and Inselberg. In 1818, Gauss was asked to carry out a geodetic survey of the state of Hanover and he devoted a large effort to carrying out and reducing the data himself. He was certainly aware of the fact that the sum of the angles of the triangle was 180° within the limits of geodetic measurements.

The fathers of non-Euclidean geometry were Nikolai Lobachevsky, who became rector of Kazan University in Russia in 1827, and János Bolyai in Transylvania, then part of Hungary. In the 1820s, they independently solved the problem of the existence of non-Euclidean geometries and showed that Euclid's fifth postulate could not be deduced from the other postulates (Lobachevsky, 1829; Bolyai, 1832). In his papers entitled *On the Principles of Geometry*, Lobachevsky also proposed an astronomical test of the geometry of space. If the geometry were hyperbolic, the minimum parallax of any object would be

$$\theta = \arctan \left(\frac{a}{R} \right), \quad (1.2)$$

where a is the radius of the Earth's orbit and R the radius of curvature of the geometry. He found a minimum value of $R \geq 1.66 \times 10^5$ AU = 2.6 light years, using an observational upper limit of 1 arcsec for the parallax of bright stars. In a prescient statement which will warm the hearts of observational astronomers, he remarked:

There is no means other than astronomical observations for judging the exactness which attaches to the calculations of ordinary geometry.

Non-Euclidean geometries were placed on a firm theoretical basis by Bernhard Riemann, who also discovered closed spherical geometries. The English-speaking world was introduced to these ideas through the works of William Clifford and Arthur Cayley. Until Albert Einstein's discovery of the General Theory of Relativity, considerations of the geometry of space and the role of gravity in defining the large-scale structure of the Universe were separate questions. After 1915, they were inextricably linked.

In that year, after a titanic intellectual struggle, Einstein discovered the definitive version of his General Theory of Relativity which describes how space-time is distorted by the presence of matter and how, in turn, matter moves along trajectories in bent space-time (Einstein, 1915, 1916). For the first time, a relativistic theory of gravity was available which enabled self-consistent models of the Universe as a whole to be constructed and, characteristically, Einstein did not hesitate to do so.

In seeking a solution of his field equations for the Universe as a whole, Einstein had explicitly in mind that Mach's principle should be incorporated into any model of the large-scale structure of the Universe. He had, however, a major problem. Without modification, the field equations predicted that the Universe was unstable. He could only find static solutions by introducing what is now known as the *cosmical* or *cosmological constant* λ , which appears as a constant in Einstein's field equations. In his great paper of 1917, Einstein showed that the introduction of the cosmological constant resulted in static solutions for the Universe as a whole which had closed, spherical geometry and a finite size (Einstein, 1917). He also believed that he had incorporated Mach's principle into General Relativity, in the sense that no solution of the equations would exist if there were no matter present. In the same year, this was, however, shown to be incorrect by Willem de Sitter, who found solutions of the equations even if there were no matter present in the Universe (de Sitter, 1917).

For many decades, the status of the cosmological constant was the subject of debate. In 1919, Einstein realised that a term involving the cosmological constant would appear in the field equations of General Relativity, quite independent of its cosmological significance (Einstein, 1919). In the derivation of the field equations, the λ -term appears as a constant of integration which is normally set equal to zero in the development of standard General Relativity. Einstein was not enthusiastic about the term, remarking that it 'detracts from the formal beauty of the theory'. Willem de Sitter wrote in 1919 that the term

... detracts from the symmetry and elegance of Einstein's original theory, one of whose chief attractions was that it explained so much without introducing any new hypotheses or empirical constant.

Others regarded it as a constant which appears in the development of the General Relativity and its value should be determined by astronomical observation.

The irony of the situation is that this debate took place *before* it was realised that the Universe is in fact non-stationary. In 1922, Aleksander Friedman published the first of two classic papers in which he discovered both static and expanding solutions of Einstein's field equations. In the first paper, Friedman found solutions for expanding universes with closed spatial geometries, including those which expand to a maximum radius and eventually collapse to a singularity (Friedman, 1922). In the second paper of 1924, he showed that there exist expanding solutions which are unbounded and which have hyperbolic geometry (Friedman, 1924). These solutions correspond exactly to the standard world models of general relativity and are known as the *Friedman world models*.

In 1
nificanc
early da
admitted
independ
of Fried
apprecia

Eins
satisfact
to Georg
regarded
life' (Ga
long how

The
bacl

The cos
Earth w
constant
verse ha
500 km:
than the
dating u
constant
time-sca
It turned
ing revi
eliminat
necessar
tures of
during tl
mologic
from det
Backgro

As th
thrust of
cal and g
its mear
cult pro
twenty-f
of by the

In that year, after a titanic intellectual struggle, Einstein discovered the definitive version of his General Theory of Relativity which describes how space-time is distorted by the presence of matter and how, in turn, matter moves along trajectories in bent space-time (Einstein, 1915, 1916). For the first time, a relativistic theory of gravity was available which enabled self-consistent models of the Universe as a whole to be constructed and, characteristically, Einstein did not hesitate to do so.

In seeking a solution of his field equations for the Universe as a whole, Einstein had explicitly in mind that Mach's principle should be incorporated into any model of the large-scale structure of the Universe. He had, however, a major problem. Without modification, the field equations predicted that the Universe was unstable. He could only find static solutions by introducing what is now known as the *cosmical or cosmological constant* λ , which appears as a constant in Einstein's field equations. In his great paper of 1917, Einstein showed that the introduction of the cosmological constant resulted in static solutions for the Universe as a whole which had closed, spherical geometry and a finite size (Einstein, 1917). He also believed that he had incorporated Mach's principle into General Relativity, in the sense that no solution of the equations would exist if there were no matter present. In the same year, this was, however, shown to be incorrect by Willem de Sitter, who found solutions of the equations even if there were no matter present in the Universe (de Sitter, 1917).

For many decades, the status of the cosmological constant was the subject of debate. In 1919, Einstein realised that a term involving the cosmological constant would appear in the field equations of General Relativity, quite independent of its cosmological significance (Einstein, 1919). In the derivation of the field equations, the λ -term appears as a constant of integration which is normally set equal to zero in the development of standard General Relativity. Einstein was not enthusiastic about the term, remarking that it 'detracts from the formal beauty of the theory'. Willem de Sitter wrote in 1919 that the term

... detracts from the symmetry and elegance of Einstein's original theory, one of whose chief attractions was that it explained so much without introducing any new hypotheses or empirical constant.

Others regarded it as a constant which appears in the development of the General Relativity and its value should be determined by astronomical observation.

The irony of the situation is that this debate took place *before* it was realised that the Universe is in fact non-stationary. In 1922, Aleksander Friedman published the first of two classic papers in which he discovered both static and expanding solutions of Einstein's field equations. In the first paper, Friedman found solutions for expanding universes with closed spatial geometries, including those which expand to a maximum radius and eventually collapse to a singularity (Friedman, 1922). In the second paper of 1924, he showed that there exist expanding solutions which are unbounded and which have hyperbolic geometry (Friedman, 1924). These solutions correspond exactly to the standard world models of general relativity and are known as the *Friedman world models*.

In 1925, Friedman died of typhoid in Leningrad before the fundamental significance of his work was appreciated. The neglect of Friedman's work in these early days is somewhat surprising since Einstein had commented, incorrectly as he admitted, on the first of the two papers in 1923. It was not until Georges Lemaître independently rediscovered the same solutions in 1927, and then became aware of Friedman's papers, that the pioneering nature of Friedman's contributions was appreciated (Lemaître, 1927).

Einstein's field equations without the cosmological constant contain perfectly satisfactory solutions in which the Universe is uniformly expanding. According to George Gamow, when the expansion of the Universe was discovered, Einstein regarded the introduction of the cosmological constant as the biggest blunder of my life' (Gamow, 1970). The cosmological constant was not consigned to oblivion for long however. As Yakov Zeldovich remarked:

The genie is out of the bottle and, once he is out, he is very difficult to put back in again.

The cosmological constant immediately found a rôle in reconciling the age of the Earth with the expansion age of the Universe as given by the inverse of Hubble's constant H_0^{-1} . If the cosmological constant is zero, all Friedman models of the Universe have ages less than H_0^{-1} . At that time, Hubble's estimate of H_0 was about $500 \text{ km s}^{-1} \text{ Mpc}^{-1}$, corresponding to $H_0^{-1} = 2 \times 10^9$ years. This time-scale was less than the age of the Earth as determined by nucleocosmochronology, that is, from dating using long-lived radioactive isotopes. A positive value of the cosmological constant can resolve this discrepancy since its effect is to stretch out the expansion time-scale of the Universe, a picture advocated by Arthur Eddington and Lemaître. It turned out that Hubble's estimate of H_0 was seriously overestimated and, following revisions in the 1950s by Walter Baade and Allan Sandage, this conflict was eliminated. Despite the fact that the cosmological constant appeared to be no longer necessary, it made regular appearances in the literature to account for various features of cosmological data, but these pieces of evidence were not compelling. Then, during the period 1995 to 2005, convincing evidence for a positive value of the cosmological constant was found from studies of very distant Type Ia supernovae and from determinations of the power spectrum of fluctuations in the Cosmic Microwave Background Radiation.

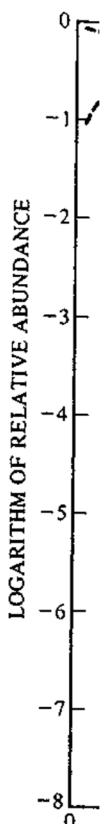
As the standard models of General Relativity became better understood, a major thrust of cosmological research became the determination of the large-scale dynamical and geometrical properties of the Universe - its rate of expansion, its deceleration, its mean density, its geometry and its age. These remained among the most difficult programmes of modern observational cosmology until, in the first years of the twenty-first century, precise estimates became available using techniques undreamt of by the pioneers of geometrical cosmology.

1.4 The Big Bang

The next major advance occurred soon after the Second World War when George Gamow realised that, in an expanding Universe, the early stages must have been very hot indeed – the temperature was so high that the dynamics of the expansion were dominated by the energy density of thermal radiation rather than by its matter content, in other words, the Universe was radiation-dominated. Following an earlier suggestion of Lemaître, he attempted to explain the origin of the chemical elements by *primordial nucleosynthesis*, that is, by nuclear fusion processes as the Universe cooled down from its very hot initial stages. The reasons for adopting this picture were twofold. Firstly, following the work of Cecilia Payne, the abundances of the chemical elements in stars seemed to be remarkably uniform and secondly it was thought that the central temperatures of the stars were not high enough for nucleosynthesis to take place. Gamow's programme was not successful because of the problem of synthesising elements heavier than helium – there are no stable isotopes with atomic mass numbers 5 and 8. Therefore, in the short time-scales available in the hot early phases of the expansion, there was not time to synthesise elements heavier than helium. Gamow's coworkers Ralph Alpher and Robert Herman showed that only deuterium, helium-3 and helium-4 were created in significant quantities (Fig. 1.6) (Alpher and Herman, 1950).

In the course of their calculations, Alpher and Herman worked out the thermal history of the Universe in some detail and predicted that there should be present in the Universe today a diffuse background of black-body radiation with temperature about 5 K, the cooled remnant of its very hot early phases (Alpher and Herman, 1948). The detection of this background radiation was far beyond the capabilities of the technology of the 1940s and the lack of success of Gamow's programme of primordial nucleosynthesis resulted in the neglect of this key prediction for many years. Furthermore, in the 1950s, Fred Hoyle discovered the triple- α resonance, which leads to the formation of carbon from three helium nuclei (Hoyle, 1954). Soon after, he and his colleagues, Margaret Burbidge, Geoffrey Burbidge and William Fowler, showed how the heavy elements could be accounted for by nucleosynthesis in stars (Burbidge et al., 1957).

Interest in what is now referred to as the *Big Bang* model of the Universe grew steadily through the 1950s and early 1960s as evidence was found for cosmological evolutionary effects in the distribution of faint radio sources (Ryle, 1955, 1958). On the theoretical side, interest was rekindled in the question of the synthesis of elements in the early Universe, not now with a view to creating all the elements, but rather to account for the cosmic abundance of helium. By 1964, it was appreciated that, wherever helium could be observed in the Universe, it is present with a very high chemical abundance, about 24% by mass. This figure far exceeded what could be explained by stellar nucleosynthesis. I remember vividly attending a course of post-graduate lectures given by Fred Hoyle in Cambridge in 1964 entitled *Problems of Extragalactic Astrophysics* in which this problem was discussed. During the lecture course, Hoyle, Roger Tayler, and John Faulkner carried out detailed computations of the expected abundance of helium produced by primordial nucleosynthesis. Within



a week c
mass is i
of the pi
Nature i
Robert V
other ele
isotope c
this way
resulted
one tent
abundan

By t
improved
from the
discover
(Penzias
second k
by the C

1.4 The Big Bang

The next major advance occurred soon after the Second World War when George Gamow realised that, in an expanding Universe, the early stages must have been very hot indeed – the temperature was so high that the dynamics of the expansion were dominated by the energy density of thermal radiation rather than by its matter content, in other words, the Universe was radiation-dominated. Following an earlier suggestion of Lemaître, he attempted to explain the origin of the chemical elements by *primordial nucleosynthesis*, that is, by nuclear fusion processes as the Universe cooled down from its very hot initial stages. The reasons for adopting this picture were twofold. Firstly, following the work of Cecilia Payne, the abundances of the chemical elements in stars seemed to be remarkably uniform and secondly it was thought that the central temperatures of the stars were not high enough for nucleosynthesis to take place. Gamow's programme was not successful because of the problem of synthesising elements heavier than helium – there are no stable isotopes with atomic mass numbers 5 and 8. Therefore, in the short time-scales available in the hot early phases of the expansion, there was not time to synthesise elements heavier than helium. Gamow's coworkers Ralph Alpher and Robert Herman showed that only deuterium, helium-3 and helium-4 were created in significant quantities (Fig. 1.6) (Alpher and Herman, 1950).

In the course of their calculations, Alpher and Herman worked out the thermal history of the Universe in some detail and predicted that there should be present in the Universe today a diffuse background of black-body radiation with temperature about 5 K, the cooled remnant of its very hot early phases (Alpher and Herman, 1948). The detection of this background radiation was far beyond the capabilities of the technology of the 1940s and the lack of success of Gamow's programme of primordial nucleosynthesis resulted in the neglect of this key prediction for many years. Furthermore, in the 1950s, Fred Hoyle discovered the triple- α resonance, which leads to the formation of carbon from three helium nuclei (Hoyle, 1954). Soon after, he and his colleagues, Margaret Burbidge, Geoffrey Burbidge and William Fowler, showed how the heavy elements could be accounted for by nucleosynthesis in stars (Burbidge et al., 1957).

Interest in what is now referred to as the *Big Bang* model of the Universe grew steadily through the 1950s and early 1960s as evidence was found for cosmological evolutionary effects in the distribution of faint radio sources (Ryle, 1955, 1958). On the theoretical side, interest was rekindled in the question of the synthesis of elements in the early Universe, not now with a view to creating all the elements, but rather to account for the cosmic abundance of helium. By 1964, it was appreciated that, wherever helium could be observed in the Universe, it is present with a very high chemical abundance, about 24% by mass. This figure far exceeded what could be explained by stellar nucleosynthesis. I remember vividly attending a course of post-graduate lectures given by Fred Hoyle in Cambridge in 1964 entitled *Problems of Extragalactic Astrophysics* in which this problem was discussed. During the lecture course, Hoyle, Roger Taylor, and John Faulkner carried out detailed computations of the expected abundance of helium produced by primordial nucleosynthesis. Within

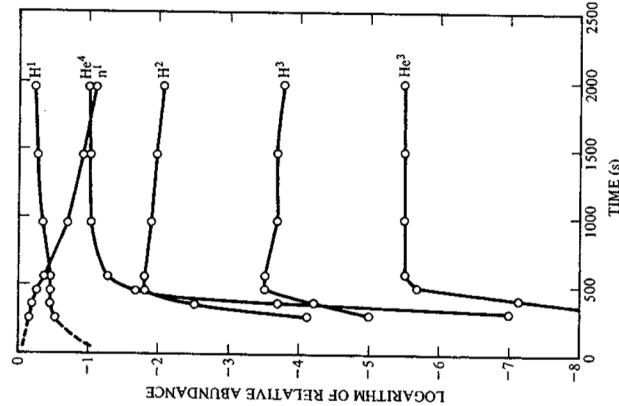


Fig. 1.6. The evolution of the fraction (by number) of the light nuclei in a radiation-dominated Universe, according to calculations by Fermi and Turkevich and published by Alpher and Herman in 1950 (Alpher and Herman, 1950). The models began with 100% of the material in the form of neutrons. The tritium ${}^3\text{H}$ and neutrons shown surviving to 2000 seconds decay radioactively with half-lives of 12.46 years and 10.25 minutes respectively

a week of the topic being raised, they had shown that about 23 to 25% of helium by mass is created by this process and that the percentage is remarkably independent of the precise initial conditions. The paper by Hoyle and Tayler was published in *Nature* in 1964 (Hoyle and Tayler, 1964). Subsequent more detailed calculations by Robert Wagoner, Fowler and Hoyle confirmed these conclusions and suggested that other elements which are difficult to account for by stellar nucleosynthesis, the light isotope of helium, ${}^3\text{He}$, deuterium, D, and lithium, ${}^7\text{Li}$, could also be accounted for in this way (Wagoner et al., 1967). Equally important, the success of these computations resulted in an upper limit to the mean baryon mass density of the Universe of about one tenth the critical density – if the density were any higher, less than the observed abundances of deuterium and helium-3 would be created primordially.

By the early 1960s, as the sensitivity of receivers for centimetre wavelengths improved, it became feasible to search for the cool background radiation left over from the early stages of the Big Bang. The predicted remnant of the Big Bang was discovered, more or less by accident, by Arno Penzias and Robert Wilson in 1965 (Penzias and Wilson, 1965). The *Cosmic Microwave Background Radiation* was the second key discovery of twentieth century observational cosmology. Observations by the Cosmic Background Explorer (COBE), launched in 1989, showed that, away

from the Galactic plane, the radiation is uniform over the sky to better than one part in 100,000 on angular scales greater than 7° and that its spectrum is of black-body form with a quite remarkable precision (Smoot et al., 1992; Fixsen et al., 1996). These observations provided compelling evidence that our Universe went through a very hot, dense phase when the matter and radiation were in thermal equilibrium in its early stages.

The upshot of these discoveries was that there were four independent pieces of evidence for the Big Bang picture of the origin and evolution of our Universe. Firstly, the expansion of the distribution of galaxies discovered by Hubble; secondly, the black-body spectrum and isotropy of the Cosmic Microwave Background Radiation; thirdly, the formation of the light elements by primordial nucleosynthesis; and fourthly, the fact that the ages of the oldest stars and nucleochronology ages were of the same order as the expansion age of the Universe. Thus, the Big Bang provided a natural framework within which to tackle the problems of galaxy and structure formation.

1.5 Galaxy and Structure Formation

The Friedman world models are isotropic and homogeneous and so the enormous diversity of structure we observe in the Universe today is absent. The next step in developing more realistic models of the Universe is to include small density perturbations into the homogeneous, isotropic models and study their development under gravity. For the case of a stationary medium, this problem was solved by James Jeans in 1902 (Jeans, 1902). The criterion for collapse is that the size of the perturbation should exceed the *Jeans' length* $\lambda_J = c_s/(G\varrho_0/\pi)^{1/2}$, where c_s is the speed of sound in the medium and ϱ_0 its density. On scales greater than the Jeans' length, the instability grows exponentially. The physical meaning of the instability criterion is that, on large enough scales, the gravitational force of attraction by the matter of the perturbation exceeds the pressure gradients which resist collapse.

The analysis was repeated for the case of an expanding medium in the 1930s by Lemaître and by Richard Tolman for the case of spherically symmetric perturbations (Lemaître, 1933; Tolman, 1934) and the solution for the general case was found by Evgenii Lifshitz in 1946 (Lifshitz, 1946). Lifshitz found that the condition for gravitational collapse is exactly the same as the Jeans' criterion at any epoch but, crucially, the growth-rate of the density perturbations is no longer exponential but only algebraic. For a Universe with the critical density, $\Omega_0 = 1$ or $\varrho_0 = 3H_0^2/8\pi G$, the density contrast $\Delta = \delta\varrho/\varrho$ grows with time as $\Delta \propto t^{2/3}$. The implication of this result is that the fluctuations from which the large-scale structure of the Universe formed cannot have grown from infinitesimal random perturbations. For this reason, Lemaître, Tolman and Lifshitz inferred that galaxies could not have formed by gravitational collapse.

From the early 1960s onwards, other authors took the point of view that the solution to the problem was to include finite perturbations into the model of the early Universe and then follow in detail how their mass spectrum would evolve with time.

The Moscow school of James Peebles at the Institute of structure in the backwards into the horizon scale. Novikov showed galaxies, the density contrast $\Delta = \delta\varrho/\varrho \sim 10^{-3}$ at the epoch (Novikov, 1972), their origin had to be explained.

The discovery of the immediate impact of the gas could be understood if the speed of sound in the gas was zero. There is no energy source for background radiation at the redshift, exactly as at $z \sim 1500$, the temperature there were sufficient to ionize all the intergalactic gas and at earlier epochs, the primordial epoch, the primordial density of recombination of the matter and radiation equals zero.

The coupling between Ray Weymann in 1969 (Weymann, 1969), Zeldovich and Sazanov in 1970, which had been a remarkable class of papers showed that the interaction were maintained as the intergalactic density determined at all epochs. The condition of the Jeans' length was known as the *Jeans radius*.

In 1968, Jose Silk calculated the waves in the radiating matter (Silk, 1968), for clusters of masses less than the critical density of recombination. Consequently, the large-scale structure formed after recombination independently put together.

from the Galactic plane, the radiation is uniform over the sky to better than one part in 100,000 on angular scales greater than 7° and that its spectrum is of black-body form with a quite remarkable precision (Smoot et al., 1992; Fixsen et al., 1996). These observations provided compelling evidence that our Universe went through a very hot, dense phase when the matter and radiation were in thermal equilibrium in its early stages.

The upshot of these discoveries was that there were four independent pieces of evidence for the Big Bang picture of the origin and evolution of our Universe. Firstly, the expansion of the distribution of galaxies discovered by Hubble; secondly, the black-body spectrum and isotropy of the Cosmic Microwave Background Radiation; thirdly, the formation of the light elements by primordial nucleosynthesis; and fourthly, the fact that the ages of the oldest stars and nucleochronology ages were of the same order as the expansion age of the Universe. Thus, the Big Bang provided a natural framework within which to tackle the problems of galaxy and structure formation.

1.5 Galaxy and Structure Formation

The Friedman world models are isotropic and homogeneous and so the enormous diversity of structure we observe in the Universe today is absent. The next step in developing more realistic models of the Universe is to include small density perturbations into the homogeneous, isotropic models and study their development under gravity. For the case of a stationary medium, this problem was solved by James Jeans in 1902 (Jeans, 1902). The criterion for collapse is that the size of the perturbation should exceed the Jeans' length $\lambda_J = c_s / (G\varrho_0/\pi)^{1/2}$, where c_s is the speed of sound in the medium and ϱ_0 its density. On scales greater than the Jeans' length, the instability grows exponentially. The physical meaning of the instability criterion is that, on large enough scales, the gravitational force of attraction by the matter of the perturbation exceeds the pressure gradients which resist collapse.

The analysis was repeated for the case of an expanding medium in the 1930s by Lemaître and by Richard Tolman for the case of spherically symmetric perturbations (Lemaître, 1933; Tolman, 1934) and the solution for the general case was found by Evgenii Lifshitz in 1946 (Lifshitz, 1946). Lifshitz found that the condition for gravitational collapse is exactly the same as the Jeans' criterion at any epoch but, crucially, the growth-rate of the density perturbations is no longer exponential but only algebraic. For a Universe with the critical density, $\Omega_0 = 1$ or $\varrho_0 = 3H_0^2/8\pi G$, the density contrast $\Delta = \delta\varrho/\varrho$ grows with time as $\Delta \propto t^{2/3}$. The implication of this result is that the fluctuations from which the large-scale structure of the Universe formed cannot have grown from infinitesimal random perturbations. For this reason, Lemaître, Tolman and Lifshitz inferred that galaxies could not have formed by gravitational collapse.

From the early 1960s onwards, other authors took the point of view that the solution to the problem was to include finite perturbations into the model of the early Universe and then follow in detail how their mass spectrum would evolve with time.

The Moscow school led by Yakov Zeldovich, Igor Novikov and their colleagues and James Peebles at Princeton pioneered this approach to the study of the development of structure in the Universe. If perturbations on a particular physical scale are tracked backwards into the past, at some large redshift, the scale of the perturbation is equal to the *horizon scale*, that is $r \approx ct$, where r is the age of the Universe. In 1964, Novikov showed that, to form structures on the scales of galaxies and clusters of galaxies, the density perturbations on the scale of the horizon had to have amplitude $\Delta = \delta\varrho/\varrho \sim 10^{-4}$ in order to guarantee the formation of galaxies by the present epoch (Novikov, 1964). These were certainly *not* infinitesimal perturbations and their origin had to be ascribed to processes occurring in the very early Universe.

The discovery of the Cosmic Microwave Background Radiation in 1965 had an immediate impact upon these studies since the thermal history of the pre-galactic gas could be worked out in detail and this was essential in order to determine how the speed of sound, and hence the Jeans' length, varied with cosmic epoch. If there is no energy input into the background radiation, the temperature of the thermal background radiation changes with scale factor a as $T = T_0/a = T_0(1+z)$, where z is redshift, exactly as in the adiabatic expansion of a photon gas. Therefore, at redshifts $z \sim 1500$, the temperature of the radiation was about 4000 K, at which temperature there were sufficient photons in the Wien region of the Planck distribution to ionise all the intergalactic hydrogen. This epoch is referred to as the *epoch of recombination* and at earlier epochs the hydrogen was fully ionised; at a correspondingly earlier epoch, the primordial helium was ionised as well. Somewhat earlier than the epoch of recombination, the inertial mass density of the radiation was equal to the mass density of the matter, $qc_s^2 = a\Gamma^4$ and so, at times earlier than the *epoch of matter and radiation equality*, the dynamics of the Universe were radiation-dominated. The coupling of matter and radiation by electron scattering was worked out by Ray Weymann in 1966 and in much more detail by Zeldovich and Rashid Sunyaev in 1969 (Weymann, 1966; Zeldovich and Sunyaev, 1969). The pioneering papers by Zeldovich and Sunyaev were based upon the theory of induced Compton scattering which had been published by Aleksander Kompaneets in 1956, long after this remarkable classified work had been completed (Kompaneets, 1956). What these papers showed was that, during the radiation-dominated epochs, the matter and radiation were maintained in very close thermal contact by Compton scattering as long as the intergalactic gas remained ionised. This enabled the speed of sound to be determined at all epochs before the epoch of recombination. Therefore, the evolution of the Jeans' length and the mass of baryonic matter within this length, what is known as the *Jeans' mass*, could be evaluated.

In 1968, Joseph Silk showed that, during the pre-recombination epochs, sound waves in the radiation-dominated plasma were damped by repeated electron scatterings (Silk, 1968). The effect of this damping was to dissipate fluctuations with masses less than about $10^{12} M_\odot$, a mass known as the *Silk mass*, by the epoch of recombination. Consequently, all fine-scale structure would be wiped out and only large-scale structures on the scale of large galaxies and clusters of galaxies could form after recombination. In the early 1970s, Zeldovich and Edward Harrison independently put together information about the spectrum of the initial fluctuations on

different physical scales and showed that observed structures in the Universe could be accounted for if the mass fluctuation spectrum had the form $\Delta(M) \propto M^{-2/3}$ in the very early Universe, corresponding to a power spectrum of initial fluctuations of the form $|\Delta_k|^2 \propto k^n$ with $n = 1$. The amplitude of this scale-free power spectrum, known as the *Harrison-Zeldovich spectrum of initial perturbations*, was inferred to be $\sim 10^{-4}$ (Harrison, 1970; Zeldovich, 1972).

A key test of these models was provided by the fact that density fluctuations at the epoch of recombination should leave some imprint upon the intensity distribution of the Cosmic Microwave Background Radiation on the sky. In the simplest picture, if the process of recombination were instantaneous, adiabatic perturbations would be expected to result in temperature fluctuations $\Delta T/T = \frac{1}{3} \Delta \phi/c^2 = \frac{1}{3} \Delta \varrho/\varrho$ on large physical scales associated with large-scale gravitational perturbations, an effect known as the *Sachs-Wolfe effect* (Sachs and Wolfe, 1967). In fact, the problem is somewhat more complicated than this, partly because the process of recombination is not instantaneous and because other physical processes come into play on angular scales of about 1° and less. These include the adiabatic compression of the perturbations and first-order Doppler scattering due to the collapse of the primordial perturbations. These predictions provided a challenge for the observers since the amplitudes of the temperature fluctuations in these early theories were in the range $\Delta T/T \sim 10^{-3} - 10^{-4}$, well within the capability of sensitive anisotropy measurements of the Cosmic Microwave Background Radiation.

In the 1970s, these concepts gave rise to two principal scenarios for the formation of structure in the Universe. The first, known as the *adiabatic* model, was based upon a picture in which the perturbations were adiabatic sound waves before the epoch of recombination and structure in the Universe formed by the fragmentation of large-scale structures which reached amplitude $\delta\varrho/\varrho \sim 1$ at relatively late epochs. A realisation of this scenario was described by Andrei Doroshkevich, Sunyaev and Zeldovich in 1974 (Doroshkevich et al., 1974).

An alternative picture was one in which the perturbations were not sound waves but *isothermal* perturbations in pressure balance with the background radiation in the pre-recombination plasma. Small mass perturbations were not damped in this picture and so perturbations on all scales survived to the recombination epoch. After that epoch, the Jeans' mass dropped to about $10^6 M_\odot$, corresponding roughly to the masses of globular clusters. Galaxies and clusters of galaxies then formed by the process of *hierarchical clustering* under the influence of perturbations on larger physical scales.

Both models predicted similar amplitudes for the density perturbations at the epoch of recombination on large physical scales and consequently similar temperature perturbations in the Cosmic Microwave Background Radiation. Their subsequent behaviour was, however, quite different. The adiabatic picture could be thought of as a 'top-down' process of galaxy formation in which the largest scale structures formed first and then smaller scale structures formed by a process of fragmentation. In contrast, the isothermal picture corresponded to a 'bottom-up' process in which small-scale objects came together to form larger structures by hierarchical clustering. In the adiabatic picture, galaxies, stars and the chemical elements all formed at

relatively
very muc

Throu
fluctuatio
being ana
His own
by that tir
conflict w
of the pe
the upper
Radiation
formation
from prin
the Unive
would hav

1.6 Hot

A solution
collaborato
a finite res
and Zeldo
preciable c
1966) and,
neutrinos c
their role i
Lyubimov
would just
Zeldov
which the I
et al., 1980
relativistic
stream free
be quickly
would oscil
nation, the l
and then ev
free-stream
with masses
as in the ad
smaller scal
great advant
fluctuations

different physical scales and showed that observed structures in the Universe could be accounted for if the mass fluctuation spectrum had the form $\Delta(M) \propto M^{-2/3}$ in the very early Universe, corresponding to a power spectrum of initial fluctuations of the form $|\Delta_k|^2 \propto k^n$ with $n = 1$. The amplitude of this scale-free power spectrum, known as the *Harrison-Zeldovich spectrum of initial perturbations*, was inferred to be $\sim 10^{-4}$ (Harrison, 1970; Zeldovich, 1972).

A key test of these models was provided by the fact that density fluctuations at the epoch of recombination should leave some imprint upon the intensity distribution of the Cosmic Microwave Background Radiation on the sky. In the simplest picture, if the process of recombination were instantaneous, adiabatic perturbations would be expected to result in temperature fluctuations $\Delta T/T = \frac{1}{3} \Delta \phi/c^2 = \frac{1}{3} \Delta \rho/\rho$ on large physical scales associated with large-scale gravitational perturbations, an effect known as the *Sachs-Wolfe effect* (Sachs and Wolfe, 1967). In fact, the problem is somewhat more complicated than this, partly because the process of recombination is not instantaneous and because other physical processes come into play on angular scales of about 1° and less. These include the adiabatic compression of the perturbations and first-order Doppler scattering due to the collapse of the primordial perturbations. These predictions provided a challenge for the observers since the amplitudes of the temperature fluctuations in these early theories were in the range $\Delta T/T \sim 10^{-3} - 10^{-4}$, well within the capability of sensitive anisotropy measurements of the Cosmic Microwave Background Radiation.

In the 1970s, these concepts gave rise to two principal scenarios for the formation of structure in the Universe. The first, known as the *adiabatic model*, was based upon a picture in which the perturbations were adiabatic sound waves before the epoch of recombination and structure in the Universe formed by the fragmentation of large-scale structures which reached amplitude $\delta_0/\rho \sim 1$ at relatively late epochs. A realisation of this scenario was described by Andrei Doroshkevich, Sunyaev and Zeldovich in 1974 (Doroshkevich et al., 1974).

An alternative picture was one in which the perturbations were not sound waves but *isothermal* perturbations in pressure balance with the background radiation in the pre-recombination plasma. Small mass perturbations were not damped in this picture and so perturbations on all scales survived to the recombination epoch. After that epoch, the Jeans' mass dropped to about $10^6 M_\odot$, corresponding roughly to the masses of globular clusters. Galaxies and clusters of galaxies then formed by the process of *hierarchical clustering* under the influence of perturbations on larger physical scales.

Both models predicted similar amplitudes for the density perturbations at the epoch of recombination on large physical scales and consequently similar temperature perturbations in the Cosmic Microwave Background Radiation. Their subsequent behaviour was, however, quite different. The adiabatic picture could be thought of as a 'top-down' process of galaxy formation in which the largest scale structures formed first and then smaller scale structures formed by a process of fragmentation. In contrast, the isothermal picture corresponded to a 'bottom-up' process in which small-scale objects came together to form larger structures by hierarchical clustering. In the adiabatic picture, galaxies, stars and the chemical elements all formed at

relatively late epochs, whereas in the isothermal picture, they could begin to form at very much earlier cosmic epochs.

Throughout the 1970s increasingly sensitive searches were made for temperature fluctuations in the Cosmic Microwave Background Radiation, these observations being analysed critically by Bruce Partridge in his review of 1980 (Partridge, 1980a). His own observations had reached sensitivities of $\Delta T/T \approx 10^{-4}$ or slightly better by that time (Partridge, 1980b). Models with low density parameters were in serious conflict with these upper limits because, in these, there is relatively little growth of the perturbations after the epoch of recombination. Thus, by the early 1980s, the upper limits to the intensity fluctuations in the Cosmic Microwave Background Radiation were beginning to constrain severely purely baryonic theories of structure formation. Furthermore, the limits to the density parameter in the form of baryons from primordial nucleosynthesis arguments showed that, if the density of matter in the Universe were close to the critical density, most of the matter in the Universe would have to be in some non-baryonic form.

1.6 Hot and Cold Dark Matter

A solution to these problems appeared in 1980 when Valentin Lyubimov and his collaborators reported experiments which suggested that the electron neutrino had a finite rest mass of about 30 eV (Lyubimov et al., 1980). In 1966, Semion Gershtein and Zeldovich had noted that relic neutrinos of finite rest mass could make an appreciable contribution to the mass density of the Universe (Gershtein and Zeldovich, 1966) and, in the 1970s, György Marx and Alex Szalay had considered the role of neutrinos of finite rest mass as candidates for the dark matter, as well as studying their role in galaxy formation (Marx and Szalay, 1972). The intriguing aspect of Lyubimov's result was that, if the relic neutrinos had this rest mass, the Universe would just be closed, $\Omega_0 = 1$.

Zeldovich and his colleagues developed a new version of the adiabatic model in which the Universe was dominated by neutrinos with finite rest mass (Doroshkevich et al., 1980). Neutrino fluctuations would begin to grow as soon as they became non-relativistic but, since the neutrinos are very weakly interacting particles, they would stream freely out of the perturbations and so small-scale density perturbations would be quickly damped out. The closely coupled matter and radiation density fluctuations would oscillate at a low level during the pre-recombination era but, after recombination, the baryonic matter would fall into the larger amplitude neutrino fluctuations and then evolve more or less as in the standard adiabatic scenario. Because of the free-streaming of the neutrinos, however, only the very largest scale perturbations with masses $M \gtrsim 10^6 M_\odot$ would survive to the epoch of recombination and so, just as in the adiabatic model, the largest scale perturbations would form first and then smaller scale structures form by the process of fragmentation. This model had the great advantage of reducing very significantly the expected amplitude of temperature fluctuations in the Cosmic Microwave Background Radiation since the perturbations

in the baryonic matter would be of low amplitude during the critical phases when the background photons were last scattered.

In 1970, Zeldovich discovered a solution for the non-linear development of a collapsing cloud and used it to show that large-scale perturbations would collapse into sheets and pancakes which resemble the large-scale filamentary structure seen in the distribution of galaxies (Zeldovich, 1970). This scenario for galaxy formation became known as the *Hot Dark Matter* picture since the neutrinos were relativistic when they decoupled from the primordial plasma.

There were, however, concerns about this picture. First of all, there were reservations about the experiments which claimed to have measured the rest mass of the electron neutrino and it appears that Lyubimov's results were erroneous – the present upper limit to the rest mass of the electron neutrino is a few electronvolts. Secondly, constraints could be set to the masses of the neutrinos if they were to constitute the dark matter in galaxies, groups and clusters of galaxies. In 1979, James Gunn and Scott Tremaine showed how the phase space constraints associated with fermions such as neutrinos could be used to set lower limits to their masses (Tremaine and Gunn, 1979). While 30 eV neutrinos could bind clusters and the haloes of giant galaxies, those needed to bind dwarf galaxies would have to have masses much greater than 30 eV. This was not necessarily a fatal flaw because it could be that some other form of dark matter was present in the haloes of the dwarf galaxies.

There was also the realisation about this time that there were several alternative possibilities for the dark matter which came from theories of elementary particles. Examples included the axions, supersymmetric particles such as the gravitino or photino and ultraweakly interacting neutrino-like particles, all of which would be relics of the very early Universe. The period 1980 to 1982 marked the period when the particle physicists began to take the early Universe very seriously as a laboratory for particle physics. According to James Peebles, Richard Bond introduced the term *Cold Dark Matter* in 1982 to encompass many of the exotic types of particle suggested by particle physicists. The matter was 'cold' in the sense that these particles decoupled from the thermal background after they had become non-relativistic.

The Cold Dark Matter scenario is similar in many ways to the isothermal model. Since the matter is very cold, perturbations are not destroyed by free streaming. Fluctuations on all scales can survive and so, when the pre-recombination Universe became matter dominated, these perturbations began to grow, decoupled from the matter and radiation. As in the Hot Dark Matter scenario, after the epoch of recombination, the baryonic matter collapsed into the growing potential wells in the dark matter and galaxies, groups and clusters formed by hierarchical clustering. In 1982, Peebles demonstrated how the presence of such particles could reduce the amplitude of the predicted fluctuations in the Cosmic Microwave Background Radiation to levels consistent with the observational upper limits (Peebles, 1982). A remarkably useful formalism for the process of hierarchical clustering was described by William Press and Paul Schechter in 1974 which gives a good description of how the mass function of objects of different masses evolves with time (Press and Schechter, 1974).

in the baryonic matter would be of low amplitude during the critical phases when the background photons were last scattered.

In 1970, Zeldovich discovered a solution for the non-linear development of a collapsing cloud and used it to show that large-scale perturbations would collapse into sheets and pancakes which resemble the large-scale filamentary structure seen in the distribution of galaxies (Zeldovich, 1970). This scenario for galaxy formation became known as the *Hot Dark Matter* picture since the neutrinos were relativistic when they decoupled from the primordial plasma.

There were, however, concerns about this picture. First of all, there were reservations about the experiments which claimed to have measured the rest mass of the electron neutrino and it appears that Lyubimov's results were erroneous – the present upper limit to the rest mass of the electron neutrino is a few electronvolts. Secondly, constraints could be set to the masses of the neutrinos if they were to constitute the dark matter in galaxies, groups and clusters of galaxies. In 1979, James Gunn and Scott Tremaine showed how the phase space constraints associated with fermions such as neutrinos could be used to set lower limits to their masses (Tremaine and Gunn, 1979). While 30 eV neutrinos could bind clusters and the haloes of giant galaxies, those needed to bind dwarf galaxies would have to have masses much greater than 30 eV. This was not necessarily a fatal flaw because it could be that some other form of dark matter was present in the haloes of the dwarf galaxies.

There was also the realisation about this time that there were several alternative possibilities for the dark matter which came from theories of elementary particles. Examples included the axions, supersymmetric particles such as the gravitino or photino and ultraweakly interacting neutrino-like particles, all of which would be relics of the very early Universe. The period 1980 to 1982 marked the period when the particle physicists began to take the early Universe very seriously as a laboratory for particle physics. According to James Peebles, Richard Bond introduced the term *Cold Dark Matter* in 1982 to encompass many of the exotic types of particle suggested by particle physicists. The matter was 'cold' in the sense that these particles decoupled from the thermal background after they had become non-relativistic.

The Cold Dark Matter scenario is similar in many ways to the isothermal model. Since the matter is very cold, perturbations are not destroyed by free streaming. Fluctuations on all scales can survive and so, when the pre-recombination Universe became matter dominated, these perturbations began to grow, decoupled from the matter and radiation. As in the Hot Dark Matter scenario, after the epoch of recombination, the baryonic matter collapsed into the growing potential wells in the dark matter and galaxies, groups and clusters formed by hierarchical clustering. In 1982, Peebles demonstrated how the presence of such particles could reduce the amplitude of the predicted fluctuations in the Cosmic Microwave Background Radiation to levels consistent with the observational upper limits (Peebles, 1982). A remarkably useful formalism for the process of hierarchical clustering was described by William Press and Paul Schechter in 1974 which gives a good description of how the mass function of objects of different masses evolves with time (Press and Schechter, 1974).

These alternative dark matter scenarios for galaxy formation were the subject of a great deal of analysis and computer simulation during the 1980s. The Hot Dark Matter picture tended to predict too much power in large-scale structures, while the Cold Dark Matter models predicted too little (Frenk, 1986). One of the most important predictions of these models was that temperature fluctuations in the Cosmic Microwave Background Radiation should be detected at the level of about one part in 10^5 and these were detected in 1992 by the COBE satellite. Fluctuations on angular scales $\theta \geq 7^\circ$ were discovered with amplitude $\Delta I/I \approx 10^{-5}$ by George Smoot and his colleagues (Smoot et al., 1992). These fluctuations correspond to physical dimensions about ten times the size of the largest holes and voids observed in the distribution of galaxies. On these large angular scales, the source of intensity fluctuations is the Sachs-Wolfe effect, that is, the gravitational redshift associated with photons originating from within the density fluctuations at the last scattering surface (Sachs and Wolfe, 1967). It can be shown that the Harrison-Zeldovich spectrum with $n = 1$ results in temperature fluctuations which are independent of angular scale for all scales greater than a few degrees, consistent with the COBE observations.

The Cold Dark Matter model became the preferred picture for galaxy and structure formation, but it needed patching up to achieve consistency with all the observations. Viable models were constructed which include a tilted power spectrum of the initial fluctuations as compared with the standard Harrison-Zeldovich spectrum, others included the cosmological constant or decaying neutrinos, yet others considered a mixture of Hot and Cold Dark Matter and others considered that the Universe might be open. All the models included biasing of the distribution of visible matter relative to that of the dominant dark matter, which defined the large-scale structure of the distribution of galaxies (Turok, 1997).

The picture changed dramatically in the final years of the twentieth century and the first few years of the twenty-first. Firstly, the Type Ia supernovae were found to be excellent 'standard candles' for the estimation of cosmological distances. The redshift-magnitude relation for these supernovae strongly suggested that the cosmological constant was not zero, but had a large positive value, the best-fitting value corresponding to a density parameter of what became known as the *dark energy* of $\Omega_\Lambda \approx 0.7$ (Knop et al., 2003; Tonry et al., 2003).

Equally important were experiments to determine the detailed power spectrum of temperature fluctuations in the Cosmic Microwave Background Radiation which were predicted to display prominent maxima on angular scales less than 1° . The predictions were refined in the context of the baryonic adiabatic and isothermal models by Sunyaev and Zeldovich in 1970 (Sunyaev and Zeldovich, 1970). It was realised that these *acoustic* fluctuations contain a great deal of information about the large-scale properties of the Universe and a number of experimental groups made very large efforts to pin down the exact shape of their power spectrum. Strong evidence for these oscillations were found in these experiment, but these endeavours were largely superseded by the first-year results of the Wilkinson Microwave Anisotropy Probe (Bennett et al., 2003). These defined in exquisite detail the power spectrum of the fluctuations in the Cosmic Microwave Background Radiation and

enabled remarkably precise estimates of cosmological parameters to be made, particularly when combined with the results of large-scale galaxy surveys such as the Anglo-Australian Telescope 2dF Survey and the Sloan Digital Sky Survey. These observations demonstrated beyond doubt that the large-scale geometry of the Universe is very close to flat, that the density parameter in dark and baryonic matter is close to 0.28 and that the density parameter in the dark energy is close to 0.72 (Tegmark et al., 2004). What is particularly impressive about these results is that they are entirely consistent with many independent astronomical estimates of the cosmological parameters. This has given rise to the concept of the *concordance values* of the cosmological parameters, a set of parameters which, within the quoted uncertainties, are in agreement with all the best estimates of their values.

From the perspective of galaxy and structure formation, the good news was that these achievements incorporate naturally the formation of structure according to the standard Λ CDM model with no biasing. Equally impressive was the fact that the determination of the two-point correlation function for galaxies determined from the large-scale galaxy surveys now overlapped the corresponding angular scales in the Cosmic Microwave Background Radiation and that these were in excellent agreement. In the most recent analyses, evidence has been found for a maximum in the two-point correlation function for galaxies corresponding to the first peak in the power spectrum of perturbations in the Cosmic Microwave Background Radiation.

The upshot of these remarkable developments is that we have now entered the era of precision cosmology in which cosmological parameters can be estimated with confidence to better than 5% and much deeper cosmological questions can be addressed by the present and future generations of observations and experiments.

1.7 The Very Early Universe

Despite the undoubtedly success of the concordance model, it raises as many problems as it solves. The picture is incomplete in the sense that, within the context of the standard world models, the initial conditions have to be put in by hand in order to create the Universe as we observe it today. How did these initial conditions arise? The resolution of these problems will undoubtedly give insight into the laws of physics under physical conditions which at the moment can only be studied by cosmological observations.

- *The horizon problem.* This problem can be restated, ‘Why is the Universe so isotropic?’ (Dicke, 1961). At earlier cosmological epochs, the particle horizon $r \sim ct$ encompassed less and less mass and so the scale over which particles could be causally connected became smaller and smaller. A vivid example of this problem is to work out how far light could have travelled along the last scattering layer at $z = 1000$ since the Big Bang. Regions of the sky separated by angular distances greater than 2° could not have been in causal communication. Why then is the Cosmic Microwave Background Radiation so isotropic?

enabled remarkably precise estimates of cosmological parameters to be made, particularly when combined with the results of large-scale galaxy surveys such as the Anglo-Australian Telescope 2dF Survey and the Sloan Digital Sky Survey. These observations demonstrated beyond doubt that the large-scale geometry of the Universe is very close to flat, that the density parameter in dark and baryonic matter is close to 0.28 and that the density parameter in the dark energy is close to 0.72 (Tegmark et al., 2004). What is particularly impressive about these results is that they are entirely consistent with many independent astronomical estimates of the cosmological parameters. This has given rise to the concept of the *concordance values* of the cosmological parameters, a set of parameters which, within the quoted uncertainties, are in agreement with all the best estimates of their values.

From the perspective of galaxy and structure formation, the good news was that these achievements incorporate naturally the formation of structure according to the standard Λ CDM model with no biasing. Equally impressive was the fact that the determination of the two-point correlation function for galaxies determined from the large-scale galaxy surveys now overlapped the corresponding angular scales in the Cosmic Microwave Background Radiation and that these were in excellent agreement. In the most recent analyses, evidence has been found for a maximum in the two-point correlation function for galaxies corresponding to the first peak in the power spectrum of perturbations in the Cosmic Microwave Background Radiation.

The upshot of these remarkable developments is that we have now entered the era of precision cosmology in which cosmological parameters can be estimated with confidence to better than 5% and much deeper cosmological questions can be addressed by the present and future generations of observations and experiments.

1.7 The Very Early Universe

Despite the undoubted success of the concordance model, it raises as many problems as it solves. The picture is incomplete in the sense that, within the context of the standard world models, the initial conditions have to be put in by hand in order to create the Universe as we observe it today. How did these initial conditions arise? The resolution of these problems will undoubtedly give insight into the laws of physics under physical conditions which at the moment can only be studied by cosmological observations.

- *The horizon problem.* This problem can be restated, 'Why is the Universe so isotropic?' (Dicke, 1961). At earlier cosmological epochs, the particle horizon $r \sim ct$ encompassed less and less mass and so the scale over which particles could be causally connected became smaller and smaller. A vivid example of this problem is to work out how far light could have travelled along the last scattering layer at $z = 1000$ since the Big Bang. Regions of the sky separated by angular distances greater than 2° could not have been in causal communication. Why then is the Cosmic Microwave Background Radiation so isotropic?

- *The flatness problem.* This problem arises from the fact that, according to the standard world models, if the Universe were set up with a value of the density parameter differing even slightly from the critical value $\Omega_0 = 1$, it would diverge very rapidly from $\Omega_0 = 1$ at later epochs (Dicke, 1961; Dicke and Peebles, 1979). There is nothing in the standard world models that would lead us to prefer any particular value of Ω_0 . Why then in the density parameter close to one today? This is sometimes referred to as the *fine-tuning problem*.
- *The baryon asymmetry problem.* The baryon asymmetry problem arises from the fact that the photon-to-baryon ratio today is $N_\gamma/N_B \approx 10^9$. If photons were neither created or destroyed, this ratio is conserved as the Universe expands. At temperature $T \approx 10^{10}$ K, electron-positron pair production takes place from the photon field. At a correspondingly higher temperature, baryon-antibaryon pair production takes place with the result that there must have been a very small asymmetry in the baryon-antibaryon ratio in the very early Universe if we are to end up with the correct photon-to-baryon ratio at the present day. If the Universe had been symmetric with respect to matter and antimatter, the photon-to-baryon ratio would now be about 10^{18} , in gross contradiction with the observed value (Zeldovich, 1965). Therefore, there must be some mechanism in the early Universe which results in a slight asymmetry between matter and antimatter.
- *The primordial fluctuation problem.* What was the origin of the density fluctuations from which galaxies and large-scale structures formed? The amplitudes of the density perturbations when they came through the horizon had to be of finite amplitude, $\Delta = \delta\varrho/\varrho \sim 10^{-4}$, on a very wide range of mass scales. These cannot have originated as statistical fluctuations in the numbers of particles on, say, the scales of superclusters of galaxies. There must be some physical mechanism which generated finite amplitude perturbations with power spectrum close to $P(k) \propto k$ in the early Universe.
- *The values of the cosmological parameters.* The horizon and flatness problems were recognised before compelling evidence was found for the finite value of the cosmological constant, or in modern parlance, the density parameter of the dark energy Ω_A . The Universe seems to be geometrically flat and so the sum of the density parameters in the matter and the dark energy must sum to unity, $\Omega_A + \Omega_m = 0.72 + 0.28 = 1$. Even if the sum of these two parameters were not precisely unity, it is a surprise that the two parameters are of the same order of magnitude at the present epoch because the matter density evolves with redshift as $(1+z)^3$, while the dark energy density parameter is unchanged with cosmic epoch. Why then do we live at an epoch when they have more or less the same value?
- A further problem concerns the present value of the density parameter of the dark energy Ω_A which can be estimated using simple concepts from quantum field theory. The value found is about 10^{10} times greater than permissible values at the present epoch. This is quite a problem, but it should not be passed over lightly. If the inflationary picture of the very early Universe is taken seriously, this is exactly the type of force which drove the inflationary expansion.

- *The nature of dark matter and dark energy.* As if these problems were not serious enough, they are compounded by the fact that the nature of the dark matter and the dark energy are unknown. One of the consequences of precision cosmology is the troubling result that we do not understand the nature of about 95% of the material which drives the large-scale dynamics of the Universe.

The first suggestion that some of these problems might be resolved by appeal to particle physics was made by Sakharov in 1967 who suggested that the baryon-antibaryon asymmetry might be associated with the type of symmetry-breaking observed in the decays of the K mesons, in other words, that the asymmetry is associated with the type of symmetry-breaking which occurs in Grand Unified Theories of elementary particles in the early Universe (Sakharov, 1967).

The most important conceptual development for contemporary cosmology came in 1981 with Alan Guth's proposal of the *inflationary model* for the very early Universe (Guth, 1981). There had been earlier suggestions foreshadowing his proposal. For example, Zeldovich had noted in 1968 that there is a physical interpretation of the cosmological constant Λ associated with the zero-point fluctuations of a vacuum. Andrei Linde in 1974 and Sydney Bludman and Malvin Ruderman in 1977 had shown that the scalar Higgs fields, which had been introduced to give the W^\pm and Z^0 particles mass, have similar properties to those which would result in a positive cosmological constant.

In Guth's paper of 1981, he realised that if the Universe went through an early exponential expansion phase, this would solve both the problem of the isotropy of the Universe on a large scale and would also drive the Universe towards a flat spatial geometry. The effects of the exponential expansion is to drive neighbouring particles apart at an exponentially increasing rate so that, although they were in causal contact in the very early Universe, the exponential inflation quickly moves them far beyond their local horizons and can account for the large-scale isotropy of the Universe by the end of the inflation epoch. The exponential expansion also straightens out the geometry of the Universe, however curved it may have been in its initial stages. At the end of this phase of exponential inflation, the Universe transforms into the standard Friedman world model, which, since it has very precisely flat geometry, must have $\Omega_0 = 1$. In Guth's original picture, the transformation to the Friedman solution took place through a first-order phase transition but this created too many magnetic monopoles. The model was revised in 1982 by Linde and by Andreas Albrecht and Paul Steinhardt who showed how the transition to the Friedman solutions could be smooth and continuous and so avoid many of the problems associated with Guth's proposal (Linde, 1982, 1983; Albrecht and Steinhardt, 1982).

The original hope that a physical realisation for the inflationary expansion could be found within the context of particle physics beyond the standard model has not been achieved, but the underlying concepts of the inflationary picture have been used to define the necessary properties of the *inflaton* potential needed to create the Universe as we know it.

Since 1982, the inflationary scenario for the early evolution of the Universe between the epochs when it was only 10^{-34} to 10^{-32} seconds old has been studied

very intens
the realisati
also amplifi
Gibbons w
de Sitter spa
and Hawkin
inflationary
the 1982 Nu
the spectrum
result natura
to the Harris
to Andrew L

Althoug
tions ne
is owed
a possib
are belie
formed t
microwa

The nature of dark matter and dark energy. As if these problems were not serious enough, they are compounded by the fact that the nature of the dark matter and the dark energy are unknown. One of the consequences of precision cosmology is the troubling result that we do not understand the nature of about 95% of the material which drives the large-scale dynamics of the Universe.

The first suggestion that some of these problems might be resolved by appeal to particle physics was made by Sakharov in 1967 who suggested that the baryon–antibaryon asymmetry might be associated with the type of symmetry-breaking observed in the decays of the K mesons; in other words, that the asymmetry is associated with the type of symmetry-breaking which occurs in Grand Unified Theories of elementary particles in the early Universe (Sakharov, 1967).

The most important conceptual development for contemporary cosmology came in 1981 with Alan Guth's proposal of the *inflationary model* for the very early Universe (Guth, 1981). There had been earlier suggestions foreshadowing his proposal. For example, Zeldovich had noted in 1968 that there is a physical interpretation of the cosmological constant Λ associated with the zero-point fluctuations of a vacuum. Andrei Linde in 1974 and Sidney Bludman and Malvin Ruderman in 1977 had shown that the scalar Higgs fields, which had been introduced to give the W^\pm and Z^0 particles mass, have similar properties to those which would result in a positive cosmological constant.

In Guth's paper of 1981, he realised that if the Universe went through an early exponential expansion phase, this would solve both the problem of the isotropy of the Universe on a large scale and would also drive the Universe towards a flat spatial geometry. The effects of the exponential expansion is to drive neighbouring particles apart at an exponentially increasing rate so that, although they were in causal contact in the very early Universe, the exponential inflation quickly moves them far beyond their local horizons and can account for the large-scale isotropy of the Universe by the end of the inflation epoch. The exponential expansion also straightens out the geometry of the Universe, however curved it may have been in its initial stages. At the end of this phase of exponential inflation, the Universe transforms into the standard Friedman world model, which, since it has very precisely flat geometry, must have $\Omega_0 = 1$. In Guth's original picture, the transformation to the Friedman solution took place through a first-order phase transition but this created too many magnetic monopoles. The model was revised in 1982 by Linde and by Andreas Albrecht and Paul Steinhardt who showed how the transition to the Friedman solutions could be smooth and continuous and so avoid many of the problems associated with Guth's proposal (Linde, 1982, 1983; Albrecht and Steinhardt, 1982).

The original hope that a physical realisation for the inflationary expansion could be found within the context of particle physics beyond the standard model has not been achieved, but the underlying concepts of the inflationary picture have been used to define the necessary properties of the *inflation* potential needed to create the Universe as we know it.

Since 1982, the inflationary scenario for the early evolution of the Universe between the epochs when it was only 10^{-34} to 10^{-32} seconds old has been studied

very intensively. Among the further successes claimed for the theory has been the realisation that quantum fluctuations in the fields which drive the inflation are also amplified during the inflationary era. In 1977, Stephen Hawking and Gary Gibbons worked out the important result that quantum fluctuations in expanding de Sitter space produce thermal radiation with a well-defined temperature (Gibbons and Hawking, 1977). This acted as a stimulus to apply similar ideas to the new inflationary picture with a view to estimating the perturbation spectrum. Following the 1982 Nuffield Workshop held in Cambridge, the key result was established that the spectrum of quantum fluctuations of the vacuum Higgs fields were scale-free and result naturally in adiabatic curvature perturbations with spectrum strikingly similar to the Harrison–Zeldovich spectrum with $n \approx 1$ (Gibbons et al., 1983). According to Andrew Liddle and David Lyth:

Although introduced to resolve problems associated with the initial conditions needed for the Big Bang cosmology, inflation's lasting prominence is owed to a property discovered soon after its introduction. It provides a possible explanation for the initial inhomogeneities in the Universe that are believed to have led to all the structures we see, from the earliest objects formed to the clustering of galaxies to the observed irregularities in the microwave background (Liddle and Lyth, 2000).