

REVIEW ARTICLE

A survey of dark matter and related topics in cosmology

Bing-Lin Young

Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, USA

Email: young@iastate.edu

Received April 14, 2016; accepted May 2, 2016

This article presents an extensive review of the status of the search of the dark matter. The first eight sections are devoted to topics in dark matter and its experimental searches, and the rest to selected topics in astrophysics and cosmology, which are intended to supply some of the needed background for students in particle physics. Sections 9 and 13 are introductory cosmology. The three astrophysical topics, Big Bang nucleosynthesis Section 10, Boltzmann transport equation and freeze out of massive particles Section 11, and CMB anisotropy Section 12 can all be studied in analytical approaches when reasonable approximations are made. Their original analytical forms, to which this article follows very closely, were given by particle physicists. Dark matter is an evolving subject requiring timely update to stay current. Hence a review of such a subject matter would undoubtedly have something wanting when it appears in print. It is hoped that this review can form a humble basis for those graduate students who would like to pursue the subject of dark matter. The reader can use the extensive table of contents to see in some details the materials covered in the article.

Keywords dark matter, CMB anisotropy, Boltzmann transport equation, freeze out of massive particles

PACS numbers 12, 13, 95

Contents				
		3	Milky Way and galactic dark matter density profile	23
I	A survey of dark matter	4		
1	Introduction	4	3.1 A blurb of the Milky Way	23
2	Observational evidence	8	3.2 Dark matter in the Milky Way	23
2.1	Galactic rotation curve	9	4 Dark matter candidates	26
2.2	Large scale structures, hot and cold dark matter	11	4.1 Dark matter: Kinematic types and production mechanisms, a broad characterization	26
2.3	Gravitational lensing, collisions of galaxy clusters	12	4.2 Dark matter: Particle types	28
2.3.1	Gravitational Lensing and dark matter	13	4.2.1 Broaden the scope: Dark sectors – a brief description	30
2.3.2	Collisions of clusters and galaxies	14	4.3 Remarks on viable dark matter candidates	31
2.4	Dark matter in galaxies and a cosmic map of dark matter	18	4.4 Ranges and bounds of masses of dark matter candidates	32
2.4.1	Dark matter halo	18	5 WIMPs	33
2.4.2	A cosmic map of dark matter	18	5.1 Weakly interacting massive particles (WIMPs)	33
2.5	Cosmic microwave background anisotropy	19	5.2 Supersymmetric particles as WIMPs	33
2.6	Primordial nucleosynthesis: The baryon content of the universe	20	5.3 WIMP physics and cosmology	34
2.7	Challenges to CDM and alternatives to dark matter	20	5.3.1 Thermal relics	35
2.7.1	Challenges to CDM	20	5.3.2 Hot relics	36
2.7.2	Alternatives to dark matter	22	5.3.3 Cold relics	37
2.8	Finding a solution	22	5.3.4 The WIMP miracle	38

5.4	Comments on thermal vs. non-thermal production	39	9.1.8	A brief comment on measuring the Hubble expansion rate	78
6	Dark matter searches I – Mostly WIMPs	39	9.1.9	Newtonian limit and gravitational potential	78
6.1	Worldwide deep underground laboratories	40	9.2	Kinetic theory – Thermodynamics	79
6.2	Direct searches – Dark matter in terrestrial laboratories	42	9.2.1	Equilibrium distributions	79
6.2.1	A cursory look of the basic formulation	42	9.2.2	Chemical potential	80
6.2.2	Direct search experiments	44	9.2.3	Entropy	80
6.3	Indirect searches – Galactic dark matter	46	9.3	Relativistic and non-relativistic behaviors of particle densities	81
6.4	High energy collider – Accelerator production	47	9.3.1	Ultra-relativistic limit	81
7	Dark matter searches II – Light particles	49	9.3.2	Non-relativistic limit	82
7.1	Searches for the axion and axion-like particles	49	9.3.3	Internal degrees of freedom g_j	83
7.2	KeV sterile neutrinos	50	9.3.4	Some detailed properties	84
7.2.1	A possible candidate of 7 keV sterile neutrino	52	9.4	Beyond equilibrium, decoupling of particles or original of species	86
8	Summary of the status of dark matter searches	53	9.4.1	Mechanism of equilibrium	87
8.1	Direct searches	53	9.4.2	Decoupling – Breakaway from equilibrium	87
8.1.1	Direct search: DAMA/LIBRA	53	9.5	Thermal history of the early universe	88
8.1.2	Direct search: CDMS-II and SuperCDMS	55	9.5.1	Separation of neutrino and photon temperatures	89
8.1.3	Direct search: CoGeNT	55	9.5.2	Temperature chronology of the early universe	90
8.1.4	Direct search: CRESST	56	9.5.3	Baryon-photon ratio in the equilibrium era	92
8.2	Indirect searches	57	9.5.4	Matter-radiation equality	93
8.2.1	Indirect searches: Charged particles, positrons and antiproton	58	9.5.5	Recombination and the last scattering surface	93
8.2.2	Indirect searches: Photons	60	9.5.6	Decoupling of the photon and the last scattering surface	96
8.2.3	Indirect searches: Neutrinos	62	9.5.7	A cheat sheet of the cosmo-history & a map of the universe	97
8.2.4	Anti-deuteriums	63	10	Big Bang nucleosynthesis: Primordial helium	98
8.2.5	Searches of other anti-particles	63	10.1	Introduction	99
8.3	Summary	63	10.2	Estimate decoupling	101
8.3.1	Status of experimental searches	63	10.2.1	Cosmic expansion rate	101
8.3.2	A glance at the future	64	10.2.2	Interaction rate	102
8.3.3	Some miscellaneous theoretical remarks	66	10.2.3	Approximate decoupling temperature T_f and neutron freezing fraction	102
II	Related Topics in Cosmology	67	10.3	Differential equation and initial condition for the neutron fraction	102
9	An introductory cosmology for pedestrians	68	10.3.1	Nucleon conversion rates	104
9.1	The Einstein field equation and FLRW metric	68	10.3.2	Solution of the differential equation	107
9.1.1	Review of some generality	68	10.4	Freezing out neutron fraction	109
9.1.2	The FLRW metric	70	10.5	Effect of neutron decay	110
9.1.3	The Einstein equation in the FLRW metric	71	10.6	Arresting neutron decay – Formation of light elements	111
9.1.4	Cosmological redshift	75	10.6.1	Deuterium bottleneck	113
9.1.5	Energy components in the early universe as functions of the redshift	76	10.7	Primordial fraction of helium-4	114
9.1.6	Effect of the curvature term and the flat space	76	10.8	Counting the baryonic matter and the primordial lithium problem	115
9.1.7	Age of the universe	77			

11	Freeze out of massive particles	116	12.6	Small fluctuations and linear cosmological perturbation	137
11.1	Boltzmann transport equation for particle production and annihilation	116	12.6.1	Linear metric perturbations and their decomposition	138
11.2	The rate of change of particle number density	117	12.6.2	Perturbation in the stress-energy tensor	139
11.3	Digression – Riccati equation and some of its properties	120	12.6.3	Cosmological gauge transformation	142
11.3.1	Transforming to a linear equation	120	12.6.4	Perturbation to cosmological equations: Number of equations and solutions	144
11.3.2	Multiple solutions	120	12.7	Gauge fixing	146
11.3.3	Boundary condition and the unique solution	121	12.7.1	Newtonian gauge	147
11.4	Solution of the particle number density	121	12.7.2	Synchronous gauge	148
11.5	Freeze out: freeze out temperature and particle abundance	123	12.7.3	Remarks on other gauges and gauge conversions, and gauge solutions	149
11.5.1	Relic abundance due to freeze out	123	12.8	Very early cosmic perturbation – Outside the Horizon	149
11.5.2	The freeze out epoch	123	12.8.1	Issues involved – A preliminary discussion	149
11.6	Numerical value of λ and its Physical implications	124	12.8.2	Fourier decomposition and the stochastic property	150
11.6.1	Estimate of λ	124	12.8.3	Conservation outside the horizon	152
11.6.2	Fraction of critical density	125	12.8.4	Adiabatic solutions	153
11.6.3	Mass and annihilation cross section of massive particle	126	12.9	Evolution of scalar perturbations in the hydrodynamic limit	156
12	CMB anisotropy	126	12.9.1	Setting up the system of equations	156
12.1	Classification of CMB anisotropy and key experiments	127	12.9.2	Hydrodynamic equations and initial conditions	157
12.1.1	Classification of CMB anisotropy	127	12.9.3	The approach to analytic solutions	160
12.1.2	Key experiments	128	12.9.4	Solutions in the long wavelength regime	164
12.2	Secondary anisotropy	129	12.9.5	Solutions in the short wavelength regime	169
12.2.1	Anisotropy due to Earth motion	129	12.9.6	Interpolating between long and short wavelengths (ILS) – Transfer functions	171
12.2.2	The Sunyaev–Zel’dovich effect	130	12.10	Temperature anisotropy due to scalar perturbations	171
12.2.3	Integrated Sachs–Wolfe effect	130	12.10.1	Sudden transparency approximation – Following the photon trajectory	172
12.3	Primary CMB anisotropy	130	12.10.2	Temperature fluctuation due to scalar perturbations	174
12.3.1	Acoustic oscillation and modes of perturbations	130	12.10.3	Explicit expression for the temperature multipole coefficient	182
12.3.2	Diffusion damping	131	12.10.4	Multipole coefficients $C_{TT,\ell}$	185
12.4	Formulation of anisotropy	132	13	Cosmological distances and times	187
12.4.1	Properties of $a_{\ell m}$, Gaussian distribution, and angular power spectrum	132	13.1	Introduction	187
12.4.2	Observed angular power spectrum	133			
12.5	Angular sensitivity, multipoles, and angular scales	134			
12.5.1	Analysis by spherical harmonics	134			
12.5.2	Analysis by Fourier transform	135			
12.5.3	Some properties of the Spherical Bessel function of the first kind	135			

13.2	The FLRW metric	187	B.1.1	Klein–Nishina formula	204
13.2.1	The past and future light-cones	188	B.1.2	Thomson scattering cross section	205
13.2.2	The Λ CDM model	188	B.2	Weak interactions cross sections	205
13.3	Distance & time, comoving radial coordinate	189	B.2.1	Muon lifetime	205
13.3.1	Light propagation and null path in the FLRW metric	189	B.2.2	Leptonic elastic cross sections	205
13.3.2	Comoving Radial coordinate	190	B.2.3	Leptonic inelastic cross sections	205
13.3.3	Lookback time and light traveling time $\Delta t = t_0 - t_1$	190	B.2.4	More neutrino leptonic and annihilation cross sections	206
13.4	Hubble scales, age of the universe, and explicit forms of scale factor	190	B.2.5	Neutrino nucleon elastic cross sections	206
13.4.1	Hubble time τ_H and Hubble length D_H	190	B.2.6	Two-body neutrino nucleon inelastic cross sections	206
13.4.2	Ages of the universe τ_U and the matter-radiation epoch t_{EQ}	191	B.2.7	Orders of magnitude of typical cross sections	206
13.4.3	Functional form of the scale factor a	191	Appendix C	Useful constants and units conversion	206
13.5	Proper distance and related length measures	194	C.1	Natural units and units conversion	206
13.5.1	Proper distance, d_p , comoving distance d_C , and the Hubble expansion	194	C.1.1	Units conversion	207
13.5.2	Particle horizon d_{ph}	195	C.2	Masses and binding energies	207
13.5.3	Cosmic event horizon	196	C.3	Useful constants	208
13.6	Physical distance measures	196	C.4	Planck quantities	208
13.6.1	Angular diameter distance d_A and comoving angular diameter distance d_{cA}	196	Acknowledgements		209
13.6.2	The luminosity distance d_L	197	Remarks		209
13.7	Recapitulating results in a flat universe $\kappa = 0$	198	References and notes		209
13.7.1	Summary of results	198			
13.7.2	Explicit examples: LSS and later epoches	198			
13.7.3	Cosmic event horizon and a sketch of the long-term cosmic future	198			
13.7.4	Results for low redshifts	200			
13.7.5	Remarks on the cubic equation relating time and the Hubble scale factor	200			
13.8	Summary table	201			
Appendix A	Phase space, invariants, and the Liouville’s theorem	201			
A.1	Lorentz invariant volume element of the momentum space	202			
A.2	Lorentz invariant volume element of the configuration space	203			
A.3	Lorentz invariant phase space elements	204			
A.4	Liouville’s theorem	204			
Appendix B	Useful cross section formulae	204			
B.1	Electromagnetic interaction cross sections	204			

Part I

A survey of dark matter

1 Introduction

“Physics thrives on crisis. We all recall the great progress made while finding a way out of the various crises of the past.”

– Steven Weinberg [1]

“The natural world does not recognize the artificial separation of academia into traditional disciplines.”

– David Gross [2]

“The studies of the very large (cosmology) and the very small (elementary particles) are coming together.”¹⁾

– David Schramm

¹⁾I was not able to find the original source of this quotation although I have known it for a long time. The closest to a formal source that I can find is a press release on Dec. 2, 2005, on the inauguration symposium of the Kavli Institute for Cosmological Physics at the University of Chicago, which was also an event in honor of the memory of the late David Schramm. The news release can be found in <http://www-news.uchicago.edu/releases/05/051202.newviews.shtml> But quotations of this sentence appeared before this news release. I read it in the literature and then quote this sentence in my talk given in July 2004 at the International Workshop/School on Frontiers on High Energy Physics, which can be found at the ITP download website at “Download/HEP_2004_7_2-10/yangbl/Underground2.ppt”.

“Nowhere is the inherent unity of science better illustrated than in the interplay between cosmology, the study of the largest things in the Universe, and particle physics, the study of the smallest things.” – E. W. Kolb [3]

One of the extraordinary intellectual achievements in human history is the establishment of the standard model of quarks and leptons which took place during the second half of the 20th century. It is an answer to a question as old as civilization itself: What are the fundamental elements of matter of the physical world? The standard model (SM) gives clear description of the fundamental elements of all matter that are observed in terrestrial laboratories, and is an exact theory formulated in a precise mathematical framework to prescribe how fundamental particles behave. The last critical missing piece, i.e., the Higgs boson, has been observed in two LHC experiments [4–7] in 2012.

However, the SM is not perfect. With experimental inputs required to fix 19 parameters appearing in the theory, it is designed to represent physics down to 10^{-17} cm in the length scale or up to several hundreds of GeV in the energy scale. Both theoretical arguments and experimental observations indicate that the SM is not the end of the story line in the description of the microscopic world. There are problems concerning the theoretical structure of the SM: the gauge hierarchy problem; the flavor problem; the mass problem; and the strong CP problem. It is not clear whether or not all these problems are originated from a structural problem of the standard model. Near the end of the last century, the mass problem is specifically manifested by the discovery of massive neutrinos. Although finite neutrino mass terms can be introduced conventionally, in the form of Higgs terms without a significant modification of the theory, it makes the SM to appear rather contrived.²⁾ In addition to this direct manifestation due to massive neutrinos, compelling evidences which ascertain that there are more beyond the SM can be found when one gazes beyond the microscopic world to look into the very large scale of the cosmos: The presence of the dark matter and dark energy. The dark matter and dark energy are now integral parts of the matter-energy mix that make up the universe,³⁾ yet they are not required in the successful SM as we know it.

The SM has been extensively tested in terrestrial laboratories. If our understanding of the fundamental con-

stituents of all matter is complete, then quarks and leptons albeit massive neutrinos, together with other necessary components of the theory, should be all that are needed for the construction of a theory to study cosmic phenomena. But it is not the case. Once we go outside the terrestrial confinement to study phenomena of cosmic sizes, like galaxies clusters, we found that quarks, leptons, gauge bosons, and Higgs boson, the complete list of the standard model components, are woefully inadequate in counting the basic ingredients of the universe.

Another equally consequential milestone has also been established in the 20th century, based on the study of the opposite end of the natural world, i.e., the very large scale universe. The concordance model, or the Λ CDM Big Bang model, provides the best description of the universe and its evolution.⁴⁾ It is the answer to another set of fundamental questions: What is the essence of the universe to which we owed our existence, how is it formed, and what is its eventual fate? This standard model of cosmology is based on the Einstein field equation of general relativity in the FLRW metric (see Section 9.1), together with the inflationary scenario right after the Big Bang to give a flat geometry, augmented, in addition to quarks and leptons, etc., with cold dark matter and a cosmological constant-like component as the dark energy. The dark components of the universe dominates the energy-matter budget of the universe today and even more so in the future, and they are not contained in the SM of quarks and leptons. Including the dark components the Λ CDM model can provide a satisfactory description of high statistics astrophysics observations, such as the cosmic microwave background radiation (CMB), its temperature anisotropy that are imprinted by the dark components of the universe, etc. [9]. Assuming the primordial temperature fluctuations to be adiabatic with a power law spectrum, data on the temperature alone require a significant amount of non-baryonic cold dark matter. So the composition of the universe contains a lot more than what we have been able to fathom from the study of the microscopic scale physics by experiments performed on Earth.

Figure 1.1 shows the composition of the universe in the present epoch and at the *last scattering surface* when the universe is 380 000 years old, when the photon was decoupled from other components of the universe and able to propagate freely. The composition at the present epoch shown in the graph on the right comes from observation, and that of the last scattering surface, shown in the left, is from theoretical calculation. The difference in the compositions of the two epoches is due the different rate of changes of the various matter-energy densities as

²⁾In spite of this assertion, the introduction of a set of right-handed neutrinos and neutrino mass terms in the SM is a viable phenomenological approach and has been worked out in the literature.

³⁾In this work we take the mainstream cosmological view accepting the existence of dark matter and dark energy. A brief remark will be made at the end of the next chapter on the alternate paradigm of the modified newtonian dynamics (MOND).

⁴⁾To maintain a proper perspective, one should keep in mind the presently known theoretical and observational paradoxes of the standard cosmological model. See, for example, [8].

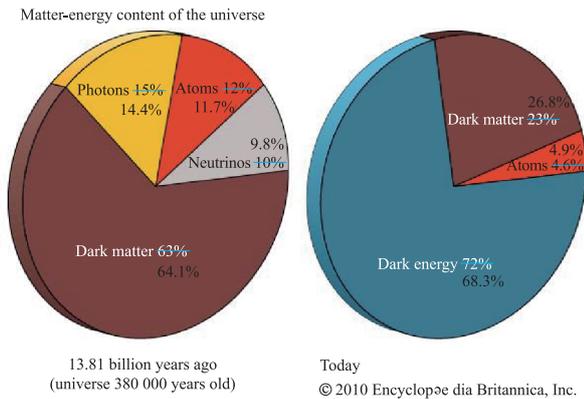


Fig. 1.1 The composition of the universe: The graph on the right is for today and that on the left is for the last scattering surface where the universe is at the age of 380 000 years when matter and radiation began to decouple. For more explanation of the figure see the relevant text. The unmodified graphs, which is based on older data set, is originated from the 2010 Encyclopedia Britannica which is the last printed edition of this authoritative English general reference books after 244 years of printing.

the universe undergoes Hubble expansion. This will be discussed in Section 9.

Two sets of data of the present epoch, and consequently two sets of results for the last scattering surface, are presented in Fig. 1.1 to show the rapid progress made in the observational cosmology. The numbers of smaller fonts which are also shown to be crossed out by short light blue lines are the WMAP 5-year data [10], which is the WMAP third data release made in March, 2008, incorporated in the 2012 *Review of Particle Physics* [11] are: age of the universe 13.75 ± 0.13 Gyr, dark energy $(73 \pm 3)\%$, dark matter $(11.1 \pm 0.6)h^{-2}\%$, baryonic matter $(4.5 \pm 0.3)\%$, and $h = 0.710 \pm 0.025$ which is the scale factor for the Hubble expansion rate. They have been superseded by the most recent WMAP 9-year results (4th data release, January, 2010) [10] and the Planck 2013 results [12]. The more recent data are given by numbers of larger fonts. As reflected in the current (2015) *Review of Particle Physics* 2014 [13], the updated numbers are: age of the universe 13.81 ± 0.05 Gyr, dark energy $(68.5 + 1.7 - 1.6)\%$, cold dark matter $(11.98 \pm 0.261.1)h^{-2}\%$, ordinary baryonic matter⁵⁾ $(4.99 \pm 0.22)\%$, and $h = 0.673 \pm 0.012$.

⁵⁾The baryonic matter, or ordinary matter, refers to atoms and their constituents, including the leptons. However, the observed baryon matter in the forms of luminous galaxies and diffused gas is less than 1%. Where lying the rest, which is the majority of the baryon matter, has long been a mystery. However, recent X-ray observations have found the evidence of this part of the ordinary matter to be in the form of rarified inter-galaxy hydrogen atoms, about six per cubic meter. In comparison, in the interstellar space there are a million atoms per cubic meter. See the Chandra X-ray Observatory press release on May 11, 2010, [14]. More details can be found in [15].

Another 0.44% consists of photons and neutrinos.⁶⁾ The two data sets are different but agree within 1σ .

From the discussion above we see that about 95% of the universe consists of energy-matter fields unknown to the SM. The right panel of Fig. 1.1 shows the composition of the universe at the time when the radiation and matter fields are decoupled about thirteen billion years ago when the universe was 380 000 years after the big bang. The universe became transparent to photons after 380 000 years. The dark matter, then, was the dominant component of this early universe and hence should play an important role already in the evolution of the early universe, such as the formation of galaxies, etc. To summarize the current status of this fundamental theory, we quote the WMAP 7-year results released on January 26, 2010 [16]: “*WMAP now places 50% tighter limits on the standard model of cosmology (Cold Dark Matter and a Cosmological Constant in a flat universe), and there is no compelling sign of deviations from this model.*” And “*WMAP has detected a key signature of inflation.*”⁷⁾

Evidences of the existence of dark matter come from the behavior of galaxies and clusters due to the gravitational effect, as revealed in a variety of astrophysical observations. The accumulation of evidences show more and more clearly that a large part of the matter of the universe is in a nonluminous form which may also be very weakly interacting with elements of the SM, so as to be able to escape terrestrial laboratory identification [17]. Today, more than five sixth of the matter in the universe is dark, while less than one sixth of the matter is made of quarks, leptons, photons, and neutrinos. This conclusion is consistent with the result comes from the Big Bang nucleosynthesis which constrains the baryonic matter to lie between 13.3% to 17% of the matter component of the universe.

But what is dark matter? To find out the properties of the dark matter we need appropriate laboratory tools, and the tools available have all been developed in the terrestrial environment for the study of ordinary particles. So, does it make sense to study dark matter using the reductionist’s tools of particle physics?

The approaches in astrophysics studies have always been the usage of the laws and tools of physics, including

⁶⁾The vast majority of cosmic neutrinos can be found in the cosmic microwave background radiation (CMBR). Although the cosmic neutrino background has not been directly observed, the theory of the light elements production such as deuterium, helium, etc., i.e., the big bang nucleosynthesis, has provided a strong evidence for its existence with properties consistent with what we known about neutrinos in particle physics.

⁷⁾WMAP data place tight constraints on the hypothesized burst of growth in the first trillionth of a second of the universe, known as *inflation*, when ripples in the fabric of space may have been created. The 7-year data provide evidences that the large-scale fluctuations are slightly more intense than the small-scale ones, which is a prediction common to many inflation models.

microscopic physics [18]: to make observations of galaxies, to study the birth and evolution and death of stars, even collisions of galaxy clusters, etc. On the approach to study the dark matter let us first look at the relations between elementary particles and cosmology. These two successful fundamental theories operate on very different length and time scales. Particle physics is on the scale of 10^{-17} cm and cosmology on the order of maximally 10^{28} cm. Their difference is as large as 45 orders of magnitude. On the time scale reactions among fundamental particles lasts less than a zeptosecond (10^{-21} s), while astrophysical processes last for many millions of years. The difference of the two time scales is the order of 10^{35} . Can they possibly be related? In an inclusive paradigm, the Big Bang theory suggests that they should be [19]. Let us trace the universe backward in time to the beginning close to the Big Bang. Our present reachable world must be started from a very tiny spacial point where the law of particle physics should rule if it can be extended to near the Planck energy scale in the form of a theoretical framework beyond the standard model (BSM). Then the dark matter must be a part of BSM. Following the argument of cosmology, at some time during the evolution of the universe, if massive enough, the dark component of the matter fields would be decoupled from the rest of the regular particles before the epoch during which the radiation and the normal matter are decoupled. But dark matter continues to exert deciding influence due to the gravitational effect. Such a unified view of particle physics and cosmology provides the motivation to look for dark matter in the framework of particle physics. And, at the same time, a refined theory in the next step of particle physics in synergy with cosmology may emerge, ideally, without the need to construct a much higher energy accelerator than LHC to guide us.

The story of the dark matter, as it usually told, began in the early 1930s by the assertion of the Swiss astronomer Fritz Zwicky in 1933 [20] that there exists in the universe extra mass beyond the luminous ones, as indicated by the observed orbital velocities of galaxies in galaxy clusters⁸⁾ The extra mass is not in any form of

⁸⁾The actual storyline, not as cut-and-dries as this description would imply, is apparently more complicated. It has been recognized that in 1932 Jan Oort had made similar observations on rotation curves of stars in galaxies. Zwicky's 1933 results were based on observations made in the larger setting of galaxy clusters. There are reports on the existence of possible dark matter even earlier by Knut Lundmark in 1930 and by Jacobus Kapteyn in as early as 1922. A description of Lundmark's work can be found in the listing of the "Physics Forums Insights": <https://www.physicsforums.com/insights/>. Reference to Kapteyn can be found in lecture notes on dark matter given in the Helsinki university, for instance, by a search on "[~xfiles/cosmology/12/cosmo2012_07.pdf](#)", and in [21]. A description of the history of dark matter searches can be found in [22]. See also brief but clear descriptions, on dark matter in general as well, in "[www.eclipse.net/~cmmiller/DM/](#)".

the known matter found in terrestrial laboratories. In the last two decades significant progress has been made in understanding properties of the dark matter, mostly by learning what it is not:

- They are not baryonic and carry no electric or color charges.
- They are not composed of known particles of the standard model.
- They do not interact electromagnetically and thus neither emit nor absorb light.
- They should be mostly cold and/or warm, i.e., non-relativistic or at least not ultra-relativistic.
- They do interact via gravity. The gravitational effect of the dark matter is ubiquitous in various aspects of the structure of the universe.

In spite of our ignorance as stated, the dark matter may be studied in some of the existing framework of particle physics and there are theoretical models that contain possible candidates of the dark matter. Presently, we are in the experimentally driven stage and a plethora of experiments have been performed, in progress, or under construction in search for dark matter particles, from which we can also learn its properties. In responding to experimental findings, theoretical frameworks of dark matter are modified and new ideas proposed. Ultimately the challenge is to find out where in a theoretical framework that the dark matter may fit.

This lecture notes claim very little original work although some detailed calculations have been performed and demonstrated. The detailed calculations are carried out mostly in Part II in topics of cosmology/astrophysics. They are for the purpose of providing a better understanding, or checking some of the results given in the literature, or for my own benefit in trying to gain more insights, to a specific topic. This is the reason why part II is lengthy. Almost all figures in Part I of the dark matter survey chapters are borrowed from the cosmology/astrophysics literature and from webpages I was able to access. For some of them, the sources are given. I apologize for the ones to which the sources are silent. I made the majority of the figures in Part II.

I would like to say briefly how the numerical works are performed. Numerical calculations are performed and most of the figures in Part II drawn using the commercial program called Mathcad which does not use any particular programmatic language. Numerical works are performed directly from the input mathematic formulae in the regular form as one usually writes them. The program runs fast on a modern desk top or laptop PC. The program can also plot graphs quickly in various ways, but unfortunately, not quite in the publication quality

by themselves.⁹⁾

Materials presented in Part I on the dark matter will very likely be replaced by updates, better results, in both theory and experiment. Although it is often said that *the goal of physics is to understand nature at the most fundamental level*, during the exploration stage of a topic many aspects are transitory in nature. This should not discourage students from vigorously pursue them. Consider the following statement that attributed to Niels Bohr: *Science aims not to understand nature but what we can meaningfully say about it.*¹⁰⁾ All efforts in search for dark matter are trying to say meaningfully what it is or what it is not.

A few words about the bibliography. The literature on dark matter and cosmology is huge and quickly enlarging, including many excellent review works. Hence, there is no way to claim inclusive. But I would like to mention two articles on dark matter [24] and [25]. The first one gives a good treatment of particle dark matter which is the attention of the majority of experimental searches of dark matter and the article has been highly cited. The second is one of the most recent review on dark matter which is concise and targeted to graduate students. I would like to recommend to readers who are interested in the present notes to also read this TASI lecture notes.

In addition, the 2014 SLAC Summer Institute lectures [26] is also a good source of current materials on the subject targeted at advanced students. Since only transparencies of the lectures are available, they are challenging learning tools for beginners. However they may serve as a guiding measure as to what students are expected to know in eyes of experts.

Before we plunge into the details we make the following remark to conclude this introduction. The plethora of particles in many different models proposed over the years exhibit the highly active state of the field. But this robustness also indicates the lack of experimental constraints on new particles beyond the standard model. So particle physics and astrophysics are facing the challenge together.

2 Observational evidence

Although dark matter has not yet been observed in any laboratories, either in terrestrial settings or in satellites,

⁹⁾Here a disclaimer is in order. This comment is not an advertisement for Mathcad, but just wants to reveal how numerical calculations are performed here. In the author's opinion Mathcad is a convenient tool to do quick calculations. It is very easy to debug for not using an arcane programming language. I am sure readers have their own favorable tools to do numerical calculations, such as the Mathematica.

¹⁰⁾Both statements in italic can be found in [23].

there exists a string of strong evidence of its existence which are entirely astronomical in nature. Cold non-baryonic dark matter came out as a common answer to a number of diverse astrophysical and cosmological phenomena which cannot be explained by the existing cosmological theory without the presence of such a matter component. The observations include galaxy rotation curves, behaviors of X-ray gases in galaxy clusters, lensing images involving galaxies and clusters, galaxy redshift surveys, cosmic microwave background anisotropy, etc. We discuss briefly each type of observations in the sections below.

The earliest evidence of dark matter was discovered in the 1930s. With continuous corroborating observations in various astronomical systems, in the early 1980s the astrophysical community was convinced that the gravitational force needed to hold together galaxies and clusters of galaxies are due to invisible masses. It has been generally postulated that the dark matter is located in a massive spherical halo enshrouding each galaxy.

Briefly, the observational evidence in the various systems are the followings: In spiral galaxies, stars and clouds of gas in the disk move in approximate circular orbits. Radial variation of the orbit speed can be readily measured through Doppler shifts and are found to be nearly constant out to large radial distance, in apparent violation of the law of Newtonian gravity. In our own galaxy, the Milky Way, because we are located in the galaxy disk it is difficult to measure the rotation curve at large radius, but other approaches can be used to estimate the mass it contains. For elliptical galaxies, as they are not supported by rotation, other dynamical mass indicators, not the rotation curve, are needed to measure their total masses. For instance some of the largest elliptical galaxies have an envelope of hot X-ray emitting gas. From the radial temperature distribution of the hot gas, the total mass of the parent elliptical galaxy can be calculated. For clusters of galaxies Doppler effect of individual galaxies, can also be used to infer the existence of dark matter on a much large scale. Zwicky, for example, found in 1933 that about 10 times as much mass as observed in the form of visible light was necessary in order to keep the individual galaxies within the cluster gravitationally bound. Dynamically, in the interactions of galaxies, dark matter is manifested in gravitational lensing.

The early history of the discovery of dark matter, in varying detail and emphasis, can be readily found on internet. For an example we refer to [27],¹¹⁾ which gives the following list of historical events, to which we refer for more details and references.

¹¹⁾This is a talk given by Primack in 2009. It highlights some of the key publications in the early stage of the evidence of dark matter.

Early history of dark matter

- 1922 - Kapteyn: “dark matter” in Milky Way disk
- 1933, 1937 - Zwicky: “dunkle (kalte) materie” in Coma cluster
- 1937 - Smith: “great mass of internebular material” in Virgo cluster
- 1937 - Holmberg: galaxy mass $5 \times 10^{11} M_{sun}$ from handful of pairs
- 1939 - Babcock observes rising rotation curve for M31¹²⁾
- 1940s - large cluster σV confirmed by many observers
- 1957 - van de Hulst: high HI rotation curve for M31
- 1959 - Kahn & Woltjer: MWy-M31 infall $\Rightarrow M_{LocalGroup} = 1.8 \times 10^{12} M_{sun}$
- 1970 - Rubin & Ford: M31 flat optical rotation curve
- 1973 - Ostriker & Peebles: halos stabilize galactic disks
- 1974 - Einasto, Kaasik & Saar; Ostriker, Peebles & Yahil: summarize evidence that galaxy M/L increases with radius
- 1975, 78 - Roberts; Bosma: extended flat HI rotation curves
- 1978 - Mathews: X-rays reveal enormous mass of Virgo cluster
- 1979 - Faber & Gallagher: convincing evidence for dark matter
- 1980 - Most astronomers are convinced that dark matter exists around galaxies and clusters.

2.1 Galactic rotation curve

The general pattern of motions of stars in a galaxy, as well as galaxies in a cluster, provide the bulk of the early evidence of this non-baryonic matter. A cartoon sketch of the pattern is given in Fig. 2.1. This figure shows the stars’ rotation velocity-vs-distance relationship, known as the rotation curve, of a typical spiral galaxy, in which the individual stars move around the center of the galaxy in a circular orbit. The horizontal axis is the distance from the center of the galaxy and the vertical axis is the velocity of individual stars. The dashed line A is the predicted rotation curve based on the observed mass distribution of the galaxy. The solid line B is the observed rotation curve. Note that the observed rotation curve becomes flattened when the distance of a star from the center increases, and the flattening of the rotation velocity extends to a large distance where very little visible matter can be found. Hence there is either non-luminous matter, i.e., the *dark matter*, presenting in the halo of the galaxy, or the Newtonian law of gravitation is no longer valid in the cosmic scale. The existence of dark matter with a certain required property is the favored explanation. It is required in some cases, that the dark matter

¹²⁾M31 refers to the spiral galaxy Andromeda which is the nearest major galaxy to the Milky Way. Andromeda is the largest galaxy of the local group, which includes also the Milky Way, the Triangulum Galaxy, and about 44 other smaller galaxies.

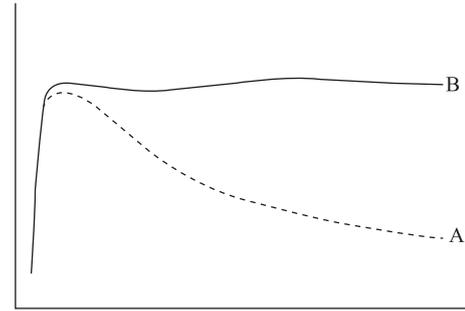


Fig. 2.1 The typical rotation of stars in a spiral galaxy. The vertical axis is the velocity of the stars and horizontal distance to the center of the galaxy.

exists beyond the visible limit of a galaxy by as far as 10 times further.

The feature of the curves A and B of Fig. 2.1 can be reproduced crudely by the following simple mass distributions, using freshman mechanics. Let $M_A(r)$ and $M_B(r)$ be two spherical symmetric mass distributions which give rise to total masses contained within a sphere of radius r of the forms,

$$M_A(r) = M_T \left(\frac{r}{R}\right)^3 \theta(R - r) + M_T \theta(r - R),$$

$$M_B(r) = M_T \left(\frac{r}{R}\right)^3 \theta(R - r) + M_T \left(\frac{r}{R}\right) \theta(r - R), \quad (2.1)$$

where R is a fixed radius. For $M_A(r)$, within the spheres of radius R the mass density is constant with a total mass M_T . Outside the sphere of radius R the mass density vanishes and the total mass is constant for any sphere of $r > R$. The mass density in this case is given by

$$\rho_A(r) = \left(\frac{M_T}{\frac{4\pi}{3} R^3}\right) \theta(R - r). \quad (2.2)$$

This represents the case of a galaxy consists of baryon matter confined in a sphere of radius R . For $M_B(r)$, the mass distribution is the same as $M_A(r)$ for $r \leq R$. Outside the sphere of radius R the mass density in $M_B(r)$ is non-vanishing but decrease like r^{-2} . The mass density is

$$\rho_B(r) = \rho_A(r) + \left(\frac{M_T}{4\pi R}\right) \frac{1}{r^2} \theta(r - R). \quad (2.3)$$

So the total mass in case B in a sphere of radius r increases with r , even for $r > R$.

Now we can compute the rotational curves for the two galaxies. An object of a certain mass at a distance r from the center moving in a circular orbit with a velocity $v(r)$ satisfies the equation

$$\frac{v(r)^2}{r} = G_N \frac{M(r)}{r^2}, \quad (2.4)$$

Then the velocities of the two mass distributions are given by

$$\begin{aligned}
 v_A(r) &= \sqrt{\frac{G_N M_T}{R}} \left[\frac{r}{R} \theta(R-r) + \sqrt{\frac{R}{r}} \theta(r-R) \right] \\
 &\sim \frac{1}{\sqrt{r}} \quad \text{for } r > R, \\
 v_B(r) &= \sqrt{\frac{G_N M_T}{R}} \left[\frac{r}{R} \theta(R-r) + \theta(r-R) \right] \\
 &\sim \text{const} \quad \text{for } r > R.
 \end{aligned} \tag{2.5}$$

$v_A(r)$ and $v_B(r)$ characterize respectively the two velocity curves of Fig. 2.1. For $r < R$, both velocity curves raise linearly with r . For $r > R$, curve A decreases like \sqrt{r} , while Curve B stays constant. Note that the mass distribution M_B can be cut off at some radius $R_c > R$ in order to have a finite total mass for the system. But the rotation curve of an object inside the sphere of radius R_c is not affected. So the dark matter puzzlement is that the observed mass distribution is given by ρ_A while the observed rotation curve is given by v_B . The dark matter makes up the difference between the two mass distributions.¹³⁾

Another feature we should note is that if we only take into account of the mass distribution $\rho_A(r)$, i.e., the mass distribution is confined within a sphere of radius R , but take the rotation curve $v_B(r)$. Then something peculiar will happen: all objects outside the sphere of radius $2R$ will have velocities greater than the escape velocity. This confined mass distribution does not have enough gravitational force to hold on the objects of constant rotation velocity too far out. In a physical galaxy this is to say that these stars cannot be bound to the galaxy. For the mass distribution $\rho_A(r)$, the escape velocity v_{esc} can be calculated straightforwardly. Let the escape velocity happens at a radius r_e , where the total energy of the start vanishes:

$$\begin{aligned}
 \frac{1}{2} m v_{\text{esc}}^2 - G_N \frac{m M_T}{r_e} &= 0, \\
 v_{\text{esc}}^2 &= \frac{2 G_N M_T}{r_e}.
 \end{aligned} \tag{2.6}$$

Comparing v_{esc} with v_B given in Eq. (2.5), we have

$$r_e = 2R. \tag{2.7}$$

In observation, however, stars in a typical spiral galaxy are found to exist at radius of many times of R where the rotation curve flattens out.

The features discussed above can also be applied to clusters of galaxies. Galaxy clusters are gravitationally

¹³⁾Our example as given is artificial as the difference of the two mass distributions happens in the region $r > R$. We can take M_T in mass distribution B larger than that in mass distribution A so that there can have dark matter inside the sphere of radius R .

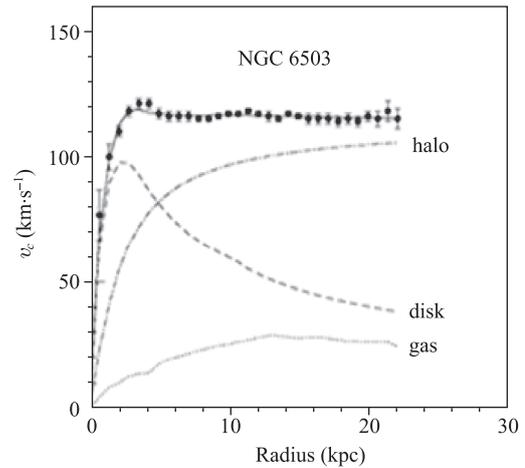


Fig. 2.2 The fitted total velocity curve is given by $v_c = \sqrt{v_{\text{halo}}^2 + v_{\text{disk}}^2 + v_{\text{gas}}^2}$.

bound; i.e., galaxies within a cluster orbit one another. For example, the Milky Way is a member of a small cluster of more than 50 members, dominated by the Milky Way and the Andromeda galaxy (designated as M31 or NGC224). This cluster is called the Local Group.¹⁴⁾ Figure 2.2 is a plot of the observed rotation curve of the spiral galaxy cluster NGC 6503.^{15),16)} The dashed line, labeled by “disk”, is the rotation curve based on the visible mass of the galaxy disk. The gas distributed in the galaxy gives a rotation curve as show by the light solid line labeled as “gas”. The two together are far short of the observed curve which consists of the data points fitted by a solid line, especially at large radius. The dash-dotted curve labeled as “halo” is the velocity contribution from the expected dark matter. The dark matter dominates the halo.

From their rotation curves, some of the galaxies, such as Low Surface Brightness (LSB) galaxies, are dominated by dark matter which makes 95% of these galaxies. They are diffused galaxies with surface brightness an order of magnitude lower than the ambient night sky viewed from

¹⁴⁾The Andromeda galaxy is the nearest spiral galaxy to the Milky way, about 1.5 million light-years from Earth. Andromeda and the Milky Way, together with the Triangulum galaxy and about 50 other smaller (dwarf) galaxies form the Local Group. Andromeda and the Milky Way dominant the Local Group. Andromeda is the largest galaxy in the Local group, containing about 10^{12} stars more than those in the Milky way which has about 200-400 billion (10^{11}) stars.

¹⁵⁾NGC 6503 is a dwarf spiral galaxy which spans about 3×10^4 light-years and is approximately 17 million light-year away from us. A dwarf galaxy has a fewer number of stars than the Milky Way, less than a few tens of billions, and therefore a fraction of the mass.

¹⁶⁾NGC stands for “New General Catalogue” of deep sky objects. Astronomic objects are designated with letters and numbers. NGC together with numerals is one of the designations. The others are IC (Index Catalogue) and M (Messier). The Andromeda galaxy mentioned above is denoted as NGC 224 or M31.

Earth. They are mostly dwarf galaxies with baryonic matter in the form of neutral gaseous hydrogen rather than stars. See Fn. 2.1. The Milky Way consists of many satellite dwarf galaxies many are discovered in the past decade.

2.2 Large scale structures, hot and cold dark matter

The large structure of the universe has been investigated in galaxy surveys and theoretical studies. Galaxy surveys, also known as redshift surveys, measure the number of galaxies per unit of solid angle in redshift bins, resulting in a 3D map of the distribution of matter within a field of the sky for the study of statistical properties of large scale structure of the universe. Together with cosmic microwave background (CMB) measurements, strong constraints on cosmological parameters, such as the average matter density and the Hubble constant can be obtained. There are a number of galaxy surveys completed or in progress: the CfA Redshift Survey, 2dF Galaxy Redshift Survey, Sloan Digital Sky Survey, DEEP2 Redshift Survey, and VIMOS-VLT DEEP Survey. Theoretical studies use large numerical simulations based on theoretical models to predict the structure of the universe. Below is a brief sketch of the results.

Let us first start at the very large scale of the universe. In the very large structure the universe looks uniform. But galaxies, which have various sizes and different numbers of stars, do not have a uniform distribution like dots on a gride. The main feature of the large structure of the universe is that of walls and voids:

- Walls: Walls are cosmic regions which contain a typical cosmic mean density matter abundance.¹⁷⁾ Furthermore, the wall possesses two sub-features;
 - Clusters: These are highly matter concentrated zones where walls meet and where one finds galaxy clusters, which in turn consist of many galaxies.
 - Filaments: They are branching arms of walls. These arms can greatly stretched for tens of megaparsecs (3.262×10^6 lyrs or 3.086×10^{22} m).
- Voids: Voids are vast regions which have very low cosmic mean density, less than one tenth of the average density of the universe. They are commonly 10 megaparsecs stretched in a given directions.

The walls and filaments can be broken down further into superclusters, clusters, galaxy groups, galaxies, and stars. A numerical simulation of the large structure of the

¹⁷⁾The Cosmic mean density is usually defined as number of galaxies, not the total matter mass, in a unit volume.

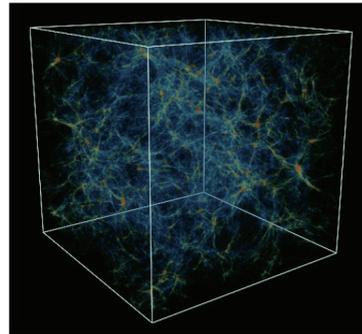


Fig. 2.3 A numerical simulation of the large structure of the universe. Notice the presence of walls, filaments, and voids. The red spots where walls intersect are expected to contain clusters and superclusters. The simulation contains both cold and hot dark matter. See later discussions of the current section.

universe is shown in Fig. 2.3¹⁸⁾, which shows the features mentioned above.

If we start looking from the smaller scale to ever larger ones, we see that stars tend to congregated into galaxies, galaxies to groups, to clusters, etc. In general clusters are made of 50 to 1000 galaxies, having a diameter up to 10 megaparsecs. A well-known super-structure is the Sloan Great Wall observed by the Sloan Digital Sky Survey (SDSS). This hierarchical structure resembles that of the very small scale of physics, starting from quarks to go on to hadrons, in particular, the stable protons and neutron, atoms and molecules, then to visible larger classical objects. Here we see the necessity of dynamics in the case of the cosmos. This is of course the gravity with proper mass distributions.

In an expanding universe, the space is constantly stretched and the distance between any two points fixed on the fibre of the space will be ever increasing. Without some force to hold them up they will be forever separated. Individual ordinary objects made of protons, neutrinos, and electron, or a group of atoms, are bound up by electromagnetic force, while a group of stars are held together by their gravity if sufficient mass are contained in them. There are different scenarios within the framework of the theory of gravitational instability for the formation of structures from small matter density fluctuations. Typical scenarios are the clustering *bottom-up* called the *hierarchical clustering theory*, the fragmentation *top-down* called the *pancake theory*, and the *hybrid scenarios* somewhere between the two extremes.

The pancake scenario ascertains that in the early universe radiation smooths out matter fluctuations to damp out small-scale fluctuations so as to form supergalactic pancakes. The pancakes accrete matter to grow and

¹⁸⁾This figure is adopted from the website: <http://www.astro.virginia.edu/jh8h/Foundations/chapter14.html>

eventually collapse to fragment into galaxies. This type of theories predicts the existence of large sheets of galaxies with low matter density voids between galaxy sheets, while clusters of galaxies will form at the intersection of the sheets. In the hierarchical clustering theory, the order of the formation of the cosmic structure is reversed. Fluctuations survived on all relevant scales. Small-scale system, i.e., star-cluster sized objects of small galaxies, are formed first. Then by clustering due to gravitational attraction, they merge into systems of larger and larger sizes, e.g., galaxies, galaxy clusters, filaments, and walls. Voids also appear.

All theoretical scenarios require the presence of a significant amount of dark matter, up to 90% of the total mass, of different kinematic type for different scenarios. Crudely one can argue for the existence of the dark matter in the following way: In a cluster, for instance, each galaxy is moving at some velocity relative to others. However, generally kinetic energies of the galaxies in the cluster appear to be large enough that the visible matter in the cluster is not sufficient to provide enough gravitational potential to prevent individual galaxies from escaping so as to hold the cluster together. Aside from revising general relativity, a natural way to explain why large galaxy clusters can exist is that there is a large quantity of non-luminous matter present in the cluster to make up the apparent deficiency of the baryonic mass. Furthermore, ordinary baryonic matter had too high a temperature, and hence too much pressure, left over from the Big Bang. Therefore, with baryonic matter alone it is difficult for the aggregated mass to collapse to form smaller structures.

The bottom-up scenario requires that dark matter, when made of particles, consists of heavy, slow moving, non-relativistic particles, which are known as the *cold dark matter* (CDM). The CDM would encourage matter to clump into small regions. Hence smaller, dwarf galaxies are predicated to permeate cosmically. The top-down scenario requires dark matter to be composed of weakly interacting relativistic particles known as *hot dark matter*. Because of their high velocity small structures can be smoothed out to form large pancake structures to start structure formation. The cold and hot dark matter models predict different large structures of the universe as depicted in Fig. 2.4¹⁹⁾.

Observation shows that stars in galaxies are of 10 to 14 billion year old, and many clusters are still in the process of formation. This gives evidence that the universe proceeds hierarchically, galaxies are formed before clusters and hence favors the bottom-up scenario. This bottom up scenario of structure formation requires cold dark matter to make it to work. Baryonic matter had

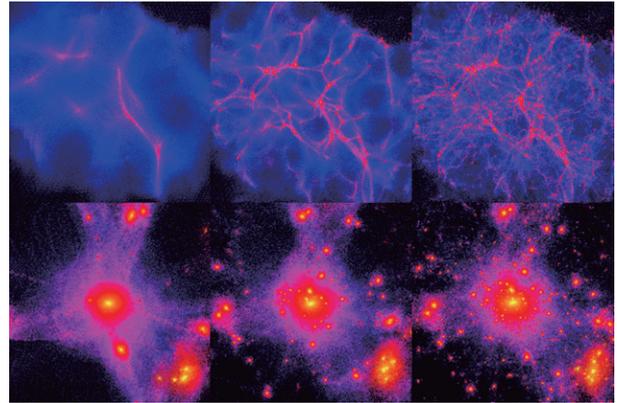


Fig. 2.4 Numerical simulations of the large structure of the universe. The upper panel is the cold dark matter model, the lower panel is the hot dark matter model.

too high a temperature, and therefore too much pressure, left over from the Big Bang, to collapse to form smaller structures. The dark matter serves as the compactor, enabling the collapse of matter to form structures.

Large computer simulations, involving billions of dark matter particles, have been carried out [28]. They confirm that the cold dark matter model of structure formation is consistent with the large-scale structures observed through galaxy surveys, such as the Sloan Digital Sky Survey, 2dF Galaxy Redshift Survey, and Lyman-alpha forest. These studies have been crucial in constructing the Λ CDM model.

Although the cold dark model is successful with large-scale structures of the universe, it faces a number of difficulties of small-scale problems related to the galaxy formation, such as missing satellites, too-big-to-fail, cusp-core, etc. In general the cold dark matter model predicts too many dwarf galaxies contradicting to the observation. But there are resolutions to all or some of these problems from the inclusion of baryonic physics of feedback or dynamical fraction; or the inclusion of some hot component in the dark matter, such as the warm dark matter.

2.3 Gravitational lensing, collisions of galaxy clusters

Gravity lensing refers to the phenomenon of light bending around a massive object, such as a cluster of galaxies, as a consequence of general relativity. It is a geometric effect sensitive to all types of matter, independent of any astrophysical assumptions. It is a measurement of mass not resorting to dynamics. Gravitational lensing distorts the image of background objects, such as a galaxy, in two ways, called convergence and shear. The convergence magnifies the background object by increasing its size and the shear stretches the image of an object tan-

¹⁹⁾The image is adopted from <http://burro.astr.cwru.edu/Academics/Astr222/Cosmo/Structure/darkmatter.html>

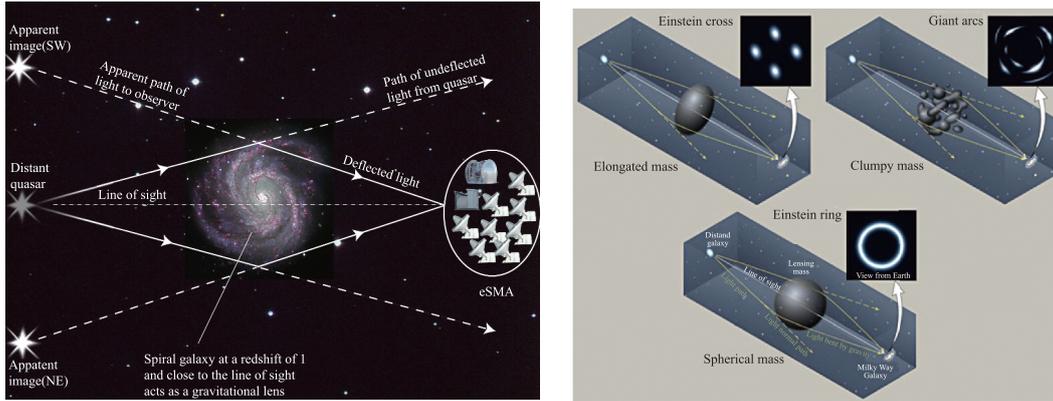


Fig. 2.5 Artists' rendition of the gravity lensing effect. The right panel shows the possibility of different lensing images depending on the mass distribution of the foreground that give rise to the lensing images.

gentially around the foreground mass.

Normally, the gravitational lensing effect is used to reveal the existence of galaxies in the background eclipsed by objects in the foreground, such as a galaxy or a cluster of galaxies. In the probe of dark matter, the roles of the foreground and background matter distributions are switched. The appearance of lensing images of background galaxies reveals the existence of a foreground mass distribution which is otherwise invisible. The foreground mass distribution is generally referred to as the *lensing mass*. In summary: For the dark matter search, the gravitational lensing

- provides a unique probe to the distribution of matter, luminous or dark;
- serves as a calibrator for the relation between luminous and dark matter.

In general, a galaxy or cluster of galaxies is composed of three components of material: solid stars, diffused gas, and dark matter. Each can be studied in a unique way: Stars are visible optically, gas by x-rays, and dark matter by gravity. The gravity effect can be studied, besides the rotation curve, by gravitational lensing after the gravity effect of stars and gas is separated out. Gravitational lensing is proved to be an effective, robust tool for probing dark matter and investigating some of the properties of dark matter. It provides many of the more recent new discoveries on behaviors of dark matter as revealed in collisions of galaxy clusters.

2.3.1 Gravitational Lensing and dark matter

Figure 2.5²⁰⁾ depicts the gravity lensing effect that a background star, which lies behind a visible foreground galaxy or galaxy cluster, can reveal itself as mirage images together with the visible foreground objects. The

²⁰⁾The figure on the left is adopted from <http://stevenasimpson.com/illo9.html>.

left panel shows the lensing of a star which forms two images. Viewed from a suitable angle, multiple images more than two can also be seen when the lensing mass is sufficiently large. The right panel shows the lensing of a galaxy in the background of a cluster of galaxies. The cluster with its huge mass of invisible dark matter makes easily identifiable lensing images. As the foreground is so massive, it bends light rays emitted from a background galaxy significantly by about half a degree, producing a strong warp in the images. In this example the lensing image forms a (blue) ring, called the *Einstein ring* which is formed by the multiple images of the same distant background galaxy. This is an example of the *strong gravitational lensing*, in which multiple images, arcs, or Einstein rings can be formed when the lensing mass density greater than a critical value.

As an example of the strong lensing effect an image observed by the Hubble Space Telescope is shown in left panel of Fig. 2.6.²¹⁾ The foreground is the galaxy cluster CL0024+1654 which is about 4 billion light years away and shows as yellow bright spots. Also shown are multiple blue images of a single background galaxy. The huge mass of the invisible dark matter in the foreground cluster bends the light of a background galaxy to form five images shown in blue, one in the center and four arrayed around as the Einstein ring [29]. The image shown on the right panel is the so-called smiley face of the Hubble image of the galaxy cluster SDSS J1038+4849.

From Earth, along most lines of sight, the deflection of the star light from a single background source is often too weak to be detected. A technique called Weak gravitational lensing [30] has been developed. The technique applies statistical analysis in vast galaxy surveys. By a systematic alignment of background sources, the lensing

²¹⁾The figure on the right was discovered among Hubble's data by the amateur astronomer Judy Schmidt and was submitted to the Hubble hidden treasure competition. It first appeared in the Washington Post on March 9, 2025.



Fig. 2.6 **Left:** NASA images of galaxy cluster CL0024+1654 from Large Synoptic Survey Telescope (LSST): <http://www.lsst.org/lsst/public>. **Right:** Hubble image of galaxy cluster SDSS J1038+4849. The “eyes” are two galaxies.

mass’ minute distortions of images of the background galaxy can be analyzed. In examining the apparent image deformation of the background galaxies, the mean distribution of dark matter can be characterized statistically, and the mass-to-light ratios that correspond to dark matter densities predicted by other large-scale structure measurements can be determined.

A strong evidence of dark matter is the Hubble Space Telescope image of the galaxy cluster CL0024+17, also known as Zw CL0024+1652, as shown in Fig. 2.7. The left panel shows the Hubble images of the foreground clusters and the lensing. The most visible part of the images is many spectacular galaxies that are part of CL0024+17, typically appearing tan in color. Almost all small brown spots are a galaxy. Next, around the center of the cluster shows several unusual and repeated galaxy shapes, typically shown in blue. These are the multiple images of distant background galaxies, resulted from the lensing of the foreground cluster. In addition to these prominent features, there are also many distorted faint images of the distant background galaxies, shown all over in blue, which indicate the presence of the dark matter.

The right panel of Fig. 2.7 shows the imposition of the dark matter image together with the Hubble image. The dark matter distribution is shown as diffused blue images around the center region and as a ring-like band. Because of their mutual gravitational attraction, dark matter and visible material are generally expected to be together. In most regions of the universe this is found to be the case [31]. The dark matter distributions shown in the right panel of Fig. 2.7, however, does not match with that of the stars and hot gas. This is the first time that dark matter has been found to be in a distribution that is substantially different from that of the ordinary matter. The ring is about 2.6 million light

years across and located about 5 billion years away from Earth, when the universe is only 2/3 of its present age. This discovery, announced in May 2007, is considered to be among the strongest evidence of the existence of dark matter.²²⁾ An interpretation of the formation of the huge dark matter ring is that it is a transient feature formed when CL0024+17 collided with another cluster of galaxies about one billion years ago with the collision axis happened to line up with the line of sight of the earth today. Due to the gravitational effect, a ring distribution of dark matter is left as a result. A publication five year prior [32] has already discussed the collision of clusters during their history. This kind of dark matter distribution is very difficult to explain in terms of a modified theory of gravity [33], which offers an alternative interpretation of the observed behavior of the galaxy rotation curve.

2.3.2 Collisions of clusters and galaxies

Studies of collisions between galaxies and clusters of galaxies are robust tools in confirming the existence of dark matter and to learn dynamic properties of dark matter self-interactions, giving the fact that dark matter have very weak interaction with ordinary matter. Furthermore, collision studies are also excellent testing ground for ideas about dark matter.

When two groups of dark matter pass through each other, say the dark matter halo of a galaxy moves through a background dark matter particles, self interaction of the dark matter can lead to the deceleration and evaporation of the halo. In both occurrences the centroid of the galaxy can be shifted post-collision. To

²²⁾See, the Hubble news release made on May 15, 2007 (#STSci-2007-17) <http://hubblesite.org/newscenter/archiv/releases/2007/17/>.

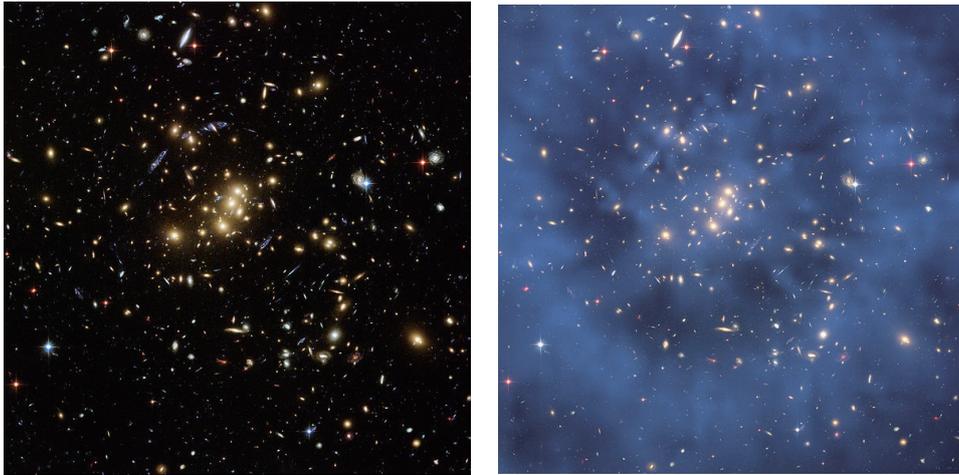


Fig. 2.7 **Left:** The Hubble image of the CL0024+17 (ZwCL0024+1652) galaxy cluster. **Right:** The gravity map superimposed on the Hubble image which shows a dark matter distribution in the central region and a thick ring. Credit: NASA, ESA, M. J. Jee and H. Ford, Johns Hopkins University.

be more explicit, if the dark matter self-interaction gives rise to frequent interactions but with very small momentum transfers, e.g., due to long range interactions by the exchange of light particles in the intermediate state, deceleration of the dark matter halo will be resulted because of the creation of a drag force as a result of the interaction. The dark matter emerging from the collision will be slowed down. However, in the opposite case of rare interactions with large momentum transfer, e.g., the very short range contact interaction, due to the exchange of heavy intermediate state particle, evaporation of dark matter will be seen as the final state dark matter particle can be scatter away and get lost. In both situations, the collision can never be violent enough to completely separate the dark halo from the original galaxy. In majority cases, the matter components of either group, clouds of gas, stars, and dark matter remain bound by the same gravitational potentials pre- and post-collision. We refer to [34] for more details.

Below we briefly describe three most fruitful studies of collisions of clusters of galaxies to date. The first one concerns the well-known Bullet cluster in 2004 and the next two are new results published in early 2015.

2.3.2.1 The Bullet cluster

Another strong evidence of the dark matter from gravitational lensing is the Hubble's observation of the Bullet cluster of galaxies (1E 0657-56) which consists of two colliding clusters of galaxies in the process of moving away from each other. This is shown in Fig. 2.8. The detailed mass distributions mapped out by lensing effect indicates that the distribution of the baryonic matter is different from that of the dark matter. This feature can be understood as the result of a close encounter of the two

galaxies. As they passed through each other at a speed of 16 million km per hour, the luminous parts of the two clusters interact and are slowed down. The dark matter components of the two clusters do not interact significantly except for the gravitational effect, they passed through each other without much disruption. This difference in interactions causes the dark matter parts to move ahead of their luminous counterparts, separating each cluster into two components: dark matter in the lead and luminous matter lagging behind.

In X-ray observations, which map the gas distribution, indeed, they show that much of the baryonic matter in the system is concentrated in the center of the visible system, while weak gravitational lensing observations show that much of the total mass, mostly dark matter, resides outside the central region of the baryonic gas. This is shown in Fig. 2.8. The left panel of Fig. 2.8 shows the image of the visible galaxies together with the fitted dark matter distribution. Most of the baryonic mass is in the hot gas detected by the Chandra X-ray Observatory and shown in pink. The dark matter, which makes up most of the total mass and dominates the gravitational effect, is shown in blue. As exhibited in the figure the mass ratios of dark matter to gas to stars is 70:10:1. The separation of the distributions of dark and baryonic matter is a result of the collision. The right panel illustrates the distributions of the three types of matter. In this example, the collision axis is perpendicular to the earth line of sight.

The observation of the mass distribution of Bullet cluster was announced in August 2006, providing the best evidence to date for the existence of dark matter. The news release can be found in [35] and the published articles in [36]. This spatial offset of the center of the

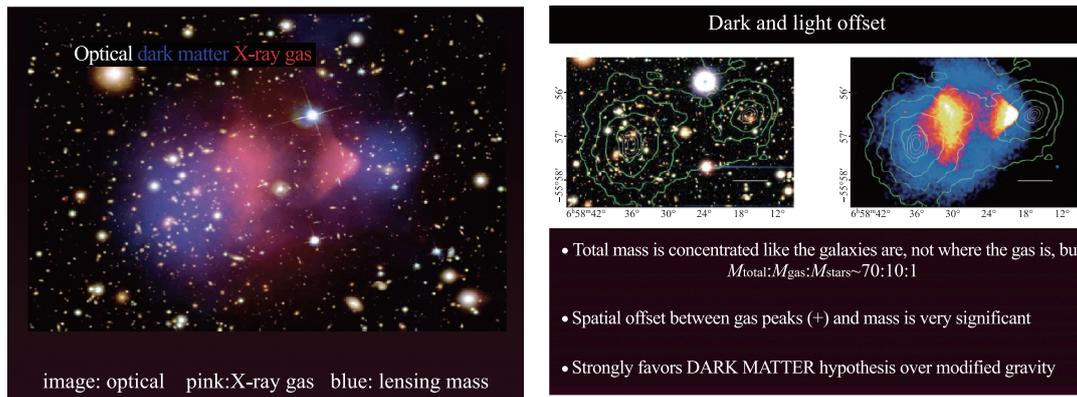


Fig. 2.8 The Bullet cluster which shows the mass distribution after the collision of two galaxies. The organ color is the baryonic matter and blue the dark matter. X-ray: NASA/CXC/CfA/M.Markevitch *et al.*; Optical: NASA/STScI; Magellan/U.Arizona/D.Clowe *et al.*; Lensing Map: NASA/STScI; ESO WFI; Magellan/U.Arizona/D.Clowe *et al.*

total mass from the center of the baryonic mass peaks is difficult to explain by an alteration of the gravitational force law.²³⁾ Observations of other galaxy cluster collisions, such as MACS J0025.4-1222 [39], also show significant displacement between their center of visible matter and their gravitational mass. The lack of deceleration of the dark matter provides a constraint on the dark matter self-interaction strength which is expressed in terms of the self-interaction cross section for long range forces $\sigma_{\text{DM}}/m < 1.25 \text{ cm}^2/\text{g}$ (68% confidence limit or 1σ significance).

Since the Bullet cluster, half a dozen of galaxy cluster collision have been discovered. There are inherent restrictions in the analysis of galaxy collisions due to uncertainties in the determination of the original mass of the clusters and the 3D collision geometry. The latter affect the determination of the angle of the motion in relation with the line of sight of observation, the impact parameter of the collision, and the impact velocity. No tighter constraints have been obtained from the newer studies.

2.3.2.2 Statistical analysis of multiple cluster collision events

The situation has changed more recently. Broad range of studies have been carried out and interesting results extending the study of dynamic behaviors of dark matter obtained. In [40] the authors, using data from Chandra and Hubble Space Telescopes, have investigated 72 cases of collisions between galaxy clusters, including both major and minor mergers, at different angle to Earth observers and taking place at different times. Combining

these measurements statistically, the study confirms the existence of dark matter at the statistical significance of 7.6σ .²⁴⁾

Multiple views of the collisions have been studied by tracking gas clouds by Chandra X-ray image, visible stars by Hubble telescope, and the dark matter by gravitational lensing. The distributions of the three types of matter can thus be mapped out. It was found that the gas clouds are greatly slowed down or even stopped by the collision as expected due to their interactions that produced a drag effect. The stars mostly slide passing one another unless suffering headon collisions which is extremely rare, also as expected because stars interact only gravitational among one another. The study found that the dark passes through every thing obviously unaffected, similar to the stars. This indicates not only very weak interactions of dark matter with ordinary matter, but also that there is also very little self-interactions among dark matter particles themselves besides gravity. This allows a more stringent bound to be set for the dark matter self-interaction cross section: $\sigma_{\text{DM}}/m < 0.47 \text{ cm}^2/\text{g}$ at 95% CL, which is a significant improvement from the Bullet result. Based on this limit, some of the dark matter model proposed in the literature will be ruled out. We will come back to it more later. A cartoon to show the relative positions of the three matter components is given in Fig. 2.9 which is taken straightly from the original article [40].

²³⁾M. Milgrom, the proposer of the modified Newtonian dynamics (MOND) claimed otherwise. See his refutation in an on-line article [37]. A detailed study of various versions of MOND in light of the Bullet cluster can be found in [38].

²⁴⁾The percentage confidence limit and the number of σ significance are related as follows. Let the number of σ be denoted as n_σ and the corresponding confidence limit as $CL(n_\sigma)$. Then $CL(n_\sigma) = \text{erf}(n_\sigma/\sqrt{2})$, where $\text{erf}(z) = (2/\sqrt{2}) \int_0^z \exp(-x^2)dx$ is the error function. So for 7.6σ significance the confidence limit is $CL(7.6) = (1 - 2.964) \times 10^{-14}$. In high energy physics a discovery event requires no less than 5σ which corresponds to a confidence limit $(1 - 5.733) \times 10^{-7}$.

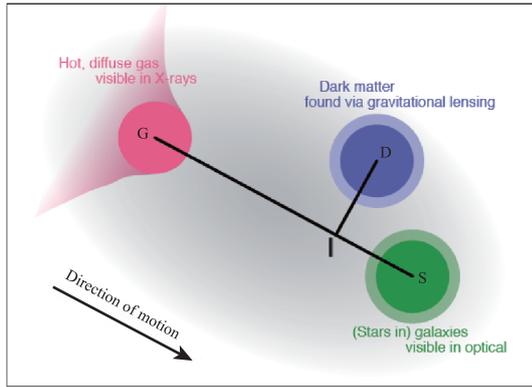


Fig. 2.9 A typical post-collision positions of the three types of matter of a galaxy cluster. This cartoon is taken directly from [40] where detailed explanation of the cartoon can be found.

2.3.2.3 A quartet colliding galaxies in Abell 3827

A simultaneous collision of four galaxies in the galaxy cluster Abell 3827 has been carried out using images from Hubble and from the European Southern Observatory’s (ESO) Very Large Telescope (VLT). The study [41] traces out the mass lying within the system and compare positions and distribution of the luminous stars with that of the dark matter. The image given in the article is reproduced in Fig. 2.10.²⁵⁾

As shown in Fig. 2.10 in one of the four galaxies, i.e., the one on the left marked as N1, the dark matter clump lags behind the visible part of the galaxy it surrounds by 5000 light years, corresponding to 4.7×10^{10} m. This striking feature that the mass associated with the galaxy N1 is offset from its stars has never been observed in a system involving a single galaxy. The offset, in the present case involving four galaxies, can be understood as a result of dark matter self-interaction, in addition to gravity, although an astrophysical interpretation cannot be completely ruled. Astrophysical effects aside, this is the first signal of non-gravity self-interaction among dark matter ever being observed. The effect can be translated into a self-interaction cross section

$$\sigma_{\text{DM}}/m = (1.7 \pm 0.7) \times 10^{-4} \left(\frac{t_{\text{infall}}}{10^9 \text{ yrs}} \right)^{-2} \text{ cm}^2/\text{g}, \quad (2.8)$$

where t_{infall} is the infall duration which must be less than the life time of the universe at the redshift of the cluster, $z = 0.099$, i.e., $\sim 10^{10}$ yrs. But likely $t_{\text{infall}} \lesssim 10^9$ yrs,

²⁵⁾This figure is taken from the ESO news release at <http://www.eso.org/public/usa/news/eso1514/#4>, while the remarks on the Milky Way stars and the lensing images are added in relation to Fig. 1 of [41]. One can find a 1 minute video on the website.

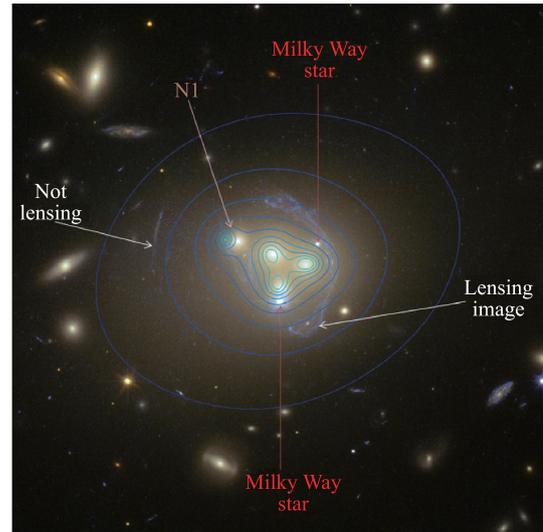


Fig. 2.10 Hubble image of galaxy cluster Abell 3827. The dark matter distribution in the cluster is shown with blue contour lines. The dark matter clump for the galaxy at the left is significantly displaced from the position of the galaxy itself, possibly implying dark matter-dark matter interactions of an unknown nature are taking place.

a lower bound on the interaction cross section can be obtain:

$$\sigma_{\text{DM}}/m \gtrsim (1.7 \pm 0.7) \times 10^{-4} \text{ cm}^2/\text{g}. \quad (2.9)$$

This dark matter non-gravity self-interaction appears to contradict the result of the recent survey of 72 galaxy cluster collisions discussed earlier. The previous result indicates that dark matter interacts very little, if at all, with itself other than gravity.²⁶⁾ But the two studies focused on systems of very different scales. The newer study zooms in on individual colliding galaxies, while the earlier paper on entire clusters which may not be so sensitive to the small offset found in galaxy collision. It could also be that the collision in the newer study has lasted longer than the previous study to allow even a tiny effect to build up and become observable. The two studies have provided a fascinating insight to dark matter properties for their possible self-interactions over two different cosmic scales. Taken literally, the two studies have provided the upper and lower bounds on dark matter non-gravitational. Since this is the first studies of its kind we expect more observational studies to follow and refined numerical simulations to be performed to clarify the situation.

²⁶⁾We note that there is a large overlap among authors of the two articles.

2.4 Dark matter in galaxies and a cosmic map of dark matter

2.4.1 Dark matter halo

Up to now the dark matter has been discussed qualitatively. The arguments for its existence based on the line of evidence in the whole spectrum of astrophysical objects is compelling, ranging from galactic scale of dwarf galaxies and spiral galaxies, to much a larger scale of galaxy clusters, and to the cosmological scale. Quantitatively, dark matter is now part of the standard theoretical framework for the study of all relevant cosmological behaviors, such as the critical questions of structure formation. In this framework the universe's energy-mass budget is dominated by a cosmological constant and the cold dark matter. Small density perturbations will then grow, due to gravitational instability, to form the bound structure known as the dark matter halo, which in turn provides the platform for the detailed structure of the universe to be developed. In a rigorous mathematical form, this framework also serve as a test of the dark matter paradigm of cosmology.

There are explicit halo profiles traditionally used in numerical simulations. One commonly used is the two-parameter NFW profile ρ_{NFW} [42, 43] and its generalization of three parameters ρ_{gNFW} . The two-parameter NFW profile takes the form

$$\rho_{\text{NFW}}(r) = \frac{\rho_c \delta_c}{\frac{r}{r_s} \left(1 + \frac{r}{r_s}\right)^2}, \quad (2.10)$$

where r is the radial distance to the center of the galaxy, ρ_c is the critical density, δ_c is a dimensionless constant, and r_s is a characteristic radius. The generalized NFW profile is

$$\rho_{\text{gNFW}}(r) = \frac{\rho_c \delta_c}{\left(\frac{r}{r_s}\right)^\gamma \left(1 + \frac{r}{r_s}\right)^{3-\gamma}}. \quad (2.11)$$

Another profile is the Einasto profile [44, 45]

$$\rho_{\text{Einasto}}(r) = \rho_0 \exp\left(-\frac{2}{\alpha} \left(\left(\frac{r}{r_s}\right)^\alpha - 1\right)\right), \quad (2.12)$$

where ρ_0 and α are constant. A comparison of the NFW and Einasto profiles in the Planck data can be found in [46]. A fourth profile, originally proposed for halos of dwarf galaxies is given in [47], having the form

$$\rho_{\text{Burkert}}(r) = \frac{\rho_0}{\left(1 + \frac{r}{r_0}\right) \left(1 + \frac{r^2}{r_0^2}\right)} \quad (2.13)$$

Plots to compare the first three profiles can be found in [48] and are reproduced in Fig. 2.11.

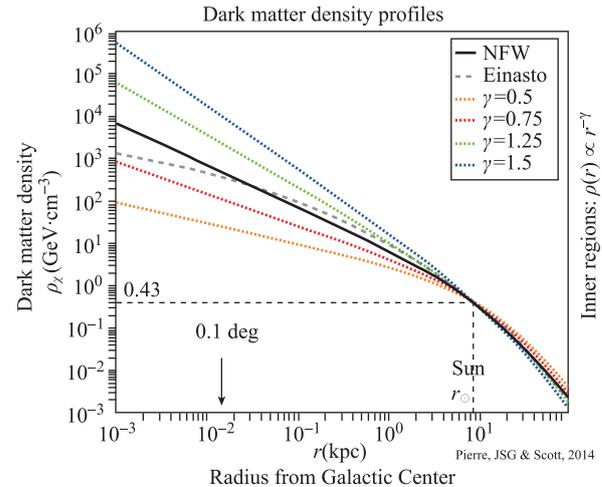


Fig. 2.11 Comparison the dark matter density profiles, NFW, gNFW, and Einasto.

2.4.2 A cosmic map of dark matter

Weak lensing enables the reconstruction of the spatial distribution of mass density across the sky for both luminous and dark matter, providing a powerful tool aiding the study of cosmology. Thus it is possible to identify the dark matter halo and to study the connection between the baryonic matter (stars and gas clouds) and dark matter by correlating the three imaging results, i.e., the lensing, optical, and X-ray maps. The first data release of a comprehensive correlated mapping was made by the Dark Energy Survey (DES) collaboration in April, 13, 2015. Details are given in [49]²⁷⁾ and summarized in the accompanying map shown in Fig. 2.12. The red areas represent the highest concentrations of dark matter, the orange and yellow the next highest. The blue areas are voids where low densities of dark matter and clusters of galaxies are found. The existence of dark matter filamentary structures in the red, orange, and yellow colors are clearly manifested. The gray dots represent galaxy Clusters, and the bigger the dot, the larger the cluster. It is also evident that the clusters are preferentially found in the reddish areas, and orange and yellow filament structures. They are scarcely found in the blue voids. So the baryonic matter, which is concentrated in clusters of galaxies, follows the dark matter density.

The map in Fig. 2.12 shows data in 3% of the total areas to be scanned in 5 years running time. The DES collaboration has measured how the dark matter and ordinary matter evolve together over the cosmic time, with the dark matter enveloping galaxies and clusters

²⁷⁾A FermiLab press release can be found in http://www.fnal.gov/pub/presspass/press_releases/2015/Mapping-The-Cosmos-20150413.html/. A scientific American report in the layman language is given at <http://blogs.scientificamerican.com/cocktail-party-physics/2015/04/13/new-dark-matter-map-confirms-current-theory>

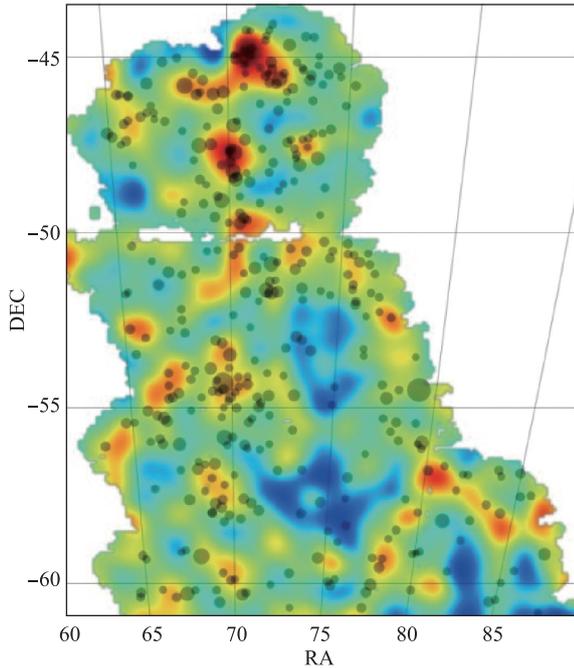


Fig. 2.12 The Dark Energy Survey (DES) Science Verification (SV) mass map along with foreground galaxy clusters.

of galaxies of various types. Much more data will be available over the next few years for more stringent tests of theoretical models and the dark matter paradigm it self. Undoubtedly, the resultant inventory of galaxies and clusters of galaxies will be very useful, e.g., providing more samples of galaxies and clusters collision for study.

2.5 Cosmic microwave background anisotropy

The cosmic microwave background (CMB) is the relic radiation of the early universe left at 380 000 years after the big bang, when photons decouple from the baryon matter and redshift down to the microwave frequency range due to the expansion of the universe over its life time of 13.81 billion years as being observed presently. The temperature is not entirely homogeneous and isotropic over the universe, although the dominant feature is. After the last-scattering, CMB photons stream freely in the universe. The relic temperature fluctuations reflects what happened in the early universe at the time of photon decoupling and manifest today as minute CMB temperature variation across the sky, known as the *CMB anisotropy*. Hence the anisotropy on a given angular scale observed today is related to the density perturbations of the early universe. The CMB anisotropy can be understood as acoustic oscillation in the photon-baryon plasma prior to the emission of the CMB with gravity providing the restoring force [17]. Baryonic matter interacts significantly with radiation whereas dark matter does not. But both types of matter can affect the

oscillations by their gravitational effect, and they produce different effects on the CMB anisotropy. The power spectrum of the CMB anisotropy, displayed in Fig. 2.13, shows clearly a large main peak and smaller successive peaks, resolved down clearly to moment $\ell = 3000$ which shows clearly 5 peaks. A detailed discussion of the CMB anisotropy and the newer data can be found in Section 12.

The peaks and valleys are signatures of relevant physics. Let us describe it briefly below. A detailed discussion is given in Section 12. The temperature fluctuation over the whole sky can be expressed by a multi-pole expansion in terms of the complete set of the angular functions of the spherical harmonics $Y_{\ell m}$. The expansion coefficients determine the power spectral function C_{ℓ} which is the vertical axis of Fig. 2.13. The horizontal axis ℓ which is the rank of the multiple moment is proportional to the inverse of the angular coverage of the sky or the angular resolution of a CMB anisotropy measurement, we have

$$\ell \approx \frac{180^{\circ}}{\theta_{\text{res}}(\text{degrees})}, \quad (2.14)$$

which is treated as a continuous variable. We describe the physics of the first few peaks below:

- The first peak centered at $\ell_1 = 200$ contains information about the total amount of energy-matter in the universe: dark energy, dark matter, baryonic matter, photons, neutrinos, and any additional stuff which might exist but not yet known. The size and location of the peak is related to the geometry of the universe. It tells us how “flat” it is. The position of the first maximum favors the total energy-matter density ratio $\Omega_{\text{total}} = 1$. Combined the data of CMB, supernovae, and large scale structure, densities of the various energy-matter components can

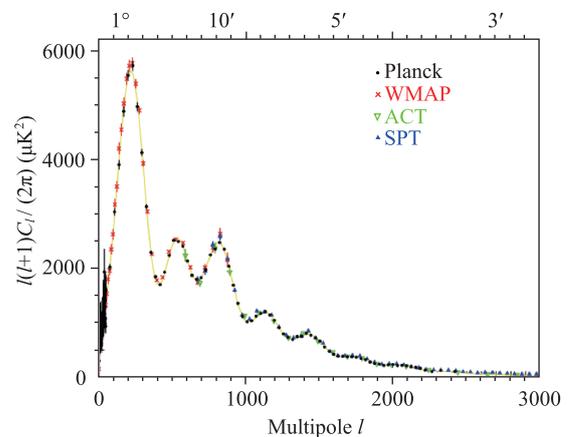


Fig. 2.13 The power spectrum of the cosmic microwave background radiation temperature anisotropy in terms of multipole moments. The plot is taken from [50], the CMB review of The Review of Particle Physics [13].

be obtained which allows the calculation of the age of the universe which be discussed in Sections 9 and 13.

- The first peak covers the ℓ -value of the order of $\Delta\ell \approx 250$, which corresponds to an angular size greater than 1° of the sky (the Sun is 0.5° on the sky!). This is much larger than the temperature fluctuations we are seeing. Hence it is not produced by sound waves in the early universe, but by the total energy density of the early universe that drives the expansion. We note that the largest component dark energy acts as a repulsive force and has the following effects: The universe expands with an increasing speed and freezes the cosmic web which smooths the Hubble flow.
- The second peak around $\ell_2 \approx 500$ corresponding to an angular resolution of $\theta_{\text{res}} \approx 0.36^\circ$ is due to sound waves, which tells us how much ordinary matter there is in the universe.
- The third peak at $\ell_3 \approx 700$ with an angular resolution $\theta_{\text{res}} \approx 0.24$ is related to both the ordinary and dark matter. The difference between the third and second peaks give us the density of the dark matter in the early universe.

As shown in Fig. 2.13, the Λ CDM model with a prominent component of dark matter fits the observation very well.

2.6 Primordial Nucleosynthesis: The baryon content of the universe

So far all lines of evidence of the existence of dark matter come from the gravitational effect. There is a direct evidence, coming from an entirely different line of argument that is not gravitational, for the existence of extra matter other than the baryonic type. This is the nucleosynthesis which is one of the most successful calculation in astrophysics based on well-established framework of fundamental physics.

The primordial nucleosynthesis, or Big Bang nucleosynthesis (BBN), refers to the production of light nuclei other than the nucleon, during the early phases of the universe. The BBN took place a short moments, the order of a hundred seconds, after the Big Bang and is responsible for the formation of a series of light elements beyond hydrogen: a heavier isotope of hydrogen, i.e., the deuterium (D), the helium isotopes ^3He and ^4He , the lithium isotopes ^6Li and ^7Li . In addition to these stable nuclei some other unstable isotopes were also produced during the primordial nucleosynthesis, such as tritium (^3H), beryllium-7 (^7Be), and beryllium-8 (^8Be). These unstable isotopes either decayed or were fused with other nuclei to form stable isotopes.

The BBN calculation predicts the abundances of the light nuclei up to the lithium, that include deuterium, helium-3, helium-4, and lithium. As stated in the Introduction, it constrains the abundance of the baryon matter to lie between 13.3% to 17% of the matter component of the universe. The rest 86.7% to 83%, i.e., the majority of the matter must be non-baryonic. A discussion of the BBN calculation is given in Section 10.

2.7 Challenges to CDM and alternatives to dark matter

2.7.1 Challenges to CDM

The cold dark matter (CDM) model consisting of weakly interacting particles has been shown to be greatly successful in the cosmological scale where gravitational force dominates in the formation of cosmic structures and their growth. On smaller scales, galactic sub-galactic, however, the simplest, dissipationless or collisionless dark matter CMD model faces challenges because some of its predictions are in tension with observations. However, tests at the smaller scales is critical for the confirmation for the CDM model. An analysis of the challenges and discussions of possible resolutions can be found in [51],²⁸⁾ The major point of contention is that the CMD model predicts dark matter overdensity around the center of the galaxy, especially in dwarf galaxies. Below is a brief summary of the challenges:

- The *cusp-core* problem. Galaxies and their clusters are assumed being enveloped essentially with spherical halos of dark matter. Cosmological simulations suggest that the halo would assuming a cusp structure in the dark matter distribution, i.e., sharply peaked at the center. The NFW dark matter profile given in Eq. (2.10) takes this form, proportional to r^{-1} near the center. However, the observed rotational velocities does not show a high concentration of dark matter at the galactic center. Most galaxies have a flat constant core dark matter profile. This cusp-core problem is mostly serious for small galaxies, such as low-surface-brightness (LSB) galaxies.²⁹⁾

²⁸⁾A summary of the article accessible to the layman can be found in the Phys.org news at <http://phys.org/news/2015-02-small-scale-cold-dark.html>

²⁹⁾The LSB galaxies are diffused, mostly dwarf, galaxies with surface brightness, view from the earth, at least an order of magnitude lower than the ambient night sky. The rotation curve measurements show a very large mass-to-light ratio, indicating that stars and luminous gas contribute very little to the overall mass budget of an LSB galaxy. No less than 95% of their mass is non-baryonic and the baryonic matter is mostly neutral hydrogen gas rather than stars. The fraction of the baryonic matter in these galaxies is much lower than the average of the universe of the baryonic matter at about 15%. The LBS shows no large overdensity of stars in the center such as the bulge of stars in the case of normal galaxies like the Milky Way. There appears no supernova activity in LSB galaxies.

- The *missing satellite* problem. The simple CDM predicts a large amount of substructures such as satellite or dwarf galaxies around a massive galaxy, each with their own dark matter halo. The physics of such a predicted feature is straightforward. The non-relativistic nature of the CDM particle tends to preserve primordial fluctuations of various scales, and hence the many small fluctuations which form subhalos and therefore substructures. So the CDM model predicts that large galaxies, such as the Milky Way, have a large number of satellite galaxies, of the order of 50–200. There were 9 dwarf satellite galaxies known to the Milky Way before 2000. However, the observation of satellite structures can be difficult because of their low surface brightness, so many subhalos may remain dark and hence undetected. Improved observations in the last decade found many more satellite galaxies for the Milky Way. In the 2005s about 15 ultra-faint satellite structures have been discovered by SDSS that covered only 20% of the sky. The very recent DES found at least 3 more dwarf satellite in 3% of their data³⁰⁾ So the missing satellite problem is greatly alleviated.
- The *too-big-to-fail* problem. The simple CDM model predicts the existence of massive dark matter subhalos, which conflicts with observations. Observed massive satellites are hosted by halos which are much less than those predicted. Since massive dark matter halos are expected to host galaxies, the case that massive dark matter halos remain to be invisible is not a viable way out of the difficulty. This problem was first identified in the Milky way [52]³¹⁾ but later was shown to be a problem common to other known galaxies [53].³²⁾
- *Solutions by baryonic physics.* Small scales are complicated by physics of galaxy formation based on complicated baryonic processes. When details of galaxy formation and the supernova feedback effect are taken into account in theoretical simulations, the overdensity problem around the center of dwarf galaxies are largely resolved for larger dwarf galaxies with stellar mass M_* greater than $10^7 M_\odot$. However, further tests for smaller dwarf galaxies, especially for stellar mass in the range of $10^6 M_\odot$ – $10^7 M_\odot$ is needed.
- *Solutions by dark matter physics-WDM.* This approach is looking for solutions in relaxing the kinematic and/or dynamic properties of the non-relativistic and collisionless dark matter. The warm dark matter (WDM) model refers to the scenario of dark matter particles with a free-streaming length comparable to the size of the region which will eventually evolved into a dwarf galaxy. Then these small density perturbations tend to be erased but more sizable perturbations are not affected. So the WDM model is similar to the CDM model for large-scale structures in cosmological size, galaxy clusters and large galaxies. But WDM model predicts less abundance of dwarf galaxies and may reduce the dark matter density in the center region of large galaxies, but the collisionless WDM can still lead to the cuspcore problem. In solving the problems of the CDM, different problems require different range of mass of the WDM particle which are non-overlapping. The overall required mass range is from 0.75 to 2 KeV. However the recent Lyman- α forest data os SDSS requires the mass of the WDM to be 4 KeV. At this mass value the WDM particle is essentially a CMD. With this constraint imposed, as concluded in [55] that the WDM does no better than the CDM.
- *Solutions by dark matter physics-SIDM.* Self-interacting dark Matter (SIDM) is a scenario in which CMD particles has weak interactions with the baryonic matter but can interact among themselves with the strength of the order the nuclear interaction [56]. The type of models are proposed to resolve the difficulties of the CDM encountered on the galactic and smaller scales. In the SIDM model, in the very dense central region of a dark matter halo the elastic scattering among CDM can be sufficiently frequent so as to redistribute the energy among the dark matter particles to create an isothermal, approximately constant core density. The elastic scattering cross section required is around $\sigma_{\text{SIDM}}/m = 1 \text{ cm}^2/\text{g}$, where m is the mass of the dark matter particle. Some early numerical studies suggested that this idea did not work. However, in recent fully cosmological simulations indicated that the SIDM model

Reference [51] presents a summary of possible resolutions to the challenges to the simple CDM model. A very recent concise summary can be found in [54]. We briefly describe them below and refer the details to these references.

³⁰⁾The Cambridge University research news release on March 10, 2015 on the discovery from the DES data of 9 new substructures associated with the Milky Way, in which at least 3 are dwarf galaxies. See <http://www.cam.ac.uk/research/news/welcome-to-the-neighbourhood-new-dwarf-galaxies-discovered-in-orbit-around-the-milky-way>

³¹⁾The name was given under the influence at the time relating to the global economic crisis. The definition of TBTF given in Investopedia is “The idea that a business has become so large and ingrained in the economy that a government will provide assistance to prevent it from failure. “Too big to fail” describes the belief that if an enormous company fails, it will have a disastrous ripple effect throughout the economy.” See <http://www.investopedia.com/terms/t/too-big-to-fail.asp>

³²⁾A nice commenting article accessible to layman, entitled “Galaxies that are too big to fail, but fail anyway” can be found in <http://www.preposterousuniverse.com/blog/2014/07/18/galaxies-that-are-too-big-to-fail-but-fail-anyway/>

appears to be much more hopeful than the WDM model. There exists a viable window in mass and scattering cross section, in the range $\sigma_{\text{SIDM}}/m = 0.1\text{--}0.5 \text{ cm}^2/\text{g}$, that the SIDM model can produce CDM halo core density files that are approximately the size for Milky Way dwarf galaxies, spiral galaxies, and galaxy clusters. Further detailed theoretical work is needed [51].

- Other solutions in dark matter physics There other approaches to reduce the central halo dark matter density, which rely on dark matter particle properties. The models include particle decay, particle-antiparticle annihilation, flavor-mixed quantum states. See [51] for references.

2.7.2 Alternatives to dark matter

The dark matter explanation of the astrophysical phenomenology presented so far is based on the strict validity of general relativity, the standard theory of gravity. General relativity and the classical Newtonian gravity have tested in the scale of the solar system, not in the galactic scale and beyond. Hence in the absence of the experimental confirmation of its existence, or the theoretical proof of the capability to resolve all observational difficulties, dark matter will not be the only game in town. Other possibilities are modifications of the theory of gravity itself. There are two different directions in the modification of the theory, which include the modified Newtonian dynamics (MOND) and quantum gravity.

MOND was first proposed in [57]. It is a non-relativistic theory and valid for the galactic dynamics, not topics of cosmology. Its subsequent development of relativistic extensions, in particular, the TeVeS (Tensor-vector-scalar) theory [58], allow it to deal with subjects, such as structure formation, etc. For a recent exhaustive review of MOND/TeVeS in relation to the current astrophysical observations we refer to [59]. See also [37] and [38]. More recently, a survey on confronting MOND/TeVeS with strong gravitational lensing over galactic scales has been performed in [60]. The conclusion is that in order to accommodate all observations, a dark matter component is needed even in the MOND/TeVeS framework.^{33), 34)}

³³⁾We should mention that there are works ascertaining that MOND is a particular case of CDM with a specific dark matter distribution function to reconcile CDM and MOND. See for instance arXiv: 0811.3143 [astro-ph] and arXiv: 1310.6801 [astro-ph.CO]. However, arXiv: 1404.7525 [astro-ph.CO], which compares the two approaches concludes that CDM and MOND are two incommensurable paradigms.

³⁴⁾The debate on CMD vs MOND is interesting. The number of MOND leaning researchers is small in comparison with those in the CDM camp which includes a sizable number of particle physicists. But there is a MOND-dedicated group severely criticizing the failures of the CDM model, in the galactic scale which MOND is every successful.

There are other alternatives proposed in the literature. The modified gravity (MOG) was proposed in 2006 [61] which has been tested against observations. For more comments, see [60].

From the view point of particle physics the dark matter scenario is very attractive. The existence of new particles, especially those in a new category, opens up a new paradigm and points to a direction in searching for the physics beyond the standard model. For astrophysics and cosmology, dark matter is a huge repay to particle physics which has been providing various tools and basic ingredients for the study of cosmology and astrophysics. Dark matter and also dark energy make possible for close connections between particle physics and cosmology, as foretelling in the third quote given at the beginning of this notes. Below we take the dark matter as given and focus on it entirely.

2.8 Finding a solution

The dark matter is a problem of multi-element and multi-dimension which requires complementary approaches to study it:

- To find out what the DM stuff is made of. This is a particle physics problem, in conjunction with searching for new physics. It involves both theoretical and experimental efforts. The search effort includes both astrophysical and accelerator experiments. This will be a focus of the present notes.
- To study in detail its effects on galaxies, groups and clusters of galaxies, etc., assuming that it exists. The tool for this effort is the N-body simulation, employing heavily calculational tool such as supercomputers. It has already yielded important information on kinematic properties of dark particles and provided important constraints on particle physics search, such as ruling out the possibility of hot dark matter, deciding on cold dark matter vs. warm dark matter, resurrecting the interest in sterile neutrinos, etc. For a recent summary of the N-body simulation on DM effects, see [62].
- It is clear that another indispensable component in this overall effort is the interface of the above approaches. To illustrate, suppose LHC found a new neutral, weakly interacting massive particle, say, a supersymmetric particle. One would naturally ask if it is a dark matter particle. It may or not be one. This particle have to be under intense study of its properties, searched for in astrophysical experiments, and goes through the test in the N-body simulation, etc. Its detailed properties, such as the mass, spin-parity, decay channels, and certain scattering cross sections can be studied at LHC. The accelerator results have to be compared and matched

up with those obtained from the dark matter search experiments in which we can be certain that the new particle involved is a dark matter particle.

3 Milky Way and galactic dark matter density profile

There are a number of reasons, in both astrophysics and dark matter particle searches, that it is important to map out the dark matter the Milky Way:

- The Milky Way allows us to make a close up observation of dark matter and provides an independent piece of information for the existence of dark matter.
- It is important for the design of experimental searches for dark matter particles. Direct searches probe the dark matter in the solar system. Indirect searches depend crucially regions of high concentration of dark matter, such as centers of galaxies and clusters.
- It enables detail comparisons of simulations and observations of distributions of dark matter subhalos which evolve into satellite galaxies.
- It allows the study of galaxy formation and evolution of large disk galaxies using Milky Way as a concrete entity.

3.1 A blurb of the Milky Way

Galaxies, which come in various sizes and shapes, are the building units of the cosmos and hence the windows to the universe. Masses of galaxies range from $10^9 M_\odot$ to $10^{13} M_\odot$, known as dwarf galaxies, medium-range galaxies, and gigantic galaxies. Dwarf galaxies contain as low as several billion of stars, and medium ranged galaxies hundreds of billion stars. Supergiant galaxies, the IC 1101 for example, contain up to a hundred trillion (10^{14}) stars and have the size expanding to the order of 200 kly. Galaxies grouping together to form clusters which contain anywhere from 50 to thousands of galaxies. Groups of clusters form superclusters. The Local Group, which comprises three major galaxies, the Milky Way, Andromeda, and Triangulum, plus more than 50 dwarf galaxies, is an outlying member of the Virgo Local Supercluster. The local group has the size of 10 Mly and the local Supercluster is about 110 Mly across. The local Supercluster centers around the Virgo Cluster which is about 10 Mpc from the Milky Way. In between Superclusters are enormous voids of space where there are few galaxies.

Our own galaxy, the Milky Way, is a medium-sized barred spiral galaxy of total mass about the mass of a trillion sun, $10^{12} M_\odot$, with a diameter about 100 kly or 31 kpc, containing 300 billion stars given or taken 100

billion. It has a central disc of the thickness of about 2 kly or 0.6 kpc. Our sun is about 26 kly or 8 kpc to the galactic center.

Below we give a cartoonist's description of some of the relevant features of the Milky Way. Let us have some idea on how the dark matter may be distributed in our galaxy and what its global environment looks like. These are shown in the cartoons of Fig. 3.1 which are taken from Refs. [63] and [64]. The left panel depicts the distribution of the dark matter over the galaxy. It is extended far out in the halo. The right panel shows some of the key features of the Milky Way galaxy. Dark matter searches have to be conducted in such a setting. Direct searches which are performed in terrestrial laboratories are the measurement of dark matter in the solar system. Indirect searches can have events which take place at other parts of our galaxy or even other galaxies far from the earth, e.g., the centers, etc., where high concentration of dark matter is expected.

Another feature of the galactic environment is the existence of satellite dwarf galaxies in the vicinity of the Milky Way and within the local group. In Fig. 3.2, some of the known satellite galaxies are shown. A number of the satellite galaxies have been discovered quite recently and many are dark matter-dominated with mass-to-light ratios approaching to 1000. With improved observations, still more satellites will become visible. Up to 9 new satellite galaxies have been discovered in the the early part of (2015).³⁵⁾

3.2 Dark matter in the Milky Way

The local dark matter radial density around the sun is, within a factor 2–3:

$$\rho_\chi \approx 0.39 \text{ GeV/cm}^3 = 6.95 \times 10^{-25} \text{ g/cm}^3 \quad (3.1)$$

as quoted in [65], the review article on dark matter given in [13]. As we have seen earlier in Section 2.4.1 that the value $\rho_\chi \approx 0.43 \text{ GeV/cm}^3$ has also been used. In comparison with the critical density of the universe $\rho_c = 1.0538h^2 \times 10^{-5} \text{ GeV/cm}^3 \approx 0.5 \times 10^{-5} \text{ GeV/cm}^3$, it shows a high concentration of dark matter in the galaxy. The distribution of the dark matter in the galaxy is modeled in several forms. The most widely applied one is the NFW (Navarro–Freek–White) dark matter radial profile [42, 43] which has discussed early in Section 2.4.1. Assuming that the motion of the dark matter is similar to that of the ordinary matter, since the solar system moves with a velocity of 240 km/s and the galactic escape velocity is no more than 610 km/s, we have that the galactic dark matter particles are non-relativistic with the beta

³⁵⁾A list of the 35 satellite of the Milky Way can be found in http://en.wikipedia.org/wiki/List_of_satellite_galaxies_of_the_Milky_Way

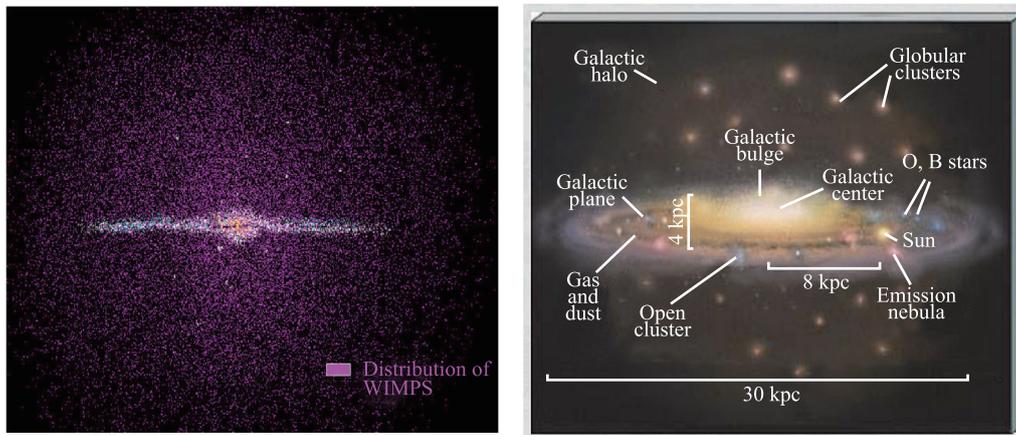


Fig. 3.1 Cartoons of the our galaxy. **Left:** The distribution of the dark matter particles. The figure is taken from Ref. [63]. **Right:** Some of the features of our galaxy which is taken from Ref. [64].

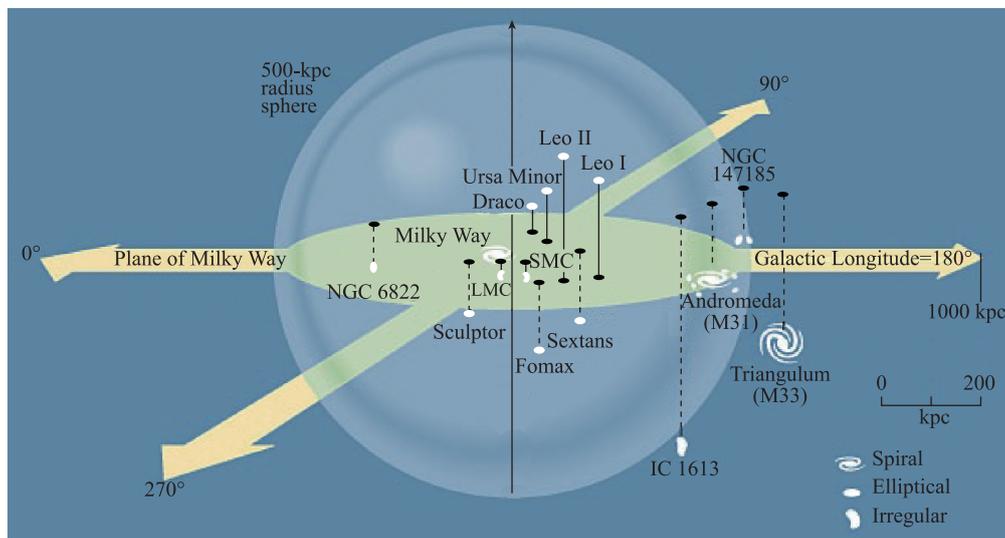


Fig. 3.2 A cartoon depicting some of the recently discovered dwarf galaxies, within a radius of 500 kpc of the galactic center, which orbit the Milky Way. The picture is taken from http://www.ualberta.ca/pogosyan/teaching/ASTRO_122/lect23/lecture23.html.

factor

$$\beta_\chi \approx \frac{v_\chi}{c} \approx 10^{-3}. \quad (3.2)$$

This number is generally used in calculation of the dark matter reaction rate.

There are several parametric forms of the radial profile of the Milky Way. Early efforts in obtaining dark matter radial profile in the Milky way are based on fittings assuming specific parametric functional forms. This approach is prone to introduce bias into the result.

Historically the extraction of dark matter distribution in the Milky Way from astrophysical observations has always been a challenge, especially in the core of the galaxy, including the region of the solar system. In the

inner region of our galaxy, which is defined as the radial region of $r < 20$ kpc following [66], while the bulk of the galactic stars lie within 18 kpc radius and the sun is about 8 kpc from the galactic center, the baryonic contribution to the galactic mass is very significant. This makes it difficult even to ascertain the presence of dark matter. The difficulty in assessing the dark matter distribution lies in the fact that the solar system, where observations are made, lies in the galactic core and moves with it, it is therefore difficult to determine the distance and the rotational velocity of other stars in the region. In addition, there is the complication that there is no general consensus on the exact distribution of stars in the Milky Way. The common practice in many studies of the dark matter in the inner Milky Way is to choose a

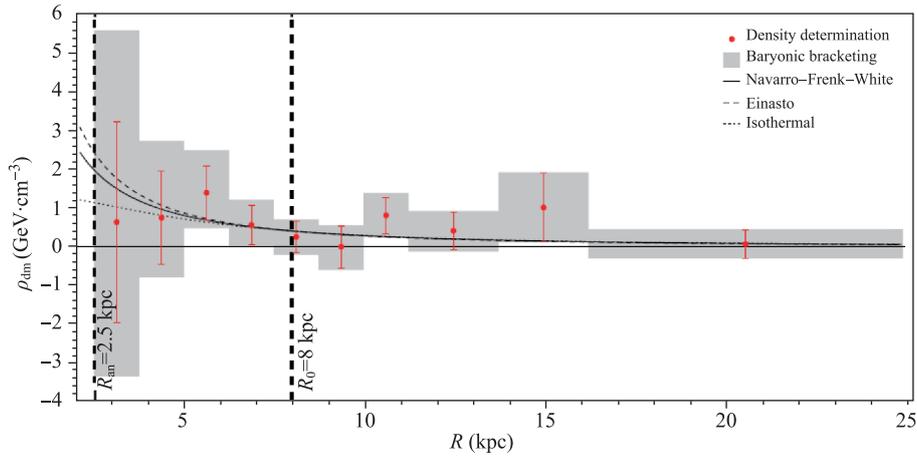


Fig. 3.3 The Milky Way’s dark matter radial profile is taken from [67]. The red points represent the 1σ measurement from observation. The grey region represents all baryonic models with their 1σ uncertainties. The curves are model profiles as indicated: The solid line is for NFW with the scale radius $r_s = 20$ pc. The dashed line is for Einasto with $r_s = 20$ pc and $\alpha = 0.17$. For more details, see [67].

model of the morphological distribution of the baryonic matter, and hence the results are likely model dependent. Hence, to date the dark matter profile of the Milky Way remains practically unconstrained in spite of persistent theoretical effort and progress in observations.

A model independent study was made in [66], in which 2780 measurements have been used to study the motion of interstellar gas and stars under various different stars morphologies to extra the baryonic part of the matter and their rotation curves. And then compare the rotational curves with the effect of the said visible mass. The authors reached the following conclusion: The need of extra matter other than visible star is manifested at all radii from 2.5 kpc to 30 kpc. The significance limit below 3 kpc is small, but it raises to 5σ for radii beyond 6–7 kpc, which includes the location where one finds the solar system.^{36), 37)}

A convincing evidence of the presence of dark matter, as discussed above, is an important step towards a detailed description of the dark matter distribution in the inner region of our galaxy. This will facilitate experiments, terrestrial and celestial, in search for the dark matter and, therefore, tests of the dark matter paradigm. Based on the latest kinematic data and observed distributions of gas and stars, without assuming a predefined forms, the dark matter distribution in the inner region

of our galaxy has been mapped out in a recent article [67]. At a given radial distance, ranging from 2.5 to 25 kp., the dark matter density mapped out varies within a sizable ranges of possible values obtained from many possible baryonic models. The commonly used dark matter density models, NFW and Einasto, fall roughly in the middle of the allowed range. In the neighborhood of the sun, the density is around $0.4 \text{ GeV}/\text{cm}^3$, consistent with Eq. (3.1). This is about 1 proton per 2.3 cm^3 , or $0.011M_\odot$ per pc^3 . The uncertainty of the density so determined is significant. Normalized to a local density ρ_L defined at the sun of $R_S = 8$ kpc to the above mentioned value, i.e., $\rho_L \equiv \rho_{NFW}(R_S) \approx 0.4 \text{ GeV}/\text{cm}^3$, the fitting for the NFW profile of Eq. (2.10), gives the parameter $r_s = 20$ kpc and $\rho_0\delta_0 = 0.314 \text{ GeV}/\text{cm}^3$. In Fig. 3.3, a comparison of the dark matter radial density profile of several fittings as given in [67] is shown. We give the fit result of the NFW and Einasto dark matter radial profiles, (2.10) and (2.12), by normalizing the density value $0.4 \text{ GeV}/\text{cm}^3$ as the position of the Sun for $r = 8$ kpc:

$$\begin{aligned} \rho_{NFW}(r) &= \rho_{0N} \left(\frac{r}{r_s} \left(1 + \frac{r}{r_s} \right)^2 \right)^{-1}, \\ \rho_{Einasto}(r) &= \rho_{0E} \exp \left(-\frac{2}{\alpha} \left(\frac{r}{r_s} \right)^\alpha \right), \end{aligned} \quad (3.3)$$

where $r_s = 20$, $\rho_{0N} = 0.3136 \text{ GeV}/\text{cm}^3$, $\alpha = 0.17$, and $\rho_{0E} = 9.428 \times 10^3 \text{ GeV}/\text{cm}^3$.

Future observational data from, e.g, the Gaia Observatory³⁸⁾ which will produce a three dimensional map of

³⁶⁾A succinct summary of the result of [66] and a discussion of its significance can be found in a Science Comment from *PhysicsWorld* entitled “Dark matter seen in the Milky Way’s core”. See <http://physicsworld.com/cws/article/news/2015/feb/10/dark-matter-seen-in-the-milky-ways-core>

³⁷⁾Two short comments refuting the result of [66] can be found in the *High-Energy Physics Literature Database*: arXiv: 1503.07501 [astro-ph.GA] and arXiv: 1503.07813 [astro-ph.GA]. The response from the authors of [66] can be found in arXiv: 1503.08784 [astro-ph.GA].

³⁸⁾Gaia is a space observatory of the European Space Agency (ESA). Its mission is to construct a 3D catalog of around 1 billion astronomical objects, mainly stars, about 1% of the Milky Way population. Details can be found in Gaia’s website, http://www.esa.int/Our_Activities/Space_Science/Gaia_overview

stars in Milky Way and monitor and their motions in unprecedented detail, will help improve the accuracy in the extraction of the dark matter profile.

4 Dark matter candidates

As types of particles in the usual sense of the word, there is very little information on the dark matter particles individually. We do not know how massive they are, how they are produced, and what forces they will interact, besides gravity. But we do have a significant amount of knowledge about them in their collective effect through their actions in the universe. We know they are a major part of the matter component of the universe, where in the universe they are likely to reside, they are mostly cold kinematically, they have rather weak interactions with ordinary known particles, and they may interact among themselves. Let us first summarize the general properties of dark matter as one of more types elementary particle.

- They carry no electric or color charges. Their lacking of electric charge comes from the fact that there is no evidence that they involve in anything to do with the photon. They have no color charge because there is no evidence of the existence of exotic isotopes. Hence if dark matter is a distribution of congregates of elementary particles, they must be composed of neutral particles interacting very weakly with ordinary particles. Hence they must be particles beyond the standard model.
- They must be produced in the early universe thermally or non-thermally by decays of their parent particles before the radiation to matter dominance transition. They have to have the correct relic abundance.
- They have to be stable against the cosmic time scale, so that they are still around today. They can be of multi-components, made of more than one types of particles.
- Their clustering properties should allow explanations of both the observed large and small scale structures. Theoretically the characteristics of these structures can be studied in astrophysical N-body simulations. Experimentally, its overwhelming presence, i.e., nearly 27% of the whole universe today and nearly 85% of the matter world, should not allow it to conceal its tracks and hide its identity without an end.

Various candidates have been proposed, in particular, when dark matter came into the focus in astrophysics and particle physics starting the 1970s. Historically, many candidates have been offered as unknown astrophysical objects and hypothetical particles in theoretic

cal frameworks of models beyond the standard model of quarks and leptons: massive compact halo objects (MACHOs), primordial black holes, neutrinos, axions, weakly interacting massive supersymmetric particles (WIMPs), universal extra dimensional Kaluza Klein excitations, strongly self-interacting particles, and other exotic objects. In these broad categories, the dark matter can be baryonic or non-baryonic. MACHOs are the baryonic type, while neutrinos, WIMPs, and axions are non-baryonic types. The masses of these possible candidates expands a range over 80 orders of magnitude [25]. Some more discussion on mass bounds will be given in Section 4.4. Over the years some of the proposed candidates have been ruled out, but new possibilities are added to the candidate list.

According to their physical properties, there are different classification of dark matter. We describe them below.

4.1 Dark matter: Kinematic types and production mechanisms, a broad characterization

The kinematic properties of the dark matter affect the CMB anisotropy spectrum and play important roles in the structure formation of the universe. The dark matter can be divided into three kinematic categories [68], according to its velocity at the time of its decoupling which has a direct bearing on galaxy and cluster formation, and therefore small and large structures of the universe.³⁹⁾

- Hot dark matter (HDM). The HDM is composed of abundant light particles. An obvious candidate is the normal light neutrino. The mass of a HDM is the order eV or less, $m_{\text{HDM}} \lesssim 1 \text{ eV}$. It is relativistic at the time of the decoupling and remains relativistic when galaxy formation takes place at the cosmic photon temperature of tens K . Since highly relativistic particles generate a high pressure, they will smooth out small dense knots of matter. As a consequence, as being demonstrated in N-body simulations, HDM will hinder the formation of galaxies by destroying small scale structure of the universe, and hence conflicts with the observed scale structure of the universe. So HDM is disfavored. Hot dark matter, while it certainly exists in our universe in the form of neutrinos, can only be a small part of the story. according to current observational data, the contribution of the standard model neutrinos to the mass of the universe no more than 0.25%.
- Cold dark matter (CDM). The cold dark matter is at the other end of the dark matter mass-velocity

³⁹⁾The terms *hot*, *warm*, and *cold dark matter* was introduced in 1983 [69, 70]. See also <http://ned.ipac.caltech.edu/level5/Primack4/frames.html> As already noted earlier, a brief history of dark matter can be found in [27].

spectrum. It is non-relativistic at decoupling. Its mass can be in the GeV to TeV order or even larger. There are many CMD candidates, including weakly interacting massive particles like neutralinos, WIMPZILLAs, solitons, etc. Since non-relativistic particle exerts very little pressure and will diffuse a negligible distance in comparison with the size of the universe, a rich varieties of small scale structures of the universe is expected. Indeed, N-body simulations with non-interacting cold dark matter show that the domination of CDM can leads to the existence of many dwarf satellites in galaxies and galaxy clusters. It also predicts the formation of cusped distributions, i.e., a sharp increase in the dark matter concentration at the center of the galaxy, which is known as the *cuspy halo problem*. Although CMD is the favored candidate for dark matter, from existing observations these features of small scale structures, which are not in the universe's portfolio, are challenges that have to be met. Let us recall that these problems and the *too-big-to-fail* problem of CMD together with the possible resolution have been discussed in Section 2.7.

- Warm dark matter (WDM). The WDM is something in between the HDM and CDM, consisting of particles of mass of the KeV order or higher, $m_{\text{WDM}} \gtrsim 1$ keV which may interact even weaker than neutrino. It is relativistic at decoupling, but non-relativistic at the radiation-to-matter dominance transition. There are possible candidates for WDM, including sterile neutrino, light gravitinos and photino. Although WDM is expect to predict a more smooth dark matter profile, allows less dwarf satellites, and alleviates the cuspy halo problem, actual simulations indicate that WDM does not do much better than CDM [54, 55].

It is clear that the hot dark matter has been ruled out as a major part of the dark matter. The warm dark matter has been vigorously pursued. There is a series of ongoing conferences, starting from 2010, dedicated to the study of warm dark matter [71]. But the cold dark matter scenario is robust and still preferred. Observations have found evidence of more dwarf satellite galaxies in the Milky Way local group and there are ways to alleviate the cuspy halo and too-big-to-fail problems in more sophistic N-body simulations with interacting cold dark matter particles. As a historical record, a concise summary of the status of CDM vs. WDM as of 2011 is given in [62].

The dark matter can also be classified according to its production mechanism as relics of the early universe, thermal vs. non-thermal. For a general discussion see [72]. A recent review of the thermal and non-thermal productions of dark matter, with an emphasis on the

latter, can be found in [73].

- *Thermal relics*. These are particles in thermal equilibrium in the early universe, due to scattering, creation, and annihilation processes involving the particle under consideration. They have the equilibrium thermal distribution, i.e, the number densities is proportional T^3 , $n \sim T^3$.⁴⁰⁾ As the universe expands, the cosmic temperature drops and the density of the particle decreases so does the annihilation rate of the particle. When the particle annihilation rate drops below the expansion rate of the universe, the particle annihilation becomes ineffective and the particle freezes in to decouple from the cosmic heat bath, and the particle number density tracks the Hubble expansion, i.e., the density is proportional to $\sim a^{-3} \sim T^3$. The resultant freeze out particle forms the dark matter. The decoupling can happen either when the particle is relativistic or non-relativistic.

An example of the relativistic decoupling is the freeze out of the neutrino which decouples at around the cosmic temperature of $T \sim 1$ MeV, much larger than the mass of the neutrino. As for the abundance of the neutrino is concerned, it's number density behaves as T^3 before or after the freeze out. So it is as if the neutrino is always in equilibrium. But the freeze out neutrino has a lower temperature than the cosmic heat bath, because the neutrino dumps its entropy into the cosmic heat bath, the photon temperature, which defined the cosmic temperature, is increased relative to the neutrino. A more systematic discussion can be found in Section 9.4 and Section 9.5.

The case of non-relativistic freeze out produces the cold dark matter. Prior to freeze out, being non-relativistic, the particle density function has an exponentially suppressed density distribution, i.e., $n \sim \exp(-m_{\text{DM}}/T)$, where m_{DM} is the mass of the particle. After the freeze out, similar to the relativistic case, the density function follows the Hubble flow and it reverses back to the T^3 behavior. The final dark matter density distribution depends on the particle annihilation cross section, the smaller the annihilation rate the larger the density function. Decoupling is an important mechanism for the survival of the relic of a massive particle.⁴¹⁾ If the particle were allowed to maintain equilibrium to maintain the exponentially suppressed density distribution form, it would eventually be depleted entirely. We will come back to this topic later in more details

⁴⁰⁾A discussion of thermal distribution functions is given in Section 9.3.

⁴¹⁾In the discussion of decoupling, Section 5.2 of [74] is entitled *Freeze Out: Origin of Species*, which we borrow as part of the title of the present subsection.

in Section 5.3 and 11.

This simple and elegant mechanism of thermal production and freeze out give rise to the *WIMP miracle*. It is widely accepted and forms the theoretical basis of most dark matter search experiments. The WIMP miracle depends on two assumptions, one in cosmology and the other in particle physics. In cosmological, it assumes, as in the standard cosmology, that the universe is radiation dominated prior to the freeze out, and in the particle physics, it requires that the dark matter annihilation into standard particles has the order or a weak interaction cross section. More discussions will be given later in Section 5.3 and 11.

According to unitarity, the mass of a thermal relic is bound above by 340 TeV. For some more details see Section 4.4.

- *Non-thermal relics.* These are dark matter particles that have a non-thermal history. They are produced non-thermally and have never been in equilibrium with other particles in the universe so that their energy distributions are different from a normal thermal distribution.

Seeking for a non-thermal production of dark matter is motivated by both theoretical and experimental considerations. The two assumptions needed for the WIMP miracle are not done deals. In the cosmological side, while radiation dominance is established in the period of the BBN to the matter-radiation equality, i.e., in the cosmic temperature range of 1 MeV to 1eV, radiation dominance not been probed. In the particle physics side, large regions of the parameter space for the WIMP have already been ruled out by various direct and indirect WIMP searches. So new possibilities should be considered.

There are a number of possible candidates for non-thermal relics, including axions emitted by cosmic strings; very heavy aggregates of particles by the name of WIMPZILLAs which are superheavy, of the order of 10^{12} to 10^{16} GeV. There also a class of models consist of scalar fields which are called moduli which can decay into, e.g., winos which are the dark matter. A systematical discussion of a large class of non-thermal dark matter can be found in a recent article [75] To illustrate the motivation and the physical reason for such model, we parallel phrase a paragraph of the Introduction section of the paper:

A well-motivated alternative to the standard thermal cosmological history mentioned above is that of a non-thermal cosmological history, in which BBN is preceded by a phase of pressureless matter domination. Such a situation is pre-

dicted in many top-down theories for new physics, e.g., low-energy limits of supergravity and string/M-theory compactifications. These theories, under some very mild assumptions, contain gravitationally coupled scalars called moduli. When the Hubble parameter drops below moduli masses, moduli begin coherent oscillations and behave as pressure-less matter, dominating the energy density of the universe until the longest-lived one (ϕ) decays to reheat the universe. In these cosmological histories, an electroweak-scale Wino provides a natural candidate for supersymmetric (SUSY) DM, provided that the modulus dominated phase ends at temperatures below a GeV or so [1, 2].

Since they are produced non-thermally, they do not obey the unitarity limit mentioned above.

4.2 Dark matter: Particle types

A more detailed category of the dark matter is by individual particle types, appearing in well-motivated theoretical ideas. Most dark matter candidate particles are also known for their own reasons, as they have been proposed as solutions of some existing problems other than the dark matter. There is also the possibility that dark matter is something not yet on the radar screen of our present theory, not yet being discussed by anyone. The known particle type include prominently the neutrino.

To be a candidate of the dark matter, a particle has to satisfy some generic constraints:

- It has to be stable on the cosmic time scale so that it is still around today.
- It has to have no strong or electromagnetic interactions.
- Taken together, one of more candidates, they have to have the required relic density.
- Based on what are known to us, the baryonic type cannot be a very significant part of the mix of the dark matter.

Well-qualified candidates of non-baryonic dark matter are all from theories beyond the standard models. They include a wide class of *weakly interacting massive particles* (WIMPs)⁴²⁾, which include neutrinos, neutralinos,

⁴²⁾The acronym *WIMP* has entered the domain of common vocabulary and can be found in the electronic dictionary/Thesaurus *The Free Dictionary by Farlex*. See also the NASA web site entitled *Imagine the Universe* http://imagine.gsfc.nasa.gov/docs/teachers/galaxies/imagine/dark_matter.html which is intended for laymen and school teachers.

the axion, etc. Another class of particle are the superWIMPs (*Superweakly interacting massive particles*) which have annihilation cross sections much smaller than that of the weak interaction reactions. The particles includes sterile (right-handed) neutrinos, gravitino, extra dimensions, etc. Below we briefly describe particle candidates of the dark matter, together with comments on them.

- Neutralinos ($\tilde{\chi}$). For R-parity conserving SUSY, in the minimal supersymmetric standard model (MSSM), neutral superpartners consist of four neutral fields: the gaugino Z , photino $\tilde{\gamma}$, and Higgsinos \tilde{H}_1^0 and \tilde{H}_2^0 . Their mixtures form 4 Majorana fermionic mass eigenstates. The neutralino refers to the lightest of the four mass eigenstates. And in most SUSY models, it is the lightest superparticle, which is stable due to R-parity conservation. The desirable parameter of the neutralino are: mass of the order of $m_{\tilde{\chi}} \sim M_{\text{SUSY}} \sim 0.1 - 1$ TeV and sub-weak interaction strength $\sim 10^{-4} \sigma_{\text{weak}}$.
- Ordinary neutrinos (ν). Neutrinos, including those in the standard model, have been proposed early on as a possible candidate of dark matter. Since the standard model neutrinos are relativistic objects, they are the hot type, hence disfavored from the large scale structure formation and the analysis of CMB anisotropy. Furthermore, the present oscillation experiments and astrophysical observation show that conventional neutrinos have very small masses and, with their known number density, can contribute no more than a small percent to the total matter of the universe. The present observation gives the fraction of light neutrinos to be $\Omega_\nu < 0.0055$ [13].
- Sneutrinos ($\tilde{\nu}$). Sneutrinos have large scattering and annihilation cross sections and therefore should be readily produced in a hadron collider if the available energy is over their production threshold. However, the search for them at the Tevatron has been negative and this implies that they are heavy, likely to be hundreds of GeV. This high mass makes it awkward to be the lightest sparticles. There is yet no news of sparticles sighting from LHC either.
- Heavy neutrinos (N). Any fourth generation $SU(2)$ neutrinos will have a mass greater than $m_Z/2$ as required by LEP data. Such a heavy neutrino will not be stable if it has any mixing with the light leptons. The decay of the heavy neutrino can be highly suppressed if it mix only with a sterile neutrino. But such models are rather contrived.
- Sterile neutrinos (ν_R). In the construction of mass terms for neutrinos, right-handed neutrino fields can be introduced to extend the standard models. Since

the right-handed neutrinos are in the singlet representations of all the standard model symmetry groups, their standard model quantum numbers are all vanishing. So they do not directly interact with standard model particles, and hence they are sterile. Right-handed neutrinos can be either the Dirac or Majorana type. To be Majorana, because they have null quantum numbers so there is nothing to prevent them to be their own antiparticles. Masses of Majorana particles are obtained from the see-saw mechanism which gives naturally very small mass to the left-handed neutrino and much larger mass to the right-handed ones. Another consequence of this model is the presence of more Higgs fields than that of the minimal standard model. Sterile neutrinos are generally considered in the class of WDM.

- Axion. The axion is a pseu-Nambu-Goldstone boson, a hypothetic particle arising from the breaking of the Peccei-Quinn $U(1)$ symmetry. It was proposed in 1977 as a solution to the strong CP problem. In the early universe axions can be produced non-thermally from QCD phase transition. A boson condensate formed by axions which works naturally as a cold dark matter. For a review we refer to [76] and [77].
- Axino or saxion (\tilde{a}) and gravitino (\tilde{G}). They can be warm (\sim KeV) or cold. They are interesting and viable candidates of the dark matter. They are not directly testable, but LHC may be able to provide some hint of their existence.
- Universal extra dimensions (UED). This is associated with theories beyond the standard model by the extension of the space to a higher dimensionality. This class of models, which can happen in string theory and theory of extra dimension, provide another interesting possibility for the dark matter. They are generally known as the Kaluza-Klein states (KKS). The lightest of such states, referred to as *LKP*, may be stable to be a candidate of the dark matter. Mass of LKP are in the range of 400 GeV – 1.2 TeV, comparable to the neutralino, but well above the current experimental constraint. They are likely to be testable. There are many works on Kaluza-Klein dark matter, a review of LKP as a dark matter candidate can be found, e.g., in [78].
- More exotic particles. There are other candidates which happen in beyond the standard model theories such as the little Higgs and Technicolor.
- WIMPZILLA (supermassive dark matter). The dark matter candidates discussed above are thermal relics of the early universe, except the axion. There is an upper limit imposed by unitarity on how massive a thermal relic dark matter particle can be [79]. The

limit is 340 TeV. However, in addition to the thermal produced dark matter, there is also the possibility of non-thermal dark matter which is composed of supermassive states produced non-thermally in the early universe. These are called WIMPZILLA [80, 81]. They have masses which are many orders of magnitude higher than the scale of the standard model, in the range of 10^{12} to 10^{16} GeV. There are several production mechanisms for WIMPZILLA. They can be accreted and accumulated in the center of the sun. Their signal is the very high energy neutrinos produced by their annihilations. But their required cross sections seem to have been ruled out by limit imposed by direct searches experiments.

- Solitons. This includes Q-balls and F-balls, both non-topological solitons. They are considered to be exotic dark matter candidates. A Q-ball can arise in theories with scalar field carrying a conserving $U(1)$ charge. The Q-ball is the ground state of the theory in the sector of fixed charge. Loosely speaking it is a finite sized stable blob which contains a large number of particles due to the attractive force among them. The proposal of the Q-ball as a component of the dark matter can be found in [82].

The F-ball arises from the breakdown of an approximate Z_2 symmetry. The simplest type is a bubble of false vacuum surround by a domain wall with many zero-mode fermions attached. The earliest suggestion of the F-ball to be a dark matter can be found in [83].

In addition to individual particles, large astronomical objects have been proposed as dark matter candidates during early dark matter searches. They are baryonic in nature and most are no longer viable because of the bound on baryonic matter by BBN and CMB anisotropy. But they should be and have been searched for to assure that there is no surprises. For completeness we mentioned them below:

- Massive compact halo objects (MACHOs). This is a possibility that dark matter involves normal baryonic matter include very dim stars, brown dwarfs which have small masses less than 10% of that of the sun; or perhaps small, dense chunks of heavy elements, which are known collectively as “MACHOs”. Gravitational lensing is a way to search for them. However, studies of big bang nucleosynthesis have shown convincingly that baryonic matter such as MACHOs cannot be more than a small fraction of the total dark matter.
- Massive compact objects (MCOs). Similar to MACHOs, these are clumps of dense ordinary matter which do not emit a measurable amount of radiation.

Such objects, if distributed in the universe, can very likely be revealed by lensing effect in the study of supernovae. The light from many supernovae studied have traveled 5 billion years to reach Earth. Encountering a MCO, the light will be dispersed. Observations of 300 distant supernovae show no sign of MCOs with a mass greater than 1% of the mass of the sun [84]. It is hence concluded in [84] that MCOs of mass no smaller than 10% of that of Earth can be ruled out. So MCOs cannot be a significant part of the dark matter.

- Black holes and primordial black holes. Galaxy-sized, very massive black holes as dark matter are ruled out by gravitation lensing data. But tiny black holes are still a possibility [85]. In general, however, black holes, irrespective of their sizes, cannot constitute a significant part of the dark matter, unless they are formed in the very early universe before the era of nucleosynthesis. The reason is similar to that against MACHOs as given above. If they are created after the era of nucleosynthesis, the black holes have to be included in the baryon density, while created before they will be counted in the density of the dark matter.

In the early universe very small black holes less than the solar mass, called the primordial black hole (PBH) can be produced. They would gravitate, float through the cosmos and form clusters. PBH are non-relativistic and collisionless, so they would be an interesting candidate of the CMD. A small PBH could have the size of an atom but the mass of an asteroid which is the order of 10^{17} Kg or more. When a PBH pass through a star, such as the sun, its gravitational effect can set the star to oscillate and therefore observable [86].

4.2.1 *Broaden the scope: Dark sectors – a brief description*

Adding to the list of dark matter candidate is the relatively new development call Dark Sectors. Since all evidence for dark matter comes from its gravitational effect, so the dark matter particle or particles may come from a hidden sector that does not interact with the known standard model forces. The dark matter particles have interactions among themselves. Different kinds of models can be constructed. A simple example is that the dark sector is a pure supersymmetric $SU(N)$ gauge theory. In such model the WIMPless miracle can happen so that the result of the WIMP miracle, which be discussed in some detail later, will be obtained. In addition, the dark sector particles can have strong self interactions which is suggested by recent cosmological observation in galaxy collisions. We refer to [87, 88].

Another focus of the dark sector refers generally to physics sectors that contains dark matter particles but not the WIMP type [89]. They include the well-motivated axion and axion like particles (ALPs), and broad hidden sectors which are not related to standard model physics but motivated by the fact that the dark components, dark energy and dark matter are the major constituents of the universe and therefore likely (why not) to have their own full structures. This opens up a completely new world and can have myriad possibilities, including new particles and interactions. But they are assumed to have similar theoretical framework of ordinary particles, such as the gauge principle and field-theoretical structure, which will serve as portals to this new world. The following are some of the possibilities. For more details and references, see [89].

- Dark photon. This is to assume the existence of a new $U(1)$ gauge bosonic field, denoted as A' , which couples very weakly to (electrically) charged particles through kinetic mixing with the normal photon, producing an effective interaction with the normal sector, e.g., $\sim A'_\mu J_{EM}^\mu$, where J_{EM}^μ is the electromagnetic current of charged normal particles.
- New, light, weakly-coupled particles that have no conventional strong or electroweak interactions. This is motivated by the null result of the various searches, including dark matter and accelerator, of the popular WIMP scenario which contains weak-scale (hundreds of GeV) weakly interacting particles WIMPs. In addition, there are also hints of possible problems with the cold dark matter model Λ CDM in astrophysics and cosmology considerations. Lighter dark matter of the MeV-GeV (Sub-GeV) range, made of scalar particles, is allowed theoretically. This light particle sector can be quite complicated involving many different species of particles and they can also be thermal relics.

4.3 Remarks on viable dark matter candidates

Dark matter candidates, such as the neutralino in supersymmetry, the lightest Kaluza–Klein particle in the superstring theory and theories of extra dimensions, and other particles from beyond the standard model are generally quite massive and weakly interacting with particles of visible sector. So they are referred to as *weakly interacting massive particles* (WIMPs). They are well-motivated by independent considerations of particle physics [90]. Since mostly of them are imbedded in a well-defined theoretical framework, for solving particle physics problems, such as the neutralino, their effects can be worked out theoretically and their laboratory tests can be systematically carried out. Their astrophysical behavior is roughly the following: In the every early uni-

verse they are in thermal and chemical equilibrium with the ordinary particles until the cosmic temperature drops below the mass of the WIMP, they then decouple and their density freezes out. It turns out that WIMPs have naturally the right density required for the cold dark matter. This we will see in some detail later. Hence this is the class of dark matter candidates which have received the most attention and been most thoroughly studied. The mass of a WIMP is estimated to lie in a wide range of 10 GeV to 1 TeV.

Not all particle physics motivated dark matter candidates are WIMPs. The axion is an example of non-WIMP particle dark matter candidate. WIMP together with axion are generally considered as mainstream candidates of the cold dark matter. Others types of particles are non-WIMPs, which can be cold or warm and can be thermal, such as WIMPZILLA, gravitinos, superWIMP, etc. are referred to as exotics. There are also candidates produced non-thermally. Because the attention has mostly been the search of the dark matter, effects of the dynamics among the dark matter particles themselves are largely neglected.

Recent developments further enlarge the possibility of dark matter candidates, motivated by both experimental and theoretical reasons. On the experimental side is the failure of uncovering any signals of dark matter from the myriad experimental searches using diverse technologies, resulting in ruling out significant parts of the WIMP parameter space. Astrophysical observations have suggested the possibility of significant interactions among dark matter particles themselves. On the theoretical side is the opening up of new possibilities in producing dark matter relics. In addition to WIMP, the early SIMPs (Strongly Interacting Massive Particles) system [91] and SIDM [56], meaning *Self-Interacting Dark Matter*, are more actively studied. We only cite the most recent articles of the two schemes and more references can be found in them. Recently a new mechanism have been proposed to produced thermalized dark matter that takes place in the hidden sector of higher dimensional theories. The hidden sector particles involved, which have strongly self-interaction, is called *SIMP* (strongly interacting massive particles) [92] and models realizing the scheme have been proposed, see, e.g., [93].⁴³⁾

Following [94] we summarize in Table 4.1 the most attractive dark matter candidates, both WIMP and non-WIMP. Several other relevant informational items are also given: the motivation of their existence, some of their general properties, and the methods of their detection.

⁴³⁾Note the difference between SIMPs and SIMP. However, it should be noted that in some SIMPs articles the name SIMP has been used.

Table 4.1 Summary of dark matter candidates and their related physics. The table is adopted from [94]. GHP stands for *gauge hierarchy problem* and NFPF for *new physics flavor problem*. $\checkmark\checkmark$ means *generic detection signals*, and \checkmark means *signal possible*.

	WIMPs	Super WIMPs	Light \hat{G}	Hidden DM	Sterile ν	Axions
Motivation	GHP	GHP	GHP NFPF	GHP NFPF	ν Mass	Strong CP
Naturally Correct Ω	Yes	Yes	No	Possible	No	No
Production Mechanism	Freeze Out	Decay	Thermal	Various	Various	Various
Mass Range	GeV–TeV	GeV–TeV	eV–keV	eV–keV	keV	μeV –meV
Temperature	Cold	Cold/Warm	Cold/Warm	Cold/Warm	Cold	Warm
Collisional				\checkmark		
Early Universe		$\checkmark\checkmark$		\checkmark		
Direct Detection	$\checkmark\checkmark$			\checkmark		$\checkmark\checkmark$
Indirect Detection	$\checkmark\checkmark$	\checkmark		\checkmark	$\checkmark\checkmark$	
Particle Colliders	$\checkmark\checkmark$	$\checkmark\checkmark$	$\checkmark\checkmark$	\checkmark		

4.4 Ranges and bounds of masses of dark matter candidates

In spite of the clear conditions on what the dark matter cannot be, the number of allowed candidates is huge with limited restrictions on their properties, such as their mass and reaction cross sections. A summary of the mass ranges of the various dark matter candidate particles vs their interaction cross sections is given in Fig. 4.1 which is an updated version given originally in [95].

Taking the plot shown in the lower panel of Fig. 4.1, we see the mass varies about 30 orders of magnitude, from 10^{-15} to 10^{15} GeV. The reaction cross sections of the majority of candidate particles cover a comparable range of 32 orders of magnitude, from 10^{-31} to 10 pb. Note that the WIMP includes several different kinds of particles: Kaluza-Klein particles, little Higgs, as well as SUSY particles. The SUSY Neutralinos which are included in WIMP have a more restrictive mass and cross section range than general WIMP particle.

Interestingly, there exist Upper and lower bounds on the mass of certain categories of dark matter particle, based on general argument, largely model independent. The upper bound is the unitarity bound which is applicable to the thermal relics of WIMPs. The lower bound is known as the Tremaine–Gunn bound which applies to particles such as axions. We briefly discuss them below. We should emphasize that the unitarity bound is not applicable to non-thermal dark matter particles.

Unitarity bound

For the thermal dark matter particle, the unitarity relation places an upper bound on its mass bound is given in [79]. It is derived for particles which are once in thermal equilibrium during their time evolution in the early universe. The partial-wave unitarity of the S-matrix

limits the annihilation cross section, which bounds the relic abundance of the dark matter particle and hence its mass. Two bounds can be obtained. It excludes (a) stable elementary, point particles of mass greater than M_{max} and (b) composite extended objects of the radii smaller than r_{min} . [79] gives M_{max} and $r_{\text{min}} = 7.5 \times 10^{-7}$ fm. In particular, the mass upper bound of a thermalized dark matter is given by⁴⁴⁾

$$m_{\text{TDM}}(\text{in TeV}) \leq 10^3 \cdot \sqrt{\frac{\Omega_{\text{TDM}} h^2}{1.7 \sqrt{X_f}}},$$

$$X_f = \frac{m_{\text{TDM}}}{T_f}, \quad (4.1)$$

where T_f is the freezing out temperature of the dark matter and it is estimated $X \approx 28$. In the case that the dark matter saturates the mass density and the reduce Hubble constant is unity, i.e., $\Omega_{\text{TDM}} h^2 = 1$, we obtain the bound of [79]. If we take the most recent observational value $\Omega_{\text{cdm}} h^2 = 0.12$ [13], we have $m_{\text{TDM}} \leq 115$ TeV. A discussion on the maximum masses of possible dark matter particle can be found in the 2014 SLAC Summer Institute [26] where a much smaller bound of 30 TeV was stated.⁴⁵⁾ Although the upper bound is rather large, in the tens or hundred TeV value which makes it not very restrictive, it is interesting that such a bound exists. In case of multi-component cold dark matter which consists of more than one species of massive particles, Eq. (4.1) is satisfied by each individual species, then each Ωh^2 is reduced and the corresponding mass bound decreases too. Since the mass bound goes like $\sqrt{\Omega h^2}$, the decrease in the individual mass bound is mild.

⁴⁴⁾Equation (12), [79]. The mass bound given is for a Majorana fermion. See [79] for details.

⁴⁵⁾It is a Project Presentation entitled *Maximum mass of DM particles* given in the afternoon of August 14, 2014.

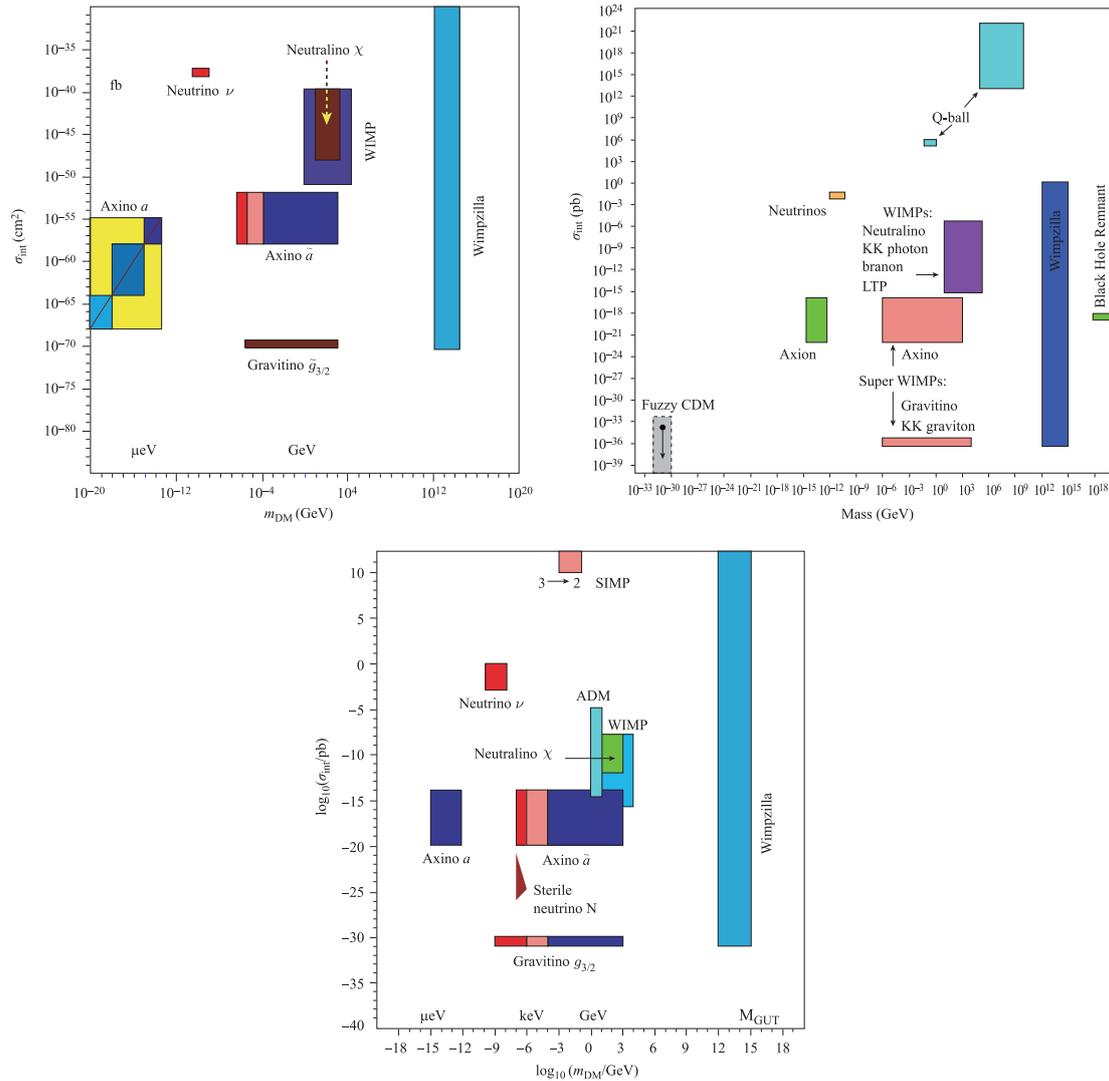


Fig. 4.1 Mass vs. interaction cross section of dark matter candidates. The upper left panel is an updated version of a similar plot given in [95]. The upper right panel is given by Park quoted in [96]. The lower panel is a further updated as given in [73] for several well-motivated dark matter candidates. We note that the time span of the three plots is over a decade and the established parameter space have not changed much. But new candidates have been added. In the bottom panel SIMP stands for Strongly Interacting Massive Particles (Section 4.3) and ADM for Asymmetric Dark Matter (Section 8.3.3). Fuzzy CDM is a very light hypothetical scalar particle of 10^{-22} eV, proposed to solve the cusp halo problem.

We should note that the unitarity bound does not apply to dark matter particles which are non-thermal.

Tremaine–Gunn bound

The Tremaine–Gunn bound [97] refers to the lower mass bound of dark matter particles. It is first derived for fermionic dark matter particles, based on the phase-space density evolution argument which, in the simple robust form, is a consideration of the average phase-space density of the dark matter particle in a given astronomical object in relation to the phase-space density of degenerate Fermi gas. The argument has been generalized to bosons [98]. The Tremaine–Gunn bound has been

used widely to put a lower limit of sterile neutrinos as dark matter particles. More discussion can be found in Section 7.2.

5 WIMPs

5.1 Weakly interacting massive particles (WIMPs)

On broad theoretical considerations of particle physics, WIMPs are the most desirable candidates of the dark matter. First, the comparable densities of the baryon and dark components, $\Omega_{DM}/\Omega_B \approx 5.4$ [13], i.e., densities

of the dark matter and ordinary matter being comparable, suggests that the baryonic matter and the dark matter are related. Second, as WIMPs are mostly particles which appear in theories beyond the SM, establishing such a relation would provide solutions to some of the outstanding fundamental problems of particle physics today. With the dominant component of DM being WIMPs, one can establish a deeper connection between microscopic fundamental particles and the large scale universe.⁴⁶⁾

5.2 Supersymmetric particles as WIMPs

Supersymmetry, which is motivated by the gauge hierarchy problem, the unification of coupling constants, and string theory, provides a concrete framework to extend the SM through an encompassing symmetry principle. The minimal supersymmetric standard model (MSSM), which, with the discrete symmetry of the R-parity conservation, has all the desirable features to give us a well-motivated WIMP candidate in the form of the lightest stable super particle such as the neutralino. The R-parity of a particle is defined by its baryon number B , lepton number L , and spin s ,

$$R = (-1)^{3(B-L)+2s}. \quad (5.1)$$

It is straightforward to see that all SM particles have positive R-parity and their super partners negative R-parity. If the R-parity is conserved the lightest super particle (LSP) will be absolutely stable. The LSP, called the neutralino and denoted by $\tilde{\chi}^0$, is a mixture of the super partners of the photon, Z-boson (mixture of \tilde{W}^0 and \tilde{B}^0), and super partners of the Higgs particle (\tilde{H}_1^0 and \tilde{H}_2^0),

$$\tilde{\chi}^0 = a_1 \tilde{W}^0 + a_2 \tilde{B}^0 + a_3 \tilde{H}_1^0 + a_4 \tilde{H}_2^0. \quad (5.2)$$

A very attractive feature of the MSSM is the gauge unification which unifies the three coupling strengths of the SM, for the strong and electroweak interactions, at high energies and, therefore, has the possibility of achieving grand unification. Figure 5.1, a frequently shown graph, describes the coupling unification of the MSSM at high energies. The left panel shows that in the absence of supersymmetry the three running coupling constants of the SM do not intersect at the same energy, while the right panel shows that in MSSM the three running couplings do intersect at around 10^{16} GeV, a three orders of magnitudes below the Planck energy.

⁴⁶⁾We have tried to emphasize the connection between particle physics and cosmology. This connection has been well recognized by now. For instance, note the Ph.D. thesis by M. E. C. Swanson by the title: *Particle Physics in the Sky and Astrophysics Underground: Connection the Universe's Largest and Smallest Scales*, arXiv: 0808.0002 [astro-ph].

Aesthetically satisfying, WIMPs as the dark matter provide a common solution to problems arising from the high energy and low energy connections facing particle physics and cosmology. In particle physics, supersymmetry shields the low energy region, i.e., the standard model of a few hundreds of GeV, from the strong influence of the higher energy scale of grand unification. And thus resolves the hierarchy problem. For cosmology, supersymmetry becomes the dark side of the universe. The super partners together with the ordinary particles form the universe as we know it.

Here a brief general comment on SUSY is in order. SUSY contains a large class of specific models or scenarios, all involve a large number of free parameter. Even the MSSM has a large parameter space. The number of parameters can be greatly reduced in some of the specific model subject to particular conditions. The most popular and widely studied are the minimal supergravity (mSUGRA) model and the constrained MSSM (CMSSM). We refer to [100] for a detailed discussion of dark matter candidates in these two models.

5.3 WIMP physics and cosmology

Following the conventional practice, let us denote generically the dark matter particle by χ and its antiparticle $\bar{\chi}$. χ and $\bar{\chi}$ are the same if the dark matter consists of Majorana particles. In discussions of the general nature, χ denotes both dark matter particle and its anti-particle. Participating in the evolution of the universe, χ is assumed to be thermally produced and therefore, at high temperatures $T \gg m_\chi$, in thermal equilibrium with the thermal bath of the universe.

In the early universe χ maintains its equilibrium with ordinary particles through high rates of annihilation and production processes, which take the form

$$\chi + \bar{\chi} \leftrightarrow \varphi + \bar{\varphi}, \quad (5.3)$$

together with the cross channel elastic scattering processes, where φ and $\bar{\varphi}$ denote an ordinary, standard model, particle and its antiparticle, such as leptons, quarks, the photon, etc. All particles in equilibrium are at the same temperature. The number density of χ in equilibrium at a given temperature is given by

$$n_\chi^{(\text{eq})} = \frac{g}{(2\pi)^3} \int f(p) d^3p, \quad (5.4)$$

where g is the number of internal degrees of freedom, i.e., the spin states $2S+1$, of χ . The function $f(p)$ is the Dirac-Fermi or Bose-Einstein distribution,⁴⁷⁾ depending on the statistical property of χ , as a function of the magnitude of the three-momentum p of χ . For massive χ , at

⁴⁷⁾See a discussion in Section 9, in particular, related to Eq. (9.98).

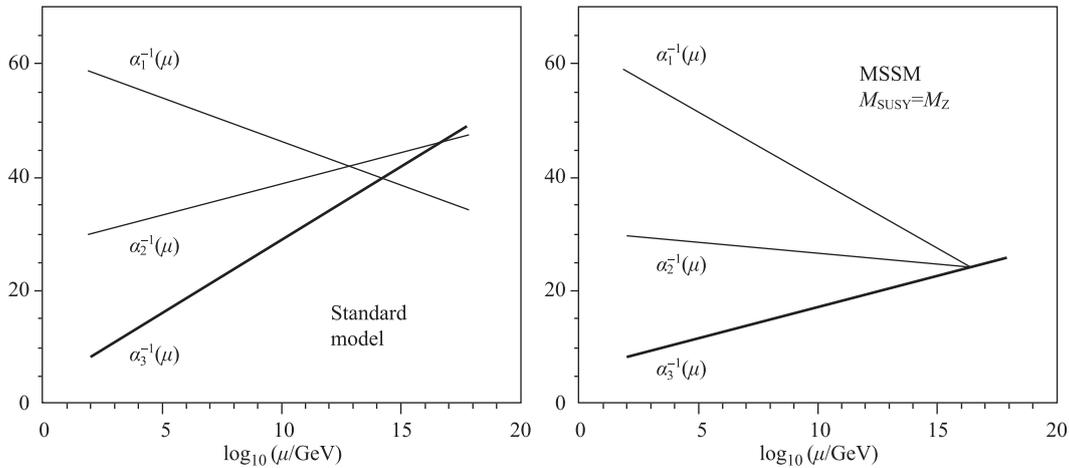


Fig. 5.1 High energy behavior of the SU(3), SU(2), and U(1) couplings constant. The left panel is for the standard model in which the couplings are not unified, while, as shown in the right panel, in the minimal supersymmetric model the couplings are unified around 10^{16} GeV. These figures are from Ref. [99] but rearranged.

high temperatures, $T \gg m_\chi$, we have $n_\chi^{(\text{eq})} \propto T^3$. At lower temperature $T \ll m_\chi$, because of the suppression of the production of $\chi \bar{\chi}$ pair from $\varphi \bar{\varphi}$ pair, while the annihilation rate maintains, the number density is exponentially suppressed $n_\chi^{(\text{eq})} \propto \exp(-m_\chi/T)$. When the temperature of the universe drops sufficiently low, the density of χ would be exponentially small if the equilibrium maintains. Fortunately, the cosmological dynamics intervenes. When the universe expands and cools down, the reduced number density of χ makes it harder and harder for χ and $\bar{\chi}$ to find each other to annihilate, and to find ordinary particles to scatter around. Eventually, when the temperature of the universe becomes sufficiently low, the annihilation process stops, χ decouples from the cosmic thermal bath consisting of photons and other relativistic particles. Then, except for the possibly very occasional annihilation into and scattering off ordinary particles, χ is decoupled from the rest of the universe. But it continues to expand freely with the Hubble flow so as to maintain a constant number in a comoving volume a^3 , i.e., $a^3 n_\chi = \text{constant}$. The number density of χ resumes the T^3 behavior. Below we present a qualitative argument for the physics of thermal decoupling or freeze-out of particles, and hence the appearance of thermal relics. A detailed treatment can be found in Section 11.

5.3.1 Thermal relics

In this subsection we outline the formulation in dealing with thermal relics for both relativistic and non-relativistic cases. In the relativistic case the relics result in a hot dark matter and in the non-relativistic case a cold dark matter. We assume the symmetric case that the dark matter particle and its anti-particle have identi-

cal density functions $n_{\bar{\chi}} = n_\chi$. We start with the Boltzmann transport equation (BTE) which deals with the time evolution of the number density of the particle χ , n_χ , to be derived in Section 11,

$$\frac{dn_\chi}{dt} = -3Hn_\chi - \langle v\sigma_{\text{ann}} \rangle (n_\chi^2 - n_\chi^{(\text{eq})2}), \quad (5.5)$$

where $n_\chi^{(\text{eq})}$ is given in Eq. (5.4), H the Hubble expansion rate, σ_{ann} the $\chi\bar{\chi}$ annihilation cross section of Eq. (5.3) summed over all relevant final states, and v the magnitude of the relative velocity of χ and $\bar{\chi}$. $\langle v\sigma_{\text{ann}} \rangle$ is the thermal average of $v\sigma_{\text{ann}}$. It is straightforward to check that both sides of the above equation has the expected engineering dimension of inverse volume divided by time, i.e., $\text{cm}^{-3}\cdot\text{s}^{-1}$. For the later usage we define two rates of annihilation of χ and $\bar{\chi}$ ⁴⁸⁾.

$$\begin{aligned} \Gamma_{\text{ann}} &= n_\chi \langle v\sigma_{\text{ann}} \rangle, \\ \Gamma_{\text{ann}}^{(\text{eq})} &= n_\chi^{(\text{eq})} \langle v\sigma_{\text{ann}} \rangle. \end{aligned} \quad (5.6)$$

The time rate of change of the total number of particles in a comoving volume a^3 , $a^3 n_\chi$ has a simpler form,

$$\frac{d(a^3 n_\chi)}{dt} = -\langle v\sigma_{\text{ann}} \rangle a^3 (n_\chi^2 - n_\chi^{(\text{eq})2}), \quad (5.7)$$

which can be rewritten as

$$\frac{1}{2} \frac{d(a^3 n_\chi)^2}{dt} = -\Gamma_{\text{ann}} \left((a^3 n_\chi)^2 - (a^3 n_\chi^{(\text{eq})})^2 \right). \quad (5.8)$$

⁴⁸⁾Let us recall that the expressions of the equilibrium forms of the number density in the relativistic and non-relativistic limit are

$$\begin{aligned} n_{\text{rel}}^{(\text{eq})} &= \frac{3}{4} g \frac{\zeta(3)}{\pi^2} T^3, \quad n_{\text{rel}}^{(\text{eq})} = g \frac{\zeta(3)}{\pi^2} T^3, \\ n_{\text{non}}^{(\text{eq})} &= g_\chi m_\chi \left(2\pi \frac{m_\chi}{T} \right)^{-3/2} \exp\left(-\frac{m_\chi}{T}\right). \end{aligned}$$

For details see, Section 9.3.

Since the comoving volume a^3 is not a measurable quantity and the cosmic time is awkward to work with in the present setting. We replace a^3 by the entropy density s by defining

$$Y \equiv \frac{n_\chi}{s}, \quad Y_{\text{eq}} \equiv \frac{n_\chi^{(\text{eq})}}{s}. \quad (5.9)$$

The BTE Eq. (5.7) becomes

$$\frac{dY}{dt} = -\langle v_\chi \sigma_{\text{ann}} \rangle s (Y^2 - Y_{\text{eq}}^2). \quad (5.10)$$

The more convenient variable is

$$x \equiv \frac{m_\chi}{T}. \quad (5.11)$$

Because of the entropy conservation in a comoving volume, $a^3 s \sim \text{const.}$ Eq. (9.115), replacing a^3 by s^{-1} is natural. The variable x can be used to differentiate directly a relativistic case from a non-relativistic one. Crudely, $x = 2$ can be used as a criterion for the two cases, $x < 2$ the relativistic case and $x > 2$ the non-relativistic case.

To obtain the BTE in terms of x , we need the time derivative of the cosmic temperature

$$\frac{d}{dt} = -x \left(\frac{\dot{T}}{T} \right) \frac{d}{dx}, \quad (5.12)$$

where the conventional notation $\dot{T} \equiv dT/dt$ is used. To calculate \dot{T}/T we use the explicit expression of s given in Eq. (9.137), i.e., $s = (2\pi^2/45)g_{s*}T^3$, and $a^3 s \sim \text{const.}$ which gives $\dot{s} = -3Hs$,

$$\frac{\dot{T}}{T} = -H \left(1 + \frac{T}{3} \frac{d}{dT} \ln(g_{s*}) \right)^{-1}. \quad (5.13)$$

Then⁴⁹⁾ the BTE Eq. (5.10) can be written in various equivalent forms which can be found in the literature, e.g.,

$$\begin{aligned} \frac{dY}{dx} &= \frac{\langle v \sigma_{\text{ann}} \rangle}{3H} \frac{ds}{dx} (Y^2 - Y_{\text{eq}}^2) \\ &= -\frac{\langle v_\chi \sigma_{\text{ann}} \rangle}{H} \frac{s}{x} \left(1 + \frac{T}{3} \frac{d}{dT} \ln(g_{s*}) \right) (Y^2 - Y_{\text{eq}}^2) \\ &= -\sqrt{\frac{\pi}{45}} g_{\text{eff}}^{1/2} m_\chi M_{\text{P}} \frac{\langle v \sigma_{\text{ann}} \rangle}{x^2} (Y^2 - Y_{\text{eq}}^2), \\ g_{\text{eff}}^{1/2} &= \frac{g_{s*}}{\sqrt{g_{\rho*}}} \left[1 + \frac{T}{3} \frac{d}{dT} \ln(g_{s*}) \right], \end{aligned} \quad (5.14)$$

where $M_{\text{P}} = 1/\sqrt{G_{\text{N}}}$ is the Planck mass. We have used the expression for the entropy as quoted in the line above

⁴⁹⁾The expression of the first equality below can be found in [25] which is related to the second equality by the identity $ds/dx = -(3s/x)(1 + (T/3)d(\ln(g_{s*}))/dT)$.

Eq. (5.13), and the expression for the total energy density, Eq. (9.136), i.e., $\rho = (\pi^2/30)g_{\rho*}T^4$ in the expression of the Hubble expansion rate $H \equiv \sqrt{(8\pi G_{\text{N}}/3)\rho}$, assuming radiation dominance. From the expression of the second quality sign of Eq. (5.14) we can rewrite the BTE in yet another form which is particularly revealing of the physical properties of the BTE,

$$\frac{x}{Y_{\text{eq}}} \frac{dY}{dx} = -\frac{\Gamma_{\text{ann}}^{(\text{eq})}}{H} \left[1 + \frac{T}{3} \frac{d \ln(g_{s*})}{dT} \right] \left[\left(\frac{Y}{Y_{\text{eq}}} \right)^2 - 1 \right]. \quad (5.15)$$

We will come back to this equation later.⁵⁰⁾

The BTE, in any of the forms given above, in their full generality, having temperature dependent g_{s*} and $\Gamma_{\text{ann}}^{(\text{eq})}$, has to be solved numerically. They are the non-linear, Riccati equation, which have no known general analytic solutions.⁵¹⁾ However, knowing the behavior of the equations, an approximate solution can be obtained.

The ratio $\Gamma_{\text{ann}}^{(\text{eq})}/H$ is the controlling factor in the solution of the BTE. In the early universe of high temperature, $T \gg m_\chi$, the $\chi\bar{\chi}$ equilibrium annihilation rate $\Gamma_{\text{ann}}^{(\text{eq})}$ is much larger than the universe expansion rate H . When χ starts in equilibrium, the large coefficient $\Gamma_{\text{ann}}^{(\text{eq})}/H$ in BTE Eq. (5.15) enforces χ to stay in equilibrium due to the negative feedback effect. As the universe expands both H and $\Gamma^{(\text{eq})}$ decrease, with the latter slower than the former, the coefficient $\Gamma_{\text{ann}}^{(\text{eq})}/H$ decreases.⁵²⁾ As this tendency continues, at some point $\Gamma_{\text{ann}}^{(\text{eq})}/H$ will become sufficiently small, Y is able to break away from Y_{eq} . This will happen in the case of massive particle and it is crucial mechanism for the existence of relics. If a particle species is unable to break away from its equilibrium state, its number density would become exponentially small and there would be no relics left eventually. The temperature which begins the departure of n_χ from the equilibrium form is denoted as the freeze-out temperature T_f with the x -variable x_f :

$$\Gamma_{\text{ann}}^{(\text{eq})} = H|_{x=x_f}. \quad (5.16)$$

For $x_f < 2$, the freeze-out takes place when χ is relativistic resulting in a hot relic, while for $x_f > 2$ the freeze-out for non-relativistic χ which is a cold relic.

5.3.2 Hot relics

This is the case when the particle freezes out when it is still relativistic, so $x_f \lesssim 2$.⁵³⁾ At the decoupling or

⁵⁰⁾Equation (5.15) is given in [74] Eq. (5.26) and in [25] Eq. (3), while the possible temperature dependence of g_{s*} is neglected.

⁵¹⁾which we will discuss in some detail in Section 11.3.

⁵²⁾Since $n_\chi^{(\text{eq})} \sim T^3$ and in a radiation dominated era $H \sim T^2$, $\Gamma_{\text{ann}}^{(\text{eq})}/H$ decreases like T or faster.

⁵³⁾The discussion in the present subsection follows that given in [74], pp 122–123.

freeze-out χ is still relativistic and

$$Y_{\text{eq}}(x_f) = \frac{n_{\chi}^{(\text{eq})}(x_f)}{s(x_f)} = \frac{45\zeta(3)}{2\pi^4} \frac{g_{\chi^*}}{g_{s^*}(T_f)} \quad (x_f < 2) \quad (5.17)$$

is constant.⁵⁴⁾ Then the asymptotic value of Y , Y_{∞} , is insensitive to the temperature. We have

$$Y_{\infty} = Y_{\text{eq}}(x_f) = \frac{45\zeta(3)}{2\pi^4} \frac{g_{\chi^*}}{g_{s^*}(T_f)} \quad (x_f < 2). \quad (5.18)$$

To calculate the number density at the present epoch, we can identify $Y_0 = T_{\infty}$ and call the present entropy s_0 [13]

$$n_{\chi 0} = s_0 Y_{\infty} = s_0 \frac{45\zeta(3)}{2\pi^4} \frac{g_{\chi^*}}{g_{s^*}(T_f)} \quad (x_f < 2). \quad (5.19)$$

In the present case the equilibrium density of χ , $n_{\chi}^{(\text{eq})} \sim T^3$, and the cosmic entropy density $s \sim T^3$. Then Y_{eq} is time independent. For relativistic particles the number density of χ and the entropy of the system are given in Eqs. (9.119) and (9.122),

$$\begin{aligned} n_{\chi}(T) &= g_{\chi^*} \frac{\zeta(3)}{\pi^2} T^3, \\ s(T) &= g_{s^*} \frac{2\pi^2}{45} T^3, \end{aligned} \quad (5.20)$$

where $\zeta(3) = 1.20206$. $g_{\chi^*} = g_{\chi}$ for boson and $g_{\chi^*} = 3g_{\chi}/4$ for fermion, where g_{χ} is the spin degrees of freedom of χ . g_s is the effective degrees of freedom for the entropy of the system at the temperature T , given by

$$g_{s^*} \equiv g_B + \frac{7}{8} g_F, \quad (5.21)$$

where g_B and g_F are respectively the total degrees of the freedom of the contributions to the entropy from all bosons and fermions. We have

$$Y(T) = \frac{n_{\chi}(T)}{s(T)} = \frac{45\zeta(3)}{2\pi^4} \frac{g_{\chi^*}}{g_{s^*}}, \quad (5.22)$$

which can be treated as a constant. Therefore, the asymptotic value of Y_{∞} which determines the relic density is given by Eq. (5.18). The number density and mass density of the hot relic χ , and its fractional contribution to the universe energy budget are respectively, of χ at the present time are

$$\begin{aligned} n_{\chi 0} &= s_0 Y_{\infty} = \frac{45\zeta(3)}{2\pi^4} \frac{g_{\chi^*}}{g_{s^*}} s_0, \\ \rho_{\chi 0} &= m_{\chi} n_{\chi 0} = \frac{45\zeta(3)}{2\pi^4} \frac{g_{\chi^*}}{g_{s^*}} s_0 \cdot m_{\chi}, \\ \Omega_{\chi} h^2 &= \frac{m_{\chi} n_{\chi 0}}{\rho_c h^{-2}} = \frac{45\zeta(3)}{2\pi^4} \frac{g_{\chi^*}}{g_s} \frac{s_0 \cdot m_{\chi}}{\rho_c h^{-2}}. \end{aligned} \quad (5.23)$$

⁵⁴⁾We recall that at x_f , $n_{\chi}^{(\text{eq})} = g_{\chi^*}(\zeta(3)/\pi^2)T^3$, where $g_{\chi^*} = (3/4)g_{\chi}$ for fermions and $g_{\chi^*} = (3/4)g_{\chi}$ for bosons, and $s = g_{s^*}(T_f)(2\pi^2/45)T^3$, while $g_{s^*}(T_f)$ is g_{s^*} evaluated at T_f . Then $Y_{\text{eq}} = (45\zeta(3)/(2\pi^4))(g_{\chi^*}/g_{s^*}(T_f))$.

where s_0 is the present entropy density and ρ_c the critical density of the universe. Substituting $s_0 = 2891.2 \text{ cm}^{-3}$ and $\rho_c = 1.0538 \times 10^4 h^2 \text{ eV/cm}^3$ [13], we obtain the numerical expressions

$$\begin{aligned} n_{\chi 0} &= 8.0276 \times 10^2 \frac{g_{\chi^*}}{g_{s^*}}, \\ \rho_{\chi 0} &= 8.0276 \times 10^2 \frac{g_{\chi^*}}{g_{s^*}} \left(\frac{m_{\chi}}{\text{1 eV}} \right) \text{ eV/cm}^{-3}, \\ \Omega_{\chi} h^2 &= 7.6178 \times 10^{-2} \frac{g_{\chi^*}}{g_{s^*}} \left(\frac{m_{\chi}}{\text{1 eV}} \right). \end{aligned} \quad (5.24)$$

This implies that there is an upper bound for hot relics. From the third expression of the above equation with the upper bound of the left-handed side $\Omega_{\chi} < 1$, we have

$$m_{\chi \text{hot}} < 13.13 \left(\frac{g_{s^*}}{g_{\chi^*}} \right) h^2 \text{ eV}. \quad (5.25)$$

Since the hot relics can not over-saturate the required dark matter, with $\Omega_{\text{dm}} h^2 = 0.1198$, we have $m_{\chi \text{hot}} < 1.6(g_{s^*}/g_{\chi^*}) \text{ eV}$.

We can use Eq. (5.24) to derive the usual contribution of neutrinos to the mass budget of the universe. The neutrino decouples at the temperature around 1 MeV (See Section 9.5). The particles which contribute to the entropy include the photon three species of neutrinos and anti-neutrinos, and the electron and positron. As shown in the second row of Table 9.4, $g_{s^*} = 10.75$. For the two-component neutrino theory $g_{\nu^*} = 2(3/4)$, we obtain from the third equation in Eq. (5.24)

$$\Omega_{\nu} h^2 = \frac{\sum m_{\nu}}{94.1 \text{ eV}} \quad (5.26)$$

The current neutrino mass bound given by PDG [13] is $\sum m_{\nu} < 0.23 \text{ eV}$ which leads to $\Omega_{\nu} h^2 < 0.00245$, at least a factor 50 smaller than the current dark matter value of $\Omega_{\text{DM}} h^2 = 0.1198$.

5.3.3 Cold relics

For non-relativistic decoupling the freeze-out temperature is smaller than the mass of $x_f > 2$. The decoupling is a result of the time development of the number density of the massive particle under consideration. In the following we first describe this time development quantitatively, then do the calculation following the approach of [101]⁵⁵⁾ based on an expression in Eq. (5.14) together with the facilitating approximation. Let us repeat the argument given below Eq. (5.15). After the thermal production of χ at high temperature, $T \gg m_{\chi}$ or $x \ll 1$, χ is relativistic and in thermal equilibrium with the heat bath of the universe. So it has an equilibrium distribution with its number density proportional to T^3 , which gives $Y = Y_{\text{eq}}$. In this stage of the development, the annihilation rate is much greater than the Hubble expansion rate. So Y is forced to track Y_{eq} very closely. When

⁵⁵⁾[101], pp 74–77.

the temperature decreases, x increases, the equilibrium annihilation rate decreases like x^{-3} (T^3) or faster, while H decreases like x^{-2} (T^2) in a radiation dominated situation. So the ratio Γ_{ann}/H decreases as the temperature decreases. When the temperature decreases below m_χ or $x > 2$ the equilibrium number density takes the non-relativistic form of the Maxwell-Boltzmann distribution which decreases exponential decreasing, $\sim \exp -x$. Around $x_f > 2$ Y departs from Y_{eq} and gets away from Y_{eq} more and more as temperature decreases further. At sufficiently low temperature the density of the particle is so diluted that the annihilation process is no long effective, and the number of particles in a comoving volume freezes to a constant and becomes a relic of the expanding universe. The solution for x_f from Eq. (5.16) depends on the size of the annihilation cross section, Since the larger the annihilation cross section, the larger the annihilation rate, and the later the particle is able to maintain its non-relativistic distribution, hence the lower its relic density. A graphic description of the decoupling process is depicted in Fig. 5.2 [74]. A detailed demonstration of the decoupling can be found in Section 11. See, in particular, Fig. 11.1.

Below we will make a quantitative estimate of the expected relic density. Let us rewrite the third equality of Eq. (5.14) as

$$\frac{dY}{dx} = -\frac{\tilde{\lambda}}{x^2}(Y^2 - Y_{\text{eq}}^2),$$

$$\tilde{\lambda} \equiv \sqrt{\frac{\pi}{45}} g_{\text{eff}}^{1/2} m_\chi M_{\text{P}} \langle v\sigma_{\text{ann}} \rangle. \quad (5.27)$$

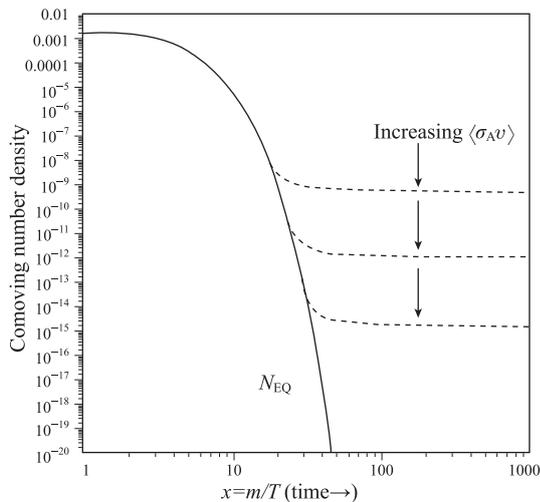


Fig. 5.2 The WIMP comoving number density as in early universe as the temperature decreases. The dashed curve shows the density after decoupling [74]. Note the obvious feature that x_f increases with the annihilation cross section which leads to the decrease of the relic density.

We will approximate $\tilde{\lambda}$ as a constant and focus on the above equation in the region of $T \leq T_f$.⁵⁵⁾ After freeze-out Y is much larger than Y_{eq} which is exponentially suppressed and can therefore be neglected. Equation (5.27) can be rewritten as

$$\frac{dY}{dx} = -\tilde{\lambda} \frac{Y^2}{x^2}. \quad (5.28)$$

Integrating the equation from x_f to ∞ , we have

$$\frac{1}{Y_\infty} - \frac{1}{Y_f} = \frac{\tilde{\lambda}}{x_f}, \quad (5.29)$$

where Y_∞ and Y_f are respectively the values of Y asymptotically and at the freeze-out point. In general $Y_f \gg Y_\infty$, we have the approximate asymptotic value of Y_∞ which allows to determine the freeze-out density,

$$Y_\infty \simeq \frac{x_f}{\tilde{\lambda}}. \quad (5.30)$$

To obtain the freeze-out density we multiply Y_∞ by the entropy density s_1 , where T_1 , which is much higher than the present temperature T_0 , is a sufficiently later temperature when Y reaches its asymptotic value. After T_1 the density will follow the Hubble flow to scale like a^{-3} . Hence the number density at the present time is given by, using $s_1 = (2\pi^2/45)g_{s^*}T_1^3$,

$$n_{\chi 0} = Y_\infty s_1 \left(\frac{a_1}{a_0} \right)^3$$

$$= \sqrt{\frac{4\pi^3}{45}} \frac{g_*^{1/2}}{m_\chi M_{\text{P}} \langle v\sigma_\chi \rangle} T_0^3 \left(\frac{a_1 T_1}{a_0 T_0} \right)^3,$$

$$g_*^{1/2} \equiv \frac{g_{s^*}}{g_{\text{eff}}^{1/2}} = g_{\rho^*}^{1/2} \left[1 + \frac{T}{3} \frac{d}{dT} \ln(g_{s^*}) \right]^{-1}, \quad (5.31)$$

where Eq. (5.14) for $g_{\text{eff}}^{1/2}$ has been used.

5.3.4 The WIMP miracle

To proceed further we have to estimate the values of x_f , g_* , and $(a_1 T_1)^3 / (a_0 T_0)^3$. x_f is the order of a few tens as indicated in Fig. 5.2. We normalize x_f by 10. If nothing happens between T_1 and T_0 , the cosmic temperature which is defined by the photon temperature, which would simply track a^{-1} . Then aT would be constant and $(a_1 T_1) / (a_0 T_0)$ would be unity. But this is not the case. A familiar example is the temperature of the photon and that of the neutrinos. When the cosmic temperature is a few MeV or higher, the photon and neutrinos are in thermal equilibrium and they have the same temperature. Neutrinos are decoupled at about 1 MeV from the photon before the electron-positron annihilation which heated up the cosmic heat bath and hence the photon. Neutrinos are not affected. Hence the cosmic temperature

becomes higher than that of neutrinos. The same mechanism works in the decoupling of χ . Between 100 GeV and 1 MeV, a number of standard model particles are annihilated and hence heat up the cosmic temperature, so $a_0 T_0$ is much larger than $a_1 T_1$. See Table 9.4 for a list of the various channels of annihilation involving standard model particles. Ref. [101] gives $(a_1 T_1)^3 / (a_0 T_0)^3 \approx 1/30$. For the same reason g_* is estimated to be of the order 100. We now make a numerical estimate of the mass fraction of χ at the present epoch and express it in the various units one finds in the literature:

$$\begin{aligned} \Omega_\chi h^2 &\equiv m_\chi n_{\chi 0} / \rho_c = \sqrt{\frac{4\pi^3}{45}} \frac{g_*^{1/2}}{M_P} \frac{x_f}{\langle v\sigma_\chi \rangle} \frac{T_0^3 h^2}{\rho_c} \left(\frac{a_1 T_1}{a_0 T_0}\right)^3 \\ &= 0.282 \left(\frac{x_f}{10}\right) \left(\frac{g_*}{100}\right)^{1/2} \frac{10^{-37} \text{ cm}^2}{\langle v\sigma_\chi \rangle} \\ &= 0.847 \left(\frac{x_f}{10}\right) \left(\frac{g_*}{100}\right)^{1/2} \frac{10^{-27} \text{ cm}^3 \text{ s}^{-1}}{\langle v\sigma_\chi \rangle} \\ &= 0.752 \left(\frac{x_f}{10}\right) \left(\frac{g_*}{100}\right)^{1/2} \frac{10^{-10} \text{ GeV}^{-2}}{\langle v\sigma_\chi \rangle}. \end{aligned} \quad (5.32)$$

Following the usual approach, We can set

$$\langle v\sigma_\chi \rangle = a + \frac{3}{x_f} b, \quad (5.33)$$

where a denotes the s-wave and b the p-wave annihilation cross sections. The average velocity can be taken as $0.27c$. Let us rewrite the second line of Eq. (5.32)⁵⁶⁾

$$\Omega_\chi h^2 = 0.1 \times \left(\frac{x_f}{10}\right) \left(\frac{g_*}{100}\right)^{1/2} \frac{0.282 \text{ pb}}{a + \frac{3}{x_f} b} \quad (5.34)$$

where pb is picobarn (10^{-36} cm^2). Since the contribution of the cold dark matter is $\Omega_{\text{dm}} h^2 = 0.11$, which, according to Eq. (5.34), requires the annihilation cross section to be of the order of pb. This is highly non-trivial. It is the cross section expected from many hypothetical particles in models beyond the standard model, including WIMPs. Hence it is called the WIMP miracle.

5.4 Comments on thermal vs. non-thermal production

Most of the dark matter candidates, such as WIMPs, are produced thermally, as already illustrated in Fig. 5.2. However dark matter candidates can also be produced non-thermally, such as the axion, superWIMP, and WIMPZILLA. Let us briefly summarize their difference.

Thermally produced dark matter refers to the dark matter candidates which are in thermal equilibrium at the early universe. They evolve according to the thermal equilibrium distribution until their reaction rate drops below the Hubble expansion rate, they lost contact with

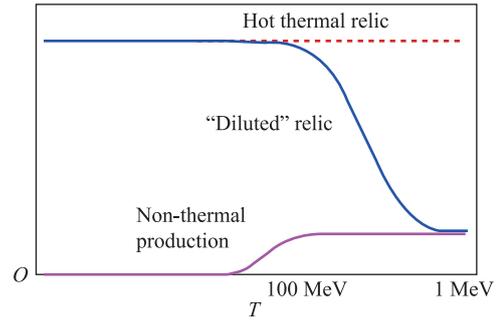


Fig. 5.3 The blue curve is the evolution of thermal relics as a function of the temperature, while the pink curve is that of the non-thermal production of particle which has vanishing number density at very high temperatures. This figure is taken from Ref. [102].

the rest of the universe and are decoupled. Then they expand freely with the Hubble flow and maintain the T^3 behavior of their number density. This kind of particles are also referred to as thermal relics of the early universe. During their thermal evolutions as described above, they are governed by the laws of thermodynamics and cosmology. Otherwise, no free adjustments can be made. This is the reason why the WIMP miracle is so highly valued and suggests that WIMPs should be taken as a serious candidate for the cold dark matter.

Non-thermal dark matter particles are produced through the decays of their parent particles which are present in the very early universe. So the dark matter particle may have a very small or even vanishing number density initially, and increase their number with time as more and more their parent particles decay. Eventually, the non-thermal dark matter particles reach the required number density. From there on they will also expand freely with the Hubble flow and their number density maintains the $T^3 \sim a^{-3}(t)$ behavior. In this way they have never entered equilibrium with the rest of the universe. The difference of the density evolution between thermal relics, which include hot and cold species, and non-thermal production of dark matter particles is illustrated in Fig. 5.3.⁵⁷⁾

6 Dark matter searches I – Mostly WIMPs

As stated earlier, dark matter researches are presently experimentally driven. The astrophysical observations of the effect of dark matter, although critical in providing convincing evidences of its existence, is just half of the story. From the view point of fundamental physics, observations of actions of dark matter in contexts outside gravity are indispensable for identifying the dark matter

⁵⁶⁾This expression is Eq. (3.60) of Ref. [101].

⁵⁷⁾This figure is taken from Ref. [102].

particle and studying its properties. Hence we have to put the dark matter in an appropriate particle physics framework to do the experimental or theoretical study.

The general framework we work in is that the dark matter consists of individual particles, subject to the categorization established in our experience with ordinary particles: having a definite mass and a set of quantum numbers. If the dark matter interacts via only gravity, there is not much can be said from the standpoint of particle physics. However, in addition to the universal gravity, they may also interact, albeit weakly, with ordinary particles. If so, dark matter particles can be put into an extended theoretical framework together with ordinary particles. In this way they can collide elastically with and annihilate each other into ordinary particles. If there is a theoretical framework which is imbedded with both dark matter particles and ordinary particles, we can write down the elastic scattering and annihilation cross sections of dark matter particles and ordinary particles. Then their detection can be searched in appropriate cosmos settings and terrestrial laboratories.

We discuss the search for dark matter particles in this and the next two chapters.

Dark matter drives the dynamics of galaxies and clusters, and generates the growth of large-scale structures. The halo of our own galactic, the Milky Way, should be filled with dark matter particles, with an energy density of the order of 0.4 GeV/cm^3 . We can then try to detect their presence by the signals of their elastic scattering off and annihilation into ordinary particles. In high energy accelerators they may be produced in pairs in collisions of ordinary, SM particles. In any searches, one has to study and verify all aspects of predictions that a relevant theoretical framework can make, in tight control of backgrounds and uncertainties, in order to claim the discover of a dark matter particle. These are not easy tasks to perform, of course. In the particle physics perspective, the most interesting candidates of dark matter are WIMPs together with the axion and sterile neutrinos. These classes of dark matter candidates are well-motivated. The study of what they can do in cosmology is equally important in particle physics.

There are broadly three approaches to search for dark matter particles and study their properties, including WIMPs and other types, not relying on their gravitational effects:

- Direct searches in deep underground laboratories;
- indirect detections in deep underground laboratories, under water or ice, large area telescopes, and satellite experiments;
- accelerator productions of dark matter particles.

These approaches are independent, complementary, and necessary. Direct and indirect searches are necessary in

establishing the existence of dark matter particles. The direct search is to detect signals of their scattering off the nucleus and performed in terrestrial laboratories, relevant to dark matter around Earth. The indirect search is to detect their annihilation into ordinary particles as shown in Eq. (5.3), relevant to dark matter located in our local galaxy where concentrated dark matter distributions are expected. These approaches allow us to establish their existence in the cosmic setting and also get to know what ordinary particles they can interact with. The accelerator production of dark matter particle enables us to study their detailed properties under controlled environment, find out their detailed properties, and decide what particle physics theoretical framework they can be fitted in. In the direct and indirect search for the dark matter particle, the experiments are generally designed with some specific assumptions on properties of the dark matter in mind, and will therefore subject to corresponding uncertainties.

There are a plethora of experiments in search for galactic dark matter in terrestrial laboratories and in the galactic environment, outside the collider and gravitational approaches. They are mostly on going, or in construction. The numerous experiments and the wide varieties of state-of-art technologies employed reflect the fact of how little we know about the dark matter. Unlike collider experiments, which involve hundreds or even thousands of collaborators with the deployment of gigantic detectors, dark matter experiments, currently, are generally moderate efforts, involving a few tens collaborators and working with moderate-sized detectors. The detectors designed are required to be very sensitive to a certain type of signals, with strong background suppression. Hence, most of the present generation of dark matter experiments are a proof-of-existence exploration which is a logical approach for getting a good peek of the unknown in as many different possibilities as possible.

6.1 Worldwide deep underground laboratories

An important fact to consider in the direct search for dark matter candidates is the suppression of any possible backgrounds. A very serious source of the background is the energetic neutrons produced by the penetrating cosmic ray muons originated in the atmosphere, leading to interactions which may be indistinguishable from dark matter events. Since this type of background are controllable, it should be suppressed as much as possible. For terrestrial experiments the practical way to enforce background suppression is to degrade the cosmic muon intensity by shielding a detector with sufficient amount of Earth material, solid or liquid. Hence experiments of direct search for dark matter have to be carried out in a deep underground or deep ice/underwater laboratory. A number of underground laboratories exist worldwide,

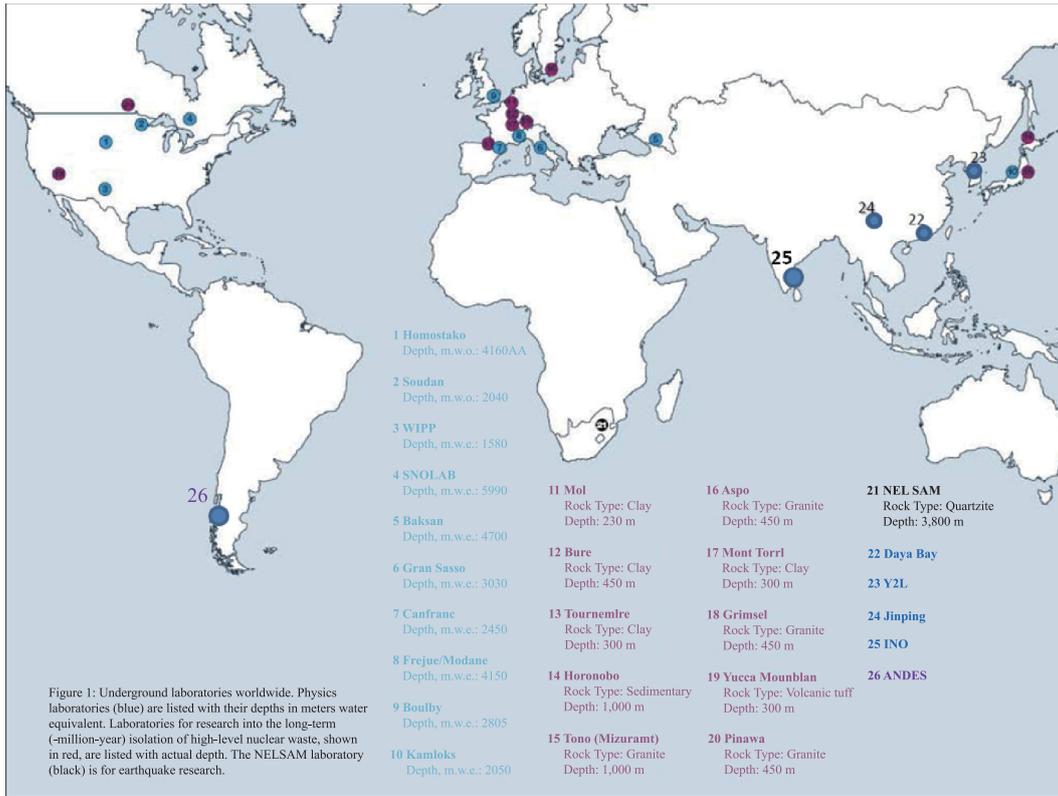


Fig. 6.1 Worldwide distribution of underground physics research laboratories. It is noted that #26, the ANDES Deep Underground Laboratory, being presently a proposal, will be the only science underground laboratory in the southern hemisphere.

which are summarized in Figs. 6.1⁵⁸⁾ and 6.2⁵⁹⁾ Figure 6.1 shows the worldwide locations of existing deep underground laboratories. Figure 6.2 gives the relevant physical parameters of those laboratories that are active in physics and astrophysics research, their depths in meters of standard rock, depths in meters of water equivalent (MWE), and their remaining cosmic ray muon intensities.

Note that the underground laboratories #22 in Fig. 6.1, which did not appear on the original plots, is the Daya Bay underground laboratory, housing the ongoing reactor neutrino experiment of the same name. It is the first experiment which determined accurately the neutrino mixing angle θ_{13} . Underground laboratory #24, the China Jinping Deep Underground Laboratory (CJPL), is a relatively new addition. It has the world's deepest

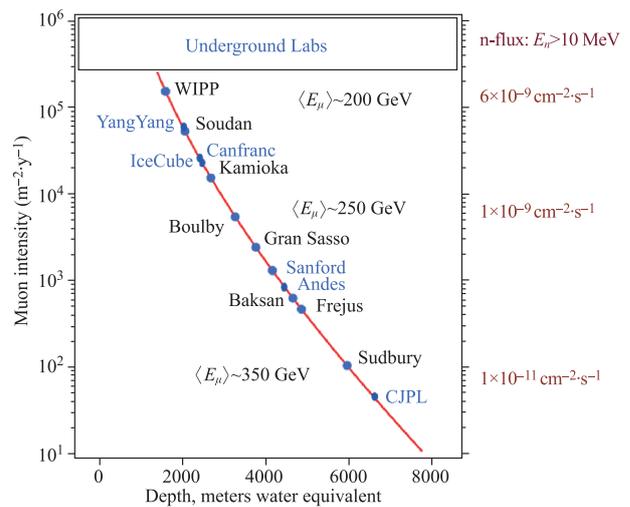


Fig. 6.2 Deep underground science laboratories together with their rock overburdens (top scale), meters water equivalent (bottom scale), and expected cosmic ray muon fluxes and neutron fluxes (vertical scales). This figure is adopted from P. Cushman's talk in [26].

overburden of 2400 m rock or 6720 MWE. Its measured cosmic muon intensity is about 40 per meter square per

⁵⁸⁾The present figure is an updated version of the plot which, appearing in 2007 in an article entitled Deep Science, and containing 21 sites. It be found in <http://www.deepscience.org/contents/facilities.shtml>

⁵⁹⁾A plot of this kind also appeared in the 2007 "Deep Science" article and can be found in http://www.deepscience.org/contents/underground_universe.shtml The present figure is adopted from P. Cushman's talk in Ref. [26], entitled *WIMP Direct Detection Searches: Background Technology*.

year. An early internal discussion of CJPL can be found in TAUP 2011 [104] and a more recent summary is given in [105]. #26, the ANDES Deep Underground Laboratory, is in the proposal stage. If built, it will be the first underground science laboratory in the southern hemisphere. It has the unique capability to settle the controversy of the annual modulation data which have appeared in three dark matter experiments, in which DAMA has the most and first data, to determine critically if the modulation is due to dark matter or not. A Summary of the world's underground laboratories in the first decade of the new century can be found in TAUP 2011 [104]. For a more recent collection of descriptions of deep and large underground laboratories, see [106].

6.2 Direct searches – Dark matter in terrestrial laboratories

According to the model of the cosmological structure formation, the luminous matter of a galaxy is gravitationally bound to a more massive, sprawling halo of dark matter. In the Milky Way the solar system and therefore Earth would constantly be passing through a flux of dark matter particles distributed in the halo of the galaxy. Once in a while the dark matter particles, such as a WIMP, could interact with ordinary matter entities, and, in the process, a nucleus is knocked off the ordinary matter entity, signaling its encounter with a dark matter particle.

Given the weak interaction scale of the expected dark matter-ordinary matter scattering, according to the WIMP miracle, galactic dark matter particle can deposit a certain amount of energy in the detector. The order of magnitude estimate of this energy deposit is not large, in the order of 1–100 KeV which we shall show later. This amount of energy deposition is measurable with a suitably sensitive detector. The detection signatures involve heat, light, or ionization. Roughly the detection rate can be estimated as

$$R = \varpi n_{\chi} \langle v_{\chi} \sigma_{\text{ela}} \rangle \quad (6.1)$$

where ϖ is a numerical constant, n_{χ} the WIMP number density, v_{χ} the WIMP velocity relative to the detector in question, and σ_{ela} the elastic scattering cross section. Note that Eq. (6.1) has the right dimension of inverse time. The majority of current experiments use one of the following three detector technologies individually or in combinations: cryogenic heat production, cryogenic ionization, and scintillation light. There are other technologies proposed that have the potential to apply to the case of very small cross sections of the order of 10^{-8} to 10^{-10} pb.

Direct searches are usually for WIMP mass in the range of the order of several GeV to hundreds of GeV. The experimental search results are usually present in

plots of cross section vs WIMP mass. The sensitivity for low WIMP mass is constrained by the threshold effect of the detector technology used due to low nucleus recoil kinetic energy, as will be explained in Eq. (6.2) below when we make a cursory look of the formulation of direct WIMP searches. The high WIMP mass sensitivity is limited by Eq. (6.1) due to the fact that the fixed energy density of the galactic dark matter, which is about ~ 0.4 GeV/cm³, makes the WIMP number density n_{χ} small for very large WIMP mass.

Both cryogenic and scintillation light detectors are capable of distinguishing the signal of dark matter particles scattered off nuclei from the background. Background events are ordinary particles scattered off electrons contained in the detector target that are made of ordinary atoms. A schematic description of the process of the nucleus-WIMP scattering, together with the background events of atomic electron scattered by photons and electrons, is shown in the left panel of Fig. 6.3.

6.2.1 A cursory look of the basic formulation

We will take a quick look of some of the basic formulae of experimental analysis necessary in the search of dark matter particles. This is by no means a comprehensive review of the formulation of the physical analysis. We do not even try to push the argument to its logical conclusion. But going through some of the arguments to be present below allows us to see the physics involved. We follow the treatment of [107–110], to which we refer for more details.

Let us first summarize the properties of the local dark matter distribution. As we have alluded to before, the WIMP local density in the galactic halo is about $\rho_{\chi 0} \approx 0.4$ GeV/c²/cm³. We now consider the case of a WIMP of mass m_{χ} diffusing in the solar system and scattered elastically off a nucleus of mass m_A . We would like to know the rate of this types of encounters. In the earth rest frame, called simply the earth system, the nucleus recoils in the direction of angle θ_r with respect to the initial WIMP incident direction. See the right panel of Fig. 6.3 for an illustration. The initial velocity of χ relative to the detector in question, denoted by v , is dominated by the galactic rotation of the solar system which is about 220 km/s.^{60),61)} So the process can be treated

⁶⁰⁾Note that the earth orbital speed is about 30 km/s and the maximal tangential speed due to the rotation of Earth around its axis is about 0.5 km/s. So to our lowest order estimate we neglect the earth motion. However, in the detail consideration, such as the DAMA/LIBRA experiment, which we will discussed later, it is exactly the earth motion which is supposedly to provide the signal.

⁶¹⁾Recently the National Astronomical Observatory of Japan (NAOJ) announced an updated value of the solar galactic rotation velocity of 240 km/s. See www.nao.ac.jp/E/release/2012/10/03/mass-of-dark-matter-revealed-by-precise-measurements-of-the-galaxy.html. But this difference does not affect the discussion given here.

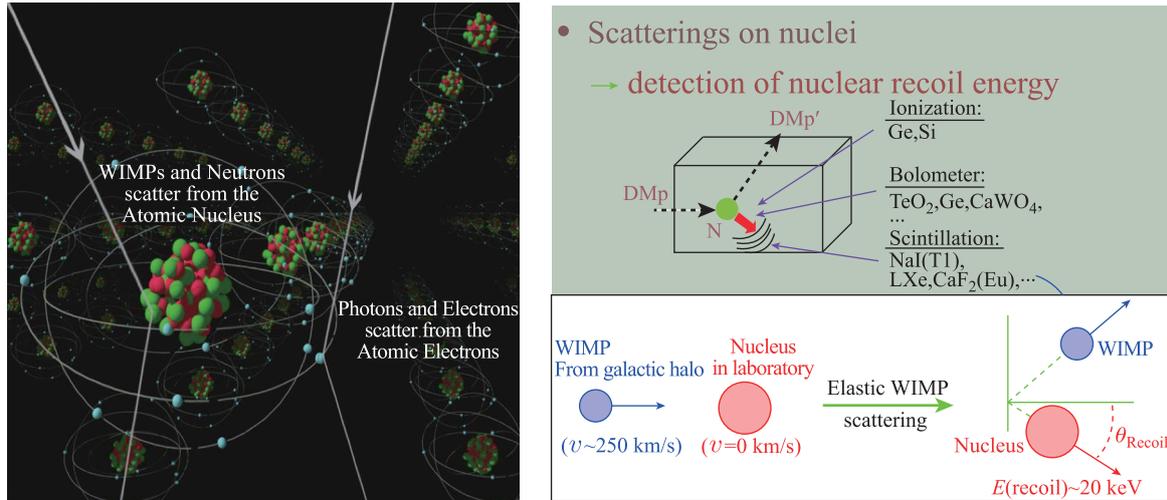


Fig. 6.3 Schematic diagram for the WIMP-nucleus scattering and detection. A single scatter nuclear recoil event of ionization which about 3 times lower than the electron recoil from the dominant radioactive background of gamma rays and electrons.

non-relativistically.

Consider the elastic scatter of a WIMP of mass m_χ with a nucleus of nucleon number A and mass m_A . It is straightforward to calculate the nucleus recoil momentum q_R , which is the momentum transfer to the nucleus in this collision process,

$$\begin{aligned} q_r &= 2m_R v \cos \theta_r \\ m_R &= \frac{m_A m_\chi}{m_\chi + m_A}, \end{aligned} \quad (6.2)$$

where m_R is the reduced mass of the $\chi - N$ system. The kinetic energy of the recoiled nucleus is

$$E_r = \frac{q_r^2}{2m_A} = \left(\frac{1}{2}m_\chi v^2\right) \left(\frac{4m_\chi m_A}{(m_A + m_\chi)^2}\right) \cos^2 \theta_r. \quad (6.3)$$

In the limiting cases, we have

$$E_r \approx \frac{1}{2}m_\chi v^2 \cos^2 \theta_r \begin{cases} 4 \frac{m_A}{m_\chi} & \text{for } m_\chi \gg m_A, \\ 1 & \text{for } m_\chi = m_A, \\ 4 \frac{m_\chi}{m_A} & \text{for } m_\chi \ll m_A. \end{cases} \quad (6.4)$$

So the maximal nuclear recoil occurs when the mass of the WIMP equals to that of the nuclear target. Since $v^2 \sim 10^{-6}$, generally the target nucleus gains a very small kinetic energy. The maximal nuclear recoil kinetic energy is $10^{-6}m_\chi$ which happens in the forward scattering. Hence the recoil kinetic energy lies in the range of 0 to 100 keV, for a nuclear target and WIMP masses of the order of 100 GeV.

The detection rate necessarily depends the reaction cross section of the WIMPs and nucleus collision. The situation is complicated by a number of factors. The

cross section can be *spin-independent* or *spin-dependent*. There is also the coherent effect at low energies. The form factor due to the finite size of the nucleus which will also be different for the spin-dependent and spin-independent interactions, etc. A detailed discussion can be found in [107]. The differential cross section of the recoiled nucleus can be parameterized in terms of the nucleus momentum transfer

$$\frac{d\sigma(q_r)}{dq_r^2} = \frac{\sigma_0}{q_{rM}^2} F^2(q_r), \quad (6.5)$$

where

$$q_{rM} \equiv 2m_R v \quad (6.6)$$

is the maximal momentum transfer which takes place in the case of forward scattering. $F(q_r)$ is a form factor normalized to $F(0) = 1$, and σ_0 is the total recoil cross section [111] which consists of a spin-independent part σ_{0id} and a spin-dependent one σ_{0de} :

$$\begin{aligned} \sigma_0 &\equiv \sigma_{0id} + \sigma_{0de}, \\ \sigma_{0id} &= \frac{4\mu^2}{\pi} (f_p N_p + f_n N_n)^2, \\ \sigma_{0de} &= \frac{4\mu^2}{\pi} \frac{8G_F^2 (J+1)}{J} (a_p \langle S_p \rangle + a_n \langle S_n \rangle). \end{aligned} \quad (6.7)$$

In σ_{0id} , f_p and f_n are the WIMP couplings to the protons and neutrons respectively and they are about the same magnitude in most theoretical models. So the spin-independent cross section has a coherent enhancement by A^2 , where A is the usual atomic number of the target nucleus. The spin-dependent cross section is proportional to the spin average of the nucleons and there is no coherent enhancement effect.

σ_0 can also be defined in terms of the differential cross section at zero momentum transfer by invert Eq. (6.5),

$$\sigma_0 = q_{rM}^2 \left. \frac{d\sigma(q_r)}{dq_r^2} \right|_{q_r=0}. \quad (6.8)$$

For the detection of the dark matter particles, a rate per unit mass interval R is defined in the following form

$$dR = n_\chi \left(\frac{v}{m_A} \right) \frac{d\sigma(q_r)}{dq_r^2} dq_r^2 \tilde{f}_1(\mathbf{v}) d^3v. \quad (6.9)$$

Note that all variables are defined in the rest frame the detector that contains the scattering targets. Let us explain the various factors:

- n_χ is the number density of dark matter particle given by $n_\chi = \rho_\chi/m_\chi$, with the galactic dark matter density $\rho_\chi \approx 0.4 \text{ GeV/cm}^3$.
- The factor v/m_A defines the rate per unit mass interval. The product $v\sigma$ is the effective interaction volume per unit time.
- $\tilde{f}_1(\mathbf{v})$ is the velocity distribution of the dark matter particle in the earth (rest) system.

In the galactic (rest) frame the velocity distribution takes the Maxwellian form with the most probable speed v_0 ,

$$f_1(\mathbf{v}') = \left(\frac{1}{\sqrt{\pi}v_0} \right)^3 \exp\left(-\frac{\mathbf{v}'^2}{v_0^2}\right), \quad (6.10)$$

where \mathbf{v}' is the dark matter velocity vector and $f_1(\mathbf{v}')$ is normalized to unity when it is integrated over the whole velocity ranges $-\infty < v'_j < \infty$, $j = x, y$, and z . The dark matter velocity vectors in the galactic and Earth (rest) frames are related by the earth's galactic velocity \mathbf{v}_E and they have the following expressions respectively,

$$\begin{aligned} \mathbf{v}' &= \mathbf{v} + \mathbf{v}_E, \\ \mathbf{v}_E &= v_0(1.05 + 0.07 \cos \omega t), \end{aligned} \quad (6.11)$$

where $1.05v_0$ is the galactic velocity of the sun and the term $0.07v_0 \cos \omega t$, with $\omega = 2\pi/\text{year} = 6.39 \times 10^{-8} \pi s^{-1}$ is due the correction introduced by the rotation of Earth around the sun [110]. We can now write

$$\tilde{f}_1(\mathbf{v}) = f_1(\mathbf{v} + \mathbf{v}_E). \quad (6.12)$$

Integrating over the WIMP velocity distribution for fixed nucleon recoil energy, we have

$$\frac{dR}{dE_r} = \frac{n_\chi \sigma_0}{2m_R^2 m_\chi} F^2(q_r) \int_{v \geq v_{\min}} d^3v \frac{f_1(\mathbf{v} + \mathbf{v}_E)}{v}, \quad (6.13)$$

where

$$v_{\min} = \sqrt{\frac{m_A E_r}{2m_R^2}} \quad (6.14)$$

is the minimal velocity of the dark matter particle can have in the earth (rest) frame for a fixed nucleus energy E_r . The Integration of the WIMP velocity distribution can be done simply

$$\begin{aligned} & \int_{v \geq v_{\min}} \frac{f_1(\mathbf{v} + \mathbf{v}_E)}{v} d^3v \\ &= \frac{1}{2v_E} \left[\text{erf}\left(\frac{v_{\min} + v_E}{v_0}\right) - \text{erf}\left(\frac{v_{\min} - v_E}{v_0}\right) \right], \end{aligned} \quad (6.15)$$

where $\text{erf}(x_0)$ is the error function defined by

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2} dx. \quad (6.16)$$

We obtain the expression for differential rate

$$\begin{aligned} \frac{dR}{dE_r} &= \frac{n_\chi \sigma_0}{4m_R^2 v_E} F^2(\sqrt{2m_A E_r}) \left[\text{erf}\left(\frac{v_{\min} + v_E}{v_0}\right) \right. \\ & \quad \left. - \text{erf}\left(\frac{v_{\min} - v_E}{v_0}\right) \right], \end{aligned} \quad (6.17)$$

Let us note that the relevant galactic velocity of the dark matter particle can not be arbitrarily large, bound above by its escape velocity $v_{\text{esc}} = 650 \text{ km/s}$. The dark matter particle can no long stay in the galaxy if its velocity exceeds v_{esc} . So to be exact, the existence of v_{esc} should be taken into account as the upper limit of the above velocity integration. However the contribution from the velocity regime higher than the escape velocity is generally very small, and the upper limit of integration can be taken as infinite.

Going far enough for our purpose, we will stop pursue the subject further. To continue the discussion will push us into the consideration of the nuclear form factor. Discussions so far have provided an illustration of the complications involved in the physics of the detection of dark matter particles.

6.2.2 Direct search experiments

Let recall that direct search experiments look for energy deposited in detectors by dark matter particles present in the Milky Way halo. A large number of experiments are running, some completed, and some in the process of construction, or being proposed. Lists of experiments, both direct and indirect searches, are given in Fig. 6.4. The lists include experiments shown in the Dark Matter Portal [112] and Dark Matter Hub [113]. Both provide hyperlinks to individual experiments. The lists in Fig. 6.4 also contain names not included in the Dark Matter Portal and Hub.⁶²⁾

⁶²⁾There are other websites which give lists of dark matter search experiments with hyperlinks to homepages provided, for instance in Wikipedia: https://en.wikipedia.org/wiki/Category:Experiments_for_dark_matter_search

Direct dark matter search

ADMX	ArDM	CDMS	COUPP	CoGeNT	CRESST	CUORE
DAMA	DAMIC	DarkSide	DEAP/CLEAN	DM-TPC	DM ice	Drift
Edelweiss	Eureca	IGEX	LIBRA	LUX	MIMAC	NAIAD
NEWAGE	ORPHEUS	PandaX	Picasso	ROSEBUD	SIMPLE	TEXONO
UKDMC	Xenon	XMASS	WARP	ZEPLIN	ANAIS	CDEX
DARWIN	FUNK	KIMS	CASPAR	HDMS	Majorana	Ultima
COSM	LZ	MAX/XAX	SIGN	CAST	LAXO	

Indirect dark matter search

AMANDA	AMS	ANTARES	BAIKAL	BESS	CAPRICE	GAPS
GLAST	HEAT	IceCube	IMAX	MACRO	Nestor	NINA
PAMELA	Super-K	ACTS	ATIC-2	CTA	EGRET	FermiLAT
HESS	MAGIC	MASS	PPB-BETS	VERITAS	Whipple-GC	

Fig. 6.4 Revised list of experiments of direct and indirect searches for dark matter given in [112] and [113]. Some on the list have been completed and some others are proposals which may not be carried out. Also some are combined to form a larger detectors and some names have been changed. The indirect search experiment GLAST has been changed to FermiLAT, so experiment has been double counted in the list. To the indirect search list the *ALPS* Collaboration should be added.

We make a few remarks on the above list:

- Some of the direct search experiments have been combined. The DAMA and LIBRA are merged to call DAMA/LIBRA. The LUX and ZEPLIN are combined to be referred to as LZ.
- ADMX is a direct search experiment for the axion. FUNK is for the direct search for light-weight dark matter particles called hidden photons.
- There are several new additions to direct search experiments in Asia in addition to Super-K: CDEX, PANDA-X, TEXONO, and KIMS. CDEX and Panda-X are experiments housed in the world’s deepest underground laboratory, China Jinping Underground Laboratory (CJPL) [105]. CDEX has its first result published in 2013 and Panda-X in 2014. The KIMS collaboration is housed in the Yangyang underground laboratory in South Korea.⁶³⁾ KIMS has its first physics result published in 2012 [114].
- Also note that some of the indirect searches are balloon experiments.
- In the list there are proposed dark matter experiments that had been mentioned at one time or another, but did not seem to go forward are also included. They include direct searches CASPER,

⁶³⁾The Yangyang Laboratory is 700 m underground, similar to the Soudan and Kamioka laboratories, all have an overburden of about 2000 MWE.

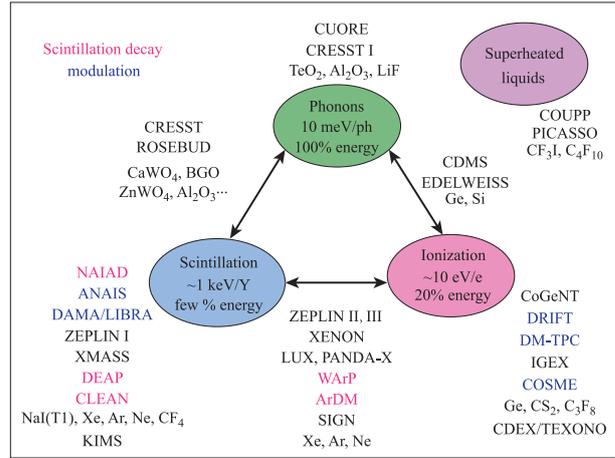


Fig. 6.5 Technologies and detector material employed by direct dark matter search experiments [116].

CAST, COSM, HDMS, IAXO, Majorana, MAX/XAX, Ultima, etc., and indirect searches ACTS, ATIC-2, EGRET, GLAST, PPB-BETS, Whipple-GC, etc.

- Discussions of some of the newer experiments and updates of the existing ones as of 2012 can be found in IDM2012 [115].

The Current results, lacking any clear signals from more than 20 running experiments, necessarily points to the requirement of going into new directions and applying new technologies and approaches to further searches. Another new experimental trend is to increase the size of the detector, which can increase the data statistics and embrace more than one type of technology in a given detector. This can be achieved, as in some next generation experiments, by joining forces of different collaborations. An example is the LZ collaboration, which combines the efforts of LUX and ZEPLIN and is expected to have three-order of magnitude improvement over the current WIMP sensitivity. Another example is the EURECA (European Underground Rare Event Calorimeter Array) experiment, which will include the future European 1 ton cryogenic dark matter search and three other existing collaborations, EDELWEISS, CRESST, and ROSEBUD, and plus possibly new members.⁶⁴⁾

Detection technologies, as mentioned above, used by various experiments can be illustrated in a cartoon shown in Fig. 6.5 [116]. We refer to [116] and [115] for the description of a number of the experiments. Detailed descriptions of many of the experiments can also be found

⁶⁴⁾To manage the multiple efforts, a network of European national government agencies responsible for coordinating and funding national research efforts in astroparticle physics, including dark matter searches, Called ASPERA (ASTroParticle ERAnet), has been established since July 2006.

in their homepages that are hyperlinked through [112] and [113].

6.3 Indirect searches – Galactic dark matter

Indirect searches for dark matter look for the occasional dark matter annihilation or decay into standard model particles. They can take place in wide regions of the cosmos, in Earth, the sun, the halo of the Milky Way, and halos of other galaxies. A cartoon description of galactic locations where higher concentration of dark matter are expected is given in Fig. 6.6.

Indirect searches are sensitive to the interactions of dark matter particles with normal particles and can probe the dark matter annihilation cross section. The manifestation of a signal is usually an anomalous event of cosmic rays. As a result of the annihilation, the final state particles to be studied include the proton, light nuclei, electron, photons, neutrinos and their antiparticles. The anti-particle signal can be particularly sensitive because of their rarity in normal cosmic ray events. The detectors used for the study are either land-based, satellite-based, under water, or imbedded in thick ice, employing state-of-the-art technology. The cosmic ray has had a very long history in particle physics research and has essentially initiated the experimental study of elementary particles.⁶⁵⁾ The indirect search of dark matter complements the direct search. It may be able to explore dark matter of unusual situations, such as of very high masses, unusual couplings, etc. We should mention, however, on general ground, the γ -ray background is not fully understood, so signals based on anomalous behavior of γ -event may not be conclusive.

WIMPs, such as neutralinos, can annihilate into ordinary particle-antiparticle pairs, e.g., $\gamma\gamma$, $\nu\bar{\nu}$, e^+e^- , $p\bar{p}$, etc. Due to the non-relativistic nature of the dark matter particle, the pair of final state particles are monoenergetic of the energy $m_\chi/2$. A schematic diagram for the annihilation processes is given in Fig. 6.7. Let us describe some of the interesting signals:

- Final state neutrinos can be a good signal, especially for the concentration of dark matter particles gravitationally trapped in the sun and in the center of Earth. Being neutrinos they can easily escape the surrounding dense material of their production sources to reach detectors near Earth surface.
- An interesting class of signals is the excess of antiparticles including positrons, antiprotons, and possibly anti-deuteriums, as anti-particles are very rare in nature.

⁶⁵⁾For a readily accessible and concise historical description, including the space program as part of the cosmic ray study, we refer to the very readable historical introduction of [117].

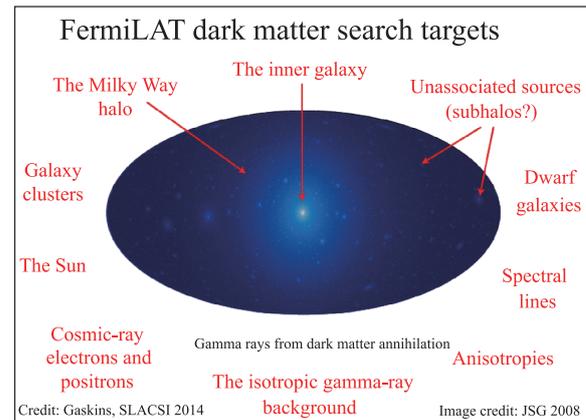


Fig. 6.6 A cartoon description of Galactic localities where higher concentration of dark matter is expected.

- When the directional information of the final state particles is available, pointing in the direction of a dense medium, such as the center of our galaxy, nearby dwarf galaxy, the sun, and the center of Earth where higher density of dark matters is expected, such events can add to the weight for the dark matter signal.
- Since annihilation processes are flavor blind, muon neutrinos and anti-neutrinos resulting from annihilation in the earth may cause excess muons to show up in very deep underground detectors.

The event rate of the annihilation process depends on the annihilation cross section and the square of the dark matter particle density,

$$R_{\text{ann}} \propto \int n_\chi^2 \langle v_\chi \sigma_{\text{ann}} \rangle dV. \quad (6.18)$$

To identify the signal, it is necessary to distinguish the WIMP annihilation process from possible astrophysical backgrounds. Unlike the case of the direct detection, the dark matter-ordinary matter scattering will happen in the detector, the annihilation process, however, can take place in another part of the galaxy, or even a nearby galaxy. While the wide possibility for signals is an advantage of indirect searches, astrophysical backgrounds can be serious and have to be understood and screened out.

Below we describe possible signals in slightly more detail together with some of their experiments. Note that all these signatures of signals are model dependent. A list of experiments of indirect searches can be found in Fig. 6.4.

- Antimatter, e^+ , \bar{p} , \bar{D} , etc., from pair annihilations of WIMPs in galactic halos from within a few kpc. The detectors include MASS (Matter Antimatter Space

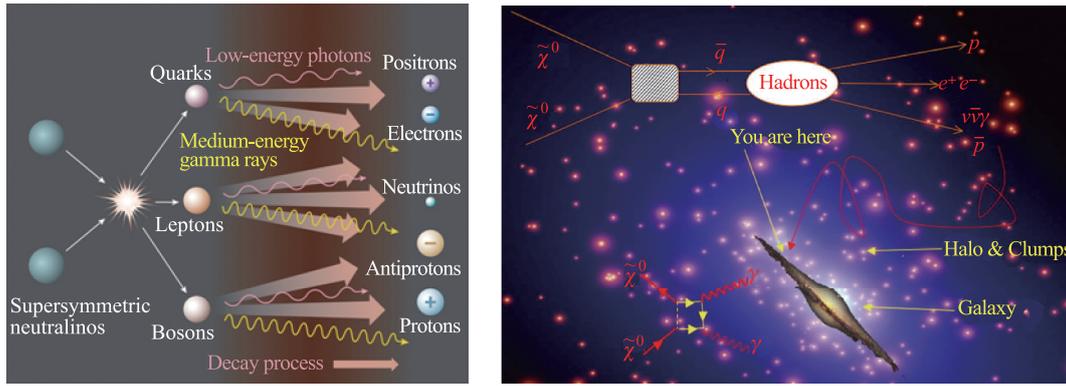


Fig. 6.7 Schematic diagram for the WIMP-WIMP annihilation into particle-antiparticle pairs. As signals of the dark matter, possible dark matter particle decays should be included.

Spectrometer), PAMELA (Payload for Antimatter Matter Exploration and Light nuclei Astrophysics), Fermi Gamma-ray Space Telescope, etc.

- Annihilation of WIMPs into high energy gamma ray pairs in the galactic halo or in extra-galactic sources, especially in the galactic center. The experimental collaborations include GLAST, VERITAS, Whipple GC (Galactic Center) Detection.
- Annihilation into high energy neutrinos taking place in dense media, such as the center of the galaxy, sun, or Earth. The high energy neutrino can easily leave the dense media to reach a detector. The annihilation rate is expected to be high because the possibly high density of dark matter captured there due to gravity. The detectors include: IceCube, Super-Kamiokande, etc.
- There are other ideas: Traces of WIMP annihilations in the WMAP haze [118], dwarf galaxies, rich clusters, etc.

The instruments for indirect searches for dark matter involve usually various forms of telescopes and satellite observatories, including telescopes to search for excess gamma rays, neutrino telescopes for high energy neutrinos, and antiparticle observation for positrons and antiprotons, etc. Classified according to the type of telescopes, we have, for example:

- Gamma-ray telescope: EGRET, Fermi, and MAGIC.
- Neutrino telescope: IceCube, Super-K, and ANTARES.
- Antiparticle telescope: PAMELA for access positrons and antiprotons. Other anti-particle detectors include AMS and GAPS.

Indirect searches are understandably less competitive than direct searches, partly due to data accumulation and background problem. In case dark matter particles

do not annihilate or decay, indirect searches would fail, of course. For a summary we refer to [119].

6.4 High energy collider – Accelerator production

If dark matter particles have been produced in the high temperature environment of the early universe, we should be able to produce them at high energy particle accelerators such as LHC, unless they are exceedingly heavy, say of mass above the TeV scale, or they are leptophilic to interact with leptons only. An important feature of the production of dark matter particles in an accelerator is that normal particles, in conjunction with or being related to dark matter particles, may also be produced. This can facilitate the understanding of the physics behind the dark matter. Extensive searches for new particles had been conducted in high energy colliders in the past, such as LEP and Tevatron. The lack of any positive results could be due to insufficient energies to produce them. In such a case, the much more powerful LHC may be able to produce them.

Let us make a brief remark on SUSY WIMPs relevant to the dark matter search. Experimental signatures of SUSY WIMPs depend on how SUSY is broken and the energy scale at which the symmetry breaking takes place. In a given symmetry breaking scheme, there may be several different models which require different sets of parameters to describe them. Therefore the search of WIMPs has to be made in a given scenario of the SUSY theory. There are a few ways to achieve the symmetry breaking. For example, in GMSB (gauge-mediated super-symmetry breaking), the symmetry breaking is achieved by introducing new super chiral multiplets, called messengers, of mass much less than M_{Planck} , which couple to the ultimate source of symmetry breaking. In the case of R-parity conservation, all super-partners decay into the next lightest super-particle (NLSP) which is either the slepton or sneutrino. The

sneutrino, it decays into a gravitino plus a photon. To solve the SUSY flavor problem, it needs light gravitino with a mass of the order of 1 keV. This does not naturally yield cold dark matter, but it can be a candidate of the warm dark matter.

A large number of works on how to identify dark matter particles in LHC have appeared in the literature and our brief summary here cannot do justice to the subject. For more details we refer to [120]. Mini reviews on both supersymmetry theory and experiment can be found in [13].

The following is a quick tour of LHC: It is a proton on proton collider with the maximal designed energy of 7 TeV for each beam. So the maximal total center of mass energy is 14 TeV. The LHC designed luminosity is $10^{34} \text{ cm}^{-2}\cdot\text{s}^{-1}$, 2 orders of magnitude higher than that of the Tevatron which is at $10^{32} \text{ cm}^{-2}\cdot\text{s}^{-1}$. There is a plan to increase the LHC luminosity by another one order of magnitude to $10^{35} \text{ cm}^{-2}\cdot\text{s}^{-1}$ by 2018.⁶⁶⁾ With the original designed luminosity of $10^{34} \text{ cm}^{-2}\cdot\text{s}^{-1}$, the events rate for the production of standard model particles is quite high, e.g., the $t\bar{t}$ quark pair production is about 1.6 per second or 1.7×10^7 a year running at half of the time. The two large general detectors, ATLAS and CMS, and two smaller collaborations, ALICE and LHCb, have been designed to look for the Higgs particle(s), studies of general features of the standard model, and to search for supersymmetric particles and any other beyond the standard model particles that are within its energy reach, including dark matter particles.

The first operational run were made in 2010–2013 during which the discovery of the Higgs particle was announced on July 4, 2012 with the mass very close to 125 GeV. A new particle and some rare decays have also been discovered, all consistent with the expectation of the standard model. Upgrade to higher collision energy has been made in the period of 2013–2015. The second operational run began on April 5, 2015 with the initial collision energy of 13 TeV and with the expectation of the full designed energy of 14 TeV later. Collection of physics data restarted on June 3, 2015.⁶⁷⁾

The WIMP miracle makes it plausible that LHC may be able to produce dark matter particles. In the case of neutralino dark matter, it can be produced directly or through the decays of other sparticles. Unless WIMPs are very heavy they should be produced copiously at LHC. However, if WIMPs turns out to be very light, their search will become difficult due the overwhelming standard physics background.

The dominant production mechanism of sparticles at

⁶⁶⁾This upgrade of luminosity is called super LHC, or sLHC. Details can be found on the LHC website [121].

⁶⁷⁾An accessible write-up on general aspects of LHC, targeted for the layman but also helpful for scientists, can be found in https://en.wikipedia.org/wiki/Large_Hadron_Collider

LHC involves the parton process of sparticle pair productions:

$$\begin{aligned} q + q' &\rightarrow \tilde{q} + \tilde{q}', \\ q + g &\rightarrow \tilde{q} + \tilde{g}', \\ g + g' &\rightarrow \tilde{g} + \tilde{g}', \end{aligned} \quad (6.19)$$

where \tilde{q} and \tilde{g} are respectively the squark and gluino. In the MSSM the squarks and gluinos decay into quarks, gluons, neutralinos, and charginos. And the heavier neutralinos and charginos decay in turn to the lightest neutralinos ($\tilde{\chi}_1^0$), plus Higgs, Z, leptons, etc. Looking at the final state products, the reaction take the form,

$$p + p \rightarrow \tilde{\chi}_1^0 + \bar{\tilde{\chi}}_1^0 + X, \quad (6.20)$$

where X represents multiple leptons and hadron jets, and $\tilde{\chi}_1^0$ and $\bar{\tilde{\chi}}_1^0$ show as missing transverse energy. The missing energy is an azimuthal imbalance in calorimeter energy as shown in Fig. 6.8.⁶⁸⁾

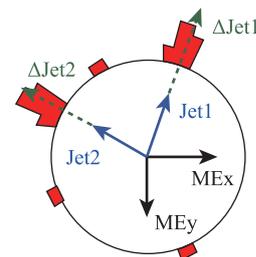


Fig. 6.8 A cartoon to indicate an event with unbalance energy deposit in the calorimeter which implies some energy is missing in the event that can represent a particle or particles escaping the detector. This means that the missing particle(s) does not or has very weak interactions with the detector material.

⁶⁸⁾In the hadron collider, the colliding reactions are in forms of parton-parton collisions. Even though the initial momenta of the colliding particles can not be determined, it is known that the initial transverse momenta in the plan that is perpendicular to the beam axis is very close to zero. So the sum of the transverse momenta of the final state will have to vanish too. A neutral particle leaving the detector without a trace can reveal its presence as a transverse momentum imbalance of the final state. The measurement usually takes in a calorimeter which registers the amount of energy deposited in it. Minimal ionizing particles, such as muons, leave little energy in the calorimeter. These cause a significant vector sum of the transverse energy of all of the detected particles. The imbalance, i.e., the negative of the vector sum in the transverse plane then corresponds to the missing transverse energy (MET).

However, being a great “trick” as it is, since the missing transverse energy measures the vector sum of all particles leaving the detector, we obtain no information concerning the energy/mass/direction of the individual particles involved nor the total number of particles involved. There are also instrumental sources for the missing transverse energy, because of some imperfections of the calorimeter, which may have regions of cracks due to support structures, transition regions between components, etc.

Typically we expect the neutralino event rate to be 10^5 per fb^{-1} integrated luminosity [120], while the ordinary particle event rates is 10^8 – 10^9 , i.e., 3 to 4 orders of magnitude higher. And these events involve ordinary particles can be the background of the WIMP signal. So to identify a signal can be a significant challenge.

If a model is confirmed from the collider data, its predictions for cosmology can be extracted, e.g., using the model parameter to calculate the relic dark matter density and compare it with the WMAP result to determine whether or not the WIMP is indeed a candidate of the dark matter particle, or if other candidates should be looked for. Hence the identification of a WIMP signal at a hadron collider is not equivalent to the discovery of a dark matter particle. It is the beginning of a process that are necessary in order to ascertain that the WIMP particle provides, at least a part of, the gravitation effect of the dark matter which shows so conspicuously in the cosmos.

7 Dark matter searches II – Light particles

In this section we discuss briefly the search for light dark matter candidates, which we define to lie in the broad mass region below 1 GeV. The candidates include the axion, sterile neutrino, axino, SIMP, dark photon, gravitino, fuzzy DM, and possibly others. We will focus on the following interesting candidates, the axion and axion-like particles, and KeV sterile neutrino. They are highly motivated in particle physics for physics beyond the standard model. A large amount of theoretical studies can be found on them.

7.1 Searches for the axion and axion-like particles

The axion was originally proposed for solving the strong CP problem and has been generalized to particles of similar properties. To be specific, the original axion is also referred to as the QCD axion. There are two typical models, the DFSZ model and KSVZ model. Let us sketch very brief some of the simple axion dynamics. For more details we refer to [122] and [123]. The interaction Lagrangian of the axion with ordinary particles can be formulated in terms of a current, arising through its mixing with the neutral pion (π^0) and eta particle (η), and its coupling to photons. Denoting the axion field by ϕ_a , we have the axion effective interaction Lagrangian,

$$\begin{aligned} \mathcal{L}_a &= G_f J_\mu \partial^\mu \phi_a + \frac{g_{a\gamma}}{4} F^{\mu\nu} \tilde{F}_{\mu\nu} \phi_a, \\ G_f &= \frac{g_f}{f_a}, \\ G_{a\gamma} &= \frac{\alpha}{2\pi} \frac{g_\gamma}{f_a}, \end{aligned} \tag{7.1}$$

where J_μ is the Noether-like current from the spontaneous breaking of a global $U(1)$ symmetry, α the electromagnetic fine structure constant, $F^{\mu\nu}$ is the electromagnetic field tensor and $\tilde{F}_{\mu\nu} \equiv \epsilon_{\mu\nu\lambda\tau} F^{\lambda\tau}$ its dual. The quantity f_a , having the energy dimension, is the axion decay constant which provides an energy scale for the axion physics. The coupling parameters g_f and g_γ are model dependent, but they are expected to be the order of unity. For the axion associated with the standard model, we expect $f_a \sim V_{\text{SM}}$, where $V_{\text{SM}} = (\sqrt{2}G_F)^{-1/2} = 247$ GeV is the standard model symmetry breaking scale, and G_F the Fermi constant. Such an axion is referred to as the standard axion. The second term on the right-handed of the axion interaction Lagrangian Eq. (7.1) defines the interaction of the axion with two photons. It plays a crucial role in the axion physics, as we see below, ranging from the search for axion to its effect in the evolution of stars.

The first term on the right-handed side of the first equation in Eq. (7.1) is generally the coupling to fermions. A particularly interesting possibility is the coupling to the electron-positron pair. This coupling can have significant effect in astrophysical processes. An electron-positron coupling can happen in in the DSFZ model, although not in the hadronic model KSVZ. The coupling is in the form $\bar{\psi}_e \gamma^\mu \gamma_5 \psi_e \partial_\mu \phi_a$ which is equivalent to a Yukawa coupling, taking the form

$$\begin{aligned} \mathcal{L}_{ae} &\equiv g_{ae} \bar{\psi}_e i \gamma_5 \psi_e \phi_a, \\ g_{ae} &\equiv \frac{C_e m_e}{f_a}, \end{aligned} \tag{7.2}$$

where ψ_e is the electron field and C_e is a constant related to the ratio of the vacuum expectation values of the two Higgs fields in the model.

The axion mass can be estimated from its mixing with π^0 by the relation $f_a m_a \sim f_{\pi^0} m_{\pi^0}$,

$$m_a = \frac{\sqrt{z}}{1+z} \frac{f_\pi m_\pi}{f_a} \approx \frac{6.0 \text{ eV}}{f_a / (10^6 \text{ GeV})}, \tag{7.3}$$

where $z = m_u/m_d \approx 0.56$, $m_{\pi^0} = 135$ MeV, and $f_{\pi^0} = 92$ MeV is the pion decay constant, and m_u and m_d the up and down quark masses. The above mass expression is written for the case of axion in the meV (10^{-3} eV) range, then the axion decay constant has to be very large in the order of 10^{10} GeV. For a keV axions, the axion decay constant will be the order of 10^4 GeV. For the standard axion with $f_a \sim 250$ GeV, the mass of the axion is about 24 keV. So even with large uncertainties in f_a , axion would be a light dark matter.

The standard axion and those of the variant models (called variant axions), in which quarks of different flavors but the same chirality are assigned with different PQ charges so that they can couple to different Higgs fields, have a mass of the order of hundreds of keV. But they

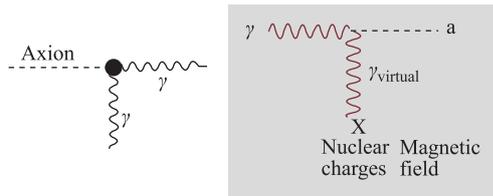


Fig. 7.1 **Left:** The axion and two-photon vertex. **Right:** The Primakoff effect. The virtual photon interact with a nucleus or a magnetic field which is represented by a cross.

have been ruled out experimentally. There remains the possibility of very light axions, often called the *invisible axion*, which has the scale factor $f_a \gg V_{\text{SM}}$, such as the example given in Eq. (7.3). Presently axion searches are focused on the invisible axion with extended dynamics. The most useful coupling of the axion to ordinary particles is with two photons, i.e., the effective $\gamma\gamma\phi_a$ coupling term given in Eq. (7.1), also depicted in the left panel of Fig. 7.1. The two-photon coupling allows the detection of the axion by the Primakoff effect, which involves one virtual photon originated from a nucleus or a region of strong magnetic field. The coupling can be significantly enhanced for nucleus with high atomic number Z . The Primakoff effect is shown in the right panel of Fig. 7.1. The Primakoff photon serves as a signal for the axion.

The Primakoff effect can be the mechanism to produce axions in the early universe which is the environment of dense photons, together with strong magnetic fields and many charged particles. Once produced, the axions can stay in the cosmos long enough, and because the axion can carry away the energy of a stars by free streaming out of its core, it can affect the astrophysical processes of the star. Recently, Ref. [124] using massive stars of 8–12 solar mass to place a bound on the axion-photon coupling:

$$g_{a\gamma\gamma} < 0.8 \times 10^{-10} \text{ GeV}^{-1}. \quad (7.4)$$

We have from Eqs. (7.1) and (7.3), $f_a \gtrsim 3.6 \times 10^6 \text{ GeV}$ and $m_a \lesssim 1 \text{ eV}$.⁶⁹⁾ This limit agrees with, but not yet competitive to, the exiting ones. However it points again to the interesting direction which uses stars as a laboratory for particle physics studies, independent whether of not axion is a dark matter candidate. For a fuller description of the possibilities we refer to [123] and [125].

Constraints on axions have been established through astrophysical and cosmological bounds, as well as laboratory experiments. The laboratory experiments include CAST (CERN Axion Solar Telescope), CARRACK (Axion Research with Rydberg Atoms in a Resonant Cav-

⁶⁹⁾This again says that the axion is not the standard axion. Its physics scale is much higher than that of the standard model. So the physics of the axion, if axion exists, barring some fortuitous reason, should be at a much high energy scale.

ity), ADMX (Axion Dark Matter eXperiment), PVLAS (acronym for its name in Italy), and early experiments RBF and UH. A large span of the mass value has been covered, ranging from sub-MeV to μeV , and the majority of the region has been excluded. But some of the results are quite model dependent. The current bounds can be found in [89], which we adopt in Fig. 7.2 and refer for more details. Some of the features of Fig. 7.2 are summarized below:

- The dark intervals marked as ADMX, Helioscopes and Telescope/EBL are their approximate search regions. Helioscopes refers to CAST and others. Telescope/EBL refers FermiLAT and Chandra X-ray Telescope, where EBL stands for *Extragalactic Background Light*.
- The two green regions show the reach of the planned future upgrades of the respective experiments.
- The Beam Dump represent roughly the exclusion range for standard or variant axions.
- The “Globular Cluster” and “White Dwarfs” results uses the DFSZ model with $C_e = 1/6$. $g_{a\gamma}$ and g_{ae} refer respectively to calculations use axion-photon and axion-electron couplings.
- The “Cold DM” is quite uncertain.

For the current status of the axion searches, let us quote from [89]: “*Direct searches for such particles and calculations of their effect on the cooling of stars and on the supernova SN1987A exclude most values of $f_a < 10^9 \text{ GeV}$. Some of these constrain only the axion coupling to photons ($g_{a\gamma\gamma}$), while others constrain the axion coupling to electrons (g_{ae}). Recent and future laboratory test (the latter shown in light green) can probe $f_a \lesssim 10^9 \text{ GeV}$ or $f_a \gtrsim 10^{12}$ (possibly higher) but intermediate values are more challenging.*”

We have just presented a limited scope of the axion search. Because of its strong theoretical motivation and elegant theoretical formulation, axion has its enthusiastic followers. Ongoing updates of the mass and coupling limit are available in the literature, such as [89]. For recent summer school lectures on the axion theory and experiment, see [26]. For research directions, long term future perspectives, and other aspects of axions we refer to [126].

7.2 KeV sterile neutrinos

A class of particles with the possibility of mass in the KeV order as candidates of WDM has attracted a significant amount of attention. They include sterile neutrinos, light neutralinos, gravitons, majorons, etc. The sterile neutrino is particularly attractive for the particle physics consideration and astrophysics as well. In particle physics, the neutrino oscillation and therefore mas-

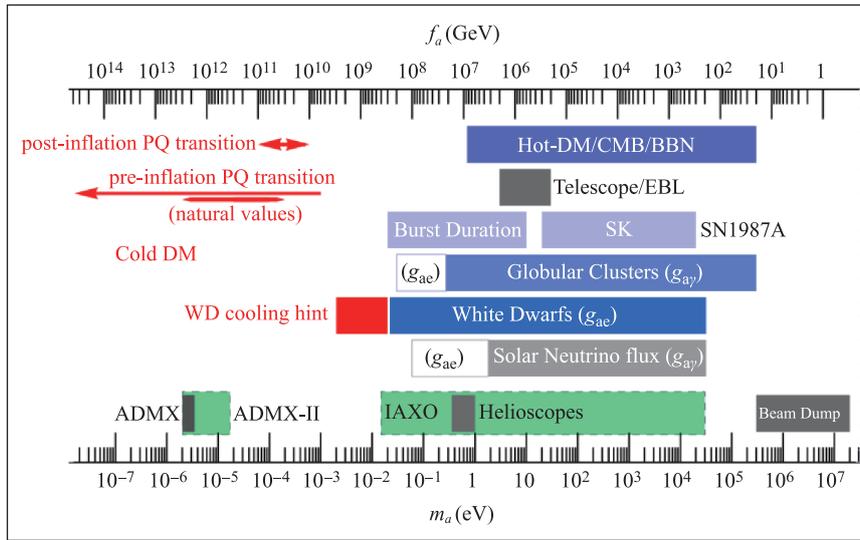


Fig. 7.2 Axion exclusion ranges based on various astrophysical and cosmological observations, and laboratory experiments. See the related text for the explanation.

sive neutrino suggests the right-handed and hence sterile neutrinos, in addition to the three standard model neutrinos. For astrophysics, the motivation is the small scale structure problem of the cold dark matter scenario. The sterile neutrinos, outside the standard model will have very weak interactions with SM particles. Despite the fact that searches for sterile neutrinos of the eV order have been pursued for two decades, in relation to the LSND events [127, 128], and they are largely fruitless, the interests in sterile neutrino is still high. There are good reasons for the persisting interests in the subject in particle physics. On the theoretical side, the idea of the sterile neutrino is very attractive, as it can naturally appear in many theories beyond the standard model as right-handed states. And they can be a realization of the Majorana fermion, a bold and elegant theoretical structure. Experimentally, for some years the MiniBooNE experiment has data that are consistent with LSND, implying the existence of extra neutrino states. For details, we refer to the recent MiniBooNE analysis [129, 130], where a brief summary of the phenomenological status of sterile neutrinos can be found.

There is a large body of literature associated with the subject of sterile neutrinos. To be a dark matter particle, fine tune is generally needed to obtain the right production rate in the early universe. For details, we refer to the website [131] and the recent white paper on the light sterile neutrino [132].

Our discussion below will focus on a simple massive neutrino model constructed by a straightforward extension of the minimal standard model with the addition of a set of right-handed neutrinos. The model is referred to as the ν MSM (ν -minimal standard model) [133, 134].

We will briefly describe the model and some of its phenomenology.

As describe in ν MSM [133, 134], let N right-handed neutral leptons be introduced to the standard model and they are referred to as right-handed neutrinos. Being leptons they are color $SU(3)$ singlet; being right-handed they must be weak isospin $SU(2)$ singlet; and being electric charge neutral their hypercharge $U(1)$ quantum number are zero. These say that (i) these right-handed neutrinos have no standard model interactions; (ii) they can give rise to mass terms with the ordinary left-handed neutrinos through the Higgs mechanism; and (iii) they can be Majorana particles.⁷⁰⁾ The part of the Lagrangian contains the right-handed neutrinos can be written conveniently by setting the charged leptons in diagonal form and the Majorana mass terms for the right-handed neutrino also in the diagonal form.

For the ν MSM, the number of right-handed neutrinos is set to be 3, denoted as N_j , $j = 1, 2, 3$. There are

⁷⁰⁾In theoretical model constructions there is a general unstated principle, that is, once you decide the symmetry of your model, you should include all terms which are allowed in your Lagrangian dynamics as dictated by the symmetry. With this principle, the addition of the right-handed neutrinos is natural. The standard model does not include this term for two reasons. One is that the right-handed neutrino have not standard model quantum numbers so there is no reason to include them. Another could be that before the neutrino oscillation has been confirmed, most people thought that the standard model neutrinos would be massless. It should be remarked that the neutrino mass terms so added does not have to be treated at the fundamental Lagrangian level. They could be effective terms induced by physics beyond the standard model. In retrospect, ν -MSM should have been studied in details years ago, even only on its merit as a very economical way to introduce the Majorana fermions and, therefore, the see-saw mechanism which allows naturally very small neutrino masses.

18 new parameters in addition to those of the standard model. The new parameters are 3 Dirac masses, 3 Majorana masses, 6 mixing angles, and 6 CP violation phases. The physics is arranged in the following way: N_1 gain a mass of order keV, $M_1 \sim \text{keV}$. N_2 and N_3 give masses to the ordinary, active neutrinos and generate baryon asymmetry. Their masses are of the order of the standard model energy scale, i.e., M_2 and M_3 are hundreds of GeV. The sterile neutrinos N_1 has a small mixing and therefore can couple to the standard model particles via charged and neutral currents. The interaction strength, depending on the mixing angle, can be super-weak. All three sterile neutrinos are unstable. N_1 has a lifetimes $\tau_1 > \tau_U$, where $\tau_U \sim 10^{17}$ seconds is the age of the universe. And the lifetimes of N_2 and N_3 are shorter than 0.1 second.

As a dark matter candidate N_1 is quite robust. The production of N_1 can be either thermal or non-thermal and it may or may not be present initial in the early universe. Take the case that N_1 is not present initially in the early universe with its presence through non-thermal production, through lepton and quark annihilations,

$$\ell + \bar{\ell}, \quad q + \bar{q} \rightarrow \nu + N_1. \quad (7.5)$$

N_1 can decay, via the mixing with ordinary neutrino by the one-loop effect involving the W-boson and charged lepton intermediate states, leading to the decay process

$$N_1 \rightarrow \nu + \gamma \quad (7.6)$$

with a lifetime of the form [135]

$$\tau_{\nu_s} = 1.8 \times 10^{21} (\sin \theta)^{-2} \left(\frac{1 \text{ keV}}{m_{N_1}} \right)^5, \quad (7.7)$$

where θ is the mixing angle. With the age of the universe 4.34×10^{17} s, a keV sterile neutrino looks like a stable particle in the present epoch. This decay channel provides the signal for the search for N_1 . The final state γ shows up as a monochromatic photon with the energy $E_\gamma = M_1/2$. This keV X-ray can be searched in the sky. There are also particle physics consequences of this model. In addition to the X-ray line from the radiative decay of N_1 as discussed above, the Majorana nature of the sterile neutrino can cause neutrinoless double β -decay. N_1 can also participate in hadron weak decays, in particular, in kaon decays showing as a missing energy.

The sterile neutrino has been a very active topic in both particle physics and astrophysics. However they are interested in different mass ranges. particle physics, motivated by LSND with not-so-small mixing with the active neutrinos, focuses mostly in the eV mass region,

while being a proposed candidate for WDM, the interesting mass range is in the keV range. Astrophysical searches for keV sterile neutrinos have been made through different channels, such X-rays spectra of galaxies and clusters, the supernova explosion, and other constraints. The X-ray search has excluded a significant range of the N_1 mass-mixing angle space.⁷¹⁾ Figure 7.3 which is adopted from [139]⁷²⁾ is an example of the exclusion plots in the ν MSM model. Note that the allowed region is significantly constrained: The mass of N_1 is greater than 1 keV and the mixing angle is very small $\theta < 10^{-4}$ (see the figure caption). The obvious feature of the constraint are the following: The dark regions above and below are excluded by either too many or too little dark matter particles. We should note that the exclusion by “not enough DM” is due to the requirement that the sterile neutrino is the whole part of the dark matter. In the case of multi-particle scenario, in which sterile neutrino is one of the several species of dark matter particles, this exclusion should be removed. The blue region in the upper right is the exclusion due to the non-observation of X-ray that would have resulted from the dark matter decay. The left pink region is the limit established from the Tremaine–Gunn bound [97] and the Lyman- α observational limit. For more details on the bounds, including references, we refer to [136].⁷³⁾ For a summary of past experiments and a list of the future ones, in both particle physics and astrophysics, we refer to [132]. A website which provides a collection of references to the various aspects of sterile neutrinos and useful links can be found in [137].

7.2.1 A possible candidate of 7 keV sterile neutrino

In 2014 two research collaborations [138] and [139] announced the discovery a weak X-ray line of $E_\gamma = 3.5$ KeV, in a stacked XMM-Newton Observatory spectrum of 73 galaxy clusters. This line has not been known previously to be present in the spectra of galaxies or galaxy clusters, and it has the clear tendency to become stronger toward the centers of the assumed source, and absent in the data set of the deep blank sky. This anomalous x-ray line, if not spurious, can be understood from the decay of a dark matter particle. The mass of the dark matter particle is twice the energy of the X-ray, i.e., 7 keV. Details of the interpretation of this events can be found given in [139] and [140].

This event is shown in Fig. 7.3 as the blue data point. It requires a rather small mixing angle of the order of $\theta \approx 3.5 \times 10^{-6}$ in the ν MSM.

⁷¹⁾Let us recall that the energy of the X-ray ranges from 100 eV to 100 keV. The soft X-ray lies below 10 keV and hard X-ray above 10 keV.

⁷²⁾The figure appears originally in [136].

⁷³⁾Fig. 2, [138]. Various other bounds can be found in [137].

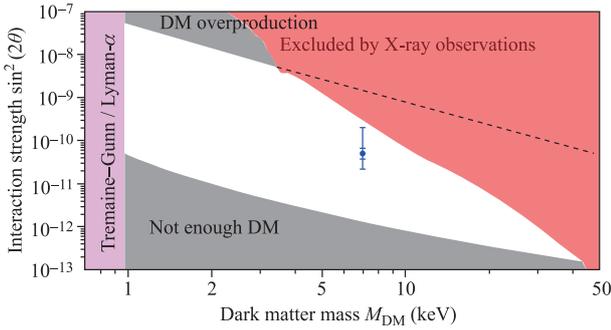


Fig. 7.3 Constraint plot of the keV sterile neutrino in the ν MSM model from astrophysics data. On the vertical axis θ is the mixing angle of the sterile neutrino with the standard model neutrino and it essentially indicates the coupling strength. The allowed region is the unshaded region without a color. The colored-shaded regions are excluded based on different types of data as indicated in the graph. See the text and see [136]. The blue data point at $M_{\text{DM}} = 7$ keV is the 7 keV sterile neutrino discussed in Section 7.2.1.

8 Summary of the status of dark matter searches

To summarize a developing subject, with a number of experiments running, new experiments commencing, and therefore the possibility of new results emerging at any-time, is not a very ensuring endeavor. Serious updates are necessary from time to time, and hence some of the results shown below might be out of date. In any event, this chapter summarizing the status of dark matter searches has the characteristics of tentativeness to exhibit the rapid development of topics of dark matter.

Most searches of dark matter performed to date, direct, indirect, and accelerator, have set increasingly tighter limits. There are also tantalizing signals claimed. But they seem to conflict with other observations and, therefore, need to be understood and confirmed independently. We first describe briefly below those experiments that have made (tentative) claims of a signal, then the summary results of most of the relevant experiments are given. More details can be found in reports given in recent conferences, several have held in first half of 2015.⁷⁴⁾

⁷⁴⁾Let us mention some of the conferences: *Workshop on Off-the-Beaten-Track Dark Matter and Astrophysical Probes of Fundamental Physics*, ICTP, Trieste, Italy, April 13–17, 2015, <http://indico.ictp.it/event/a14282/>; DARK MALT 2015, *Dark Matter: Astrophysical Probes, Laboratory Tests, and Theory Aspects*, Feb. 2–27, 2015, MIAPP, Munich, Germany, <http://www.munich-iapp.de/scientific-programme/programmes-2015/dark-matter/>; NDM15, *Neutrinos and Dark Matter in Nuclear Physics 2015*, Jun. 1–5, 2015, Jyväskylä, Finland, <https://www.jyu.fi/en/congress/ndm15>, <https://indico.cern.ch/event/394248/timetable/#all.detailed>

8.1 Direct searches

Most of the effort in direct searches of dark matter is focused on the searches of well-motivated WIMPs. Their detection is based on searching for nuclear recoils of the KeV magnitude in Earth based detectors. Excesses events of this type, which may be interpreted as candidates of dark matter particles, have been reported quite early by the DAMA/NaI collaboration in 1997. Now 4 collaborations have reported data which contain possible candidates of dark matter: DAMA [141], CDMS [142], CoGeNT [143], and CRESST [144]. All four sets of data are in regions of light WIMPs of mass 6–30 GeV, covering the scattering cross section range $5 \times 10^{-4} - 4 \times 10^{-7}$ pb. However, their allowed mass-cross section parameter spaces are largely non-overlap, except for CoGeNT and CDMS. Furthermore, they lie in regions of exclusion of some other experiments. Below we discuss these four sets of data.

8.1.1 Direct search: DAMA/LIBRA

The DAMA is an observatory for rare processes, based on highly radio pure scintillators. It is model independent and one of earliest detectors used in direct search for Dark Matter particles in the galactic halo along the earth orbital around the sun. The detector is 1400 m underground, located in the Gran Sasso underground laboratory in Italy near the French boarder. The first generation detector the DAMA/NaI has been replaced by the present general detector DAMA/LIBRA.

Let us describe briefly the physics underlying the DAMA detector. In the Milky Way, or the Galaxy, the sun moves at a velocity $v_{\text{sun}} \approx 232$ km/s towards the star Vega near Hercules, with respect to the Local Standard of Rest. Because of the orbital motion around the sun ($v_{\text{orb}} \approx 30$ km/s), the magnitude of Earth's velocity in the Galaxy has an annual modulation which gives rise to a corresponding modulation in the dark matter flux crossing Earth. So Earth should encounter a higher flux of dark matter particles in the halo around June 2nd, and by a lower one around December 2nd. on the former date, Earth's orbital velocity is added to that of the solar system with respect to the Galaxy, while on the latter date, the two velocities are subtracted. Since the event rate of the interaction of dark matter and ordinary particles depends linearly on the relative velocity of the dark matter and the DAMA detector, this will give rise to a seasonal modulation in the event number of the DAMA experiment. This is depicted in the upper panel of Fig. 8.1.

In 1997 DAMA/NaI became the first dark matter search experiment to claim a direct signal of the dark matter [141]. The data accumulation prior to 2013 is 1.17 ton-year data over 13 annual cycles [145] and shows

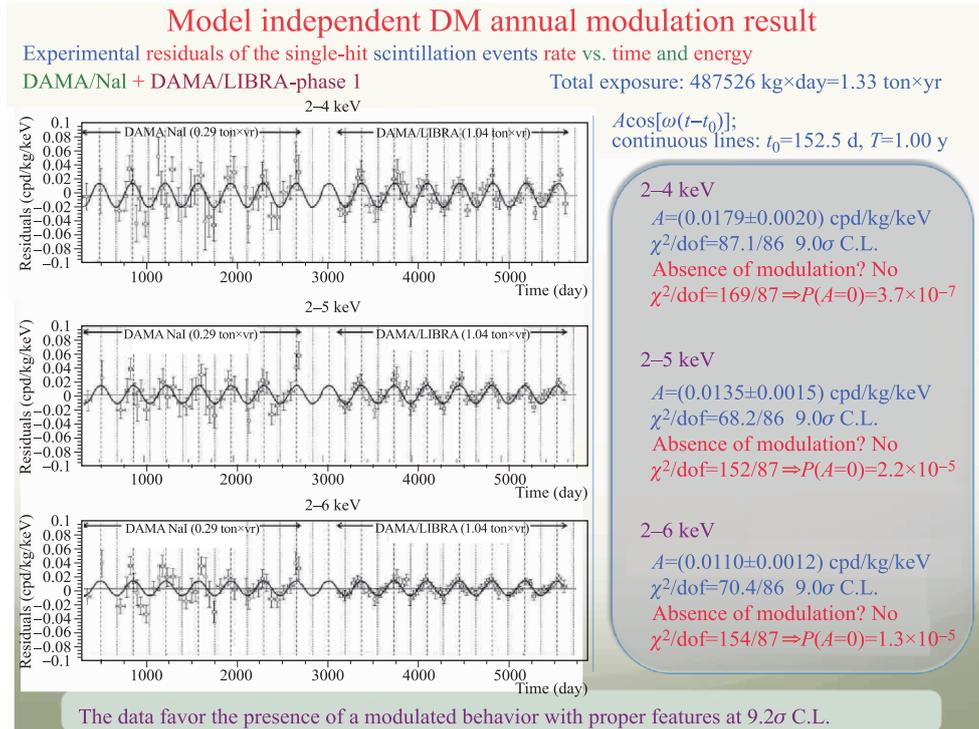
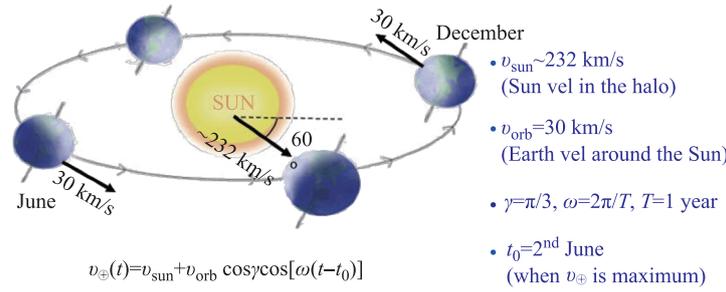


Fig. 8.1 Upper: The velocity modulation of an Earth laboratory. Lower: The DAMA/LIBRA and DAMA/NaI annual modulation data and analysis. The plots are adopted from [145] and [146], which together with other presentations made on DAMA can be found in Downloads of the collaboration homepage [147].

consistently the annual modulation in the counting rate. The most recent data accumulation of 1.33 ton-yr has been reported in [146] which can be found on DAMA's download page [147], as shown in the lower panel of Fig. 8.1, where the result of the data analysis is also given. The annual modulation can be interpreted as the seasonal dependence of the WIMP velocity in the earth rest frame. The modulation is an 9.2 σ effect. With all known systemic and background effects address [145], the collaboration ascertains that the observed annual modulation effect is a model independent evidence of the presence of dark matter particles in the galactic halo. The DAMA collaboration will release a new annual modulation at the end of the DAMA/LIBRA-phase 1. DAMA/LIBRA-phase 2 will begin with some improved hardware and software.

The DAMA data can be fitted to a WIMP-nucleon cross section which is compatible with a MSSM model.

It is pointed out in [148]⁷⁵⁾ that if interpreted in the DM framework, the DAMA data imply either a WIMP of $m_{\chi} = 50 \text{ GeV}$ with the elastic WIMP-proton cross section $\sigma_{\chi p} \approx 7 \times 10^{-6} \text{ pb}$, or a lighter WIMP of mass 6-10 GeV with $\sigma_{\chi p} \approx 7 \times 10^{-3} \text{ pb}$. However, the DAMA data is not compatible with current upper limits from other experiments which have higher solutions, as will be shown later. Proposals addressing the conflict of DAMA with other experiments can be found in the literature, but none has been totally satisfactory. We will describe some of the proposals below. On the experimental side, at least two modulation experiments have been planned in an effort to verify the DAMA result. See [116] for more details and references.

There are suggestions that the DAMA/LIBRA data is due to the cosmic ray background which also varies sea-

⁷⁵⁾See [148], pp 9-10.

sonally [149]. However, it was found in a recent detailed simulation [150] that the muon background contribution is rather small and can not explain the DAMA/LIBRA data. An effort to resolve this quandary by a theoretical framework can be found in [151] which proposed a solution by a specific dynamics: A Dirac dark matter particle interacts with ordinary particle through the exchange of a light pseudoscalar. This coupling enhances significantly the interaction strength between the dark matter particle and the proton over that of the neutron. So the protophilic property of the pseudoscalar particle enables the interaction of the dark matter particle with a nucleus through the unpaired protons or unpaired neutrons, much more significantly for the former than for the latter. The DAMA/LIBRA experiment uses Na(11) and I(53) as targets containing odd number of protons as indicated in the round brackets, while other conflicting experiments, such as Xenon100, use Si(14), Ge(32), and Xe(54) target which contain even number of protons. Hence, in this scenario, only the DAMA/LIBRA experiment is sensitive to the presence of the dark matter.

Despite the impressive consistency of the DAMA data collected for so long a period of time, the community is still not convinced that dark matter has been discovered, and the DAMA data is not a seasonally environment effect. The particular environmental effect in question arises from cosmic rays which is known to have a seasonal varying intensity. Towards a resolution of the conundrum requires a direct demonstration that the DAMA modulation is not caused by such an effect. This can be done by duplicating the DAMA experiment in the south hemisphere. The cosmic ray seasonal oscillation in the south hemisphere is the opposite to that of the north hemisphere, and hence the effect of the current DAMA data, while the supposedly dark matter effect will be the same in both hemisphere. Indeed, there is one ongoing experiments Called DM Ice17 running along with Ice-Cube, since July 2011. The first report, which is on its background data of the first two years running, can be found in [152]. Another experiment is proposed for a deep underground science laboratory ANDES in South American.

8.1.2 Direct search: CDMS-II and SuperCDMS

The Cryogenic Dark Matter Search (CDMS) employs germanium and silicon detectors to search for WIMP dark matter particles. The first general experiment CDMS-I was located in a tunnel under the Stanford University campus. The second generation experiment, i.e., the CDMS-II has been relocated in the Soudan Mine in Norther Minnesota, U.S.A. since 2003. It is about 700 m underground which provides a much better shield against cosmogenic events such as the cosmic ray muons. The

CDMS-II detector is designed to detect minute phonon signals generated within a detector crystal due to elastic collisions of the detector nuclei and the WIMP. The energy deposited in the detector by an incident WIMP may be as low as a few tens KeV. The detector is maintained at a very low base temperature of 10 mK in order to suppress possible energy depositions by thermal activities of detector nuclei. A further upgrade CDMS detector, called the SuperCDMS, which has been running since 2012, will be moved to an even deep underground laboratory, the SNO Laboratory which is about 2000 m underground and the detector volume will be increased too.

The first observation of possible candidates of dark matter was made in 2007 and the announcement of two events for possible candidates was made in 2009, while the expected background is 0.9 ± 0.2 events. The CDMS II results have been published in 2013 [142], and the SuperCDMS results in 2014 [153], They are summarized in Fig. 8.2. In the 2013 results, the CDMSII Si provides the data for the region of possible signals, shown as the light blue region (at the 68% C.L.) in Fig. 8.2, covering a region of possible WIMP mass between 6 and 20 GeV and WIMP-nucleon scattering cross section $10^{-42} - 4 \times 10^{-40}$ cm², with the maximum likelihood point at 8.6 GeV and WIMP-nucleon cross section 1.9×10^{-41} cm² (not shown in the present figure). The dark curve designated as “this result” shows the exclusion region of the SuperCDMS results. An upper limit on the spin-independent WIMP-nucleon scattering cross section of 1.2×10^{-42} cm² at the WIMP mass 8 GeV.

The SuperCDMS will move to the deepest underground Laboratory in North America, the SNOLAB and will be renamed as SuperCDMS-SNOLAB. This new home will allow a 100 times reduction of the cosmogenic neutron background. It is expected to be able to reach the cross section bound of 10^{-46} cm² or 10^{-10} pb for WIMPs of mass in the 30–100 GeV range.

We note that the U.S. Science funding agencies, DOE and NSF have identified three next generation dark matter search experiments to represent the US effort in the search for the dark matter: LZ and SuperCDMS-SNOLAB for low and high mass WIMPs search, and ADMX-Gen2 for the search of axions.

Another new generation of cryogenic experiment, the 1 ton EURECA (European UNderground Rare Event Calorimeter Array), which will replace EDELWEISS, CRESST, and ROSEBUD, is under planning. Its designed reach is comparable to that of SuperCDMS, but eventually down to 10^{-11} pb [155].

8.1.3 Direct search: CoGeNT

The CoGeNT (Germanium Neutrino Technology) collaboration, a dedicated search for low mass WIMPs, uses

New limits from SuperCDMS Ge iZIPs

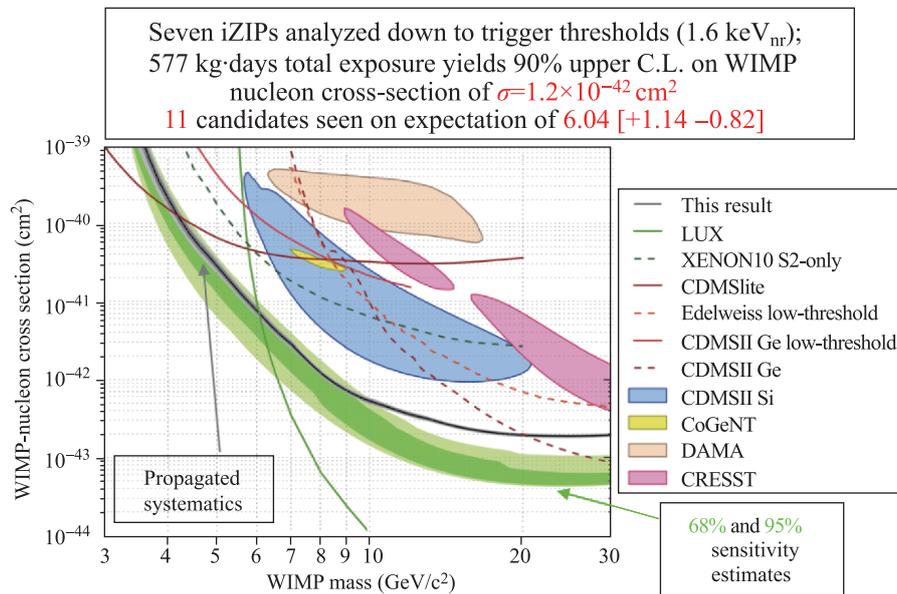


Fig. 8.2 The parameter space of allowed and excluded regions. The basic plot of the various exclusion curves and allowed regions is given in [153]. Taken together with the inserts of explanations of the colored curves and regions, the present plot is adopted from [154]. Note the prominent dark exclusion curve marked as “this result”.

high-purity and low radioactivity Ge for ionization signal. It is also located in the Soudan Mine. The detector is tuned to detect WIMP mass around 10 GeV, much lower than most of the other direct search experiments. The techniques employed [143, 156] are also ideal to search for annual modulation signatures due to low mass WIMPs as DAMA/LIBRA. The experiment started data collection in December, 2009. The CoGeNT data set has been analyzed for the annual modulation effect. With 450 days running, a modulation was found in the data set of $0.5\text{--}3.0\text{ keVee}^{76)}$ [157]. The modulation plot of [143] is reproduced in the upper panel of Fig. 8.3. Recently an independent analysis has been made on the time-dependent properties of the CoGeNT data, using a variety of statistical tests, the authors confirmed the presence of annual modulation within the statistical limit [158] as shown in the lower panel of Fig. 8.3. Several publications have studied the compatibility between DAMA and CoGeNT. In the WIMP mass-cross section plot they do not overlap, but the CoGeNT allowed region is very close to the low mass region of DAMA. Due to the astrophysics uncertainties and the complications involved in extracting the cross section, it is difficult to reach a clear cut conclusion. We refer to two publication for more details, [159] and [160], where relevant references can be found.

The CoGeNT data which favors a light WIMP is in contrast to the majority of theoretical models which favor heavy WIMPs of hundreds of GeV. Similar to

DAMA, CoGeNT is also in tension with other experimental results that have not seen any candidate events. The data from XENON100 [161] have excluded the allowed regions of both DAMA/LIBRA and CoGeNT. There are ways to avoid the direct conflicts among these experiments if non-conventional astrophysical scenarios are assumed.

An improved analysis on more data, which include 1200 days running since Dec. 3, 2009, has been presented in [162]. Preference for annual modulation continues in the bulk counting rate. But the amplitude of the oscillation is 4-7 times larger than expected a standard WIMP. The possibility of a non-Maxwellian local halo velocity distribution can help to reconcile it with other reported observation. More independent analysis of the CoGeNT data can be found in the literature, for example, [163] which concludes that the CoGeNT data show a preference for light dark matter recoils at less than 1σ .

8.1.4 Direct search: CRESST

The CRESST-II (Cryogenic Rare Event Search with Superconduction Thermometers) is a direct detection experiment via WIMP scattered off target nuclei in CaWO_4 crystals, in which nuclear recoils produce both phonons and scintillation signals. The phonon provides measurement of the energy deposition of the interaction. The scintillation light yield, together with the energy deposit, discriminates different interactions, and serves to screen out background events as well. Completing 730 kg days data, the CRESST-II phase 1 [144] found 67 acceptable

⁷⁶⁾keVee is the electron equivalent energy units. It is the energy of an event if it were due to an electron recoil.

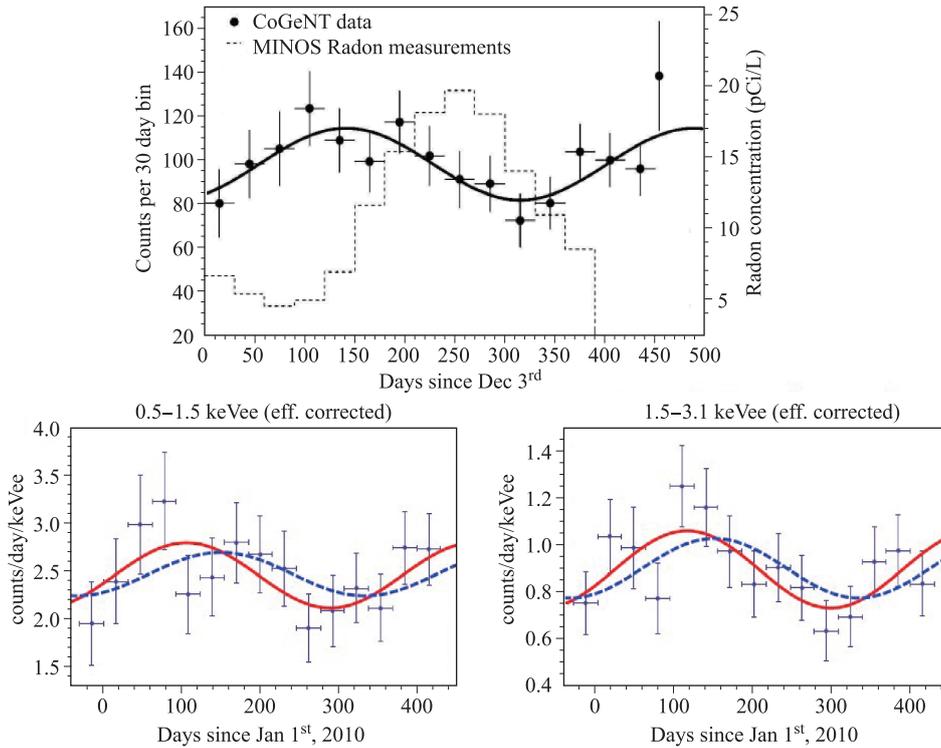


Fig. 8.3 CoGeNT annual modulation: The upper panel is taken from the CoGeNT publication [143]. The lower panel shows the result of an independent analysis performed in [158].

events which fall into two regions marked as M1 (high mass region) and M2 (low mass region) in Fig. 8.4. Details of the experiment can be found in [164] where the collaboration summarized their earlier results together with other existing data.

Updated data based on the upgraded CRESST-II detector were published in 2014 [165] in which new results are announced as summarized in Fig. 8.4: (i) A limit is set on spin-independent WIMP-nucleon scattering which probes a new region of parameter space for WIMP masses below 3 GeV. (ii) An exclusion limit is obtained, shown as the red solid line in Fig. 8.4. (iii) The former allowed low mass region M2 of the phase 1 is clearly excluded.

8.2 Indirect searches

The goal of indirect searches, shared with that of direct searches, is to study what type of particle the dark matter is made of and how it interacts with the standard model particles. An important aspect that differentiates indirect searches from direct searches is that the source of the former can be various astrophysical settings, such as the center of the Galaxy, some nearby dwarf galaxies, etc., not restricted to the solar system. Hence the indirect searches able us to study how the dark matter is distributed elsewhere. There are no compelling evidences up to date that can be called a signal of dark

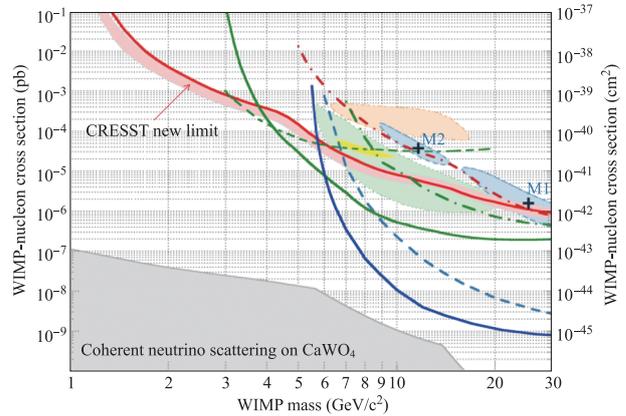


Fig. 8.4 The WIMP parameter space of CRESST-II together with allowed and exclusion regions of other experiments. The figure is taken from [165]. The allowed regions are given by CRESST, DAMA, and CoGeNT. The solid red curve is the new limit of exclusion regions. The former allowed low mass region M2 is clearly excluded. The areas above the various curves are the exclusive regions of the corresponding experiments. We refer to [165] for details.

matter. However, there are several unusual observations which can not be explained by known astrophysics. They are generally called *anomalies* which will come back to them at the end of this section.

Wide scopes of discussions of the current status of

many of indirect search the experiments can be found in lectures of the 2014 SLAC Summer Institute. Summary talks there given by Dan Hooper and Jennifer Siegal-Gaskink and can be found in [26]. A review article which provides many details on indirect searches can be found in [166].

The physical processes and detection mechanism can be seen from the cartoon of Fig. 6.7, to which the possibility of decay of dark matter particle should be included. The detections are classified according to the ordinary particles detected, mainly charged or neutral. The charged particles are focused on positrons and anti-protons for reducing background events, and the neutral particles on photons and neutrinos. The general feature of a signal event is the excess of these charged or neutral normal particles over the cosmic ray background. Hence in indirect searches, it is crucial to know the behavior of the regular cosmic ray.

8.2.1 Indirect searches: Charged particles, positrons and antiproton

The charged particles searched for dark matter signals include positrons and antiprotons which do not exit in the primordial fluxes of the cosmic ray, and generally are secondary in nature. The experimental collaboration involved include the major detectors PAMELA and AMS, and also FermiLAT, HESS, and ATIC, CREAM, BESS, etc.⁷⁷⁾ For WIMP particles annihilation and decay, the flux shapes of the anti-particles e^+ and \bar{p} are determined by the WIMP mass and the annihilation or decay channels. For the annihilation channel the features of the possible excess of positron and anti-proton are illustrated in Fig. 8.5. The required signature is that the excess occurs in a limited energy range of the positron and anti-proton, depending on the mass of the dark matter particle. And eventually the energy spectrum should return to their normal shape and magnitude when the effect of the dark matter particle is exhausted.

The general feature of the charge particles of the cosmic ray as of the middle of last year, a data land dominated by PAMELA, was that there is an increase in the ratio of positron to positron-plus-electron when energy increases, while the anti-proton to proton ratio follows the expectation of the regular cosmic ray. But the errors are quite sizable toward the higher energy end that were available. This different behaviors for leptons and

hadrons lend to some model builders the proposal of a leptophilic dark matter particle which is not overall too satisfactory. See the discussion given in [166]. These earlier data are shown in the upper panel of Fig. 8.6.

The new AMS02 data, published in 2014, have greater accuracy and extend to much higher energy. It is shown in the lower panel of Fig. 8.6. Let us discuss the positron first and then the anti-proton data.

8.2.1.1 Positron data

There are interesting features in the positron data: The positron fraction has been raising since below 7 GeV. It starts a downward trend at 275 ± 32 GeV as marked as zero crossing, where the slope of the curve becomes negative. The central value of the last data point at 500 GeV clearly shows this trend, although the accuracy of the data there and below have also deteriorated markedly. Taking the central values of the data points and comparing them with the plot of the left panel of Fig. 8.5, one is strongly tempted to say that this is the indication of the presence of a heavy particle. However, there are also suggestions that positron event can be explained by conventional astrophysical sources due the presence of nearby pulsars radiation.⁷⁸⁾

Understandably, the new AMS02 data have prompted a burst of theoretical activities in confront theoretical models with the accurate new data available. We mention a few of the works below and apologize for all omissions. A detailed theoretical analysis of the new AMS02 positron data has been carried out in [167]. It has looked into both the dark matter scenarios and the pulsars alternative. The conclusion of the article is the following: Subjected by the stringent requirement of gamma ray and CMB bounds, the fulfilling dark matter species has a mass between 0.5 and 1 TeV. The excess position is explained by the annihilation of the dark matter particle into four leptons through a light scalar or vector mediator. However, the excess may also be explained by one of more of five nearby pulsars. This general conclusion has also arrived at by a earlier paper [168]. Another article [169] has done a studied particularly on the pulsar interpretation of the new AMS02 data and found that any of the four pulsars they selected, or the combination of them can do the job. In the case of the multiple pulsars contribution, the positron fraction will show a particular structure at high energies.

8.2.1.2 Anti-proton data

The anti-proton in cosmic ray is an important messenger for energetic phenomena of astrophysics as an impor-

⁷⁷⁾The acronyms are PAMELA: Payload for Antimatter Matter Exploration and Light-nuclei, AMS: Alpha Magnetic Spectrometer, FermiLAT: Fermi Large Area Telescope, HESS: High Energy Stereoscopic System, ATIC: Advanced Thin Ionization Calorimeter, CREAM: Cosmic Ray Energetics and Mass, BESS: Balloon-Borne Experiment with Superconducting Spectrometer. Note that FermiLAT, formerly called GLAST, is mainly a high energy photon detector, but it is also an electron detector and has contributed to the early positron data.

⁷⁸⁾For non-astronomers, a pulsar is a rotating, magnetized neutron star. As a neutron star forms it rotates rapidly. Electromagnetic radiation and Charged particles are created out and accelerated near the star's magnetic poles.

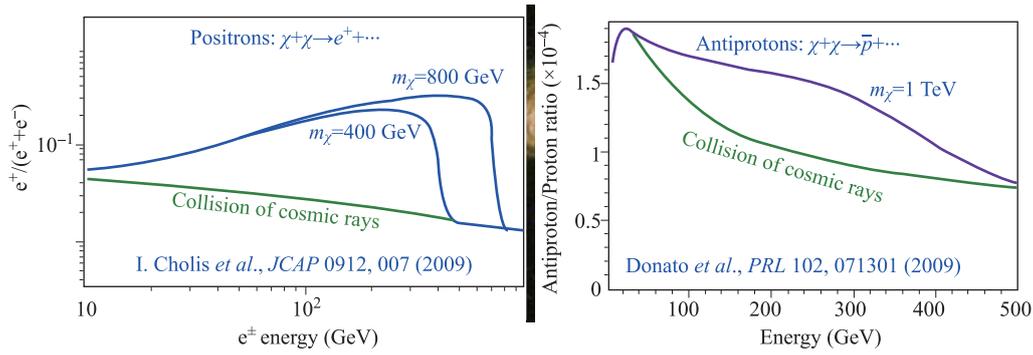


Fig. 8.5 Illustrations of excess positron and anti-proton events over the regular cosmic ray background, due to a WIMP particle. This figure is adopted from the talk by Ting given at the AMS Day at CERN [170].

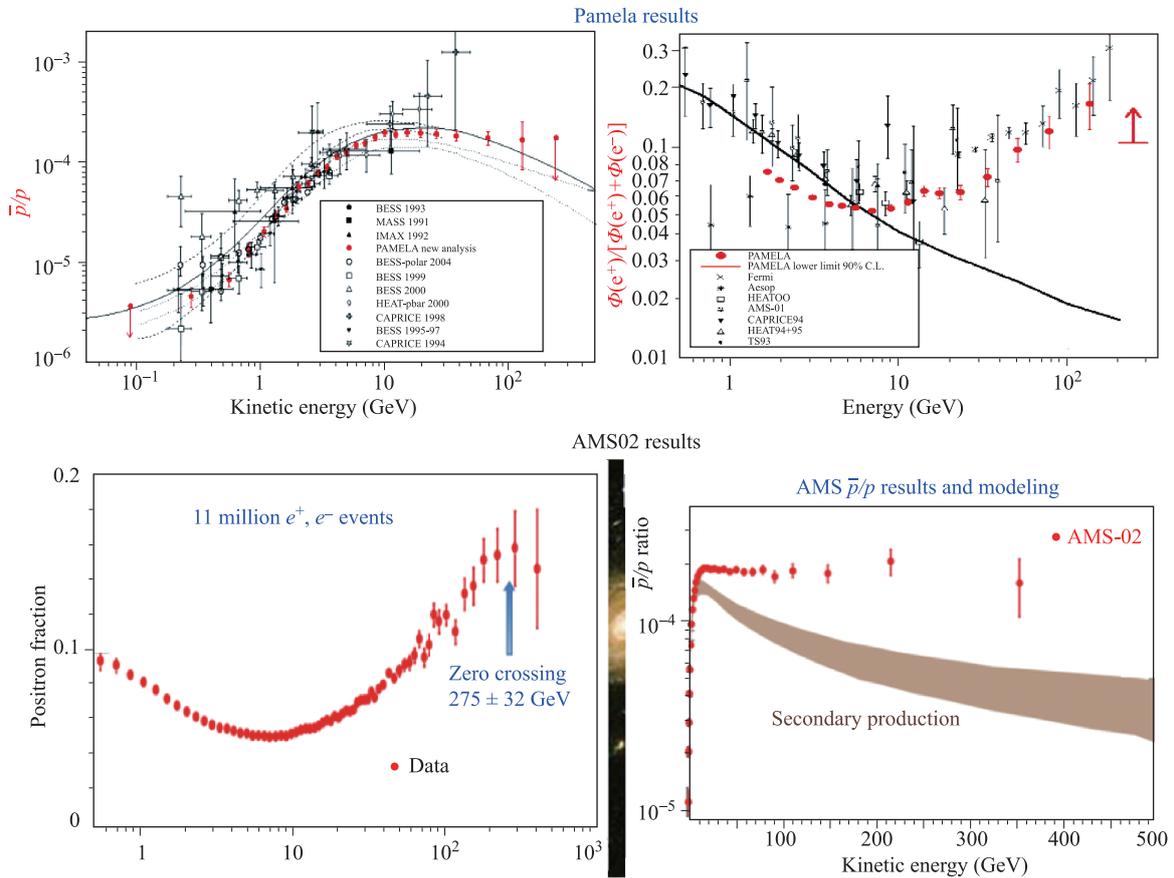


Fig. 8.6 Upper panel: The older data on positron and anti-proton prior to AMS02. The dark curves are the expected trends of the regular cosmic ray. Lower panel: The AMS02 data. These figures is adopted from the talk by Ting given at the AMS Day at CERN [170].

tant diagnostic tool for cosmic ray sources and propagation properties. In indirect searches of dark matter, it is one of the prime channels of the dark matter annihilation, resulting from hadronization of primary quarks, and through standard model processes, of gauge bosons or leptons. Prior to the new AMS02 data, the secondary antiproton, which is produced by the collision of cosmic rays primaries with interstellar medium by ordinary

hadronic processes, is sufficient to account for the bulk of the existing anti-proton data, no exotic processes such as the intervention of a dark matter particle needed. Several papers reanalyzing the situation in light of the new data have appeared. We summarize briefly two of them, which reflect likely the generally conclusions on this matter. The first paper [171] which finds no unambiguous evidence for a significant excess with respect to expecta-

tion. The flat anti-proton to proton ratio implies a flatter energy dependence of the diffusion coefficient. The paper also gives an assessment to allow some components of dark matter annihilation or decay and new stringent constraints have to be imposed. Another article [172] takes the flat anti-proton to proton ratio as an excess and concludes that it calls for a dark matter particle of the order of TeV mass. In particular, the Wino dark matter with mass of about 3 TeV, which has a thermal relic abundance is consistent with the present dark matter abundance, can account for the anti-proton excess given by AMS-02.

8.2.1.3 A tentative conclusion on charged particle

The AMS02 precision data have brought a lot of excitement to the community and established tighter constrain on related physics. The situation of the charged particle excesses is fundamentally unresolved. This reflects the heavy gravity of the challenges in finding dark matter and learning what it is, emphasizing again the experimental nature of the subject. To conclude, let us quote from Ting's talk [170] on what more are needed:

To identify the dark matter signal we need

- i) Measurement of e^+ , e^- , and \bar{p} .
- ii) Precise knowledge of the comic ray fluxes (...).
- iii) Propagation and acceleration (Li, B/C, ...).

8.2.2 Indirect searches: Photons

The photon signature of the dark matter is the anomalous photon events over the cosmic background. Both prompt line photons and diffuse gamma rays can be produced by the annihilation and decay of dark matter. Diffuse gamma rays, having continuous energy distribution, are due to complicated processes of secondary photons of dark matter annihilation into charged particles, which are hadronized to produce, say, high energy neutral pions which, in turn, decay into two photons of high energies. Another source of diffuse gamma rays comes from the internal bremsstrahlung of charged particles produced in the annihilation process. The prompt photons, which are monochromatic, are results of dark matter annihilation directly into a pair of mono-energetic photons, or due to dark matter decays. in the case of annihilation, each photon has the energy of the mass of the dark matter particle, as the annihilation process takes place at very low kinetic energy of the dark matter particle. The prompt photon, with very little cosmic background, is generally considered as a smoking gun for the dark matter [173, 174]. As the propagation of the photon from its production to its detection is largely not interfered, the spatial distribution of the prompt photon follows that of the dark matter particles. Hence there are higher possibility of finding these photons would be where higher

concentration of dark matter is expected, such as the center of the galaxy and the Galactic satellite dwarf galaxies called dwarf spheroidals (dSphs). The Galactic core is expect to have a dark matter density several orders higher than that of the solar system (see Fig. 2.11), while in the dSphs the average dark matter density is a couple of magnitude higher.

Several experiments have been running, including the space telescope FermiLAT, ground-based high energy gamma-ray detector HESS, VERITAS, MAGIC, and air shower ARGO-YBJ, etc.,⁷⁹⁾ while FermiLAT is, up-to-date, the major gamma detectors.

Dark matter signals have been searched in dSphs [175–177], galaxy clusters [178, 179], the Galactic (Milky Way) halo [180], and extragalactic gamma-ray background (EGRB/IGRB) [181, 182]. The experimental data is has been dominated by FermiLAT and no Dark matter signals have been uncovered, although some limits have been derived by some of the data.

To amend the extremely compressed summary given in the preceding paragraph, let us mention how some of the limits are obtained by the FermiLAT collaboration to show the extensive work that has been performed. FermiLAT combined observations on Milky Way's 10 satellite galaxies which are among the most promising regions for diffuse γ -ray signals for the dark matter, to increase the data sample. Joint likelihood analysis of 24 months observational data of 10 dSphs has not uncovered any dark matter signals. But robust upper limits are obtained on three channels of dark matter annihilation cross sections [175]. The results are shown in Fig. 8.7 which are adopted from [175].

Our extremely simplified discussion above can not do justice to these important topics. Fortunately there are excellent reviews on them, see, for example, [166] for detailed considerations of these topics and for ample references.

8.2.2.1 Tantalizing signs

There are tantalizing data samples which hind the possibility of dark matter signals. They are from the observation of the Galactic center and those of the extragalactic gamma-ray background (EGRB).

Observations of the Galactic center

The Milky way galactic center is predicted to contain very high density of dark matter. But since it is also a very complicated astrophysical region, the necessary search for dark matter there is quite challenging.

⁷⁹⁾The acronyms are, FermiLAT: Fermi Large Area Telescope; HESS: High Energy Stereoscopic System; VERITAS: Very Energetic Radiation Imaging Telescope Array System; MAGIC: Major Atmospheric Gamma-Ray Imaging Chrekov; ARGO-YBJ: the Yang-Ba-Jing Cosmic Ray Observatory

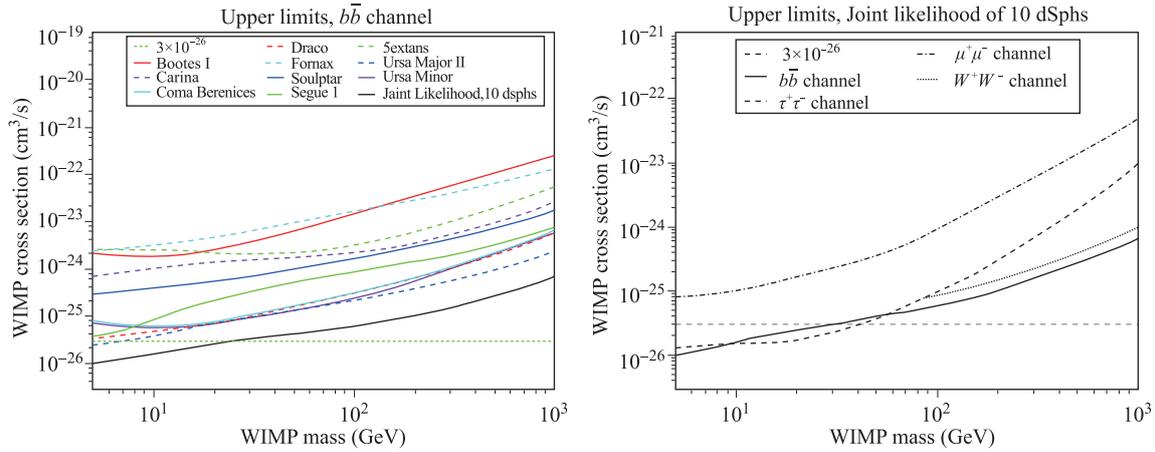


Fig. 8.7 FermiLAT limits drawn from the joint analysis of data on the observation of 10 dSphs [175].

Studying the morphology and spectrum of the gamma ray emission from the Galactic center of three years data from FermiLAT, a group of authors found the evidence of a spatially extended component which occurs at energies between 300 MeV and 10 GeV [183]. And the extended component has been independently confirmed [184] and [185]. According to [183] the extended emission can be accounted for by either the annihilation of dark matters, or by the collision of high energy protons that are accelerated by the supermassive black hole of the Milky Way. When interpreted as dark matter events, the emission spectrum requires a dark matter particle of mass 7–12 GeV if the annihilation is dominated by lepton channels, and 25–45 GeV if annihilation is dominated by hadrons. The authors of [183] have fitted the data using a combination of annihilation channels, and presented explicit fittings in three different scenarios. One of the scenarios in which the dark matter annihilation is dominated by $b\bar{b}$ is shown in Fig. 8.8. The other two scenarios can fit data well.

A recent updated analysis [186] of the relevant FermiLAT data found that it is a compelling case to interpret the excess as an the annihilation of dark matter particle. The data can be well fitted by a dark matter particle of 36-51 GeV mass with an annihilation cross section $v\sigma = (1 - 3) \times 10^{-26} \text{ cm}^3/\text{s}$. This agrees with the earlier analysis as given above.

Observations of Spectral Lines

Monochromatic photon lines are very strong evidence of exotic events, and hence the possibility of dark matter particles. We discuss two separate observations:

- The 130 GeV line
Independent analysis of earlier FermiLAT data, containing 43 months observations, were made by three groups [174, 187], and [188], and found indications

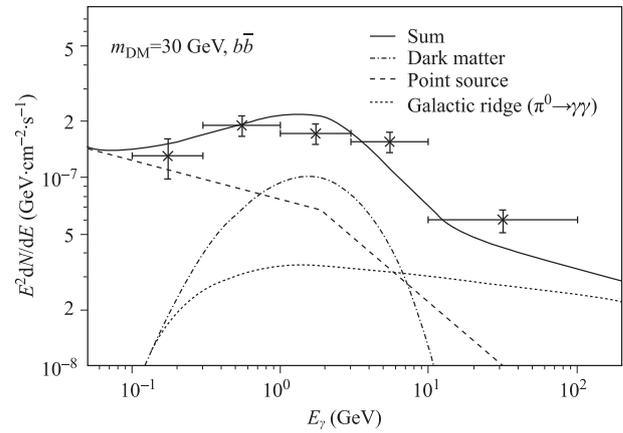


Fig. 8.8 The figure is adopted from [183]. The result of a 30 GeV dark matter particle annihilates into the $b\bar{b}$ channel with an annihilation cross section is $v\sigma \sim 6 \times 10^{-27} \text{ cm}^3/\text{s}$.

of a γ line at $E_\gamma \approx 130 \text{ GeV}$. As stated in [174] the effect is 4.6σ base on the observation of 50 photons. If this is interpreted as a WIMP annihilation into a γ pair, the observation corresponds to a WIMP mass $m_\chi = 129.8 \pm 2.4_{-13}^{+7} \text{ GeV}$ and an average annihilation rate $\langle v\sigma \rangle = (1.27 \pm 0.32_{-0.28}^{+0.18}) \times 10^{-27} \text{ cm}^3/\text{s}$.

Very recently, the FermiLAT collaboration has published their updated search for monochromatic spectral lines from Galactic dark matter in the energy range of 200 MeV to 500 GeV, using 5.8 years data [189]. They found no strong evidence for the existence of a line spectrum. They have also investigated, in particular, the 133 GeV line, based on their new data and new event reconstruction and selection algorithms. They found that the confidence level of the existence of such a monochromatic photon line has significantly lowered from its earlier value, dropping to 0.73σ .

- The 511 keV line

The observation of the Galactic 511 keV emission is one of the fascinating story lines of the development of particle physics and astrophysics. The first observation took place in the early 1970's by balloon-born experiments. It is the first γ -ray line ever identified that originates outside the solar system, from the general direction of the Galactic center. The accepted explanation of this spectral line is the e^+e^- annihilation taking place in the Galactic center. From its flux on Earth of $10^{-3}/(\text{cm}^2 \cdot \text{s})$, together with the distance between Earth and the Galactic center at 8 kpc, it implies an annihilation rate of $\sim 1 \times 10^{43} e^+$ per second. The power released is about $10^4 L_\odot$, where $L_\odot = 2.83 \times 10^{26} \text{ W}$ is the Solar luminosity. Many experimental and theoretical studies have been made, but the main source of the positron has not been identified. Although conventional astrophysical sources such as type Ia supernovae, etc., are generally regarded to be plausible candidate sources, dark matter annihilation has been proposed. For a review of subject we refer to [190].

8.2.3 Indirect searches: Neutrinos

Neutrinos and anti-neutrinos produced in annihilation of DM particles serve as a good signal for its parent particles. Although neutrinos are difficult to detect, they carry valuable information with them. An advantage of the neutrino is its very long mean free path due to its weak interactions. For heavy DM particles one expects to see high energy neutrinos coming from the galactic center, where a concentration of DM is generally ex-

pected. Unlike photons, neutrinos can be detected in a well-controlled environment of underground laboratories, under water, or in ice. The water and ice serve as detection media of charge current interaction of neutrinos to produce charge leptons. The muon final state is especially revealing by its Cherenkov radiation in water and ice and their huge penetration length. An added advantage is the important directional information. When the incoming neutrino energy is high enough, its traveling direction is preserved by the outgoing muon. To alleviate the background problem of cosmic rays, upward going neutrino signal in a detector can take advantage of the whole Earth as its shield. Its background is neutrinos from other sources, such as the atmospheric neutrinos, which form the major background. Another advantage of the neutrino probe is the existence of large detectors for the neutrino physics which can be used to search for neutrinos from DM particles.

Presently the online neutrino detectors include Super-Kamiokande (Super-K), IceCube, and ANTARES. To date there are no signals observed in the neutrino channel yet from Super-K, IceCube, and AMANDA, where AMANDA is the predecessor of IceCube. The absence of high energy neutrinos from the galactic center and from the sun can set useful constraints on the DM particles mass and annihilation cross sections. Some earlier results and analysis can be found in [191, 192], and [193]. For the most recent results from Super-K searching for dark matter in the sun can be found in [194] and the results are shown in Fig. 8.9, which we adopted from [194].

The most recent results from IceCube searching for dark matter from the Galactic center can be found in

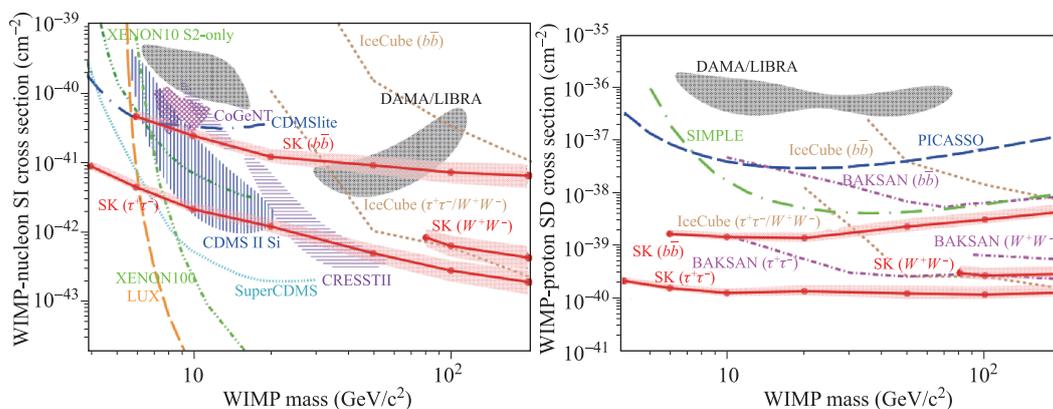


Fig. 8.9 Comparison of the allowed and exclusion limits obtained from direct and indirect experiments. The colored regions are allowed regions of the parameter space from the direct search experiments DAMA, CoGeNT, CDMS, and CRESST. The curves mark the exclusion areas. Recall that Xenon, PICASSO, SIMPLE, and LUX are direct experiments. BAKSAN, together with SK and IceCube are indirect searches. Let us remind ourselves that direct experiments measure the dark matter-nucleon collisions and indirect experiments dark matter annihilations. Hence the indirect searches have to mark the particular channel of final state that the limit is for. **Left:** Spin-independent cases. **Right:** Spin-dependent cases. This figure is adopted from [194], to which we refer for details.

[195]. The plot showing the result is given Fig. 8.10 which is adopted from [195].

8.2.4 Anti-deuteriums

Indirect searches are generally focused on γ -rays, positrons, anti-protons, and neutrinos, as discussed above. However, these possibly copious channels may suffer from more severe background due to uncertainties of common astrophysical processes. The cosmic ray anti-deuteriums, which can appear as secondary particles from WIMP annihilation and which has very little expect background, has been proposed as a promising signal for an indirect search for dark matter [196]. The mechanism of production is the coalescence of an anti-proton and an anti-neutron originated from the annihilation of WIMP dark matter particles. In order for the \bar{p} and \bar{n} to from an \bar{d} , they have to have similar momenta to move in the same direction. This rather stringent phase space requirement implies that the production cross section will be small. Fortunately, according to calculation [196], the flux of the \bar{d} so produced is peaked at low kinetic energies, which is a fraction of a GeV. Since the cosmic ray \bar{d} production is generally from the spallation of high energy cosmic ray protons, there is very little cosmic ray background in this low energy regime. Hence the detection of a low energy anti-deuterium can be considered a smoking gun for the existence of a WIMP dark matter. Recent considerations of the \bar{d} production suggests the possibility of other scenarios, for instance, the coalesce mechanism may be enhance and WIMP originated \bar{d} may also appear at higher energies [197].

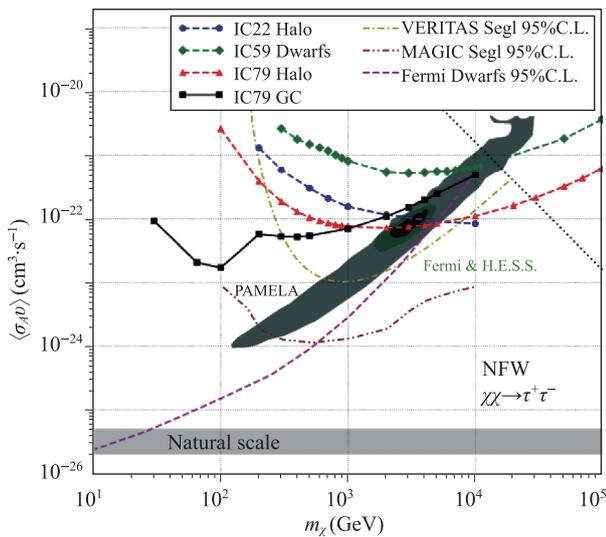


Fig. 8.10 Comparison of 4 IceCube measurements and results of other indirect searches. The grey-shaded region is the dark matter interpretation of the PAMELA positron excess. The figure is adopted from [195] to which we refer for details.

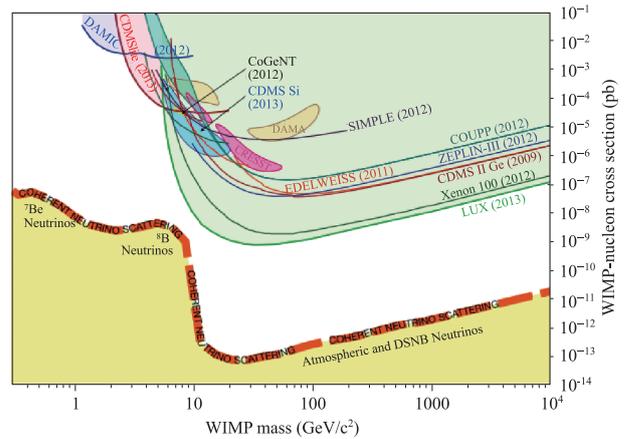


Fig. 8.11 A summary of the present status of experimental searches of dark matter candidates for the case of spin independent scattering. Both allowed and exclusion regions are shown. The upper right shaded area is the excluded region which contain all the areas allowed by DAMA/LIBRA, Co-GeNT, CDMS-Si, and CRESST. Reproduced from Ref. [199].

There are upper bounds for the \bar{d} search given by the BESS experiment. The experiments for further searching for \bar{d} include GASP and AMS-2. We refer to [119] for more discussions and references. An updated review on both theoretical and experimental aspects of cosmic-ray anti-deuteriums search is given in [198].

8.2.5 Searches of other anti-particles

The are searches of other kind of antiparticles for dark matter signal, such as antihelium-4. The experiments include BESS (Balloon-borne Experiment with superconducting Spectrometer) which has carried out antimatter search since 1993. It has collected more than 5 billion cosmic ray events as of the end of 2011 and found no evidence of antihelium, which sets a bound of 10^{-7} for antihelium to helium ratio.

8.3 Summary

8.3.1 Status of experimental searches

8.3.1.1 Summary-direct searches

The present status of direct searches are summarized in Fig. 8.11 for the case of spin-independent cross sections. The figure is adopted from [199]. A similar plot emphasizing the region of exclusion can be found in [65].⁸⁰ We note that for the majority of the mass region, except for WIMP mass less than 8 GeV, LUX provides the strongest limit. For low mass below 8 GeV, Xenon 100,

⁸⁰The plot shown in Fig. 8.18 is given in the 2014 version of the PDG review article 25. *Dark Matter*. A newer plot is available in the September 2015 updated of the PDG review article.

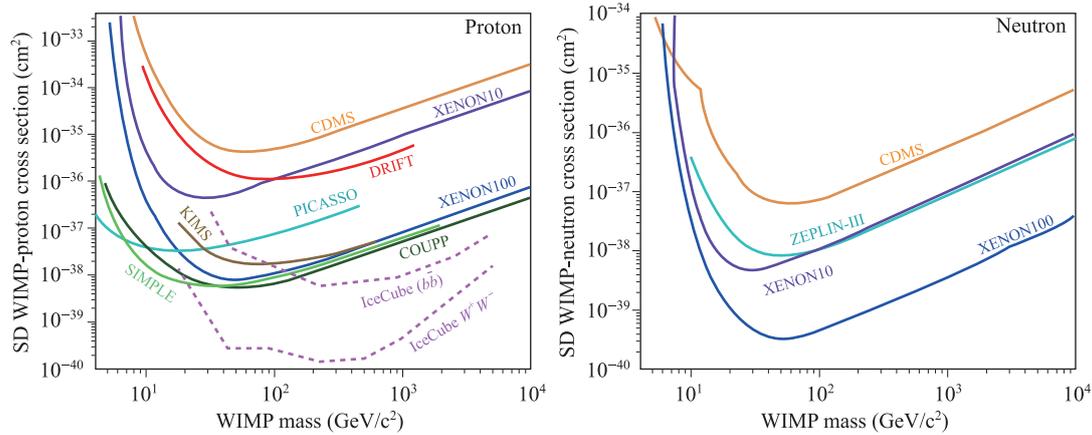


Fig. 8.12 Left: Regions of exclusion for the proton-WIMP scatter cross section. Right: Regions of exclusion for the neutron-WIMP scatter cross section.

SuperCDMS, and DAMIC provide the best constraints. The overall region of exclusion contains all the four allowed parameter spaces: DAMA/LIBRA, CoGeNT, CDMS-Si, and CRESST. The two sets of experiments are in serious tension. The thick orange bars mark the dark matter mass-cross section region where the direct search experiment begins to be sensitive to neutrino scatterings, imposing strong constraints on reach of direct dark matter search experiments in the future.

Bounds of exclusion obtained from various experiments for spin dependent WIMP-nucleon scatterings, are shown in Fig. 8.12. The figure is adopted from Ref. [200]. Note that the indirect search experiment IceCube has provided rather strong exclusion bounds in the case of WIMP scattering against proton.

8.3.1.2 Summary-indirect searches

A very prominent feature presently in indirect searches is the presence of Excesses and anomalies, in addition to various exclusion bounds obtained. We have discussed most of the excesses and anomalies in Section 8.2. We summarize them by showing a summary table of the talk given by Hooper in [26] shown in Fig. 8.13, augmented by the new data from AMS02 discussed in Section 8.2. These anomalous events can be originated from unknown conventional source or from the annihilation or decay of dark matter particles. According to Hooper [26], for a candidate of the dark matter particle, the last one on the above list of anomalies, i.e., the GeV excess from the Galactic Center, is particularly compelling, among the many indirect detection anomalies having appeared over the years, it can be interpreted as a WIMP of 31–40 GeV having an annihilation cross section of the order of $\langle v\sigma \rangle \approx (1.7\text{--}2.3) \times 10^{-26} \text{ cm}^3/\text{s}$, agreeing with the expectation of the WIMP miracle. Future observations on dwarfs, cosmic-ray anti-protons, etc., are needed to unravel the true nature of this inviting event.

Excesses and anomalies (circa 2014)			
Anomaly	Info. content	Astrophysics?	Dark matter?
511 keV	Modest	Semi-plausible (LMXBs)	Non-standard, plausible
Radio Bkg.	Modest	Non-standard ($z \sim 6$)	Non-standard
Positron excess	Modest	Plausible (pulsars)	Non-standard, plausible
130 GeV line	Low significance	NA	Non-standard, plausible
3.5 keV line	Modest	Unknown atomic line?	Fairly standard
GeV excess	High	None apparent	Standard

Caveat: Every entry in this table is highly subjective
 Advice: Learn the details for yourself and make your own evaluations

Fig. 8.13 List of Excesses and anomalies in indirect search experiments as of early 2014 before the release of the new AMS-02 data. To this list we should add the possible excess of anti-protons shown in the data of AMS02 as well as the new FermiLAT data which further weaken the significance of the 130 GeV γ -line.

8.3.2 A glance at the future

The existence of dark matter, as manifested in its gravitational effect, is unequivocal. Almost any dark matter theory tangible has been proposed. The experimental situation presently is also unsettling. A few years ago there were the inklings among physicists that dark matter signals might be discovered soon. But now the reality sets in and people are prepared for the long haul.

Large detectors for direct searches are coming online or in plan for the next few years. They are designed to probe the dark matter mass vs cross section to one or two orders of magnitude better in the elastic scattering of the dark matter particle with nucleons. They include XENON1T, XENONnT, DarkSide, LUX-ZAPLIN (LZ), and DARWIN. The XENON1T [201], which expands from XENON100 at Gran Sasso just coming online, has been inaugurated on November 11, 2015. Its sensitivity is expected to improve that of LUX, which provides the strongest exclusion limits to date, by a fac-

tor of the order of 50 for the WIMP mass $m_\chi > 6$ GeV. Further expansions to XENONnT is planned to take place in 2018. The sensitivity plots of XENON1T and its expansion XENONnT, together with the existing limits are given in Fig. 8.14. A further expansion beyond the XENONnT experiment is the multi-ton DARWIN (DARK matter WImp search with Noble liquids) project at the 20 t scale [202, 203]. The sensitivity of DARWIN is expected to approach the neutrino coherent scattering limit.

The LZ (LUX-Zeplin) of Xe detectors is another next generation (Generation 2) large dark matter experiment, with a sensitivity similar to XENON1T. It has passed the Critical Decision Step 1 and become an official project of the US DOE. The LZ experiment will be housed at the Sanford Underground Research Facility (SURF), located at the former Homestake gold mine, in Lead, South Dakota, USA. The experiment will involve two stages. The first stage operates a 1.5–3 tonne detector while the second stage a 20 tonne detector. Its designed sensitivity can reach 5×10^{-49} cm² for 100 GeV WIMPs. The most recent status of LZ can be found in [204].

The 2013 Snowmass CF1 Summary [205], has provided a roadmap for WIMP direct searches up to the year 2020. Compiled sensitivity reaches on pin independent cross sections as functions of the WIMP mass, for both presently established and future planned, are shown Fig. 8.15 [205]. On future generations of experiments, it includes the projected sensitivities for DarkSide G2, LZ, as well as the neutrino coherent scattering limit. A similar plot of the sensitivity reaches including XENONnT and DARWIN is shown in Fig. 8.16 which, as adopted from [205], appeared in [203] and other publications.

We note that since the report was published near the end of 2013, some of the planned reaches, given in dotted/dashed curves, have already been delivered. For example LUX, represented by the dashed green curve pro-

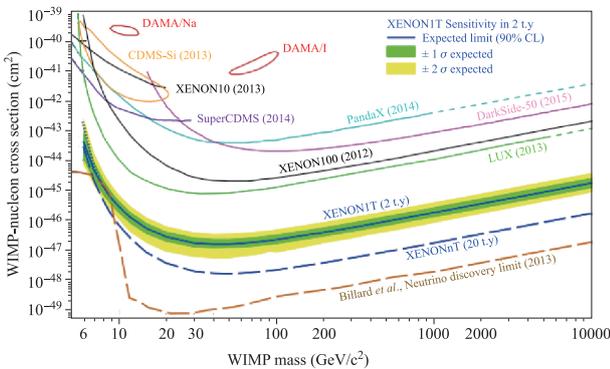


Fig. 8.14 Sensitivity curves of the online XENON1T together the existing results and its future expansions XENONnT. The figure and its caption shown below the figure are from [201].

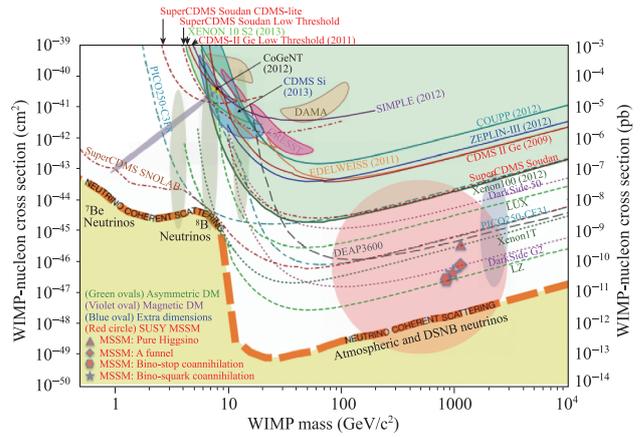


Fig. 8.15 Figure and its caption are from Ref. [205]. The solid curves are the established limits, and the dotted/dashed curves are the planned ones as of the end of 2013. The LUX bound represent by the green dashed curve is now an established bound which is the strongest for WIMP mass greater than 8 GeV.

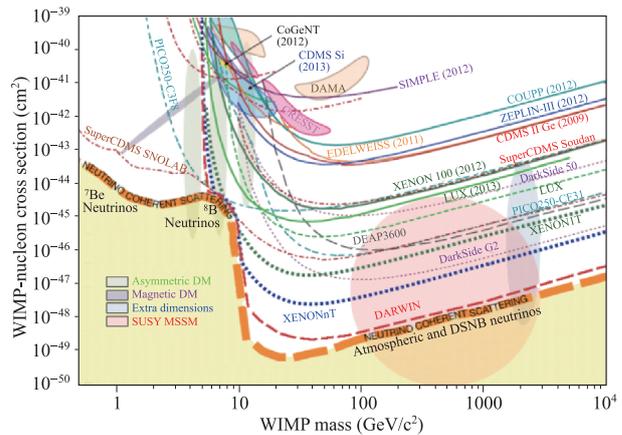


Fig. 8.16 The figure is reproduced from Ref. [203] which, in turn, adopted it from Ref. [205].

vides now the strongest exclusion bounds to date. We should remark that these detectors are capable to probe most of the SUSY parameter space.

The experimental progresses made in both the direct and indirect searches are impressive. The direct detection sensitivity has followed a Moore-like law, doubling the sensitivity every 20 months. This is shown in Fig. 8.17 [206]. The future development is on track in consolidating different efforts of smaller experiments into larger and more versatile ones. A few large experiments, in the traditional high energy physics approach, covering all the important technologies, competing yet collaborating, enabling checks made on one another, may create a concerted efforts to unveil this deep mystery we are facing.

Sensitivity roughly doubled every ~20 months for the past decade

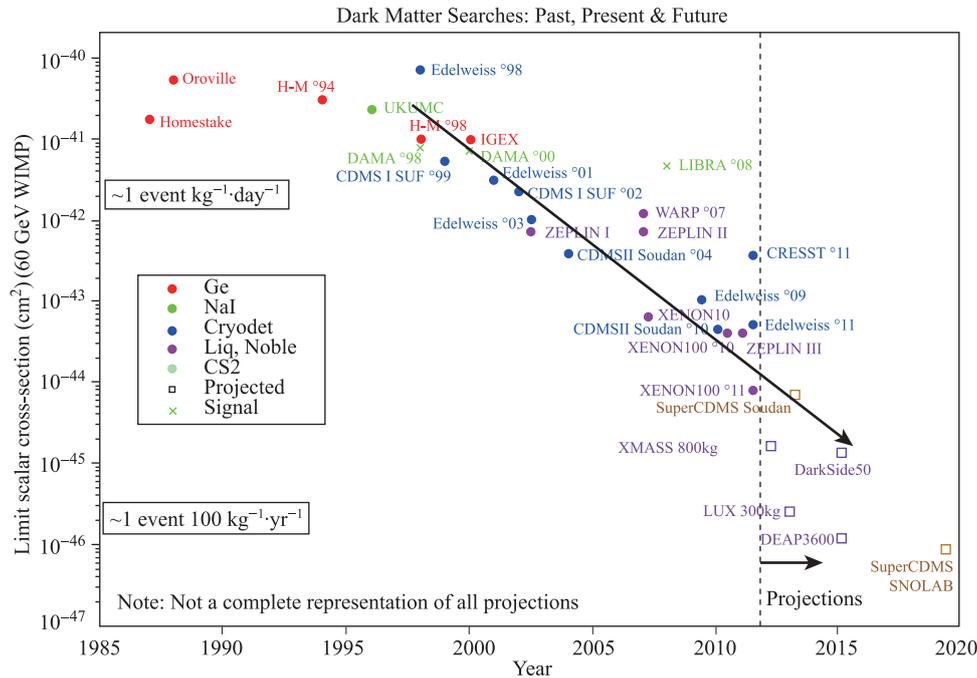


Fig. 8.17 Moore law of the direct detection of dark matter. Roughly the experimental sensitivity doubles every 20 months in the past decade. The figure is reproduced from [206].

8.3.3 Some miscellaneous theoretical remarks

We have not concerned ourselves much with the particle theory aspects of dark matter which is a very large subject by itself. Below we make a few remarks to just touch on some of the current issues, in particular, about supersymmetry. The remarks will be of a tentative nature, likely to be altered when newer data appear from both dark matter search experiments and particle physics experiments carried out at the on-going Run 2, LHC.

8.3.3.1 Supersymmetry

Supersymmetry motivates the dark matter search experiments for WIMPs. The present exclusion region of the SUSY parameter space in the mass vs WIMP-nucleon scattering cross section plot is shown in Fig. 8.18. The lower right area shows the parameter space of some of the SUSY schemes and the region of exclusion by dark matter direct searches. Direct searches have eliminated more and more portion of the SUSY parameter space. The present strongest bounds are provided by LUX. As indicated in Fig. 8.15, future experiments are able to cover the whole SUSY parameter space.

With the operation of LHC, a systematic test of Supersymmetry has begun. The status of the Run 1 on the search for new physics, including SUSY, can be summarized as follows: No statistically significant sign for SUSY particles has been uncovered. Neither the pre-

cision measurements of the B -meson system, nor the search for an extended Higgs sector, has found any sign of new physics. More generally, all precision measurements of Run 1 show good agreement with the standard model.

More specifically let us mention the following: The low value of the mass of the Higgs particle discovered at LHC is seen to disfavor the minimal supersymmetric model. The recent measurement of the rare decay $B_s \rightarrow \mu\bar{\mu}$ branching ratio from LHCb [207] and the updated rare decay $B^- \rightarrow \tau^- \nu$ branching ratio from Belle [208], both of which agree with the standard model predictions, put further pressures on SUSY. However, the actual situation is not yet clear. For example, [209] found that the two highly constrained SUSY models, i.e., the constrained minimal supersymmetric standard model (CMSSM) and the non-universal Higgs model (NUHM) are still surviving but in a reduced parameter space.

At the present time, one cannot say what will happen to SUSY. The LHC's first phase proton running at 7 TeV and then 8 TeV, with the discovery of the Higgs particle, is just the beginning.⁸¹⁾ The next phase of proton running has already begun in 2015 at the upgraded energy of 13.5 TeV. A more vigorous test of SUSY will be executed.

⁸¹⁾For a brief description of this first phase of LHC see the LHC news release on Dec. 17, 2012: <http://press.web.cern.ch/press-releases/2012/12/first-lhc-protons-run-ends-new-milestone>

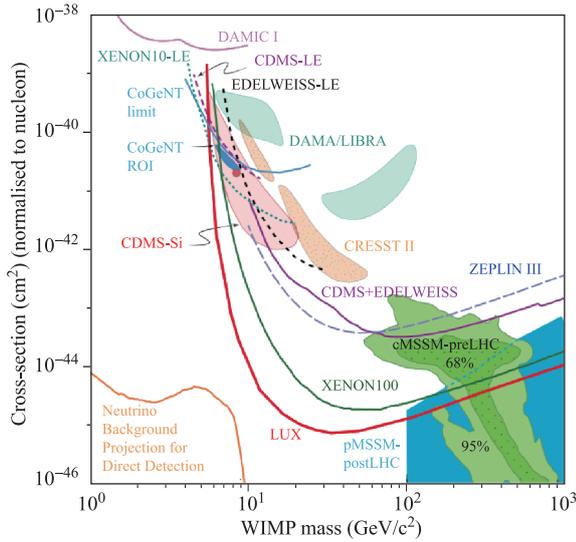


Fig. 8.18 The current exclusion region of the SUSY parameter space. This is essentially the same exclusion region of Fig. 8.11. This figure is reproduced from Ref. [65]. We note that a newer version of this plot exists in the September 2015 updated version of the article. In the newer plot, the CRESS II signal regions are removed because of updated data of the CRESS collaboration. Also, exclusion curves from Dark Side 50, CRESS, etc. are shown. The SUSY exclusion region is deemphasized. Otherwise the newer plot is similar to the present one.

A summary of highlights of LHC can be found in [210]. A more pedagogical presentation of LHC’s results so far, together with lectures on SUSY and dark matter can be found in [211]. A recent review of the status SUSY after Run 1 can be found in [212].

8.3.3.2 Other developments of dark matter models

We focus on the possibility of light dark matter particles, which are particles of the order of several GeV or lighter, not the WIMPs of hundreds of GeV commonly expected in SUSY, and in addition to the sterile neutrino and axion discussed earlier. There are both theoretical and experimental motivations for this class of light dark matter. Experimentally, the null result of the search of SUSY particle at LHC will push the SUSY scale higher and makes it even less accessible experimentally. Also from existing dark matter search experiments, all claims of signals from direct searches, DAMA, CoGeNT, and CRESST, are in the regime of light dark matter particles. Theoretically, much high mass scale of SUSY makes it less attractive. Furthermore, the original work of Lee and Weinberg [213] suggested a lower mass bound of 4 GeV for thermal dark matter relics.⁸²⁾ So low mass dark

⁸²⁾In Lee and Weinberg [213] the lower mass bound is 2 GeV, with the assumption that matter provides the whole critical density. In the present case of dark matter is about one quarter of the critical density, the lower mass bound is raised to 4 GeV.

mass particle is plausible.

The closeness of the baryon and dark matter density, i.e., $\rho_{DM} = 4.5\rho_b$ suggests that there might be a close connection between baryon and dark matter particles, while in the common WIMP scenario, a connection is not necessary and the closeness of the two densities is accidental. The well-known origin of the baryon density clearly illustrate the difference. The baryon density is a result of the baryon asymmetry as a consequence of CP violation. But the WIMP scenario, motivated by hierarchy, does not have such a connection. The asymmetric dark matter (ADM) [214] is to assume that the dark matter and the baryon have a common origin, both are results of the CP violation mechanism. ADM models typically result in similar number density of the baryon and the ADM. This says that the dark matter particle will have a mass of the order of 4 GeV, agreeing with the result of the Lee–Weinberg decoupling mechanism. This kind of light particles will entail different search strategy from the common WIMPs. For a review of the ADM we refer to [215]. A review of light dark matter can be found in [216].

Another class of models is referred to as the dark sector [217], which assumes that the dark matter is a complicate particle sector by itself, not just one species of particles as commonly assumed.

Part II

Related topics in cosmology

In this part we present several topics in cosmology relevant to the discussion of dark matter. Since they are also fundamental as parts of the corner stones which establish the present picture of the universe, it is useful to see how the argument goes for readers who have not acquired the expertise of modern cosmology, such as most students in particles physics. We try to give as much details as we think to be useful or to point out where more details can be found. Hence Part II is lengthy. If you are familiar with cosmology, or not interested in the details, or simply comfortable with statements made in dark matter discussions, these chapters are not for you.

Cosmology has become a precision science, e.g., with the determination of the age of the universe to within 0.4% in 1σ , i.e., 13.81 ± 0.05 Gyr [13], the CMB anisotropy in one part in 10^5 , etc. Many reviews of cosmology can be found in the literature and good textbooks of different levels are also available. In addition to the classics, such as [74, 218–220]¹⁾, etc., a number of newer books have appeared in the new millennium after

¹⁾This book which emphasizes inflation has an arXiv version available online: arXiv: hep-th/0503203.

the discovery of dark energy and other modern developments being made, e.g., [101, 221–223], and more recently [224].²⁾ A collection of entries A–Z, including key ideas, relevant terms, scientists involved, observational projects, etc., can be found in [225], which can serve as a quick reference to various topics in cosmology.

The theoretical framework of cosmology is quite concise as can be seen from these textbooks. The subject is mostly organized into three parts: the fundamental dynamic equations of general relativity and the FLRW metric; applications of thermodynamics when consideration of material is involved for the study of behavior of the cosmos, such as the thermal history of the early universe, etc.; the application of relevant physics laws, perturbation theory on CMB anisotropy and large structure formation. The detailed behavior of the universe is governed by laws of physics, of course. In particular, particle physics and nuclear physics provide the basic physical mechanism for the inner operation of the universe.

The theoretical and computational structures of cosmology have some parallelism to particle physics. The homogeneous isotropic FLRW, which serves the starting point and is referred to as the background universe, does not concern with individual structure of the universe. It can perhaps be thought of as the analogy of a (very complicated) free field theory in particle physics. Then there is the symmetry principle which is the invariance under general 4-dimensional coordinate transformation of the comoving frame. Because of the expansion of the universe, consequences of the invariance of the coordinate transformation become very non-trivial. The analogy in particle physics is the gauge transformation and gauge invariance. Then there is the perturbative calculations for the CMB anisotropy and large structures formation due to growth of small perturbations. Perturbation in particle physics is due to interactions among different fields contained in the theory. In cosmology, perturbations are performed in background universe in the FLRW metric and the energy-momentum tensors. One may visualize the parallelism of these two types of perturbations to imply some form of interactions, if one considers John Wheeler’s famous words: “*Spacetime tells matter how to move; matter tells spacetime how to curve.*” [226].

On the observational side, Hubble’s observation of a linear relation between the redshift and distance of observed galaxies in the late 1920s [227] is the cornerstone in the establishment of the expanding universe, and observations by Zwicky [20] and others, made in the late 1920s and early 1930s, of the flattening of the galaxy rotation curve when the distance of the galaxy increases, as an indication of the presence of dark mat-

ter, are a milestone of our understanding of the cosmos in the early development of cosmology. Prior to the 1960s, most of the information concerning the universe has been obtained from the redshift and distances observation of distant galaxies. In the mid-1960 the discovery [228] of a nearly isotropic background of microwave radiation, known as the cosmic microwave background (CMB) had greatly expanded the scope of astrophysical observations and prompted many subsequent measurements. But for many years the detailed spectrum of this background radiation was not determined. The Planck spectrum of the microwave background radiation as an ideal blackbody radiation was settled by the Cosmic Background Explorer Satellite (COBE) experiment in the early 1990s [229] and ushered in the era of precision cosmology.³⁾ The discovery of the dark energy in the late 1990s [231, 232] put cosmology on a new trajectory in the understanding of the cosmo.

As already mentioned, we work in the natural units which is extended to include the temperature sector by taking the Boltzmann constant $k_B (= 8.6173 \times 10^{-5} \text{ eV/K}) \equiv 1$. So there is again only one unit and all fundamental units are related as given in [74].⁴⁾

9 An introductory cosmology for pedestrians

9.1 The Einstein field equation and FLRW metric

9.1.1 Review of some generality

We begin by making a quick review to delineate some the general known facts so as to define our notation. The metric of a space is denote by the covariant and contravariant symmetric metric tensors, $g_{\mu\nu}$ and $g^{\mu\nu}$ respectively, which are related by

$$\begin{aligned} g^{\mu\lambda} g_{\lambda\nu} &= \delta_{\nu}^{\mu}, \\ g_{\mu\nu} &= g_{\nu\mu}, \quad g^{\mu\nu} = g^{\nu\mu} \end{aligned} \quad (9.1)$$

Under a general coordinate transformation $x \rightarrow x'$, then a tensor is transformed as

$$\begin{aligned} V_{\nu\dots}^{\mu\dots}(x) &\rightarrow V_{\nu'\dots}^{\mu'\dots}(x') \\ &= \frac{\partial x'^{\mu}}{\partial x^{\lambda}} \cdots \frac{\partial x^{\sigma}}{\partial x'^{\nu}} \cdots V_{\sigma\dots}^{\lambda\dots}(x), \end{aligned} \quad (9.2)$$

where $V_{\nu'\dots}^{\mu'\dots}(x')$ means a functional form change of $V_{\nu\dots}^{\mu\dots}$ and being evaluated at the transformed space-time point x' .

³⁾For a detailed first look of the CMB, see Section 2 [221]. For an interesting historical description of the discovery of Penzias and Wilson, see Weinberg’s classic science writing *The First Three Minutes* [230].

⁴⁾See pp 499–500, Appendix A of [74] and we reproduce the relations among the different units in Table C.2, Section C.1.

²⁾This is far from a complete list of the available books on cosmology. A search of books available on Amazon, <http://www.amazon.com>, will reveal many other books.

Next we define the covariant derivative of the above tensor

$$V_{\nu \dots ; \lambda}^{\mu \dots}(x) = \frac{\partial}{\partial x^\lambda} V_{\nu \dots}^{\mu \dots}(x) + \Gamma_{\lambda \sigma}^\mu V_{\nu \dots}^{\sigma \dots}(x) + \dots - \Gamma_{\nu \lambda}^\sigma V_{\sigma \dots}^{\mu \dots}(x) - \dots, \quad (9.3)$$

where $\Gamma_{\mu\nu}^\lambda$, called the affine connection, is given by

$$\Gamma_{\mu\nu}^\lambda = \frac{1}{2} g^{\lambda\sigma} \left(\frac{\partial g_{\mu\sigma}}{\partial x^\nu} + \frac{\partial g_{\sigma\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma} \right). \quad (9.4)$$

It is symmetric with respect to the lower indices μ and ν because of the symmetry of $g_{\mu\nu}$.

The Einstein field equation which is a fundamental equation of cosmology is given by

$$R_{\mu\nu} = -8\pi G_N S_{\mu\nu} = -8\pi G_N \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T_\lambda^\lambda \right) - \Lambda g_{\mu\nu}, \quad (9.5)$$

where Λ is the cosmological constant and G_N the gravitational constant. The other symbols involved are explained below. The Einstein field equation can also be written in the form

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R_\lambda^\lambda = -8\pi G_N T_{\mu\nu} + \Lambda g_{\mu\nu}, \quad (9.6)$$

where we have used Eq. (9.5) to express T_λ^λ in terms of R_λ^λ .

$R_{\mu\nu}$ is the Ricci tensor⁵⁾ expressed in terms of the affine connection of Eq. (9.4),

$$R_{\mu\nu} = \frac{\partial \Gamma_{\lambda\mu}^\lambda}{\partial x^\nu} - \frac{\partial \Gamma_{\mu\nu}^\lambda}{\partial x^\lambda} + \Gamma_{\mu\sigma}^\lambda \Gamma_{\lambda\nu}^\sigma - \Gamma_{\mu\nu}^\lambda \Gamma_{\lambda\sigma}^\sigma. \quad (9.7)$$

The Ricci tensor is symmetric with respect to the interchange of its two indices, although not explicitly. It can be simply demonstrated that

$$R_{\mu\nu} = R_{\nu\mu}. \quad (9.8)$$

$T^{\mu\nu}$ is the *energy-momentum tensor*, also known as the *stress-energy tensor*, which provides the information on the energy density, momentum density, pressure, and stress. It is given by, for a perfect fluid of frictionless continuum,⁶⁾

$$T^{\mu\nu} = \mathcal{P} g^{\mu\nu} + (\mathcal{P} + \rho) u^\mu u^\nu, \quad (9.9)$$

where u^μ is a velocity four-vector. The conservation of energy and momentum is defined by the vanishing covariant derivation of the energy-momentum tensor,

$$T_{;\nu}^{\mu\nu} = \frac{\partial}{\partial x^\nu} T^{\mu\nu} + \Gamma_{\nu\lambda}^\mu T^{\lambda\nu} + \Gamma_{\nu\lambda}^\nu T^{\mu\lambda} = 0. \quad (9.10)$$

⁵⁾For the Ricci tensor we follow the definition given in [221]. The definition of the Ricci tensor given in some works is the negative of the expression Eq. (9.7)

⁶⁾Let us repeat that we use the natural units $c = 1$. We note the proper, engineering dimensionality of the energy density and pressure are the same: $[\text{mass}][\text{time}]^{-2}[\text{length}]^{-1}$. Hence the mass density has the same engineering dimension as pressure divided by c^2 , i.e., \mathcal{P}/c^2 .

Here we note that a constant term in $T^{\mu\nu}$ proportional to the metric tensor $g^{\mu\nu}$ makes no contribution to the energy-momentum conservation equation due to the vanishing of the covariant derivative of the metric tensor, as an identity,

$$g_{;\lambda}^{\mu\nu} = \frac{\partial}{\partial x^\lambda} g^{\mu\nu} + \Gamma_{\lambda\sigma}^\mu g^{\sigma\nu} + \Gamma_{\lambda\sigma}^\nu g^{\mu\sigma} \equiv 0. \quad (9.11)$$

Because of this we can absorb the cosmological constant term of the Einstein equation (9.5) in the energy-momentum tensor

$$T^{\mu\nu} \rightarrow T^{\mu\nu} - g^{\mu\nu} \tilde{\Lambda}, \quad \tilde{\Lambda} \equiv \frac{\Lambda}{8\pi G_N}. \quad (9.12)$$

Then the term $g_{\mu\nu} \Lambda$ on the right-handed side of Eq. (9.5) can be dropped.⁷⁾

To conclude this subsection let us comment on the fact that the energy momentum conservation law Eq. (9.10) is not an independent relationship, but derivable from in the Einstein field equation Eq. (9.6). This follows from the differential Bianchi identity, rewritten in the form of contravariant tensor,

$$\left(R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} g^{\lambda\sigma} R_{\lambda\sigma} \right)_{;\mu} = 0. \quad (9.15)$$

Taking the covariant derivative of the Einstein field equation in the form of Eq. (9.6) and using the fact the covariant derivative of the metric tensor vanishes $g_{;\mu}^{\mu\nu} = 0$, we obtain the energy-momentum conservation relation Eq. (9.10). A demonstration of the Bianchi identity can be found in [218].⁸⁾ Although the energy-momentum conservation relation is not independent of the Einstein field equation, it has the important relationship that the conservation of energy and momentum hold separately in isolated systems of cosmic constituents which do not

⁷⁾Let us mention that in many works the Einstein field equation is written in a form which looks different from that of Eq. (9.5), i.e.,

$$R'_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R' + \Lambda g_{\mu\nu} = 8\pi G_N T_{\mu\nu}, \quad (9.13)$$

where $R' = g^{\mu\nu} R'_{\mu\nu}$. The sign difference between Eqs. (9.5) and (9.13) can be explained by the sign difference in defining the Ricci tensor, $R'_{\mu\nu} = -R_{\mu\nu}$, as already noted in Fn. 5). The other difference can be explained by contracting the μ and ν indices of the Einstein equation to express the trace of the Ricci tensor $R = R_\mu^\mu$ in terms of the trace of the energy-momentum tensor T_μ^μ and Λ ,

$$8\pi G_N T_\lambda^\lambda = R + 4\Lambda = -R' + 4\Lambda.$$

Then Eq. (9.13) is just Eq. (9.5). Furthermore, the Einstein equation is often written in a more concise form by defining $G_{\mu\nu} = R'_{\mu\nu} - (1/2)g_{\mu\nu} R'$,

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G_N T_{\mu\nu}. \quad (9.14)$$

⁸⁾See, pp 146–147, [218].

exchange energy and momentum with other system. An example is the system of neutrinos, after the decoupling of neutrinos at the cosmic temperature of 10^{10} K or 1 MeV, the energy and momentum are conserved within the neutrinos themselves.

9.1.2 The FLRW metric

The expanding universe represented by the fact that galaxies fly away from one another comes from Hubble's systematic observation of redshifts of galaxies. To a good approximation, at a scale of 300 million light-years larger, the universe looks isotropic and homogeneous to all comoving observers. We will define the *comoving observer* below. An isotropic and homogeneous expanding universe can be described uniquely by the Friedmann–Lemaître–Robertson–Walker (FLRW) model [218, 221].^{9),10)} In this model the covariant metric tensor $g_{\mu\nu}$ in the Cartesian system of the comoving frame is given by

$$\begin{aligned} g_{00} &= -1, \\ g_{0j} &= g_{j0} = 0, \\ g_{ij} &= g_{ji} = a^2(t) \left(\delta_{ij} + \kappa \frac{x^i x^j}{1 - \kappa r^2} \right), \end{aligned} \quad (9.16)$$

where $i, j = 1, 2, 3$ and $\mathbf{r} = (x^1, x^2, x^3)$.¹¹⁾ $a(t)$ is the FLRW scale factor describing an expanding universe and κ is the curvature constant.

For completeness we also give the contravariant metric tensor $g^{\mu\nu}$. First we note that regarding to $g_{\mu\nu}$ as a 4×4 matrix denote by $(g_{\mu\nu})$, we have the determinant of the matrix to be

$$|(g_{\mu\nu})| = -\frac{(a^2(t))^3}{1 - \kappa r^2}. \quad (9.17)$$

Then the elements of the inverse of $(g_{\mu\nu})$, denoted by $(g^{\mu\nu})$, give the matrix of the contravariant metric tensor. They are related by $g_{\mu\nu} = g_{\mu\lambda} g_{\nu\sigma} g^{\lambda\sigma}$. The elements of $(g^{\mu\nu})$ can be worked out straightforwardly as

$$\begin{aligned} g^{00} &= -1, \\ g^{0j} &= g^{j0} = 0, \\ g^{jk} &= g^{kj} = \frac{1}{a^2(t)} (\delta_{jk} - \kappa x^j x^k), \\ |(g^{\mu\nu})| &= -\frac{1 - \kappa r^2}{(a^2(t))^3}. \end{aligned} \quad (9.18)$$

It is easy to verify the relationship Eq. (9.1). The expressions of components of the metric tensor are needed in working out the explicit forms of the Einstein field equations and the equations of the energy-momentum conservation.

The FLRW line element is given by, written in the spherical coordinates in the \mathbf{r} space,¹²⁾

$$\begin{aligned} ds^2 &\equiv -g_{\mu\nu} dx^\mu dx^\nu \\ &= dt^2 - a^2(t) \left((d\mathbf{r})^2 + \kappa \frac{(\mathbf{r} \cdot d\mathbf{r})^2}{1 - \kappa r^2} \right), \end{aligned} \quad (9.19)$$

Using the identities

$$\begin{aligned} (d\mathbf{r})^2 &= dr^2 + r^2 d\Omega^2, \\ \mathbf{r} \cdot d\mathbf{r} &= r dr, \\ d\Omega^2 &= d^2\theta + \sin^2\theta d^2\phi, \end{aligned} \quad (9.20)$$

we obtain the usual spherical form of the FLRW line element

$$ds^2 = dt^2 - a^2(t) \left(\frac{dr^2}{1 - \kappa r^2} + r^2 d\Omega^2 \right). \quad (9.21)$$

This line element defines a simpler metric of the space under consideration,

$$\begin{aligned} g_{00} &= -1, \\ g_{rr} &= \frac{a^2(t)}{1 - \kappa r^2}, \\ g_{\theta\theta} &= a^2(t)r^2, \\ g_{\phi\phi} &= a^2(t)r^2 \sin^2\theta, \end{aligned} \quad (9.22)$$

and all non-diagonal elements are zero.

The coordinates x^j , $j = 1, 2, 3$ or equivalently r , θ , and ϕ are the comoving coordinates. Individual objects, such as galaxies, which participate in the expansion of the universe, follow geodesics of constant values of x^j or r , θ , and ϕ . $\sqrt{ds^2}$ is called the *proper time*, and t the cosmological proper time or generally *cosmic time* for simplicity. Hence a comoving observer is defined as a stationary observer in a comoving frame which moves with the Hubble flow. The cosmological proper time is the time measured by a comoving observer.

Since κ always appear together with $x^j x^k$, we can scale the comoving coordinates x^j without affecting the metric, Eq. (9.16), so that κ takes one of the following three values:

$$\kappa = \begin{cases} +1 & \text{spherical, positive curvature, closed;} \\ -1 & \text{hyperspherical, negative curvature, open;} \\ 0 & \text{Euclidean, no curvature, flat.} \end{cases} \quad (9.23)$$

¹²⁾If we use the conformal time η , $dt = a(t)d\eta$, the dt^2 term is replaced by $a^2(t)d\eta^2$. The line element looks more symmetric but then g_{00} and g^{00} become more complicate and have to be replaced respectively by $a^2(t)$ and $a^{-2}(t)$.

⁹⁾§1.1 pp 2–4, [221]; §13.5 pp 395–403, [218].

¹⁰⁾ Accessible quick readings of the attributions to FLRW can be found in, e.g., <http://en.wikipedia.org/wiki/Alexander_Friedmann> for Friedmann, <http://en.wikipedia.org/wiki/Georges_Lemaître> for Lemaître, <http://en.wikipedia.org/wiki/Howard_Percy_Robertson> for Robertson, and <http://en.wikipedia.org/wiki/Arthur_Geoffrey_Walker> for Walker.

¹¹⁾ It is noted that most metric convention used in quantum field theory is $g_{00} = 1$ for the special relativity, while many books in gravity uses $g_{00} = -1$ for general relativity.

We note that in rescaling κ and x^j , we have to rescale the FLRW scale factor $a(t)$, then the FLRW line element (9.19) is not altered. This says that although $a(t)$ is a physical quantity, its magnitude is not physical. As we will see later all measurable quantities, if related to $a(t)$, depend on the ratio of $a(t)$'s at different times. A convenient normalization of $a(t)$ is $a(t_0) = 1$, where t_0 is the time of the present epoch. We will not make this choice by keeping $a(t)$ arbitrary.

In the Euclidean case of vanishing curvature constant $\kappa = 0$, the relevant expressions can be simplified:

$$ds^2 = dt^2 - a^2(t)dr^2, \tag{9.24}$$

and the metric tensors in the Cartesian system have the simple diagonal form

$$(g_{\mu\nu})|_{\kappa=0} = \begin{pmatrix} -1 & & & \\ & a^2(t) & & \\ & & a^2(t) & \\ & & & a^2(t) \end{pmatrix},$$

$$(g^{\mu\nu})|_{\kappa=0} = \begin{pmatrix} -1 & & & \\ & \frac{1}{a^2(t)} & & \\ & & \frac{1}{a^2(t)} & \\ & & & \frac{1}{a^2(t)} \end{pmatrix}, \tag{9.25}$$

which indicates a flat space with an expanding spatial structure uniformly in all directions.

We conclude that in the FLRW model the local system, i.e., the comoving frame, simply looks like a uniformly expanding sphere in time with the radius scaled by the time-dependent factor $a(t)$. Or it looks like, in a given sphere of a radius r , the equivalent time slows down by a factor $a(t)$. Such a system which seems to be simple has very interesting features in the presence of matter and energy when coupled with the Einstein dynamics of general relativity. This is the case we will explore below and we restrict ourselves to the Euclidean case of vanishing curvature constant $\kappa = 0$. This flat, Euclidean system, which simplifies greatly calculations presented below and in subsequent chapters, agrees with the first order observational data.

9.1.3 The Einstein equation in the FLRW metric

The cosmological dynamics is governed by three sets of equations [218],¹³⁾ the Einstein equation Eq. (9.6), the energy-momentum conservation Eq. (9.10), and the equation of state which will be given in some details later. We explore each of the three equations below. To exhibit the full display of the cosmological dynamics, laws of physics, in particular thermodynamics, particle

¹³⁾See p. 473 [218].

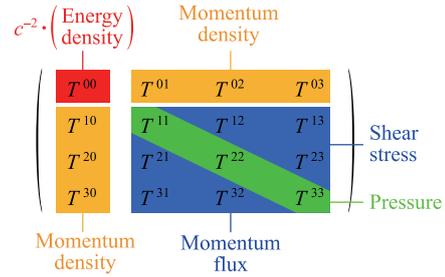


Fig. 9.1 Physical quantities represented by the energy-momentum tensor.

physics, and nuclear physics are used. Thermodynamics treats average properties of systems of large number of particles in the ideal situation of a dilute gas system. Particle and nuclear physics provide the laws of interactions among individual fundamental particles and the formation of the baryonic matter.

9.1.3.1 Energy and momentum conservation

Let us first consider the energy-momentum conservation. The energy and momentum conservation is defined in terms of the energy-momentum tensor $T_{\mu\nu}$ which appears in the Einstein field equation, It is well-known in classical physics. e.g., classical electrodynamics. The physics involved can be described in a cartoon shown in Fig. 9.1:¹⁴⁾

In the comoving frame, for a given species of particle, the isotropy and homogeneity requires that the proper energy density $\rho(t)$ and pressure density $\mathcal{P}(t)$ appearing in the energy-momentum tensor in Eq. (9.9) are function of the time only [221].¹⁵⁾ The velocity four-vector is given by

$$u^0 = -u_0 = 1, \quad u^j = u_j = 0. \tag{9.26}$$

From Eqs. (9.9) and (9.26), the components of the energy-momentum tensor in the forms of $T^{\mu\nu}$, $T^\mu_\nu = g_{\lambda\nu}T^{\mu\lambda}$, and $T_{\mu\nu} = g_{\mu\lambda}g_{\sigma\nu}T^{\lambda\sigma}$, and for the case of vanishing curvature constant $\kappa = 0$ and vanishing shear stress, can be readily calculated. They are diagonal and the non-vanishing components are

$$T^{00} = \rho(t), \quad T^0_0 = -\rho(t), \quad T_{00} = \rho(t),$$

$$T^{jk} = \delta_{jk}a^{-2}(t)\mathcal{P}(t), \quad T^j_k = \delta^j_k\mathcal{P}(t), \quad T_{jk} = \delta_{jk}a^2(t)\mathcal{P}(t),$$

$$T^\lambda_\lambda = -\rho(t) + 3\mathcal{P}(t), \tag{9.27}$$

The affine connections are also simple in the case of vanishing curvature constant and its non-vanishing components are¹⁶⁾

$$\Gamma^0_{jk} = a\dot{a}\delta_{jk},$$

$$\Gamma^j_{0k} = \frac{\dot{a}}{a}\delta_{jk}, \tag{9.28}$$

¹⁴⁾The figure is adopted from Wikipedia entitled *Stress-energy tensor* [233].

¹⁵⁾See, p. 8 [221].

¹⁶⁾The affine connections and the Ricci tensor in the spherical system with finite curvature constant $\kappa \neq 0$ can be found in [234].

where, as usual, the dot “.” over a symbol denotes the cosmic time derivative.

We can now write down the explicit form of the energy-momentum conservation law given in Eq. (9.10). Using the expressions in Eqs. (9.27) and (9.28), we have the only non-vanishing component to be

$$T^0{}_{;\nu} = \frac{\partial}{\partial x^0} T^{00} + \Gamma_{jk}^0 T^{jk} + \Gamma_{0j}^j T^{00} = 0. \quad (9.29)$$

This gives the energy-momentum conservation constraining equation

$$\dot{\rho}(t) + 3\frac{\dot{a}(t)}{a(t)}[\rho(t) + \mathcal{P}(t)] = 0, \quad (9.30)$$

which can also be rewritten as

$$\dot{\mathcal{P}}(t) - \frac{1}{a^3} \frac{d}{dt} [a^3(\rho(t) + \mathcal{P}(t))] = 0. \quad (9.31)$$

It should be noted that the simplicity of the energy-momentum conservation equation comes from two simplifications of the cosmic model that we are working with. One is the Euclidean nature of the space-time so that the metric tensor $g_{\mu\nu}$ has no spatial coordinate dependence and is diagonal, and the other is the homogeneity and isotropy so that the energy and pressure densities are everywhere the same and depend on time only.

9.1.3.2 The Einstein equation

We now work out the Einstein field equation Eq. (9.5), which is the dynamic equation governing the expansion of the universe. First we work out the components of the Ricci tensor Eq. (9.7):

$$\begin{aligned} R_{00} &= \frac{\partial}{\partial x^0} \Gamma_{j0}^j + \Gamma_{0k}^j \Gamma_{j0}^k = 3\frac{\ddot{a}}{a}, \\ R_{0j} &= R_{j0} = 0, \\ R_{jk} &= -\frac{\partial}{\partial x^0} \Gamma_{jk}^0 + \Gamma_{j\ell}^0 \Gamma_{0k}^\ell + \Gamma_{j0}^\ell \Gamma_{\ell k}^0 - \Gamma_{jk}^0 \Gamma_{0\ell}^\ell \\ &= -(2\dot{a}^2 + a\ddot{a})\delta_{jk}. \end{aligned} \quad (9.32)$$

The components of $S_{\mu\nu}$ in the right-handed side of the Einstein field equation Eq. (9.5) are

$$\begin{aligned} S_{00} &= T_{00} - \frac{1}{2}g_{00}T^\mu{}_\mu + g_{00}\frac{\Lambda}{8\pi G_N} \\ &= \frac{1}{2}(\rho + 3\mathcal{P}) - \frac{\Lambda}{8\pi G_N}, \\ S_{0j} &= S_{j0} = 0, \\ S_{jk} &= T_{jk} - \frac{1}{2}g_{jk}T^\mu{}_\mu + g_{jk}\frac{\Lambda}{8\pi G_N} \\ &= a^2 \left[\frac{1}{2}(\rho - \mathcal{P}) + \frac{\Lambda}{8\pi G_N} \right] \delta_{jk}. \end{aligned} \quad (9.33)$$

The structure of these results can be understood from the symmetry properties of homogeneity and isotropy: R_{00} and S_{00} are scalars, so they are allowed to take non-vanishing values. R_{0j} and S_{0j} are 3-dimensional spatial

vectors, they have to vanish. Finally, R_{jk} and S_{jk} are tensors of rank 2, so they have to be proportional to δ_{jk} .

From the above two sets of expressions, the Einstein field equation Eq. (9.5) gives us two independent equations. The $(\mu\nu) = (00)$ component gives

$$\frac{\ddot{a}}{a} = -\frac{4\pi G_N}{3}(\rho + 3\mathcal{P}) + \frac{\Lambda}{3}. \quad (9.34)$$

The $(\mu\nu) = (jk)$ components which are proportional to δ_{jk} give just one equation:

$$\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} = 4\pi G_N(\rho - \mathcal{P}) + \Lambda. \quad (9.35)$$

9.1.3.3 Hubble expansion rate

Substituting the \ddot{a}/a term of Eq. (9.35) by its expression given in (9.34), we obtain the expression for the Hubble expansion rate

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G_N}{3}\rho + \frac{\Lambda}{3} = \frac{8\pi G_N}{3}(\rho + \rho_\Lambda), \quad (9.36)$$

where

$$\rho_\Lambda = \frac{\Lambda}{8\pi G_N}. \quad (9.37)$$

which is the energy density due to the cosmological constant. We can define the total energy density as

$$\rho_{\text{total}} \equiv \rho + \rho_\Lambda = \rho_R + \rho_M + \rho_\Lambda, \quad (9.38)$$

where ρ_R and ρ_M are respectively the energy densities of the radiation and matter. The Hubble expansion rate is simply given by the total energy density

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G_N}{3}\rho_{\text{total}}. \quad (9.39)$$

From Eqs. (9.34) and (9.35) we also obtain the expression for the pressure:

$$\begin{aligned} \frac{2\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 &= -8\pi G_N \mathcal{P} + \Lambda \\ &\equiv -8\pi G_N (\mathcal{P} + \mathcal{P}_\Lambda), \\ &\equiv -8\pi G_N \mathcal{P}_{\text{total}} \end{aligned} \quad (9.40)$$

where

$$\mathcal{P}_\Lambda \equiv -\frac{\Lambda}{8\pi G_N}, \quad (9.41)$$

and a total pressure is defined by

$$\mathcal{P}_{\text{total}} \equiv \mathcal{P} + \mathcal{P}_\Lambda. \quad (9.42)$$

Here we make some remarks:

- In the case of non-vanishing curvature constant, the Hubble expansion rate Eq. (9.36) is modified by adding to the right-handed side by the term $-\kappa/a^2$.
- The energy-momentum conservation equation, Eq. (9.30) is not independent of the Einstein field equation and can be derived by differentiating the Hubble expansion rate equation Eq. (9.36) by time and eliminating the acceleration \ddot{a} term by the acceleration equation Eq. (9.34).

9.1.3.4 The equation of state

Each thermodynamic system has its characteristics which are expressed in terms of relationships among its variables. An elementary example is the ideal gas equation of a dilute gas system which relates the pressure, volume, with the temperature. In cosmology the equation of state of the perfect fluid is used.¹⁷⁾ The pressure and energy densities of a particle are related by the equation of state

$$\mathcal{P}_j = w_j \rho_j. \tag{9.43}$$

where j denotes the different component of the system in question: radiation, matter, and vacuum and w_j is independent of time.

The equations of state of several different types of particles and fields are given in [235] to which we refer for further discussions.¹⁸⁾ We reproduce them below in Table 9.1. We will verify some of the equation of states below when we discuss the thermodynamics of the universe. It should be noted that the pressure density of a non-relativistic particles does not exactly vanish but it is much smaller than the energy density and hence generally it is set as zero. See Eq. (9.128) below. Note that the equation of state for the cosmological constant is already indicated in Eq. (9.54), i.e., $w_\Lambda = -1$.

9.1.3.5 Some implications

We study below a few of the implications of the above equations which allow us to obtain some gross, yet characteristic features of the present model of cosmology.

Because of the linearity property of the time variation equations of the energy and pressure density, Eqs. (9.30) and (9.31) hold for individual particle species j ,

$$\begin{aligned} \dot{\rho}_j &= -3(\rho_j + \mathcal{P}_j) \frac{\dot{a}}{a} \\ \dot{\mathcal{P}}_j &= \frac{1}{a^3} \frac{d}{dt} (a^3(\rho_j + \mathcal{P}_j)). \end{aligned} \tag{9.44}$$

Summing over all particle and energy components, above

relations hold also for the overall mass and pressure densities, which can further be related to the Hubble expansion rate:

$$\begin{aligned} \dot{\rho}_{\text{total}} &= -3(\rho_{\text{total}} + \mathcal{P}_{\text{total}}) \frac{\dot{a}}{a} = \frac{3}{4\pi G_N} H \dot{H}, \\ \dot{\mathcal{P}}_{\text{total}} &= \frac{1}{a^3} \frac{d}{dt} [a^3(\rho_{\text{total}} + \mathcal{P}_{\text{total}})] \\ &= -\frac{1}{4\pi G_N} (\ddot{H} + 3H\dot{H}). \end{aligned} \tag{9.45}$$

Since the time variation of the FLRW scale factor is related to the total energy density as given by Eq. (9.36), the time behavior of the energy and pressure densities of a particle j depend also on other coexisting particle species. From the first equation in Eq. (9.44) and the equation of state Eq. (9.43) we obtain the behavior of the various forms of energy density as a function of the scale factor,

$$\rho_j \sim a^{-3(1+w_j)} \tag{9.46}$$

We can rewrite the acceleration and velocity equations Eqs. (9.35) and (9.34) as

$$\begin{aligned} \frac{\ddot{a}}{a} + 2 \left(\frac{\dot{a}}{a} \right)^2 &= 4\pi G_N (\rho_{\text{total}} - \mathcal{P}_{\text{total}}) = 3H^2 + \dot{H}, \\ \frac{\ddot{a}}{a} - \left(\frac{\dot{H}}{H} \right)^2 &= -4\pi G_N (\rho_{\text{total}} + \mathcal{P}_{\text{total}}) = \dot{H}, \\ \frac{\ddot{a}}{a} &= -\frac{4\pi G_N}{3} (\rho_{\text{total}} + 3\mathcal{P}_{\text{total}}) = H^2 + \dot{H}, \end{aligned} \tag{9.47}$$

which will be used in the discussion of gauge transformation in the cosmological perturbation theory. We see from the last of the above relations that matter and radiation decelerate the expansion of the universe, while the cosmological constant, or the vacuum energy, accelerate it. Hence if one form of energy density dominates, the Hubble expansion rate in the form of acceleration can be summarized as follows:

$$\ddot{a} \begin{cases} < 0 & \text{for } \rho_{\text{total}} + 3\mathcal{P}_{\text{total}} > 0 & \text{expansion slowing down, matter \& radiation dominance;} \\ = 0 & \text{for } \rho_{\text{total}} + 3\mathcal{P}_{\text{total}} = 0 & \text{expansion in a constant rate;} \\ > 0 & \text{for } \rho_{\text{total}} + 3\mathcal{P}_{\text{total}} < 0 & \text{expansion speeding up, vacuum energy dominance.} \end{cases} \tag{9.48}$$

We can obtain the time dependence of the scale factor

a when a given type of energy density, radiation, matter, or cosmological constant, dominates the Hubble expansion. Substituting Eq. (9.46) into Eq. (9.39) and taking the positive root, we can integrate over the time to obtain

$$a \sim t^{\frac{2}{3(1+w_j)}} \tag{9.49}$$

¹⁷⁾Roughly speaking, a perfect fluid can be characterized by its energy and pressure density in the rest frame. A perfect fluid has no shear stresses, or viscosity, or heat conduction. Its pressure is isotropic. And its energy-momentum tensor is diagonal which is the case Eq. (9.9) together with Eq. (9.26).

¹⁸⁾Reference [235], pp 22–23.

Table 9.1 The equation of state, the ratio of pressure density to energy density \mathcal{P}_j/ρ_j .

w_j	Particle species j
1	Free massless scalar fields
$\frac{1}{3}$	Radiation and relativistic matter
0	Dust, i.e., non-relativistic matter
$-\frac{1}{3}$	Curvature energy
$-\frac{2}{3}$	Domain walls
-1	Cosmological constant

which is valid, of course for a limited range of time during the age of the universe. During the cosmic evolution, the Hubble expansion may be dominated by different types of energy densities at different epochs. So the expansion rate will change as a function of time.

- For vacuum dominance, $1 + w_\Lambda = 0$, in which case the above result is not valid, we have a constant energy density which leads to a constant Hubble expansion rate $H = \sqrt{\frac{\Lambda}{3}}$. Then, from the Hubble expansion rate Eq. (9.39) we have

$$a|_{\text{vd}} \sim e^{Ht}. \tag{9.50}$$

One can check that this is consistent with the acceleration equation, Eq. (9.47), which gives $\ddot{a}/a = (8\pi G_N/3)\rho_\Lambda$. This exponential expansion is referred to as a de Sitter space. It is relevant to the inflationary cosmology.

- For a radiation dominated epoch, $w = 1/3$, then $\ddot{a}/a = -(8\pi G_N/3)\rho_R$, the expansion will slow down.

$$a|_{\text{rd}} \sim t^{1/2}, \quad H|_{\text{rd}} = \frac{\dot{a}}{a} = \frac{1}{2t}. \tag{9.51}$$

- For matter dominated, $w = 0$, $\ddot{a}/a = -(4\pi G_N/3)\rho_M$, the expansion also slows down,

$$a|_{\text{md}} \sim t^{2/3}, \quad H|_{\text{md}} = \frac{\dot{a}}{a} = \frac{2}{3t}. \tag{9.52}$$

We summarize in Table 9.2 the relationship among the three types of energy densities, the scale factor, the time variable, and the Hubble expansion rate. In the sixth column T stands for the temperature of the radiation heat bath of the universe, necessarily involving the photon. We shall discuss the temperature in more detail below in the next section, Section 9.2, on thermodynamics. We include the temperature here to show its relationship with the other parameters. These variables are together

Table 9.2 Relationship among the energy density, scale factor, and time variable. T is the temperature at the corresponding time or scale factor.

Dominance	a	ρ_R	ρ_M	ρ_Λ	$T^\text{①}$	H	$p^\text{②}$
		a^{-4}	a^{-3}	Constant	a^{-1}		a^{-1}
Radiation	$t^{\frac{1}{2}}$	t^{-2}	$t^{-\frac{3}{2}}$	Constant	$t^{-\frac{1}{2}}$	$1/(2t)$	
Matter	$t^{\frac{2}{3}}$	$t^{-\frac{8}{3}}$	t^{-2}	Constant	$t^{-\frac{2}{3}}$	$2/(3t)$	
Vacuum	e^{Ht}	e^{-4Ht}	e^{-3Ht}	Constant	e^{-Ht}	$\sqrt{\Lambda/3}$	

①Based on entropy conservation to be discussed in the topic of entropy below.

②The momentum of a particle is always scaled as the inverse of the scale parameter as the universe expands, $p \sim a^{-1}$. See, [221]: Eq. (1.1.23) and the discussion on p. 109.

forming the central parameter sets for the description of the evolution of the universe.

Let us briefly digress on the table. Firstly, we see that as the universe expands the densities decrease as definite powers of the scale parameter, for matter as $1/a^3$, for radiation as $1/a^4$, and the vacuum energy density remaining constant. Because $\rho_R/\rho_M \sim 1/a$, the matter density will overwhelm that of the radiation at some point in the expansion process of the universe even if the latter is much higher than the latter to begin with, which is the case of our universe. Secondly, as a consequence of the relationship between energy densities and the scale parameter, the Friedman model also determines the time behavior of the scale parameter and hence the time behaviors of energy densities. This fact is important in the chronological account of the evolution of the universe.

Redefining the energy and pressure densities

We note that the cosmological constant can be incorporated into the expression of the energy-momentum tensor, not as a separate term as given in the Einstein equation of Eq. (9.5), by redefining the energy and pressure densities in Eq. (9.9)

$$\begin{aligned} \rho &\rightarrow \rho_{\text{total}} = \rho_M + \rho_R + \rho_\Lambda, \\ \mathcal{P} &\rightarrow \mathcal{P}_{\text{total}} = \mathcal{P}_R + \mathcal{P}_\Lambda, \end{aligned} \tag{9.53}$$

where $\mathcal{P}_M = 0$ for non-relativistic matter components. Then all the above expressions containing the cosmological constant Λ hold true without explicitly separating out the cosmological constant Λ . We also note that energy and pressure densities due to the cosmological constant are related by

$$\mathcal{P}_\Lambda = -\rho_\Lambda. \tag{9.54}$$

From now on we shall drop the subscript ‘‘total’’ and use ρ instead of ρ_{total} to denote the total energy density,

and likewise for the total pressure, including the vacuum, radiation, and matter terms. In the most general case they may also contain the equivalent terms due to the curvature constant, as we will comment on later.

9.1.4 Cosmological redshift

Let us consider the propagation of a light signal emitted from some galaxy and traveling to us. Because of homogeneity and isotropy we can take our own location to be the origin of the coordinate system and orient the space coordinates in any direction that is convenient. Since the light travels in a null path, and in the present case, with a constant θ and ϕ . Then we have from Eq. (9.21), with the possibility of a curvature constant taking into account in the present discussion, in order to show that the result is independent of the curvature constant,

$$d^2s = 0 = dt^2 - \frac{a^2(t)}{1 - \kappa r^2} dr^2, \quad \frac{dt}{a(t)} = -\frac{dr}{\sqrt{1 - \kappa r^2}}. \tag{9.55}$$

The negative sign on the right-handed side of the second expression is due the fact that we have to take the negative solution in calculating $\sqrt{dt^2}$, for a photon to propagate from a radial location at r towards the origin where the observer is located in the comoving frame. Note the interesting features of the second equation: The left-handed side is a function of t and the right-handed side is a function of r .

The derivation of the cosmological redshift, which is straightforward from Eq. (9.55), is given below, following, for example, [218]¹⁹⁾. Let us consider a characteristic point of a light wave, say the crest, which leaves a galaxy located at the radial coordinate r_1 at time t_1 , and arrives at an Earth detector at $r = 0$ and t_0 , where t_0 is the time of the present epoch, which serves as the reference time. Then we have

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_0^{r_1} \frac{dr}{\sqrt{1 - \kappa r^2}} = \frac{1}{\sqrt{\kappa}} \arcsin(\sqrt{\kappa}r_1) = \begin{cases} \arcsin r_1 & \kappa = 1, \\ r_1 & \kappa = 0, \\ \operatorname{arcsinh} r_1 & \kappa = -1. \end{cases} \tag{9.56}$$

Then we consider the next light wave crest which leaves the galaxy again at r_1 but at time $t_1 + \delta t_1$, and arrive at the detector at $t_0 + \delta t_0$. For the second light wave we obtain an expression similar to Eq. (9.56) with but the integration limits of the left-handed side being replaced by $t_1 + \delta t_1$ and $t_0 + \delta t_0$, while the right-handed side is unchanged. Hence we have

$$\int_{t_1 + \delta t_1}^{t_0 + \delta t_0} \frac{dt}{a(t)} = \int_{t_1}^{t_0} \frac{dt}{a(t)}. \tag{9.57}$$

The above integrations can be rewritten as

$$\int_{t_E}^{t_0 + \delta t_0} \frac{dt}{a(t)} = \int_{t_1}^{t_1 + \delta t_1} \frac{dt}{a(t)}. \tag{9.58}$$

For small δt_1 and δt_0 we have²⁰⁾

$$\frac{\delta t_0}{a(t_0)} = \frac{\delta t_1}{a(t_1)}. \tag{9.59}$$

In the natural units these time intervals are just the respectively wavelengths of their wave propagations at the emission and observation points in the comoving frame, $\delta t_1 = \lambda_1$ and $\delta t_0 = \lambda_0$, we can write Eq. (9.59) to obtain the relationship between the scale factors and wavelength/frequencies at r_1 and $r = 0$, and define the redshift z_1 :

$$\frac{\delta t_0}{\delta t_1} = \frac{a(t_0)}{a(t_1)} = \frac{\lambda_0}{\lambda_1} = \frac{\nu_1}{\nu_0} \equiv 1 + z. \tag{9.60}$$

Defining $a_0 \equiv a(t_0)$ and $a_1 \equiv a(t_1)$, we can express the scale factor as a function the redshift,

$$a_1 = \frac{a_0}{1 + z_1}. \tag{9.61}$$

Since a_0 is a constant we have the important relation

$$\frac{da}{a} = -\frac{dz}{1 + z}. \tag{9.62}$$

Let us note that the special relativistic flat space Doppler effect is given by

$$1 + z_{Dop} = \gamma(1 + v_{//}) = 1 + v_{//} + v^2 + \dots, \tag{9.63}$$

where v is the magnitude of the velocity and $v_{//}$ is the component of the velocity along the line of sight. With the Hubble law, the cosmological frequency shift finds a natural explanation in term of the Doppler effect for small r_1 . To the first order in r_1 , we have²¹⁾, from Eqs. (9.61) and (9.56),

$$z_1 \sim (t_0 - t_1)H(t_0) \sim \frac{d}{dt}(r_1 a(t))|_{t_0} \rightarrow v_{//}, \tag{9.64}$$

which agrees with the above expression of the Doppler effect to the first order in velocity. However, for distant galaxies, light frequencies are also affected by the effect of gravitation. It is not useful or even correct to characterize the frequency shift by the effect of special relativity only. We can also interpret the result as follows: δt_1 is the time interval between two photons sent from the source and δt_0 is therefore the time interval for

²⁰⁾Being the periods of their respective wave propagations, δt_1 and δt_0 are of the order of 10^{-15} for the visible light. Hence unless for very very long wave length Eq. (9.59) is justified.

²¹⁾See p. 417 [218].

¹⁹⁾See pp 415–418 [218].

an observer receiving the photons. From

$$\delta t_0 = (1 + z_1)\delta t_1, \tag{9.65}$$

we see that the receiving time interval is stretched by a factor $1 + z_1$. This result will be used in the discussion of the luminosity distance later Section 13.

9.1.5 Energy components in the early universe as functions of the redshift

From now on we will modify our notation slightly. The overall energy and pressure density will carry no sub-

scripts, dropping the superscript *total*,

$$\rho_{total} \rightarrow \rho = \rho_\Lambda + \rho_R + \rho_M \tag{9.66}$$

and densities of specific types are given a definite subscript as indicated above.

We can now express all densities, starting from some earlier times which will be discussed later, by their respective present values together with the corresponding redshift, using the behavior of the densities in the scale parameter as given in Table 9.2:

$$\rho_j = \begin{cases} \rho_R(z) = \left(\frac{a_0}{a(z)}\right)^4 \rho_{R0} = (1+z)^4 \rho_{R0}, & \text{radiation plus relativistic particles,} \\ \rho_M(z) = \left(\frac{a_0}{a(z)}\right)^3 \rho_{M0} = (1+z)^3 \rho_{M0}, & \text{matter, all nonrelativistic particles,} \\ \rho_\Lambda(z) = \rho_\Lambda, & \text{cosmological constant,} \end{cases} \tag{9.67}$$

where ρ_{R0} , ρ_{M0} , and ρ_Λ are respectively the present densities of radiation, matter, and vacuum. Knowing the present density of a given form, we can obtain the information of the densities at early times when the redshift at that time is known, although the detailed information of the densities is a complicated process to obtain.

We can rewrite the Hubble's law Eq. (9.39) as

$$\begin{aligned} H^2(z) &= \left(\frac{\dot{a}(z)}{a(z)}\right)^2 = \frac{8\pi G_N}{3} \rho(z) \\ &= \frac{8\pi G_N}{3} [(1+z)^4 \rho_{R0} + (1+z)^3 \rho_{M0} + \rho_\Lambda] \\ &= H_0^2 [(1+z)^4 \Omega_R + (1+z)^3 \Omega_M + \Omega_\Lambda], \end{aligned} \tag{9.68}$$

where H_0 is the presently observed value of the Hubble parameter which can be found in [13], and

$$\Omega_j \equiv \frac{\rho_{j0}}{\rho_c} = \begin{cases} \Omega_R, & \text{radiation,} \\ \Omega_M, & \text{matter,} \\ \Omega_\Lambda, & \text{for cosmological constant,} \end{cases} \tag{9.69}$$

and

$$\rho_c \equiv \frac{3H_0^2}{8\pi G_N}, \tag{9.70}$$

which is called the critical density of collapse, or simply the *critical density* of the universe.

The meaning of the critical density as defined above is clear. If the total energy density of the present epoch is greater than ρ_c , or $\Omega_\Lambda + \Omega_R + \Omega_M > 1$, then we have a non-vanishing curvature $\kappa = +1$ and the universe will

eventually collapse to a big crunch. The condition for a flat universe is that the total energy density, including the dark energy, radiation, and non-relativistic matter, saturates the Hubble expansion rate without the need of a curvature term. Defined for the present epoch, this is $\Omega_\Lambda + \Omega_R + \Omega_M = 1$. These will become clear in the subsection below when we consider the effect of the curvature. Numerically the critical density is given by [13] as

$$\begin{aligned} \rho_c &= 1.87847(19) \times 10^{-26} h^2 \text{ kg/m}^3 \\ &= 10.5375(11) h^2 \text{ GeV/m}^3. \end{aligned} \tag{9.71}$$

Taking $h = 0.673$, we have mnemonically the various densities in terms of number of protons per m^3 : the critical density is equivalent to 4 and three quarters protons, the dark energy about 3 and a quarter, the cold dark matter about 1 and a quarter, and the baryon matter about a quarter.

9.1.6 Effect of the curvature term and the flat space

So far we have ignored the curvature term entirely. Let us consider the effect of the curvature in some detail. In the presence of a finite curvature, $\kappa \neq 0$, we have to modify the density formula that enters the expressions of the Hubble expansion rate Eq. (9.39) and therefore the total energy density (9.66) by adding an effective curvature density term,

$$\begin{aligned} \rho_\kappa(z) &= -\frac{3}{8\pi G_N} \frac{\kappa c^2}{a^2(z)} = (1+z)^2 \rho_{\kappa 0}, \\ \rho_{\kappa 0} &= -\frac{3}{8\pi G_N} \frac{\kappa c^2}{a_0^2}, \end{aligned} \tag{9.72}$$

where $\rho_{\kappa 0}$ is the curvature density of the present epoch. Then the total density entering the Hubble expansion rate becomes

$$\rho \rightarrow (z)\rho_{\Lambda}(z) + \rho_R(z) + \rho_M(z) + \rho_{\kappa}(z). \quad (9.73)$$

The corresponding ratio of the present curvature density to the critical density is

$$\Omega_{\kappa} = \frac{\rho_{\kappa 0}}{\rho_c}. \quad (9.74)$$

Now we can write as an identity

$$\Omega_{\Lambda} + \Omega_R + \Omega_M + \Omega_{\kappa} = 1, \quad (9.75)$$

which allows us to determine Ω_{κ} and therefore κ given the energy densities Ω_{Λ} , Ω_R , and Ω_M .

We can write the density fraction of the curvature at an epoch of arbitrary redshift z as

$$\begin{aligned} \Omega_{\kappa}(z) &\equiv \frac{\rho_{\kappa}(z)}{\rho(z)} = \frac{8\pi G_N}{3H^2(z)}\rho_{\kappa}(z) \\ &= \frac{H_0^2}{H^2(z)}(1+z)^2\Omega_{\kappa}. \end{aligned} \quad (9.76)$$

We note that in the very early universe radiation energy dominates. Then $H^2(z) \sim (1+z)^4$ which says

$$\Omega_{\kappa}(z) \sim (1+z)^{-2}\Omega_{\kappa}. \quad (9.77)$$

This gives rise to the well-known flatness problem. For any finite value of the curvature parameter, i.e., any value of Ω_{κ} at the present epoch, the curvature fraction to the effective total energy density is negligibly small at the early universe of $z \gg 1$. Running the argument in the reversed direction with $\Omega_{\kappa} \sim (1+z)^2\Omega_{\kappa}(z)$, we have a z^2 growth in the curvature density fraction. From the fact that the observed matter-energy density today ρ_0 is close to the critical density ρ_c , this requires a very small curvature density fraction in the early universe. This gives raise to a fine tuning problem unless $\kappa = 0$. Furthermore, a finite curvature constant allows the determination of the scale factor at the present time, a_0 , which is unphysical, from the second expression of Eq. (9.72).

As stated in the 2013 results of the Planck Collaboration [236] that there is “no compelling evidence for deviation from the base Λ CDM model” and the limit of the value²²⁾ of Ω_{κ} is quite small. We will continue to ignore the curvature term in discussions below.

²²⁾See §6 and Table 10, [236].

9.1.7 Age of the universe

We follow [222]²³⁾ to use Eq. (9.68) to estimate the age of the universe. Let us consider

$$\begin{aligned} H &= \frac{d}{dt} \ln \frac{a(t)}{a_0} \\ &= \frac{d}{dt} \ln \frac{1}{1+z} \\ &= -\frac{1}{1+z} \frac{dz}{dt}, \end{aligned} \quad (9.78)$$

which together with Eq. (9.68) gives

$$dt = -\left[H_0(1+z)\sqrt{(1+z)^4\Omega_R + (1+z)^3\Omega_M + \Omega_{\Lambda}} \right]^{-1} dz. \quad (9.79)$$

Integrating the time variable from some early time t_e corresponding to the redshift z_e to the present time t_0 corresponding to the redshift 0, we have

$$\begin{aligned} \tau_e &= t_0 - t_e \\ &= \frac{1}{H_0} \int_0^{z_e} \frac{dz}{(1+z)\sqrt{(1+z)^4\Omega_R + (1+z)^3\Omega_M + \Omega_{\Lambda}}}. \end{aligned} \quad (9.80)$$

The age of the universe is obtained by putting $t_e = 0$ and $z_e \rightarrow \infty$,

$$\tau_U = \frac{1}{H_0} \int_0^{\infty} \frac{dz}{(1+z)\sqrt{(1+z)^4\Omega_R + (1+z)^3\Omega_M + \Omega_{\Lambda}}}. \quad (9.81)$$

The most recent astrophysical data [13] give $H_0^{-1} = 9.777752h^{-1}$ Gyr, $h = 0.673$, $\Omega_M = 0.315$, $\Omega_{\Lambda} = 0.685$, $\Omega_{\gamma} = 4.8 \times 10^{-5}$, and $0.9 \times 10^{-3} < \Omega_{\nu} < 0.048$, where only the central values are quoted here. The errors in most of the quantities are small, except for the uncertainty in the neutrino energy density. However because the smallness of its value, the contribution of the radiation are negligible in our estimate of the universe age. There are uncertainties in this simplified estimate, such as abrupt changes in the radiation and matter densities during the early evolution of the universe. But these complications happened in the early history of the universe, their effect will be small to the determination of the age of the universe. Another uncertainty we will ignore is the contribution of the neutrino. With the approximation as stated the numerical integration of Eq. (9.81) gives

$$\tau_U = 1.381 \times 10^{10} \text{ yr}, \quad (9.82)$$

which agrees with the age of the universe given in [13]: $\tau_U^{exp} = 13.81 \pm 0.05$ Gyr.

²³⁾See, p. 76, [222].

9.1.8 A brief comment on measuring the Hubble expansion rate

Since the Hubble expansion rate is such a fundamental quantity, let us discuss briefly how it is measured. Recall the basic Hubble expansion law. Consider a point which is at a comoving radial distance r from the origin at the cosmic time t . The actual distance is

$$d(t) = a(t)r. \tag{9.83}$$

Then the point will move away with a velocity

$$v(t) = \dot{d}(t) = \dot{a}(t)r = H(t)d(t). \tag{9.84}$$

Since $H(t)$ is independent of r , a function of time only, this velocity-distance relationship is universal for all points at a given time. This relation suggests a seemingly simple way to measure $H_0 = H(t_0)$, where t_0 is the cosmic time of the present epoch: measuring both the distance and receding velocity of an astronomic object, such as a galaxy, and dividing the latter by the former. Then check the universal value of H_0 by repeating the measurements of many objects at different distance and positions. In order to obtain a precise value for $H(t_0)$ the measurements have to be performed on samples of galaxies far enough away so that motions due to local gravitational effect, the so-called *peculiar motion*, can be neglected. The receding velocity measurement is relatively simple by spectroscopic observation to obtain the redshift of the object. However, the precise distance of a galaxy can be difficult to determine in many situation.

9.1.9 Newtonian limit and gravitational potential

In this subsection we explore briefly the contact of the Newtonian gravity with the Einstein theory and therefore the presence of gravitational potential in the Einstein field equation.²⁴⁾ The Newtonian theory is a limiting form of general relativity and from the discussions presented so far in the above, we do not see where the Newtonian gravitational potential may reside. We will see below the gravitational potential can be introduced through perturbation in the metric tensor. More details on cosmological perturbation theory will be given in Section 12 when the CMB anisotropy is discussed.

In the Newtonian theory two important frameworks involve the equation of motion of a massive point particles in a gravitational force field and the Poisson equation of gravitational potential due to the presence of a mass distribution. They are respectively given by

$$\begin{aligned} \frac{d^2 \mathbf{x}}{dt^2} &= -\nabla\Phi(\mathbf{x}), \\ \nabla^2\Phi(\mathbf{x}) &= 4\pi G_N \cdot \rho(\mathbf{x}), \end{aligned} \tag{9.85}$$

²⁴⁾Parts of the discussions here follow that given in [218], §3.4.

where \mathbf{x} is the trajectory of a massive particle in a local inertial system $\Phi(\mathbf{x})$, and $\rho(\mathbf{x})$ is the density of a mass distribution providing the source of the gravitational potential. The Einstein field equation Eq. (9.5) or (9.13) is the law of gravity in general relativity and the Newtonian gravitational potential is replaced by the ten components 4-dimensional symmetric metric tensor $g_{\mu\nu}$.

In the case of purely gravitational force, the equation of motion in the general case can be written as

$$\begin{aligned} \frac{d^2 x^\mu}{d\tau^2} + \Gamma_{\nu\lambda}^\mu \frac{dx^\nu}{d\tau} \frac{dx^\lambda}{dx} &= 0, \\ d\tau^2 &= -g_{\mu\nu} dx^\mu dx^\nu, \end{aligned} \tag{9.86}$$

where $d\tau$ is the proper time. Let us consider the case of a weak stationary gravitational field in which the particle under consideration moves slowly, so that we can neglect $dx^j/d\tau$, $j = 1, 2, 3$, in comparison with $dx^0/d\tau = dt/d\tau$. The first expression of Eq. (9.86) becomes

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma_{00}^\mu \left(\frac{dt}{d\tau} \right)^2 = 0. \tag{9.87}$$

In the stationary case the time derivative can be omitted, the affine connection appearing above is simply

$$\Gamma_{00}^\mu = -\frac{1}{2} \frac{\partial g_{00}}{\partial x_\mu}. \tag{9.88}$$

We now introduce the gravitational potential as a perturbation to elements of the metric tensor Eq. (9.16). For weak gravitational potential we can write

$$g_{00} = -1 - 2\phi(\mathbf{x}), \quad |\phi(\mathbf{x})| \ll 1. \tag{9.89}$$

Equation (9.87) gives rise to two equations,

$$\begin{aligned} \frac{d^2 \mathbf{x}}{d\tau^2} &= -\left(\frac{dt}{d\tau} \right)^2 \nabla\phi(x), \\ \frac{d^2 t}{d\tau^2} &= 0. \end{aligned} \tag{9.90}$$

The first expression gives the usual particle acceleration equation in Eq. (9.85) when the perturbation term in g_{00} is interpreted as the gravitational potential

$$\frac{d^2 \mathbf{x}}{dt^2} = -\nabla\phi(\mathbf{x}), \tag{9.91}$$

which is the first equation of Eq. (9.85).

We continue the discussion of the case of stationary weak gravitational potential for the Einstein field equation Eq. (9.5). Consider the time-time component of the Ricci tensor:

$$\begin{aligned} R_{00} &= \frac{\partial \Gamma_{\lambda 0}^\lambda}{\partial x^0} - \frac{\partial \Gamma_{00}^\lambda}{\partial x^\lambda} + \Gamma_{0\sigma}^\lambda \Gamma_{\lambda 0}^\sigma - \Gamma_{00}^\lambda \Gamma_{\lambda\sigma}^\sigma \\ &= \frac{\partial \Gamma_{j0}^j}{\partial x^0} - \frac{\partial \Gamma_{00}^j}{\partial x^j} + \Gamma_{0j}^0 \Gamma_{00}^j + \Gamma_{0k}^j \Gamma_{j0}^k - \Gamma_{00}^0 \Gamma_{0j}^j - \Gamma_{00}^j \Gamma_{jk}^k, \end{aligned} \tag{9.92}$$

where no approximation has been made up to this point. Now we have to specify some details about the metric tensor. Since we are looking for local stationary effects, we will neglect the expansion of the universe and therefore drop the Hubble scale factor $a(t)$. The metric tensor can be written as

$$\begin{aligned} g_{\mu\nu} &= \eta_{\mu\nu} + \delta g_{\mu\nu}(\mathbf{x}), \quad |\delta g_{\mu\nu}| \ll 1, \\ \eta_{00} &= -1, \quad \eta_{0j} = \eta_{j0} = 0, \quad \eta_{jk} = \delta_{jk}, \\ \delta g_{00} &= -2\phi(\mathbf{x}). \end{aligned} \tag{9.93}$$

The affine connections $\Gamma_{\nu\lambda}^\mu$ given in Eq. (9.4) are first order in the metric perturbation $\delta g_{\mu\nu}$. Therefore, all four terms in products of two affine connections, which are second order in the perturbation, will be ignored. The term in the time derivative of the affine connection can also be omitted. Using Eq. (9.5), we have

$$\begin{aligned} R_{00} &\approx -\frac{\partial \Gamma_{00}^j}{\partial x^j} = \frac{1}{2} \frac{\partial}{\partial x^j} \frac{\partial g_{00}}{\partial x^j} \\ &= -\nabla^2 \phi(\mathbf{x}). \end{aligned} \tag{9.94}$$

The right-handed side of the Einstein equation Eq. (9.5), together with Eq. (9.27) and the omission of the cosmological constant term, gives

$$-8\pi G_N \left(T_{00} - \frac{1}{2} g_{00} T_\lambda^\lambda \right) = -4\pi G_N [\rho(\mathbf{x}) + 3\mathcal{P}(\mathbf{x})]. \tag{9.95}$$

For non-relativistic matter the pressure term $3\mathcal{P}$ can be omitted. Equations (9.94) and (9.95) give the Poisson's equation

$$\nabla^2 \phi(\mathbf{x}) = 4\pi G_N \cdot \rho(\mathbf{x}). \tag{9.96}$$

This concludes our demonstration of the Newtonian limit and the possibility of explicitly introducing the gravitational potential.

Let us see to what extent the validity of the assumption of weak gravitational potential holds. The engineering dimension of the gravitational potential is energy divided by mass, i.e., the dimension of velocity-squared which is dimensionless in the natural unit. This is the reason why it is a pure number as appearing in Eq. (9.89). Hence to calculate the strength of a gravitational potential in Eq. (9.89), we calculate the potential in any units divided by the light-velocity-squared. Let us take the gravitational potential at the surface of the sun, denoted by $\phi_{s\odot}$, as an example. Given the solar mass $M_\odot = 1.9885 \times 10^{30}$ kg, the solar (equatorial) radius $R_\odot = 6.9551 \times 10^8$ m, and $G_N = 6.6738 \times 10^{-11}$ m³·kg⁻¹·s⁻², we have

$$\begin{aligned} \phi_{s\odot} &\rightarrow \frac{G_N M_\odot}{c^2 R_\odot} \\ &= 2.12 \times 10^{-6}. \end{aligned} \tag{9.97}$$

Discussions for additional astrophysical settings can be found in [218]: The gravitational potential at the surface of the earth is 10^{-9} and that of a white dwarf star 10^{-4} . Therefore, in general the assumption of a first order weak gravitational potential is valid.

9.2 Kinetic theory – Thermodynamics

Much of the early history of the universe can be described by equilibrium thermodynamics, although some important events taking place in the cosmos, such as the decoupling of massive particles, are governed by non-equilibrium dynamics. Below we review briefly the key features of the equilibrium thermodynamics.

9.2.1 Equilibrium distributions

In the ideal gas approximation, in which particles are considered to be diluted and weakly interacting, the energy distribution of particles species j of mass m_j under consideration in a heat bath of temperature T have the following *equilibrium* distribution functions for fermions and bosons respectively,

$$\begin{aligned} f_{Fj}(p) &= \frac{1}{\exp\left(\frac{E_j - \mu_j}{T}\right) + 1}, \\ f_{Bj}(p) &= \frac{1}{\exp\left(\frac{E_j - \mu_j}{T}\right) - 1}, \end{aligned} \tag{9.98}$$

where $E_j = \sqrt{p^2 + m_j^2}$ is the energy of the particle, p the magnitude of the three-momentum, and μ_j its chemical potential. We denote generically the two statistical types of particles as

$$f_{j\pm}(p) = \frac{1}{\exp\left(\frac{E_j - \mu_j}{T}\right) \pm 1}, \tag{9.99}$$

where “+” is for fermions and “−” bosons. So we will use the two types of subscripts $F(B)$ and + (−) interchangeably. The number density of the particle species j is given by

$$n_{j\pm} = \frac{g_j}{(2\pi)^3} \int f_{j\pm}(p) d^3p, \tag{9.100}$$

the energy density

$$\rho_{j\pm} = \frac{g_j}{(2\pi)^3} \int E_j f_{j\pm}(p) d^3p, \tag{9.101}$$

and the pressure density

$$\mathcal{P}_{j\pm} = \frac{g_j}{(2\pi)^3} \int \frac{p^2}{3E_j} f_{j\pm}(p) d^3p. \tag{9.102}$$

The integration is over the whole momentum space. g_j is the internal degrees of freedom of the particle species j ,

generally the number of spin states times other internal quantum numbers, if any, such as the color, isospin, etc.

Let us remark that the number density distribution functions, Eq. (9.99) or Eq. (9.98) are Lorentz invariant quantities. A demonstration of this fact will be given in Appendix A.

9.2.2 Chemical potential

The chemical potential μ is a characteristic of a system of congregation of particles. It is an important concept in thermodynamics for the description of systems of gases. In this elementary physical setting, the chemical potential regulates the flow of particles among systems in contact, while the temperature regulates the flow of energy among them. For instance two systems of gases in contact, originally at temperatures T_1 and T_2 with chemical potentials μ_1 and μ_2 , that can exchange both energy and particles, are said to be in combined thermal and diffusive equilibrium when $T_1 = T_2$ and $\mu_1 = \mu_2$. For applications in cosmology, for a system which is in chemical or diffusive equilibrium, the chemical potentials of all particles involved in a reaction²⁵⁾ are conserved additively, similar to the conservation of additive quantum numbers of the particles involved. For an illustration let us consider the following reaction,



Then if the particles have nonzero chemical potentials, the chemical equilibrium requires

$$\mu_A + \mu_B = \mu_C + \mu_D + \dots \quad (9.104)$$

This relation shows that if there is no restriction on the number of particles of a given species to appear in reactions, the chemical potential of the species vanishes. The Photon is such a particle species as multi-photon emissions are allowed in a process for any charged particle. So the photon has to have a vanishing chemical potential, $\mu_\gamma = 0$. As a consequence, π^0 and η , which can decay into two photons, have vanishing chemical potentials. These results can also be obtained by the fact that particles which are their own antiparticles, such as the photon, π^0 , and η^0 , have vanishing chemical potentials, because the chemical potential of an antiparticle is the negative of that of the particle.²⁶⁾ For a discussion of the chemical potential in thermodynamics, see [237].

In the early universe, under equilibrium in the relativistic case the chemical potentials can be generally neglected [218].²⁷⁾ Then the three density functions,

²⁵⁾Let us note that in order to maintain equilibrium the reaction has to take place rapidly enough.

²⁶⁾This can be seen straightforwardly from the crossing relation. Moving particle B in Eq. (9.103) to the right-handed side we have the reaction $A \rightarrow \bar{B} + C + D + \dots$ and the equation corresponding to Eq. (9.104) is $\mu_A = -\mu_{\bar{B}} + \mu_C + \mu_D + \dots$ which says $\mu_{\bar{B}} = -\mu_B$.

Eqs. (9.100), (9.101), and (9.102), are functions of their energy and the external temperature variable T . However, in some non-relativistic cases, in situations the presence of the chemical potential are crucial and lead to considerations of quantities which are independent of the chemical potentials of the particles involved, such as the Saha equations in baryogenesis and the last scattering surface to be considered later. Nevertheless, we do not need to know the explicit form of the chemical potential in these cases. Hence in general we do not have to concern ourselves with the chemical potential.

For a more detailed discussion of the chemical potential, we refer to [218].²⁷⁾ and [74] which contains discussions involving chemical potentials in several cases.

9.2.3 Entropy

Here we introduce the important quantity entropy, $S(T, V)$, which is a function of the volume V and the temperature T . We are concerned with the situation of the early universe, in which the reaction rate among particles are large than the universe expansion rate so that thermal equilibrium can be maintained. The entropy is defined in terms of the differential equation

$$\begin{aligned} dS(T, V) &= \frac{\partial S(T, V)}{\partial T} dT + \frac{\partial S(T, V)}{\partial V} dV \\ &= \frac{1}{T} [d(\rho(T)V) + \mathcal{P}(T)dV]. \end{aligned} \quad (9.105)$$

The above equations allow us to identify

$$\begin{aligned} \frac{\partial S(T, V)}{\partial T} &= \frac{V}{T} \frac{d\rho(T)}{dT}, \\ \frac{\partial S(T, V)}{\partial V} &= \frac{1}{T} [\rho(T) + \mathcal{P}(T)]. \end{aligned} \quad (9.106)$$

It is reasonable to assume that the derivatives involved are continuous so that the mixed second order derivative with respect to T and V is immaterial in the order of the differentiation variables, i.e.,

$$\frac{\partial^2 S(T, V)}{\partial T \partial V} = \frac{\partial^2 S(T, V)}{\partial V \partial T}, \quad (9.107)$$

which allows us to obtain the following relation, remembering that the densities ρ and \mathcal{P} are not functions of the volume but depend only on the temperature,

$$\frac{d\mathcal{P}(T)}{dT} = \frac{1}{T} [\rho(T) + \mathcal{P}(T)]. \quad (9.108)$$

The time dependence of \mathcal{P} can be obtained through that of the temperature, we have

$$\frac{d\mathcal{P}}{dt} = \dot{T} \frac{d\mathcal{P}}{dT} = \frac{\dot{T}}{T} [\rho(t) + \mathcal{P}(t)]. \quad (9.109)$$

²⁷⁾See [218], pp 530–531.

Comparing this equation with the energy conservation equation Eq. (9.31), we obtain

$$\frac{1}{a(t)^3} \frac{d}{dt} \{a(t)^3 [\rho(t) + \mathcal{P}(t)]\} = \frac{\dot{T}(t)}{T(t)} [\rho(t) + \mathcal{P}(t)], \tag{9.110}$$

which gives the important conservation equation

$$\frac{d}{dt} \left(a(t)^3 \left(\frac{\rho(t) + \mathcal{P}(t)}{T(t)} \right) \right) = 0. \tag{9.111}$$

We define the entropy density

$$s(t) \equiv \frac{\rho(t) + \mathcal{P}(t)}{T}. \tag{9.112}$$

The expression Eq. (9.111) is interpreted as the entropy conservation in a comoving volume $a(t)^3$,

$$a^3(t)s(t) = \text{constant}. \tag{9.113}$$

The entropy density of individual species of particles, from Eqs. (9.101) and (9.102), expressed as functions of time, is given by

$$\begin{aligned} s_{j\pm}(t) &= \frac{\rho_{j\pm}(t) + \mathcal{P}_{j\pm}(t)}{T(t)} \\ &= \frac{g_j}{(2\pi)^3} \frac{1}{T} \int \frac{3E_j^2 + p^2}{3E_j} f_{j\pm}(p). \end{aligned} \tag{9.114}$$

Let us recall that “+” is for fermions and “-” bosons. We can also define the entropy functions as functions of the scale factor a or the temperature T .

The energy and pressures densities that enter the entropy density are those species of particles which are in equilibrium with the heat bath of temperature T , i.e., the temperature of the photon. From the entropy conservation equation the last column of Table 9.2 can be deduced in the case of no entropy production. The argument is as follows. As will be shown in Eq. (9.122) below, the photon contribution is proportional to T^3 , then the total entropy in a comoving volume a^3 can be written as

$$a^3 s \sim a^3 T^3 = \text{constant}. \tag{9.115}$$

Hence the conservation of entropy in a comoving volume requires that

$$T \sim a^{-1}. \tag{9.116}$$

We note that this relation can already be obtained quite early by using the energy-momentum conservation equation Eq. (9.30) and the equation of state Eq. (9.43) for the case of radiation. It is independent of the form of the dominant energy density of the universe. The difference due to the form of the dominant energy is reflected

in the time dependence of the scale factor and therefore the temperature as functions of time, as summarized in Table 9.2. This relation Eq. (9.116) says that the expansion of the universe is adiabatic and it enables us to relate the cosmic temperature with the time:

$$\frac{d(aT)}{dt} = 0, \tag{9.117}$$

which gives

$$\frac{1}{T} \frac{dT}{dt} = -\frac{1}{a} \frac{da}{dt} = -\sqrt{\frac{8\pi G_N}{3}} \rho. \tag{9.118}$$

This explains Footnote 1 of Table 9.2. We shall make use of this relation in later discussions.

The second law of thermodynamics dictates that the total entropy of the universe does not decrease. The cosmos total entropy will increase or stay constant during the evolution of the universe. Since the various known physical processes took place during the cosmic evolution do not generate much entropy, except in the extremely brief period of inflation. It is generally assumed that all the entropy of the universe is generated during inflation and then stays constant afterward.

9.3 Relativistic and non-relativistic behaviors of particle densities

We now come back to the various densities to discuss their properties in some details.

9.3.1 Ultra-relativistic limit

The ultra-relativistic limit is the case $T \gg m$ in which the particle mass can be neglected in the integrands of Eqs. (9.100), (9.101), and (9.102). Under this condition and in the absence of chemical potential, $\mu_j = 0$, all integrals involved are analytically integrable. We have²⁸⁾

$$\begin{aligned} n_{BRj} &= g_j \frac{\zeta(3)}{\pi^2} T^3, \\ n_{FRj} &= \frac{3}{4} g_j \frac{\zeta(3)}{\pi^2} T^3, \end{aligned} \tag{9.119}$$

²⁸⁾The integrals involved can be performed as given below

$$\begin{aligned} \int_0^\infty \frac{x^\nu}{e^x - 1} dx &= \Gamma(\nu)\zeta(\nu), \\ \int_0^\infty \frac{x^\nu}{e^x + 1} dx &= \left(1 - \frac{1}{2^{\nu-1}}\right) \Gamma(\nu)\zeta(\nu), \end{aligned}$$

where the Γ function is given by $\Gamma(n) = (n-1)!$ for an integral value of n . The Riemann zeta function is defined by

$$\zeta(n) \equiv \frac{1}{\Gamma(n)} \int_0^\infty \frac{x^{n-1}}{e^x - 1} dx,$$

are given by: $\zeta(\nu)$ functions for integral n are given explicitly as $\zeta(0) = -\frac{1}{2}$, $\zeta(1) = \infty$, $\zeta(2) = \frac{\pi^2}{6}$, $\zeta(3) = 1.020206$, $\zeta(4) = \frac{\pi^4}{90}$, $\zeta(5) = 1.03693$, $\zeta(6) = \frac{\pi^6}{945}$.

where $\zeta(x)$ is the Riemann zeta function and $\zeta(3) = 1.20206$. The different species of particles are only distinguished by their internal degrees of freedom.

The energy densities Eq. (9.101) are given by

$$\begin{aligned} \rho_{BRj} &= g_j \frac{\pi^2}{30} T^4, \\ \rho_{FRj} &= \frac{7}{8} g_j \frac{\pi^2}{30} T^4. \end{aligned} \quad (9.120)$$

The pressure densities Eq. (9.102) are

$$\begin{aligned} \mathcal{P}_{BRj} &= \frac{1}{3} \rho_{BRj}, \\ \mathcal{P}_{FRj} &= \frac{1}{3} \rho_{FRj}. \end{aligned} \quad (9.121)$$

This energy and pressure relationship is already advertised early as the equation of state for massless particles with $w = 1/3$. The entropy densities Eq. (9.114) are

$$\begin{aligned} s_{Bj} &= \frac{\rho_{BRj} + \mathcal{P}_{BRj}}{T} = g_j \frac{2\pi^2}{45} T^3, \\ s_{Fj} &= \frac{\rho_{FRj} + \mathcal{P}_{FRj}}{T} = \frac{7}{8} g_j \frac{2\pi^2}{45} T^3. \end{aligned} \quad (9.122)$$

To gain some insight of the dense state of the matter in the early universe, let us examine the particle density function in some detail. Taking the relativistic boson for an example, we rewrite the boson particle density in the proper units of number of particle per unit volume cm^{-3} . As discussed in Appendix C.1, the number density of particles is obtained simply by multiplying expressions in Eq. (9.119) by the numerical factor $N_{\text{MeV}} = (5.06773 \times 10^{10})^3 = 1.3015 \times 10^{32}$ when the temperature is given in units of MeV, or by the factor $N_{\text{Kel}} = (4.36704 \times 10^{10})^3 = 8.3284 \times 10^{31}$ when T is units of 10^{10} Kelvin, where these conversion factors can be read off from Table C.2. So we have

$$\begin{aligned} n_{BRj}(T) &= 1.5851 \times 10^{31} g_j \left(\frac{T}{1\text{MeV}} \right)^3 \text{cm}^{-3} \\ &= 1.0143 \times 10^{31} g_j \left(\frac{T}{10^{10} \text{K}} \right)^3 \text{cm}^{-3}. \end{aligned} \quad (9.123)$$

For the photon its density at the present epoch of 2.725 K the number density is 410.5 cm^{-3} . At the early universe of $T = 1 \text{ MeV}$, which is $1.16045 \times 10^{10} \text{ K}$, the photon density is $3.2 \times 10^{31} \text{ cm}^{-3}$.

We are also interested in knowing the number distribution of particles with the minimal energy E_m at a given temperature T , we write the particle number density function as

$$\begin{aligned} n_{BRj}(E_m, T) &= n_{BRj}(T) D_B(\epsilon_T) \\ &= 1.5851 \times 10^{31} g_j \left(\frac{T}{1 \text{ MeV}} \right)^3 D_B(\epsilon_T) \text{cm}^{-3}, \end{aligned} \quad (9.124)$$

where $n_{BRj}(T)$ is the total number of density in units of MeV^3 as given in Eq. (9.119)

$$\begin{aligned} D_B(\epsilon_T) &= \frac{1}{2\xi(3)} \int_{\epsilon_T}^{\infty} \frac{y^2}{e^y - 1} dy, \\ \epsilon_T &= \frac{E_m}{T}, \end{aligned} \quad (9.125)$$

Normalized to unity, $D_B(0) = 1$, $D_B(\epsilon_T)$ is a universal curve giving the fraction of particle with energy equal or greater than $E_m = \epsilon_T T$.

The universal function $D_B(\epsilon_T)$ is plotted in Fig. 9.2. For a given value of T the majority of the photons have the energy of T or less. There are about 10% of the photons with energy greater than $5T$. The fraction drops precipitously when the energy of the photon increases further. But there are significant number of photons of higher energies. There more than 1% of the photon with energy greater than $8T$, 0.23% with energies greater than $10T$, $4 \times 10^{-5}\%$ with energy greater than $20T$, etc.

9.3.2 Non-relativistic limit

For the non-relativistic case, $T/m \ll 1$, we can expand the energy expression in Eq. (9.98) in the non-relativistic form $E_j \approx m_j + p^2/(2m_j)$ and drop the ± 1 terms in the denominators. So we do not have to make the distinction of fermions and bosons. The integrals in the density functions Eqs. (9.100)–(9.102), ignoring the chemical potentials, are straightforwardly integrable to give the Maxwell–Boltzmann particle number density distribution

$$n_{Nj} = g_j m_j^3 \left(\frac{T}{2\pi m_j} \right)^{3/2} e^{-\frac{m_j}{T}}, \quad (9.126)$$

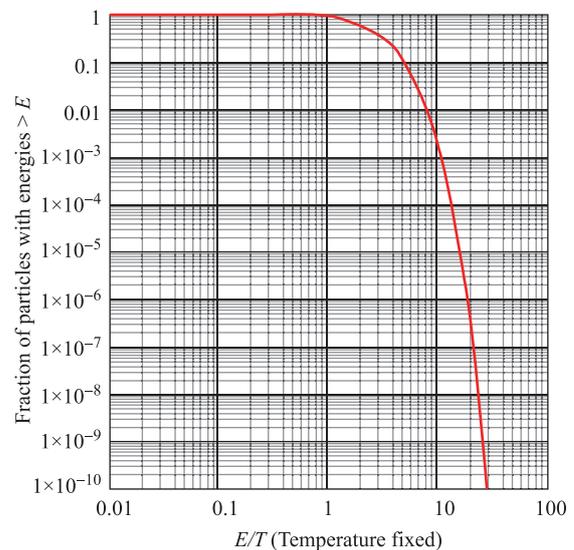


Fig. 9.2 Plot of the particle fraction, $D_B(\epsilon_T)$, with energy greater than ϵ_T .

the energy density

$$\begin{aligned} \rho_{Nj} &= g_j \left(m_j + \frac{3}{2}T \right) m_j^3 \left(\frac{T}{2\pi m_j} \right)^{3/2} e^{-\frac{m_j}{T}}, \\ &= \left(m_j + \frac{3}{2}T \right) n_{BNj} \end{aligned} \quad (9.127)$$

and the pressure density

$$\begin{aligned} \mathcal{P}_{Nj} &= g_j T m_j^3 \left(\frac{T}{2\pi m_j} \right)^{3/2} e^{-\frac{m_j}{T}} \\ &= n_{Nj} T. \end{aligned} \quad (9.128)$$

The equation of state becomes

$$\mathcal{P}_j = \frac{T}{m_j + 3T/2} \approx \frac{T}{m_j} \rho_j, \quad (9.129)$$

where higher order terms in $\frac{T}{m}$ are dropped. Since $T \ll m$, we can set the pressure to zero as noted in Table 9.1.

In circumstances where we need to include the effect of the chemical potential, such as the tinny excess of baryon and leptons than antibaryon and antileptons, which requires nonvanishing chemical potential. Then the non-relativistic density functions become

$$\begin{aligned} n_{Nj}(\mu_j) &= g_j m_j^3 \left(\frac{T}{2\pi m_j} \right)^{3/2} e^{-\frac{m_j - \mu_j}{T}}, \\ \rho_{Nj}(\mu_j) &= g_j \left(m_j + \frac{3}{2}T \right) m_j^3 \left(\frac{T}{2\pi m_j} \right)^{3/2} e^{-\frac{m_j - \mu_j}{T}} \\ &= n_{Nj}(\mu_j) \left(m_j + \frac{3}{2}T \right), \\ \mathcal{P}_{Nj}(\mu_j) &= g_j T m_j^3 \left(\frac{T}{2\pi m_j} \right)^{3/2} e^{-\frac{m_j - \mu_j}{T}} \\ &= n_{Nj}(\mu_j) T. \end{aligned} \quad (9.130)$$

So the relations between the particle number density and those of the energy and pressure density remain the same as the case of vanishing chemical potential. The chemical potential of an antiparticle is the negative of that of the particle.

We note some salient of points of the non-relativistic results in light of the elementary kinetic theory of gas in classical mechanics:

- From Eqs. (9.126), (9.127), and (9.130) we have the average energy per particle

$$\bar{\varepsilon}_c = m + \frac{3}{2}T. \quad (9.131)$$

Restoring the proper units, we have the first term to be the “rest energy” mc^2 of the particle. The second term $\frac{3}{2}k_B T^3$ is the kinetic energy resulting from the well-known equal partition of energy of a classical particle, i.e., $T/2$ in each spatial degree of freedom. So we have

$$\frac{1}{2}m\bar{v}_c^2 = \frac{3}{2}T, \quad (9.132)$$

which is just the kinetic energy of a monatomic molecule $(\gamma - 1)^{-1}k_B T$ for the specific heat ratio $\gamma = 5/3$.

- Similarly the average pressure per particle is given by

$$\bar{\mathcal{P}}_c = T = \frac{1}{3}m\bar{v}_c^2.$$

We summarize the analytic results of the relativistic and non-relativistic cases in Table 9.3.

9.3.3 Internal degrees of freedom g_j

We give some examples of the counting of the internal degrees of freedom. For massless particles the number of spin states is always 2 as only the maximum helicity states are physical. So for the photon $g_\gamma = 2$; gluons $g_g = 2 \times 8$, where 8 is the number of color states of gluons; for the quark and antiquark of a given flavor $g_q = g_{\bar{q}} = 2 \times 3 \times 2$, where 3 is the number of color states of the quark. The total number of degrees of freedom for the three families of neutrinos and antineutrinos is $g_\nu = 3 \times 2$ as neutrinos are left-handed and hence each neutrino species just has 1 degree of freedom. Composite particles and bound states, such as hadrons and molecules which can act in circumstances in a whole as individual entities, are counted similarly by ignoring their internal structure. The proton and neutron each has $g_{p,n} = 2$, and similarly for their anti-particles. Pions of a definite charge state is

Table 9.3 Relativistic and non-relativistic forms of number, energy and pressure density functions.

	Relativistic ($T \gg m$)		Non-relativistic ($T \ll m$)	
	Bosons	Fermions	Fermions/Bosons	per particle
n_j	$g_j \frac{\zeta(3)}{\pi^2} T^3$	$\left(\frac{3}{4}\right) g_j \frac{\zeta(3)}{\pi^2} T^3$	$g_j m_j^3 \left(\frac{T}{2\pi m_j}\right)^{3/2} e^{-(m_j - \mu_j)/T}$	-
ρ_j	$g_j \frac{\pi^2}{30} T^4$	$\left(\frac{7}{8}\right) g_j \frac{\pi^2}{30} T^4$	$\left(m + \frac{3}{2}T\right) n_j$	$m + \frac{3}{2}T$
\mathcal{P}_j	$\rho_j/3$	$\rho_j/3$	$T n_j$	T
$s_j = \frac{1}{T} (\rho_j + \mathcal{P})$	$g_j \frac{2\pi^2}{45} T^3$	$\left(\frac{7}{8}\right) g_j \frac{2\pi^2}{45} T^3$	-	-

Table 9.4 The Boson, fermion, and total degrees of freedom contributing to the energy density function of effective massless particles. The last two columns will be used in Section 11 in the dark matter calculation.

Temperature ^①	Particles added	$4g_{B*}$	$4\left(\frac{7}{8}\right)g_{F*}$	$4g_*^{\text{②}}$	$\mathcal{C}_X^{\text{③}}$ channels	$\mathcal{C}_X/\sqrt{g_*^{\text{④}}}$ channels
$m_e > T$	$\gamma + \nu's$	8	$21\left(\frac{4}{11}\right)^{4/3\text{⑤}}$	13.45	4	2.2
$m_\mu > T > m_e$	e^\pm	8	$14+21\left(\frac{4}{11}\right)^{4/3}$	27.45	5	1.9
$m_\pi > T > m_\mu$	μ^\pm	8	49	57	6	1.6
$T_{\text{cd}}^{\text{⑥}} > T > m_\pi$	π^\pm, π^0	20	49	69	6	1.4
$m_c > T > T_{\text{cd}}$	$u, \bar{u}, d, \bar{d}, s, \bar{s}$, + gluons removing π 's	72	175	247	15	1.9
$m_\tau > T > m_c$	c, \bar{c}	72	217	289	18	2.1
$m_b > T > m_\tau$	$\tau, \bar{\tau}$	72	231	303	19	2.2
$m_{W,Z} > T > m_b$	b, \bar{b}	72	273	345	22	2.4
$m_{\text{Higgs}} > T > m_{W,Z}$	W^\pm, Z	108	273	381	24	2.5
$m_t > T > m_{\text{Higgs}}$	H^0	112	273	385		
$T > m_t$	t, \bar{t}	112	315	427		

^①The temperature here can be considered as that defined for the photon.

^② $g_* = g_{B*} + \frac{7}{8}g_{F*}$ is the total effective degrees of freedom.

^③This is the total number of channels of particle and antiparticles normal particles pairs created in the heavy particle and antiparticle annihilation, to be used in Section 11.6.

^④This will be used in Section 11.6.

^⑤The neutrino temperature is $T_\nu = \left(\frac{4}{11}\right)^{1/3} T_\gamma$ below 1 MeV.

^⑥ T_{cd} is the confinement-deconfinement temperature between quarks and hadrons, taken to lie between 150 and 400 MeV. The inclusion of the pions is to take into account of the gluon effect below the confinement temperature but above the pion mass.

$g_{\pi^\pm, \pi^0} = 1$. Nucleus are also counted similarly: e.g., the helium-4 and helium-3 have respectively spin 0 and 1/2, then $g(^4\text{He}) = 1$ and $g(^3\text{He}) = 2$. Deuteron has spin 1, so $g_D = 3$. For the hydrogen atom, the $1s$ ground state has two hyperfine states of respectively total spins 0 and 1. The spin 0 configuration has only one state while the spin 1 configuration has 3 states. So the counting of the internal degrees of freedom of the ground state hydrogen atom is $g_{H1s} = 1 + 3 = 4$.

Following [238], we summarize in Table 9.4 the international degrees of freedom entering the energy density, referring to the standard model particle spectrum. A complication should be kept in mind in that, as will be discussed later, some of particle species may evolve with a temperature which is different from that of the photon. We should note that in Table 9.4 the effective degrees of freedom in one temperature period to the next change abruptly. This is an approximation, of course. In reality they change continuously. This will be discussed in the next two sections Section 9.4 and Section 9.5 below in some details.

We also note that in the above table the Higgs mass lies below the top quark mass.²⁹⁾ We cannot give the degrees of freedom for energies above the standard model

²⁹⁾The Higgs mass is $m_{H^0} = 125.7 \pm 0.4$ GeV and that of the top quark $m_t = 173.21 \pm 0.51 \pm 0.71$ GeV. See [13] for details.

symmetry breaking energy scale. They are model dependent.

9.3.4 Some detailed properties

Most of the following discussion of some of the details is of no immediate direct relevance to the present goal of the study of dark matter here, we present them below for the purpose of completeness. Figure 9.3 shows the ratios of exact and limiting expressions of the particle number densities, Eqs. (9.100), (9.119), and (9.126). The curves show how the onsets of the validity of the limiting expressions, relativistic and non-relativistic. The temperature is expressed in units of the particle mass T/m . Let us explain the five curves in the figure:

- **The dashed-dotted curve** (black). The dashed-dotted curve in the middle of the figure is the ratio of the exact expressions of fermion over boson. We see that there is no difference between bosons and fermions for low temperature $T/m < 0.2$ which is reflected by the fact that the fermion and boson have the same non-relativistic form. But this fermion and boson identity sets in much earlier than the validity of the non-relativistic limiting expression. For higher temperature the fermion drops to 75% of the boson for $T/m > 8$ which is comparable to the be-

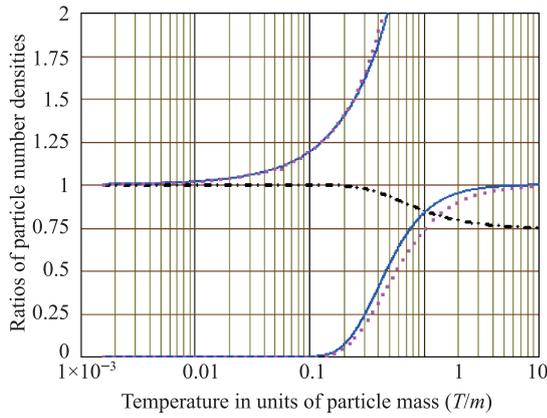


Fig. 9.3 Ratios of the exact and limiting expressions of particle number densities. The dash-dotted (black) curve: ratio of exact particle number densities of fermion over boson; Solid (blue) curve: the ratio of expressions exact over limiting for the fermion; dotted (red) curve: the ratio of expressions exact over limiting for the boson.

gining of temperature range in which the relativistic limiting expressions become exact.

- **Solid curves** (blue). The two solid curves are the fermionic exact expression over its limiting expressions. Above the ratio 1 is for the non-relativistic case and below the ultra-relativistic case.
- **Dotted curves** (red). The two dotted curves are similar to the solid dashed curves but for the boson.

We note that the relativistic limiting expressions are larger than the exact one in the intermediate values of T/m , while the non-relativistic limiting expressions are lower. The fermion approaches to the relativistic limiting form for $T/m > 5$ and the boson for $T/m > 8$. For the non-relativistic case, the approach to the limiting form is slow, roughly for $T/m < 0.01$.

Plots of ratios of energy densities and those of pressure densities show similar features as ratios of number densities as given in Fig. 9.3. The approach to the ultra-relativistic limiting forms are faster than those of the non-relativistic limiting forms.

It is also interesting to examine the energy behavior of the particle number distribution functions. We can write the energy differential distribution of the particle number density,

$$\frac{\partial n_{j\pm}(E)}{\partial E} = \frac{g_j}{2\pi^2} E \sqrt{E^2 - m_j^2} f_{j\pm}(E) \quad (9.133)$$

In Fig. 9.4, we plot the differential distribution Eq. (9.133) as a function of the particle energy E at fixed temperature and a given particle mass. Both the energy and mass are normalized to the temperature. We also normalize the particle mass in terms of the temperature. We show four sets of curves and each set consists

of two curve in solid and dotted lines. The solid lines are for bosons and dotted lines for fermions. Each set of curves has a different point of on set, at $E/T = 0, 1, 3$ and 5 which are their respectively m/T . The zero mass $m/T = 0$ lines are the universal curves for the radiation. The heights of the curves are arbitrary, but normalized with respect to the height of the boson radiation curve. Most features of the curves can be easily understood. For a given temperature and mass, the fermion curve is lower than the corresponding boson curve and the fermion curve peaks at a higher value of the energy than the boson curve, which reflects the fact of the difference of the fermionic vs bosonic statistics. The rapid decrease in the heights of the curves as the mass increases reflects the fact that it become more difficult to excite a state when its mass increases.

From Figs. 9.3 and 9.4 we can summarize some of the general features of particle number density in relation with the temperature and particle energy:

- The fermion and boson distributions are approach to each other when the temperature is low in comparison with the mass $T/m \leq 0.2$.
- The expressions of the non-relativistic limit is valid approximately for $T/m \lesssim 0.02$.
- The expression of the ultra-relativistic limit is valid approximately for $T/m \gtrsim 5$.
- To know quantitatively the energy distribution of number function of a particle, the mass of the particle has to be given. For the case of radiation, or practically massive particle for $T/m > 5$, as can be seen from Fig. 9.4, the maximum occurs at $E/T \simeq 1.594$ for bosons and $E/T \simeq 2.218$ for the fermions. It can also be found that half of the particles have energy greater than $2.35T$ in the case of the boson and $2.84T$ in the case of the fermion.

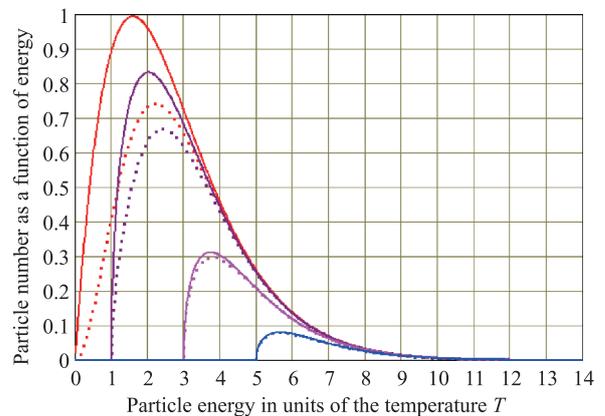


Fig. 9.4 Particle number distribution in the particle energy at fixed mass and temperature. See the relevant text for the identification of different sets of curves.

- For massive particles at a fixed temperature, the height of the maximum of the particle number distribution function decreases as the ratio of the mass over the temperature increases. The difference between the energy where the height of distribution is maximal and the threshold m decreases.
- The average energy per particle, i.e., ρ/n , is linear in the temperature for $T/m \gtrsim 5$ as given in the relativistic situation Eqs. (9.119) and (9.120),

$$\langle E \rangle_{\text{rel}} = \begin{cases} \frac{\pi^4}{30\zeta(3)}T = 2.701T \text{ for bosons,} \\ \frac{7\pi^4}{180\zeta(3)}T = 3.151T \text{ for fermions.} \end{cases} \quad (9.134)$$

In the non-relativistic case for $T/m \lesssim 0.02$ the average energy per particle, as given Eqs. (9.126) and (9.127), is $m + 3T/2$ where $3T/2$ is the average kinetic energy. The intermediate range of T/m extrapolates between the relativistic and non-relativistic results.

- There is a long tail of high energy particles in the particle number density function that depends on the mass of the particle. In the case of radiation or ultra-relativistic particle ($T/m \gtrsim 5$) there are more than 1% of the particles have energies greater than $8T$. For the massive case, i.e., for the temperature to be comparable or smaller than the particle mass, the minimal energy of the 1% higher energy particles moves to higher and higher E/T values. Let us denote the minimal of the higher energy particle to be E_{min} . For example, for $m/T = 1$ we have $E_{\text{min}} = 8.55T = 8.55m$ for bosons and $E_{\text{min}} = 8.75T = 8.75m$. For $m/T = 3$, E_{min} is $9.94T = 3.31m$ for the boson and $9.97T = 3.32$ for the fermion. For $m/T > 5$ boson and fermions behave similarly and $E_{\text{min}} \simeq 11.6 = 2.3m$. For $m/T = 10$ we have $E_{\text{min}} = 16.2T = 1.6m$, and for $m/T = 20$, $E_{\text{min}} = 26T = 1.3m$. So E_{min} moves gradually to the threshold value of m . Hence most particles have very little kinetic energies which is expected. For very high m/T , the whole features becomes academic as number density becomes very small and the particle will decouple to the heat reservoir and is no longer in equilibrium with the heat bath. This is a topic we will discuss below.

9.4 Beyond equilibrium, decoupling of particles or original of species

In the very early universe the temperature is high, much larger than the masses of all particles present. Hence all particles are relativistic and can be treated as massless.

All species of particles, including the photon, have the same temperature and the energy density expression is simply. The total energy density is

$$\rho = g_* \left(\frac{\pi^2}{30} \right) T^4, \\ g_* = \sum_{j_B, \text{bosons}} g_{j_B} + \left(\frac{7}{8} \right) \sum_{j_F, \text{fermions}} g_{j_F}. \quad (9.135)$$

However, during the evolution of the universe, as temperature decreases non-trivial features arise due to episodes in which parts of the content of the universe departs from thermal equilibrium. They include baryogenesis, neutrino decoupling from the cosmos heat bath, primordial nucleosynthesis, photon reheating due to electron-positron annihilation, etc. The temperature ranges and the universe ages for these important cosmos events are: about 3 ~ 4 MeV for $\nu_\mu(\bar{\nu}_\mu)$ and $\nu_\tau(\bar{\nu}_\tau)$ decoupling, about 2 MeV for $\nu_e(\bar{\nu}_e)$ decoupling when the universe is about 0.1 second old, about 1 MeV for the proton-neutron freeze out at the universe age of 1 second, at $m_e/3$ for photon reheating when the universe is 40 second in age, and at about 0.1 MeV for primordial nucleosynthesis when the universe is about three minutes old, etc. Particle species which are decoupled but still relativistic, such as the neutrino, will track the expansion of the universe as they did before the decoupling but may have a temperature which is different from, i.e., lower than, the temperature of particles which are in thermal equilibrium including the photon. We will explain this below. With radiation dominance, in general form the total energy density can be written as

$$\rho = g_{\rho*} \left(\frac{\pi^2}{30} \right) T^4, \\ g_{\rho*} = \sum_{j_B, \text{bosons}} g_{j_B} \left(\frac{T_{j_B}}{T} \right)^4 + \left(\frac{7}{8} \right) \sum_{j_F, \text{fermions}} g_{j_F} \left(\frac{T_{j_F}}{T} \right)^4, \quad (9.136)$$

where T is the photon temperature which is taken to be the temperature of the cosmic heat bath. The other particles may have different temperatures as exhibited explicitly in the above expression. $g_{\rho*}$ is the effective degrees of freedom relative to the degree of freedom of the photon. The analogous expression for the total entropy is

$$s = \sum_j \frac{1}{T_j} (\rho_j + P_j) = g_{s*} \left(\frac{2\pi^2}{45} \right) T^3, \\ g_{s*} = \sum_{j_B, \text{bosons}} g_{j_B} \left(\frac{T_{j_B}}{T} \right)^3 + \left(\frac{7}{8} \right) \sum_{j_F, \text{fermions}} g_{j_F} \left(\frac{T_{j_F}}{T} \right)^3. \quad (9.137)$$

Let us remark that if a relativistic species is decoupled from the rest of the cosmos, its entropy in a comoving

volume a^3 will conserve by itself. The total entropy of all the rest of relativistic particles which couple thermally with the photon will also be conserved in a co-moving volume. Hence the total entropy given in the above expression can be written as a sum of separately conserved components. All relativistic particles species in a given component will have a common temperature. The entropy of the universe is usually referred to be the component containing the photon. The entropy of the photon $s_\gamma = (2\pi^2/45)T^3$ is referred to as the fiducial entropy [239].

9.4.1 Mechanism of equilibrium

What is the mechanism for maintaining thermal equilibrium among different species of particles and for a given species of particles with the thermal heat bath? Here another component of cosmology comes into play: interactions among all relevant particles. Due to elastic and inelastic collisions, energy, momentum, and particle numbers are maintained statistically. Let us focus on the stage of the early universe after the hadronization of quarks and gluons, and most of the short lived hadrons disappear through decays. This begins the period when we know the underline physics governing the particles involved, modulo the dark matter which is supposedly not to participate in the common interactions of the standard model particles. The only baryons present are protons and neutrons; the leptons are electrons, muons and neutrinos; and the ubiquitous photons. The stable particles are the photon, proton, electron, three species of neutrinos as required in the standard model. Their equilibrium can be maintained by the following groups of electroweak elastic scattering, production, and annihilation processes:

$$\begin{aligned} p(\bar{p}) + \gamma &\rightarrow p(\bar{p}) + \gamma, \\ p(\bar{p}) + e^\pm &\rightarrow p(\bar{p}) + e^\pm, \\ p(\bar{p}) + \nu_j(\bar{\nu}_j) &\rightarrow p(\bar{p}) + \nu_j(\bar{\nu}_j), \end{aligned} \tag{9.138}$$

for the elastic scattering involving nucleons, and

$$\begin{aligned} e^\pm + \gamma &\rightarrow e^\pm + \gamma, \\ e^\pm + \nu_j &\rightarrow e^\pm + \nu_j, \\ \nu_j + \nu'_j &\rightarrow \nu_j + \nu'_j, \\ \bar{\nu}_j + \bar{\nu}'_j &\rightarrow \bar{\nu}_j + \bar{\nu}'_j, \\ \nu_j + \bar{\nu}'_j &\rightarrow \nu_j + \bar{\nu}'_j, \end{aligned} \tag{9.139}$$

for the elastic scattering involving leptons, and

$$\begin{aligned} p + \bar{p} &\rightleftharpoons \gamma + \gamma, \quad e^- + e^+, \quad \nu_j + \bar{\nu}_j, \\ e^- + e^+ &\rightleftharpoons \gamma + \gamma, \quad \nu_j + \bar{\nu}_j, \\ \nu_j + \bar{\nu}_j &\rightleftharpoons \nu'_j + \bar{\nu}'_j, \end{aligned} \tag{9.140}$$

for the annihilation and production processes. In the first two reactions of Eq. (9.140), in order for the production

processes to be significant, the temperature of the universe has to be sufficiently high, no less than the sum of the masses of the massive particles in the final states.

Unstable particles will mostly disappear if they are left long enough, at a given epoch, we have to compare their lifetime with the time of the early universe under consideration to ascertain whether or not they can be continuously produced. The muon and anti-muon, though having the short mean lifetime $\tau_\mu = 2.197 \times 10^{-6}$ s, can be produced through particles production mechanism to have a presence in the very early universe if the temperature is high enough. We will come back to it in some detail later. The neutron can live much longer with the mean free lifetime $\tau_n = 880.3$ s or 14.67 min. The processes for them to be in equilibrium with the stable particle are, for the muons

$$\begin{aligned} e^-(e^+) + \bar{\nu}_e(\nu_e) &\rightleftharpoons \mu^-(\mu^+) + \bar{\nu}_\mu(\nu_\mu), \\ e^-(e^+) + \nu_\mu(\bar{\nu}_\mu) &\rightleftharpoons \mu^-(\mu^+) + \nu_e(\bar{\nu}_e), \\ \mu^- + \mu^+ &\rightleftharpoons \gamma + \gamma, \quad e^- + e^+, \quad \nu_j + \bar{\nu}_j, \\ p + \bar{\nu}_\mu &\rightleftharpoons \mu^+ + n, \\ n + \nu_\mu &\rightleftharpoons p + \mu^-, \end{aligned} \tag{9.141}$$

and for the neutron and proton conversion

$$\begin{aligned} n + \nu_e &\rightleftharpoons p + e^-, \\ n + e^+ &\rightleftharpoons p + \bar{\nu}_e, \\ n &\rightleftharpoons p + e^- + \bar{\nu}_e. \end{aligned} \tag{9.142}$$

When the universe expands its temperature drops as the inverse of the scale factor and eventually a massive particle move into the non-relativistic regime if the particle maintains to be in equilibrium. Then the number density of the particle will drop exponentially as given by Eq. (9.126). For example the number ratio of a non-relativistic particles, fermions or bosons, of mass m , and the photon is, from Eqs. (9.119) and (9.126),

$$\eta_m(T) = \frac{n_m}{n_\gamma} = \frac{\sqrt{2\pi}}{8\zeta(3)} g_m \left(\frac{m}{T}\right)^{3/2} \exp\left(-\frac{m}{T}\right), \tag{9.143}$$

which would already drop to about to 10^{-11} for $T = m/29$. For proton this is about $T \approx 33$ MeV or 3.8×10^{11} K. For the electron this is $T \approx 1.8 \times 10^4$ eV or 2.1×10^8 K. Hence if massive particles are in equilibrium with the background radiation today, they would become insignificant in number.

9.4.2 Decoupling – Breakaway from equilibrium

Fortunately the cosmos has a mechanism to break away from the bondage of equilibrium. Roughly speaking when the reaction rate, such as those shown in Eqs. (9.138)–(9.142) is slower than the cosmos expansion rate, equilibrium can no long be maintained. The relevant particles in question will decouple from the background radiation,

called freeze out. Then the energy density and therefore the number density will decrease as a^{-3} for massive particles as indicated in Eq. (9.46). Let us discuss this in some detail.

Let us first consider the universe expansion rate which is given by the Hubble expansion rate $H = \dot{a}/a$. A useful scenario that also allows us to make analytic approximation is the case when the decoupling takes place in the radiation dominated era. From Eq. (9.36), dropping the curvature term, we have in the radiation dominated era, with the effective energy density given by (9.135),

$$H = \sqrt{\frac{8\pi G_N}{3}\rho} = \sqrt{\frac{4\pi^3}{45}\sqrt{g_*}} \left(\frac{T^2}{M_P}\right) \sim T^2 \sim a^{-2}, \quad (9.144)$$

where we have used the Planck mass $M_P = 1.2209 \times 10^{19}$ GeV defined by $G_N^{-1} = M_P^2$.³⁰⁾ So the Hubble expansion rate decreases as T^2 as the universe expands and the temperature decreases.

Next let us estimate the particle reaction rate and compare it with the Hubble expansion rate. The particle reaction rate if defined as

$$\Gamma = n\langle v\sigma \rangle. \quad (9.146)$$

n is the density of the particle under consideration which behaves as $n \sim T^3 \sim a^{-3}$. $\langle v\sigma \rangle$ is the average of the velocity of the particle v times its relevant reaction cross section σ which behaves generally as a power of the cosmic temperature

$$\langle v\sigma \rangle \sim T^{\xi_T}, \quad (9.147)$$

where $\xi_T \geq 0$. We note that the reaction rate Γ has indeed the dimension of inverse time and behaves as

$$\Gamma \sim T^{3+\xi_T} \sim a^{-(3+\xi_T)}. \quad (9.148)$$

At sufficiently early in the universe, the reaction rate will be very large due to the high particle density, it will be larger than the Hubble expansion rate, $\Gamma > H$.³¹⁾

³⁰⁾We can rewrite the Hubble expansion rate as

$$H = \sqrt{\frac{4\pi^3}{45}g_*}M_P \left(\frac{T}{M_P}\right)^2 = 2.066 \times 10^5 \sqrt{g_*} \left(\frac{T}{1 \text{ GeV}}\right)^2 \text{ s}^{-1}, \quad (9.145)$$

where we have used the Planck time $t_P = M_P^{-1} = 5.39123 \times 10^{-44}$ s given in Table C.6. The value of g_* at various temperature range can be found in Table 9.4. The above relation will be given later in Eq. (10.8).

³¹⁾It should be remarked that the universe expansion rate is smaller than particle reaction rate is not true during the inflationary epoch, the universe expands exponentially and the particles in the universe will not be able to maintain in equilibrium. Since the inflation lasted for only a very small time interval it could not affect the particles equilibrium in the subsequent evolution of the universe.

The expansion of the universe does not hinder the particle interactions, so the particle can maintain its equilibrium with the cosmic heat bath. As the universe expands, both Γ and H decreases, but the decrease in Γ is faster by at least one power of the cosmic temperature T , and in some cases much faster, say, by 3 powers of the temperature, T^3 . Such is the case when the scattering involves neutrinos. Then there will be a temperature T_f , called decoupling or freeze out temperature, at which

$$\left(\frac{\Gamma}{H}\right)_{T=T_f} \approx 1. \quad (9.149)$$

Below T_f , the universe expansion rate overtakes that of the particle reaction rate, it will be less and less probable for the particle to find one another to interact in order to maintain equilibrium. The particle will eventually be decoupled from the rest of the particles of the universe, in particular from the photon. Then the particle will be streaming freely and maintains a constant number density in a comoving volume so that its number density decreases like a^{-3} , maintaining a constant ratio with the photon number density which determines the cosmic temperature and decreases also as a^{-3} .

We will make a more detailed illustration on how the decouple happens when we consider below the Big Bang nucleosynthesis and the decoupling of massive particles later in Section 11, where we will show a derivation of Eq. (9.149).

9.5 Thermal history of the early universe

In the evolution of the universe, when the cosmic temperature decreases, some of the particles species will decouple when their reaction rates with the rest of the universe, as defined in Eq. (9.146), to maintain thermal equilibrium become slower than the rate of expansion of the universe. The history of the early universe contains records of such milestone events. We start at the cosmic temperature below 10^{12} K which corresponds to an energy below 90 MeV. At such a temperature, the energy is not large enough to produce muons, for instance, through the production process $\nu_\mu + e^- \rightarrow \mu^- + \nu_e$, and, due its unstable nature with a sufficiently short lifetime $\tau_\mu = 2.1970 \times 10^{-6}$ s, most of the muons from earlier times have disappeared. So we can neglect the effect of the muon. The particles which can contribute to the cosmic energy density include γ , e^\pm , ν_j and $\bar{\nu}_j$, for $j = e, \mu$, and τ , and a tiny fraction of baryons. Our discussion below follows that of [221]³²⁾ and [218]. In order to have the argument valid at temperature around or below the electron mass, we will not neglect the electron mass. But it is justified to ignore the masses of the neutrinos. We

³²⁾See, pp 151–154, [221].

will also ignore the effect of the baryons, except for two topics in which baryons are discussed specifically.

9.5.1 Separation of neutrino and photon temperatures

In this subsection we will derive the temperature difference of the neutrino and photon.³³⁾ The conservation of entropy $s(T)$ in a comoving volume a^3 allows us to connect the cosmic time with its temperature as follows. From the conservation of entropy, $a^3 s(T) = \text{constant}$, we have

$$3\frac{\dot{a}}{a} + \frac{\dot{s}}{s} = 0. \tag{9.150}$$

We rewrite the above equation,

$$3H + \frac{1}{s(T)} \frac{ds(T)}{dT} \frac{dT}{dt} = 0. \tag{9.151}$$

Then

$$t = - \int \frac{1}{s(T)} \frac{ds(T)}{dT} \frac{dT}{3H} + \text{const}. \tag{9.152}$$

Let us first estimate the time-temperature relationship assuming radiation dominance. In the situation of radiation dominance, we have an analytic expression for the total energy density and the Hubble expansion rate,

$$H = \sqrt{\frac{8\pi G_N}{3} \rho(T)}$$

$$\rho(T) = g_* \frac{\pi^2}{30} T^4 = \frac{g_*}{2} \rho_\gamma(T), \tag{9.153}$$

where the total degrees of freedom for the energy density is

$$g_* = 2 + \left(\frac{7}{8}\right) 2 \times 2 + \left(\frac{7}{8}\right) 3 \times 2 = \frac{43}{4}. \tag{9.154}$$

The first term on the right-handed side is from the photon, the second term from the electron and positron, and the third from the three flavors of neutrinos. This value of g_* appears in the second row of Table 9.4.

The entropy density $s(T)$ is proportional to T^3 for relativistic particles, then

$$\frac{1}{s(T)} \frac{ds(T)}{dT} = \frac{3}{T}, \tag{9.155}$$

which is independent the actual energy content of the universe. We can integrate Eq. (9.152), with the Newton constant G_N being related to the Planck mass

$M_P, G_N = M_P^{-2}$ in natural units, to obtain, using Eq. (9.154),³⁴⁾

$$t = \sqrt{\frac{45}{16\pi^3 g_*}} \frac{M_P}{T^2} + \text{const}$$

$$= 0.73818 \left(\frac{1 \text{ MeV}}{T}\right)^2 \text{ s} + \text{const}$$

$$= 0.99416 \left(\frac{10^{10} \text{ }^\circ\text{K}}{T}\right)^2 \text{ s} + \text{const}. \tag{9.156}$$

Following [221] we set the time to be zero when the temperature of the universe is 10^{12} K and call this time \tilde{t} . For simplification we also replace the factor 0.99416 by unity. Then we have the simple time-temperature relationship holds in the radiation dominated epoch for the cosmic temperature below 10^{12} K,

$$\frac{\tilde{t}}{1 \text{ s}} \simeq \left(\frac{10^{10} \text{ K}}{T}\right)^2 - 10^{-4}. \tag{9.157}$$

Let us remark that this relation can be obtained more quickly by the radiation dominance relation $t = 1/(2H)$ as shown in Table 9.2.

The time-temperature relation given in Eq. (9.156) will become problematic when the cosmic temperature decreases below the electron mass. Then we have to include the effect of the electron mass, but still neglect masses of neutrinos. In addition, there is another complication which is the freezing out of neutrinos which we discuss below.

The neutrino decoupling temperature can be estimated using Eq. (9.149). The reactions that maintaining neutrino equilibrium with the electron and positron system are the neutrino-electron scattering listed in Eq. (9.139). For the electron neutrino the scattering is mediated by both charge and neutral currents, and for the muon and tau neutrinos by the neutral current only. So the decoupling of the muon and tau neutrinos may have slightly higher temperatures. The estimate made below will not differentiate among the three neutrino species. The elastic neutrino-electron scattering cross section is given in Eq. (B.8), with $\xi_T = 2$ in Eq. (9.147), we write

$$\langle v\sigma_{\nu e} \rangle = n_\nu \xi_{cs} G_F^2 T^2, \tag{9.158}$$

where ξ_{cs} is a numerical factor of the order of unity. The energy entering in Eq. (9.147) is the total 4-momentum squared S of the initial state, which is proportional to T^2 . The velocity entering in Eq. (9.158) is the order of 1 in natural units. Using the Hubble expansion rate

³³⁾Our order of magnitude estimate of the neutrino reaction rate Γ and the Hubble expansion rate H that enter in Eq. (9.149) is different from that given in [221].

³⁴⁾The second line in the equation below is obtained by converting the MeV into second using the relation $1 \text{ MeV}^{-1} = 6.5822 \times 10^{-22} \text{ s}$.

Eq. (9.144), the effective degrees of freedom Eq. (9.154), and the reaction rate (9.146), we have

$$\frac{\Gamma}{H} = \frac{9}{4} \frac{\zeta(3)}{\pi^2} \left(\frac{43\pi}{5}\right)^{-1/2} \frac{G_F^2 M_P}{\pi} \xi_{sc} T^3 \quad (9.159)$$

which gives for $\Gamma/H = 1$ the neutrino decoupling temperature

$$\left(\frac{T}{1 \text{ MeV}}\right) \Big|_{\nu\text{decoupling}} = \frac{3.30}{(\xi_{cs})^{1/3}}. \quad (9.160)$$

So the neutrino decoupling temperature is the order of 1 MeV, or around 10^{10} °K. Hence when we consider the early history of the universe across the 1 MeV regime we have to take into consideration of the effect of the neutrino decoupling.³⁵⁾

To calculate the temperature of the neutrino we note that when the neutrino decouples its free expansion with the universe makes its number in a comoving volume a^3 a constant. Then its number density maintains the equilibrium form, Eq. (9.119) which leads to the following relationship of neutrino temperature T_ν and the FLRW scale factor

$$T_\nu \propto a^{-1} \quad (9.161)$$

which is an important fact to recognize in the calculation of the neutrino temperature relative to that of the photon.

However, looking at the photon and electron and positron system ($\gamma - e^\pm$), which are in equilibrium, in decreasing temperature below m_e , the electron mass is no long negligible in the calculation of the entropy of the system, and hence the entropy is no longer proportional to T^3 . The conservation of entropy in a comoving volume does not lead to $T_\gamma \propto a^{-1}$, where T_γ is the temperature of the $\gamma - e^\pm$ system. Hence T_ν and T_γ are no longer the same. The entropy of the $\gamma - e^\pm$ system is given by, including the electron mass,

$$\begin{aligned} s_{e\gamma}(T) &= \frac{1}{T} (\rho_\gamma + \mathcal{P}_\gamma) + \frac{2}{T} (\rho_e + \mathcal{P}_e) \\ &= \frac{4}{3} \frac{\pi^2}{15} T^3 + \frac{4}{2\pi^2} \frac{1}{T} \int_0^\infty \left(\sqrt{p^2 + m_e^2} \right. \\ &\quad \left. + \frac{p^2}{3\sqrt{p^2 + m_e^2}} \right) \frac{1}{e^{\sqrt{p^2 + m_e^2}/T} + 1} p^2 dp \\ &\equiv s_\gamma(T) \mathcal{S}_e(T), \end{aligned} \quad (9.162)$$

³⁵⁾The exact energy for $\Gamma/H = 1$ depends on the value of ξ_{SC} . Since the dependence is $\xi_{SC}^{1/3}$, this is not a strong dependence. So the value of $T_{D\nu}$ given in the literature varies. For example, [221] (p. 153) gives the value 10^{10} °K or about 1 MeV, [235] (p. 112) gives also 1 MeV, while [222] (p. 155) gives 4 MeV.

where

$$\begin{aligned} s_\gamma(T) &= \frac{4\pi^2}{45} T^3, \\ \mathcal{S}_e(T) &= 1 + \frac{45}{2\pi^4} \int_0^\infty \left(\sqrt{y^2 + \frac{m_e^2}{T^2}} \right. \\ &\quad \left. + \frac{y^2}{3\sqrt{y^2 + \frac{m_e^2}{T^2}}} \right) \frac{y^2}{e^{\sqrt{y^2 + \frac{m_e^2}{T^2}} + 1}} dy. \end{aligned} \quad (9.163)$$

$s_\gamma(T)$ is the entropy of the photon. We have neglected the effect of the chemical potentials of the electron and positron. The integral in $\mathcal{S}_e(T)$ can be obtained analytically at $T = 0$ and $T \rightarrow \infty$,

$$\begin{aligned} \mathcal{S}_e(0) &= 1, \\ \mathcal{S}_e(\infty) &= \frac{11}{4}, \end{aligned} \quad (9.164)$$

where we have used the results of integrals given in Footnote 28) to obtain the second equality. Combining the conservation of entropy in a comoving volume and Eq. (9.161), we have

$$\begin{aligned} a^3 s_\gamma(T) \mathcal{S}_e(T) &\propto T_\nu^{-3} s_\gamma(T) \mathcal{S}_e(T) \\ &\propto T_\nu^{-3} T^3 \mathcal{S}_e(T) \propto \text{const.} \end{aligned} \quad (9.165)$$

We define the “const” by taking very large T when $T_\nu = T$, then we have

$$\begin{aligned} T_\nu &= \left(\frac{\mathcal{S}_e(T)}{\mathcal{S}_e(\infty)}\right)^{1/3} T \\ &= \left(\frac{4}{11}\right)^{1/3} (\mathcal{S}_e(T))^{1/3} T \\ &\xrightarrow{T \ll 1 \text{ MeV}} \left(\frac{4}{11}\right)^{1/3} T. \end{aligned} \quad (9.166)$$

The neutrino will maintain its number density in the form Eq. (9.119) and its energy density Eq. (9.120) with the temperature T_ν which tracks with that of the photon according to Eq. (9.166).

The above discussion of neutrino is an example of the situation in which the particle species which is frozen out when it is still relativistic, i.e., the freeze out temperature T_f is greater than the mass of the particle, $T_f > m$, and the universe is radiation dominated. The calculation of the relic abundance, which is similar to that of the relic photon, is relatively straightforward as shown above. Another category of freeze out takes place when the particle species becomes non-relativistic, $m > T_f$. We will discuss it in great detail in Section 11 where we consider the freeze out of heavy dark matter particles.

9.5.2 Temperature chronology of the early universe

When the neutrino starts to decouple, the dependence of the time on the cosmos temperature will change, no

longer given by Eq. (9.156). We have to recalibrate their relationship, due to two influencing factors. One factor is the decoupling of the neutrino which is no longer included in the calculation of the entropy of the universe that contains the photon as a component. The temperature of the photon and that of the neutrino are no longer the same, as demonstrated in the previous subsection. The other factor is that the mass of the electron is not negligible so that the universe is no longer radiation dominated. These Changes affect both the entropy and the Hubble expansion rate, the latter through the change in the energy density. Therefore the two expressions, Eqs. (9.153) and (9.155) are modified.

First, the cosmic energy density:

$$\begin{aligned} \rho &= \rho_\gamma + \rho_\nu + \rho_e \\ &= 2 \times \frac{\pi^2}{30} T^4 + \frac{7}{8} \times 6 \times \frac{\pi^2}{30} T_\nu^4 \\ &\quad + 4 \times \frac{1}{2\pi^2} \int_0^\infty \frac{\sqrt{p^2 + m_e^2}}{e^{\sqrt{p^2 + m_e^2}/T} + 1} p^2 dp \\ &\equiv \rho_\gamma(T) \mathcal{E}_e(T), \end{aligned} \tag{9.167}$$

where $\rho_\gamma(T)$ is the energy density of the photon,

$$\begin{aligned} \rho_\gamma(T) &= \frac{\pi^2}{15} T^4, \\ \mathcal{E}_e(T) &\equiv 1 + \frac{21}{8} \left(\frac{4}{11} \mathcal{S}_e(T) \right)^{4/3} \\ &\quad + \frac{14}{8} \left(\frac{120}{7\pi^4} \right) \int_0^\infty \frac{\sqrt{y^2 + \frac{m_e^2}{T^2}} y^2}{e^{\sqrt{y^2 + \frac{m_e^2}{T^2}} + 1} dy}. \end{aligned} \tag{9.168}$$

For $T \gg m_e$, $\mathcal{E}_e(T) \rightarrow g_*/2 = 43/8$ given in Eq. (9.154) as expected. We can again use Eq. (9.152) to obtain the cosmic time as a function of the temperature. Here we have, from (9.162),

$$\begin{aligned} \frac{1}{s(T)} \frac{ds(T)}{dT} &= \frac{1}{s_\gamma(T)} \frac{ds_\gamma(T)}{dT} + \frac{1}{S_e(T)} \frac{dS_e(T)}{dT} \\ &= \frac{1}{T} \left[3 + \frac{T}{S_e(T)} \frac{dS_e(T)}{dT} \right], \end{aligned} \tag{9.169}$$

where the first expression of Eq. (9.163), $s_\gamma \sim T^3$, has been used. From Eqs. (9.167) and (9.168), we write

$$3H = 3\sqrt{\frac{8\pi G_N}{3} \rho(T)} = \sqrt{\frac{8\pi^3}{5}} \sqrt{\varepsilon_e(T)} \frac{T^2}{M_P}. \tag{9.170}$$

From Eq. (9.152) we have the cosmic time for the temperature less than the electron mass

$$t_{<} = -M_P \sqrt{\frac{5}{8\pi^3}} \int \frac{1}{T^3} \left[3 + \frac{T}{S_e(T)} \frac{dS_e(T)}{dT} \right] \frac{1}{\sqrt{\varepsilon_e(T)}} dT. \tag{9.171}$$

To be precise we should add a constant term to the right-handed side of the above expression, but the constant term is quite small.

We can combine the two cosmic time functions $t_{>}$ and $t_{<}$, where $t_{>}$ is given by Eqs. (9.156), to form a chronological order of temperatures vs time, defining $t = 0$ at $T = 10^{12}$ K where the neutrino and the photon have the common temperature. Since neutrinos decouple at around 1 MeV, it still has the same temperature as the photon above a few times 10^{10} K. After decoupling, neutrinos are free streaming, following the Hubble expansion of the universe. Hence the time-temperature relationship of the neutrino follows Eq. (9.171) before and after the decoupling. Approximating the numerical factor 0.99416 by 1, we write

$$T_\nu(t) \approx \sqrt{\frac{1 \text{ s}}{t + 10^{-4}}} \times 10^{10} \text{ K}. \tag{9.172}$$

To obtain the photon temperature as a function of the cosmic time after the neutrino decoupling is more complicate. We have to numerically invert Eq. (9.171), or to invert Eq. (9.166) by replacing $T_\nu(t)$ by Eq. (9.172). For an approximate expression, we take Eq. (9.166) and replace $S_e(T)$ by $S_e(T_\nu)$ to obtain

$$T_\gamma(t) \approx \left[\frac{4}{11} S_e(T_\nu(t)) \right]^{-1/3} T_\nu(t), \tag{9.173}$$

where $T_\nu(t)$ is given in Eq. (9.172). The conditions at both high and low temperatures are satisfied, i.e., at high temperatures we have $T_\nu = T_\gamma$ and at low temperature $T_\nu = (4/11)^{1/3} T_\gamma$. Numerically, the error is a few per cent. We plot the two temperatures in Fig. 9.5.

The temperature difference between the photon and neutrino has a clear physical reason due to entropy conservation (in a co-moving volume). Before the decoupling, the neutrino contributes to a component of the

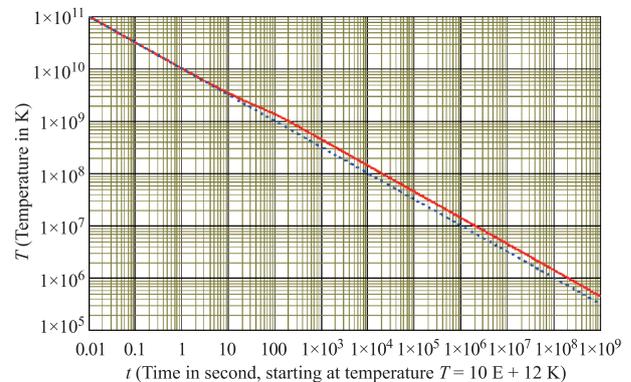


Fig. 9.5 The photon and neutrino temperature as a function of the cosmic time, with starting time to be set at the cosmic temperature 10^{12} °K. The solid red line is the photon temperature T_γ and the dotted blue line is the neutrino temperature T_ν .

cosmic entropy together with the photon and electron-positron. The three individual systems, are in equilibrium and have a common temperature. After the decoupling, the 3 flavors of neutrinos form a separated system which has its own entropy $s_\nu(T_\nu)$ which is conserved in a co-moving volume, where the FLRW scale factor evolves according to $s_\nu(T_\nu) a^{-1}$. The $\gamma - e^\pm$ system forms another conserving entropy with its own temperature. When the universe evolves and the cosmic temperature decreases, the contribution of e^\pm decreases because of their mass. In order to maintain the entropy, the temperature T_γ of the system $\gamma - e^\pm$ has to increase. The usual way to describe this is to say that the e^\pm system dumps its entropy into the photon system. Hence the photon temperature will be higher and the photon number density will be larger. The increase in the photon number density can also be understood due to the electron-positron annihilation into photons $e^- + e^+ \rightarrow 2\gamma$. The total entropy of the universe, i.e., the sum of the neutrino entropy and that of the $\gamma - e^\pm$ system are the same in a co-moving volume before and after the decoupling. This is straightforward to verify by taking the temperature after the decoupling to be sufficiently low so that the e^\pm is negligible.

Figure 9.5 shows a general feature of the variation of the cosmo temperature as a function of time. In most of the period of the early universe, the energy is dominated by radiations and the temperature varies as $T \sim t^{-1/2}$ except the region where a massive species of particle cools down from being relativistic to non-relativistic and transfer energies to the pool of radiation to raise the temperature of the universe and the temperature varies slower than $t^{-1/2}$. Eventually the particle drops out of the picture and the temperature resume its $t^{-1/2}$ behavior, which lies higher than the original $t^{-1/2}$ curve. The region that the temperature deviates from $t^{-1/2}$ lies in the region around the mass of the particle in question, starting from temperature which is several times of the particle mass down to a fraction of the mass.

When the cosmic temperature decreases to about 10^9 K, the temperatures and number densities of the photon and neutrino track each other in the relationship

$$\begin{aligned} T_\nu &= \left(\frac{4}{11}\right)^{1/3} T_\gamma = 1.9454 \left(\frac{T}{2.7255 \text{ K}}\right) \text{ K}, \\ n_\nu &= 0.4091 n_\gamma = 168.0 \left(\frac{T}{2.7255 \text{ K}}\right)^3 \text{ cm}^{-3}, \\ n_\gamma &= \frac{2\xi(3)}{\pi^2} T^3 = 410.7 \left(\frac{T}{2.7255 \text{ K}}\right)^3 \text{ cm}^{-3}, \\ \rho_\gamma &= \frac{\pi^2}{15} T^4 = 0.2606 \left(\frac{T}{2.7255 \text{ K}}\right)^4 \text{ eV/cm}^3, \end{aligned} \quad (9.174)$$

where the temperature of the current epoch is $T_0 = 2.7255$ K corresponding to 410 photons per cm^3 , 168

neutrinos and 58 of them each flavor, and the photon energy density is 0.2606 eV per cm^3 . Most of these photons are in the microwave frequency range, about 1 to 100 GHz. They form the *cosmic microwave background* (CMB). It is the oldest light of the universe, embedded in the sky when the universe is about 380 000 years old, left over during the recombination period, which we will come back briefly in Section 9.5.6. Its observation forms the cornerstone of the Big Bang model of the universe. It is largely homogeneous and isotropic with tiny fluctuation of the order of 10^{-5} . The CMB anisotropy will be discussed in some detail in Section 12.

9.5.3 Baryon-photon ratio in the equilibrium era

At early enough time and with sufficiently high temperature, there is an era that the radiation and the baryonic part of the matter are in close interaction and, therefore, thermal equilibrium, say for the temperature much less than the proton mass of 10^{13} K. Photons are in free expansion and preserves their Planck black-body distribution, while the baryons, i.e., protons and neutrons become non-relativistic and follow the Maxwell-Boltzmann distribution. Hence the temperature of the photon would go like a^{-1} and the photon number density decreases as a^{-3} . Being in the non-relativistic regime, the baryon number density has the form proportional $\exp\{-p^2/(2m_P T)\}$. Then the temperature would seem to decrease like a^{-2} . But the two temperatures should be the same as they are in equilibrium. So how exactly will the temperature behave?

To settle this question let us examine the conservation of baryon number and the second law of thermodynamics.³⁶⁾ In a given comoving volume a^3 both the entropy and the baryon density are conserved. This also says that the entropy per unit of baryon is conserved. Denoting the entropy per unit baryon by σ_B , we have, considering the dominant contribution for the photon and baryon, the second law of thermodynamics gives³⁷⁾

$$d\sigma_B = \frac{1}{T} \left[d\left(\frac{\epsilon}{n_b}\right) + \mathcal{P} d\left(\frac{1}{n_b}\right) \right], \quad (9.175)$$

where ϵ is the total energy density of the system of photons plus baryons and \mathcal{P} is the pressure density. n_b is the baryon number density and $1/n_b$ is the volume per baryon, therefore $d(1/n_b)$ provides the dV term in the second law of thermodynamics. The energy density and pressure density of the photon plus baryon are given by

$$\begin{aligned} \epsilon &= \frac{\pi^2}{15} T^4 + \frac{3}{2} n_b \mathcal{N}_b T, \\ \mathcal{P} &= \frac{1}{3} \frac{\pi^3}{15} T^4 + n_b \mathcal{N}_b T, \end{aligned} \quad (9.176)$$

³⁶⁾We follow the argument given in [221], p. 109–110.

³⁷⁾Recall the second law of thermodynamics: $ds = (1/T)(dU + PdV)$.

where \mathcal{N}_b , which is the number of particles per baryon to take into account the possibility of the presence of light nuclei other than the proton and neutron, is of the order of unity. In the above two equations, the first terms on the right-handed sides are from photons, while the second terms are from the thermal contributions of the non-relativistic baryonic particles present in the system. We note from Eqs. (9.131) and (??) that the energy and pressure per particle for non-relativistic particle do not depend on the chemical potential of the particle. The solution of Eq. (9.175) is

$$\begin{aligned} \sigma_B &= \frac{4}{3} \frac{\pi^2 T^3}{15 n_b} + \mathcal{N}_b \ln \left(\frac{T^{3/2}}{n_b C} \right) \\ &= \frac{2\pi^4}{45\zeta(3)} \frac{n_\gamma}{n_b} + \mathcal{N}_b \ln \left(\frac{\pi^2}{2\zeta(3)} \frac{n_\gamma}{n_b} \frac{1}{CT^{3/2}} \right), \end{aligned} \quad (9.177)$$

where C is an integration constant and n_γ is the photon density given in (9.119).

Since this entropy density is conserved, σ_B of Eq. (9.177) must be nearly constant as a function of the temperature. But the two terms on the right-handed side have different temperature behaviors. The first term requires that n_b tracks with n_γ , i.e., both behaves as T^3 . The second term requires that $n_b \sim T^{3/2}$. A possible way to resolve this conflict is that one of the two terms dominates and being constant in T . The present value of n_γ/n_b is the order of 10^9 (see, Table C.5). If n_b tracks n_γ from this early time down to the present epoch, we can set $n_\gamma/n_b \sim 10^9$. Then the first term is $\sim 10^9$ which is definitely dominant. In comparison, the second term is minuscule in the temperature under consideration, and the integration constant becomes irrelevant.

In order for the second, logarithmic term to dominate we have to have a very small or very large integration constant C . And σ_B is undetermined. We don't consider this physical. So we conclude that during the equilibrium epoch of non-relativistic baryon, the baryon number density tracks with that of the photon in their temperature variation

$$n_b \sim n_\gamma \sim T^3 \sim a^{-3}. \quad (9.178)$$

The present value of the ratio of the baryon and photon number is [13] $\eta \equiv \frac{n_b}{n_\gamma} = 6.05 \times 10^{-10}$ and this value has been maintain since the relatively early universe.

9.5.4 Matter-radiation equality

At the very early universe, the radiation energy dominates the expansion of the universe. Densities of all forms of energy decrease with the expansion of the universe, with the radiation component decreasing faster going like $a^{-4} \sim T^4$, while the non-relativistic matter component $a^{-3} \sim T^3$. Because of the faster decrease, at some temperature period in the the expansion of the universe, the

matter component will becomes dominant. The cross over temperature of the two forms of energies, called the *matter-radiation equality temperature* T_{eq} , can be readily calculated from their values at the present epoch, which we will do below.

The matter energy density, including both the dark matter and baryonic matter, is given by

$$\rho_M = \Omega_M \rho_c \left(\frac{T}{T_0} \right)^3, \quad (9.179)$$

where $T_0 = 2.725$ K is the temperature of the present epoch, $\Omega_M = 0.133h^{-2}$, and the critical density $\rho_c = 1.05375 \times 10^{-2} h^2 \text{ MeV/cm}^3$ [11].

For the radiation energy we include both the photon and neutrino contributions. It is not unreasonable to assume that neutrinos are relativistic at this energy unless the equality temperature is extremely low. We will be able to check this assumption when we obtain T_{eq} . We calculate the radiation energy density according to the expression given in Eq. (9.119),

$$\rho_R = \frac{\pi^2}{15} \left[1 + 3 \frac{7}{8} \left(\frac{4}{11} \right)^{4/3} \right] T^4. \quad (9.180)$$

The equality temperature is given by $\rho_M = \rho_R$. For the convenience of calculation we convert the critical density in units of K^4 , $\rho_c = 1.47457 \times 10^6 \text{ K}^4$. Then we obtain

$$T_{\text{eq}} \approx 8760 \text{ K}. \quad (9.181)$$

The corresponding redshift is

$$z_{\text{eq}} \approx \frac{8760 \text{ K}}{T_0} - 1 \approx 3220. \quad (9.182)$$

9.5.5 Recombination and the last scattering surface

The recombination is the period during which the universe undergoes a fundamental change in which the matter component transforms from ionized gases to neutral atoms. As a consequence the photon reduces its interactions with the electron, proton, and other forms of ions. When there is a sufficient amount of conversion of charged ions into neutral atoms, the photon will no longer have significant interactions and therefore it decouples from matter and can freely stream in the cosmos. So the universe becomes transparent to the photon, and for every observe in the cosmos, there is sphere of space, in which the photon carries the information of its last interactions. The outer most surface of the sphere is called the *last scattering surface*. Space outside the last scattering surface (LSS) is opaque to any observers, because the scattering with charged particles the that the photon have suffered there will erase any memory it may have before the scattering. In this subsection we estimate the temperature and age of the universe when this happens,

and the redshift of the LSS when we try to make a observation of it.

Let us give some more details of the physics of the last scattering surface. When the universe is sufficiently expanded and the temperature becomes significantly low, the content of the universe is similar to what it has presently, but in different proportions and forms: baryon and non-baryon matters, radiation, and dark energy. The matter component is made of the dark matter which has decoupled and the baryon matter which is in complex nuclei and atomic forms, and in close contact with the photon. The radiation component is made of the photon and neutrinos. The photon and neutrino evolve at different temperatures as discussed in Section 9.5.1. Originally, the baryon matter consists of protons, neutrons, and electrons. When the temperature decreases, it first goes through the energy regime of strong interactions undergone the processes of the formation of light nuclei and subsequently photo-dissociated. When the universe further expands, the temperature and energy of the photon decrease more, and eventually it does not have enough energy to breakup the light nuclei. Then the light nuclei will stay. As we will see in Section 10, that at the temperature of a fraction of the nuclear energy scale, less than one tenth of 1 MeV, or the temperature below 10^9 K, nucleosynthesis is completed. The matter component now consists of dark matter, protons, helium nuclei, and electrons, plus very small fractions of other light nuclei. The matter components, not including the dark matter, are in thermal equilibrium with the photon. 24% of the baryon is now in the form of helium 4. Since the helium 4 ionization energy being 24.587 eV, the helium atom will be formed when the photon energy drops roughly to above the eV scale.

At the meantime the hydrogen formation and dissociation are also progressing, following the reaction process



where H denotes collectively the ground state as well as all excited states of the hydrogen atom. Because of the large ratio of photon to proton, there are sufficient number of higher energy photon to ionize the hydrogen atom even at the temperature of the order of 1 eV. But more and more hydrogen atoms are formed to stay as the temperature decreases more and more. When most of the protons are in the form of neutral hydrogen atoms, cosmic photons are free to stream and the universe become transparent. In the following we calculate the fraction of the free proton left as a function of the temperature. Let us recall that at the temperature of the order of 1 eV, the proton, electron, and hydrogen atoms are non-relativistic and their number densities are given by the non-relativistic expressions given in Eq. (9.130).

First we note that the chemical potentials of the parti-

cles involved in Eq. (9.183) have to be taken into account and they play an important role. Since the universe is electrical charge neutral, the number density of the proton and that of the electron are equal. In view that huge difference in their masses the number density equality of the proton and electron can be maintained only with the help of their chemical potentials. Hence Eq. (9.183) says

$$\mu_p + \mu_e = \mu_H. \quad (9.184)$$

We note that H in the right-handed side of Eq. (9.183) represents a large number of states of the hydrogen atom, which include the $1s$ ground state and all the excited states. Furthermore, there are also rare states, such as the fine structure, Lamb shift, and hyperfine structure states. As all atomic states of the hydrogen atom are related by photoemission and since the photon has a zero chemical potential, all hydrogen atomic states have the same chemical potential as that of the ground state. Then the number ratio of the n -th excited state to that of the ground state is $\exp(-(E_1 - E_n)/T)$, where E_n and E_1 are the excitation energies of the n -th and ground states. With $E_n = 13.6/n^2$ eV, $E_1 - E_n \geq E_1 - E_2 = 10.2$ eV. We see that for the temperature around 1 eV the number densities of all excited states, $n \geq 2$ are highly suppressed in comparison with the ground state. So we need only to consider the effect of the ground state of hydrogen atom in Eq. (9.183), greatly simplifies the discussion.

Now we want to calculate the free proton fraction of the total number of protons as a function of the temperature, defining

$$X \equiv \frac{n_p}{n_p + n_H}, \quad (9.185)$$

where n_p is the number density of the free proton and n_H the number density of the hydrogen atom. As will be discussed in Section 10, after the nucleosynthesis, protons, free and bound in the hydrogen atom, are 76% of the total baryons, 24% are in the form of helium 4 plus other rare light nuclei. So we have

$$n_p + n_H = 0.76n_b = 0.76\eta n_\gamma, \quad (9.186)$$

where n_γ is the photon number density and $\eta = n_b/n_\gamma = 6.05 \times 10^{-10}$ as discussed in Section 9.5.3. In the set up so far, we cannot work out the free proton fraction X as defined in Eq. (9.185) without knowing the chemical potential involved. Fortunately there is a way to get around the hurdle as follows. From Eq. (9.184), we see that the following expression is independent the chemical potentials involved,

$$\frac{n_H}{n_p^2} = \frac{n_H}{n_p n_e} = \left(\frac{m_e T}{2\pi} \right)^{-3/2} e^{Q_H/T}, \quad (9.187)$$

where

$$Q_H = m_p + m_e - m_H = 13.6 \text{ eV} \quad (9.188)$$

is the excitation energy of the ground state of the hydrogen atom. For Eq. (9.187) the relationship $n_p = n_e$ has been used in the right-handed side of the first equality. To obtain the second equality we have used Eq. (9.130) and made the approximation $m_H = m_p$. For the internal degrees of freedom, we have $g_p = 2$, $g_e = 2$, and $g_H = 4$.³⁸⁾

Using Eqs. (9.186) and (9.187) we consider the follow quantity which is independent of the chemical potentials, and known as the *Saha's equation*

$$\begin{aligned} S &\equiv \frac{(n_p + n_H)n_H}{n_p^2} = \frac{1}{X^2} - \frac{1}{X} \\ &= 0.76\eta n_\gamma(T) \left(\frac{m_e T}{2\pi}\right)^{-3/2} e^{Q_H/T} \\ &= 0.76\eta \frac{2\zeta(3)}{\pi^2} \left(\frac{2\pi Q_H}{m_e}\right)^{3/2} \left(\frac{Q_H}{T}\right)^{-3/2} e^{Q_H/T} \\ &= 2.50146^{-16} \left(\frac{Q_H}{T}\right)^{-3/2} e^{Q_H/T}, \end{aligned} \tag{9.189}$$

where we call S the *Saha factor*. The baryon over photon number ratio $\eta = 6.05 \times 10^{-10}$ can be found in [11] as already mentioned in Section 9.5.3. We can now solve the Saha's equation to obtain the free proton fraction

$$X = \frac{1}{2S} \left(-1 + \sqrt{1 + 4S}\right). \tag{9.190}$$

This solution of the root of a quadratical equation is generally dull and trivial. However the rapidly varying function S as a function of the temperature can make it interesting. Let us first examine its behavior for small and large S :

$$X = \begin{cases} 1 - S + \mathcal{O}(S^2), & S \ll 1, \\ \frac{1}{\sqrt{S}} (1 + \mathcal{O}(S^{-1})), & S \gg 1. \end{cases} \tag{9.191}$$

Because of its exponential behavior, S varies very quickly, which makes X to drop precipitously when S increase from the value 1. We plot in Fig. 9.6 X against T and z , with the temperature and with the redshift which are related by

$$1 + z = \frac{T}{T_0}, \tag{9.192}$$

where $T_0 = 2.725$ K is the temperature of the present epoch. As the blue dotted and pink dash-dotted curves show that the Saha factor, as an exponential function is a very steep function of the temperature or redshift, varying several orders of magnitude when the temperature

or the redshift vary just a factor of a few. The curves in Fig. 9.6 show clearly the features given in Eq. (9.191). When S is below unity, X stays close to 1. When S increases over unity as T or z decreases, X decreases precipitously. We list in Table 9.5 several values and the corresponding values of T and z .

Since during this period, which has the temperature of several thousands of Kelvin, the universe is already matter-dominated, we can calculate the age of the universe at each of these temperatures. From Eq. (9.78), we have

$$t(z_1) = \frac{1}{H_0} \int_{z_1}^{\infty} \frac{dz}{(1+z) \sqrt{(1+z)^4 \Omega_R + (1+z)^3 \Omega_M + \Omega_\Lambda}}, \tag{9.193}$$

where values of the various quantities are given below Eq. (9.81). The age of the universe at the several redshift values are given in Table 9.5. We have comments as follows:

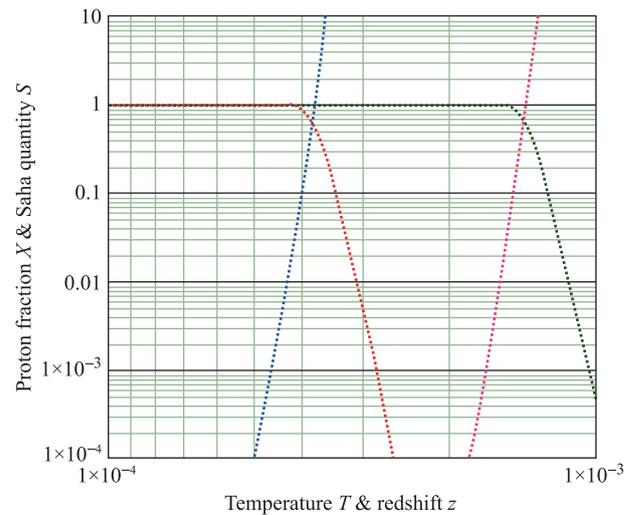


Fig. 9.6 Proton fraction and the Saha factor. The horizontal axis denotes the temperature T and redshift z . Note that T and z decrease from left to right, or in the direction of increasing time as the universe expands. The red solid curve is the free proton fraction as a function of the temperature, the green dashed curve the proton fraction as a function of the redshift, the blue dotted curve the Saha factor as a function of the temperature, and the pink dash-dotted curve the Saha factor as a function of the redshift.

Table 9.5 The free proton fraction, i.e., the ratio of free proton number density to the total proton number density, and the time of its occurrence.

X	0.99	0.5	0.4	0.3	0.2	0.1	0.01
T (K)	4295	3737	3680	3616	3537	3418	3093
z	1575	1370	1349	1326	1297	1253	1134
t (Megayear)	0.221	0.280	0.287	0.295	0.307	0.325	0.383

³⁸⁾The hydrogen ground states have the triplet spin 1 and singlet spin 0 states. We treat these spin states as degenerate by neglecting the hyperfine structure due to the spin-spin interaction arising from the intrinsic magnetic moments of the proton and electron.

- The universe ages at the various temperature and redshift given in Table 9.5 are different from those given [221]³⁹⁾. Our age values are also different from those given in [240]⁴⁰⁾ for the same reason.
- In the present calculation of the age of the universe at various redshift values, the value of the redshift is the order of 10^3 . So it is tempting to drop both the vacuum and radiation contributions and the integral in Eq. (9.193) can be done analytically. In doing so, a sizable error can occur.
- The present calculation is for values of X not too far from unity when the electron is still in equilibrium with the photons. However for small values of X and therefore lower temperature the electron will go out of equilibrium, the Saha equation is no longer valid. We have to use the full Boltzmann equation in dealing with temperature dependence of the free proton ratio. With the exact calculation one finds that X decreases much slower as the temperature decreases and eventually flattens out at a value of a few times 10^{-4} . However, the Saha equation gives a good estimate of the redshift value at the recombination. We refer to [221] and [101] for the details.⁴¹⁾ We will illustrate the use of the Boltzmann equation in Section 11 in the discussion of the freeze out of massive particles.

9.5.6 Decoupling of the photon and the last scattering surface

When the recombination is completed, the photon will lose contact with matter. It no longer suffers scattering with the electron and other charged particles, dominated by the scattering with the electron, to become free-streaming. Hence the universe now becomes transparent and the photon is able to carry the information of the universe that is imprinted on it during the last scattering.

The decoupling of the photon takes place when the rate of the elastic electron-photon scattering, or the Compton scattering,

$$e^- + \gamma \rightarrow e^- + \gamma \quad (9.194)$$

falls below the rate of the expansion of the universe. This also means that the mean free path of the electron is longer than the Hubble length given by

$$d_H = \frac{c}{H}, \quad (9.195)$$

where we restore the appearance of the speed of the light in order to explicitly show the proper units it has. The

³⁹⁾See, Table 2.2, p. 124 in [221] at similar z values. The difference is due the different values of cosmic parameters used here and those in [221].

⁴⁰⁾Table 9.1, p. 195.

⁴¹⁾See [221] pp 116–129 and [101] pp 70–73.

mean free path of the electron is given by

$$\lambda_e = \frac{1}{n_e \sigma_{\text{Th}}}, \quad (9.196)$$

where n_e is the electron number density and σ_{Th} the Thomson scattering cross section. The Thomson scattering is the low energy Compton scattering in the limit of zero energy photons. It has a constant cross section proportional to the classic radius of the electron which can be found in Appendix B.1.2.

$$\begin{aligned} \sigma_{\text{Th}} &= 6.65256 \times 10^{-25} \text{ cm}^2 \\ &= 1.2687 \times 10^{-23} \text{ K}^{-2}. \end{aligned} \quad (9.197)$$

The decoupling temperature is given by the condition $\lambda_e = d_H$, i.e.,

$$H = n_e c \sigma_{\text{Th}}. \quad (9.198)$$

We can start directly with this equation, which says that the electron reaction rate $n_e \sigma_{\text{Th}}$ equals the universe expansion rate H . σ_{Th} is given Eq. (9.197) and the other factors entering in Eq. (9.198) are

$$\begin{aligned} n_e &= b_p \eta n_\gamma X_e, \\ H &= H_0 \sqrt{\Omega_\Lambda + \Omega_M \left(\frac{T}{T_0}\right)^3 + \Omega_R \left(\frac{T}{T_0}\right)^4}. \end{aligned} \quad (9.199)$$

With the parameters $b_p = 0.76$, $\Omega_M = 0.26$, $\Omega_R = 4.8 \times 10^{-5}$, $\Omega_\Lambda = 1 - \Omega_M - \Omega_R$, $\eta = 6.23 \times 10^{-10}$, and $X_e = X$ given by Eq. (9.190), we obtain the approximate electron decoupling temperature

$$T_{\text{ed}}^{(A)} = 3112 \text{ K}, \quad (9.200)$$

which corresponds to the approximate redshift, $z_{\text{ed}} = (T_{\text{ed}}/T_0) - 1$, and the age

$$\begin{aligned} z_{\text{ed}}^{(A)} &= 1142, \\ \tau_{\text{ed}}^{(A)} &= 0.38 \text{ Megayear} \end{aligned} \quad (9.201)$$

and the free proton or electron fraction is about 0.1.

The values given above overestimate the decoupling epoch. The Boltzmann calculation gives

$$\begin{aligned} T_{\text{ed}} &\approx 3000 \text{ K}, \\ z_{\text{ed}} &\approx 1100, \\ \tau_{\text{ed}} &\approx 0.42 \text{ Megayear} \end{aligned} \quad (9.202)$$

and the corresponding electron or proton fraction is the order of 0.4. So the decoupling happens in the middle of the recombination era, while in the Saha approach it happens near the end of the recombination era. Although the Saha approach is a poor estimate of the proton or electron fraction because of the rapid decreasing of the fraction as a function of the temperature, its results on

the decoupling temperature, and therefore redshift, has only a 4% error. The estimate of the age of the universe at the last scattering surface is also quite good using the Saha approach. So the last scattering surface is a spherical surface of about 13.6 billion years, during which the universe lives most of its time.

The decoupling of the cosmic photon with the charged media, which consists of charged particles of electrons and protons, does not happen instantaneously, but over a period of time. Therefore the LSS is not a single surface but has a finite depth during which the mean free path of

the photon increases rapidly. The first year WMAP [241] determines that the LSS depth is about 115 000 years, while the CMB photons start to form when the universe has the age of about 372 000 years, and the formation of CMG is completed at the universe age of 487 000 years.

9.5.7 A cheat sheet of the cosmo-history & a map of the universe

A cartoon of the cosmic timeline from the Big Bang to the present is shown in Fig. 9.7. In the table below that follows, Table 9.6, we list the various events hap-

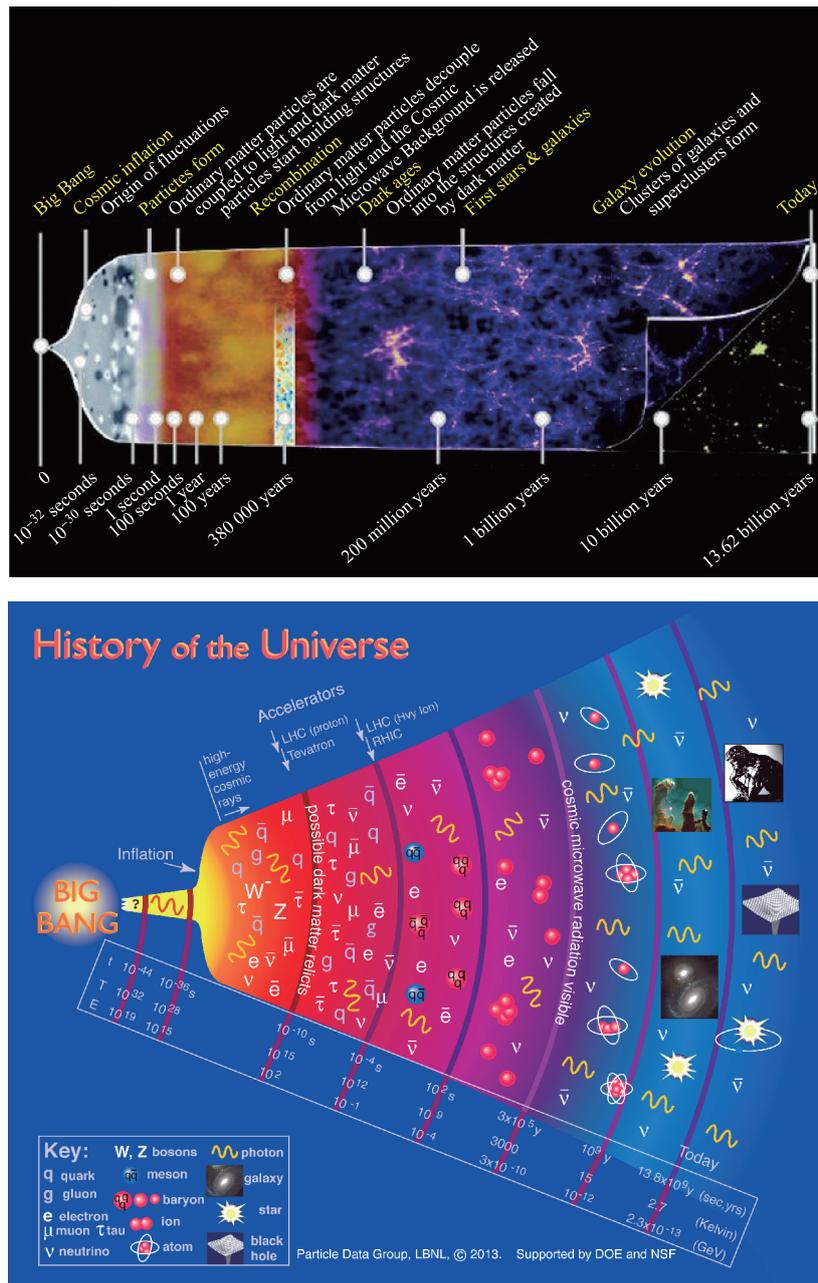


Fig. 9.7 The cosmic Chronology of the standard model of the universe from the Big Bang to the present epoch.

Table 9.6 Historical landmarks in the evolution of the universe. Gyr is giga years and Myr million years. The time-temperature relation $t = (10^{10} \text{ K}/T)^2 \text{ sec.}$, Eq. (9.171) is useful in the radiation dominated era. The various numbers given should be considered as approximations that provide order of magnitude estimates.

Cosmic event	Redshift z	Temperature (K) (energy)	Age τ_U
	Comment		
Planck era		$\gtrsim 10^{32}$ ($\gtrsim 10^{19}$ GeV)	$\lesssim 10^{-43}$ s
Inflation era			10^{-36} – 10^{-34} s
	Seeding density perturbations		
End of inflation, reheating			$\sim 10^{-32}$ s
	Vacuum energy transferred to radiation & matter		
Undelineated physics	Undelineated but important physics took place Including possibly SUSY breaking at $\sim 10^{11}$ – 10^{13} GeV, decoupling of dark matter, etc.		
Physics of SM (EW era) Antiparticles disappearing	$\sim 10^{16}$ (10^3 GeV) Order of the physics listed here is not clear Physics between reheating and ν -decoupling not delineated		
Neutrino decoupling		1.6×10^{10} (1 MeV)	~ 0.4 s
	$T_\nu = \left(\frac{4}{11}\right)^{1/3} T_\gamma$		
Neutron-proton freeze out	4×10^9	10^{10}	~ 1 s
Freeze out of $2\gamma \rightarrow e^+e^-$	8×10^8	2×10^9	40 s
Nucleosynthesis	3×10^8	10^9	$\sim 2 \times 10^2$ s
	Synthesis of light elements: ^4He , D, ^3He , Li		
Matter-radiation equality	3220	8800 (< 1 eV)	$\sim 50\,000$ yr
Before recombination	Baryonic plasma (e^- + nuclei) due to Coulomb interaction Isotropic \mathcal{P}_γ due to Thomson scattering with baryonic plasma		
Recombination	End of γ -baryon plasma thermal equilibrium		
	1370	3740	0.24 Myr
Photon decoupling	1100	3000	0.38 Myr
Last scattering surface	1100	3000	0.38 Myr
	Universe transparent to photons		
Dark age			0.35–400 Myr
First star	70		30 Myr
Reionization by first stars	11	30	100–400 Myr
Λ -radiation equality	9	25	540 Myr
Structure formation	7		1 Gyr
	Formation of stars, galaxies, clusters		
Λ -matter equality Solar system formation	0.42	3.9	9.2 Gyr
Present epoch	0	2.725	13.81 Gyr

pened during the evolution of the universe, listing their redshifts, temperatures, and ages. They tell us about the universe at their respective times/epochs. Our numbers may not be precise, but they show the chronological orders and basic parameters of the various important landmarks of the cosmic history.

A conformal map of the universe depicting the recent discoveries up to the beginning of this 21st century can be found in [242], which is largely for experts.

10 Big Bang nucleosynthesis: Primordial helium

The theory of the Big Bang nucleosynthesis (BBN) was proposed in 1948 [243]⁴²⁾ which ascertains that the Big

⁴²⁾There is an interesting history to this paper. See <https://en.wikipedia.org/wiki/Alpher-Bethe-Gamow_paper> for a brief description and additional references.

Bang created the chemical elements hydrogen, helium, etc., in the right amount in the early universe. In the 1970 it became clear that the cosmic baryon density calculated on the base of the BBN was much less than that obtained from the universe expansion rate. This arose as a serious challenge to the BBN theory. With the introduction of the dark matter, the disagreement was removed. And it is the first indication of the existence of non-baryon matter other than the evidence from rotation curves of galaxies.

According to the Big Bang theory, nuclear reactions taking place in the first few minutes produced the first three elements, the hydrogen, helium, lithium, and their isotopes. Heavier elements are produced in stars much later. In this chapter we discuss helium production that took place in the early universe during the first three minutes to illustrate a general method for treating the decoupling of the baryon and the mechanism of light elements production, mostly the helium-4. We also outline the production of other rare ones, such as deuterium⁴³⁾ Let us quote from the brief review on Big-Bang Nucleosynthesis (BBN) given in the 2011 Particle Data Book [11] “Predictions of the abundances of the light elements, D, ³He, ⁴He, and ⁷Li, synthesized at the end of the ‘first three minutes’, are in good overall agreement with the primordial abundances inferred from observational data, thus validating the standard hot Big-Bang cosmology This is particularly impressive given that these abundances span nine orders of magnitude - from ⁴He/H \approx 0.08 down to ⁷Li/H \approx 10^{-10} (ratios by number). Thus BBN provides powerful constraints on possible deviations from the standard cosmology . . . , and on new physics beyond the Standard Model”

In BBN the determination of the abundances of the light elements are dependent on the baryon to photon ratio η which is now measured quite accurately by the anisotropy of the cosmic microwave background CMB. From η , the fraction of the baryonic matter $\Omega_b h^2$ is determined. It turned out that this is much smaller than the fraction of all non-relativistic matter that should contribute to the critical density of the universe. This discrepancy provides a quantitative evidence of the existence of non-baryonic matter. Furthermore, BBN can also serve as a test of the standard model of cosmology, with the inclusion of dark matter, and as a probe in the dark matter study [244, 245]. A particular relevant point is the lithium problem which we will return to at the end of this Section.

10.1 Introduction

Let us elaborate briefly what are stated above. Currently,

⁴³⁾Most of the discussion we focus on concerns with the deuterium nucleus, the deuteron. Following the convention usage, we will use deuterium to represent both the atom and its nucleus.

the most abundant chemical element is the hydrogen which constitute about three quarter, in mass, of the baryon matter.⁴⁴⁾ The next widely distributed element in the cosmos is the helium-4 which forms about one quarter of the cosmic baryon mass. The fact that the origin of the hydrogen lies in the early universe is quite clear. Convincing arguments can also be made for the case that the majority of the other light elements, the deuterium, etc. are also of cosmic origin, most of them are unlikely formed by stars burning [74, 218].⁴⁵⁾ The formation of nuclei involves nuclear physics processes of the energy scale of MeV. For instance, the binding energy of the helium-4 is 28.3 MeV and that of the deuterium 2.23 MeV, their formation must be at energies smaller than their binding energies. Hence the temperature range of the BBN is around 10^{10} K or lower, recalling that 1 MeV is 1.16×10^{10} K. Under such high temperatures, there are very few tools available for us to probe the physics involved. It is still a photon fog state when photons are actively interacting with all the charged particles present at the time, which will smooth out most of the memories that the photon may initially have. Furthermore, the neutrino is also in the verge of being decoupled and they cannot offer useful clues for us to study. BBN, which allows the prediction of the abundances of light elements being produced in the early universe, together with the cosmic microwave background (CMB) temperature fluctuation, offers stringent probes of the standard model of the universe [246] in this energy regime.

There are recent reviews on the subject which provide more details than given in the PDG article referred above. Several review articles are given in [221] which are accessible from the high energy Archive. For the readers' convenience, they are listed in [247].

The physical argument for the BBN goes roughly as follows. When the universe temperature decreases to the QCD scale of the order of a hundred MeV, hadrons are formed and the nucleons, protons and neutrons, are in equilibrium and they become non-relativistic when the temperature is below their mass. When the temperature decreases to the order of tens of MeV which is in the nuclear physics energy scale, light nuclei begin to form by the fusion of protons and neutrons and they are all in

⁴⁴⁾Baryon matter is made of protons and neutrons, or quarks and antiquarks. It is the ordinary matter, which also include other short-lived particles, unstable over the cosmic time scale. It also include electrons, in numbers equal to that of the proton, which do not contribute significantly to the present mass of the universe because they are about 2000 times lighter than protons. Baryon matter can be found in many different forms: gaseous clouds, neutron atoms and molecules, ionized plasmas, frozen condensations of comets. It also exists in dense and hot environments such as stars, planets, stellar remnants like white dwarfs, neutron stars, and black holes. Except for stars, star remnants, and planets, baryon matter consists mainly hydrogen and helium.

⁴⁵⁾See, §15.7, particularly pp 545–546 in [218], and §1.6, pp 15–16 in [74].

equilibrium, although their number are rather small. We expect that the lightest nucleus, i.e., the deuterium, consisting of a proton and a neutron, appears first through two-body interaction processes. The direct formation of multi-nucleon nuclei, such the tritium, helium, etc., are very rare because multi-body reactions are highly suppressed and are proportional to the product of several powers of nucleon densities which are small to begin with. For instance, the direct production of helium from nucleons, $2p + 2n \rightarrow {}^4\text{He} + \gamma s$, requires an interaction of four-body initial state and the product of four nucleon densities functions, two for the proton and two for the neutron. The nucleon density functions are very small when the four-body reaction can take place, giving rise to negligibly small reaction rates. Hence the effective way for BBN is to create nuclei of higher and higher atomic numbers, step by step, $A = 2, 3, 4, \dots$, through a series of two-body reactions.

The formation of nuclei begins with that of the deuterium when the temperature is still above its bind energy of 2234.52 ± 0.20 keV. In this temperature range the fusion of nucleons into deuteriums and the deuterium dissociation are equally probably. As the temperature decreases sufficiently below the deuterium nuclear binding energy, the deuteron fusion rate overtake the rate of its dissociation, more and more deuterons are accumulated. The formation of the deuterium becomes the first step of a series of reactions that form the helium and a few other light elements.

This body of helium so produced is referred to as the primordial helium and the process, and in general the BBN, as cosmological nucleosynthesis. This is to be distinguished from nucleosynthesis of heavier elements which takes place inside of stars during the latter's evolution. However, for two reasons the procedure of cosmological nucleosynthesis cannot continue very long so as to produce very heavy isotopes. One is that there is no stable nuclei of mass numbers $A = 5$ and 8 to support the process of the step by step increase in the mass number of the nuclei formation beyond $A = 7$. And the other is that after the formation of the helium, the temperature has fell sufficiently low and there is very little possibility to form nuclei of much higher atomic numbers. BBN can form nuclei as heavy as ${}^7\text{Li}$ and heavier nuclei have to be await for a long while to be produced in star burning.

The calculation of the primordial helium abundance constituents some of the most quantitative works in cosmology. It deals with a theory which goes beyond equilibrium and makes use of calculations used in particle and nuclear physics. The success of the calculation of the helium abundance in the early universe forms a part of the foundation of the Big Bang model of the cosmos and provides good probes to aspects of particle physics, such as the number of species of neutrino or, generally,

massless particles. Discussions on BBN of helium and other light elements, which are accessible to beginners can be found in [218] and [221].⁴⁶⁾ Further discussions can be found, fore example, in [248], [249], and [250].

Precise calculations of BBN require detailed numerical work, in which the physics of BBN is difficult to reveal. However, analytic analysis assisted by nominal numerical computation are given in several works: [251–253]. We follow closely the approach given in [251] which will be referred to as BBF. We will also adopt some of the arguments give in [252] and [253]. The approach of BBF [251] involves necessarily approximations which can be justified on physics ground. It deals only with the abundance of helium 4. Analytic calculations of the abundances of other very rare light elements can be found in [252] and [253]. In this Section our goal is to show, using examples, the type of calculations employed and the physics involved in this branch of fundamental science. Our calculation concerns mostly on the proton-neutron conversion through inelastic collisions at high temperatures due to weak interactions and the subsequent decoupling of the proton and neutron when temperature decreases, remember that neutron has a higher mass than the proton and, therefore, equilibrium statistics will favor to have more proton than neutrons as cosmos evolves so that its temperature decreases as time increases. The road map of the calculation is the followings:

- We will first estimate the energy scale at which the proton and neutron begin to decouple and therefore the rough abundance ratio of neutron to proton at the decoupling. This has to happen sufficiently early in the universe in order to make sure that the number density of (free) neutrons remains high enough for the nucleus formation, in the wake of decays of free neutrons.
- We then demonstrate decoupling quantitatively through the solution of a first order differential equation, mostly by analytical approach, but with the assistance of nominal numerical work. The effect of the decoupling is that the neutron and proton densities can maintain a constant ratio independent of the decrease of the cosmos' temperature when the cosmos time increases. This is done by ignoring the decay of the neutron.
- The decay channel of the neutron cannot be decoupled, of course. Hence if it left in the free state, the neutron will disappear eventually. We demonstrate this effect by including the decay channel of the neutron into the solution of the differential equation.

⁴⁶⁾See [218] pp 545–556 and [221] pp 159–173. A concise summary of the history of the calculation of the production of light elements in BBN can be found in footnote 1, pp 159–160, [221] where several recent reviews of the subject are also given.

- Nature has an ingenious way to prevent the decay of the neutron by storing them in nuclei. Then we have to calculate when the fusion of neutrons and protons to form a nuclei begins. Once it starts, the fusion, which is a nuclear (strong interaction) interaction with possibly some electromagnetic process, will proceed very quickly. Then all the undecayed free neutrons will be captured almost instantly into deuteriums and possibly other light nuclei.

10.2 Estimate decoupling

We first observe that the amount of primordial helium formed depends on the amount of the deuterium available, which in turn depends on the neutron to proton ratio at the time when the deuterium begins to form. In the early universe, say the temperature of 10^{12} K, or the order of a hundred MeV, all particles present are kept in thermal equilibrium through rapid collision via weak interactions. when the universe temperature decreases to the order of MeV, the weak interaction rates are no longer fast enough in comparison with the universe expansion rate, and the neutron and proton are decoupled to freeze out. So we have to calculate the neutron concentration at freeze out.

The neutron and proton maintain their equilibrium by their interactions with members of the first lepton family,

$$\begin{aligned} n + \nu_e &\rightleftharpoons p + e^-, \\ n + e^+ &\rightleftharpoons p + \bar{\nu}_e, \\ n &\rightleftharpoons p + e^- + \bar{\nu}_e. \end{aligned} \tag{10.1}$$

Let us start at the temperature of about one trillion Kelvin, i.e., $T \approx 100$ MeV or 1.1605×10^{12} K. It is reasonable to assume that all the baryons are in the form of protons and neutrons. The proton and neutron are non-relativistic while maintain thermal equilibrium due to weak interactions according to Eq. (10.1). Their thermal distributions are given by Eq. (9.130) and they are assumed to have negligible chemical potentials. Since neutron and proton have about the same mass, their densities ratio at the thermal and chemical equilibrium is expected to take the simple Boltzmann-Maxwell form given by

$$\left(\frac{n_n(T)}{n_p(T)}\right)_{\text{eq}} \simeq \exp\left(-\frac{Q}{T}\right), \tag{10.2}$$

where

$$Q = m_n - m_p = 1.2933 \text{ MeV}, \tag{10.3}$$

and the proton and neutron masses are $m_p = 938.2720$ MeV and $m_n = 939.5653$ MeV.

Since at this high temperature baryons are overwhelmingly in the form of neutrons and protons. The equilibrium baryon fraction of the neutron is given by

$$\begin{aligned} X_n^{(\text{eq})}(T) &\equiv \left(\frac{n_n(T)}{n_p(T) + n_n(T)}\right)_{\text{eq}} \\ &= \frac{1}{1 + e^{Q/T}}. \end{aligned} \tag{10.4}$$

We will verify this relation with a quantitative calculation later. Equation (10.2) forms the initial condition which allows us to estimate the neutron fraction at the temperature of freeze out. We need to calculate the universe expansion rate and the $n - p$ reaction rate.

10.2.1 Cosmic expansion rate

First we consider the cosmos expansion rate, the Hubble expansion parameter. The universe is now dominated by relativistic particles: photons, electrons, positrons, and the three families of neutrinos and antineutrinos, and they are in equilibrium at the same temperature. The total internal degrees of freedom for the energy density content of the universe is given by Eq. (9.154). Because of the mass information provided by neutrino oscillation and the upper bound of the electron neutrino, we know that the masses of all three flavors of neutrinos are of the order of electron volt or smaller. So they are ultra-relativistic. Then by Eqs. (9.120) and (9.36), ignoring the cosmological constant and the curvature term, the total energy density and the Hubble expansion rate are

$$\rho = g_* \frac{\pi^2}{30} T^4 \tag{10.5}$$

and

$$\begin{aligned} H &= \frac{\dot{a}}{a} = \sqrt{\frac{8\pi G_N}{3} \rho} \\ &= \sqrt{\frac{4\pi^3}{45} \sqrt{g_*} \frac{T^2}{M_P}} \\ &= 1.66015 \sqrt{g_*} \frac{T^2}{M_P}, \end{aligned} \tag{10.6}$$

where M_P is the Planck mass given in the natural units by

$$M_P = \frac{1}{\sqrt{G_N}} = 1.2209 \times 10^{22} \text{ MeV}. \tag{10.7}$$

Generally speaking, the Hubble expansion rate is suppressed by a factor of the Planck mass. Expressing energies in units of MeV and convert energy into second inverse, we have

$$H = 0.20658 \sqrt{g_*} \left(\frac{T}{1\text{MeV}}\right)^2 \text{ s}^{-1}, \tag{10.8}$$

which has been given earlier in Eq. (9.145). Below $T = 100$ MeV and till m_e with $g_* = 43/4$ as given in Eq. (9.154) we obtain $H = 6.7732 \times 10^3 \text{ sec}^{-1}$. Below $T = m_e$, the electron and positron degrees of freedom are removed and we have $g_* = 29/4$ which makes H just about 18% smaller.

10.2.2 Interaction rate

Next we estimate the rate of the reactions listed in Eq. (10.1) which maintain the equilibrium state among the nucleons, i.e., protons and neutrons. The average rate can be written as

$$\Gamma_{np} = n_\ell \langle v_\ell \sigma_{np} \rangle, \tag{10.9}$$

where n_ℓ is the lepton, i.e., electron or neutrino, number density, v_ℓ the lepton velocity, and σ_{np} the low energy leptonic nucleon weak interaction scatter cross section for Eq. (10.1). At this energy the leptons are relativistic, hence n_ℓ takes the form of Eq. (9.119), and their energy is the order of T , then $n_\ell \sim T^3$. The velocity is $v_\ell \sim c = 1$. From Eqs. (B.12) and (B.13) where the lepton energy is the order of the heat bath temperature T , the weak cross section is given by $\sigma \sim G_F^2 T^2$. Therefore, we have⁴⁷⁾

$$\Gamma_{np} = \xi_{np} G_F^2 T^5, \tag{10.10}$$

where $G_F = 1.16637 \times 10^{-5} \text{ GeV}^{-2} = 1.16637 \times 10^{-11} \text{ MeV}^{-2}$ is the Fermi constant. ξ_{np} is a numerical constant. As can be seen from Eqs. (B.12) and (B.13) it is of the order unit, e.g., $\xi_{np} = (g_V^2 + g_A^2)/\pi \approx 1.83$.

10.2.3 Approximate decoupling temperature T_f and neutron freezing fraction

From Eqs. (10.8) and (10.10), we see that the reaction rate has a temperature dependence of T^5 while the cosmos expansion rate depends on T^2 . At sufficiently high temperatures which exit in the early universe the reaction rate overwhelms the cosmos expansion, and at low enough temperatures the reverse is true. The critical temperature T_f below which the nucleons species begin to freeze out and decouple from each other. The freeze out temperature is determined by the condition $\Gamma_{np} = H$, which gives

$$\begin{aligned} T_f &= \left(\frac{4\pi^3}{45} g_* \right)^{1/6} \left(\frac{1}{\xi_{np}} \right)^{1/3} \left(\frac{1}{M_P G_F^2} \right)^{1/3} \\ &= 1.000 \left(\frac{\sqrt{g_*}}{\xi_{np}} \right)^{1/3} \text{ MeV}. \end{aligned} \tag{10.11}$$

⁴⁷⁾In the formula below we should multiple the right-handed side by a factor $\xi_\ell = g_\ell \zeta(3)/(3/4\pi^2)$, where $g_\ell = 2$ and $\zeta(3) = 1.202$, which appears in the formula of the leptonic number density. But for a quick estimate, we ignore it.

The decoupling temperature is the MeV order. From the above discussions on the values of g_* and ξ_{np} , we have

$$T_f \approx 1.2 \text{ MeV}. \tag{10.12}$$

The neutron to proton density ratio, Eq. (10.2), is of the order of 0.34. The baryon fraction of the neutron Eq. (10.4) is

$$X_n(T_f)_{\text{eq}} \approx 0.26. \tag{10.13}$$

So there is a significant number of neutrons remained to participate in the primordial nucleosynthesis. This estimate is rather crude and a more careful calculation is required.

10.3 Differential equation and initial condition for the neutron fraction

Now we make a quantitative calculation of the freezing fraction of the neutron. At the starting temperature of $T = 100^\circ \text{ K}$, the neutron to proton ratio is close to unity, but decreases exponentially when the universe expands and the temperature deceases. Let us form the respective neutron and proton fractions with respect to the total baryon number density, taking into account of the fact that at this temperature the baryons are exclusively in the forms of nucleons, i.e., neutrons plus protons,

$$\begin{aligned} X_n(T) &= \frac{n_n(T)}{n_p(T) + n_n(T)} \\ X_p(T) &= \frac{n_p(T)}{n_p(T) + n_n(T)} = 1 - X_n(T). \end{aligned} \tag{10.14}$$

We focus on the properties of the neutron fraction $X_n(T)$ which has the equilibrium value, approximating the value $m_n/m_p \approx 1$,

$$\begin{aligned} X_n^{(\text{eq})}(T) &= \frac{1}{1 + \left(\frac{n_p(T)}{n_n(T)} \right)_{\text{eq}}} \\ &= \frac{1}{1 + \exp(Q/T)}. \end{aligned} \tag{10.15}$$

The above expressions have already been discussed in Appendix B.2 where we estimate the nucleon decoupling temperature. The formation of nuclei which are stable under dissociation will take some time to happen after the temperature drops much blow $T = 100$ MeV. A differential equation for $X_n(T)$ will be derived below, under the effects of the reactions of Eq. (10.1) which cause conversions between the neutron and proton.

Notice that We have expressed the nucleon densities and their fractions in the above as functions of the temperature. However, they can be expressed as functions of the time, as the temperature is a function of the time t of the cosmic evolution. So the two types of expressions, $X_n(T)$, etc., and $X_n(t)$, etc., are equivalent and we use them interchangeably.

There are advantages in considering the baryon fraction into the neutron fraction $X_n(T)$ and the proton fraction $X_p(T)$, rather than the number density densities themselves. To deal with the densities we have to consider $a^3 n_n(T)$ and $a^3 n_p(T)$ which involves the complication of the time variation of the comoving volume. Furthermore, the controlling factors of the time variation of the baryon fractions are easier to see. Denoting the rate of conversion of a neutron into a proton by $\lambda_{n \rightarrow p}(t)$ and its inverse, the rate of a proton transferring into a neutron, by $\lambda_{p \rightarrow n}(t)$, we have the rate of change of the neutron fraction to be⁴⁸⁾

$$\frac{dX_n(t)}{dt} = -\lambda_{n \rightarrow p}(t)X_n(t) + \lambda_{p \rightarrow n}(t)X_p(t). \quad (10.16)$$

Similarly,

$$\frac{dX_p(t)}{dt} = -\lambda_{p \rightarrow n}(t)X_p(t) + \lambda_{n \rightarrow p}(t)X_n(t). \quad (10.17)$$

Clearly, we have

$$\frac{dX_n(t)}{dt} + \frac{dX_p(t)}{dt} = 0 \quad (10.18)$$

as expected. We will discuss the rates $\lambda_{n \rightarrow p}(t)$ and $\lambda_{p \rightarrow n}(t)$ in more detail in the following subsection. The differential equation for the rate of change of the proton fraction can be obtained from Eq. (10.16) by interchange the subscripts n and p . It is clearly that the sum of the rate of neutron and that of the proton vanishes, as it should be.

Using the second expression of Eq. (10.14), we obtain a first order differential equation for the rate of change of the neutron fraction,

$$\left(\frac{d}{dt} + \Lambda_{np}(t)\right) X_n(t) = \lambda_{p \rightarrow n}(t) \Lambda_{np}(t) = \lambda_{n \rightarrow p}(t) + \lambda_{p \rightarrow n}(t). \quad (10.19)$$

Suppose we are given the initial condition $X_n(t_0)$ at some time t_0 . A solution of differential equation of the following form is obtained:

$$\begin{aligned} X_n(t) &= I(t, t_0) \left(\int_{t_0}^t I(t', t_0)^{-1} \lambda_{p \rightarrow n}(t') dt' + X_n(t_0) \right) \\ &= \int_{t_0}^t I(t, t') \lambda_{p \rightarrow n}(t') dt' + I(t, t_0) X_n(t_0), \end{aligned} \quad (10.20)$$

where

$$I(t_2, t_1) = \exp\left(-\int_{t_1}^{t_2} \Lambda_{np}(t') dt'\right). \quad (10.21)$$

⁴⁸⁾For more discussions, see pp 65–66, [101]; pp 161–162, [221]; and p. 548, [218].

A description of how to obtain the solution is given in the footnote.⁴⁹⁾

We note that the function $I(t, t')$ satisfies the following identity:

$$I(t, t') = \frac{1}{\Lambda_{np}(t')} \frac{dI(t, t')}{dt} \quad (10.22)$$

Applying this identity in Eq. (10.20) and performing the partial integration, we can rewrite $X_n(t)$,

$$\begin{aligned} X_n(t) &= \frac{\lambda_{p \rightarrow n}(t)}{\Lambda_{np}(t)} - \int_{t_0}^t I(t, t') \frac{d}{dt'} \left(\frac{\lambda_{p \rightarrow n}(t')}{\Lambda_{np}(t')} \right) dt' + \\ &I(t, t_0) \left(X_n(t_0) - \frac{\lambda_{p \rightarrow n}(t_0)}{\Lambda_{np}(t_0)} \right). \end{aligned} \quad (10.23)$$

Equation (10.22) can be used repeatedly to write $X_n(t)$ in powers of time derivatives and $\Lambda_{np}(t)^{-1}$.

Let us examine Eq. (10.23) in some details. We define

$$\tilde{X}_n(t) = \frac{\lambda_{p \rightarrow n}(t)}{\Lambda_{np}(t)} - \int_{t_0}^t I(t, t') \frac{d}{dt'} \left(\frac{\lambda_{p \rightarrow n}(t')}{\Lambda_{np}(t')} \right) dt', \quad (10.24)$$

which consists of the first two terms of the right-handed side of Eq. (10.23) and is independent of the initial condition. It can be easily verify that $\tilde{X}_n(t)$ satisfies the same differential equation as $X_n(t)$, i.e.,

$$\frac{d\tilde{X}_n(t)}{dt} = -\Lambda_{np}(t)\tilde{X}_n(t) + \lambda_{p \rightarrow n}(t). \quad (10.25)$$

The difference between $\tilde{X}_n(t)$ and $X_n(t)$ is their initial conditions. As shown in Eq. (10.23), $X_n(t)$ has an arbitrary initial condition $X_n(t_0)$ at t_0 , while the initial condition of $\tilde{X}_n(t)$ is given by the nucleon conversion rates at t_0 :

$$\tilde{X}_n(t_0) = \frac{\lambda_{p \rightarrow n}(t_0)}{\Lambda_{np}(t_0)}. \quad (10.26)$$

⁴⁹⁾The form of the differential equation Eq. (10.19) suggests that we write the solution as

$$X_n(t) = I(t, t_0)g(t),$$

where $I(t, t_0)$ is given in Eq. (10.21). It is straightforward to show that the function $g(t)$ satisfies a simpler differential equation

$$\frac{dg(t)}{dt} = I(t, t_0)^{-1} \lambda_{p \rightarrow n}(t)$$

The solution for $g(t)$ is straightforward

$$g(t) = \int_{t_0}^t I(t', t_0)^{-1} \lambda_{p \rightarrow n}(t') dt' + \text{const.}$$

The initial condition, $X_n(t_0) = g(t_0)$, says that the “const.” above is just $X_n(t_0)$. Putting together everything concerned we obtain the solution Eq. (10.20).

Hence if $X_n(t)$ satisfies an initial condition as that of $\tilde{X}_n(t_0)$ above, $X_n(t)$ is just $\tilde{X}_n(t)$. From the numerical view point, as will be shown below, for $T > 10$ MeV the sum of the nucleon conversion rate $\Lambda_{np}(t)$ will be very large and therefore $I(t, t_0)$ is very small. So the initial condition dependent terms in Eqs. (10.20) and (10.23) are very small. Hence the solution is almost independent of the initial condition and $\tilde{X}_n(t)$ is practically the needed solution.

10.3.1 Nucleon conversion rates

The $n \rightleftharpoons p$ conversions go through weak interaction. As we have seen above, the rates $\lambda_{n \rightarrow p}$ and $\lambda_{p \rightarrow n}$ are essential in the calculation of the baryonic fraction of the neutron. We discuss them below.

For a order of magnitude estimate, an average interaction rate has been defined in Eq. (10.9), and the temperature is taken to be the energy scale of the scattering cross section. However, for a quantified calculation, we have to take into account of the energy of the individual particles involved and integrate over the phase space of the relevant particles. The temperature enters the calculation through the thermal distribution function. Let us describe in some details of the calculation of the nucleon conversion rates:

- The conversion rates $\lambda_{n \rightarrow p}$ and $\lambda_{p \rightarrow n}$ are for individual nucleons, integrated over all available leptons momenta.
- For the cosmic temperature below 100 MeV, the nucleons are non-relativistic with small kinetic energies. We can neglect the kinetic energy of the nucleon in the initial state as well as the recoil of the nucleon in the final state, so only their masses enter the calculation.
- The leptons involved are relativistic and we have to treat them properly. For the rate of individual nucleons, We have to integrate over the momentum spectrum of the lepton in the initial states, weighted by the momentum distribution determined by the thermal distribution function given in Eq. (9.98).
- For the lepton in the final state, in the stationary nucleons approximation, the final state lepton energy is determined by that of the initial lepton by energy conservation. There is also a factor $1 - f$ multiplying the cross section due to Pauli blocking, where f is the distribution function of the final state lepton. The Pauli blocking factor provides a counting of the unfilled states at a given energy, because the final state lepton is externally introduced into the system through nucleon conversion reactions and, therefore, the Pauli exclusion principle has to be imposed externally.

- Because of the highly relativistic nature of the lepton, the velocity of the lepton can be set to be $c = 1$.

Putting all these together and referring to the cross section formula in Eqs. (B.12) and (B.13) the expressions of the rate of conversion of the reactions of Eq. (10.1) are given below. For further details, we refer to [218]⁵⁰⁾

$$\begin{aligned} \lambda(n + \nu_e \rightarrow p + e^-) &= A \int p_e E_e p_\nu^2 dp_\nu f_\nu (1 - f_e), \\ \lambda(n + e^+ \rightarrow p + \bar{\nu}_e) &= A \int E_\nu^2 p_e^2 dp_e f_e (1 - f_\nu), \\ \lambda(n \rightarrow p + e^- + \bar{\nu}_e) &= A \int p_e E_e p_\nu^2 dp_\nu (1 - f_e)(1 - f_\nu), \end{aligned} \tag{10.27}$$

while the reversed conversion $p \rightarrow n$ by

$$\begin{aligned} \lambda(p + e^- \rightarrow n + \nu_e) &= A \int E_\nu^2 p_e^2 dp_e f_e (1 - f_\nu), \\ \lambda(p + \bar{\nu}_e \rightarrow n + e^+) &= A \int p_e E_e p_\nu^2 dp_\nu f_\nu (1 - f_e), \\ \lambda(p + e^- + \bar{\nu}_e \rightarrow n) &= A \int p_e E_e p_\nu^2 dp_\nu f_e f_\nu, \end{aligned} \tag{10.28}$$

where f_e is the thermal distribution function for the electron or positron and f_ν that of the electron neutrino or electron antineutrino, as given by the form of Eq. (9.98), and

$$\begin{aligned} A &\equiv = \frac{(g_V^2 + 3g_A^2)G_F^2}{2\pi^3} \cos^2 \theta_c \\ &= 1.8167 \times 10^{-2} \text{ MeV}^{-5} \cdot \text{s}^{-1}, \end{aligned} \tag{10.29}$$

where $g_V = 1$, $g_A^2 = 1.257$ is the correction to the axial vector charge of the nucleon, and θ_c is the Cabibbo angle, $\cos \theta_c = 0.9745$.⁵¹⁾ The energies of the final state electron (positron) and electron neutrino (electron antineutrino) are related as follows:

$$\begin{aligned} E_e &= Q + E_\nu, \quad \text{for } n + \nu_e \rightleftharpoons p + e^-, \\ E_\nu &= Q + E_e, \quad \text{for } n + e^+ \rightleftharpoons p + \bar{\nu}_e, \\ E_e &= Q - E_\nu, \quad \text{for } n \rightleftharpoons p + e^- + \bar{\nu}_e, \end{aligned} \tag{10.30}$$

where $Q = 1.2933$ MeV is the neutron-proton mass difference given in Eq. (10.3), which is the energy release in the neutron to proton conversion.

We note that the quantity A defined in Eq. (10.29) is made of two factors: a factor $G_F^2 (g_V^2 + 3g_A^2) \cos^2 \theta_c / \pi$ from the cross section formulas and another factor $1/(2\pi^2)$ from the momentum phase space integration

⁵⁰⁾See [218], pp 547–548.

⁵¹⁾To be more precise we follow [221] to include the Cabibbo angle which is a 5% effect. See, Eq. (3.2.6), p. 161, [221]. The contribution of the Cabibbo angle is omitted from the cross section formulae in Eqs. (B.12) and (B.13).

over the Fermi distribution function with the angular integration being performed. We do not distinguish distribution functions of the particle from the antiparticle, as we have neglected their chemical potential. However, the chemical potentials of the electron and positron cannot be strictly zero as there is a very small excess of number of electrons over positrons in the universe. However, this difference between the electron and positron will have little effect, hence we will omit the chemical potential in the discussion below.

In order to proceed with an analytic approach, some simplification has to be made in the calculation of the reaction rates. BBF made the following approximations:

- Using low temperature approximation such that the lepton energy is larger than the temperature. So we have

$$f_j \approx e^{-E_j/T}, \quad 1 - f_j \approx 1. \quad (10.31)$$

This kind of approximations will also be used in the derivation of the Boltzmann equation in the calculation of the massive particle decoupling given in Section 11 below.

- The above approximation allows us to extend the upper limit of the momentum integration to infinite to simplify the result. The cross section formulae used in Eqs. (10.27) and (10.28) are not valid for very large lepton momentum. But the error so introduced is negligible due to the exponential suppression of large lepton momenta present in the initial lepton distribution function. Then the neutron to proton conversion rates become

$$\begin{aligned} \lambda_B(n + \nu_e \rightarrow p + e^-) &\approx A \int_0^\infty p_e E_e p_\nu^2 dp_\nu e^{-E_\nu/T}, \\ \lambda_B(n + e^+ \rightarrow p + \bar{\nu}_e) &\approx A \int_0^\infty E_\nu^2 p_e^2 dp_e e^{-E_e/T}, \\ \lambda_B(n \rightarrow p + e^- + \bar{\nu}_e) &\approx A \int_0^{Q-m_e} p_e E_e p_\nu^2 dp_\nu, \end{aligned} \quad (10.32)$$

where the subscript “B” denotes the BBF approximation. The approximate expression of the neutron to proton conversion rate is the sum of the three expressions above, i.e.,

$$\begin{aligned} \lambda_B(n \rightarrow p) &= \lambda_B(n + \nu_e \rightarrow p + e^-) \\ &+ \lambda_B(n + e^+ \rightarrow p + \bar{\nu}_e) \\ &+ \lambda_B(n \rightarrow p + e^- + \bar{\nu}_e). \end{aligned} \quad (10.33)$$

- As to be demonstrated later, the neutron decay channel contributes very little when the temperature is greater than 0.3 MeV due to the long lifetime

of the neutron, $\tau_n = 885.7 \pm 0.5$ sec. So we will first neglect neutron decay process to further simplify the calculation of the nucleon conversion rate.

$$\begin{aligned} \lambda_{B1}(n \rightarrow p) &\equiv \lambda_B(n + \nu_e \rightarrow p + e^-) \\ &+ \lambda_B(n + e^+ \rightarrow p + \bar{\nu}_e). \end{aligned} \quad (10.34)$$

Similarly for the proton to neutron conversion rate. The omission of the neutron decay effect allows us to demonstrate more clearly the decoupling between neutron and proton. We will put back the neutron decay effect when we later calculate the neutron fraction available for deuterium formation.

Before we proceed further with the approximation we pause to examine how good the approximation is by comparing numerically the BBF approximate, i.e., Eq. (10.33) including the neutron decay channel or Eq. (10.34) omitting the neutron decay effect, with the exact expressions given by Eqs. (10.27) and (10.28). The rate expressions of the latter are given in [218] and [221]. As already stated before, the BBF approximation allows us to obtain an analytic result and the physics is more transparent when we put the neutron decay channel back into the nucleon conversion rates.

Compact expressions for the exact nucleon conversion rates, including the neutron decay, Eqs. (10.27) and (10.28) are given in [218] and [221].⁵²⁾ We reproduce them below with a slight simplification by means of the change of the variable of integration,⁵³⁾

$$\begin{aligned} \lambda_W(n \rightarrow p) &= A \int_{m_e}^\infty \left(1 - \frac{m_e^2}{q^2}\right)^{1/2} \left\{ \frac{(q-Q)^2 q^2}{[1 + e^{(q-Q)/T_\nu}] (1 + e^{-q/T_\gamma})} \right. \\ &\quad \left. + \frac{(q+Q)^2 q^2}{[1 + e^{-(q+Q)/T_\nu}] (1 + e^{q/T_\gamma})} \right\} dq, \end{aligned} \quad (10.35)$$

and

$$\begin{aligned} \lambda_W(p \rightarrow n) &= A \int_{m_e}^\infty \left(1 - \frac{m_e^2}{q^2}\right)^{1/2} \left\{ \frac{(q-Q)^2 q^2}{[1 + e^{-(q-Q)/T_\nu}] (1 + e^{q/T_\gamma})} \right. \\ &\quad \left. + \frac{(q+Q)^2 q^2}{[1 + e^{(q+Q)/T_\nu}] (1 + e^{-q/T_\gamma})} \right\} dq \\ &= A e^{-\left(\frac{Q}{T_\nu}\right)} \int_{m_e}^\infty \left(1 - \frac{m_e^2}{q^2}\right)^{1/2} \\ &\quad \left\{ \frac{(q-Q)^2 q^2}{[1 + e^{(q-Q)/T_\nu}] (1 + e^{-q/T_\gamma})} e^{q\left(\frac{1}{T_\nu} - \frac{1}{T_\gamma}\right)} \right. \end{aligned}$$

⁵²⁾[218], p. 348 and [221], pp 160–161, in particular, Eqs. (3.2.4) and (3.2.5). A description of how the expressions are obtained can be found in [218], pp 546–548.

⁵³⁾The expression in Eqs. (10.35) and (10.36) are obtained from n-p conversion rates of [218] and [221] by making the change of the variable of integration, $q \rightarrow q - Q$.

$$+ \frac{(q+Q)^2 q^2}{[1 + e^{-(q+Q)/T_\nu}] (1 + e^{q/T_\gamma})} e^{-q(\frac{1}{T_\nu} - \frac{1}{T_\gamma})} \Bigg\} dq, \tag{10.36}$$

where the subscript “W” denotes expressions lifted off [218] and [221]. T_ν denotes the temperature the neutrino and T_γ that of the photon. As discuss in Section 9, the two temperatures start to depart from each other no more than 10% at the temperature of 0.4 MeV and reach the relation $T_\nu = (4/11)^{1/3} T_\gamma$ to within 10% at about $T = 0.1$ MeV. In light of our approximation we will not distinguish the two temperature and set $T_\nu = T_\gamma = T$. We note that $T_\nu = T_\gamma$ is valid for the cosmic temperature above a few MeV. Then the expressions given in Eqs. (10.35) and (10.36) satisfies the relation of the detailed balance

$$\lambda_W(p \rightarrow n)|_{T_\nu=T_\gamma=T} = e^{-Q/T} \lambda_W(n \rightarrow p)|_{T_\nu=T_\gamma=T}, \tag{10.37}$$

i.e.,

$$e^{-m_p/T} \lambda_W(p \rightarrow n)|_{T_\nu=T_\gamma=T} = e^{-m_n/T} \lambda_W(n \rightarrow p)|_{T_\nu=T_\gamma=T}, \tag{10.38}$$

We can rewrite Eqs. (10.35) and (10.36) as

$$\begin{aligned} & \lambda_W(n \rightarrow p) \\ &= AT^5 \int_x^\infty \left(1 - \frac{x^2}{z^2}\right)^{1/2} \left\{ \frac{[z - (Q/m_e)x]^2 z^2}{[1 + e^{(z-(Q/m_e)x)](1 + e^{-z})}] (1 + e^{-z})} \right. \\ & \quad \left. + \frac{[z + (Q/m_e)x]^2 z^2}{[1 + e^{-(z+(Q/m_e)x)](1 + e^z)] (1 + e^z)} \right\} dz \end{aligned} \tag{10.39}$$

and

$$\begin{aligned} \lambda_W(p \rightarrow n) &= AT^5 e^{-(Q/m_e)x} \int_x^\infty \left(1 - \frac{x^2}{z^2}\right)^{1/2} \\ & \quad \times \left\{ \frac{[z - (Q/m_e)x]^2 z^2}{[1 + e^{(z-(Q/m_e)x)](1 + e^{-z})}] (1 + e^{-z})} \right. \\ & \quad \left. + \frac{[z + (Q/m_e)x]^2 z^2}{[1 + e^{-(z+(Q/m_e)x)](1 + e^z)] (1 + e^z)} \right\} dz, \\ &= e^{-(Q/m_e)x} \lambda_W(n \rightarrow p), \end{aligned} \tag{10.40}$$

where $x \equiv m_e/T$ and the integration variable is changed to $z \equiv q/T$.

First let us look at the high temperature behavior of the nucleon conversion rates. For $T \gg m_e$, we can set $x = 0$ in Eq. (10.39) to obtain⁵⁴⁾

$$\begin{aligned} \lambda_W(n \rightarrow p)|_{T \gg m_e} &= \frac{7\pi^4}{15} AT^5, \\ \lambda_W(p \rightarrow n)|_{T \gg m_e} &= \frac{7\pi^4}{15} AT^5 e^{-Q/T}. \end{aligned} \tag{10.41}$$

⁵⁴⁾The resultant integrals by setting $x = 0$ in Eq. (10.39) are elementary to perform when the following identity is applied: $\int_0^\infty \frac{y^4}{(1+e^y)(1+e^{-y})} dy = \frac{7\pi^4}{30}$.

If the small term Q/T in the exponential $e^{-Q/T}$ is neglected, the proton to neutron and neutron to proton rates are the same as expected.

The high temperature form of the BBF approximation Eq. (10.32) can be obtained similarly by dropping its dependence on m_e and Q . The integrals can be done straightforwardly and we obtain

$$\begin{aligned} \lambda_B(n + \nu_e \rightarrow p + e^-)|_{T \gg m_e} &\approx \lambda_B(n + e^+ \rightarrow p + \hat{\nu}_e)|_{T \gg m_e} \\ &= (4!) AT^5 \end{aligned} \tag{10.42}$$

So in the BBF approximation the high temperature expression for the neutron to proton conversion is

$$\lambda_{B1}(n \rightarrow p) = 2(4!) AT^5, \tag{10.43}$$

which agrees with the exact expression given in Eq. (10.41) within 6%.

At sufficiently low temperature, the two-body reactions contribution to the neutron-proton conversion rates are suppressed exponentially and the conversion rate is dominated by the neutron decay process. Let us consider this case of the neutron to proton conversion. For the BBF approximation the low temperature behavior given in Eq. (10.32) is dominated by the neutron decay channel, as the contributions of scattering channels are suppressed exponentially by $e^{-E/T}$. Numerically we have

$$\begin{aligned} \lambda_B(n \rightarrow p)|_{T \ll m_e} &\approx \lambda_B(n \rightarrow p + e^- + \bar{\nu}_e) \\ &\approx 1.04 \times 10^{-3} \text{ s}^{-1}. \end{aligned} \tag{10.44}$$

We can estimate the low temperature behavior of the Weinberg expression given in Eq. (10.39) by considering the case $x \gg 1$. The dominant contribution is given by

$$\begin{aligned} \lambda_W(n \rightarrow p)|_{T \ll m_e} &\approx A \int_{m_e}^Q (q^2 - m_e^2)^{1/2} q(Q - q)^2 dq \\ &= 1.13 \times 10^{-3} \text{ s}^{-1}. \end{aligned} \tag{10.45}$$

The two estimates agree within in 9%. Hence we can conclude that the BBF approximation and that the exact expression for the neutron to proton conversion rates agree better than 10%. We can make similar conclusion for the proton to neutron conversion rate.

In Fig. 10.1 we plot the neutron to proton conversion rate using the value of $A = 1.8167 \times 10^{-2} \text{ MeV}^{-5} \cdot \text{s}^{-1}$ as given in Eq. (10.29). The horizontal axis is the temperature in units of MeV, where $1 \text{ MeV} = 1.1605 \times 10^{10} \text{ K}$. The vertical axis is the neutron to proton conversion rate in second inverse. The meanings of the various curves are explained in the caption. We note particularly the dotted (red) and dashed (blue) curves. The latter is the high temperature expression $\lambda_W(n \rightarrow p)|_{T \gg m_e}$ given in Eq. (10.41) and the former the exact expression $\lambda_W(n \rightarrow p)$ of Eq. (10.35). The two curves are close to

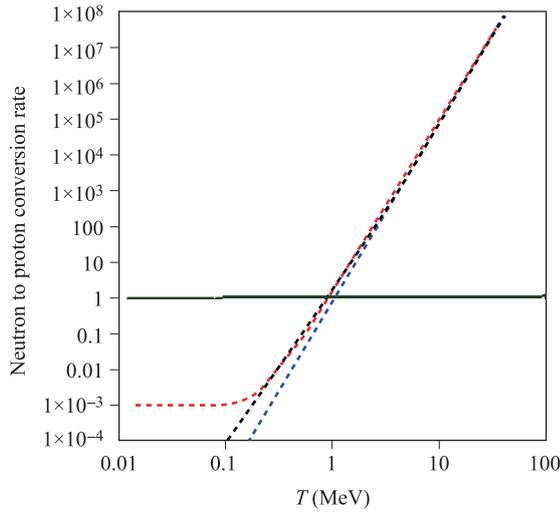


Fig. 10.1 The neutron to proton rate $\lambda(n \rightarrow p)$. The horizontal axis is the temperature in units of MeV (1 MeV = 1.1605×10^{10} K). The various curves are, the dotted (red) curve: the Weinberg exact expression given in Eq. (10.35), the dashed curve (blue): The high temperature approximation given in Eq. (10.41); the dash-dotted (black) curve: The BBF approximate expression without the neutron decay contribution as given in (10.34); The solid (dark green, almost horizontal) curve: The ratio of the BBF full approximation given in Eq. (10.32) to the Weinberg expression Eq. (10.35).

each other down to $T = 1$ MeV, and the high temperature expression is exactly valid for $T \geq 5$ MeV. Note that $T = 5$ MeV is only about 4 times of the neutron-proton mass difference Q which is the largest energy scale in the expression. We also note the exact conversion rate as given by the dotted (red) curve becomes constant for $T < 0.1$ MeV, with the value $1.04 \times 10^{-3} \text{ s}^{-1}$. This constant value is dominated by the contribution of the neutron decay process.

A numerical comparison is made of the full approximation of BBF as given in Eq. (10.32) with the exact expressions of [221] and [218] as given in Eq. (10.35). The ratio of the former over the latter is plotted as the solid (dark green, almost horizontal) curve in Fig. 10.1. The former is larger by no more than 10%. So the use of the BBF approximation is justified within such an approximation. So much for the digression, we can now proceed with Eq. (10.23).

10.3.2 Solution of the differential equation

To obtain an analytic solution for the neutron fraction as a function of the temperature we have to make further simplification of the neutron-proton conversion rates. With the approximate expressions of BBF given in Eq. (10.32) together with Eq. (10.30), and further setting $m_e = 0$, the integrals of the first two expressions in

Eq. (10.32) can be analytically integrated:

$$\begin{aligned} \lambda(n + \nu_e \rightarrow p + e^-) &= \lambda(n + \nu_e \rightarrow p + e^-) \\ &= AT^3(24T^2 + 12QT + 2Q^2) \end{aligned} \quad (10.46)$$

The limit of setting $m_e = 0$ has almost no effect on the rate $\lambda(n + \nu_e \rightarrow p + e^-)$ even down to low temperature of 0.01 MeV, neither will it significantly affect the rate $\lambda(n + e^+ \rightarrow p + \bar{\nu}_e)$ for $T > 1$ MeV. The difference at $T = m_e$ is less than 6%.⁵⁵⁾

The neutron decay contribution, neglecting the Pauli blocking factors of the final state electron and neutrino, is just the neutron decay width, or the inverse of the neutron lifetime. Here we have to keep the electron mass. Performing the integral of the third expression of Eq. (10.32), we obtain the BBF expression⁵⁶⁾,

$$\begin{aligned} \Gamma_n &= \lambda(n \rightarrow p + e^- + \bar{\nu}_e) \\ &= A \int_0^\Delta p_\nu E_\nu p_e^2 dp_e \\ &= A \left(\frac{1}{5} \Delta \left[\frac{1}{6} Q^4 - \frac{3}{4} m_e^2 Q^2 - \frac{2}{3} m_e^4 \right] \right. \\ &\quad \left. + \frac{1}{4} m_e^4 Q \ln \frac{Q + \Delta}{m_e} \right), \end{aligned} \quad (10.47)$$

where $\Delta \equiv \sqrt{Q^2 - m_e^2}$. The numerical value of Γ_n is already given in Eq. (10.44). The neutron lifetime thus calculated is

$$\begin{aligned} \tau_n^{\text{Th}} &= \frac{1}{\Gamma_n} \\ &= 965.6 \text{ s}. \end{aligned} \quad (10.48)$$

The experimental value of the neutron life-time found in [11] is

$$\tau_n^{\text{exp}} = 885.7 \pm 0.8 \text{ s}. \quad (10.49)$$

The calculated value Eq. (10.44) is 9% too large. To correct this error, we follow BBF to normalize the value of A by Eq. (10.47) by using the inverse of the experimental neutron lifetime Eq. (10.49) for the left-handed side. Hence we define A , following BBF,⁵⁷⁾

$$\begin{aligned} A &= \frac{a}{4\tau_n^{\text{exp}}} Q^{-5}, \\ a &= 254. \end{aligned} \quad (10.50)$$

⁵⁵⁾For much smaller T the difference for physics m_e and $m_e = 0$ for the rate $\lambda(n + e^+ \rightarrow p + \bar{\nu}_e)$ is very dramatic. However, their magnitudes are very small and, fortunately this region of the temperature does not contribute the physics of helium production.

⁵⁶⁾See Eq. (2.23), [251].

⁵⁷⁾BBF used the neutron lifetime of 986 ± 16 sec and hence had the value of a to be 255 which is very close to our value of 254 given in Eq. (10.50).

From Eqs. (10.46) and (10.50), we have the approximate neutron to proton conversion rate defined in Eq. (10.34) to be⁵⁸⁾

$$\lambda_{B1}(n \rightarrow p) = \frac{a}{\tau_n^{\text{exp}}} \left(\frac{T}{Q}\right)^3 \left[12 \left(\frac{T}{Q}\right)^2 + 6 \frac{T}{Q} + 1\right] \equiv \frac{a}{\tau_n^{\text{exp}}} \left(\frac{12}{y^5} + \frac{6}{y^4} + \frac{1}{y^3}\right), \quad (10.51)$$

where

$$y \equiv \frac{Q}{T}. \quad (10.52)$$

The proton to neutron conversion rate is

$$\lambda_{B1}(p \rightarrow n) = e^{-y} \lambda(n \rightarrow p)_{\text{BBF1}} = e^{-y} \frac{a}{\tau_n^{\text{exp}}} \left(\frac{12}{y^5} + \frac{6}{y^4} + \frac{1}{y^3}\right), \quad (10.53)$$

The sum of the $n \rightarrow p$ and $p \rightarrow n$ rates, defined in the second expression of Eq. (10.19), is given by

$$\Lambda_{\text{np1}} = \frac{a}{\tau_n^{\text{exp}}} \left(\frac{12}{y^5} + \frac{6}{y^4} + \frac{1}{y^3}\right) (1 + e^{-y}), \quad (10.54)$$

We now proceed to calculate \tilde{X} defined in Eq. (10.24). First we calculate the function $I(t, t')$ defined in Eq. (10.21). It is more convenient to rewrite $\Lambda_{\text{np1}}(t)$ and $I(t, t')$ as functions of the variable $y = Q/T$ as defined in Eq. (10.52),

$$I(y, y') = \exp\left(-\int_{y'}^{y'} \Lambda_{\text{np1}}(y'') \frac{dt'}{dy'} dy''\right), \quad (10.55)$$

where, from Eqs. (9.151) and (9.153), the Jacobian of the variable change is

$$\begin{aligned} \frac{dt}{dy} &= \frac{dt}{dT} \frac{dT}{dy} \\ &= \frac{Q}{y^2} \left[\frac{T}{3s(T)} \frac{ds(T)}{dT} \right] \frac{1}{TH} \\ &= \frac{M_P}{Q^2} \sqrt{\frac{45}{4\pi^3 g_*}} \left[\frac{T}{3s(T)} \frac{ds(T)}{dT} \right] y. \end{aligned} \quad (10.56)$$

For radiation dominance the entropy is proportional to T^3 , therefore the factor depending on the entropy, which is contained in the parentheses in the right-handed side of the above equation, is just unity. However, as discussed in §9.5 the situation is more complicated during this epoch around the $T = 1$ MeV, during which period some of the epic events of the cosmos took place. They include the decoupling of the neutrino and the annihilation of the electron and positron to reheat the photon gas. For a brief time period the entropy has a more complicated temperature dependence than the T^3 form and

the temperature of the neutrino and that of the photon start to depart continuously. However, for the interest of simplification so as to push an analytic solution as much as possible, we keep the T^3 form of the entropy. The complication in the entropy function does not spoil the validity of this simplification. Then the Jacobian is proportional to y

$$\frac{dt}{dy} = \frac{M_P}{Q^2} \left(\sqrt{\frac{45}{4\pi^3 g_*}}\right) y \quad (10.57)$$

and the integral in the exponent of Eq. (10.55) can be worked out,

$$\begin{aligned} &\int_{y'}^y \Lambda_{\text{np1}}(y'') \frac{dt}{dy''} dy'' \\ &= \frac{M_P}{Q^2} \sqrt{\frac{45}{4\pi^3 g_*}} \frac{a}{\tau_n^{\text{exp}}} \int_{y'}^y \left(\frac{12}{y''^4} + \frac{6}{y''^3} + \frac{1}{y''^2}\right) (1 + e^{-y''}) dy'' \\ &\equiv -b [\mathcal{K}(y) - \mathcal{K}(y')], \end{aligned} \quad (10.58)$$

where, because of the specific form of the integrant, the integral has an analytic expression,

$$\begin{aligned} \mathcal{K}(y) &= \frac{4}{y^3} + \frac{3}{y^2} + \frac{1}{y} + \left(\frac{4}{y^3} + \frac{1}{y^2}\right) e^{-y} \\ b &\equiv \frac{M_P}{Q^2} \sqrt{\frac{45}{4\pi^3 g_*}} \frac{a}{\tau_n^{\text{exp}}}. \end{aligned} \quad (10.59)$$

Then

$$I(y, y') = \exp(b(\mathcal{K}(y) - \mathcal{K}(y'))). \quad (10.60)$$

Let us come back to $X_n(t)$ given in Eq. (10.23) again to clarify a point on the initial condition raised earlier. We can now see the justification of the omission of the third term on the right-handed side of Eq. (10.23) so that $X_n(t)$ is practically $\tilde{X}_n(x)$ given in (10.24). The omitted term is $I(t, t_0)$ times a factor which is of order one. Since the initial time t_0 is very small and so the corresponding cosmic temperature T_0 is very high and so y_0 is very small, then $I(t, t_0)$ or $I(y, y_0)$ is highly suppressed for almost any t (y) after, but not too close to, t_0 (y_0). So the arbitrary initial condition will not influence the solution of the differential equation.

Another clarification we would like to make concerns the solution $\tilde{X}_n(t)$, Eq. (10.24). As we have discussed earlier, neglecting the neutrino and photon temperature difference, the neutron to proton and the proton to neutron conversion rates are related by detailed balance as shown in Eqs. (10.39) and (10.40). The first term on the right-handed side of (10.24) is just $\tilde{X}_n(t_0)$, i.e., the initial value of $\tilde{X}_n(t)$,

$$\tilde{X}_n(t_0) \equiv \frac{\lambda_{p \rightarrow n}(t_0)}{\Lambda_{\text{np}}(t_0)} = \frac{1}{1 + e^{y_0}}, \quad (10.61)$$

⁵⁸⁾This is Eq. (2.27) of BBF.

which is the equilibrium solution of $X_n(t)$ given in (10.4)

$$X_n^{(\text{eq})}(t) \equiv \frac{\lambda_{p \rightarrow n}(t)}{A_{\text{np}}(t)} = \frac{1}{1 + e^y}. \tag{10.62}$$

BBF uses the approximation of the photon and neutrino temperature degeneracy and neglects the neutron decay to demonstrate the approach of decoupling of the proton and neutron system at some temperature, and shows that their densities ratio becomes constant when temperature decreases further. Starting at $T \gg Q$ and hence $y \ll 1$, we have $X_n(t) \approx X_n^{(\text{eq})} \approx 1/2$. When time increases to cause T to decrease so that y is not significantly different from 1, then $X_n(t)$ will move away from the equilibrium solution and reaches a constant solution of the freeze out value. For additional discussions we refer to [221]⁵⁹⁾.

10.4 Freezing out neutron fraction

We can now write down the solution for the neutron fraction, \tilde{X}_n , expressed as a function of temperature through the variable y . We first note from Eq. (10.62)

$$\begin{aligned} \frac{d}{dy} \left(\frac{\lambda_{p \rightarrow n}(y)}{A_{\text{np}}(y)} \right) &= -e^y X_n^{(\text{eq})}(y)^2 \\ &= -\frac{e^y}{(1 + e^y)^2}. \end{aligned} \tag{10.63}$$

Then Eq. (10.24) gives

$$\begin{aligned} \tilde{X}_n(y) &= X_n^{(\text{eq})}(y) + \int_{y_0}^y e^{y'} X_n^{(\text{eq})}(y')^2 \exp(b(\mathcal{K}(y) - \mathcal{K}(y'))) dy' \\ &= \frac{1}{1 + e^y} + \int_{y_0}^y \frac{e^{y'}}{(1 + e^{y'})^2} \exp(b(\mathcal{K}(y) - \mathcal{K}(y'))) dy'. \end{aligned} \tag{10.64}$$

The initial condition is given by

$$\begin{aligned} \tilde{X}_n(y_0) &= X_n^{(\text{eq})}(y_0) \\ &= \frac{1}{1 + e^{y_0}}. \end{aligned} \tag{10.65}$$

We plot the neutron fraction in Fig. 10.2 as both a function of $y = Q/T$ and that of the temperature T . Since the initial temperature is very high we take $y_0 = 0$. We have use the experimental neutron lifetime $\tau_n \approx 886$ s and the total particle internal degree freedom $g_* = 43/4$ in the approximate case of degenerated photon and neutrino temperature.

Let us notice some of the properties of the neutron fraction as given above. At the temperature of $T = 10^{12}$ K or about 86 MeV which is much greater than the neutron and proton mass difference, the neutron fraction

is about 1/2, i.e., the baryon content is composed of about equal number of proton and neutrons. As discussed above the solution \tilde{X}_n tracks the equilibrium solution at high temperatures down to $T = 3 \times 10^{10}$ K or about 2.6 MeV where $\tilde{X}_n \approx 0.38$. When the temperature decreases further, \tilde{X}_n departs from the equilibrium value and approaches to an asymptotic value, while the equilibrium value goes exponentially to zero.

The significance of the initial condition Eq. (10.65) is pointed out in [218] and [221].⁶⁰⁾ This is an interesting point worthy of further elaboration. The initial condition Eq. (10.65) is independent of any detailed model of the early universe and does not require the choice of a specific function in order to start the process of nucleosynthesis, even though such a possibility exists according to Eq. (10.23). We can trace back the origin of the independence of the initial condition as follows: The singular behavior⁶¹⁾ of the \mathcal{K} function Eq. (10.59) at very small time and therefore very large temperature washes out whatever the initial condition is. The statistical thermal distribution of the particle contents of the early universe at high temperature can provide its own initial condition as shown in Eq. (10.24). It is also important to note that the self-consistent initial condition makes it possible for the leptons involved to have a very small or vanishing chemical potential.

The asymptotic value of the neutron fraction is what we are after for. As given in \tilde{X}_n above, it is

$$\tilde{X}_n(\infty) = 0.143, \tag{10.66}$$

which is approached slowly at low temperatures from the value of slightly over 0.15 at $T = 10^9$ K or 0.1 MeV. This asymptotic value depends on the input parameters and we discuss below their effects and possible physics implications.

One parameters is the particle species internal degrees of freedom g_* which determines the energy content of the universe at the time when the nucleosynthesis starts to takes place. The other is the neutron lifetime τ_n which in the present approach represents the weak interaction strength. Let us consider the effect of the latter first. Increasing or decreasing the neutron lifetime means respectively to decrease or increase the weak interaction strength. If we decrease the weak interaction strength, the neutron-proton conversion rates are weakened and, as the conversion rate depends on T^5 , the departure from the equilibrium solution will take place at a higher temperature and the asymptotic value of the neutron fraction will be larger. Since the weak interaction strength can be measured accurately in the terrestrial laboratory,

⁶⁰⁾See the discussions in [218] below Eq. (15.7.21) on p. 550, and [221] below Eq. (3.2.13) on p. 163.

⁶¹⁾The origin of the self-consistent initial condition was pointed out in [254], see Ref. 109a, Section 15, [218].

⁵⁹⁾See the discussion below Eq. (3.2.12), pp 162–163, [221].

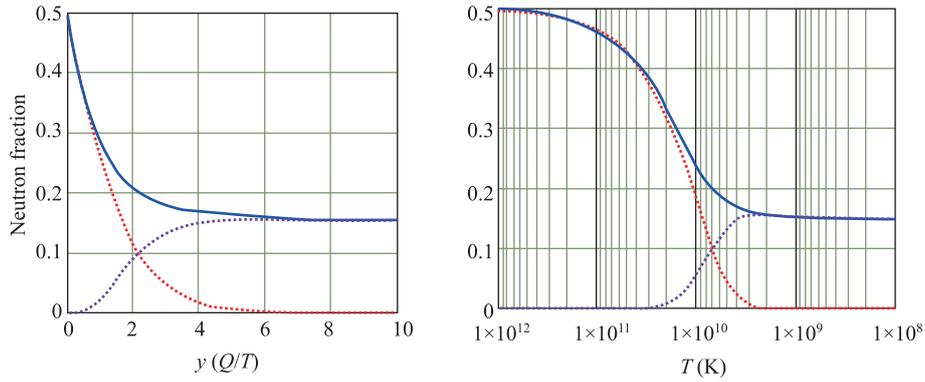


Fig. 10.2 The neutron fraction \tilde{X}_n as a function y in the left panel and a function of the temperature T in the right panel. The solid curves (blue) are the neutron fraction \tilde{X}_n , the dotted curves (red) are the equilibrium solution $X_n^{(eq)}$, and the dashed curve (brown) the second term of the right-handed side of the neutron fraction solution Eq. (10.64).

if it was determined that one had to adjust the weak interaction strength in order to gain a proper understanding of the helium production in the early universe, it would suggest that the weak interaction strength might be changed in the environment of high temperature heat bath so that the interaction strength has a temperature dependence. The terrestrial laboratory can be regarded as a very low temperature heat bath.

Another parameter is the particle degrees of freedom g_* which enters indirectly at the very late stage of the calculation of X_n , as the Jacobian of the variable change from time to temperature which is given in Eq. (10.56). Larger g_* , which means the universe has a high energy content, leads to a faster decrease of temperature as time increase. This implies that reaction rates will generally decrease faster. So, the departure from the equilibrium solution will again happen at a higher temperature and the asymptotic value of X_n is higher.

Numerical studies of \tilde{X}_n bears out the above physical arguments: increasing τ_n or/and g_* increase the value of y where the solution departs from that of the equilibrium and therefore a higher asymptotic value for \tilde{x}_n will be obtained. Decreasing τ_n or/and g_* will have the opposite effect. The effect of g_* on the primordial helium production has been used effectively on constraining the number of neutrino flavors [255]. Each flavor of neutrino contributes to g_* by a factor of 7/4. See [256] for a review and additional references.

The result given in Eq. (10.66) is the important first part of the story. In order to arrive at a proper prediction of the helium abundance We need to know the effect of the decay of the neutron in the neutron-proton conversion rates.

10.5 Effect of neutron decay

The effect of the neutron decay is included in the BBF treatment by multiplying \tilde{X}_n by the exponential decay

factor,

$$\tilde{X}_n(y) \rightarrow e^{-t/\tau_n} \tilde{X}_n(y). \tag{10.67}$$

Here, we include the effect of the neutron decay by restoring directly its contribution to the total np conversion rate Λ_{np} in Eq. (10.24), since the ratio $\lambda_{p \rightarrow n}(y)/\Lambda_{np}(y)$ is not changed. This is done, from Eqs. (10.21), (10.24), (10.58), and (10.59), by putting back the decay contribution to the right-handed side of Eq. (10.64) with the following replacement:

$$I(y, y') \rightarrow \exp(b(\mathcal{K}(y) - \mathcal{K}(y'))) \xi_I(y, y'), \tag{10.68}$$

with

$$\begin{aligned} \xi_I(y, y') &= \exp\left(-\int_{t'}^t \lambda(n \rightarrow p + e^- + \bar{\nu}_e)(1 + e^{-y''}) dt''\right) \\ &= \exp(b(\delta\mathcal{K}(y) - \delta\mathcal{K}(y'))), \end{aligned} \tag{10.69}$$

where

$$\delta\mathcal{K}(y) \equiv \frac{1}{a} \left[-\frac{y^2}{2} + (1 + y)e^{-y} \right]. \tag{10.70}$$

The variable y corresponds to the time t and y' the time t' . So the effect of the inclusion of the neutron decay process is to replace the \mathcal{K} function in (10.60) by

$$\mathcal{K}(y) \rightarrow \hat{\mathcal{K}}(y) = \mathcal{K}(y) + \delta\mathcal{K}(y) \tag{10.71}$$

and

$$\begin{aligned} \tilde{X}_n(y) &\rightarrow \hat{X}_n^{(d)}(y) \\ &= \frac{1}{1 + e^y} + \int_{y_0}^y \frac{e^{y'}}{(1 + e^{y'})^2} \exp\left(b\left(\hat{\mathcal{K}}(y) - \hat{\mathcal{K}}(y')\right)\right) dy'. \end{aligned} \tag{10.72}$$

We plot the modified neutron fractions $\hat{X}_n^{(d)}$ and the original \tilde{X}_n in Fig. 10.3. The various curves are explained in the figure caption.

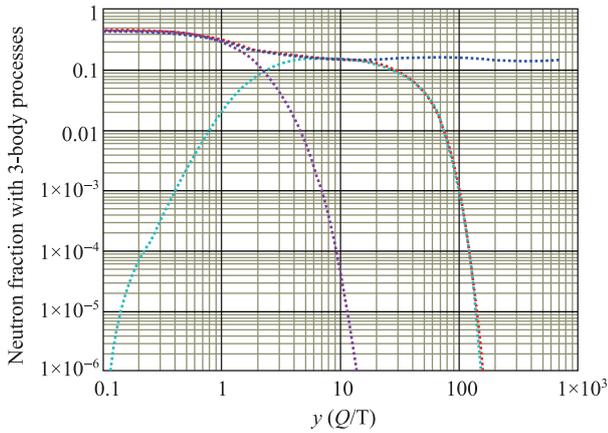


Fig. 10.3 The neutron fractions \tilde{X}_n and $\hat{X}_n^{(d)}$ as functions of y . $\hat{X}_n^{(d)}$ takes into account of the three body process of the neutron decay while \tilde{X}_n does not. The red solid curve: $\hat{X}_n^{(d)}$; The blue dotted curve: \tilde{X}_n ; The brown dashed curve: The equilibrium solution, i.e., the first term of the second line of Eq. (10.72); and the cyan dotted-dashed curve: the second term of the right-handed side of the second line of Eq. (10.72).

Comparing the red curve ($\hat{X}_n^{(d)}(y)$ including the effect with neutron decay) and the blue dotted curve ($\tilde{X}_n(y)$ without the neutron decay effect), we see that the two curves are on top of each other for $y < 10$ or $T > 1.5 \times 10^9$ K, so in this temperature region the neutron decay process has very little effect. Beyond $y > 30$ or $T < 5 \times 10^8$ K, the red curve $\hat{X}_n^{(d)}(y)$ is suppressed exponentially for larger y due to the neutron decay effect. Let us examine Eq. (10.72) more closely. The term $\exp(b\hat{K}(y))$ can be factored out of the integral. From Eqs. (10.71) and (10.70), we identify the term $-by^2/(2a)$ to be just the exponential factor due the neutron decay $-t/\tau_n$. Hence the integral on the right-handed side of Eq. (10.72) is proportional to $\exp(-t/\tau_n)$, recalling Eqs. (10.47) and (10.48). Actually one can readily see the appearance of the exponential decay term from Eq. (10.69). The first term on the right-handed side of the first line contributes to a term

$$\exp\left(-\frac{t-t'}{\tau_n}\right). \tag{10.73}$$

The first part, which is $\exp(-t/\tau_n)$, gives the exponential decay factor of the neutron and can be factored out of the integral. The second part $\exp(t'/\tau_n)$ which depends on the integration variable has to be integrated.

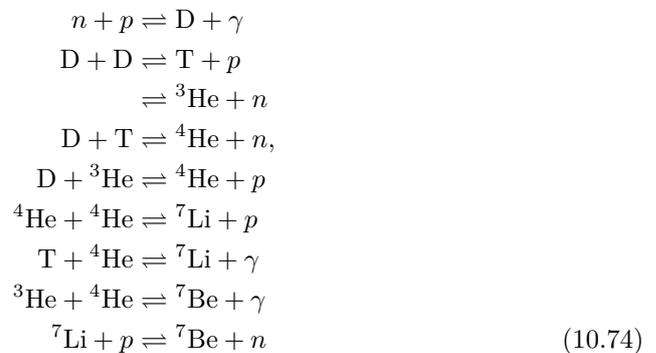
Let us make two remarks. The first is that BBF did not go into any details of the effect of the neutron decay channel other than stating that the inclusion of the effect of the neutron decay is to multiply the rate equation \tilde{X}_n by the exponential decay factor $\exp(-t/\tau_n)$, i.e., Eq. (10.67). So in BBF this exponential decay factor will also multiply the equilibrium solution. However, we see

in the above discussion that the exponential decay factor should not affect the equilibrium solution. However, the difference between the results with and without the exponential decay factor is small because of the rapid decreasing nature of the equilibrium solution at large y .

The second remark is that our solution including the neutron decay channel is close to the treatment of [221]. We reproduce approximately its Table 3.2⁶²⁾ down to the temperature of 6×10^8 K which corresponds to an energy of 0.05 MeV. Our result is larger when the temperature is further decreased. The origin of this difference is very likely due the fact that we have used a value for the number of degrees freedom g_* larger than it is warranted at this low temperature. At this lower temperature, the e^- and e^+ are rapidly disappear due to annihilation and the energy content of the universe is dominated by the photon and the neutrino where the neutrino has a lower temperature than the photon. So g_* should be taken as $2 + 6 \times (7/8) \times (4/11)^{4/3} = 3.363$ instead of $43/4 = 10.75$ which is used as stated. Numerical work shows that using the lower value of g_* brings our results more in lines with that of [221]. However, at higher temperature the use of the larger number of g_* is justified.

10.6 Arresting neutron decay – Formation of light elements

Nature has an ingenious mechanism to stop neutrons from decay by storing them in bound states through nuclear interactions to form complex nuclei. The formation process starts from the simplest multi-nucleon nucleus, i.e., the deuterium D (^2H) to go to more complicated ones, i.e., tritium T (^3H), helium-3 (^3He), helium-4 (^4He), etc.⁶³⁾ The reactions that form the light elements are two-body reactions given by



Although helium-4 can, in principle, be formed directly from the four-body collision involving two neutrons and

⁶²⁾See p. 164, [221].

⁶³⁾The binding energies of these light elements: the deuterium $\varepsilon_D = 2.23452$ MeV, the tritium $\varepsilon_T = 8.4818$ MeV, helium-3 $\varepsilon_{^3\text{He}} = 7.7180$, and the helium-4 $\varepsilon_{^4\text{He}} = 28.301$ MeV. For elements of higher atomic number, the binding energy per nucleon is usually given.

two protons in the initial state, i.e., $2n + 2p \rightleftharpoons {}^4\text{He} + \gamma$, the rate is too small to be effective, due to small cross sections, small phase space, and the four powers of a small nucleon density functions. So the formation of light elements starts with the formation of the deuterium which can be accomplished by two-body collisions. The nuclear reactions in the formation of light elements up to ${}^7\text{Be}$ and ${}^7\text{Li}$ are depicted graphically in Fig. 10.4 which we adopt from [257]. Although ${}^7\text{Be}$ can be formed in the process of primordial process but its half-life is only 53 days, so it is not one of the primordial light elements.

The reaction time of the deuterium formation $n + p \rightarrow D + \gamma$ is typically the electromagnetic interaction time of the order 10^{-16} seconds, which is much faster than the universe expansion time. And at the temperature of the order of $T = 1$ MeV, the reaction rate of the deuterium formation is also larger than the Hubble expansion rate. So the deuterium is in equilibrium with the nucleon with the deuterium formation and dissociation the same rate. As the universe expands further its temperature drops and the nuclear dissociation rate of the deuterium decreases. So we need to calculate the time when the neutron and proton begin to fuse to form the deuteron and the reversed deuteron dissociation process becomes small in comparison. Since the formation reaction is of the electromagnetic strength, it completes very rapidly characterized by the eV time scale which is 10^{-16} s as already stated, once the effective formation starts it completes very quickly and all free neutron are converted into deuterons almost instantly. This hence gives the amount of neutrons which is available for the formation of the helium eventually.

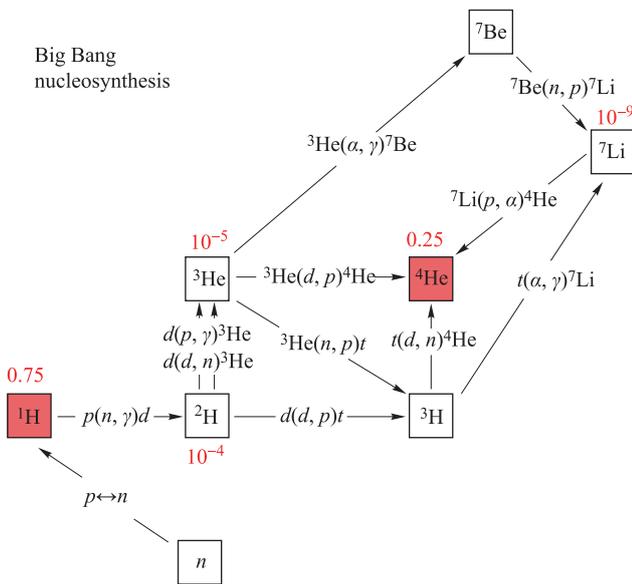


Fig. 10.4 The nuclear reactions of the nucleosynthesis of light elements. This figure is reproduced from [257].

Let us now elaborate some more on the fact that the primordial formation of light element stops at lithium. On the formation reactions, a crucial factor is the absence of the relatively stable element of atomic numbers 5 and 8. Without an $A = 5$ element, the highest atomic number that can be formed by two-body reactions is $A = 8$ from two helium-4. The resultant nucleus is the beryllium-8 which is extremely unstable with a half-life of 7×10^{-17} s. After ${}^7\text{Li}$ the lightest stable element is ${}^9\text{Be}$. Although ${}^7\text{Li}$ plus a deuterium can form a ${}^9\text{Be}$, due to the very small density of ${}^7\text{Li}$, and by this time the temperature of the universe is also low, no meaningful amount of ${}^9\text{Be}$ can be formed.

The nucleons and light elements are in thermal and chemical equilibrium through the reactions listed in Eq. (10.74). The number densities of most of the elements are very small and follow the non-relativistic, Maxwell-Boltzmann distribution of Eq. (9.126),

$$n_j(T_\gamma) = g_j \left(\frac{m_j T_\gamma}{2\pi} \right)^{3/2} e^{-(m_j - \mu_j)/T_\gamma}, \quad (10.75)$$

where the chemical potentials are not negligible. Since the temperature involved in maintaining the equilibrium is that of the photons as indicated by the first expression of Eq. (10.74), so the temperature variable above is explicitly marked to avoid any confusion. The internal (spin) degrees of freedom are: $g_p = 2$, $g_n = 2$, $g_D = 3$, $g_T = 2$, $g_{{}^3\text{He}} = 2$, and $g_{{}^4\text{He}} = 1$. Denote the total baryon density number by n_b which includes the free nucleons, p and n , and all nucleons bound in nuclei. We define the baryon-photon ratio,

$$\eta \equiv \frac{n_b}{n_\gamma} \equiv \eta_{10} \times 10^{-10},$$

$$n_\gamma = \frac{2\zeta(3)}{\pi^2} T_\gamma^3, \quad (10.76)$$

where $\eta_{10} = 6.23 \pm 0.17$ [11].

To carry out the calculation we have to work with an expression which is independent of the chemical potential of the particles involved. The deuterium formation reaction, i.e., the first reaction of Eq. (10.74), offers such a possibility. We first observe that the chemical potential of the photon always vanishes, because the possibility of multi-photon process for a given initial state, e.g., in the final state of the first reaction in Eq. (10.74), multiple photons are allowed even though their probabilities are very small in comparison with the single photon process. The thermal and chemical equilibrium of the first reaction in Eq. (10.74) says that the sum of the chemical potentials of the proton and neutron equals to that of the deuteron, $\mu_p + \mu_n = \mu_D$. This leads us to consider the Saha equation,

$$\frac{n_n n_p}{n_D} = \frac{g_n g_p}{g_D} \left(\frac{m_n m_p}{m_D} \cdot \frac{T_\gamma}{2\pi} \right)^{3/2} e^{-\varepsilon_D/T_\gamma}, \quad (10.77)$$

where ε_D is the nuclear binding energy of the deuterium⁶³.

$$\varepsilon_D = m_n + m_p - m_D = 2.23452 \text{ MeV.} \quad (10.78)$$

We consider the ratios of individual nucleons and nuclei possibly formed divided by the total baryon number density. Defining

$$X_j \equiv \frac{n_j}{n_b}, \quad (10.79)$$

we have

$$X_n + X_p + 2X_D + 3X_T + 3X_{^3\text{He}} + 4X_{^4\text{He}} + \dots = 1. \quad (10.80)$$

The Saha equation can be rewritten as

$$\begin{aligned} G_{np} &= \frac{X_n X_p}{X_D} \\ &= \frac{\sqrt{\pi}}{12\zeta(3)} \frac{10^{10}}{\eta_{10}} \left(\frac{m_n}{T_\gamma}\right)^{3/2} e^{-\varepsilon_D/T_\gamma} \\ &= 3.5388 \times 10^{13} \eta_{10}^{-1} \left(\frac{1 \text{ MeV}}{T_\gamma}\right)^{3/2} e^{-\varepsilon_D/T_\gamma}, \end{aligned} \quad (10.81)$$

where we have made the approximation $m_D = 2m_p$. From

$$X_D = G_{np}^{-1} X_n X_p, \quad (10.82)$$

we use Eq. (10.80) to eliminate X_p and obtain

$$X_D = \frac{G_{np}^{-1} X_n}{1 + 2G_{np}^{-1} X_n} (1 - X_n - 3X_T - 3X_{^3\text{He}} - 4X_{^4\text{He}} - \dots). \quad (10.83)$$

Similarly expressions can be obtained for other nucleus fractions X_T , etc.

10.6.1 Deuterium bottleneck

Let us examine briefly the behavior of G_{np} given in Eq. (10.81). Its exponential form makes it a rapidly changing function of T_γ . It reaches a maximum at $T_\gamma = 1.5$ MeV with the magnitude of the order of 10^{12} for $\eta_{10} = 6.23$. At $T_\gamma = 0.1$ MeV, G_{np} decreases by 8 orders of magnitude to $G_{np} \approx 3.5 \times 10^4$. So G_{np}^{-1} is very small and rapidly varying in this range of temperature. This implies that all the nucleus fractions, including X_D , except X_p and X_n , are small and rapidly varying with T_γ . When T_γ decreases further G_{np} decreases quickly to unity at $T_\gamma \approx 0.067$ MeV. The formation of the deuterium will become important somewhere in the temperature range below 0.1 MeV. In the general study of X_D the $G_{np}^{-1} X_n$ term in the numerator of Eq. (10.83) and in all the nuclei fractions on the right-handed side can be neglected. We can write

$$X_D \lesssim G_{np}^{-1} X_n (1 - X_n). \quad (10.84)$$

Now we can study the behavior of X_D as a function of the temperature. For $T_\gamma = 1$ MeV, $G_{np} \sim 10^{12}$ and X_D is about 10^{-13} for $X_n = 0.2$. At $T_\gamma = 0.1$ MeV, $G_{np} \approx 3.5 \times 10^4$ and $X_D \approx 4.5 \times 10^{-6}$. At such a temperature the density of deuteriums is still too small to provide a sufficient reservoir for the formation of the helium. As the temperature decreases further X_D increases dramatically. For $T_\gamma = 0.09, 0.08,$ and 0.07 we obtain respectively $X_D = 4.6 \times 10^{-5}, 8.6 \times 10^{-4},$ and 3.8×10^{-2} .

Let us interject some physical considerations here. First, the physical reason of the smallness of X_D in the temperature range of 1 to 0.1 MeV is the high entropy per baryon as determined by the ratio of baryon to photon, with the photon energy equal or greater than the binding energy of the deuterium. Such photons are energetic enough to dissociate the deuteron into their neutron and proton components once it is formed.⁶⁴ Consider the case of $T_\gamma < \varepsilon_D$ so that $\exp(\varepsilon_D/T_\gamma) \gg 1$. We can obtain the number of such photons by analytically integrating Eq. (9.100) by approximating the photon distribution function by $f_\gamma(E) \approx \exp(-E/T_\gamma)$. Roughly, this approximation is valid already for $T_\gamma = 1$ MeV which gives $\exp(\varepsilon_D/1 \text{ (MeV)}) = 9.34$. Then we have

$$n_\gamma(E \geq \varepsilon_D) \approx g_\gamma \frac{T_\gamma^3}{2\pi^2} \left[\left(\frac{\varepsilon_D}{T_\gamma}\right)^2 + 2\left(\frac{\varepsilon_D}{T_\gamma}\right) + 2 \right] e^{-(\varepsilon_D/T_\gamma)}. \quad (10.85)$$

The entropy relative to the deuterium is measured by

$$\begin{aligned} \frac{n_\gamma(E \geq \varepsilon_D)}{n_D} &= \frac{n_\gamma(E \geq \varepsilon_D)}{n_b X_D} \\ &\approx \frac{10^{10}}{\zeta(3)\eta_{10} X_D} \left(\frac{\varepsilon_D}{T_\gamma}\right)^2 e^{-(\varepsilon_D/T_\gamma)}, \end{aligned} \quad (10.86)$$

where we have replace the factor $1/2 + (\varepsilon/T_\gamma)^{-1} + (\varepsilon/T_\gamma)^{-2}$ by 1. It is an approximation good within a factor 2 for $T_\gamma < 1$. Substituting X_D given in Eq. (10.84) and using G_{np} defined in Eq. (10.81) we have

$$\begin{aligned} \frac{n_\gamma(E \geq \varepsilon_D)}{n_D} &\approx \frac{1}{X_n(1 - X_n)} \left(\frac{10^{10}}{\zeta(3)\eta_{10}}\right)^2 \frac{\sqrt{\pi}}{12} \left(\frac{m_n}{\varepsilon_D}\right)^{3/2} \left(\frac{\varepsilon_D}{T_\gamma}\right)^{7/2} e^{-2\varepsilon_D/T_\gamma} \\ &= 2.271 \times 10^{21} \frac{1}{X_n(1 - X_n)} \left(\frac{\varepsilon_D}{T_\gamma}\right)^{7/2} e^{-2\varepsilon_D/T_\gamma}. \end{aligned} \quad (10.87)$$

Taking $X_n \approx 0.1$, the above expression shows that for every deuterium nucleus there are about 4.8×10^{21} photons at $T_\gamma = 1$ MeV, 6.3×10^{20} at 0.5 MeV and 5.2×10^7 at 0.1 MeV. The ratio decreases rapidly as the temperature decreases. The entropy to deuterium ratio becomes 1 at

⁶⁴Here we follow the discussion given in [253].

0.07 MeV. We expect that this is the ball park temperature scale when the deuterium start to form. Because of the rapidly decrease of the entropy to deuterium ratio, the formation of the deuterium can be completed rather quickly once it begins.

The fact that the formation of helium 4 depends on the formation of the deuterium which can only happen at a low temperature blow 0.1 MeV is known as the deuterium bottleneck because of smallness of the amount of deuteriums available at earlier time. If helium-4 can be formed independent of the deuterium, the appearance of helium-4 would have happened much sooner. We discuss this briefly below. From the consideration of thermal equilibrium alone, we can define the baryon fraction of helium-4

$$X_{4\text{He}} = \frac{n_{4\text{He}}}{n_b}, \tag{10.88}$$

and form the ratio

$$G_{4\text{He}} = \frac{X_p^2 X_n^2}{X_{4\text{He}}} \sim e^{-\epsilon_{4\text{He}}/T}, \tag{10.89}$$

similar to the discussion of the deuterium in Eq. (10.81), where $\epsilon_{4\text{He}}$ is the binding energy of helium 4. Because of the exponential factor $\exp(-\epsilon_{4\text{He}}/T)$, $G_{4\text{He}}$ decreases very fast as the temperature decreases. At $T = 0.3$ MeV, $G_{4\text{He}}$ is 8.5 with a very steep slope⁶⁵, it seems that helium-4 could be formed at around $T = 0.3$ MeV. But the reaction for the direct formation, $2p + 2n \rightarrow {}^4\text{He} + \gamma$, is fantastically small, this channel of formation is not valid and it has to wait till there are a sufficient number of deuteriums to continue the series of reactions listed in Eq. (10.74). So this is also a factor contributing to the deuterium bottleneck.

Another remark we would like to make is the following. From $T = 0.1$ MeV to 0.07 MeV, the deuterium density increases by about four orders of magnitude. Is this physically reasonable? Let us look how long it takes for such a variation of temperature to occur. We can estimate this time spend as follows. As we have discussed in §9.5, after the cosmic temperature decreases below 1 MeV the photon has been reheated due to the electron-positron annihilation, while neutrinos are decoupled and expand with the universe by themselves. The photon and neutrinos evolve with different temperatures $T_\nu = (4/11)^{1/3} T_\gamma$. One of the approaches to do the calculation is to track the neutrino temperature when the universe expands and then convert it to that of the photon which defines the cosmic temperature. When neutrinos are decoupled, they expand freely with the universe so that $a^3 T_\nu^3 = \text{const}$. We have

$$\frac{\dot{a}}{a} + \frac{1}{T_\nu} \frac{dT_\nu}{dt} = \sqrt{\frac{8\pi G_N}{3}} \rho(T_\nu) + \frac{1}{T_\nu} \frac{dT_\nu}{dt} = 0. \tag{10.90}$$

⁶⁵ $G_{4\text{He}}$ is the order of 10^{-2} at 0.28 MeV and 10^3 at 0.32 MeV.

Below $T = 0.1$ MeV the universe is dominated by photons and neutrinos and the temperatures of the two species of particles track each other by the relation $T_\gamma = (11/4)^{1/3} T_\nu$, and

$$\begin{aligned} \rho(T) &= \rho_\gamma(T_\gamma) + \rho_\nu(T_\nu) = g'_* \frac{\pi^2}{30} T_\nu^4, \\ g'_* &= 2 \left(\frac{11}{4}\right)^{4/3} + 6 \left(\frac{7}{8}\right) = 12.956. \end{aligned} \tag{10.91}$$

Integrating Eq. (10.90), we obtain⁶⁶

$$\begin{aligned} t &= \sqrt{\frac{45}{16\pi^3 g'_*}} \frac{M_P}{T_\nu^2} + t_0 \\ &= \sqrt{\frac{45}{16\pi^3 g'_*}} \left(\frac{11}{4}\right)^{2/3} \frac{M_P}{T_\gamma^2} + t_0 \\ &= 1.3199 \left(\frac{1 \text{ MeV}}{T_\gamma}\right)^2 \text{ s} + t_0, \end{aligned} \tag{10.92}$$

where t_0 is estimated in BBF to be $t_0 = 2$ s. Hence the time lapse between 0.1 MeV and 0.07 MeV is about 137 second. This is a very long period of time for neutron fusion with proton to form the deuterium by strong/electromagnetic interaction in the MeV/eV energy scale. Their corresponding interaction time scales are respectively 10^{-22} and 10^{-16} seconds.

10.7 Primordial fraction of helium-4

The calculation of the neutrons capture time into deuteriums has two important features. One is that it contains many details which require, in particular, the examination of rates involving the reaction sequence of Eq. (10.74). The second is that the rapid variation of the various quantities involved in the temperature region of 0.1 MeV, which we have seen in our order of magnitude estimate, requires careful numerical works. Both are too involved for our attempt of a simple exposition of these topic. However, we expect the formation of ${}^4\text{He}$ to take place when the deuterium is no long being overwhelmed by the entropy, that is when the cosmic temperature is below 0.1 MeV. This has been discussed in the paragraph below Eq. (10.87). This critical temperature of the formation of ${}^4\text{He}$, denoted as $T_{\gamma,c}$, can be found in the literature and is also given in BBF,

$$T_{\gamma,c} = \frac{\epsilon_d}{26} \approx 0.086 \text{ MeV}, \tag{10.93}$$

which lies indeed in the temperature range as expected above. According to Eq. (10.92), the corresponding critical time is

$$t_c = 180 \text{ s}, \tag{10.94}$$

⁶⁶The first and second lines of the equation below are Eq. (3.10) of BBF.

which is three minutes after the big bang. We have the neutron fraction at the critical temperature

$$\hat{X}_n(Q/T_{\gamma,c}) = 0.136 \quad (10.95)$$

with the neutron decay channel included in the integration. Another results is

$$\tilde{X}_n^{(d)}(Q/T_{\gamma,c}) = 0.123, \quad (10.96)$$

which is just the result of BBF given its equation (3.43).

We can now obtain the baryon fraction of helium-4. Once the deuterium begins to form, all neutrons are captured into deuterium very quickly due to the very rapid strong/electromagnetic interaction process. The rest of the series of reactions in Eq. (10.74) can also be completed very quickly. Because of the large binding energy of the helium-4 most of the deuteriums are converted into it. The total number of the deuterium nucleus fraction is just \hat{X}_n or $\tilde{X}_n^{(d)}$. The helium-4 formation is given by the first five reactions of Eq. (10.74), they are effectively $D + D \rightarrow {}^4\text{He} + \gamma$. We have the number of the helium-4 to be just one-half of the number of the deuterium. Then the primordial baryon number fraction of helium-4, which is denoted as Y_p with the subscript P to mean ‘‘primordial’’, is 4 times of the fraction of helium-nucleus, we have

$$Y_p = 4X_{{}^4\text{He}} = \begin{cases} 2\hat{X}_n(Q/T_{\gamma,c}) = 0.272, \\ 2\tilde{X}_n^{(d)}(Q/T_{\gamma,c}) = 0.246, \end{cases} \quad (10.97)$$

where the value $2\tilde{X}_n^{(d)}(Q/T_{\gamma,c}) = 0.246$ is the BBF result and $2\hat{X}_n(Q/T_{\gamma,c}) = 0.272$ agrees with the result given in [221]⁶⁷⁾. The recent observational data give the value of Y_p in the range of 0.247 to 0.252 with typically rather small errors [11]. The value given in Eq. (10.97) for $\hat{X}_n^{(d)}$ lies in this observation range. This agreement should not be overstated in view of the approximations made in the calculation. It just says that the approximates which are made to enable us to do analytic calculations are not unreasonable.

10.8 Counting the baryonic matter and the primordial lithium problem

The predicted abundances of ${}^4\text{He}$, D , ${}^3\text{He}$, and ${}^7\text{Li}$ by the standard model of Big-Bang cosmology and the observational values are shown in Fig. 10.5 which is taken from [11]. The predicted values as represented by vertical bands are in the 95% range. The observational values are shown as horizontal bands, where the smaller bands with solid boundary-lines include 2σ statistical errors and the larger bands with dashed boundary-lines include 2σ statistical plus systematic errors. For ${}^4\text{He}$ the systematic

⁶⁷⁾See, page 167, [221].

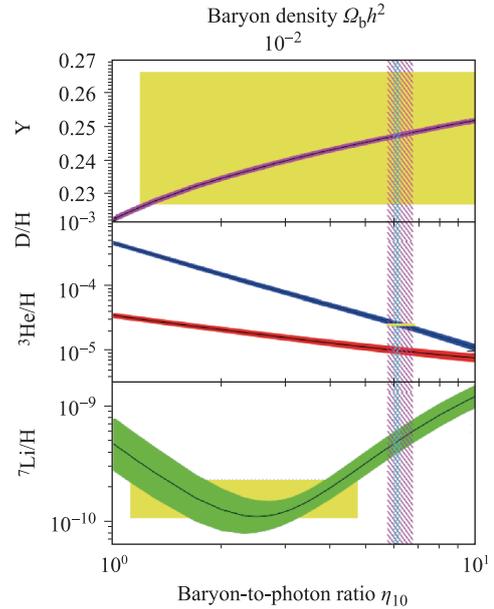


Fig. 10.5 The predicted abundances of light elements in Big-Bang nucleosynthesis and their observational values. The fractions of the deuterium, ${}^3\text{He}$, and lithium-7 are relative to the hydrogen. The subscript p means primordial. Note that $\eta \times 10^{10}$ is η_{10} in the notation used here.

error dominates. The wider vertical band marked by inclined red line segments indicates the range predicted by the BBN concordance theory. The narrow vertical band marked by short blue line segments shows the predictions according to the CMB measurement of the cosmic baryon density. The predicted and observational values for helium-4 and deuterium agree very well. There are yet no observation data for helium-3. We will come to the Lithium-7 later below. The scale of the upper horizontal axis Ωh^2 and that of the lower horizontal axis $\eta \times 10^{10} \equiv \eta_{10}$ are related by

$$\begin{aligned} \Omega_b h^2 &\equiv \frac{\rho_b}{\rho_c} h^2 \\ &= \frac{n_\gamma m_N}{\rho_c} \eta_{10} \\ &= 3.65 \times 10^{-3} \eta_{10} \left(\frac{m_N}{0.938 \text{ GeV}} \right), \end{aligned} \quad (10.98)$$

where m_N is the effective nucleon mass which is close to the proton mass and

$$\begin{aligned} \rho_c &= \frac{3H_0^2}{8\pi G_N} \\ &= 1.05375 \times 10^{-5} h^2 (\text{GeV}/c^2) \text{cm}^{-3} \end{aligned} \quad (10.99)$$

is the critical density.

The BBN result is an earlier evidence, no later than the 1970s, for non-baryonic matter before the WMAP precision measurement. As indicated in Fig. 10.5, The

deuterium fraction is quite sensitive to η and serves as a good determination of it. The deuterium data gives η_{10} in the range of 5.0–6.8 which corresponds to

$$\Omega_b h^2 = 0.018 - 0.024, \quad \Omega_b = 0.042 - 0.053,$$

where we have taken the current value of the scale factor for Hubble expansion rate $h = 0.673$. However, the fraction of all non-relativistic matter Ω_m which contributes to the critical density obtained from other studies is much larger. The study of galaxy clusters gives $\Omega_m \approx 0.2$ and the study of redshift-distance relation of type I supernovae gives $\Omega_m \approx 0.3$. This difference, which is close to a factor 5, makes it very difficult to identify Ω_b with Ω_m . Hence extra matter not in the form of baryon should exist.

The lithium-7 data, which agrees marginally with helium-4 and deuterium, requires that η_{10} to be no more than 5. This conflicts with the high precision CMB measurement of the baryon density which requires η_{10} to be no less than 6. As shown in Fig. 10.5, the CMB determination of the baryon to photon ratio η predicts the lithium-7 ratio to be no smaller than 5×10^{-10} , while the observed value is no greater than 3×10^{-10} . This disagreement is usually referred to as the lithium problem. Assuming that the origin of the disagreement is not experimental in origin, we have to have new physics for its explanation. Indeed there are scenarios in the supersymmetric standard model which can account for the overabundance of lithium-7, and the lithium problem can be related to the dark matter. For further discussions for probes of the standard BBN with new physics, we refer to [244, 245].

11 Freeze out of massive particles

In this Section we discuss in some details the freeze out of particles in the early universe, assuming the standard

cosmology. This is relevant in the treatment of dark matter such as WIMPs in establishing the widely known WIMP miracle. Good treatment of the subject can be found in, e.g., [74], [221], and [101].⁶⁸⁾ The original article [213], which is the basis of the WIMP miracle, is concise and also readable, starting from the Boltzmann transport equation. We begin by deriving the Boltzmann transport equation which we have relied on the discussion of the physics of thermal relics WIMPs in §5.3.

11.1 Boltzmann transport equation for particle production and annihilation

We consider the two-body scattering process, involving particles denoted as 1, 2, 3, and 4,⁶⁹⁾

$$1 + 2 \rightleftharpoons 3 + 4. \quad (11.1)$$

This process describes the annihilation of particles 1 and 2 into 3 and 4, as well as the reversed reaction of the production of particles 1 and 2 from 3 and 4. In the discussion of dark matter, particles, 1 and 2 represent the dark matter particle and its antiparticle. In the case of Majorana particles, the particle and anti-particle are the same. Particles 3 and 4 are normal particles, which can be photons, charged leptons, neutrinos, quarks, etc., and their anti-particles. A particular example is the reaction of dark matter particles, X and \bar{X} annihilation into a pair of lighter normal particles, say a lepton-antilepton pair

$$X + \bar{X} \leftrightarrow \ell + \bar{\ell}. \quad (11.2)$$

The present discussion is not restricted to the production of normal leptonic particles. of course.

We denote the number densities of the particles involved by n_j , $j = 1, \dots, 4$. Let us fix our attention on particle 1. The rate of change of number density of particle 1 in the expanding universe in the comoving volume a^3 is given by the Boltzmann equation

$$\begin{aligned} \frac{1}{a^3} \frac{d}{dt} (n_1 a^3) = & \sum_{\text{spins}} \int \dots \int \prod_{j=1}^4 \left(\frac{d^4 p_j}{(2\pi)^3} \delta_+(p_j^2 - m_j^2) \right) (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4) |\mathcal{M}|^2 \\ & \times (f_3 f_4 (1 \pm f_1) (1 \pm f_2) - f_1 f_2 (1 \pm f_3) (1 \pm f_4)), \end{aligned} \quad (11.3)$$

where \mathcal{M} is the matrix element of the reaction of Eq. (11.1) and the summation is over the spins states of all particles involved. The δ_+ means to take the positive root in the solution of the δ -function, and f_j are the particle distribution function given in Eq. (9.98). The factor f_j is associated with particles in the initial state of the reaction, while $(1 \pm f_j)$ for particles in the final state. $1 - f_j$ is for the fermion and $1 + f_j$ for the boson.

They represent respectively the effects of Pauli blocking and the boson enhancement.

The first term in the second line of the right-handed side of Eq. (11.3) is the production of particles 1 and 2 by particle 3 and 4. It increases the number of particle 1

⁶⁸⁾We follow the discussion given in [101], pp 58–61 and 73–78. See also [74], pp 119–130 and [221], pp 185–196.

⁶⁹⁾Our starting argument follows that given in Section 3, [101].

and therefore has a positive sign. The second term, which represents the annihilation of particle 1 and 2 into 3 and 4, decreases the number of particles 1, and hence gets the negative sign. Let us look at the engineering dimensionality of both sides of the equation as a quick check. The left-handed side has the dimension per unit volume per unit time in the proper units. In the natural units it is energy to the fourth power. Since a two-body to two-

body reaction matrix element is dimensionless and the dimension of a δ -function is the inverse of its argument, hence the right-handed side has indeed the required dimension of the energy to the fourth power.

Since the particle distribution functions f_j , $j = 1, \dots, 4$ are independent of the particle spin states, we can move the summation sign to the matrix element term,

$$\frac{1}{a^3} \frac{d}{dt} (n_1 a^3) = \int \dots \int \prod_{j=1}^4 \left(\frac{d^4 p_j}{(2\pi)^3} \delta_+(p_j^2 - m_j^2) \right) (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4) \sum_{\text{spins}} |\mathcal{M}|^2 \times (f_3 f_4 (1 \pm f_1)(1 \pm f_2) - f_1 f_2 (1 \pm f_3)(1 \pm f_4)), \quad (11.4)$$

11.2 The rate of change of particle number density

The particle number density change rate Eq. (11.4) can be simplified in the physical situation under consideration.⁷⁰⁾ In the WIMP model, the cold dark matter particles are heavy with mass in the GeV to multi-GeV range. Before it freezes out it is already non-relativistic. Hence if it is maintained in equilibrium, its number density, denoted by n_X , will have the form of Eq. (9.126) which decreases exponentially as the cosmic temperature decreases. Then there will be very little of the particle remaining today. It has to decouple from the cosmic heat bath at some point of the cosmic evolution so as to freeze out to maintain the dilution due to Hubble expansion, $n_X \sim a^{-3}$. It is hopefully that the number density is still large enough to allow its existence to be probed experimentally. The freeze out takes place presumably below the GeV temperature range $m_X \gg T$, and the energy of the massive particle is dominated by its mass. Because of the energy conservation δ -function in Eq. (11.4), $E_1 + E_2 = E_3 + E_4$, particles 3 and 4, which are lighter ordinary standard model particles, are highly relativistic, and therefore in equilibrium with the cosmic thermal bath. So in this dominant kinematic condition, all the number density distribution functions involved can be approximated by

$$f_j = \frac{1}{e^{(E_j - \mu_j)/T} \pm 1} \rightarrow e^{-(E_j - \mu_j)/T}, \quad (11.5)$$

where μ_j is the chemical potential of the particle j . In this approximation the Pauli blocking and boson enhancement factors can be replaced by unity. This says that the particle densities are sufficiently diluted, the effects of the Pauli block and boson enhancement can be neglected. This realization of the kinematic states of the

particles involved greatly simplify the formulation to be discussed below.

In the out-of-equilibrium situation the particles involved in the reaction are not in chemical equilibrium so that

$$\mu_1 + \mu_2 \neq \mu_3 + \mu_4. \quad (11.6)$$

Now we can write the expression of the second line of Eq. (11.4)

$$f_3 f_4 (1 \pm f_1)(1 \pm f_2) - f_1 f_2 (1 \pm f_3)(1 \pm f_4) \rightarrow e^{-(E_3 - \mu_3)/T} e^{-(E_4 - \mu_4)/T} - e^{-(E_1 - \mu_1)/T} e^{-(E_2 - \mu_2)/T} = e^{-(E_1 + E_2)/T} [e^{(\mu_3 + \mu_4)/T} - e^{(\mu_1 + \mu_2)/T}], \quad (11.7)$$

where we have used the energy conservation relation $E_1 + E_2 = E_3 + E_4$.

In the spirit of the approximation of Eq. (11.7) above, the particle number densities involved can be approximated by

$$n_j \approx g_j \frac{4\pi}{(2\pi)^3} \int e^{-(E - \mu_j)/T} p^2 dp = g_j e^{\mu_j/T} \begin{cases} \frac{T^3}{\pi^2}, & T \gg m_j, \\ \left(\frac{m_j T}{2\pi}\right)^{3/2} e^{-m_j/T}, & T \ll m_j. \end{cases} \quad (11.8)$$

We also define the number density in the absence of chemical potential,

$$n_j^{(0)} \equiv n_j|_{\mu_j=0} = \begin{cases} g_j \frac{T^3}{\pi^2}, & T \gg m_j, \\ g_j \left(\frac{m_j T}{2\pi}\right)^{3/2} e^{-m_j/T}, & T \ll m_j. \end{cases} \quad (11.9)$$

⁷⁰⁾We follow the treatment of [101], pp 58–62 and pp 73–78.

Comparing the above expressions for $n_j^{(0)}$ and the exact ones given in Eqs. (9.119) and (9.126), we see that the non-relativistic approximations are the same while the relativistic expression given in the above lies between the fermion and boson cases of Eq. (9.119). So the differences are small, unimportant in light of the treatment presented here. We now arrive at the crucial approximation,

$$e^{\mu_j/T} = \frac{n_j}{n_j^{(0)}}. \tag{11.10}$$

We can rewrite Eq. (11.7)

$$f_3 f_4 (1 \pm f_1)(1 \pm f_2) - f_1 f_2 (1 \pm f_3)(1 \pm f_4) \rightarrow e^{-(E_1+E_2)/T} \left(\frac{n_3 n_4}{n_3^{(0)} n_4^{(0)}} - \frac{n_1 n_2}{n_1^{(0)} n_2^{(0)}} \right). \tag{11.11}$$

In the case of chemical equilibrium, $\mu_1 + \mu_2 = \mu_3 + \mu_4$, the above expression vanishes. This is the expected result that in thermal equilibrium the particle density in a comoving volume is constant.

Now the Boltzmann equation Eq. (11.4) becomes

$$\begin{aligned} \frac{1}{a^3} \frac{d}{dt} (n_1 a^3) &\approx \int \dots \int \prod_{j=1}^4 \left[\frac{d^4 p_j}{(2\pi)^3} \delta_+(p_j^2 - m_j^2) \right] (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4) \sum_{\text{spins}} |\mathcal{M}|^2 \\ &\times e^{-(E_1+E_2)/T} \left[\frac{n_3 n_4}{n_3^{(0)} n_4^{(0)}} - \frac{n_1 n_2}{n_1^{(0)} n_2^{(0)}} \right] \\ &= n_1^{(0)} n_2^{(0)} \left[\frac{n_3 n_4}{n_3^{(0)} n_4^{(0)}} - \frac{n_1 n_2}{n_1^{(0)} n_2^{(0)}} \right] \\ &\times \frac{1}{n_1^{(0)} n_2^{(0)}} \int \dots \int \prod_{j=1}^4 \left(\frac{d^4 p_j}{(2\pi)^3} \delta_+(p_j^2 - m_j^2) \right) (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4) \\ &\times \sum_{\text{spins}} |\mathcal{M}|^2 e^{-(E_1+E_2)/T}. \end{aligned} \tag{11.12}$$

In the second equality sign we factor out the terms involving ratios of particle number densities because they are independent of the particle momenta. Let us rewrite the last two lines of the above expression,

$$\begin{aligned} &\frac{1}{n_1^{(0)} n_2^{(0)}} \int \dots \int \prod_{j=1}^4 \left[\frac{d^4 p_j}{(2\pi)^3} \delta_+(p_j^2 - m_j^2) \right] (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4) \sum_{\text{spins}} |\mathcal{M}|^2 e^{-(E_1+E_2)/T} \\ &= \frac{g_1 g_2}{n_1^{(0)} n_2^{(0)}} \int e^{-E_1/T} \frac{d^3 p_1}{(2\pi)^3} \int e^{-E_2/T} \frac{d^3 p_2}{(2\pi)^3} \left| \frac{\mathbf{p}_1}{E_1} - \frac{\mathbf{p}_2}{E_2} \right| \\ &\times \frac{1}{2E_1 2E_2 \left| \frac{\mathbf{p}_1}{E_1} - \frac{\mathbf{p}_2}{E_2} \right|} \int \int \frac{d^3 p_3}{(2\pi)^2 2E_3} \frac{d^3 p_4}{(2\pi)^2 2E_4} (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4) \sum |\overline{\mathcal{M}}|^2 \\ &= \frac{g_1 g_2}{n_1^{(0)} n_2^{(0)}} \int e^{-E_1/T} \frac{d^3 p_1}{(2\pi)^3} \int e^{-E_2/T} \frac{d^3 p_2}{(2\pi)^3} \left| \frac{\mathbf{p}_1}{E_1} - \frac{\mathbf{p}_2}{E_2} \right| \sigma(1 + 2 \rightarrow 3 + 4) \\ &\equiv \langle v \sigma \rangle, \end{aligned} \tag{11.13}$$

where $\sum |\overline{\mathcal{M}}|^2$ is the summation over the spin states of particles 3 and 4, and average of the spin states of particles 1 and 2. To average over the spins of particles 1 and 2 is to divide $\sum_{\text{spin}} |\mathcal{M}|^2$ by the product $g_1 g_2$. We note in the above expression that the second line after the first equality sign is the cross section formula of the reaction $1 + 2 \rightarrow 3 + 4$. The term

$$\left| \frac{\mathbf{p}_1}{E_1} - \frac{\mathbf{p}_2}{E_2} \right| \tag{11.14}$$

is the relative velocity of the initial particles. As can be

seen in the non-relativistic situation, it is the absolute value of the difference of the velocities of the two particles, $|\mathbf{v}_1 - \mathbf{v}_2|$.⁷¹⁾ The form of the integrals over p_1 and p_2 is just the average over the distribution functions of

⁷¹⁾We remind the reader that the term $2E_1 2E_2 \left| \frac{\mathbf{p}_1}{E_1} - \frac{\mathbf{p}_2}{E_2} \right|$ in Eq. (11.13) is referred to as the flux density of the initial state and it is usually written in the covariant form, in either the center of mass frame or the laboratory frame, as $2\lambda(S, m_1^2, m_2^2)$, where $\lambda(a, b, c) \equiv \sqrt{a^2 + b^2 + c^2 - 2ab - 2ac - 2bc}$. In the present discussion $S = (p_1 + p_2)^2$, where p_1 and p_2 are respectively the 4-momentum of particles 1 and 2.

Eq. (11.9) of the initial particles of the product of their relative velocity and the reaction cross section. The multiplication of the factors g_1 and g_2 is just a counting of the degrees of freedom entering in the densities of particles 1 and 2. Now we obtain the differential equation for the rate of change of the number density of particle 1, averaged in a comoving volume,

$$\frac{1}{a^3} \frac{d}{dt} (a^3 n_1) = -\langle v\sigma \rangle n_1^{(0)} n_2^{(0)} \left[\frac{n_1 n_2}{n_1^{(0)} n_2^{(0)}} - \frac{n_3 n_4}{n_3^{(0)} n_4^{(0)}} \right]. \quad (11.15)$$

From Eqs. (9.151), (9.155), and (9.153), or simply from the fact $aT \sim \text{const.}$, we have

$$dt = -\frac{1}{H(T)} \frac{dT}{T}. \quad (11.16)$$

Let us introduce the variable

$$x = \frac{m_X}{T}. \quad (11.17)$$

Then we can rewrite the time derivative in terms of the derivative in x ,

$$\frac{d}{dt} = xH(T) \frac{d}{dx} = x^{-1}H(m_X) \frac{d}{dx}. \quad (11.18)$$

The second equality of the above expression follows from the fact that the massive particle decoupling takes place at high temperatures and therefore the cosmic expansion is dominated by radiation composed of relativistic ordinary particles.⁷²⁾ So the total energy density is $\rho_R \sim T^4$. Then, from Eq. (9.153), we have

$$H(T) = \sqrt{\frac{8\pi G_N}{3} \rho(T)} = x^{-2} H(m_X),$$

$$H(m_X) = \sqrt{\frac{8\pi G_N}{3} \rho(m_X)} = \sqrt{\frac{4\pi^3}{45} g_*} \left(\frac{m_X^2}{M_P} \right), \quad (11.19)$$

where g_* is the total number of degrees of freedom of effectively massless particles at the temperature under consideration. See Table 9.4 for values of g_* under various situations. In arriving at the above expression, Eq. (11.19), we have used the fact that the cosmic temperature is high energy so that the universe is radiation dominated, so the total energy density takes the form $\rho(T) = g_*(\pi^2/30)T^4$ of Eq. (9.135).

Let us come back to Eq. (11.15). In the symmetric case of the annihilation of massive particles $n_1 = n_2$ and $n_1^{(0)} = n_2^{(0)}$. For particles 3 and 4, we take $T \gg m_3, m_4$ so that they are in thermal equilibrium with the cosmic heat bath. Then n_3 and n_4 are respectively $n_3^{(0)}$

⁷²⁾Strictly speaking, we should write $H(T) = H(m_X)\sqrt{g_*(T)/g_*(m_X)}$. We make the approximation by ignoring the difference between $g_*(T)$ and $g_*(m_X)$.

and $n_4^{(0)}$. We also take particles 3 and 4 to be particle and antiparticle in the symmetric case $n_3 = n_4$. Furthermore, we assume that this is a radiation dominated epoch and particles 3 and 4 are highly relativistic so that their chemical potentials can be neglected. From Eq. (9.116), i.e., $aT \sim \text{const.}$, we define the massive particle number density in a comoving volume up to a constant⁷³⁾

$$Y(x) \equiv \frac{n_1}{T^3} \propto n_1 a^3,$$

$$Y_{\text{eq}}(x) \equiv \frac{n_1^{(0)}}{T^3} \propto n_1^{(0)} a^3. \quad (11.20)$$

We can now write the rate of change of the massive particle density Eq. (11.15) as

$$\frac{d}{dx} Y(x) = -\frac{\lambda}{x^2} [Y(x)^2 - Y_{\text{eq}}(x)^2], \quad (11.21)$$

where

$$\lambda \equiv \frac{m_X^3}{H(m_X)} \langle v\sigma \rangle = \sqrt{\frac{45}{4\pi^3 g_*}} m_X M_P \langle v\sigma \rangle. \quad (11.22)$$

All the quantities involved, Y , x , and λ , are dimensionless. We note that the product $m_X \langle v\sigma \rangle$ enters the differential equation which is a major feature of the freeze out equation to be explored later. As pointed out in [101], in many theories λ is a constant.⁷⁴⁾ We will take it to be a constant and estimate it later in this section, following the treatment given in [221].

Let us now interpret Eq. (11.21). Our derivation of the equation uses the fact that around the decoupling the approximation Eq. (11.5) is valid and where the physics of heavy particle under consideration is dominated. Therefore, instead of the approximation employed in Eq. (11.9), we can use the exact equilibrium distribution function of the massive particle without the chemical potential for $n_1^{(0)}$. Then $Y_{\text{eq}}(x)$ is known explicitly. The density function n_1 , even in the approximate form, contains the unknown chemical potential is now the function to be solved from the differential equation Eq. (11.21). Hence n_1 can stay arbitrary. Therefore we write for the heavy particle of spin s_X by its exact equilibrium form,

$$n_X^{(0)} = \frac{2s_X + 1}{(2\pi)^3} \int \frac{4\pi p^2 dp}{\exp\left(\sqrt{p^2 + m_X^2}/T\right) + (-1)^{2s_X + 1}}, \quad (11.23)$$

where the number of spin states of the massive particle

⁷³⁾In the WIMP miracle discussions in §5.3.1, the function Y is defined in an equivalent form in Eq. (5.9), i.e., $Y = n/s$, where s is the entropy density. In the case of no-entropy production $s \sim T^3$.

⁷⁴⁾See p. 75, [101].

$2s_X + 1$ is its degrees of freedom. Then we have⁷⁵⁾

$$Y_{\text{eq}}(x) = \frac{n_X^{(0)}}{T^3} = \frac{2s_X + 1}{(2\pi)^3} \int \frac{4\pi y^2 dy}{\exp(\sqrt{y^2 + x^2}) + (-1)^{2s_X + 1}}. \quad (11.24)$$

An equation equivalent to Eq. (11.21) can be obtained by going back to the time derivative of the massive particle density, which follows straightforwardly from Eq. (11.21) with the relationship given in Eqs. (11.18), (11.19), and (11.22), or directly from Eq. (11.15), to obtain

$$\frac{d}{dt}n_X(x) = -3Hn_X(x) - \langle v\sigma \rangle \left[(n_X(x))^2 - (n_X^{(0)}(x))^2 \right], \quad (11.25)$$

which is the Boltzmann transport equation (BTE) given in Eq. (5.5). The physical meaning of the BTE is clear. The terms on the right-handed side describe the different origins that cause the density variations of the massive particles. The first term on the right-handed side describes the dilution of massive particles due to the expansion of the universe and hence the dependence on the Hubble expansion rate H . The factor 3 comes from the spatial dimension of the space. This is a universal term common to all particle species. The function $Y \sim a^3 n$ defined in Eq. (11.20) removes this term from it. The second term is also a dilution effect due to the annihilation of two massive particles into normal particles, and hence the coefficient $\langle v\sigma \rangle$. This term depends on $n_X(x)^2$ because the annihilation reaction involves two massive particles. The negative sign reflects the fact that they decrease the particle density. The third term which is proportional to $n_X^{(0)}(x)^2$, with its positive sign and the $\langle v\sigma \rangle$ coefficient, represents the reversed reaction, the production of heavy particles from the annihilation of normal particles. As has been discussed in Section 5.3.1, in particular related to Eqs. (5.15) and (5.16), the second and third terms on the right-handed of Eq. (11.25) constitute a feedback that tends to maintain the particle to its equilibrium density when $\langle v\sigma \rangle/H$ is large, but can break away from it when $\langle v\sigma \rangle/H$ becomes small. We will work with Eq. (11.21) to obtain a solution for $Y(x)$. After obtaining the solution, we will refer back to Eq. (11.25) to gain some understanding of the general behavior of the solution. Equation (11.21) is a Riccati equation, which, known since the early 1700s, does not have a general closed form solution although particular forms of Riccati

equation have analytic solutions. So before we go about getting a numerical solution for Eq. (11.21), we digress to give a brief discussion of the general properties of the Riccati equation. In doing so we will also indicate how to obtain a unique solution, which is a necessary requirement when the equation can only be solve numerically.

11.3 Digression – Riccati equation and some of its properties

The Riccati equation (RE) for a function $y(x)$ takes the form

$$\frac{d}{dx}y(x) = P(x)y(x)^2 + Q(x)y(x) + R(x). \quad (11.26)$$

It is a first order non-linear differential equation. It has no known general solutions and, unlike a linear differential equation, it can have many different solutions as we will show later. Below we discuss two important properties of the equation.

11.3.1 Transforming to a linear equation

If there is an analytic solution, it can often be obtained by the following procedure which transform the Riccati equation to a linear equation. Let us define a function

$$v(x) = P(x)y(x), \quad (11.27)$$

which can be shown straightforwardly to satisfy a RE

$$\frac{d}{dx}v(x) = v(x)^2 + \left[Q(x) + \frac{d}{dx} \ln P(x) \right] v(x) + P(x)R(x). \quad (11.28)$$

Further we define

$$v(x) = -\frac{f'(x)}{f(x)}. \quad (11.29)$$

Then $f(x)$ satisfies a second order linear different equation,

$$\frac{d^2}{dx^2}f(x) - \left(Q(x) + \frac{d}{dx} \ln P(x) \right) \frac{d}{dx}f(x) + P(x)R(x)f(x) = 0. \quad (11.30)$$

If an analytic solution for $f(x)$ is obtained. We can then refer back to Eqs. (11.27) and (11.29) to obtain the analytic solution of the original Riccati equation,

$$y(x) = -\frac{1}{P(x)} \frac{d}{dx} \ln(f(x)). \quad (11.31)$$

11.3.2 Multiple solutions

If a solution of a RE is known, another solution can be readily obtained. This is particularly useful if the known

⁷⁵⁾The expression of Eq. (11.24) is the function $u_{\text{eq}}(x)$ given in Eq. (3.4.8) of [221]. Note that the variable x there is the inverse used here.

solution is a closed form function. Suppose a solution is known. Let us call it $y_1(x)$. Define

$$y(x) = y_1(x) + u(x), \tag{11.32}$$

where $u(x)$ satisfies also a RE, not the original form due to non-linearity,

$$\frac{d}{dx}u(x) = P(x)u(x)^2 + (Q(x) + 2P(x)y_1(x))u(x), \tag{11.33}$$

which does not have the term independent of $u(x)$ and hence a null function is a solution, in which case $y(x)$ is just the known solution $y_1(x)$. If $u(x) = 0$ is a unique solution of the above equation, then $y_1(x)$ is a unique solution of the original Riccati equation.

The above equation, Eq. (11.33), can be further trans-

$$z(x) = - \left(\int_{x_0}^x P(x') e^{\int_{x_0}^{x'} (Q(x'') + 2P(x'')y_1(x'')) dx''} dx' + C_1 \right) e^{-\int_{x_0}^x [Q(x') + 2P(x')y_1(x')] dx'}, \tag{11.37}$$

where C_1 is an integration constant and x_0 is arbitrary. So a second solution of the original RE can be obtained:

$$\begin{aligned} y_2(x) &= y_1(x) + \frac{1}{z(x)} \\ &= y_1(x) - \frac{e^{\int_{x_0}^x [Q(x') + 2P(x')y_1(x')] dx'}}{\int_{x_0}^x P(x') e^{\int_{x_0}^{x'} [Q(x'') + 2P(x'')y_1(x'')] dx''} dx' + C_1}. \end{aligned} \tag{11.38}$$

We can use the solution $y_2(x)$ to construct yet another solution, repeating the process, by substituting $y_2(x)$ for $y_1(x)$ in the above expression together with another integration constant C_2 . This process can be repeated as many times as allowed. So, in principle, there may exist infinite many solutions. This will be a nightmare when even the first solution has to be obtained numerically. Then to apply the Riccati equation is simply impractical.

11.3.3 Boundary condition and the unique solution

Fortunately the process can be terminated quite quickly by a boundary or initial condition. Suppose a boundary condition is imposed at x_0 , which is the lower integration limit in Eq. (11.37). This boundary condition has to be satisfied by all possible solutions of the RE, $y_1(x_0)$, $y_2(x_0)$, etc. Then we have to have

$$z(x_0)^{-1} = C_1^{-1} = 0, \tag{11.39}$$

which says that $C_1 \rightarrow \infty$. Therefore the added part to the solution, i.e., $1/z(x)$ vanishes identically. Hence

formed to a linear form by defining

$$u(x) = \frac{1}{z(x)}, \tag{11.34}$$

where $z(x)$ satisfies a first order linear differential equation

$$\frac{d}{dx}z(x) = -[Q(x) + 2P(x)y_1(x)]z(x) - P(x). \tag{11.35}$$

The solution of this equation is well-known, obtained by multiplying the integrating factor

$$e^{\int_{x_0}^x (Q(x') + 2P(x')y_1(x')) dx'} \tag{11.36}$$

to the differential equation and the solution can be obtained by inspection,

$y_1(x)$ is a unique solution. Therefore, in the case of numerical solution, if it can be constructed to satisfy a boundary/initial condition, the numerical solution will be unique. This tells us that we can construct a numerical solution for Eq. (11.21) this way and the solution is unique.

11.4 Solution of the particle number density

We now present the numerical solution for Eq. (11.21),⁷⁶⁾ using a commercial program called Mathcad. Since at high temperature $x \ll 1$ the massive particle under consideration is in equilibrium with the cosmic heat bath, its number density will have the equilibrium form. So the boundary condition can be imposed as $Y(0) = Y_{\text{eq}}(0)$. Per the discussion above on the Riccati equation, the numerical solution is unique. As will be demonstrated in the next section, the value of λ on the right-handed side of Eq. (11.21) is generally very large.

We plot in Fig. 11.1 for 5 different values of the parameter $\lambda = 10^n$, $n = 5, 6, 8, 10$ and 12. The cases of $\lambda = 10^5$ and 10^{10} are plotted in [101]⁷⁷⁾, and the case $\lambda = 10^{10}$ in [221]⁷⁸⁾, which agree with the corresponding curves given in Fig. 11.1.

We make several comments below and a couple remarks concerning the numerical work are given in the

⁷⁶⁾Although (11.21) has no closed form solution, approximation with appropriate analytic expression can be used to obtain a reasonably good solution in the case of helium production in primordial nucleosynthesis. See [258] and Section 10.

⁷⁷⁾Figure 3.5, p. 76, [101].

⁷⁸⁾Figure 3.2, p. 189, [221].

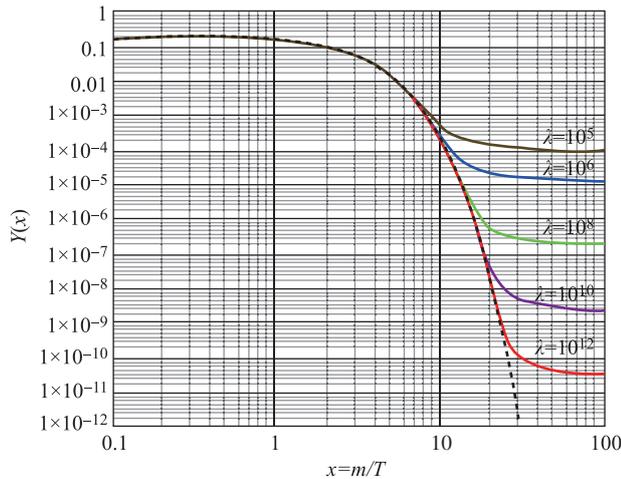


Fig. 11.1 Plot of the numerical solutions of Eq. (11.21). m is the mass of the massive particle. The solid curves of various colors are $Y(x)$ for five values of $\lambda = 10^n$, $n = 5, 6, 8, 10$, and 12 . The asymptotic value $Y(x)$ decreases as n increases. The black dashed curve is for Y_{eq} .

footnote.⁷⁹⁾

- The behavior of the curves in Fig. 11.1 can be understood as follows. $Y(x)$ starts at small x , where the cosmic temperature is much larger than the particle mass, to be the order of its equilibrium value $Y_{\text{eq}}(x)$ as the boundary condition. Because λ in the right-handed side is large and the presence of the x^{-2} which is also large, $Y(x)$ is forced to maintain its closeness to $Y_{\text{eq}}(x)$. Any significant deviation of $Y(x)$ from its equilibrium value will be amplified further by the effect of the $Y(x)^2$ term, and this

⁷⁹⁾ (1) Although, there is not much that can be said about the process of employing a canned numerical program for the general usage, such as Mathcad, there are, however, subtleties in the present case because $Y(x)$ can vary rapidly in a small range of values of x . Let us mention some technical problems in obtaining the plot of Fig. 11.1 in case a reader has the desire to reproduce it, in particular, also using Mathcad, or uses a program which has numerical instability when a small function under consideration varies rapidly. It should be noted first from Fig. 11.1 that the curves begin to drop precipitously around $x = 10$. For the largest λ value considered, i.e., $\lambda = 10^{12}$, Y changes more than 7 orders of magnitude when x changes only 1 order of magnitude: from $Y(7) = 2.8 \times 10^{-3}$ to $Y(27) = 2.1 \times 10^{-10}$. A straightforward application of Mathcad to Eq. (11.21) would not give a sensible result for even $\lambda > 10^7$. The smallness of the value of $Y(x)$ at large x causes numerical instability. However, the differential equation can be rewritten for the function $\sqrt{\lambda}Y(x)$ then Mathcad can be applied to $\lambda = 10^{13}$, but not for still larger λ . Furthermore, the value of x can only be applied for $x \leq 300$, although for small values of λ the applicable range of x can be extended to $x = 600$. We have not studied if there are other transformations of Eq. (11.21) which would allow us to extend the solution to larger values of λ and higher values of x .

(2) We would also like to remark that for technical reasons, we have imposed the boundary condition $Y(x_0) = Y_{\text{eq}}(x_0)$ at $x_0 = 0.01$ rather than at $x_0 = 0$.

would cause $Y(x)$ to have a gigantic derivative which pushes $Y(x)$ back to $Y_{\text{eq}}(x)$. When the cosmic temperature decreases passing the particle mass so that $x > 1$, $Y(x)$ will follow Y_{eq} also decreases exponentially as e^{-x} . The right-handed side becomes small, which gives a small slope for $Y(x)$, hence $Y(x)$ will change very slowly and its value practically freezes, while Y_{eq} still drops exponentially. This is the freeze out phenomenon.

- The above discussion starts out with the boundary condition that $Y(x)$ has its equilibrium value at small $x \ll 1$. How crucial is this boundary condition in the numerical solution? We have found that even if the initial condition at small x , say $x = 0.01$, is away from the equilibrium value, either greater or smaller, the differential equation Eq. (11.21) drives $Y(x)$ to its equilibrium value near $x = 0.1$ so that the right-handed side stays small. Then $Y(x)$ follows the curve as shown in Fig. 11.1. This kind of “restoring force” at small x can be understood simply from the sign of the right-handed side of Eq. (11.21). If $Y(x) > Y_{\text{eq}}$, the derivative of $Y(x)$ is negative which forces $Y(x)$ to decrease towards $Y_{\text{eq}}(x)$. If $Y(x) < Y_{\text{eq}}$, the derivative of $Y(x)$ is positive which forces $Y(x)$ to increase towards $Y_{\text{eq}}(x)$. Hence for small x , $Y(x)$ is forced to be very close to Y_{eq} .
- When λ increases the exponentially decreasing behavior of $Y(x)$ for $x > 1$, as described in the above paragraph, can be maintained to a larger value of x , and hence a small value of the asymptotic value for $Y(x)$. This can also be understood physically as follows: Increasing λ corresponds to increasing the reaction cross section for a fixed m_X and hence the reaction strength, and therefore maintaining the equilibrium state of the massive particle longer to a lower cosmic temperature. Consequently, the freeze out value of $Y(x)$ decreases. Increasing the mass of the massive particle for a fixed cross section will have the same effect.
- Another way to understand the behavior of the solution is from Eq. (11.25). We focus on the terms on the right-handed side of the equation which decides the time behavior of the massive particle number density function n_X . Let us start at an early time or at a high temperature. The temperature behaviors of the quantities in the differential equation are $n_X \sim T^3$ and $H \sim T^2$. So the first term $3Hn_X$ goes like T^5 , and the second term which is proportional to the reaction cross section like T^6 because it is proportional to the density squared. The latter, i.e., the reaction term, would be dominant at high temperature if n_X deviates from its equilibrium form. In the equilibrium form, the time derivative of the particle number density on the left-handed side is given by

the first term of the right-handed side and maintains its equilibrium form, $n_X \sim T^3$. When the temperature decreases, the terms on the right-handed side decrease, with the first term slower than the second term. At some T the first and second terms will become comparable and the particle density is allowed to deviate from the equilibrium form. When the temperature decreases further, the second term, being proportional to the particle density squared, becomes very small and the differential equation is again determined by the rate of expansion of the universe, represented by the first term of the right handed side of Eq. (11.25). So again $n_X \sim T^3$ and Y approaches a constant value, becoming freeze out. Although at both very high and very low temperatures, n_X scales like T^3 , however, with very different coefficients. In the high temperature case, the particle behaves like massless and the coefficient is determined by thermodynamics together with all particles in equilibrium with it. At very low temperature, the particle is non-relativistic and decoupled in free streaming, and the coefficient is determined by its freeze out density. The particle is simply diluted by the expansion of the universe as a function $\sim T^3$.

11.5 Freeze out: Freeze out temperature and particle abundance

Let us summarize. At very high temperature $x = m/T \ll 1$, the dark matter particle is in equilibrium with the cosmic heat bath by its rapid interactions with the rest of the particles of the universe, and it maintains its equilibrium abundance, so that its number density has the T^3 dependence, $n \sim T^3$, i.e., the particle expands with the universe, $n \sim a^{-3} \sim T^3$. As the universe expands further, T decreases and hence x increases. Eventually the cosmic temperature passes below the mass of the dark matter particle which becomes non-relativistic and its equilibrium density decreases exponentially in x , $n \sim e^{-x}$. For sufficiently large x , the equilibrium abundance becomes so rare, and the universe keeps its Hubble expansion, that the particle can no longer be able to interact with the rest of the cosmic contents or to find its antiparticle to annihilate, there begins the on-set of the freeze out. The freeze out density again expands with the universe in the form of $na^3 \sim n/T^3 \sim \text{const}$, as indicated by the first term of the right-handed side of Eq. (11.25), i.e., the particle density is governed by the Hubble expansion. Because of $T \sim a^{-1}$, the dark matter particle maintains its temperature dependence as if it is in thermal equilibrium in the relativistic form. With the solution of the Boltzmann equation, Eq. (11.21), we can determine the epoch of the freeze out and the relic abundance of the dark matter particle.

Next we make a summary of some of the relevant points in the freeze out solution. Given the BTE, a complete numerical solution as shown in Fig. 11.1 depends the boundary condition and the value of the parameter λ . The boundary condition seems unimportant other than the fact it guarantees the uniqueness of the solution. The important quantities to be determined are the freeze out epoch and the asymptotic abundance of the freeze out particle. The freeze out epoch is roughly the x value, denoted as x_f , at which $Y(x)$ begins to separate from $Y_{\text{eq}}(x)$. The freeze out particle abundance today is determined by $Y_\infty = Y(x \rightarrow \infty)$. As we have already alluded to early, these two quantities are actually independent of the boundary condition. The BTE drives the solution to the equilibrium value for $x < 1$, and then follows the unique solution of the freeze out as x increases further. This is the nature of BTE, a non-linear differential equation emerging out of the general properties of the cosmic dynamics. Hence this makes the prediction on the dark matter particle more forceful. Below we discuss the epoch of the freeze out and the asymptotic abundance.

11.5.1 Relic abundance due to freeze out

First we discuss the freeze out abundance of the massive particle. The cosmic relic abundance of massive particles due to freeze out has been studied in detail in the 1970s in a similar physical context [213, 259, 260]. The approximate asymptotic value of $Y(x)$ is given in [221]⁸⁰⁾ as

$$Y_\infty^{(\text{LW})} \sim \frac{6.1}{\lambda^{0.95}}. \tag{11.40}$$

We list in Table 11.1 the values of $Y_\infty^{(\text{LW})}$ and those of our $Y(x)$ evaluated at $x = 300$ for several values of λ . Our $Y(300)$ should be close to the asymptotic value. The last row of the Table lists values of an expression which determines the freeze out epoch to be discussed below. We see that in the range of $\lambda = 10^7$ to 10^{12} the agreement between $Y_\infty^{(\text{LW})}$ and $Y(300)$ is very well, within a few percent. For $\lambda = 10^6$ the agreement is not as good, within about 11%, and for smaller λ the difference is even larger. For much larger λ , say, at $\lambda = 10^{13}$, the agreement is within 22%. This large difference could, at least, be partially accounted for by the fact that at $x = 300$ the asymptotic value for very large λ is not yet reached.

11.5.2 The freeze out epoch

We follow the treatment given in [101]⁸¹⁾ to calculate the freeze out epoch when the massive particle decouples

⁸⁰⁾See Eq. (3.4.11), p. 189, [221] where more references can be found.

⁸¹⁾See, pp 73–78, [101].

Table 11.1 Values at freeze out. The freeze out epoch $x_f = 6.1\lambda^{0.05}$ is given in Eq. (11.44).

λ	10^6	10^7	10^8	10^{10}	10^{12}	10^{13}
$Y_\infty^{(LW)}$	1.22×10^{-5}	1.37×10^{-6}	1.53×10^{-7}	1.93×10^{-9}	2.43×10^{-11}	2.72×10^{-12}
$Y(300)$	1.09×10^{-5}	1.33×10^{-6}	1.58×10^{-7}	2.03×10^{-9}	2.54×10^{-11}	3.49×10^{-12}
$x_f = 6.1\lambda^{0.05}$	12.2	13.7	15.3	19.3	24.3	27.2

from the cosmos heat bath. Let us use T_f to denote the freeze out temperature and the corresponding x is given by $x_f \equiv m_X/T_f$. Let us consider the application of the differential equation in the region $x > x_f$. We rewrite Eq. (11.21)

$$\frac{1}{Y(x)^2} \frac{d}{dx} Y(x) = -\frac{\lambda}{x^2} \left[1 - \left(\frac{Y_{\text{eq}}(x)}{Y(x)} \right)^2 \right] \approx -\frac{\lambda}{x^2}. \tag{11.41}$$

The second line is obtained from the fact that Y_{eq} drops very quickly when the particle begins to freeze out. The term $Y_{\text{eq}}(x)/Y(x)$ becomes small when x increases passing x_f , except in a tiny region next to x_f , as can be seen from Fig. 11.1. Hence $(Y_{\text{eq}}(x)/Y(x))^2$ will make very little contribution when we integrate the equation from x_f to ∞ . The result of the integration is

$$\frac{1}{Y_\infty} - \frac{1}{Y(x_f)} = \frac{\lambda}{x_f}. \tag{11.42}$$

Since $Y(x_f)$ is an order of magnitude larger than Y_∞ as can be seen in Fig. 11.1, the $1/Y(x_f)$ term on the left-handed side can be dropped and we can relate the freeze out temperature to the asymptotic particle density,

$$Y_\infty = \frac{x_f}{\lambda}, \tag{11.43}$$

$$T_f = \frac{m_X}{x_f}.$$

From Eqs. (11.43) and (11.40), we have at the freeze out point

$$x_f = \lambda Y_\infty^{(LW)} = 6.1\lambda^{0.05}. \tag{11.44}$$

Values of x_f for several λ s can be found in Table 11.1. We note that these values agree with the usual statement that the decoupling takes place at approximately $T_f \approx m_X/20$.

We should note that the above evaluation of x_f , and hence T_f , is not exact. The freeze out or decoupling point is usually defined to be the temperature when the particle reaction rate is equal to the Hubble expansion rate.⁸²⁾ To avoid confusion, let us denote the exact freeze out temperature by T_{fe} ,

$$[n_X(T)\langle v\sigma \rangle - H(T)]|_{T=T_{fe}} = 0. \tag{11.45}$$

⁸²⁾See p. 75, [101].

The decoupling temperature obtained above does not satisfy this condition. It can be shown straightforwardly that using the decoupling temperature given in (11.43) and the definition of λ given in (11.22), together with radiation dominance so that $H(m_X) = H(T_f)x_f^2$, we obtain

$$Y_\infty T_f^3 \langle v\sigma \rangle - H(T_f) = 0. \tag{11.46}$$

$Y_\infty T_f^3$ is the freeze out asymptotic number density, not $n_X(T_{fe})$ which may not have a simple T^3 dependence on the temperature. The exact freeze out temperature T_{fe} has to be obtained by evaluating Eq. (11.45).⁸³⁾

11.6 Numerical value of λ and its Physical implications

Let us come back to λ defined in Eq. (11.22) and study its numerical value and its physical implications. First we will examine the general case by leaving the reaction cross section and the mass of the particle free. In this discussion we follow the approach of [101]. Then we follow [221] by assuming a reasonable reaction cross section to see what sort of mass range of the massive particle can be expected.

11.6.1 Estimate of λ

We assume that the massive particle annihilation is of the weak interaction strength.⁸⁴⁾ For $T < m_X$, the massive particle is non-relativistic and the thermal average cross section times the relative velocity of the massive particles in the initial state is given by,⁸⁵⁾ in the natural units,

$$\begin{aligned} \langle v\sigma \rangle &= \frac{G_F^2}{2\pi} m_X^2 \mathcal{C}_X \mathcal{F} \\ &= 5.30 \times 10^{-38} \mathcal{C}_X \mathcal{F} \left(\frac{m_X}{1 \text{ GeV}} \right)^2 \text{ cm}^2 \\ &= 5.30 \times 10^{-2} \mathcal{C}_X \mathcal{F} \left(\frac{m_X}{1 \text{ GeV}} \right)^2 \text{ pb}. \end{aligned} \tag{11.47}$$

Let us briefly discuss how the first line of the above expression is obtained. Not rely on any particular model, there are two energy scales in this situation. Being a weak cross section the factor G_F^2 , which has the dimension of inverse of energy to the 4th power, is expected.

⁸³⁾p. 83, Exercise 10, [101].

⁸⁴⁾We follow the approach of [221], p. 189.

⁸⁵⁾Eq. (3.4.9), p. 189, [221].

The other energy scale is the mass of the dark matter m_X , while most of the normal particles entering in the reactions in the final state can be considered as massless so they cannot determine the dimension of the reaction cross section. Since the left-handed side has the dimension of length squared or energy inverse squared in natural units, so we need the factor m_X^2 to get the right units for the right-handed side of the equation. The rest factors are numerical factors: 2π is a phase factor. \mathcal{C}_X is the number of open channels of normal particles entering in the massive particle annihilation reaction Eq. (11.2). \mathcal{F} , called a fudge factor in [221],⁸⁶⁾ includes all other details of the reaction which we don't expect to deviate from unity too much.

Now we can rewrite the dimensionless λ of Eq. (11.22)

$$\begin{aligned} \lambda &= \frac{1}{2\pi} \sqrt{\frac{45}{4\pi^3}} \frac{\mathcal{C}_X \mathcal{F}}{\sqrt{g_*}} G_F^2 M_P m_X^3 \\ &= 1.5923 \times 10^8 \frac{\mathcal{C}_X \mathcal{F}}{\sqrt{g_*}} \left(\frac{m_X}{1 \text{ GeV}} \right)^3. \end{aligned} \quad (11.48)$$

Both the effective degrees of freedom g_* in the Hubble expansion rate $H(T)$ parameter and the number of annihilation channels \mathcal{C}_X depend on the mass of the heavy particle. From Table 9.4 we see that g_* varies from 72.3 for $T \approx 5$ GeV to 95.3 for $T \approx 100$ GeV, about a 25% variation. So we can treat g_* a constant in this range of the cosmic temperature. In the same range of T , $\mathcal{C}_X/\sqrt{g_*}$ takes the value of around 2.1 to 2.5, so it is the order of 1. This is the justification for considering λ a constant and large.

11.6.2 Fraction of critical density

The contribution of the massive particle at the present time can be calculated if we know its density at a time after freeze out.⁸⁷⁾ Let us assume that $Y(x)$ reaches its asymptotic value at the temperature $T_1 < T_f$. Being a non-relativistic particle and frozen out, its energy density at T_1 is given by, following the expansion of the universe,

$$\rho_{X1} = 2m_X n_X(T_1) = 2m_X Y_\infty T_1^3, \quad (11.49)$$

where the factor 2 on the right-handed side comes from, for the usual case, that the Heavy particle and its antiparticle are not identical. In the case of Majorana fermions or self-conjugate bosons, the factor 2 should be removed. T_1 is presumably still quite high and it is also the temperature of the photon then. But from T_1 down to the present time, the photon will evolve with a higher temperature because of the extra heat put into the universe due to episodes of particle-antiparticle annihilation that can take place during the cosmic evolution.

The explicit example of the neutrino and photon temperature separation has been given in §9.5. It is due to the electron-positron annihilation which takes place after the neutrino is decoupled and when the universe temperature drops below 1 MeV. This episode gives a similar temperature difference between the photon and the massive particle under consideration. In addition, other annihilation processes may happen between T_1 and the cosmic temperature of 1 MeV. Consequently, the temperature difference between the photon and massive particle today can be much larger than that of the photon and neutrino.

Let T_{X0} and T_0 be respectively the present day temperature of the massive particle and that of the photon which defines the temperature of the cosmos. The current density of the massive particle is then

$$\begin{aligned} \rho_{X0} &= \rho_{X1} \frac{T_{X0}^3}{T_1^3} = 2m_X Y_\infty T_0^3 \left(\frac{T_{X0}}{T_0} \right)^3 \\ &\equiv \frac{1}{r_X} 2m_X Y_\infty T_0^3, \\ r_X &\equiv \left(\frac{T_0}{T_{X0}} \right)^3. \end{aligned} \quad (11.50)$$

What is r_X ? As mentioned above the electron-positron annihilation increases the photon temperature by a factor $(11/4)^{1/3} = 1.40$ relative to that of the neutrino. If this is the only such episode reheating of the universe after the massive particle freezes out, then $r_X = (T_0/T_{X0})^3 = 11/4 = 2.75$. This is the value used in [221].⁸⁸⁾ However, in the case that the freeze out takes place in the cosmic temperature range of $T \gtrsim m_Z$, and when all annihilations of particle-antiparticle pairs deposit their energies into the cosmos, then from Table 9.4, we have crudely $T_0/T_{X0} = (381/43)^{1/3} (11/4)^{1/3} = 2.90$ which gives $r_X = 24.3$. So the present day temperature of the massive particle is about 1/3 of that of the photon. This is roughly the value used in [101].⁸⁹⁾ There is about an order of magnitude difference in the two estimates of r_X .

We use the explicit expressions for the asymptotic Y_∞ given in Eq. (11.40) and λ given in (11.48) to calculate the present day energy density of the massive particle,

$$\begin{aligned} \rho_{X0} &= 2m_X \left(\frac{6.1}{\lambda^{0.95}} \right) \frac{T_{\gamma 0}^3}{r_X} \\ &= \left(\frac{\mathcal{C}_X \mathcal{F}}{\sqrt{g_*}} \right)^{-0.95} \left(\frac{m_X}{1 \text{ GeV}} \right)^{-1.85} \left(\frac{T_{\gamma 0}}{2.725 \text{ K}} \right)^3 \\ &\quad \times \frac{1}{r_X} \begin{cases} 3.32 \times 10^{-4} \text{ (GeV/c}^2\text{)} \cdot \text{cm}^{-3} \\ \text{or} \\ 5.92 \times 10^{-28} \text{ g} \cdot \text{cm}^{-3}. \end{cases} \end{aligned} \quad (11.51)$$

⁸⁶⁾p. 189, [221].

⁸⁷⁾We follow the approach of [101], pp 76–77.

⁸⁸⁾Eq. (3.4.12) and the discussion above it, p. 190, [221].

⁸⁹⁾Eq. (5.57) and the discussion above it, p. 76, [101].

Now we can calculate the contribution of the massive particle in terms of the fraction of the critical density.

$$\begin{aligned}\Omega_X &= \frac{\rho_{X0}}{\rho_c} \\ &= 31.50 \left(\frac{C_X \mathcal{F}}{\sqrt{g_*}} \right)^{-0.95} \left(\frac{m_X}{1 \text{ GeV}} \right)^{-1.85} \left(\frac{T_{\gamma 0}}{2.725 \text{ K}} \right)^3 \frac{h^{-2}}{r_X},\end{aligned}\quad (11.52)$$

where $\rho_c = 3H_0^2/(8\pi G_N) = 1.05375 \times 10^{-5} h^2 \text{ (GeV}/c^2) \cdot \text{cm}^{-3}$ is the critical density of the universe.

11.6.3 Mass and annihilation cross section of massive particle

We can express the unknown massive particle mass and the fudge factor in terms of Ω_X ,⁹⁰⁾ fixing $T_{\gamma 0} = 2.725 \text{ K}$,

$$\left(\frac{m_X}{1 \text{ GeV}} \right) \sqrt{\mathcal{F}} \approx 6.46 \left(\frac{C_X}{\sqrt{g_*}} \right)^{-0.51} \left(\frac{1}{r_X} \right)^{0.54} (\Omega_X h^2)^{-0.54},\quad (11.53)$$

where we have approximate $\mathcal{F}^{0.51}$ by $\sqrt{\mathcal{F}}$.

Suppose the massive particle accounts all the missing cold dark matter, $\Omega_X \approx \Omega_{\text{cdm}}$. The latter has the present epoch fractional energy density [11]

$$\begin{aligned}\Omega_{\text{cdm}} &= \Omega_m - \Omega_b - \Omega_\nu \\ &= 0.12 h^{-2}.\end{aligned}\quad (11.54)$$

Further, simplifying the expression $C_X/\sqrt{g_*}$ to 2 which can be justified by looking at Table 9.4, we obtain

$$m_X \sqrt{\mathcal{F}} \approx \begin{cases} 8.25 \text{ GeV}, & r_X = 11/4, \\ 2.54 \text{ GeV}, & r_X = 24.3. \end{cases}\quad (11.55)$$

Substituting the above value of $m_X \sqrt{\mathcal{F}}$ for $r_X = 24.3$ and $C_X \approx 20$ as shown in Table 9.4 into Eq. (11.47), we obtain

$$\langle v\sigma \rangle \approx 7 \text{ pb}.\quad (11.56)$$

Hence with a cross section of the order of 10^{-2} pb , i.e., \mathcal{F} of order 1 we have a massive particle of mass in the range of 10 GeV. For a smaller cross section, of the order of 10^{-4} pb or 10^{-40} cm^2 we have a massive particle of the order of 100 GeV. The fact that with a weak interaction cross section we are able to obtain a reasonable range of mass m_X , this is highly non-trivial and is often referred to as the WIMP miracle.

If several different kinds of particles contribute to dark matter, we have $\Omega_X < \Omega_{\text{cdm}}$. Since m_X is proportional to $(\Omega_X)^{-0.54}$, m_X will be greater than the value given in Eq. (11.55). Hence Eq. (11.55) is a lower limit of the mass of the cold dark matter particles with weak interaction cross sections.

⁹⁰⁾We follow [221] Eq. (3.4.13), p. 190.

12 CMB anisotropy

The electromagnetic radiation, owing to its special properties, is a unique tool in probing various physical phenomena. The tremendous range of the radiation wavelength accessible can provide rich information across wide areas of particle physics, nuclear physics, astrophysics, and cosmology. In cosmology, the crucial role of the photon can be illustrated by the grand unified photon spectra (GUPS) [261] of observable diffuse radiations as shown in Fig. 12.1. Noted that the wavelength covered ranges from 10^5 cm to 10^{-22} cm with the intensity varying over more than 30 orders of magnitude.

Most of the diffuse photons are found in a nearly isotropic background of the thermal spectra around the temperature of 2.75 K today, known as the cosmic microwave background (CMB). The CMB, a cornerstone of the Big Bang cosmology, is usually referred to as the after glow of the Big Bang or the radiation relic at a time when the universe is dense and hot of about 3000 K. The CMB was predicted by Alpher and Gamov in the 1940s [243]. It is very close to be an ideal blackbody spectrum, and therefore isotropic, as predicted by the Boson distribution given in Eq. (9.98), with a vanishing chemical potential $\mu_\gamma = 0$. The agreement of the theoretical thermal spectrum and that of the observation is over three decades of multi-GHz frequency regime, mostly in the microwave frequency range⁹¹⁾. The radiation intensity varies by four orders of magnitude including a maximum around 150 GHz of wavelength about 2 mm. The observed CMB spectrum is shown in Fig. 12.2.⁹²⁾ This background radiation is a necessary consequence of the Hot Big Bang theory⁹³⁾ and it provides very effective probes into the early universe back to the last scattering surface.

The CMB dominates the content of the radiation energy in and beyond our own galaxy and it presents a snapshot of the state of the universe at the last scattering surface (LSS) (see Section 9.5) with the redshift of the order of $z_L = 1100$, when the universe is 380 000 years old and became transparent and the information of the universe at the time imprinted on the CMB photon can be largely preserved. However, there are small deviations from uniformity, known as the *CMB anisotropy*

⁹¹⁾The microwave frequency ranges from 0.3 to 300 GHz, of the wavelength 1 m to 1 mm, and of photon energy 1.24 μeV to 1.24 meV. The upper end of the CMB is actually in the far infrared region, which has frequencies in the range of 300 GHz to 405 THz. In comparison, a common kitchen microwave oven operates at 2.45 GHz, with a power in the range of 700–1000 Watts.

⁹²⁾The figure is taken from <<http://ned.ipac.caltech.edu/level5/Sept05/Gawiser2.html#Figure1>>.

⁹³⁾The discovery of the CMB is a very interesting story, in the spirit of good scientific works and dramatic suspenses. For a description, see, [230], or a search of the subject on the web.

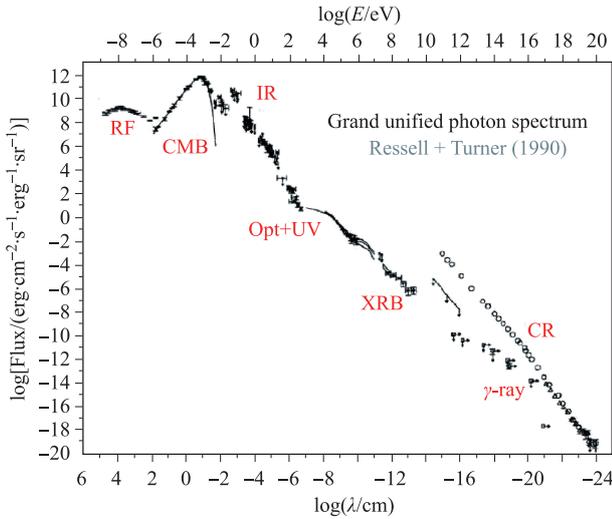


Fig. 12.1 The grand unified photon spectra [74, 261]. The figure given here is taken from [262].

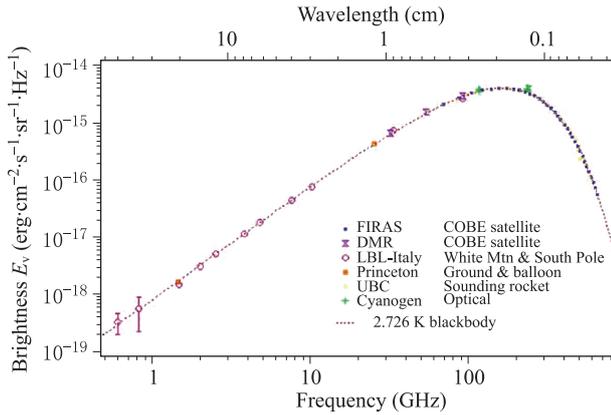


Fig. 12.2 The observed CMB distribution at the temperature 2.736 K. The vertical axis Brightness is defined as the energy density multiplied by the speed of light.

which is represented by tiny temperature fluctuations from point to point across the sky. They provide us a snapshot of the state of the universe about 380 000 years after the big bang as well as some specific features of the universe there after. Therefore, the CMB together with its anisotropy carry the information of how the universe came into being at the very first instance of the cosmos and formed the initial condition for the formation of large structures of the universe observed today. The Study of the CMB anisotropy, i.e., the deviation of CMB from perfect isotropy, plays a crucial role in many aspects of cosmology, such as constraining cosmological models and probing the natures of dark energy, modified gravity, etc. The anisotropy, observed as one part in 10^5 , presents a high accuracy measurement unusual in cosmology. Data on the anisotropy obtained from differ-

ent observational techniques are in excellent agreement. Since there are no preferred directions in the cosmos, we can only expect to obtain statistical information, not specific values, of temperature fluctuations. Hence the quantities of interest will be statistics of the observed pattern of the temperature fluctuation.

Many reviews on CMB anisotropy can be found in the published literature and on the web. We refer to the review appeared in the particle data book [11] for an expert summary. The present Section is an introduction to the subject and provides some of the mathematical tools needed in the formulation of the CMB anisotropy. But our coverage of the various relevant topics is spotty. A full exposition, especially a full scale of cosmologic perturbation theory, which is needed to handle the various sources of the anisotropy is beyond the scope of the present work. All relevant information needed can be found over several chapters in [221]. Other good treatment can be found in, e.g., [223] and [224].

The investigations of the CMB anisotropy allows us to study a wide range of properties of the universe and probe some of the underlying fundamental physics, such as inflation, dark energy, dark matter, topological defects, statistical fluctuations, etc. In broad categories, the investigations include:

- The very early initial conditions of the universe: We can study the power spectra, the types of perturbations to the Friedmann universe, and the nature of the perturbation whether it is Gaussian-like or not. We can study inflation and its possible alternatives.
- The composition of the universe: What stuffs the universe was initially made of and how much for each. What are the matter densities and neutrino masses. They are related to details of peaks and shapes of the multipole expansion of the anisotropy.
- Geometry and topology: The universe's global curvature and topological properties, such as the perturbation angular size and repeated patterns as revealed in the sky.
- Evolution: Studies of the expansion rate as a function of time; reionization which determines Hubble constant H_0 ; the dark energy equation of state w , i.e., the relation between pressure and energy densities. See, Eq. (9.43) and related discussions.
- Astrophysical effects: The Sunyaev–Zel'dovic effect (SZ effect), the effect of the foregrounds, etc.

12.1 Classification of CMB anisotropy and key experiments

12.1.1 Classification of CMB anisotropy

The study of CMB anisotropy is far from straightforward. Observed fluctuations of the CMB spectrum are

generally mixtures of effects of various sources, which can be classified into two broad categories, e.g., depending on when the anisotropy occurs, with the last scattering surface (LSS) as the demarcation: the *primary anisotropy* taking place before the LSS, and the *secondary anisotropy* taking place after the LSS. The latter is also known as the *late time* anisotropy, and the former the *early*, or *primordial* temperature variation. CMB anisotropies are sensitive to the energy and matter content of the universe and the initial conditions of the cosmological perturbation. Hence CMB anisotropies can provide important observational tests for mechanisms of the early universe that explain the origin of the cosmological fluctuations. We first give a brief summary of them and then provide some more details later.

The quantitative expression of the temperature anisotropy is defined by

$$\begin{aligned}\Delta T(\hat{n}) &\equiv T(\hat{n}) - T_0, \\ T_0 &\equiv \int T(\hat{n}) d^2\hat{n},\end{aligned}\quad (12.1)$$

where $T(\hat{n})$ is the temperature measured in the sky along the line of sight in the direction \hat{n} , and T_0 is the uniform temperature obtained by averaging over all directions \hat{n} with the effect of the earth motion removed. By definition, the average of $\Delta T(\hat{n})$ over all directions \hat{n} vanishes.

Primary anisotropy

The *primary anisotropy* is caused by effects which occur prior or at the last scattering surface. The effect left over from the early universe is the order of 10^{-5} . Specifically, the primary anisotropy consists of the following effects:

- (a) The intrinsic matter inhomogeneity which occurs in the very early universe gives rise to temperature variations.
- (b) The Doppler effect caused by velocity fluctuations in the proton-electron plasma at the last scattering surface.
- (c) The Sachs–Wolfe effect due to fluctuations in the gravitational potential which cause red- or blue-shifts at the last scattering surface.

Secondary anisotropy

The *secondary anisotropy* is caused by what happen to the CMB photons on their way from the LSS to the observer, and by the particular kinetic status of the earth (solar system) with respect to CMB. There are generally three effects:

- (a) The earth motion relative the CMB which defines a preferred axis. The distortion due to the earth motion is mainly a dipole effect with an amplitude of the order of 10^{-3} .
- (b) The CMB photons, on their way from the LSS to

the observer, are scattered against charged particles in the intergalactic media and alter their energy/temperature. This gives rise to the *Sunyaev–Zel’dovich effect* which distorts the CMB spectral also of the order of 1 mK.

(c) In traveling from the LSS to the observer, if the CMB photon passes through large scale structures, such as galaxy clusters, it will be affected by the gravitational field of the latter that lie in the path. The CMB photon will be blueshifted as it falls into the gravitational potential well and redshifted as they climb out of the well. This is called the *integrated Sachs–Wolfe effect*. If the gravitational potential fluctuates in time, the blueshift and redshift will not cancel each out, leaving a non-vanishing net integrated Sachs–Wolfe effect.

In Section 12.2 we discuss in some details the various effects of the secondary anisotropy, then in Section 12.3 those of the primary anisotropy.

12.1.2 Key experiments

Since the experimental discovery of the CMB by Penzias and Wilson in 1965 [228], many CMB experiments have been conducted to measure the radiation at different wavelengths and characterize their signatures, mostly with ground- or balloon-based detectors, reaffirming the thermal spectrum of the CMB and the presence of small amount of anisotropies. Among the experiments, three satellite measurements in the passed two decades and a half have methodically transformed in profound ways the study of the CMB into a precession science. These are the COBE, WMAP, and Planck space explorations. COBE (Cosmic Background Explorer 1989–1996) was launched in November 1989 and, within a couple of months, quickly confirmed that the CMB is a blackbody radiation [263]. The cosmic temperature was given as 2.735 ± 0.06 K to be compared with the present day value of 2.7255 ± 0.0006 K. The result of the first year observations was published in 1992 [264], which has determined, to 7σ , a statistically significant temperature anisotropy. During this period, together with other experiments, it has been determined that the universe has a flat geometry. The cosmic strings were ruled out as the leading theory of cosmic structure formation, and it suggests cosmic inflation as the right theory.

Launched in June 2001, WMAP (Wilkinson Microwave Anisotropy Probe 2001–2010) succeeded COBE as the second CMB space mission to make more accurate measurements of large scale anisotropies over the full sky. In its first year data, the WMAP team reconfirmed the big bang and inflation theories. The age of the universe was determined to be 13.7 years and the contents of the universe include 4% baryonic matter, 23% dark matter, and 73% dark energy. The new measurements also shed light on the nature of the dark energy, acting as a sort

of an anti-gravity.⁹⁴⁾ New results have been summarized and announced every tow years. The final two years results including a list of all previous news releases can be found in the final NASA news release.⁹⁵⁾ A list of all WMAP technic papers can be found in the website⁹⁶⁾.

The Planck (Planck Cosmology Probe 2009–2013) was ESA’s (European Space Agency) space mission to observe the first light of the universe to do even more detailed study of the cosmos. In contrast, both COBE and WMAP were US NASA’s scientific programs. The Planck satellite was launched in May, 2009 and ceased operation on October 23, 2013. The Planck data confirm the result of WMAP and found the universe to be slightly older than people have thought earlier at 13.8 billion years. The universe consists of 4.9% baryonic matter, 26.8% dark matter, and 68.3% dark energy. The Planck temperature map suggests that the temperature fluctuations were imprinted on the deep sky when the universe was 370 000 years old. Many of the cosmological parameters have been determined to a one percent level, making it truly a precision science. Information concerning the Planck measurement can be found in the home page of the collaboration⁹⁷⁾ We refer to [265] and [266] for the present status of the Planck results.

A comparison of the resolution of the three highly successful satellite experiments is given in Fig. 12.3⁹⁸⁾ For a list of CMB and anisotropy experiments we refer to the NASA LAMBDA (Legacy Archive for Microwave Background Data Analysis) website [267] which hosts also various information related to CMB.

12.2 Secondary anisotropy

This is also known as the late time anisotropy which has the sources causing the anisotropy to occur after the last scattering.

12.2.1 Anisotropy due to Earth motion

This is an effect of the recent universe due to the relative motion of Earth and the CMB, giving rise to an anisotropy in the observed CMB. It is a Doppler effect of local origin and has an annual modulation. A detailed treatment can be found in [221]⁹⁹⁾. The earth motion (EM) relative to the CMB provides a preferred direction and gives rise to the Doppler effect, hence an apparent

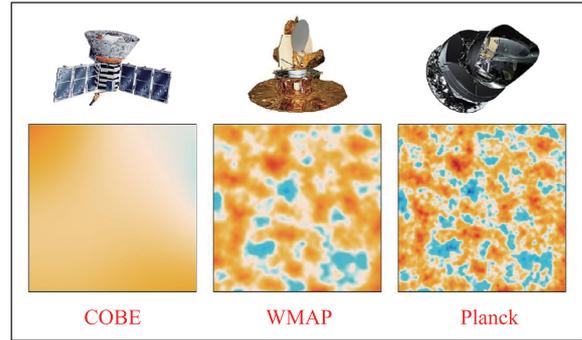


Fig. 12.3 Comparison of the resolutions of COBE, WMAP, and Planck.

anisotropy of the observed CMB is expected. We summarize the result as follows:¹⁰⁰⁾

$$\begin{aligned} \left. \frac{\Delta T(\theta)}{T} \right|_{\text{EM}} &= \frac{1}{\gamma(1 + \beta_{\text{EM}} \cos \theta)} - 1 \\ &= -\beta_{\text{EM}} P_1(\cos \theta) + \frac{\beta_{\text{EM}}^2}{6} (-1 + 4P_2(\cos \theta)) \\ &\quad + \mathcal{O}(\beta_{\text{EM}}^3). \end{aligned} \quad (12.2)$$

θ is the angle between the direction of the earth motion and that of the CMB photon being observed,¹⁰¹⁾ β_{EM} is the velocity of the earth relative to the CMB rest frame in nature units, γ_{EM} is the usual relativistic factor given by $(1 - \beta_{\text{EM}}^2)^{-1/2}$, and $P_j(\cos \theta)$, $j = 1, 2, \dots$ are Legendre polynomials. Note that the contribution to the measurement of the CMB anisotropy is dominated by the dipole form which is linear in β_{EM} . There are also monopole and quadrupole contributions as given by the second term in the second line of Eq. (12.2), but they are proportional to β_{EM}^2 . This quadrupole term is known as the kinematic quadrupole. The presence of the monopole term makes the average of the observed anisotropy over all angles non-vanishing, unlike the primordial anisotropy left over from the early universe which has a vanishing average.

The dipole effect of the earth motion was first observed in 1969 with a rather low level of accuracy. Later observations made in the 1970s established the effect beyond any doubts.¹⁰²⁾ Modern experiments such as WMAP have further clarified the situation, determined the value of β_{EM} and hence the velocity vector of the earth in the cosmic frame of reference in which the CMB is close to be isotropic, called the CMB rest frame. The motion of Earth in the CMB rest frame, known as a pe-

⁹⁴⁾See the news release of the first year results: http://map.gsfc.nasa.gov/news/PressRelease_03-064.html.

⁹⁵⁾<http://map.gsfc.nasa.gov/news/>

⁹⁶⁾http://lambda.gsfc.nasa.gov/product/map/current/map_bibliography.cfm.

⁹⁷⁾<http://www.cosmos.esa.int/web/planck/home>.

⁹⁸⁾The figure is taken from a Wikiwand article entitled *Cosmic Background Explorer*. See, http://www.wikiwand.com/en/Cosmic_Background_Explorer.

⁹⁹⁾See [221], §2.4, pp 129–132.

¹⁰⁰⁾The Legendre polynomials used below are defined by $P_1(x) = x$ and $P_2(x) = (1/2)(3x^2 - 1)$.

¹⁰¹⁾The angle θ can also be defined as the angle between the line of sight of the CMB photon and the direction of the earth motion. Then $\cos \theta$ should be replaced by $-\cos \theta$.

¹⁰²⁾There are controversies about the first discovery of the dipole anisotropy. For a discussion we refer to [268], which also gives the early observations of the dipole anisotropy.

cular velocity, is 369.0 ± 0.9 km/s [270] which gives $\beta_{EM} = 1.23 \times 10^{-3}$.¹⁰³⁾ Hence we see that the dipole anisotropy due to the earth motion is two orders of magnitude larger than the primordial anisotropy of the early universe. The monopole and quadrupole from the former, which are quadratic in β_{EM} , are about an order of magnitude smaller than the primordial anisotropy.

12.2.2 The Sunyaev–Zel’dovich effect

This is also an effect of the recent universe that affects the CMB photon due to scatterings against intergalactic electrons in galaxy clusters. When CMB photons travel from the last scattering surface to the observer on Earth, secondary anisotropy can be produced as the CMB photon goes through the intervening matter of clusters of galaxies that lie in the line of sight of the observer. Some of the low energy CMB photons are bumped into higher energy state after interacting with high energy electrons through the inverse Compton scattering process $\gamma + e^- \rightarrow \gamma + e^-$, while the energy of the outgoing photon is higher than that of the incoming one. The distortion of the CMB spectrum so produced is a function of the frequency of the CMB photon. The result is a deficit of low frequencies but excess of higher frequencies. This is known as the *Sunyaev–Zeldovich effect* (SZE) [271, 272].

Using the present epoch as an illustration, we have the ideal blackbody spectrum of the CMB has the average temperature of 2.725 K which corresponds to the photon average energy 2.35×10^{-4} eV and the average frequency 56.8 GHz. The maximum of the spectrum occurs at 160 GHz corresponding to photon temperature of 7.68 K. In comparison, for the intergalactic media, roughly 10% of the mass of galaxy clusters is made of hot ionized plasma with the temperature of the electrons mostly greater than $10^6 K \sim 86$ eV. This provides the environment for the inverse Compton scattering to redistribute the energy of some of the CMB photons. Again the spectral distortion of the CMB is the order of 1 mK.

The SZE is generally associated with galaxy clusters because of its huge mass and hence the presence of a significant number of high energy electrons. This effect does not take place in individual galaxies which does not have enough mass to cause a detectable frequency shift of the CMB photons. After passing through a massive galaxy cluster the CMB appears to be fainter at lower frequencies but brighter at higher ones. Hence The SZE can be

used to identify clusters and superclusters.¹⁰⁴⁾ The SZE can also arise from low density, warm baryonic gas that may be present between galaxy clusters.

12.2.3 Integrated Sachs–Wolfe effect

The integrated *Sachs–Wolfe effect* (ISWE) [273, 274] arises from the gravitational effect causing frequency shift of CMB photons traveling between the LSS and Earth. It is the first work that predicts the existing of anisotropy in CMB based on general properties of gravity. A straightforward explanation of the physics of the ISWE can be given as follows: As the CMB photon travels from the LSS to an observer, it may encounter matter distributions on its way and is affected by the gravitational effect of the latter. The effect will be detectable for superclusters and supervoid. Entering in a supercluster, say, the CMB photon will blueshifted and hence the energy increased because of the gravitational potential well of the cluster. Entering a supervoid the opposite happens. In a matter dominated universe the matter distribution will not be changes and the redshift and blueshift cancel each other and the net shift of frequency of the photon is zero. However in a universe dominated by dark energy or radiation, the matter distribution expands and gravitation potential well becomes shallower when the CMB photon emerges out of it. Hence the net effect if that the final CMB photon has a slightly higher temperature. The supervoid has the opposite effect. These result in an anisotropy of the CMB photon. The ISWE has been used as another evidence for the presence of dark energy. For a recent review of the ISWE see, [275].

12.3 Primary CMB anisotropy

The primary CMB anisotropy involves of acoustic oscillations and the diffusion damping. The latter is also known as collisionless damping and as Silk damping. We briefly describe them below.

12.3.1 Acoustic oscillation and modes of perturbations

The acoustic oscillations arises from the initial state of minute primordial density variation and the dynamic effects between the expansion of the photon pressure and gravitational attraction in the tightly coupled photon-baryon plasma. In the very early universe, at the temperature much higher than that of the LSS, the baryon matter, consisting of ions and electrons¹⁰⁵⁾ couple tightly with photons in the form of a cosmic plasma. Regions

¹⁰³⁾This peculiar velocity is a result of five contributions: the motion of Earth around the sun (of the speed of about 30 km/s), the motion of the sun in the local standard of rest, the motion of the local standard of rest around the Milky Way (about 220 km/s), the motion of the Milky way in the Local Group, and finally the motion of the Local Group with respect to the CMB rest frame. For more details and references we refer to [269].

¹⁰⁴⁾The PLANCK Collaboration uses precisely the SZE to hunt for galaxy clusters. see: <http://sci.esa.int/science-e/www/subject/index.cfm?fobjectid=48227> for a layman description of this program.

¹⁰⁵⁾In cosmology baryons refer to both atomic nuclei and the electron.

of slightly higher density will have slightly higher temperature, slightly higher photon pressure, and slightly higher gravitational attraction. The opposite is true for a region of slightly lower density. The higher gravity will attract more plasma which tends to collapse the region into dense haloes. But the higher photon number increases the pressure which tends to expand into neighboring regions. These two opposite effects compete to create acoustic oscillations which form characteristic peaks and valleys in the angular power spectrum. Since the photon number density is about 10 orders of magnitude higher than that of the baryon, we only need to consider the effects of the photon in order to understand the physics involved. The photon pressure is given by $P_\gamma = (1/3)\rho_\gamma \sim T^4$, which is proportional to the 4th power of the temperature. The gravity effect is proportional to the particle number density, which, for the photon $n_\gamma \sim T^3$, is proportional to the third power of the temperature. Hence the rate of change of the pressure is higher than that of the gravitational attraction. Therefore, when a higher density region gets increases to a point, the effect of the pressure will over power that of the gravity, and the region expands to become rarefied. This sets a density or pressure variation and produces a acoustic oscillations which give the CMB a characteristic peak and valley structure, analogous to sound waves with oscillation characterized by fundamentals and harmonics.

The peaks contain characteristic physical signatures. The angular scale of the first peak determines the curvature, but not the topology, of the universe. The next peak, with ratio of the odd peaks to the even peaks, determines the reduced baryon density.¹⁰⁶⁾ The third peak obtains information about the dark matter density.¹⁰⁷⁾ The locations of the peaks also give important information about the nature of the primordial density perturbations. There are two fundamental modes of density perturbations, called *adiabatic* and *isocurvature*. A general density perturbation is a mixture of the two modes, and different theories predict different mixtures. We describe briefly the two modes of perturbations. For more details we refer to some accessible literature, e.g., [277] and [278]. We following the discussion given in [277].

Adiabatic (curvature) density perturbations

For an adiabatic perturbation, which is also known as curvature perturbation or isentropic perturbation, the fractional density fluctuations of various components of the universe, radiation and matter, are simply related. More specifically,

$$\frac{1}{1 + w_j} \frac{\delta\rho_j}{\rho_j}$$

¹⁰⁶⁾See, baryons.html, [276].
¹⁰⁷⁾see, driving.html, [276].

are all equal for all cosmic components of radiation and matter, where $\delta\rho_j$ and ρ_j are respectively the density fluctuation and density of the j-th cosmic component, and w_j relates the energy and pressure densities of the jth component through the equation of state.¹⁰⁸⁾ More simply, the adiabatic condition is that the fractional number density fluctuation for all radiation and matter components, $\delta n_j/n_j$, are equal, i.e., $\delta(n_j/n_k) = 0$. Such a restricted form of density fluctuations of the early universe is predicted by the inflationary scenario with one scalar field. To date, all CMB observations are in good agreement with the adiabatic fluctuation of the anisotropy which forms a part of the standard cosmological theory, the Λ CDM model.

Isocurvature density perturbations

The isocurvature fluctuation, also known as the entropy fluctuation perturbation complements the adiabatic perturbation. It deals with the difference of fluctuation of different components of the universe, i.e.,

$$\frac{\delta n_j}{n_j} - \frac{\delta n_k}{n_k} \quad \text{or} \quad \frac{1}{1 + w_j} \frac{\delta\rho_j}{\rho_j} - \frac{1}{1 + w_k} \frac{\delta\rho_k}{\rho_k}.$$

The difference vanishes for adiabatic fluctuations. Cosmic strings would produce mostly isocurvature primordial perturbations.

The CMB spectrum can distinguish between these two modes of perturbations as the predicted peaks are located at distinctively different locations. Isocurvature density perturbations produce a series of peaks whose angular scales¹⁰⁹⁾ are roughly in the ratio 1:3:5:..., while adiabatic density perturbations produce peaks whose locations are in the ratio 1:2:3:... [279]. The most recent experimental result, given by the Planck Collaboration [266], found no evidence of any contributions from the isocurvature fluctuation, consistent with the primordial density perturbations being adiabatic, which provides a key support for inflation, and ruling out many models of structure formation involving, for example, cosmic strings.

12.3.2 Diffusion damping

The diffusion damping, also known as collisionless damping or Silk damping, is caused by two effects, when the treatment of the primordial plasma as a fluid begins to break down: The increasing mean free path of the photon as the primordial plasma becomes increasingly rarefied in an ever expanding universe the finite depth of the last scattering surface (LSS), which causes the mean free

¹⁰⁸⁾ $\mathcal{P}_j = w_j\rho_j$, $w_j = 1/3$ for the photon and neutrinos, and $w_j = 0$ for the baryon matter and dark matter. See, Section 9.1.3.
¹⁰⁹⁾The ℓ -values of the peaks to be discussed in the following section, Section 12.4.

path to increase rapidly during decoupling, even while some Compton scattering is still going on.

These effects contribute about equally to the suppression of the CMB anisotropy at small scales, and give rise to the characteristic exponential damping tail seen in the very small angular scale anisotropy.

The depth of the LSS refers to the fact that the decoupling of the photons and baryons does not happen instantaneously, but instead, requires an appreciable fraction of the age of the Universe up to that era. One method of quantifying how long this process took uses the photon visibility function (PVF). This function is defined so that, denoting the PVF by $P(t)$, the probability that a CMB photon last scattered between time t and $t + dt$ is given by $P(t)dt$.

The maximum of the PVF (the time when it is most likely that a given CMB photon last scattered) is known quite precisely. The first-year WMAP results put the time at which $P(t)$ is maximum at 372 000 years [65]. This is often taken as the “time” at which the CMB formed. However, to figure out how long it took the photons and baryons to decouple, we need a measure of the width of the PVF. The WMAP team finds that the PVF is greater than half of its maximum value (the “full width at half maximum”, or FWHM) over an interval of 115 000 years. By this measure, decoupling took place over roughly 115 000 years, and when it was complete, the universe was roughly 487 000 years old.

We summarize the various features of the CMB anisotropy in a cartoon in Fig. 12.4. The upper figure is adopted from [280] and the lower figure from [281].

12.4 Formulation of anisotropy

We begin by making a multipole expansion of the temperature anisotropy, which is also referred to as the temperature fluctuation, or temperature variation, as defined in Eq. (12.1). This is an expansion along a unit vector \hat{n} , pointing into the sky from the observational point, in terms of spherical harmonics $Y_{\ell m}(\hat{n}) \equiv Y_{\ell m}(\theta, \phi)$, where θ and ϕ are respectively the polar and azimuthal angles that define \hat{n} ,

$$\Delta T(\hat{n}) \equiv \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\theta, \phi), \quad (12.3)$$

where ΔT also depends on other variables which are not shown. The expansion coefficients $a_{\ell m}$ represents what happens at the last scattering surface (LSS), on its way to the observer, and the particular position where the observer is located as well, and

$$a_{\ell m} = \int d\Omega_{\hat{n}} Y_{\ell m}(\theta, \phi) \Delta T(\hat{n}),$$

$$d\Omega_{\hat{n}} \equiv \sin \theta d\theta d\phi, \quad (12.4)$$

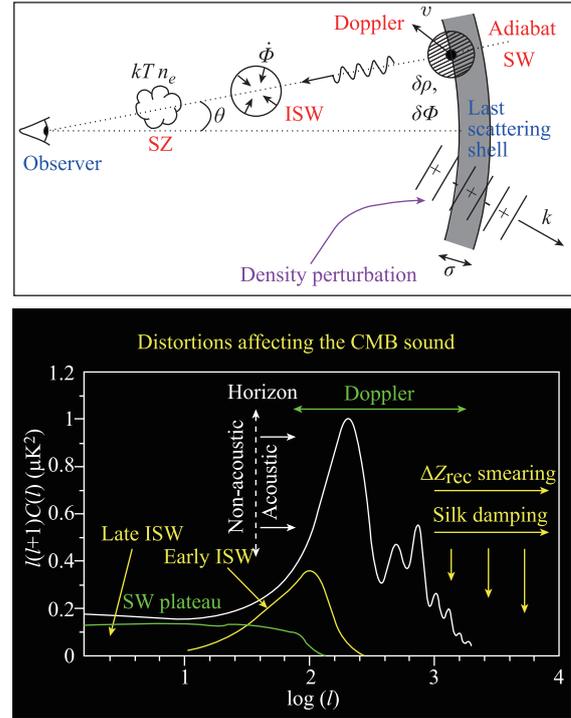


Fig. 12.4 Upper: Cartoon of CMB anisotropy time line [280]. Lower: Acoustic oscillation as a function of the inverse of the angular separation [281]. More details will be given in later sections.

Physically interesting quantities are sometimes studied under some suitable form of average. For the present case, the average can be taken either over all possible positions from which observations of a given point are made, or over times say from a single observation point. These are the ensemble average or the time average. Under rather general conditions these averages are equivalent.¹¹⁰⁾ The average of a quantity will be denoted by $\langle \dots \rangle$. T_0 is the average temperature defined by

$$\begin{aligned} \langle T(\hat{n}) \rangle &= T_0, \\ \langle \Delta T(\hat{n}) \rangle &= 0. \end{aligned} \quad (12.5)$$

So in defining the observed anisotropy, the effect of the earth motion has to be subtracted.

12.4.1 Properties of $a_{\ell m}$, Gaussian distribution, and angular power spectrum

The temperature anisotropy is thought to originate from the perturbation of the very early universe which is dominantly isotropic, in particular at the period of the exponential expansion of cosmic inflation. Given the initial conditions the time evolution of the perturbation is well-known, following the dynamic evolution governed by the

¹¹⁰⁾For more discussions of the average, we refer to [221], p. 136 and its App. D *The Ergodic Theorem*.

Einstein equation. But it is impossible to know how the perturbations are distributed spatially, especially in the environment of exponential spatial expansion. Forming a random distribution, the perturbation is a stochastic process. Hence it is the statistical properties of the perturbation that are interested to us and their measurable consequences that are expected to be uncovered by cosmological observations. Hence in the temperature fluctuation, one do not expect to predict the fluctuation at a particular point in the sky, but rather the average property of the anisotropy. In the following we will make a brief description of the statistical properties of $a_{\ell m}$, defined in Eq. (12.3), which is resulted from the properties of the Gaussian perturbation.¹¹¹⁾

Taking the complex conjugate of Eq. (12.3), from the orthogonality of the spherical harmonics, the reality condition of the temperature variation, we have

$$a_{\ell m} = (-1)^m a_{\ell -m}^* \quad (12.6)$$

The vanishing average of the temperature variation over the whole sky gives

$$\langle a_{\ell m} \rangle = 0. \quad (12.7)$$

We shall assume the expansion coefficient of the CMB temperature variation $a_{\ell m}$ to be Gaussian random variables¹¹²⁾, then

$$\begin{aligned} \langle a_{\ell m} a_{\ell' m'}^* \rangle &= \langle |a_{\ell m}|^2 \rangle \delta_{\ell \ell'} \delta_{m m'} \\ &\equiv C_\ell \delta_{\ell \ell'} \delta_{m m'}, \\ \langle a_{\ell m} a_{\ell' m'} \rangle &= (-1)^{m'} C_\ell \delta_{\ell \ell'} \delta_{m -m'}, \end{aligned} \quad (12.8)$$

where C_ℓ , which is real and non-negative, is known as the (theoretical) *angular power spectrum*¹¹³⁾. The lack of a preferred direction in the cosmos implies that C_ℓ is independent of m .

The angular power spectrum enters the average of the product of the anisotropy distributions in two directions \hat{n} and \hat{n}' :

$$\begin{aligned} \langle \Delta T(\hat{n}) \Delta T(\hat{n}') \rangle &= \sum_{\ell m} \sum_{\ell' m'} \langle a_{\ell m} a_{\ell' m'}^* \rangle Y_{\ell m}(\hat{n}) Y_{\ell' m'}^*(\hat{n}') \\ &= \sum_{\ell m} C_\ell Y_{\ell m}(\hat{n}) Y_{\ell m}^*(\hat{n}') \\ &= \frac{1}{4\pi} \sum_{\ell} (2\ell + 1) C_\ell P_\ell(\cos \theta), \end{aligned} \quad (12.9)$$

where $\cos \theta = \hat{n} \cdot \hat{n}'$. As it will become clear later that in such an expansion, θ and ℓ can be thought as conjugate

¹¹¹⁾From more detailed exposition, we refer to [221, 282], and [224].

¹¹²⁾We refer to Appendix E ‘‘Gaussian Distributions’’ in [221], for a discussion of the Gaussian variables.

¹¹³⁾ C_ℓ is called the *temperature multipole coefficient* in [221], p. 137 and p. 565.

variables. So large angular scale corresponds to small ℓ and vice versa. We can redefined the angular power spectrum as

$$C_\ell = \frac{1}{4\pi} \int \int d\Omega_{\hat{n}} d\Omega'_{\hat{n}'} P_\ell(\hat{n} \cdot \hat{n}') \langle \Delta T(\hat{n}) \Delta T(\hat{n}') \rangle, \quad (12.10)$$

which can readily be verified by means of Eq. (12.9). In specific cosmological models, such as the inflationary model, C_ℓ can be computed explicitly [283, 284]. As a system of Gaussian distributions, the power spectrum given in Eq. (12.8) and the relation Eq. (12.7) determine the average of the product of an arbitrary number of $a_{\ell m}$'s: The averages of the products of an odd number of $a_{\ell m}$'s vanishes, while that of an even number of $a_{\ell m}$'s can be expressed as a sum of products of the power spectrum. For example, the average of the product of four $a_{\ell m}$'s is decomposed as a sum of the various possible products of the average of two $a_{\ell m}$'s,

$$\begin{aligned} &\langle a_{\ell_1 m_1} a_{\ell_2 m_2}^* a_{\ell_3 m_3} a_{\ell_4 m_4}^* \rangle \\ &= \langle a_{\ell_1 m_1} a_{\ell_2 m_2}^* \rangle \langle a_{\ell_3 m_3} a_{\ell_4 m_4}^* \rangle + \langle a_{\ell_1 m_1} a_{\ell_4 m_4}^* \rangle \langle a_{\ell_2 m_2}^* a_{\ell_3 m_3} \rangle \\ &\quad + \langle a_{\ell_1 m_1} a_{\ell_3 m_3} \rangle \langle a_{\ell_2 m_2}^* a_{\ell_4 m_4}^* \rangle \\ &= C_{\ell_1} C_{\ell_3} (\delta_{\ell_1 \ell_2} \delta_{\ell_3 \ell_4} \delta_{m_1 m_2} \delta_{m_3 m_4} + \delta_{\ell_1 \ell_4} \delta_{\ell_2 \ell_3} \delta_{m_1 m_4} \delta_{m_2 m_3}) \\ &\quad + (-1)^{m_3 + m_4} C_{\ell_1} C_{\ell_2} \delta_{\ell_1 \ell_3} \delta_{\ell_2 \ell_4} \delta_{m_1 - m_3} \delta_{m_2 - m_4}. \end{aligned} \quad (12.11)$$

12.4.2 Observed angular power spectrum

The actual observation of the temperature distribution is made at a fixed observational position by varying \hat{n} over the whole sky. The observed angular power spectrum is defined according to Eq. (12.10) but without the average,

$$\begin{aligned} C_\ell^{(o)} &\equiv \frac{1}{4\pi} \int \int d\Omega_{\hat{n}} d\Omega'_{\hat{n}'} P_\ell(\hat{n} \cdot \hat{n}') (\Delta T(\hat{n}) \Delta T(\hat{n}')) \\ &= \frac{1}{4\pi} \int \int d\Omega_{\hat{n}} d\Omega'_{\hat{n}'} \frac{4\pi}{2\ell + 1} \sum_m Y_{\ell m}^*(\hat{n}) Y_{\ell m}(\hat{n}') \\ &\quad \times \sum_{\ell' m'} \sum_{\ell'' m''} a_{\ell' m'} a_{\ell'' m''}^* Y_{\ell' m'}(\hat{n}) Y_{\ell'' m''}^*(\hat{n}') \\ &= \frac{1}{2\ell + 1} \sum_m |a_{\ell m}|^2. \end{aligned} \quad (12.12)$$

So the observed angular power spectrum is approximated by averaging over the magnetic quantum number m , and hence the average of $C_\ell^{(o)}$ is just the angular power spectrum. From Eq. (12.8)

$$\begin{aligned} \langle C_\ell^{(o)} \rangle &= \frac{1}{2\ell + 1} \sum_m \langle a_{\ell m} a_{\ell m}^* \rangle \\ &= C_\ell. \end{aligned} \quad (12.13)$$

If we take $C_\ell^{(o)}$ to be C_ℓ , how good is this approximation? Let us calculate the *cosmic variance*, using Eqs. (12.8),

(12.11), (12.13), and the fact that the $a_{\ell m}$'s are Gaussian distributions,

$$\begin{aligned} & \left\langle \frac{C_{\ell}^{(o)} - C_{\ell}}{C_{\ell}} \cdot \frac{C_{\ell'}^{(o)} - C_{\ell'}}{C_{\ell'}} \right\rangle \\ &= \frac{1}{C_{\ell} C_{\ell'}} \left(\langle C_{\ell}^{(o)} C_{\ell'}^{(o)} \rangle - \langle C_{\ell}^{(o)} \rangle C_{\ell'} - C_{\ell} \langle C_{\ell'}^{(o)} \rangle + C_{\ell} C_{\ell'} \right) \\ &= \frac{1}{C_{\ell} C_{\ell'}} \left[\frac{1}{(2\ell+1)(2\ell'+1)} \sum_{mm'} \langle a_{\ell m} a_{\ell m}^* a_{\ell' m'} a_{\ell' m'}^* \rangle - C_{\ell} C_{\ell'} \right] \\ &= \left(\frac{2}{2\ell+1} \right) \delta_{\ell\ell'}. \end{aligned} \tag{12.14}$$

This result¹¹⁴⁾ demonstrates the following important properties of the cosmic variance: Cosmic variances of different ℓ values are uncorrelated and the cosmic variance decreases with increasing ℓ . For large ℓ the cosmic variance becomes very small and the $C_{\ell}^{(o)}$ is practically the theoretical angular power spectrum. So for $\ell > 5$ we will be able to extract the relevant cosmological information from the measurement of $C_{\ell}^{(o)}$. However, there is a limitation of how large a value of ℓ we can reach. For very large ℓ , e.g., $\ell > 2000$, the Sunyaev-Zeldovich effect will be too strong to allow a useful measurement of C_{ℓ} .¹¹⁵⁾

12.5 Angular sensitivity, multipoles, and angular scales

In measuring the CMB anisotropy, the sensitivity is limited by the angular resolution of the experiment, i.e., the smallest angle of observations that its instrument can resolve. The temperature fluctuation over the sky expressed in terms of the multipole expansion Eq. (12.3) allows us to determine the highest multipole value ℓ_{res} that can be reached in a given experiment, and therefore the kind of physics that can be probed and the sensitivity it has.

12.5.1 Analysis by spherical harmonics

Let us first examine the angular resolution of the spherical harmonics.¹¹⁶⁾ Since we have to define the average in terms of the summation of the magnetic quantum number m , which appears in $Y_{\ell m}(\theta, \phi)$ as a phase factor, we need only to focus on the behavior of the polar angle θ . The following is the argument for determining the angular resolution.

The θ dependence of the spherical harmonics appears in the associated Legendre function $P_{\ell}^m(\cos\theta)$, which can be written in the Fourier form, i.e., linear combinations

¹¹⁴⁾If the observation is made over a fraction of the whole sky f_{sky} the above result will be multiplied by the factor $\sqrt{f_{\text{sky}}}$, i.e., the cosmic variance is degraded by a factor $1/\sqrt{f_{\text{sky}}}$ [285].

¹¹⁵⁾See, p. 138, [221].

¹¹⁶⁾We follow the treatment given in [282].

of $\cos n\theta$ or $\sin n\theta$. The $\cos n\theta$ is for even values of m and $\sin n\theta$ odd values of m , where $n = \ell - 2k \geq 0$, $k = 0, 1, \dots$.¹¹⁷⁾ So the spherical harmonics, as determined by the part of the associated Legendre function, is an oscillating functions in θ . The oscillation increases with increasing value of ℓ , and, for a fixed ℓ , decreases when m increases. Hence for a fixed ℓ , the highest oscillating term is with the 0 magnetic quantum number, i.e., $Y_{\ell 0}$, which is a real function independent of ϕ . As θ increases from 0 to 180° the value of $Y_{\ell m}$ given in the associated Legendre function alternates between positive and negative values, separated into $\ell - |m| + 1$ regions. Plots of $P_{\ell}^m(\theta)$, $m = 0$, $\ell = 7$ and 10, in Fig. 12.5 can illustrate this clearly.¹¹⁸⁾

Therefore, in order to effectively probe the term proportional to $Y_{\ell m}(\cos\theta)$ of the temperature anisotropy, the angular resolution of an experiment has to be no worse than $\pi/(\ell + 1)$. Measured in degrees we have

$$\delta\theta_{\text{res}} \approx \frac{180^\circ}{\ell + 1}. \tag{12.15}$$

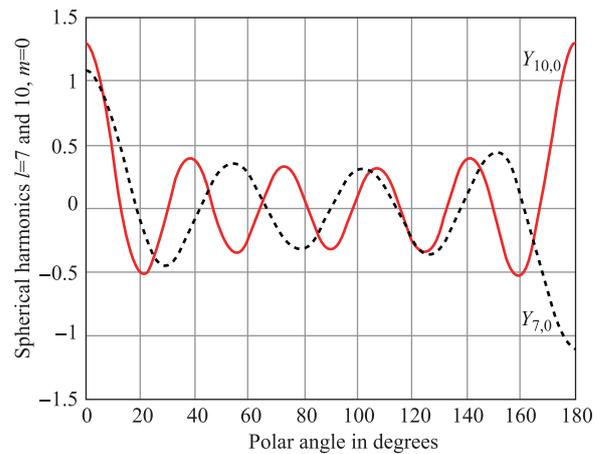


Fig. 12.5 The horizontal axis is θ in degrees. The vertical axis is $Y_{\ell 0}(\theta, \phi)$. The red solid curve is for $\ell = 10$ and the blue dotted curve for $m = 0$. Note the 11 alternate positive and negative valued regions for the case $\ell = 10$ and 8 regions for the case $\ell = 7$.¹¹⁸⁾

¹¹⁷⁾For a list of $Y_{\ell m}$ up to $\ell = 10$, see [286]. The associated Legendre polynomials are commonly expressed in powers of $\cos\theta$ and $\sin\theta$. They can be re-expressed in the Fourier form in terms of $\cos m\theta$ and $\sin m\theta$.

¹¹⁸⁾The functional forms of the two spherical harmonics can be found in [286], and for the convenience of the reader they are listed below:

$$Y_{7,0}(\theta, \phi) = \frac{1}{32} \sqrt{\frac{15}{\pi}} (429\zeta^7 - 693\zeta^5 + 315\zeta^3 - 35\zeta),$$

$$Y_{10,0}(\theta, \phi) = \frac{1}{512} \sqrt{\frac{21}{\pi}} (46189\zeta^{10} - 109395\zeta^8 + 90090\zeta^6 - 30030\zeta^4 + 3465\zeta^2 - 63)$$

$$\zeta \equiv \cos\theta.$$

Hence for an observation with the angular resolution δ_{res} , the largest value of ℓ that can be probed, for large ℓ is

$$\ell_{\text{max}} \approx \frac{180^\circ}{\delta\theta_{\text{res}}}. \tag{12.16}$$

This is not a sharp cutoff, of course. For a given experiment, it may be able to extend its sensitivity somewhat above ℓ_{res} . In an experiment the angular resolution will not be uniform, depending on the frequency of the CMB photons observed, generally the higher the frequency range, the better the angular resolution can be.

For a list of CMB and CMB anisotropy experiments, see [267]. In Table 12.1 we list the largest spherical harmonics, ℓ_{res} , that can be probed in the past experiments which shows great improvement over the past 25 years.

12.5.2 Analysis by Fourier transform

In this subsection We will analyze further the multipole expansion. Let us make a Fourier transform of the temperature anisotropy. It is more quantitative and will be in practical use for the analysis. We write

$$\Delta T(\hat{n}) \equiv f(\hat{n}x). \tag{12.17}$$

For the moment we just mention that x is the comoving distance of the LSS. We will discuss it in more detail later.

We now make a Fourier analysis

$$f(\hat{n}x) \equiv \frac{1}{(2\pi)^{3/2}} \int d^3q e^{i\mathbf{q}\cdot\hat{n}x} \tilde{f}(\mathbf{q}). \tag{12.18}$$

We make a multipole expansion of the exponential function $\exp(i\mathbf{q}\cdot\hat{n}x)$ as well as the Fourier components $\tilde{f}(\mathbf{q})$:

$$\begin{aligned} e^{i\mathbf{q}\cdot\hat{n}x} &= \sum_{\ell m} i^\ell (2\ell + 1) j_\ell(qx) P_\ell(\hat{n}\cdot\hat{q}) \\ &= 4\pi \sum_{\ell m} i^\ell j_\ell(qx) Y_{\ell m}(\hat{n}) Y_{\ell m}^*(\hat{q}), \end{aligned} \tag{12.19}$$

where $q \equiv |\mathbf{q}|$, $\hat{q} \equiv \mathbf{q}/q$. The summation runs for ℓ from 0 to ∞ and for m from $-\ell$ to ℓ . The function j_ℓ is the spherical Bessel function of the first kind which we will discuss in some detail later. Define

$$\tilde{f}(\mathbf{q}) \equiv \sum_{\ell m} \tilde{f}_{\ell m}(q) Y_{\ell m}(\hat{q}). \tag{12.20}$$

Table 12.1 Angular sensitivities of CMB anisotropy experiments.

Experiment	$\delta\theta_{\text{res}}$	ℓ_{max}
COBE [287]	7°	26
WMAP [288]	0.23°	783
BOOMRANG, DASI, MAXIMA	$10' = 0.167^\circ$	1078
Planck [289]	$5' = 0.0833^\circ$	2160

We now write, using known properties of the spherical harmonics,

$$\begin{aligned} f(\hat{n}x) &= \frac{4\pi}{(2\pi)^{3/2}} \sum_{\ell m} \sum_{\ell' m'} i^\ell \int q^2 dq d\Omega_{\hat{q}} \tilde{f}_{\ell' m'}(q) \\ &\quad Y_{\ell' m'}(\hat{q}) j_\ell(qx) Y_{\ell m}(\hat{n}) Y_{\ell m}^*(\hat{q}) \\ &= \frac{4\pi}{(2\pi)^{3/2}} \sum_{\ell m} i^\ell \int q^2 dq \tilde{f}_{\ell m}(q) j_\ell(qx) Y_{\ell m}(\hat{n}). \end{aligned} \tag{12.21}$$

Then we have, from Eq. (12.4),

$$\begin{aligned} a_{\ell m} &= \int d\Omega_{\hat{n}} Y_{\ell m}^*(\hat{n}) f(\hat{n}x) \\ &= \frac{4\pi i^\ell}{(2\pi)^{3/2}} \int q^2 dq \tilde{f}_{\ell m}(q) j_\ell(qx). \end{aligned} \tag{12.22}$$

The observed power spectrum, given in Eq. (12.12) becomes

$$c_\ell^{(o)} = \frac{2}{(2\ell + 1)\pi} \sum_m \left| \int q^2 dq \tilde{f}_{\ell m}(q) j_\ell(qx) \right|^2. \tag{12.23}$$

To proceed we have to know some of the properties of the spherical Bessel function of the first kind $j_\ell(z)$. We will study the scaled function defined by

$$j_\ell^{(r)}(z) \equiv \sqrt{\frac{\ell(\ell + 1)}{2\ell + 1}} j_\ell(z). \tag{12.24}$$

We note that the scaling is for the study of the scaled angular power spectrum, $\ell(\ell + 1)C_\ell^{(o)}$, which is in the form that the data on the angular power spectrum are generally presented.

12.5.3 Some properties of the Spherical Bessel function of the first kind

The spherical Bessel function of the first kind of order ℓ is defined as

$$\begin{aligned} j_\ell(z) &= \sqrt{\frac{\pi}{2z}} J_{\ell+\frac{1}{2}}(z) = (-z)^\ell \left(\frac{1}{z} \frac{d}{dz} \right)^\ell \frac{\sin z}{z} \\ &= -\frac{dj_{\ell-1}}{dz} + (\ell - 1) \frac{j_{\ell-1}}{z} \\ &= z^\ell \sum_{n=0}^{\infty} \frac{(-1)^n}{n!(2\ell + 2n + 1)!!} \left(\frac{z^2}{2} \right)^n, \end{aligned} \tag{12.25}$$

where $J_{\ell+\frac{1}{2}}(z)$ is the Bessel function of the first kind. The spherical Bessel function has some special properties that can be demonstrated by their graphic plots. These properties are useful in the study of the CMB anisotropy.

We need some of the limiting forms of $j_\ell(z)$:

$$j_\ell(z) \approx \begin{cases} \frac{z^\ell}{(2\ell + 1)!!} \left[1 - \frac{z^2}{2(2\ell + 1)} + \dots \right], & z \ll 1, \\ \frac{1}{z} \sin \left(z + \frac{\ell\pi}{2} \right), & z > \ell(\ell + 1). \end{cases} \tag{12.26}$$

For large ℓ ($\ell \gg 1$), neglecting the difference between $\ell + \frac{1}{2}$ and ℓ ,¹¹⁹⁾ we have

$$j_\ell(z) \stackrel{\ell \gg 1}{\approx} \begin{cases} \frac{1}{2\sqrt{\ell z}} \left(1 - \frac{z^2}{\ell^2}\right)^{-1/4} \exp\left(-\ell \left(\ln\left(1 + \sqrt{1 - \frac{z^2}{\ell^2}}\right) - \ln\left(\frac{z}{\ell}\right) - \sqrt{1 - \frac{z^2}{\ell^2}}\right)\right) \sim 0, & 0 < z < \ell, \\ \frac{1}{z} \left(1 - \frac{\ell^2}{z^2}\right)^{-1/4} \cos\left(z\sqrt{1 - \frac{\ell^2}{z^2}} - \ell \cos^{-1}\left(\frac{\ell}{z}\right) - \frac{\pi}{4}\right), & z > \ell. \end{cases} \quad (12.27)$$

From the above expressions, it can be seen that for large ℓ , in the region $z < \ell$, $j_\ell(z)$ is mostly vanishingly small, except for a small region in which z is very close to ℓ . Hence for a good approximation we can set $j_\ell(z)$ to zero for $z < \ell$. Note that the above limiting form are divergent at $z = \ell$, but spherical Bessel functions

are well defined at this particular point. The divergence is due to the approximation made that breaks down at $z = \ell$.

Later we also need first derivatives of large order spherical Bessel functions. They can be derived from Eq. (12.27):

$$j'_\ell(z) \stackrel{\ell \gg 1}{\approx} \begin{cases} \frac{1}{2\ell} \left(1 - \frac{z^2}{\ell^2}\right)^{-3/4} \exp\left(-\ell \left(1 - \sqrt{1 - \frac{z^2}{\ell^2}}\right)\right) \sim 0, & 0 < z < \ell, \\ -\frac{1}{z} \left(1 - \frac{\ell^2}{z^2}\right)^{1/4} \sin\left(z\sqrt{1 - \frac{\ell^2}{z^2}} - \ell \cos^{-1}\left(\frac{\ell}{z}\right) - \frac{\pi}{4}\right), & z > \ell. \end{cases} \quad (12.28)$$

Again for large ℓ , the first derivative of $j_\ell(z)$ in the region $z < \ell$ can be set to zero. Therefore, We need only the explicit expression for the case $z > \ell$.

In Fig. 12.6 we plot $j_\ell^{(r)}(z)$, for $\ell = 1, 10, 50, 100, 200, 500, 1000$. We summarize some of the relevant properties of the scaled spherical Bessel function and its scaled form, Eq. (12.24) as follows:

- They are oscillating functions around zero, and the oscillation becomes rather fast for $z > \ell$.
- They are very small for z small until z is close to $z = \ell$.
- The first maxima of $j_\ell^{(r)}(z)$, where the function has the largest magnitude, occurs to a good approximation at

$$z_{\max}^{(\ell)} \approx \ell(1 + \ell^{-0.69}), \quad (12.29)$$

which is valid for ℓ to be a few up to at least $\ell = 5000$.

- The integration of $j_\ell^{(r)}(z)$ over the range of z from 0 to ∞ is approximately constant, independent of ℓ , having the value approximately 0.89. The main contribution of the integration comes from a small range of values around the first maximum given in Eq. (12.29).
- For $z > \ell$ the envelopes of the values of $|j_\ell(z)|$ and $|j_\ell^{(r)}(z)|$ decrease like z^{-1} .

Properties of the spherical Bessel function have interesting implications as can be seen from Eq. (12.22). The convergence of the q -integral requires that the Fourier component of the temperature fluctuation, $\tilde{f}_{\ell m}(q)$, decreases in some power of q for large q . For $\tilde{f}_{\ell m}(q)$ decreases no less than q^{-2} the integral will be convergent because $j_\ell(qz) \sim (qx)^{-1}$ times an oscillating function as given in Eq. (12.27). Furthermore, if $\tilde{f}_{\ell m}(q)$ does not fluctuate violently, most of the contribution to the integral in Eq. (12.22) comes from a small region near

$$qx \sim z_{\max}^{(\ell)} \sim \ell, \quad (12.30)$$

for the following reasons: For the integration region below $qx \sim \ell$, j_ℓ is very small and so the contribution to the integral Eq. (12.22) is very small. In the region significantly above $qx \sim \ell$, the rapid oscillation of j_ℓ together with the fact that $q^2 \tilde{f}_{\ell m}$ goes to a constant or decreases, the contribution to the integration will also be small.

When we examine the physics at the LSS the coordinate x is the comoving distance of the LSS $d_C(z_{\text{dec}})$. In a flat universe it equals to the corresponding comoving angular diameter distance $d_A^{(c)}(z_{\text{dec}})$. Hence in flat

¹¹⁹⁾The limiting forms are usually given for the Bessel function of the first kind $J_{\ell+\frac{1}{2}}$. See formulae of large order ℓ Bessel functions given in [290], §8.452 and §8.453 respectively in p. 963 and p. 964. Leading order terms of these limiting forms can also be found in [291], formulae 9.3.2 and 9.3.3, p. 365 and p. 366.

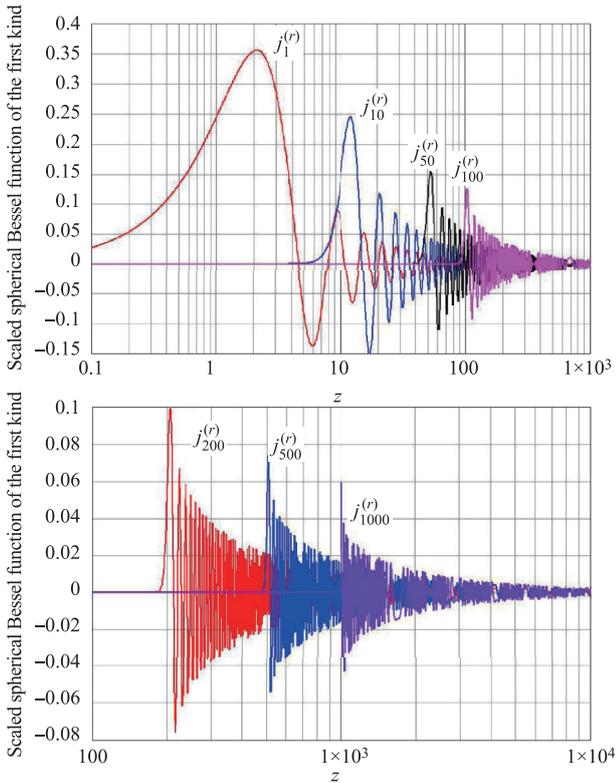


Fig. 12.6 The vertical axis is the scaled spherical Bessel function $j_\ell^{(r)}(z)$, Upper panel: $\ell = 1$ (red), $\ell = 10$ (blue), $\ell = 50$ (black), and $\ell = 100$ (pink); lower panel: $\ell = 200$ (red), $\ell = 500$ (blue), and $\ell = 1000$ (purple).

universe the dominant contribution to the power spectrum C_ℓ , from Eqs. (12.16) and (12.30), comes from the Fourier modes

$$q \approx \frac{\ell}{d_C(z_{\text{dec}})} \leq \frac{1}{d_A^{(c)}(z_{\text{dec}})} \frac{180}{\delta\theta_{\text{res}}}, \tag{12.31}$$

where

$$d_A^{(c)}(z_{\text{dec}}) = \frac{1}{H_0} \int_{\frac{1}{1+z_{\text{dec}}}}^1 \frac{dx}{\sqrt{1-\Omega_\Lambda(1-x^4)+\Omega_M(1-x)}}, \tag{12.32}$$

where Ω_Λ and Ω_M are the density fractions of the dark energy and the matter at the present epoch. We have used the fact that the sum of the present fractions of the dark energy, matter, and radiation saturates the density of the universe, $\Omega_0 = \Omega_\Lambda + \Omega_M + \Omega_R = 1$. In Section 13 we will discuss in some details the various cosmological distances.

12.6 Small fluctuations and linear cosmological perturbation

The CMB anisotropy is an indication of the existence of small deviation from cosmic uniformity in the early uni-

verse. It is the seed that leads to cosmic structures, including CMB anisotropy, the formation of galaxies, clusters, and other features which grow out of the evolution of the universe to produce the lumpiness of the presently observed universe.

The cosmos has been, so far, treated as a uniform and isotropic entity, described as a perfect fluid in an FLRW metric, with the dynamics governed by the Einstein field equation. The features of such a universe can be summarized as follows:

- The perfect fluid, which contains the matter-energy components of the universe, is described by the energy-momentum or the stress-energy tensor given in Eqs. (9.9), (9.26), and (9.27).
- The FLRW metric tensor is given in three different forms in Eqs. (9.16), (9.18), and (9.22). We will ignore the curvature constant by taking $\kappa = 0$, based on the current observation favoring a flat universe.
- The Einstein field equation, which relates the cosmic energy-momentum tensor to the metric tensor, is defined in Eqs. (9.5), (9.7), (9.4), and (9.9).

The theoretical framework, which describes small deviations from the cosmological state of homogeneity and isotropy, is the cosmological perturbation theory, augmented by the invariance of cosmological gauge transformation. Basically, the cosmological perturbation is to add small terms, which are perturbations, to the FLRW metric terms and to the perfect fluid stress-energy tensor, all of the stochastic nature. The cosmological gauge transformation is the freedom in the choice of perturbation functions due to the invariance of Einstein field equation under space-time coordinate transformations. Hence, similar to gauge transformation in field theories, the cosmological gauge transformation also provides the freedom of gauge fixing conditions. The formulation, even restricted to the first order in the perturbation terms, which is a good approximation due to smallness of the anisotropy, involves a large number of perturbative functions entailing rather lengthy algebraic manipulations. It took two Sections of [221], Sections 5 and 6, to present the details, and another chapter, Section 7, to obtain the final formula of the CMB anisotropy in analytical approximation. The perturbation involves totally 20 small parametric functions, 10 each for the metric tensor and the energy-stress tensor. We sketch the analytic result below, following closely the treatment of [221], to which we refer for more details. We note that in field theories the perturbations are small constant parameters, while in cosmology perturbations are small arbitrary functions classified according to their spatial tensor properties.

The cosmological perturbation theory provides a general approach to the *structure formation of the universe*

and it consists of two parts. The first part is the generation of the *primordial inhomogeneity* produced by the quantum fluctuation of the inflation field. This is theoretically very attractive but regarded somewhat speculative. Recently, cosmological observations have provided preliminary indication of inflation. More data will be coming in from newer experiments.

The second part concerns with the growth of the small inhomogeneity, over cosmic time, by gravitational amplification, into the presently observed structures, galaxies, cluster, as well as the cosmic microwave background radiation.

A few words about the jargons commonly used. The original FLRW metric describing a completely homogeneous and isotropic universe is referred to as the *unperturbed universe* or the *background universe*. The FLRW metric is referred to as the *unperturbed metric* or *background metric*. Quantities in the background universe are functions of the cosmic time only, while perturbative functions, which represent the inhomogeneities, depend on the cosmic time as well as the comoving coordinates. We concern ourselves only with the *linear perturbation*, which are the effects up to the first, i.e., the lowest, order in the perturbative quantities.

We will restrict ourselves to the case of the flat universe $\kappa = 0$. This is an important simplification to the formalism. Without the presence of the curvature term, different modes of the plane waver Fourier expansion of perturbation functions of the metric and the dynamic variables, which are related by the Einstein equation and the energy-momentum conservation, are linearly independent.

12.6.1 Linear metric perturbations and their decomposition

The first order perturbation to the metric tensor involves 10 small arbitrary functions which we shall explore below. Let us add a small perturbation, denoted as $h_{\mu\nu}$, to the FLRW metric,

$$\begin{aligned} g_{\mu\nu} &\equiv \bar{g}_{\mu\nu} + h_{\mu\nu}, \\ h_{\mu\nu} &= h_{\nu\mu}. \end{aligned} \tag{12.33}$$

$\bar{g}_{\mu\nu}$ is the homogenous and isotropic FLRW background metric given in Eq. (9.16), depending on the cosmic time t only via the scale factor $a(t)$ of the background metric. In the flat space of $\kappa = 0$, we have

$$\begin{aligned} \bar{g}_{00} &= \bar{g}^{00} = -1, \\ \bar{g}_{0j} &= \bar{g}_{j0} = \bar{g}^{0j} = \bar{g}^{j0} = 0, \\ \bar{g}_{jk} &= \bar{g}_{kj} = a^4 \bar{g}^{jk} = a^4 \bar{g}^{kj} = a^2 \delta_{jk}, \end{aligned} \tag{12.34}$$

which involves only one function $a(t)$, the FLRW or Hubble scale factor. $h_{\mu\nu}$ are perturbations to the elements of

the FLRW metric. Being functions of the cosmic time t and the comoving coordinates x_j , they are no longer homogeneous or isotropic functions, but maintain the symmetric property of the unperturbed metric as indicated in the second equation in (12.33).

$h_{\mu\nu}$ do not form a tensor in either the perturbed or background 4-dimensional spaces. This can be seen as follows. Since $\bar{g}^{\mu\nu}$ is the inverse of $\bar{g}_{\mu\nu}$, i.e., $\bar{g}^{\mu\lambda}\bar{g}_{\lambda\nu} = \delta_\nu^\mu$, and $g^{\mu\nu}$ the inverse of $g_{\mu\nu}$, i.e., $g^{\mu\lambda}g_{\lambda\nu} = \delta_\nu^\mu$. the two relations lead to the identity

$$h^{\mu\nu} = -\bar{g}^{\mu\lambda}\bar{g}^{\nu\rho}h_{\lambda\rho}. \tag{12.35}$$

to the leading order in the perturbative expansion. Due to the negative sign of the right-handed side, $h^{\mu\nu}$ and $h_{\mu\nu}$ are not related as a second rank tensor in either the perturbed or background space of general relativity. In the matrix form

$$h_{\mu\nu} = \begin{pmatrix} h_{00} & h_{0j} \\ h_{0j} & h_{jk} \end{pmatrix}, \quad h^{\mu\nu} = \begin{pmatrix} -h_{00} & \frac{1}{a^2}h_{0j} \\ \frac{1}{a^2}h_{0j} & -\frac{1}{a^4}h_{jk} \end{pmatrix}. \tag{12.36}$$

In general, a two-indexed symmetric function in a four dimensional space, such as $h_{\mu\nu} = h_{\nu\mu}$, is made of 10 independent functions. Dealing with a framework involving so many independent quantities is tedious and requires an efficient way to categorize them according to all possible symmetries available. A useful symmetry is their transformation property in the comoving three dimensional spatial space. The 10 quantities are divided into functions of 3-scalars, 3-vectors, and 3-tensors in the comoving spatial space. It is clear that h_{00} is a 3-scalar, h_{0j} ($j = 1, 2, 3$) a 3-vector, and h_{jk} ($j, k = 1, 2, 3$) a rank-2 3-tensor. Under rather general conditions a 3-vectors, \mathbf{V} , involving three independent components, can be divided into an irrotational (curl-less) part $\mathbf{V}_{//}$ and a solenoidal (divergenceless) part \mathbf{V}_\perp .¹²⁰⁾ In particular, the irrotational part can be written as the divergence of a scalar function, and the divergenceless part can be written as the curl of another vector function. Then

$$\begin{aligned} \mathbf{V} &\equiv \mathbf{V}_{//} + \mathbf{V}_\perp, \\ \nabla \times \mathbf{V}_{//} &= 0, \quad \mathbf{V}_{//} \equiv \nabla \phi, \\ \nabla \cdot \mathbf{V}_\perp &= 0, \quad \mathbf{V}_{//} \equiv \nabla \times \mathcal{V}, \end{aligned} \tag{12.37}$$

where \mathcal{V} is another vector function.

A rank-2 symmetric 3-tensor, having 6 independent components, can be similarly decomposed according to

¹²⁰⁾This is the Helmholtz theorem on the decomposition of vector fields. The vector under consideration is assumed to approach to zero faster than $|\mathbf{x}|^{-1}$ at infinity. We will come back to this condition later in Section 12.8.4.

their trace and divergence properties. The trace property separates the tensor terms into trace and traceless parts. The divergence property divides the terms into number of derivatives and whether or not they are divergenceless. So we can build such a tensor by two scalar functions, a vector function, and a reduced rank-2 symmetric tensor. For the scalar functions, one is multiplied by δ_{jk} , and the other subjects to two derivatives. Both contribute to the trace and divergence of the tensor. The vector function, subjected to one derivative, is divergenceless. The reduced rank-2 tensor is symmetric and traceless, and divergenceless with respect to either of its indices. So the scalar functions contribute to two components, the vector and reduced tensor contribute each also two components. Hence there are totally six independent components.

These totally 10 independent perturbative functions are all functions of the cosmic time t and the comoving spatial coordinates x^j . As functions of t they can potentially grow in magnitude to give rise to sizable structures as observed today. And as functions of x they can lead to inhomogeneity and anisotropy. In details, we have, following the notation of [221],¹²¹⁾

- h_{00} is a 3-scalar, invariant under spatial rotation. We denote

$$h_{00} = -h^{00} \equiv -E. \tag{12.38}$$

- $h_{0j} = h_{j0}$ is a 3-vector subject to the Helmholtz decomposition and we can write

$$\begin{aligned} h_{0j} &= h_{j0} = a^2 h^{0j} = a^2 h^{j0} \equiv a(\partial_j F + G_j), \\ \partial_j G_j &= 0, \end{aligned} \tag{12.39}$$

involving a scalar function F and divergenceless vector function $G - j$. We have used Eq. (12.35) which defines the relation between h_{0j} and h^{0j} . Here and below, repeated spatial indices means summing over the comoving spatial coordinates.

- h_{jk} is a symmetric 3-tensor which can be written as

$$\begin{aligned} h_{jk} &= h_{kj} = -a^4 h^{jk} = -a^4 h^{kj} \\ &\equiv a^2(A\delta_{jk} + \partial_j \partial_k B + \partial_j C_k + \partial_k C_j + D_{jk}), \\ \partial_j C_j &= 0, \\ \partial_j D_{jk} = \partial_k D_{jk} &= 0, \quad D_{jj} = 0. \end{aligned} \tag{12.40}$$

Again, Eq. (12.35) is used to relate h_{jk} and h^{jk} . The 3-tensor is made of two 3-scalar functions A and B and one divergenceless 3-vector C_j , and one divergenceless and traceless symmetric 3-tensor D_{jk} .

In summary, the metric perturbation consists of

- Four 3-scalar function, E, F, A , and B ;
- Two divergenceless 3-vectors, G_j and C_j , each made of two independent functions, and
- One divergence-less and traceless symmetric 3-tensor, D_{jk} , which consists of only two independent functions.
- Summarized in matrix form

$$\begin{aligned} \bar{g}_{\mu\nu} &= \begin{pmatrix} -1 & 0 \\ 0 & a^2 \delta_{jk} \end{pmatrix}, \quad \bar{g}^{\mu\nu} = \begin{pmatrix} -1 & 0 \\ 0 & \frac{1}{a^2} \delta_{jk} \end{pmatrix}, \\ g_{\mu\nu} &= \bar{g}_{\mu\nu} + \delta g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu} \\ &= \begin{pmatrix} -1 + h_{00} & h_{0j} \\ h_{0j} & a^2 \delta_{jk} + h_{jk} \end{pmatrix} = \begin{pmatrix} -1 - E & a(\partial_j F + G_j)(= g_{0j}) \\ g_{0j} & a^2((1 + A)\delta_{jk} + \partial_j \partial_k B + \partial_j C_k + \partial_k C_j + D_{jk}) \end{pmatrix}, \\ g^{\mu\nu} &= \bar{g}^{\mu\nu} + \delta g^{\mu\nu} = \bar{g}^{\mu\nu} - \bar{g}^{\mu\lambda} \bar{g}^{\nu\sigma} \delta g_{\lambda\sigma} \\ &= \begin{pmatrix} -1 - h_{00} & \frac{1}{a^2} h_{0j} \\ \frac{1}{a^2} h_{0j} & \frac{1}{a^2} \delta_{jk} - \frac{1}{a^4} h_{jk} \end{pmatrix} = \begin{pmatrix} -1 + E & \frac{1}{a}(\partial_j F + G_j)(= g^{0j}) \\ g^{0j} & \frac{1}{a^2}((1 - A)\delta_{jk} - \partial_j \partial_k B - \partial_j C_k - \partial_k C_j - D_{jk}) \end{pmatrix}. \end{aligned} \tag{12.41}$$

At the end of the next subsection, in the completion of the presentation on the classification of the perturbation functions and the stress-energy tensor, we will discuss the reason and the use of this classification of the perturbation functions.

12.6.2 Perturbation in the stress-energy tensor

Owing to the symmetry $T_{\mu\nu} = T_{\nu\mu}$, the perturbation to the stress-energy tensor will be made of 10 small functions. Let us start with the stress-energy tensor of the homogeneous and isotropic background universe. This unperturbed stress-energy tensor, denoted by $\bar{T}_{\mu\nu}$, is made of the homogeneous energy and pressure densi-

¹²¹⁾p. 224, [221].

ties, and the velocity 4-vector, described by a perfect fluid as given in Eq. (9.9). As to be described below, the unperturbed velocity vector is a constant 4-vector. Hence $\bar{T}_{\mu\nu}$ is determined by the unperturbed energy and pressure densities of the various components of the background universe which are functions of the cosmic time only. Furthermore, the energy and pressure densities of individual components are related by the equation of the state Eq. (9.43). The most general form of the stress-energy tensor $T_{\mu\nu}$ is more complicated, involving the density and flux of the energy, momentum, pressure, and shear-stress as depicted in Fig. 9.1. T_{00} is the energy density. $T_{0j} = T_{j0}$ is the momentum flux. And T_{jk} has the pressure density as the diagonal elements T_{jj} and the shear stress as the off-diagonal elements T_{jk} , $j \neq k$. For the background universe the energy-momentum tensor $\bar{T}_{\mu\nu}$ has only non-vanishing diagonal terms as given in Eq. (9.27), and all the non-diagonal terms are zero. For the perturbed stress-energy density, all terms can appear with the added terms being small.

The perturbed stress-energy tensor will involve the perturbed energy density

$$\rho \equiv \bar{\rho} + \delta\rho \tag{12.42}$$

and the perturbed pressure density

$$\mathcal{P} \equiv \bar{\mathcal{P}} + \delta\mathcal{P}, \tag{12.43}$$

both hold for the individual components of the universe. The perturbed velocity 4-vector is denote by

$$u_\mu \equiv \bar{u}_\mu + \delta u_\mu. \tag{12.44}$$

$\bar{\rho}$, $\bar{\mathcal{P}}$, and \bar{u}_μ are the unperturbed quantities in a homogeneous and isotropic universe, where \bar{u}_μ is given by Eq. (9.26), and $\delta\rho$, $\delta\mathcal{P}$, and δu_μ are the corresponding small perturbations. The unperturbed energy and pressure densities are functions of the cosmic time t only and defined in terms of the Hubble constant in Eqs. (9.36) and (9.40),

$$\begin{aligned} \bar{\rho} &= \frac{3}{8\pi G_N} \left(\frac{\dot{a}}{a}\right)^2 = \frac{3}{8\pi G_N} H^2, \\ \bar{\mathcal{P}} &= -\frac{1}{8\pi G_N} \left(\frac{2\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right) \\ &= -\frac{1}{8\pi G_N} (2\dot{H} + 3H^2). \end{aligned} \tag{12.45}$$

To proceed further we have to examine more closely the velocity 4-vectors. Let us list the unperturbed velocity 4-vector again (see Eq. (9.26)):

$$\begin{aligned} \bar{g}^{\mu\nu} \bar{u}_\mu \bar{u}_\nu &= -1, \\ \bar{u}^0 &= -\bar{u}_0 = 1, \\ \bar{u}^j &= \bar{u}_j = 0. \end{aligned} \tag{12.46}$$

Normalizing the perturbed velocity 4-vector similarly, we have

$$g^{\mu\nu} u_\mu u_\nu = (\bar{g}^{\mu\nu} + h^{\mu\nu})(\bar{u}_\mu + \delta u_\mu)(\bar{u}_\nu + \delta u_\nu) = -1, \tag{12.47}$$

which gives, to the first order in the perturbation functions, the following condition

$$\bar{g}^{\mu\nu} (\bar{u}_\mu \delta u_\nu + \delta u_\mu \bar{u}_\nu) + h^{\mu\nu} \bar{u}_\mu \bar{u}_\nu = 0. \tag{12.48}$$

Then we obtain,

$$\delta u^0 = \delta u_0 = -\frac{1}{2} h^{00} = \frac{1}{2} h_{00} = -\frac{1}{2} E. \tag{12.49}$$

So δu_0 is not independent, given by the metric perturbation. Only the 3-vector part, denoted as δu_j , is independent, given by one 3-scalar denoted by $\delta u^{(s)}$ and one divergenceless 3-vector $\delta u_j^{(V)}$:

$$\begin{aligned} \delta u_j &\equiv \partial_j \delta u^{(s)} + \delta u_j^{(V)}, \\ \partial_j \delta u_j^{(V)} &= 0. \end{aligned} \tag{12.50}$$

So far we have introduced 5 independent functions: $\delta\rho$, $\delta\mathcal{P}$, $\delta u^{(s)}$, and $\delta u_j^{(V)}$ with $\partial_j \delta u_j^{(V)} = 0$ which define the perturbed perfect fluid part of the stress-energy tensor. This is a direct extension of the corresponding unperturbed expression in the form of a perfect fluid, as given by Eq. (9.9), by replacing all unperturbed quantities by their perturbed counter parts,

$$T_{\mu\nu}^{(pf)} = \mathcal{P} g_{\mu\nu} + (\mathcal{P} + \rho) u_\mu u_\nu. \tag{12.51}$$

However, the most general form of a rank-2 symmetric tensor contains ten independent functions. Five more independent functions can be added to the spatial part of the perfect fluid term. Physically they are the shear-stress components, or the dissipative corrections terms [221] which contribute only to the spatial part of the energy-momentum tensor T_{jk} . We write the dissipative corrections as

$$\begin{aligned} \delta T_{\mu\nu}^{(dc)} &\equiv (1 - \delta_{0\mu})(1 - \delta_{0\nu}) (\partial_\mu \partial_\nu \pi^{(S)} + \partial_\mu \pi_\nu^{(V)} + \partial_\nu \pi_\mu^{(V)} + \pi_{\mu\nu}^{(T)}), \\ \partial_j \pi_j^{(V)} &= 0, \\ \pi_{jk}^{(T)} &= \pi_{kj}^{(T)}, \quad \pi_{jj}^{(T)} = 0, \quad \partial_j \pi_{jk}^{(T)} = \partial_k \pi_{jk}^{(T)} = 0. \end{aligned} \tag{12.52}$$

Now, we have the complete perturbed energy-momentum tensor

$$T_{\mu\nu} \equiv \bar{T}_{\mu\nu} + \delta T_{\mu\nu} = T_{\mu\nu}^{(pf)} + \delta T_{\mu\nu}^{(dc)}, \tag{12.53}$$

which consists of ten independent perturbative functions. They have a structure similar to that of the metric perturbations:

- Four 3-scalar functions $\delta\rho$, $\delta\mathcal{P}$, $\delta u^{(S)}$, and $\pi^{(S)}$;
- Two divergenceless 3-vectors, $\delta u_j^{(V)}$ and $\pi_j^{(V)}$, satisfying the conditions $\partial_j \delta u_j^{(V)} = 0$ and $\partial_j \pi_j^{(V)} = 0$;
- and
- One traceless and divergenceless symmetric tensor $\pi_{jk}^{(T)}$ satisfying the conditions $\pi_{jj}^{(T)} = 0$ and $\partial_j \pi_{jk}^{(T)} = 0$.

From Eq. (12.53), we decompose the first order perturbation terms to the energy-stress tensor according to their 3-spatial symmetry properties:¹²²⁾

$$\begin{aligned}
 \delta T_{00} &= -\bar{\rho}h_{00} + \delta\rho, \\
 \delta T_{0j} &= \delta T_{j0} = \bar{\mathcal{P}}h_{0j} - (\bar{\rho} + \bar{\mathcal{P}}) \left(\partial_j \delta u^{(S)} + \delta u_j^{(V)} \right), \quad \partial_j \delta u_j^{(V)} = 0, \\
 \delta T_{jk} &= \bar{\mathcal{P}}h_{jk} + a^2 \left(\delta\mathcal{P}\delta_{jk} + \partial_j \partial_k \pi^{(S)} + \partial_j \pi_k^{(V)} + \partial_j \pi_k^{(V)} + \pi_{jk}^{(T)} \right), \\
 \partial_j \pi_j^{(V)} &= 0, \quad \pi_{jk}^{(T)} = \pi_{jk}^{(T)}, \quad \pi_{jj}^{(T)} = 0.
 \end{aligned} \tag{12.54}$$

Summarizing in forms of matrices, we have

$$\begin{aligned}
 \bar{u}_\mu &= [-1, \quad 0], \quad \bar{u}^\mu = [1, \quad 0], \\
 u_\mu &= \bar{u}_\mu + \delta u_\mu = \left[-1 - \frac{E}{2}, \quad \partial_j \delta u^{(S)} + \delta u_j^{(V)} \right], \\
 u^\mu &= \bar{u}^\mu + \delta u^\mu = \left[1 - \frac{E}{2}, \quad \frac{1}{a^2} \left(\partial_j \delta u^{(S)} + \delta u_j^{(V)} \right) \right],
 \end{aligned} \tag{12.55}$$

and

$$\begin{aligned}
 T_{\mu\nu} &= \bar{T}_{\mu\nu} + \delta T_{\mu\nu} \\
 &= \begin{bmatrix} (1+E)\bar{\rho} + \delta\rho & a(\partial_j F + G_j)\bar{\mathcal{P}} - (\bar{\rho} + \bar{\mathcal{P}}) \left(\partial_j \delta u^{(S)} + \delta u_j^{(V)} \right) (= T_{0j}) \\ T_{0j} & a^2 \bar{\mathcal{P}} \left((1+A)\delta_{jk} + \partial_j \partial_k B + \partial_j C_k + \partial_k C_j + D_{jk} \right) + a^2 \left(\delta\mathcal{P}\delta_{jk} + \partial_j \partial_k \pi^{(S)} + \partial_j \pi_k^{(V)} + \partial_k \pi_j^{(V)} + \pi_{jk}^{(T)} \right) \end{bmatrix}.
 \end{aligned} \tag{12.56}$$

Other forms of the perturbation terms of the stress-energy tensor are given by

$$\begin{aligned}
 \delta T^{\mu\nu} &= (\bar{g}^{\mu\lambda} \delta g^{\nu\sigma} + \delta g^{\mu\lambda} \bar{g}^{\nu\sigma}) \bar{T}_{\lambda\sigma} + \bar{g}^{\mu\lambda} \bar{g}^{\nu\sigma} \delta T_{\lambda\sigma}, \\
 \delta T_\nu^\mu &= \bar{g}^{\mu\lambda} \delta T_{\lambda\nu} + \delta g^{\mu\lambda} \bar{T}_{\lambda\nu}.
 \end{aligned} \tag{12.57}$$

12.6.2.1 Comments on the tensor decomposition of perturbation functions

We have now completed the modification of the metric tensor and the energy-momentum tensor due to first order perturbation, defined by 20 independent small functions. Let us remark again that the perturbations $h_{\mu\nu}$ and $\delta T_{\mu\nu}$ are not tensors in either the original 4-space with the FLRW metric $\bar{g}_{\mu\nu}$ or in the perturbed 4-space with the metric $g_{\mu\nu}$. Both $\bar{g}_{\mu\nu}$ and $g_{\mu\nu}$ and given in Eq. (12.41).

Let us comment on the reason for making the above decomposition of the perturbation terms in terms of their

¹²²⁾The expressions given in Eq. (12.54) are just the expressions Eqs. (5.1.39)–(5.1.41), [221] p. 225.

3-space tensor properties other than the convenience in classifying them. The additional reason lies in the advantage of the Fourier decomposition of these terms as functions of the comoving coordinates x , in the form of Eq. (12.18). The Einstein field equation, which determine the relationships of the perturbation quantities hold order by order in the perturbation expansion. All terms appear in the Einstein field equation are of the same order. In the present case of the first order calculation, all terms in a given equation are first order and therefore linear in perturbation terms. Hence the different Fourier components are independent.

This property provides a good tool to simplify calculations. Since the derivative $\partial_j = \partial/\partial x_j$ means a multiplication of q_j of the corresponding Fourier component, for arbitrary q_j , the tensor nature of the various terms in the Einstein equation under consideration says that the equation will be held separately for terms of the same nature, i.e., those proportional to δ_{jk} which involves scalar perturbation functions, those having two derivatives which also involve scalar perturba-

tion function, those having one derivative which involves divergence-less vector functions, and those being trace-less and divergence-less tensor terms. Hence each of the four expressions of the Einstein field equation can potentially be separated into four equations: two consists of scalar function, one involving vector functions, and remaining one involving tensor functions. This simplifies greatly the treatment of the perturbation expansion. As we will show in the next subsection that this separation into group of terms is gauge invariant.

12.6.3 Cosmological gauge transformation

Up to now we have been working in a definite coordinate framework of reference, i.e., starting from a definite comoving coordinate system in which the unperturbed system is given by the FLRW metric together with the unperturbed dynamic quantities. However, since our formalism is covariant under four-dimensional coordinate transformations, the choice of a particular coordinate system is, although convenient, but arbitrary. As we can see below, a coordinate transformation will introduce changes in the metric and dynamic quantities. Such changes, being artifacts which do not change any of the original physical system, are called gauge transformations. It is analogous to gauge transformations in field theories due to the invariance of a Lagrangian in a symmetry transformation. In cosmology, gauge transformations involve changes of functional forms of the metric and dynamic quantities, due to the freedom of making coordinate transformation in general relativity. The differences in the various quantities before and after the gauge transformation, referred to as gauge terms, are just artifacts. So the importance of the study of gauge transformations is both to expose the gauge artifact and to explore the possibility of imposing conditions on the metric and/or dynamic terms to simplify calculations.

12.6.3.1 Gauge transformation of an arbitrary tensor

Let us consider a covariant tensor in the perturbed universe

$$\mathcal{F}_{\mu_1\mu_2\dots}(x) \equiv \bar{\mathcal{F}}_{\mu_1\mu_2\dots}(x) + \delta\mathcal{F}_{\mu_1\mu_2\dots}(x), \quad (12.58)$$

where, as in the notation of §12.6.1, $\bar{\mathcal{F}}_{\mu_1\mu_2\dots}(x)$ is the tensor function associated with the background universe and a function of the cosmic time only, and $\delta\mathcal{F}_{\mu_1\mu_2\dots}(x)$ the fluctuation of the function. Under the coordinate transformation

$$x^\mu \rightarrow x'^\mu = x^\mu + \epsilon^\mu(x), \quad (12.59)$$

where ϵ_μ consists of four independent functions. Both $\epsilon(x)^\mu$ and $\partial\epsilon(x)^\mu/\partial x^\nu$ are small functions of x of, e.g., the same order of the cosmological perturbation terms.

Being a tensor of rank n $\mathcal{F}_{\mu\nu\dots}(x)$ is transformed into $\mathcal{F}'_{\mu\nu\dots}(x')$,

$$\mathcal{F}'_{\mu_1\mu_2\dots}(x') = \mathcal{F}_{\nu_1\nu_2\dots}(x) \frac{\partial x^{\nu_1}}{\partial x'^{\mu_1}} \frac{\partial x^{\nu_2}}{\partial x'^{\mu_2}} \dots, \quad (12.60)$$

which affords a functional change of the original tensor, $\mathcal{F} \rightarrow \mathcal{F}'$, and a shift of it argument $x^\mu \rightarrow x'^\mu$. The functional change at a given point x^μ is interpreted as a gauge transformation at the point, it can be calculated as follows.¹²³⁾

We can rewrite Eq. (12.60) as

$$\mathcal{F}'_{\mu_1\mu_2\dots}(x') dx_1'^{\mu_1} dx_2'^{\mu_2} \dots = \mathcal{F}_{\nu_1\nu_2\dots}(x) dx_1^\nu dx_2^\nu \dots \quad (12.61)$$

Let us express $\mathcal{F}'_{\mu_1\mu_2\dots}(x)$ in terms of $\mathcal{F}_{\mu_1\mu_2\dots}(x)$ and ϵ^μ . For the left-handed side of Eq. (12.61), we make an expansion in x_μ

$$\mathcal{F}'_{\mu_1\mu_2\dots}(x') = \mathcal{F}'_{\mu_1\mu_2\dots}(x) + \epsilon^\nu(x) \frac{\partial}{\partial x^\nu} \mathcal{F}_{\mu_1\mu_2\dots}(x) + \mathcal{O}((\epsilon^\mu)^2). \quad (12.62)$$

In calculations below we will keep only terms to the first order in ϵ_μ . For the right-handed side, we express dx^ν in terms of dx'^μ , also to the first order in $\epsilon^\nu(x)$,

$$dx^\nu = \frac{\partial x^\nu}{\partial x'^\lambda} dx'^\lambda = \left(\delta_\lambda^\nu - \frac{\partial \epsilon^\nu(x)}{\partial x^\lambda} \right) dx'^\lambda \quad (12.63)$$

Substituting Eqs. (12.62) and (12.63) into Eq. (12.61) and identifying the coefficients of the product $dx'^{\mu_1} dx'^{\mu_2} \dots$ on the two sides, we have

$$\begin{aligned} \mathcal{F}'_{\mu_1\mu_2\dots}(x) &= \mathcal{F}_{\mu_1\mu_2\dots}(x) - \frac{\partial \epsilon^\nu(x)}{\partial x^{\mu_1}} \mathcal{F}_{\nu\mu_2\dots}(x) \\ &\quad - \frac{\partial \epsilon^\nu(x)}{\partial x^{\mu_2}} \mathcal{F}_{\mu_1\nu\dots}(x) - \dots \\ &\quad - \epsilon^\nu(x) \frac{\partial}{\partial x^\nu} \mathcal{F}_{\mu_1\mu_2\dots}(x), \\ &\equiv \mathcal{F}_{\mu_1\mu_2\dots}(x) + \Delta \mathcal{F}_{\mu_1\mu_2\dots}(x), \\ &\equiv \bar{\mathcal{F}}_{\mu_1\mu_2\dots}(x) + \delta \mathcal{F}_{\mu_1\mu_2\dots}(x) + \Delta \mathcal{F}_{\mu_1\mu_2\dots}(x). \end{aligned} \quad (12.64)$$

We note that to the lowest order in the cosmic fluctuation and the gauge transformation, the fluctuation term of the tensor $\delta\mathcal{F}_{\mu_1\mu_2\dots}(x)$ is not affected. We can now identify the term generated by the gauge transformation to be, to the first order in ϵ^μ ,

$$\begin{aligned} \Delta \mathcal{F}_{\mu_1\mu_2\dots}(x) &\equiv -\epsilon^\nu(x) \frac{\partial}{\partial x^\nu} \bar{\mathcal{F}}_{\mu_1\mu_2\dots}(x) - \frac{\partial \epsilon^\nu(x)}{\partial x^{\mu_1}} \bar{\mathcal{F}}_{\nu\mu_2\dots}(x) \\ &\quad - \frac{\partial \epsilon^\nu(x)}{\partial x^{\mu_2}} \bar{\mathcal{F}}_{\mu_1\nu\dots}(x) - \dots \end{aligned} \quad (12.65)$$

As already being stated above, $\bar{\mathcal{F}}_{\mu_1\mu_2\dots}$ is a function of the cosmic time t only. Obviously this is a first order equation as stated above.

¹²³⁾The consideration given below can be found in [292], pp 80–81.

12.6.3.2 Gauge transformation of the metric and stress-energy tensors

We can rewrite the infinitesimal coordinate shift 4-vector $\epsilon_\mu(x)$ in term of two 3-scalar functions, $\epsilon_0(x)$ and $\epsilon^{(S)}(x)$, and one divergenceless 3-vector functions $\epsilon_j(x)$:

$$\begin{aligned} \epsilon_\mu &\equiv (\epsilon_0, \epsilon_j), \\ \epsilon_j &\equiv \partial_j \epsilon^{(S)} + \epsilon_j^{(V)}, \quad \partial_j \epsilon_j^{(V)} = 0, \end{aligned} \tag{12.66}$$

and to the first order in perturbation

$$\begin{aligned} \epsilon^\mu &= \bar{g}^{\mu\nu} \epsilon_\nu, \\ \epsilon^0 &= -\epsilon_0, \quad \epsilon^j = \frac{1}{a^2} \epsilon_j \equiv \frac{1}{a^2} \left(\partial_j \epsilon^{(S)} + \epsilon_j^{(V)} \right). \end{aligned} \tag{12.67}$$

We list, from Eq. (12.65), the general terms induced by gauge transformation in a 4-scalar \mathcal{S} , 4-vector \mathcal{V}_μ , and 4-tensor $\mathcal{T}_{\mu\nu}$:

$$\begin{aligned} \Delta \mathcal{S} &= -\epsilon^0 \dot{\mathcal{S}}, \\ \Delta \mathcal{V}_\mu &= -\epsilon^0 \dot{\mathcal{V}}_\mu - \bar{\mathcal{V}}_\lambda \frac{\partial \epsilon^\lambda}{\partial x^\mu}, \\ \Delta \mathcal{T}_{\mu\nu} &= -\epsilon^0 \dot{\mathcal{T}}_{\mu\nu} - \bar{\mathcal{T}}_{\mu\lambda} \frac{\partial \epsilon^\lambda}{\partial x^\nu} - \bar{\mathcal{T}}_{\lambda\nu} \frac{\partial \epsilon^\lambda}{\partial x^\mu}. \end{aligned} \tag{12.68}$$

Below we will use $\epsilon_0 = -\epsilon^0$ instead of ϵ^0 .

In applications to scalar quantities we have, from the first expression of Eq. (12.68),

$$\Delta \rho = -\epsilon^0 \dot{\rho} = \epsilon_0 \dot{\rho}, \quad \Delta \mathcal{P} = -\epsilon^0 \dot{\mathcal{P}} = \epsilon_0 \dot{\mathcal{P}}, \tag{12.69}$$

for the energy and pressure 4-scalars densities. The scalar gauge transformation of the first expression above leads to the interesting result

$$\frac{\delta \rho_\alpha}{\dot{\rho}_\alpha} = \dots = \frac{\delta \mathcal{P}_\beta}{\dot{\mathcal{P}}_\beta} = \dots = \epsilon_0, \tag{12.70}$$

where α and β denote any energy or matter components of the universe, known as adiabatic perturbations. See the relevant discussion given in Section 12.3.1. From the second expression of Eq. (12.68), we can write

$$\Delta u_\mu = -\epsilon^0 \dot{u}_\mu - \bar{u}_\lambda \frac{\partial \epsilon^\lambda}{\partial x^\mu} = -\bar{u}_\lambda \frac{\partial \epsilon^\lambda}{\partial x^\mu}, \tag{12.71}$$

where \bar{u}_μ is given in Eq. (12.46). In components, we have

$$\begin{aligned} \Delta \delta u_0 &= \dot{\epsilon}^0 = -\dot{\epsilon}_0, \\ \Delta \delta u_j &= \partial_j \epsilon^0 = -\partial_j \epsilon_0 : \Delta \delta u^{(S)} = -\epsilon_0, \quad \Delta \delta u_j^{(V)} = 0, \end{aligned} \tag{12.72}$$

for the velocity 4-vector function. We note that for pure energy-stress tensor quantities only the gauge parameter ϵ_0 enters.

The gauge transformation of the metric tensor gives, from the third expression of Eq. (12.68),

$$\Delta g_{\mu\nu} = -\epsilon^0 \dot{g}_{\mu\nu} - \bar{g}_{\mu\lambda} \frac{\partial \epsilon^\lambda}{\partial x^\nu} - \bar{g}_{\lambda\nu} \frac{\partial \epsilon^\lambda}{\partial x^\mu}, \tag{12.73}$$

which leads to terms similar to $\delta g_{\mu\nu} \equiv h_{\mu\nu}$ due to metric perturbations. We summarize in matrix form:

$$\begin{aligned} \epsilon_\mu &= \left[\epsilon_0, \quad \partial_j \epsilon^{(S)} + \epsilon_j^{(V)} \right], \\ \epsilon^\mu &= \left[-\epsilon_0, \quad \frac{1}{a^2} \left(\partial_j \epsilon^{(S)} + \epsilon_j^{(V)} \right) \right], \\ \partial_j \epsilon_j^{(V)} &= 0, \end{aligned} \tag{12.74}$$

for the infinitesimal coordinate transformation,

$$\begin{aligned} \Delta u_\mu &= [-\dot{\epsilon}_0, \quad -\partial_j \epsilon_0], \\ \Delta u^\mu &= \left[\dot{\epsilon}_0, \quad -\frac{1}{a^2} \partial_j \epsilon^0 \right], \end{aligned} \tag{12.75}$$

for the gauge transformation of the velocity vector, and

$$\begin{aligned} \Delta g_{\mu\nu} = \Delta h_{jk} &= \begin{bmatrix} \Delta h_{00} & \Delta h_{0j} \\ \Delta h_{0j} & \Delta h_{jk} \end{bmatrix} = \begin{bmatrix} -\Delta E & a(\partial_j \Delta F + \Delta G_j) (= h_{0j}) \\ h_{0j} & a^2 (\Delta A + \partial_j \partial_k \Delta B + \partial_j \Delta C_k + \partial_k \Delta C_j + \Delta D_{jk}) \end{bmatrix} \\ &= \begin{bmatrix} -2\dot{\epsilon}_0 & -\partial_j \epsilon_0 - a^2 \partial_t \left(\frac{1}{a^2} \partial_j \epsilon^{(S)} + \frac{1}{a^2} \epsilon_j^{(V)} \right) (= \Delta h_{0j}) \\ \Delta h_{0j} & 2\epsilon_0 a^2 H \delta_{jk} - \partial_j \left(\partial_k \epsilon^{(S)} + \epsilon_k^{(V)} \right) - \partial_k \left(\partial_j \epsilon^{(S)} + \epsilon_j^{(V)} \right) \end{bmatrix} \end{aligned} \tag{12.76}$$

for the gauge transformation of the metric.

Equation (12.76) shows that the gauge transformation adds terms to the metric perturbation functions E , F , G_j , etc., listed in Eqs. (12.38), (12.39), (12.40). Denoting these additional terms due to gauge transformation by ΔE , ΔF , ΔG_j , etc., we have

$$\Delta E = 2\dot{\epsilon}_0, \tag{12.77}$$

for the 3-scalar part Δg_{00} ,

$$\Delta F = -\frac{\epsilon_0}{a} - a \partial_t \left(\frac{\epsilon^{(S)}}{a^2} \right),$$

$$\Delta G_j = -a\partial_t \left(\frac{\epsilon_j^{(V)}}{a^2} \right), \tag{12.78}$$

for the 3-vector part Δg_{0j} , and

$$\begin{aligned} \Delta A &= 2H\epsilon_0, & \Delta B &= -\frac{2}{a^2}\epsilon^{(S)}, \\ \Delta C_j &= -\frac{1}{a^2}\epsilon_j^{(V)}, & \Delta D_{jk} &= 0, \end{aligned} \tag{12.79}$$

for the 3-tensor part Δg_{jk} . We note that Eq. (12.77) is not independent, it follows from Eq. (12.49) and the

first expression of Eq. (12.72). This also serves as a consistency check.

The information obtainable from gauge transformation of $T_{\mu\nu}$

$$\Delta T_{\mu\nu} = -\epsilon^0 \dot{\bar{T}}_{\mu\nu} - \bar{T}_{\mu\lambda} \frac{\partial \epsilon^\lambda}{\partial x^\nu} - \bar{T}_{\lambda\nu} \frac{\partial \epsilon^\lambda}{\partial x^\mu} \tag{12.80}$$

is already given in Eqs. (12.69) and (12.72),¹²⁴⁾ and they agree with what are expected from Eq. (12.56). We write in matrix form,

$$\Delta T_{\mu\nu} = \begin{bmatrix} \epsilon_0 \dot{\bar{\rho}} + 2\bar{\rho}\dot{\epsilon}_0 & \bar{\rho}\partial_j \epsilon_0 - \bar{\mathcal{P}}\dot{\epsilon}_j + 2H\bar{\mathcal{P}}\epsilon_j \\ \bar{\rho}\partial_j \epsilon_0 - \bar{\mathcal{P}}\dot{\epsilon}_j + 2H\bar{\mathcal{P}}\epsilon_j & \epsilon_0 \partial_t (a^2 \bar{\mathcal{P}}) \delta_{jk} - \bar{\mathcal{P}}(\partial_j \epsilon_k + \partial_k \epsilon_j) \end{bmatrix}, \tag{12.81}$$

where ϵ_j is given in Eq. (12.66). Gauge transformation gives rise to no terms corresponding to dissipative terms of Eq. (12.52). This can be understood from the fact that the terms generated from gauge transformation is determined in part by the unperturbed energy-momentum tensor which has the form of a perfect fluid without any dissipative terms. So we have

$$\Delta\pi^{(S)} = \Delta\pi_j^{(V)} = \Delta\pi_{jk}^{(T)} = 0. \tag{12.82}$$

Now, we have completed the functional changes under gauge transformation.

12.6.3.3 Gauge invariant combinations

Let us first summarize the effect of the gauge transformation on the various scalar perturbation functions, recalling $x_0 \rightarrow x_0 + \epsilon_0$, $x_j \rightarrow x_j + \partial_j \epsilon^{(S)} + \epsilon_j^{(V)}$, $\partial_j \epsilon_j^{(V)} = 0$,

$$\begin{aligned} A &\rightarrow A + 2H\epsilon_0, & \delta\rho &\rightarrow \delta\rho + \dot{\bar{\rho}}\epsilon_0, \\ B &\rightarrow B - \frac{2}{a^2}\epsilon^{(S)}, & \delta\mathcal{P} &\rightarrow \delta\mathcal{P} + \dot{\bar{\mathcal{P}}}\epsilon_0, \\ E &\rightarrow E + 2\dot{\epsilon}_0, & \delta u^{(S)} &\rightarrow \delta u^{(S)} - \epsilon_0, \\ F &\rightarrow F - \frac{1}{a} \left(\epsilon_0 + a^2 \partial_0 \left(\frac{\epsilon^{(S)}}{a^2} \right) \right), & \pi^{(S)} &\rightarrow \pi^{(S)}. \end{aligned} \tag{12.83}$$

There are linear combinations of terms which are gauge invariant. For scalar functions we have the following gauge invariant combinations:

$$\begin{aligned} \hat{A} &\equiv A + 2H\delta u^{(S)}, \\ \hat{E} &\equiv E + 2\delta\dot{u}^{(S)}, \\ \hat{B}_F &\equiv F - \frac{a}{2}\dot{B} - \frac{1}{a}\delta u^{(S)} \equiv B_F - \frac{1}{a}\delta u^{(S)}, & B_F &\equiv F - \frac{a}{2}\dot{B}, \\ \delta\hat{\rho} &\equiv \delta\rho + \dot{\bar{\rho}}\delta u^{(S)} = \delta\rho - 3H(\bar{\rho} + \bar{\mathcal{P}})\delta u^{(S)}, \\ \delta\hat{\mathcal{P}} &\equiv \delta\mathcal{P} + \dot{\bar{\mathcal{P}}}\delta u^{(S)}, \\ \hat{\pi}^{(S)} &\equiv \pi^{(S)}, \end{aligned} \tag{12.84}$$

¹²⁴⁾Explicit expressions for ΔT_{00} , ΔT_{0j} , and ΔT_{jk} are given in Eqs. (5.3.9)–(5.3.11), [221] p. 237.

where Eq. (9.30) has been used to rewrite $\delta\hat{\rho}$. Note that all gauge invariant scalar perturbations are the respective scalar perturbation together with the scalar part of the vector fluid velocity $\delta u^{(S)}$.

Gauge invariant combinations of vector functions are

$$\begin{aligned} \hat{G}_{Cj} &\equiv G_j - a\dot{C}_j, \\ \delta\hat{u}_j &= \delta u_j, & \hat{\pi}_j^{(V)} &= \pi_j^{(V)}. \end{aligned} \tag{12.85}$$

Terms of tensor perturbation are gauge invariant:

$$\hat{D}_{jk} = D_{jk}, \quad \hat{\pi}_{jk}^{(T)} = \pi_{jk}^{(T)}. \tag{12.86}$$

To demonstrate that terms listed in Eq. (12.84) are gauge invariant, let us take the first quantity \hat{A} as an example. Using Eqs. (12.79) and (12.72), we have

$$\begin{aligned} \hat{A} &= A + 2H\delta u^{(S)} \\ &\rightarrow (A + \Delta A) + 2H(\delta u^{(S)} + \Delta u^{(S)}) \\ &= (A + 2H\epsilon_0) + 2H(\delta u^{(S)} - \epsilon_0) \\ &= A + 2H\delta u^{(S)}. \end{aligned} \tag{12.87}$$

12.6.4 Perturbation to cosmological equations: Number of equations and solutions

The Einstein equation is invariant under coordinate transformations and, under perturbation expansion, can be expanded into the zero order equation which describes the homogeneous and isotropic universe, the first order equations which involve linearly the perturbation terms examined above, and higher orders. We concern ourselves only with the first order perturbation. Involving 20 unknown functions, the derivation of the Einstein equation for the first order perturbation term is tedious. Fortunately, the comoving spatial symmetry allows the separation of the perturbation terms into three independent groups: the 3-scalars group containing 8 perturbation functions, the 3-vectors another 8 perturbation functions, and 3-tensors 4 perturbation functions.

This decomposition of perturbation functions into groups greatly simplifies the equations to be solved, although the number of independent equations increases significantly. For the discussion of the CMB anisotropy only the scalar functions are important. So we will concern us with the expressions which involve scalar functions only. The vector terms of the perturbation, satisfying equations of the decaying type and therefore decreasing in importance in the cosmic evolution, can be neglected in the present epoch. The tensor terms contribute to polarizations of CMB which is currently under intensive experimental study.

As discussed in the preceding two sections, there are eight scalar functions: $A, B, E, F, \delta\rho, \delta\mathcal{P}, \delta u^{(S)},$ and $\pi^{(S)}$. They satisfy six equations: four equations obtained from the Einstein equation and two equations from energy-momentum conservation. Let us see how we arrive at this counting of the number of equations.

- The Einstein equation is a symmetric rank-2 tensor equation in the 4-space-time. They can be separated into the $(00), (0j),$ and (jk) components. While the (00) and $(0j)$ components give one equation each, the (jk) component gives two equation. One equation of the (jk) component is proportional to δ_{jk} and the other $\partial_j\partial_k$. So we have four equations from the Einstein equation.
- The energy-momentum conservation is a 4-

dimensional vector equation. The (0) component is the energy conservation, and the (j) component the momentum conservation. So we have two equations from energy-momentum conservation.

Then the total number of independent equations is six as advertised above.

Given 8 independent unknown functions but 6 equations, how are we going to determine these unknown functions uniquely? Here the gauge invariance comes into play. From Eqs. (12.69), (12.72), and (12.77)–(12.79), we see that gauge transformations of the scalar functions are determined by two scalar gauge parameters, ϵ_0 and $\epsilon^{(S)}$. Hence by a judicious choice of the scalar gauge parameters, we can fix the values of two of the scalar perturbation functions, say, by setting them to zero, the remaining six scalar perturbation functions can therefore be uniquely determined once the initial conditions are given.

12.6.4.1 Perturbation of Einstein field equation

The four expressions for scalar perturbations obtained from the Einstein equation are listed in [221].¹²⁵⁾ They are gauge invariant as they should. They can be rewritten in manifestly gauge invariant form by expressing they in terms of the gauge invariant parametric functions defined in Eq. (12.84). We list the gauge invariant form of the four equations below.¹²⁶⁾

$$\begin{aligned} -4\pi G_N (\delta\hat{\rho} - \delta\hat{\mathcal{P}} - \nabla^2\pi^{(S)}) &= \left(-\frac{1}{2}\partial_t^2 - 3H\partial_t + \frac{1}{2a^2}\nabla^2\right) \hat{A} + \left(\frac{1}{2}H\partial_t + 2H^2 + \frac{\ddot{a}}{a}\right) \hat{E} + \frac{H}{a}\nabla^2\hat{B}_F, \\ -16\pi G_N\partial_j\partial_k\pi^{(S)} &= \frac{1}{a^2}\partial_j\partial_k (\hat{A} + \hat{E} + 2a(\partial_t + 2H)\hat{B}_F), \quad \partial_j (\partial_t\hat{A} - H\hat{E}) = 0, \\ -4\pi G_N (\delta\hat{\rho} + 3\delta\hat{\mathcal{P}} + \nabla^2\pi^{(S)}) &= \left(\frac{3}{2}\partial_t^2 + 3H\partial_t\right) \hat{A} - \left(\frac{3}{2}H\partial_t + 3\frac{\ddot{a}}{a} + \frac{1}{2a^2}\nabla^2\right) \hat{E} - \frac{1}{a}(\partial_t + H)\nabla^2\hat{B}_F. \end{aligned} \quad (12.88)$$

A useful equation which we will refer to later can be obtained from Eq. (12.88). Adding 3 times of the first equation, one half of the second equation summed over j and k , and the fourth equation, we obtain¹²⁷⁾

$$-8\pi G_N\delta\hat{\rho} = \left(-3H\partial_t + \frac{1}{a^2}\nabla^2\right) \hat{A} + 3H^2\hat{E} + 2\frac{H}{a}\nabla^2\hat{B}_F. \quad (12.89)$$

Together with the third equation of Eq. (12.88), we have

$$\partial_j \left(8\pi G_N\delta\hat{\rho} + \frac{1}{a^2}\nabla^2\hat{A} + 2\frac{H}{a}\nabla^2\hat{B}_F\right) = 0. \quad (12.90)$$

12.6.4.2 Perturbation of energy-momentum conservation equations

The conservation of energy-momentum, $T_{;\nu}^{\mu\nu} = 0$ gives two equations.¹²⁸⁾ The energy-momentum conservation

¹²⁸⁾See, Eqs. (5.1.48) and (5.1.49), p. 226, [221]. Again the identities $\bar{\rho} = (3/(8\pi G_N))(\dot{a}/a)^2$ and $\bar{\mathcal{P}} = -(1/(8\pi G_N))(2\ddot{a}/a + (\dot{a}/a)^2)$ are used to rewrite them in explicitly gauge invariant forms.

¹²⁵⁾Equations (5.1.44)–(5.1.47), p. 226, [221].
¹²⁶⁾To rewrite Eqs. (5.1.44)–(5.1.47), p. 226, [221] in manifest gauge-invariant expressions in terms of the gauge invariant functions given in Eq. (12.84), we have used Eq. (9.44) and the following zeroth order identities to replace the background energy and pressure densities: $\bar{\rho} = [3/(8\pi G_N)](\dot{a}/a)^2$, $\bar{\mathcal{P}} = -[1/(8\pi G_N)][2\ddot{a}/a + (\dot{a}/a)^2]$, and $4\pi G_N(\bar{\rho} + \bar{\mathcal{P}}) = -\dot{H} = H^2 - \ddot{a}/a$.

¹²⁷⁾In terms of original perturbation functions Eq. (12.89) is

$$-8\pi G_N\delta\rho = -2H\left(\frac{3}{2}\dot{A} + \frac{1}{2}\dot{B}\right) + \frac{1}{a^2}\nabla^2 A + 3H^2 E + \frac{2}{a}H\nabla^2 F,$$
 which we will come back in the discussion of the synchronous gauge in Section 12.7.2.

equations are not independent of the Einstein equation, but they are convenient to use. In gauge invariant quantities, we have

$$\begin{aligned} \partial_j \left(\delta \hat{\mathcal{P}} + \nabla^2 \pi^{(S)} + \frac{1}{2}(\bar{\rho} + \bar{\mathcal{P}})\hat{E} \right) &= 0, \\ \delta \hat{\rho} + 3H \left(\delta \hat{\rho} + \delta \hat{\mathcal{P}} \right) + H \nabla^2 \pi^{(S)} + \frac{3}{2}(\bar{\rho} + \bar{\mathcal{P}})\hat{A} \\ - \frac{1}{a}(\bar{\rho} + \bar{\mathcal{P}})\nabla^2 \hat{B}_F &= 0. \end{aligned} \quad (12.91)$$

Following [221], we remark that the energy-momentum conservation equations can be applied to a component or a subset of components of the universe that is decoupled from the rest of the universe, and hence do not exchange energy and momentum with the latter. Neutrinos and dark matter are such cosmic components. Neutrinos are decouple from the rest of the universe when the universe is the age of about 1 second. Dark matter is decoupled even earlier. We should also note that in the application of the energy-momentum conservation relationship to a subset of component of the universe, the original expressions should be used not their explicitly gauge invariant form. In particular, the first expression in Eq. (12.91) has the original form,

$$\partial_j \left(\delta \mathcal{P} + \nabla^2 \pi^{(S)} + (\partial_t + 3H)((\bar{\rho} + \bar{\mathcal{P}})\delta u^{(S)}) + \frac{1}{2}(\bar{\rho} + \bar{\mathcal{P}})E \right) = 0. \quad (12.92)$$

We note that this equation holds true for an individual system of the universe isolated from, not exchanging energy and momentum with, the rest of the universe, such as the dark matter. This equation will be used in the discussion of the synchronous gauge below.

To conclude this section, we notice that the differential equations derived from the Einstein equation and the energy-momentum conservation equation are real in both the configuration space of (t, \mathbf{x}) and the wave number space (t, \mathbf{q}) . This is a consequence of separating the perturbation functions according to their symmetry properties of scalar, vector, and tensor. The wave number space is the Fourier transformation of the perturbation functions, where \mathbf{q} is the comoving wave number vector. In the wave number space a spatial derivative ∂_j is replaced by iq_j . We shall come back to the wave-number space in the discussion of the evolution of the perturbation functions to be given in Section 12.8.

12.7 Gauge fixing

Because of gauge invariance of the individual equations of the Einstein field equation and the energy momentum conservation equation, perturbation functions generated by a gauge transformation, given in Eqs. (12.77)–(12.79),

and (12.82), which are called gauge terms, satisfy the Einstein field and energy-momentum conservation equations. This is a consequence of the fact that the Einstein field equation Eq. (12.88) and the energy-momentum conservation equation, Eq. (12.91) are linear in the perturbation functions. To the first order in the perturbation, given a set of perturbation functions $\delta\rho, \delta\mathcal{P}, A, \dots$, the gauge transformed set, $\delta\rho + \Delta\rho, \delta\mathcal{P} + \Delta\mathcal{P}, A + \Delta A, \dots$, is a solution as well. Then the set of gauge functions $\Delta\rho, \Delta\mathcal{P}, \Delta A, \dots$ is a solution too. Since the gauge terms $\delta\rho$, etc., are functions of arbitrary gauge parameter functions $\epsilon_\mu(x)$, these types of solutions must be spurious. This creates the gauge ambiguity.

We can also see the problem of the gauge ambiguity from the allowed solutions to the governing dynamic equations of the system under consideration. We have six equations, which include four equations from the Einstein field equation Eq. (12.88) and two equations from the energy-momentum conservation Eq. (12.91). The functions entering these six equations are six gauge invariant perturbation functions, i.e., $\delta\hat{\rho}$, etc., which are linear combinations of the eight perturbation functions $\delta\rho$, etc., as listed in Eq. (12.84). The perturbation function of the velocity potential, $\delta u^{(S)}$, appears in Eqs. (12.88) and (12.91) implicitly, but not explicitly, by helping defining the gauge invariant perturbation functions. The perturbation functions F and B appear in these equations only through their linear combination $B_F \equiv F - (a/2)\dot{B}$. So we can obtain unique solutions for the gauge invariant quantities. But, unless we can determine two of the eight perturbation functions independent of the governing dynamic equations, we will not be able to have a set of unique solutions for the system.

This ambiguity, however, can be turned into advantages, as remarked in the preceding section, by choosing a gauge which allows us to determine the complete set of solutions and possibly to simplify the calculation involved. This is to fix the gauge parameters by choosing some suitable values for $\epsilon_0, \epsilon^{(S)}$, and $\epsilon_j^{(V)}$, which allows us to set some of the metric, density, and/or velocity perturbation functions to zero. As a result, no more gauge transformations are allowed and, in the meantime, some of the equations governing the perturbation functions are simplified. Then the equations, as given in Eqs. (12.88) and (12.91) in the case of scalar perturbations, can be solved uniquely with the needed initial conditions. For the general case, since there are four gauge parameters under a coordinate shift, only four of the perturbation functions, two 3-scalars and one divergenceless 3-vector can be fixed. For the scalar sector which is relevant to the CMB anisotropy power spectrum, the two scalar gauge parameters are just right for what is necessary.

The gauge transformation of the scalar perturbations involve two of the four gauge parameters, ϵ_0 and $\epsilon^{(S)}$.

So we can fix two of the seven perturbation functions E , F , A , B , $\delta\rho$, $\delta\mathcal{P}$, and $\delta u^{(S)}$, while remaining perturbation function $\pi^{(S)}$ is gauge invariant hence not involved in gauge fixing. Let us examine the dependence of the scalar perturbations on the gauge parameters. The gauge transformation of E , A , $\delta\rho$, $\delta\mathcal{P}$, and $\delta u^{(S)}$ depend on ϵ_0 ; B depends on $\epsilon^{(S)}$; F depends on both ϵ_0 and $\epsilon^{(S)}$. So on the two functions to be chosen, either F , or B , or both have to be included. However, there may be complications to some of the gauge choice due to the specific forms of ΔE and ΔF . ΔE , given in Eq. (12.77), depends on the time derivative of ϵ_0 and it can leave undetermined a part of ϵ_0 that depends only on the spatial coordinates. The form of ΔF given in Eq. (12.78) has the following property: $\Delta F = 0$ for

$$\epsilon^{(S)}(t, x) = -a(t)^2 \int^t \frac{\epsilon_0(t', x)}{a(t')^2} dt'. \quad (12.93)$$

Then, unless the chosen gauge conditions can fix either ϵ_0 or $\epsilon^{(S)}$, a residual gauge transformation is possible, hence the perturbation functions involved can not be determined uniquely. We will explore this in the discussion of the synchronous gauge later.

There are two common choices of gauges due to their merits in calculations: the *Newtonian gauge* and the *Synchronous gauge*. Let us briefly discuss them below.

12.7.1 Newtonian gauge

The Newtonian gauge is defined by choosing the gauge functions ϵ_0 and $\epsilon^{(S)}$ so that

$$B + \Delta B = B - 2 \left(\frac{\epsilon^{(S)}}{a^2} \right) = 0, \\ F + \Delta F = F - \frac{1}{a} \left[\epsilon_0 + a^2 \partial_t \left(\frac{\epsilon^{(S)}}{a^2} \right) \right] = 0. \quad (12.94)$$

The first equation fixes $\epsilon^{(S)}$ which, together with the second equation, fixes ϵ_0 . The gauge ambiguity represented by Eq. (12.93) is removed. There is no more gauge freedom to make additional gauge transformations with non-vanishing ϵ_0 and $\epsilon^{(S)}$. In practice, the Newtonian gauge equations are obtained by setting $B = F = 0$ in Eqs. (12.88) and (12.91) and maintaining the same notations for all other perturbation functions.

Usually some of the perturbation functions are renamed, letting

$$E \equiv 2\Phi, \quad A \equiv -2\Psi. \quad (12.95)$$

$$F = 0, \quad B = 0, \quad (12.96)$$

where Φ is referred to as the Newtonian potential and Ψ the Newtonian curvature potential. They are also known as the Bardeen potentials. In this gauge, elements of the

metric scalar perturbation functions are completely fixed and simple

$$g_{00}^{(\text{NG})}|_{\text{scalar}} = -(1 + 2\Phi), \\ g_{0j}^{(\text{NG})}|_{\text{scalar}} = 0, \\ g_{jk}^{(\text{NG})}|_{\text{scalar}} = a^2(1 - 2\Psi)\delta_{jk}. \quad (12.97)$$

The line element is written as

$$ds^{(\text{NG})2}|_{\text{scalar}} = -(1 + 2\Phi)dt^2 \\ + a^2(1 - 2\Psi)\delta_{jk}dx^j dx^k, \quad (12.98)$$

and the stress-energy tensor is

$$T_{00}^{(\text{NG})}|_{\text{scalar}} = (1 + 2\Phi)|\bar{\rho} + \delta\rho, \\ T_{0j}^{(\text{NG})}|_{\text{scalar}} = -(\bar{\rho} + \mathcal{P})\partial_j u^{(S)}, \\ T_{jk}^{(\text{NG})}|_{\text{scalar}} = a^2((1 - 2\Psi)\mathcal{P} + \delta\mathcal{P})\delta_{jk} \\ + a^2\partial_j\partial_k\pi^{(S)}. \quad (12.99)$$

The Einstein field equation in the Newtonian gauge, written out in details from Eq. (12.88)¹²⁹⁾, are

$$-4\pi G_N (\delta\rho - \delta\mathcal{P} - \nabla^2\pi^{(S)}) \\ = (H\partial_t + 2\dot{H} + 6H^2)\Phi + \left(\partial_t^2 + 6H\partial_t - \frac{1}{a^2}\nabla^2\right)\Psi, \\ -8\pi G_N\partial_j\partial_k\pi^{(S)} = \frac{1}{a^2}\partial_j\partial_k(\Phi - \Psi), \\ 4\pi G_N(\bar{\rho} + \bar{\mathcal{P}})\partial_j\delta u^{(S)} = -\partial_j(H\Phi + \partial_t\Psi), \\ 4\pi G_N(\delta\rho + 3\delta\mathcal{P} + \nabla^2\pi^{(S)}) \\ = \left(3H\partial_t + 6\dot{H} + 6H^2 + \frac{1}{a^2}\nabla^2\right)\Phi + (3\partial_t^2 + 6H\partial_t)\Psi. \quad (12.100)$$

The momentum and energy conservation equations, from Eq. (12.91), are¹³⁰⁾

$$\partial_j \left(\delta\mathcal{P} + \nabla^2\pi^{(S)} + \frac{1}{a^3}\partial_t \left(a^3(\bar{\rho} + \bar{\mathcal{P}})\delta u^{(S)} \right) + (\bar{\rho} + \bar{\mathcal{P}})\Phi \right) = 0, \\ \delta\dot{\rho} + 3H(\delta\rho + \delta\mathcal{P}) + \nabla^2 \left(\frac{(\bar{\rho} + \bar{\mathcal{P}})}{a^2}\delta u^{(S)} + H\pi^{(S)} \right) \\ - 3(\bar{\rho} + \bar{\mathcal{P}})\partial_t\Psi = 0. \quad (12.101)$$

Let us mention that in the Newtonian gauge the relation given in Eq. (12.90) becomes particularly simple, setting $B = 0$ and $F = 0$ and write out in terms of the

¹²⁹⁾They can be obtained directly from Eqs. (5.1.44)–(5.1.47), [221]. Note that we have used the identity: $\ddot{a}/a = \dot{H} + H^2$.

¹³⁰⁾They can be seen more directly from Eqs. (5.1.48) and (5.1.49), [221].

perturbation function explicitly

$$\begin{aligned} & \partial_j \left(8\pi G_N \delta \hat{\rho} + \frac{1}{a^2} \nabla^2 \hat{A} - 2 \frac{H}{a^2} \nabla^2 \delta u^{(s)} \right) \\ & = \partial_j \left(8\pi G_N \delta \hat{\rho} - \frac{2}{a^2} \nabla^2 \Psi \right) = 0 \end{aligned} \quad (12.102)$$

This is a constraining equation which will be used in Section 12.8 for the derivation of a crucial conservation relation in the evolution of the early universe.

12.7.2 Synchronous gauge

The synchronous gauge, which will be used in the calculation of the CMB anisotropy, is defined by setting

$$\begin{aligned} E + \Delta E &= E + 2\dot{\epsilon}_0 = 0, \\ F + \Delta F &= F - \frac{1}{a} \left[\epsilon_0 + a^2 \partial_t \left(\frac{\epsilon^{(S)}}{a^2} \right) \right] = 0. \end{aligned} \quad (12.103)$$

The scalar part of the metric is

$$\begin{aligned} g_{00}^{(SG)}|_{\text{scalar}} &= -1, \quad g_{0j}^{(SG)}|_{\text{scalar}} = 0, \\ g_{jk}^{(SG)}|_{\text{scalar}} &= a^2 [(1 + A)\delta_{jk} + \partial_j \partial_k B] \\ &= a^2 [(1 - 2\Psi)\delta_{jk} + \partial_j \partial_k B]. \end{aligned} \quad (12.104)$$

The stress-energy tensor is

$$\begin{aligned} T_{00}^{(SG)}|_{\text{scalar}} &= \bar{\rho} + \delta\rho, \\ T_{0j}^{(SG)}|_{\text{scalar}} &= -(\bar{\rho} + \bar{\mathcal{P}}) \partial_j u^{(S)}, \\ T_{jk}^{(SG)}|_{\text{scalar}} &= a^2 \mathcal{P} [(1 + A)\delta_{jk} + \partial_j \partial_k B] \\ &\quad + a^2 (\delta\mathcal{P} + \partial_j \partial_k \pi^{(S)}) \\ &= a^2 \mathcal{P} [(1 - 2\Psi)\delta_{jk} + \partial_j \partial_k B] \\ &\quad + a^2 (\delta\mathcal{P} + \partial_j \partial_k \pi^{(S)}). \end{aligned} \quad (12.105)$$

Let us list the resultant field equations in the synchronous gauge, as obtained from Eqs. (12.88) and (12.91). They are given in [221] in terms of the original perturbation functions. The Einstein field equations are¹³¹⁾

$$\begin{aligned} & -4\pi G_N (\delta\rho - \delta\mathcal{P} - \nabla^2 \pi^{(S)}) \\ & = \frac{1}{2} \nabla^2 \left(\frac{1}{a^2} A - H\dot{B} \right) - \frac{1}{2} (\partial_t + 6H)\dot{A}, \\ & -16\pi G_N \partial_j \partial_k \pi^{(S)} = \frac{1}{a^2} \partial_j \partial_k (A - (\partial_t + 3H)\dot{B}), \\ & 8\pi G_N (\bar{\rho} + \bar{\mathcal{P}}) \partial_j \delta u^{(S)} = \partial_j \dot{A}, \\ & -4\pi G_N (\delta\rho + 3\delta\mathcal{P} + \nabla^2 \pi^{(S)}) \\ & = \frac{1}{a^2} \left[a^2 \left(\frac{3}{2} \dot{A} + \nabla^2 \dot{B} \right) \right] \equiv \frac{1}{a^2} \partial_t (a^2 \psi), \end{aligned} \quad (12.106)$$

¹³¹⁾See, in Eqs. (5.3.28)–(5.3.31), [221].

where it has been defined

$$\psi \equiv \frac{1}{2} (3\dot{A} + \nabla^2 \dot{B}). \quad (12.107)$$

The energy and momentum conservation equations are respectively¹³²⁾

$$\begin{aligned} & \delta\dot{\rho} + 3H(\delta\rho + \delta\mathcal{P}) + \nabla^2 \left(\frac{1}{a^2} (\bar{\rho} + \bar{\mathcal{P}}) \delta u^{(S)} + H\pi^{(S)} \right) \\ & + (\bar{\rho} + \bar{\mathcal{P}})\psi = 0, \\ & \partial_j \left(\delta\mathcal{P} + \nabla^2 \pi^{(S)} + (\partial_t + 3H) \left((\bar{\rho} + \bar{\mathcal{P}}) \delta u^{(S)} \right) \right) = 0. \end{aligned} \quad (12.108)$$

Equation (12.89)¹³³⁾, which will be useful later, takes the following simple form in the synchronous gauge:

$$-8\pi G_N \delta\rho = \frac{1}{a^2} \nabla^2 A - 2H\psi. \quad (12.109)$$

Let us take a close look at the synchronous gauge conditions Eq. (12.103). Both conditions are potentially problematic. The first condition which determines only the time dependent part of ϵ_0 allows the addition to ϵ_0 an arbitrary function which depends only on spatial coordinates. The second condition allows a corresponding part added to $\epsilon^{(S)}$ as shown in Eq. (12.93). This freedom is the residual gauge transformation allowed:¹³⁴⁾ The parameters of this residual gauge transformation are given by

$$\begin{aligned} \epsilon'_0(t, \mathbf{x}) &\equiv -\tau(\mathbf{x}), \\ \epsilon'^{(S)}(t, \mathbf{x}) &\equiv a^2(t) \tau(\mathbf{x}) \int^t \frac{dt'}{a^2(t')}, \end{aligned} \quad (12.110)$$

where $\tau(\mathbf{x})$ is an arbitrary time-independent function. An arbitrary gauge function $\tau(\mathbf{x})$ will cause problems because different sets of perturbation functions obtained in the synchronous gauge may not be different, but just related by a gauge transformation of the form of Eq. (12.110).¹³⁵⁾

However, this gauge ambiguity can be avoided if some suitable condition can be imposed on some of the scalar perturbation functions so that the freedom of residual gauge transformation Eq. (12.110) is removed and all perturbation functions in the synchronous gauge are fixed. This is indeed the case in the Λ CDM model in which a cold dark matter presents. As shown in [221] the scalar 4-velocity perturbation function $\epsilon^{(S)}$ of the dark matter, having the gauge transformation property as given in Eq. (12.72), is time-independent and such

¹³²⁾See, Eqs. (5.3.34) and (5.3.32), [221].

¹³³⁾See Footnote 127).

¹³⁴⁾Equation (5.3.39), [221], p. 242.

¹³⁵⁾A detailed discussion of the effect under the residual gauge transformation Eq. (12.110) can be found in [221], pp 242–243.

a gauge transformation makes the synchronous gauge unique.¹³⁶⁾

12.7.3 Remarks on other gauges and gauge conversions, and gauge solutions

12.7.3.1 Other gauges and gauge conversions

There are other gauges proposed in the literature. Different gauges can be used conveniently for different physical situation. Different gauges can be converted into one another by transformations. In the two examples given above, the gauge conditions are imposed on metric perturbations. Gauge conditions can also be defined on quantities entering in the energy-stress tensor, $\delta\rho$, $\delta\mathcal{P}$, or δu_μ .

Let us enumerate some of the other gauges appeared in the literature: The *spatially flat gauge* sets $F = 0$ and $A = 0$. There exists a non-trivial residual gauge transformation which requires additional condition to eliminate this additional gauge freedom.¹³⁷⁾ The *uniform density gauge* sets $\delta\rho = 0$, which determines ϵ_0 . The other gauge condition fixes a suitable metric perturbation term that requires either $F = 0$ or $B = 0$. The *comoving gauge* is defined by $\delta u^{(S)} = 0$ plus $F = 0$ or $B = 0$. More gauge conditions can be found in the literature, e.g., [224], [293], and for earlier works [294].

¹³⁶⁾The detailed argument can be found in [221], p. 243. For the pedagogic reason we give the argument for it here. The cold dark matter is non-relativistic, its pressure density vanishes, $\bar{\mathcal{P}}_{\text{DM}} = 0$. We can set the perturbation of its pressure to zero too. The vanishing of the dark matter pressure is maintained by the gauge transformation because of Eq. (12.68). We can also set its dissipative part to zero $\pi_{\text{DM}}^{(S)} = 0$. Then the first equation of the energy-momentum conservation, Eq. (12.92), which holds for the dark matter separately, becomes:

$$\partial_j \left(\bar{\rho}_{\text{DM}} \partial_t \delta u_{\text{DM}}^{(S)} + \delta u_{\text{DM}}^{(S)} (\partial_t + 3H) \bar{\rho}_{\text{DM}} \right) = 0.$$

Since $\bar{\rho}_{\text{DM}} \sim 1/a^3$, then $(\partial_t + 3H) \bar{\rho}_{\text{DM}} = 0$, the above expression is reduced to $\partial_j \partial_t u^{(S)} = \text{constant}$. The most general solution for this equation is that $u_{\text{DM}}^{(S)}$ is the sum of two functions, one is a function of the cosmic time only and the other a function of the comoving coordinates only. The first function gives rise to a homogeneous and isotropic velocity potential, which will be ignored because it leads to a zero mode solution. This is equivalent to drop the overall spatial derivative ∂_j in Eq. (12.92). The second function leads to a dark matter velocity potential as a function of the comoving coordinate only, called, say, $\delta u^{(S)}(\mathbf{x})$. Then a gauge transformation with the gauge parameter $\epsilon^0(\mathbf{x}) = \delta u^{(S)}(\mathbf{x})$ can be made to obtain a vanishing dark matter velocity potential $\delta u_{\text{DM}}^{(S)} = 0$, which is required physically. This then fixes the gauge completely.

¹³⁷⁾The residual gauge transformation is

$$\epsilon^{(S)} = a^2(t) \tau(\mathbf{x}),$$

where $\tau(\mathbf{x})$ is an arbitrary function of the comoving spatial coordinates only.

12.7.3.2 Gauge conversions

Although different gauges are preferred in different situations, e.g., for the reason of simplification of the calculations involved, all gauges are equivalent and one gauge can be converted into another by suitable gauge transformations. A discussion on and some examples of the conversion among different gauges can be found in [221].¹³⁸⁾

12.7.3.3 Gauge solutions

We make two comments concerning some of the general properties of gauge functions. We will use them to derive a set of solution in the next section 12.8 in the discussion of the early cosmic evolution.

- **Gauge solutions.** Since the governing equations, i.e., the Einstein field equations and the energy-momentum conservation equations, are linear and homogeneous in the perturbation functions, and are gauge invariant, as demonstrated explicitly in Eqs. (12.88) and (12.91), the gauge transformation functions, $\Delta\rho$, etc., and ΔE , etc. given in Eqs. (12.66), (12.69), (12.72), (12.77), (12.78), (12.79), and (12.82), are solutions of the Einstein Field and energy-momentum conservation solutions. This is also the reason why we can obtained solutions to these governing equations by fixing a gauge.
- **A noted point.** From Eqs. (12.78) and (12.79), it seems that arbitrary gauge functions of ϵ_0 , $\epsilon^{(S)}$, and $\epsilon_j^{(V)}$ would contribute to the perturbation function ΔB , ΔF , ΔG_j , and ΔG_j . However, these perturbation functions enter the Einstein field and energy-momentum conservation equations through their spatially derivatives, some simple functional dependence in the comoving coordinates will not contribute. As seen in Eqs. (12.88) and (12.91), F and B_j enter their second derivatives, so constant and linear in x_j in the gauge functions will have no effect. Functional restrictions on $\epsilon_j^{(V)}$ can be seen from the equations for the vector perturbation functions which we refer to [221]¹³⁹⁾

12.8 Very early cosmic perturbation – Outside the Horizon

12.8.1 Issues involved – A preliminary discussion

To begin we describe some of the salient points of the cosmological perturbation.

- The cosmological perturbations occurs at the very early time of the universe, seeded by quantum fluctuations taking place during the inflation epoch.

¹³⁸⁾See pp 243–255, [221].

¹³⁹⁾[221], Eqs. (5.1.50) and (5.1.51).

However, observations of the effects of the perturbations, such as the CMB anisotropy, is made at the present time, on what have been imprinted on the last scattering surface. The last scattering surface (LSS), where the cosmic photons decouple from the baryonic plasma, begins the epoch when the cosmic photon can propagate freely, carrying the LSS information with them unaltered to the present to be observed. Hence in order to study the effect of the cosmic perturbation, we have to know how the early perturbations have evolved as the cosmos grows.

- From the time of inflation to the much later time of observation, the universe has undergone stages of transformations. They include the reheating of the universe after the inflation, the freeze out of the dark matter, and perhaps other drastic stages of evolution of the universe. During reheating, the vacuum energy is presumably transferred to regular particles, such as quarks, gauge boson, etc. Aside from this very general picture, we know very little of the dynamics of the transformation process. on dark matter, for instance, the situation is highly uncertain. We do not know much about properties of the dark matter, what kind of particles they are, and what kind of dynamics they obey. Hence we need to have a mechanism in our calculation to ensure that the information of the very early cosmos is preserved down to an epoch when the properties of the universe becomes known, And from there on, the evolution of the universe down to LSS can be calculated with certainty.
- Perturbations to the metric and stress-energy tensor have well-defined time dependence, governed by the Einstein field equation. In contrast, their spatial dependence, which has a random distribution, is not precisely determined but defined stochastically.

The random spatial dependence entails the types of observations that can be made. In the study of the temperature deviation, for instance, one can not predict the temperature at a given point of the sky. But one can expect to measure the average temperature of the sky and the temperature anisotropy correlation for any two directions, as described respectively in Eqs. (12.5) and (12.9). Another example, say, on large structures of galaxies, one cannot ask to calculate the distance of a nearest large galaxy from the Milky Way. But one can ask to estimate the average separation between two large galaxies.¹⁴⁰⁾ These examples of what can and cannot be carried out meaningfully in cosmological studies suggest that cosmological perturbations have to be treated as stochastic variables in their spatial dependence.

- Equations governing the perturbations are linear homogeneous coupled equation involving both time and spatial derivatives. In the scalar case, there are six unknown gauge invariant scalar perturbation functions and six equations, and by fixing the gauge to determine the eight scalar perturbation functions, in the presence of both time and spatial derivatives in the six equations. This is a rather complicate situation. Fortunately, we can simplify the system by making the spatial coordinate Fourier transformation of the perturbation functions in terms of the exponential transformation

$$\exp(i\mathbf{q}\cdot\mathbf{x}) = \exp\left(i\left(\frac{\mathbf{q}}{a}\right)\cdot(a\mathbf{x})\right), \quad (12.111)$$

where $a = a(t)$ is the Hubble scale factor and $q/a \equiv |\mathbf{q}|/a$ is the physical wave-number. Since we keep only terms which are linear in the perturbation functions, the different *Fourier modes*, which are also known as the *normal modes*, with fixed q/a at a given cosmic time t are independent. Hence the Einstein field and energy-momentum conservation equations governing the perturbation functions can be expressed in terms of their Fourier modes. Furthermore, a spatial derivative ∂_j becomes a multiplication factor iq_j . So the equations of the Fourier modes can contain only time derivatives.

- The exact solution of the evolution of the cosmic perturbation functions has to be carried out in the kinetic theory using the Boltzmann equation,¹⁴¹⁾ resulting in a system of equations too complicate to be possibly solved analytically. Hence, for the purpose of making detailed comparison between theory and observations, numerical simulations are necessary. Modern computing power is no longer a challenge for such tasks. But numerical solutions are not transparent to the relevant physics involved. Working in the hydrodynamic limit [221] gives a set of analytic solutions which fit well with the observation of CMB. We follow this approach closely in the rest of this Section.

12.8.2 *Fourier decomposition and the stochastic property*

The perturbation functions are stochastic variables, characterized by the averages of their products, i.e., their n -point correlation functions at equal time, $\langle A(t, \mathbf{x})B(t, \mathbf{y})\cdots \rangle$. To solve the Einstein field and energy-momentum conservation equations we need a procedure to factor out the stochasticity dependence of the perturbation functions and disentangle some of the complications as discussed above. In carrying out this process, we can express the equations in terms of Fourier

¹⁴⁰⁾See [224], p. 85.

¹⁴¹⁾For a discussion of the kinetic theory we refer to [221], §6.1, pp 258–274.

amplitudes of the perturbation functions for a give mode q . Let us consider a typical perturbation function $X(t, \mathbf{x})$ and make a Fourier decomposition in the expanding background universe,

$$X(t, \mathbf{x}) = \sum_n \int d^3q e^{i(\mathbf{q}/a) \cdot (a\mathbf{x})} \alpha_n(\mathbf{q}) X_{nq}(t), \quad (12.112)$$

where \mathbf{x} is the comoving coordinates and \mathbf{q} the comoving wave vector. n enumerates the different mode of independent solutions and $\alpha_n(\mathbf{q})$ the corresponding stochastic variable. We need to know the statistic property of the stochastic variable which we will discuss below. The Fourier component $X_{nq}(t)$,¹⁴²⁾ labeled for the n -th solution of mode q , is an ordinary function of the cosmic time t and the wave-number $q \equiv |\mathbf{q}|$. Since the Fourier component is related to a physical quantity, it does not dependent on the direction of the wave vector, which is associated with the statistic variable.

The physical wave number is \mathbf{q}/a , from which we can define the physical wavelength of the given mode q ,

$$\lambda_q \equiv \frac{2\pi a}{q}, \quad (12.113)$$

which will be useful in devising a conservation quantity under the evolution of the universe. As expected, the wavelength, which is proportional to the Hubble expansion parameter, is stretched in time as the universe expands. During the inflationary epoch all wavelengths are stretched exponentially. In the radiation-dominated period, the stretched factor is proportional to $t^{1/2}$, while in the matter-dominated period $t^{2/3}$. See Table 9.2 for an explanation.

As indicated in Eq. (12.112), stochastic properties of the perturbation functions are contained in the scale factor of stochastic variables $\alpha_n(\mathbf{q})$, which also embodies the directional information of the wave vector. For a given mode n , $\alpha_n(\mathbf{q})$ is common to all eight scalar perturbation functions $E, F, A, B, \delta\rho, \delta\mathcal{P}, u^{(S)}$, and $\pi^{(S)}$, which have contributions from the various energy and matter components of the universe. After the choice of a gauge, there are six independent scalar perturbation functions which satisfy six linearly coupled equations. So there are six independent sets of solutions, hence $n = 1, 2, \dots, 6$. If some relationships exist among the perturbation functions, the number of independent coupled equations and hence the highest value of n will be reduced. The differential equations Eqs. (12.88) and (12.91), or Eqs. (12.100) and (12.101) in the Newtonian gauge, or Eqs. (12.106)–(12.108) in the synchronous gauge, can be rewritten in terms of their Fourier amplitudes of a given mode. As commented at the end of §12.6, the differential equations are real in both the Fourier space and the comoving coordinate space.

¹⁴²⁾The notation follows that of [221].

The stochastic variable $\alpha_n(\mathbf{q})$ can be chosen in the “*orthonormal*” basis,¹⁴³⁾ so that the two-point average of $\alpha_n(\mathbf{q})$ is orthogonal in the mode number as well as in the wave number,

$$\langle \alpha_{n_1}(\mathbf{q}_1) \alpha_{n_2}^*(\mathbf{q}_2) \rangle = \delta_{n_1 n_2} \delta^3(\mathbf{q}_1 - \mathbf{q}_2). \quad (12.114)$$

This allows us to project out a particular normal mode of a scalar perturbation function, i.e.,

$$\langle \partial_{j_1} \partial_{j_2} \dots X(t, \mathbf{x}) \alpha_n^*(\mathbf{q}) \rangle = (iq_{j_1})(iq_{j_2}) \dots e^{i\mathbf{q} \cdot \mathbf{x}} X_{nq}(t). \quad (12.115)$$

The two-point correlation function at equal time is straightforward to obtain,

$$\begin{aligned} & \langle X_1(t, \mathbf{x}) X_2(t, \mathbf{y}) \rangle \\ &= \sum_{n_1, n_2} \int d^3q_1 d^3q_2 e^{i\mathbf{q}_1 \cdot \mathbf{x}} e^{-i\mathbf{q}_2 \cdot \mathbf{y}} \\ & \quad \langle \alpha_{n_1}(\mathbf{q}_1) \alpha_{n_2}^*(\mathbf{q}_2) \rangle X_{1n_1q_1}(t) X_{2n_2q_2}^*(t) \\ &= \sum_n \int d^3q e^{i\mathbf{q} \cdot (\mathbf{x} - \mathbf{y})} X_{1nq}(t) X_{2nq}^*(t), \end{aligned} \quad (12.116)$$

which shows that the correlation function depends on the relative distance of the two comoving coordinate points under consideration, $\mathbf{x} - \mathbf{y}$, not their actual positions. This reflects the stochastic nature of the original distribution functions $X(t, \mathbf{x})$ and $Y(t, \mathbf{y})$. Let us remark that the separation of the perturbation functions into independent scalar, vector, and tensor parts requires that averages of the cross products of scalar-vector, scalar-tensor, and vector-tensor stochastic variables vanish. So the scalar, vector, and tensor perturbation functions have different stochastic variables.

Now we can straightforwardly write down the equations governing the scalar perturbation functions in their Fourier amplitudes, for a given mode n and wave-number q . The equations in the *synchronous gauge*, Eqs. (12.106) and (12.108), which will be used later, are

$$\begin{aligned} & -4\pi G_N (\delta\rho_q - \delta\mathcal{P}_q + q^2 \pi_q^{(S)}) \\ &= -\frac{q^2}{2a^2} (A_q - a^2 H \dot{B}_q) - \frac{1}{2} (\partial_t + 6H) \dot{A}_q, \\ & -16\pi G_N \pi_q^{(S)} = \frac{1}{a^2} A_q - (\partial_t + 3H) \dot{B}_q, \\ & 8\pi G_N (\bar{\rho} + \bar{\mathcal{P}}) \delta u_q^{(S)} = \dot{A}_q, \\ & -4\pi (\delta\rho_q + 3\delta\mathcal{P}_q - q^2 \pi_q^{(S)}) = \frac{1}{a^2} \partial_t (a^2 \psi_q), \end{aligned} \quad (12.117)$$

where

$$\psi_q \equiv \frac{1}{2} (3\dot{A}_q - q^2 \dot{B}_q). \quad (12.118)$$

¹⁴³⁾We refer to [221], pp 229–231 for a proof that an orthonormal basis can always be constructed.

for the Einstein field equations, and

$$\begin{aligned} &\delta\dot{\rho}_q + 3H(\delta\rho_q + \delta\mathcal{P}_q) - q^2 \left[\frac{1}{a^2}(\bar{\rho} + \bar{\mathcal{P}})\delta u_q^{(S)} + H\pi_q^{(S)} \right] \\ &+ 3(\bar{\rho} + \bar{\mathcal{P}})\psi_q = 0, \\ &\delta\mathcal{P}_q - q^2\pi_q^{(S)} + (\partial_t + 3H) \left[(\bar{\rho} + \bar{\mathcal{P}})\delta u_q^{(S)} \right] = 0, \end{aligned} \quad (12.119)$$

for the energy and momentum conservation equations. To simplify the notation we have dropped the notation for the solution mode number n and suppressed the time dependence t as well. So A_q , etc., are $A_{nq}(t)$, etc. And

$$A(t, \mathbf{x}) \equiv \sum_n \int d^3q e^{i(\mathbf{a}\mathbf{q}) \cdot (\mathbf{a}\mathbf{x})} \alpha_n(\mathbf{q}) A_{nq}, \quad (12.120)$$

etc. This notation will be used below.

12.8.3 Conservation outside the horizon

The study of perturbation functions in the wave number space is particularly useful because of the relation between the expansion of the universe and the physical wavelength. As already stated, the discussion below follows closely that given in [221] to which we refer for further consultation. During inflation the metric scale factor a grows exponentially, and the physical wavelength defined in Eq. (12.113) becomes exponentially long for a given comoving wave number q . However, the Hubble expansion rate $H = \dot{a}/a$ stays the same. Hence, during inflation and for a sufficient long period of time afterwards, most of the physical wavelengths $2\pi/(q/a)$ are longer than the Hubble length H^{-1} ,¹⁴⁴⁾ so $2\pi/(q/a) \gg H^{-1}$, then

$$\frac{q}{aH} \ll 1. \quad (12.121)$$

except for a small range of very short wavelengths. The Contribution from the very short wavelengths region, which corresponds to very large q values, will be very small, due to rapid oscillations of the exponential factor in the Fourier transformation. Fourier modes satisfying Eq. (12.121) are said to be *outside the horizon*.¹⁴⁵⁾ Contributions of the Fourier modes inside the horizon are negligible during inflation and a certain period afterwards.

In order for what has happened in inflation to maintain its effect to the much later stage of the universe, there has to be a conservation law to preserve it in the process of cosmic evolution. Such a conservation law, which does exit as we shall see below, can be stated as follows:

¹⁴⁴⁾In the natural units. See the discussion of cosmologic lengths given in Section 13.

¹⁴⁵⁾This is not the *particle horizon* or *event horizon* discussed in Section 13. However, notice that $H(t)^{-1}$ is the Hubble length at the cosmic time t . The particle horizon of the present epoch is $H_0^{-1} = H(t_0)^{-1}$ multiplied by a numerical factor of the order of unity, where t_0 is the present time of the universe.

Conservation outside the horizon: *Independent of the constituents of the universe, there exist adiabatic solutions to the Einstein field equation in which certain combinations of perturbation functions are time independent outside the horizon.*

The conserved quantities referred in the above theorem are gauge invariant combinations of perturbation functions. Let us consider the following gauge invariant quantities, for convenience in the Newtonian gauge

$$\begin{aligned} \mathcal{R} &\equiv \frac{1}{2}\hat{A} = -\Psi + H\delta u^{(S)}, \\ \zeta &\equiv \frac{1}{2}\hat{A} + \frac{\delta\hat{\rho}}{3(\bar{\rho} + \bar{\mathcal{P}})} = -\Psi + \frac{\delta\rho}{3(\bar{\rho} + \bar{\mathcal{P}})} \\ &= \mathcal{R} + \frac{1}{12\pi^2 G_N (\bar{\rho} + \bar{\mathcal{P}}) a^2} \nabla^2 \Psi, \end{aligned} \quad (12.122)$$

where the gauge invariant combinations \hat{A} and $\delta\hat{\rho}$ are given in Eq. (12.84), and the scalar perturbation function $A = -2\Psi$ enters in the 3-tensor part of the metric perturbation h_{jk} in the form $a^2 A \delta_{jk}$ as defined in Eq. (12.40). In obtaining the final result of the second expression in Eq. (12.122) we have used Eq. (12.102) by dropping the overall spatial derivative ∂_j which is just to ignore a zero mode term in the Fourier space. This can be seen by writing Eq. (12.102) in its Fourier component,

$$q_j \left(\delta\bar{\rho}_q + \frac{q^2}{4\pi G_N a^2} \Psi_q \right) = 0. \quad (12.123)$$

Dropping the multiplicative factor q_j means the omission of a term proportional to the delta function of the wave vector $\delta(\mathbf{q})$, i.e., a zero mode term.

In the wave-number space, Eq. (12.122) is

$$\begin{aligned} \mathcal{R}_q &= -\Psi_q + H\delta u_q^{(S)}, \\ \zeta_q &= \mathcal{R}_q - \frac{q^2}{12\pi G_N (\bar{\rho} + \bar{\mathcal{P}}) a^2} \Psi_q. \end{aligned} \quad (12.124)$$

Equation (12.124) shows that outside the horizon \mathcal{R}_q and ζ_q are practically identical,

$$(\mathcal{R}_q - \zeta_q) |_{\text{outside of horizon}} = 0, \quad (12.125)$$

because the coefficient of Ψ_q in the second expression is very small outside the horizon

$$\frac{q^2}{12\pi G_N (\bar{\rho} + \bar{\mathcal{P}}) a^2} = \xi \left(\frac{q}{aH} \right)^2 \ll 1, \quad (12.126)$$

where ξ is the order of unity.¹⁴⁶⁾ Being gauge invariant, this Newtonian relation between \mathcal{R}_q and ζ_q also holds in

¹⁴⁶⁾As can be seen from Table 9.1 which lists the equation of state of possible components of the very early universe. Since the early universe is dominated by radiation, not by dark matter or dark energy, $\bar{\mathcal{P}} = \bar{\rho}/3$, then $12\pi G_N (\bar{\rho} + \bar{\mathcal{P}}) = 6H^2$.

other gauges. R_q is conserved outside the horizon in the inflation theory driven by a single scalar field, so ζ_q is also conserved in this situation.

Next, we shall prove that ζ is conserved outside the horizon for a particular set of perturbations, i.e., the adiabatic perturbation which we will define when we get to it. Let us take the time derivative of ζ , using the last term in the middle expression of Eq. (12.122), we have in the wave number space

$$\begin{aligned} \dot{\zeta} &= -\dot{\Psi} + \frac{\delta\hat{\rho}_q}{3(\bar{\rho} + \bar{\mathcal{P}})} - \frac{\dot{\bar{\rho}} + \dot{\bar{\mathcal{P}}}}{3(\bar{\rho} + \bar{\mathcal{P}})^2} \delta\rho_q \\ &= -\dot{\Psi} + \frac{1}{3(\bar{\rho} + \bar{\mathcal{P}})} \left[-3H(\delta\rho_q + \delta\mathcal{P}_q) \right. \\ &\quad \left. + \frac{q^2}{a^2}(\bar{\rho} + \bar{\mathcal{P}})\delta u_q^{(S)} + q^2 H\pi_q^{(S)} + 3(\bar{\rho} + \bar{\mathcal{P}})\dot{\Psi} \right] \\ &\quad - \frac{\dot{\bar{\rho}} + \dot{\bar{\mathcal{P}}}}{3(\bar{\rho} + \bar{\mathcal{P}})^2} \delta\hat{\rho}_q, \end{aligned} \tag{12.127}$$

where we have used the first equation in Eq. (12.119) to rewrite $\delta\rho_q$. We can further simplify the expression by the relationship given in the second and third expressions of Eq. (12.100),

$$\begin{aligned} \dot{\zeta} &= \frac{1}{3(\bar{\rho} + \bar{\mathcal{P}})^2} \left(\dot{\bar{\rho}}\delta\mathcal{P}_q - \dot{\bar{\mathcal{P}}}\delta\rho_q \right) \\ &\quad + \frac{q^2}{12\pi G_N(\bar{\rho} + \bar{\mathcal{P}})a^2} \left(-\frac{3}{2}H\Phi_q - \dot{\Psi}_q + \frac{1}{2}H\Psi_q \right). \end{aligned} \tag{12.128}$$

Outside the horizon the second term on the right-handed side is very small, then we have

$$\dot{\zeta} \Big|_{\text{outside the horizon}} = \frac{1}{3(\bar{\rho} + \bar{\mathcal{P}})^2} \left(\dot{\bar{\rho}}\delta\mathcal{P}_q - \dot{\bar{\mathcal{P}}}\delta\rho_q \right). \tag{12.129}$$

In the case of *adiabatic perturbations* which means¹⁴⁷⁾

$$\frac{\delta\rho_{\alpha q}}{\dot{\bar{\rho}}_\alpha} = \dots = \frac{\delta\mathcal{P}_{\beta q}}{\dot{\bar{\mathcal{P}}}_\beta} = \dots, \tag{12.130}$$

where α and β denote any matter or energy components of the universe, ζ is conserved, irrespective of the types of energy and matter components of the universe.

$$\dot{\zeta} \Big|_{\substack{\text{outside the horizon} \\ \text{adiabatic perturbation}}} = 0. \tag{12.131}$$

Similarly, \mathcal{R} is also conserved under the same conditions.

¹⁴⁷⁾We have encountered adiabatic perturbations in Section 12.3.1 which concerns with density perturbation of different particle species of the universe.

For the later usage we rewrite \mathcal{R} and ζ in the *synchronous gauge* below by using Eq. (12.109) to replace A. For this purpose we consider $\nabla^2\mathcal{R}$ and $\nabla^2\zeta$ instead:

$$\begin{aligned} \nabla^2\mathcal{R} \Big|_{\text{SG}} &= -4\pi G_N a^2 \delta\rho + a^2 H\psi + H\nabla^2\delta u^{(S)}, \\ \nabla^2\zeta \Big|_{\text{SG}} &= -4\pi G_N a^2 \delta\rho + a^2 H\psi + \frac{\nabla^2\delta\rho}{3(\bar{\rho} + \bar{\mathcal{P}})} \\ &= -4\pi G_N a^2 \delta\rho + a^2 H\psi - \frac{H\nabla^2\delta\rho}{\dot{\bar{\rho}}}, \end{aligned} \tag{12.132}$$

where the subscript SG means *in the synchronous gauge* and $\psi \equiv (3\dot{A} + \nabla^2\dot{B})/2 \equiv -3\dot{\Psi} + \nabla^2\dot{B}/2$ is defined in Eq. (12.107). In the last line Eq. (9.45) has been used.

12.8.4 Adiabatic solutions

In this subsection we will demonstrate the existence of adiabatic solutions. We first observe that gauge transformations generate terms in density functions $\Delta\rho$ and $\Delta\mathcal{P}$ that are adiabatic like, as shown in Eq. (12.70). Together with terms generated in other perturbation functions given in Eqs. (12.69), (12.72), (12.77)–(12.79), they satisfy the Einstein field and energy-momentum conservation equations. But they can be transformed away. Hence they are gauge terms, not genuine solutions. This gauge trap can be avoided once we fix a gauge to work in. We choose the Newtonian gauge. As we have shown earlier, in the Newtonian, once we set $B = F = 0$, there does not exist additional gauge transformations that allow the system to stay in the Newtonian gauge. The solutions obtained are unique.

The trick to find an adiabatic solution is to begin in a homogeneous universe in which all the variables depend only on the cosmic time. We note two helpful points about the homogeneous universe in the Newtonian gauge relevant to our consideration.¹⁴⁸⁾

- The first point is that in a homogeneous universe the scalar perturbation functions B and F do not contribute in general. This is due the fact, as can be seen from Eqs. (12.88) and (12.91), that they appear in the equations together in the term \hat{B}_F and under two spatial derivatives ∇^2 or $\partial_j\partial_k$. Hence if we have a non-trivial coordinate transformation which maintains the cosmic homogeneity, the perturbation functions so generates is not a gauge mode and is adiabatic owing to Eq. (12.70).
- The second point is that the solution of the homogeneous universe, in which all spatial derivatives vanish, can be considered as a zero mode solution, $q = 0$, for the general, inhomogeneous universe. Hence a generalization of the zero mode solution to the case $q \neq 0$, if it exist, gives us the perturbation function that we are looking for.

¹⁴⁸⁾As noted earlier, the discussion follows that given in [221].

12.8.4.1 Adiabatic solution in the homogeneous universe $\epsilon_j(t, \mathbf{x})$

In the homogeneous universe a special gauge transformation is allowed even in the Newtonian gauge as we can see below. Let us make a change of the coordinate variables¹⁴⁹⁾

$$x_\mu \rightarrow x_\mu + \epsilon_\mu(t, \mathbf{x}),$$

$$\epsilon_\mu \equiv (\epsilon_0, \epsilon_j) = (\epsilon_0, \partial_j \epsilon^{(S)} + \epsilon_j^{(V)}), \quad \partial_j \epsilon_j^{(V)} = 0, \quad (12.133)$$

Note that the coordinate transformation is allowed to depend on the comoving coordinates, which is not spatially homogeneous. From Eqs. (12.76) and (12.77), the spatial homogeneity for $\Delta^{(H)}h_{00} = -2\dot{\epsilon}_0$, where the subscript (H) denotes the homogeneous space, says

$$\epsilon_0 = \epsilon(t) + X(\mathbf{x}), \quad (12.134)$$

where $\epsilon(t)$ and $X(\mathbf{x})$ are arbitrary small functions of t and \mathbf{x} respectively. The vanishing of the time-space component of the metric tensor generated by the gauge transformation

$$\Delta^{(H)}h_{0j} = -\partial_j \epsilon_0 - a^2 \partial_t (\epsilon_j / a^2) = 0 \quad (12.135)$$

gives a differential equation which relates ϵ_0 and ϵ_j through the arbitrary function $X(x)$ unless it vanishes,

$$\partial_t \left(\frac{\epsilon_j}{a^2} \right) = -\frac{1}{a^2} \partial_j X, \quad (12.136)$$

which has the solution

$$\epsilon_j(t, \mathbf{x}) = a^2(t) f_j(\mathbf{x}) - a^2(t) (\partial_j X(\mathbf{x})) \int_\tau^t \frac{dt'}{a^2(t')}, \quad (12.137)$$

where $f_j(\mathbf{x})$ is an arbitrary vector function of the comoving coordinates. The lower limit of the integration τ is an arbitrary cosmic time.

From Eq. (12.76), we also have

$$\begin{aligned} \Delta^{(H)}h_{jk} &= 2a^2 H \epsilon_0 \delta_{jk} - (\partial_j \epsilon_k + \partial_k \epsilon_j) \\ &= 2a^2 H (\epsilon(t) + X(\mathbf{x})) \delta_{jk} \\ &\quad - a^2 (\partial_j f_k(\mathbf{x}) + \partial_k f_j(\mathbf{x})) \\ &\quad + 2a^2 \partial_j \partial_k X(\mathbf{x}) \int_\tau^t \frac{dt'}{a^2(t')}. \end{aligned} \quad (12.138)$$

In order for $\Delta^{(H)}h_{jk}$ to be independent of the comoving coordinates \mathbf{x} , we have to have $X(\mathbf{x}) = 0$ and $f_j(\mathbf{x})$ to be linear in \mathbf{x} . So, let us write $f_j(\mathbf{x}) \equiv \omega_{jk} x_k$. We can now obtain ϵ_j and Δh_{jk} . We make a decomposition of

$$\begin{aligned} \epsilon_j(t, \mathbf{x}) &= a^2 f_j(\mathbf{x}) \equiv a^2 \omega_{jk} x_k \\ &= \frac{a^2}{3} \omega_{\ell\ell} x_j + \frac{a^2}{2} \left(\omega_{jk} + \omega_{kj} - \frac{2}{3} \omega_{\ell\ell} \delta_{jk} \right) x_k \\ &\quad + \frac{a^2}{2} (\omega_{jk} - \omega_{kj}) x_k \\ &= \frac{a^2}{6} \omega_{\ell\ell} \partial_j x^2 + \frac{a^2}{2} \partial'_j \left(x'_j \omega_{j'k} x_k - \frac{1}{3} \omega_{\ell\ell} x^2 \right) \\ &\quad + \frac{a^2}{2} (\omega_{jk} - \omega_{kj}) x_k, \end{aligned} \quad (12.139)$$

where (ω_{jk}) is a 3×3 numerical matrix and $\omega_{\ell\ell}$ its trace. We now have explicitly

$$\begin{aligned} \Delta^{(H)}h_{jk} &= 2a^2 \left[H\epsilon(t) - \frac{1}{3} \omega_{\ell\ell} \right] \delta_{jk} \\ &\quad - a^2 \left(\omega_{jk} + \omega_{kj} - \frac{2}{3} \omega_{\ell\ell} \delta_{jk} \right), \end{aligned} \quad (12.140)$$

which, as promised, is a function of the cosmic time only. The last term appearing in the second and third line of Eq (12.139), which is proportional to the antisymmetric quantity $\omega_{jk} - \omega_{kj}$, will make no contributions in the following discussions.

The results given in Eqs. (12.139) and (12.140) can be interpreted as follows. We first compare (12.139) with the standard form of ϵ_j listed in Eq. (12.66). The first term of the third line of Eq. (12.139) can be identified as $\partial_j \epsilon^{(S)}$ and the third term as $\epsilon^{(V)}$. The middle term of the right-handed side of Eq. (12.139) is ambiguous. It is both curlless and divergenceless, and therefore can be included in either $\partial_j \epsilon^{(S)}$ or $\epsilon_j^{(V)}$. The reason for this puzzling fact is that the Helmholtz decomposition of the vector field Eq. (12.37), as stated in Fn. 120), does not apply to $\epsilon_j(t, \mathbf{x})$ in Eq. (12.139) because of its asymptotic behavior in x_j . As of $\Delta^{(H)}h_{jk}$ the resultant given in Eq. (12.140) suggests that the first term of the right-handed side belongs to ΔA of Eq. (12.79), and the second term to ΔD_{jk} . It is easy to see that it satisfies the conditions imposed on D_{jk} listed in Eq. (12.40). The second term is new, not present in the general case of coordinate transformation considered for gauge transformations, see Eq. (12.79).

We can now identify all metric, density, and fluid terms generated by the coordinate transformation Eq. (12.133) and (12.134), expressible in terms of a scalar function of the cosmic time $\epsilon(t)$ and a constant 3×3 matrix ω_{jk} ,

$$\begin{aligned} \Delta^{(H)}E &= 2\dot{\epsilon}(t), \quad \Delta^{(H)}A = 2 \left[H\epsilon(t) - \frac{1}{3} \omega_{\ell\ell} \right], \\ \Delta^{(H)}D_{jk} &= -\omega_{jk} - \omega_{kj} + \frac{2}{3} \omega_{\ell\ell} \delta_{jk}, \\ \Delta^{(H)}\rho &= \dot{\rho}\epsilon(t), \quad \Delta^{(H)}\mathcal{P} = \dot{\mathcal{P}}\epsilon(t), \\ \Delta^{(H)}u_0 &= -\dot{\epsilon}(t), \quad \Delta^{(H)}u^{(S)} = -\epsilon(t). \end{aligned} \quad (12.141)$$

¹⁴⁹⁾We reproduce the discussion given in [221], p. 248 and p. 249.

All other changes vanish.

As remarked early, the coordinate transformation functions listed in Eq. (12.141) provide a set of solutions to the Einstein field and energy-momentum conservation equations in the homogenous universe. Following [221], we take the solutions as the negative of the transformation functions, i.e.,

$$\begin{aligned} \Phi(t) &\equiv -\frac{1}{2}\Delta^{(H)}E = -\dot{\epsilon}(t), \\ \Psi(t) &\equiv -\frac{1}{2}(-\Delta^{(H)}A) = H\epsilon(t) - \frac{1}{3}\omega_{\ell\ell}, \\ D_{jk} &\equiv -\Delta^{(H)}D_{jk} = \omega_{jk} + \omega_{kj} - \frac{2}{3}\omega_{\ell\ell}\delta_{jk}, \\ \delta\rho &\equiv -\Delta^{(H)}\rho = -\dot{\rho}\epsilon(t), \\ \delta\mathcal{P} &= -\Delta^{(H)}\mathcal{P} = -\dot{\mathcal{P}}\epsilon(t), \\ \delta u(t) &\equiv -\Delta^{(H)}u_0 = \dot{\epsilon}(t), \\ \delta u^{(S)}(t) &= -\Delta^{(H)}u^{(S)} = \epsilon(t). \end{aligned} \tag{12.142}$$

All other perturbation functions vanish. As can be seen from the third line of Eq. (12.142) on the solutions to the energy and pressure densities, this set of solutions is adiabatic.

It is straightforward to check that the zero mode equations of the Newtonian gauge, which involve the first and fourth expressions in Eq. (12.100) and the second expression in Eq. (12.101), are satisfied. The rest of the equations vanish identically based on the fact that they involve one or two spatial derivatives. However, the two one-derivative expressions, i.e., the third expression in Eq. (12.100) and the first expression in Eq. (12.101) are satisfied even without the spatial derivative.¹⁵⁰⁾ But the second expression of Eq. (12.100) which involves two derivatives is not satisfied without imposing the zero mode condition. This has important bearing on obtaining a unique set of solution when generalized to the case of $q \neq 0$.

12.8.4.2 Generalizing the solution and the conservation law

Let us generalize the solution to the situation of $q \neq 0$ and we will focus on the scalar functions.¹⁵¹⁾ So far $\epsilon(t)$ is an arbitrary function of the cosmic time. The removal of the condition $q = 0$ causes no problem except for the second expression in Eq. (12.100). There is, however, a natural solution to the difficulty as follows, based on the explicit form of the expression: For vanishing scalar anisotropic inertial $\pi^{(S)} = 0$, the expression is satisfied in the absence of spatial derivatives, when the condition $\Phi = \Psi$ is imposed. The condition gives rise to a first

¹⁵⁰⁾To verify the equations the following identities are convenient to use: $\ddot{a}/a = \dot{H} + H^2$, $\ddot{u}/a = \ddot{H} + 3\dot{H}H + H^3$, $8\pi G_N \bar{\rho} = 3H^2$, and $8\pi G_N \bar{\mathcal{P}} = -2\dot{H} - 3H^2$.

¹⁵¹⁾For the effect of D_{jk} we refer to [221], p. 251.

order differential equation in $\epsilon(t)$ which can be obtained from the first line of Eq. (12.142):

$$\partial_t \epsilon(t) + H\epsilon(t) = \frac{\omega_{\ell\ell}}{3} \equiv R_\omega. \tag{12.143}$$

The solution of this equation is straightforward,

$$\epsilon(t) = \frac{R_\omega}{a(t)} \int_\tau^t dt' a(t'), \quad R_\omega \equiv \frac{\omega_{\ell\ell}}{3}. \tag{12.144}$$

The non-vanishing scalar perturbations are

$$\begin{aligned} \Phi(t) = \Psi(t) &= -\delta u_0(t) = -\dot{\epsilon}(t) \\ &= R_\omega \left[-1 + \frac{H}{a(t)} \int_\tau^t dt' a(t') \right], \\ \frac{\delta\rho(t)}{\dot{\rho}} &= \frac{\delta\mathcal{P}(t)}{\dot{\mathcal{P}}} = -\delta u^{(S)} = -\epsilon(t) \\ &= -\frac{R_\omega}{a(t)} \int_\tau^t dt' a(t'). \end{aligned} \tag{12.145}$$

The second expression shows that this is an adiabatic solution.

The most general solution of Eq. (12.143) will also include a part which is the solution of the homogeneous part of the differential equation for $R_\omega = 0$. The solution takes the form

$$\tilde{\epsilon}(t) = \frac{\mathcal{C}}{a(t)}, \tag{12.146}$$

which is the difference of two solutions for a non-vanishing R_ω with different lower integration limits, say τ and τ' . This $R_\omega = 0$ solutions leads to the following perturbation functions

$$\begin{aligned} \tilde{\Phi} = \tilde{\Psi} &= -\dot{\tilde{\epsilon}} = H\frac{\mathcal{C}}{a}, \\ \frac{\delta\tilde{\rho}}{\dot{\rho}} &= \frac{\delta\tilde{\mathcal{P}}}{\dot{\mathcal{P}}} = -\tilde{\epsilon} = -\frac{\mathcal{C}}{a}. \end{aligned} \tag{12.147}$$

For an expanding universe, these are “decay” solutions, which decrease in time. They have no significance in the evolution of the universe and hence will be neglected.

The perturbation functions given in Eqs. (12.145) and (12.147) satisfy the Newtonian gauge equations for wavelength outside the horizon. Now we can make the generalization of the zero mode solution: R_ω , which is originally a constant factor multiplying all perturbation functions, can be extended to be a function of q : $R_\omega \rightarrow R_{\omega q}$. In the wave-number space, we have the q -dependence perturbation functions

$$\begin{aligned} \Phi_q(t) = \Psi_q(t) &= -\delta_q u_0(t) = -\dot{\epsilon}_q(t) \\ &= R_{\omega q} \left[-1 + \frac{H}{a(t)} \int_\tau^t dt' a(t') \right], \\ \frac{\delta\rho_q(t)}{\dot{\rho}} &= \frac{\delta\mathcal{P}_q(t)}{\dot{\mathcal{P}}} = -\delta u_q^{(S)}(t) = -\epsilon_q(t) \\ &= -\frac{R_{\omega q}}{a(t)} \int_\tau^t dt' a(t'). \end{aligned} \tag{12.148}$$

We can obtain the conserved quantities discussed in Section 12.8.3. The conserved quantities \mathcal{R}_q and ζ_q defined in Eq. (12.124) are, in the present solution

$$\begin{aligned}\mathcal{R}_q &= -\Psi_q + H\delta u_q^{(S)} = R_{\omega q}, \\ \zeta_q &= -\Psi_q + \frac{\delta\rho_q}{3(\bar{\rho} + \bar{P})} = R_{\omega q}.\end{aligned}\quad (12.149)$$

The presence of the q -dependence determined by $R_{\omega q}$ leads to spatial dependence in the perturbation functions. This spatial dependence is conserved in the evolution of the universe for each individual wave-numbers as long as their wave-lengths are outside the horizon.

12.9 Evolution of scalar perturbations in the hydrodynamic limit

To compute the evolution of the cosmological fluctuation, following [221], we start from a point of the cosmic time when we are more or less certain about the composition of the universe. This initial stage is taken to be at the temperature of 10^9 K, from which the evolution of the various perturbation functions can be worked out analytically to the comparatively recent time. Then the perturbations will grow in size and the linear perturbation can no longer be valid. However, in the presently consideration we are interested in the cosmologic fluctuation in the CMB temperature, we need only to work out analytically the evolution of the perturbation down to the last scattering surface at the temperature about 3000 K when the linear perturbation is still valid. Passing the last scattering surface (LSS), the CMB photon can propagate freely to the present epoch and preserves what have been imprinted on the LSS.¹⁵²⁾

At 10^9 K, the universe is about a couple of minutes old and is made of four main components: photons, neutrinos, dark matter, and the baryonic matter which includes electrons, while the present-day dominant dark energy forms just a very small fraction of the universe and can be neglected. These major matter-radiation components have roughly the following status:

- Neutrinos have decoupled (which took place at $T \approx 10^{10}$ K) and become cosmic free streamers.
- The synthesis of light nuclei is about to completed.
- Assumed to be of the cold category, the dark matter has “long” decoupled from the universe and become non-relativistic, pressureless.
- The photon and baryon plasma are still tightly coupled by the electromagnetic interaction due to Thomson scattering. But the photon conversion into electron-positron pairs has frozen out.

¹⁵²⁾The titles of this section is adopted from [221], §6.2, p. 274. Although the “hydrodynamic limit” is not generally a familiar terminology in particle physics, it is a broad subject in classic physics and applied mathematics. For a quick view of it, please see [295].

Let us recall that the LSS is at the cosmic age of 0.38 million years, the cosmic temperature of 3000 K, and the redshift $z = 1100$. For a brief glance of the milestones of the universe, see Table 9.6. Needless to say, the various energy-matter components are never totally separated from one another because of the gravitational effect as dictated by the Einstein field equation which guides the behavior of the universe when all other interactions becomes unimportant.

Kinetic descriptions of the various cosmic components are necessary for a precision treatment of the evolution of the universe. For a discussion of the basic formulation of the kinetic theory, see [221].¹⁵³⁾ In an exact approach, numerical simulations of the Boltzmann equations involved are necessary.¹⁵⁴⁾ In order to better see the underline physics involved, however, we follow the approach of [221] which provides an analytic treatment as an approximation to the system of the four kinds of energy-matter.¹⁵⁵⁾ Even in this type of approximation, the system under consideration is already quite complicated. In some stages of the treatment, the approximations seemed to be drastic. But the resultant expression for the CMB power spectrum describes the observation well. Below we will sketch the approaches to the approximation and refer to [221] for details.¹⁵⁶⁾

12.9.1 Setting up the system of equations

The synchronous gauge proves to be convenient for the study of the cosmic evolution. We will first list the equations that govern the evolutions of the different species of energy-matter in the synchronous gauge. Let us recall that in the Einstein field equation Eq. (12.106) all energy-matter components of the universe enter into each of the four equations. So all equations will involve the dark matter, neutrinos, photons, and the baryonic matter. For the energy-momentum conservation equation Eq. (12.108), individual components holds their own equations, under the common effect of the gravity represented by ψ . In this very early stage of the universe, the dark energy is a very small component of the universe and therefore negligible in the present consideration.

The following approximations will be made:

- In the hydrodynamic limit, the anisotropic inertial will be neglected, so we set $\pi^{(S)} = 0$ for all energy-matter components.

¹⁵³⁾See [221], §6.1, pp 258–274.

¹⁵⁴⁾In addition to the Boltzmann codes, CMBFast [296] and CAMB [297], mentioned in [221] (p. 257) there are other codes developed more recently, such as CMBEASY [298] and CLASS. A series of four articles on CLASS [302] that also provide comparisons of different codes can be found in [303].

¹⁵⁵⁾The approximate solutions occupy most of Section 6 of [221].

¹⁵⁶⁾In Sections 6 and 7 of [221], both kinetic and hydrodynamic approaches are presented. Here we follow the hydrodynamic approach of [221] only.

- We will work with cold dark matter and so its fluid velocity can be set to zero $\delta u_D^{(S)} = 0$. This amounts to fix the dark matter on the comoving coordinates.
- For neutrinos, the fluid velocity and anisotropic inertial make negligible contributions at the very early time outside the horizon. And their energy and momentum densities perturbations are related by $\delta \mathcal{P}_\nu = \delta \rho_\nu/3$, as if neutrinos are in local thermal equilibrium. Furthermore, in later time when the universe becomes matter-dominated, the overall effect of the neutrino becomes increasingly small.
- Because of the significant electromagnetic interaction of the Thomson scattering, the baryon and photon form a tightly coupled plasma so they have the same fluid velocity potential: $\delta u_B^{(S)} = \delta u_\gamma^{(S)}$. The photon is in local thermal equilibrium which gives $\mathcal{P}_\gamma = \rho_\gamma/3$. The baryonic matter is non-relativistic, so its pressure density vanishes $\mathcal{P}_B = 0$.

Under the approximations as described above, we have the following quantities in our cosmological system:

- 4 energy densities perturbations which include the of photon $\delta \rho_\gamma$, neutrino $\delta \rho_\nu$, baryonic matter $\delta \rho_B$, and dark matter $\delta \rho_D$;
- 2 fluid velocity perturbations of the photon and baryonic matter δu_γ and neutrino δu_ν ; and
- 2 metric perturbations which represent the gravitational potentials A and B .

In the synchronous gauge, the perturbations in the density and velocity potential are determined by the fourth and third expressions of the Einstein field equation given in Eq. (12.106). The density perturbation depends on a particular combination of metric perturbations defined in Eq. (12.107): $\psi = (3\dot{A} + \nabla^2 \dot{B})/2$. As we will see below, the seven quantities, which include ψ , four functions of density perturbation, and two functions of fluid velocity potential perturbation, form a closed system of seven differential equations and seven unknowns. So a complete set of solutions are allowed. Since ψ is determined by the time derivative of A and the time and spatial derivatives of B , the first and second expressions in Eq. (12.106) are needed to determine A and B individually.

12.9.2 Hydrodynamic equations and initial conditions

12.9.2.1 The differential equations

Now we list the differential equations that govern the cosmic evolution. All expressions are in the wave-number space. The cosmic components are explicitly labeled.

12.9.2.1.1 The gravitational field equation

The fourth equation of Eq. (12.106):

$$\frac{1}{a^2} \partial_t (a^2 \psi_q) = -4\pi G_N (\delta \rho_{Dq} + \delta \rho_{Bq} + 2\delta \rho_{\gamma q} + 2\delta \rho_{\nu q}). \quad (12.150)$$

The photon and neutrinos are relativistic, $\delta \mathcal{P}_\gamma = (1/3)\delta \rho_\gamma$ and $\delta \mathcal{P}_\nu = (1/3)\delta \rho_\nu$ for adiabatic perturbation. The baryon matter and dark matter are non-relativistic, $\delta \mathcal{P}_B = \delta \mathcal{P}_D = 0$.

12.9.2.1.2 Energy conservation equations

Each of the four cosmic components has its energy density evolution equation. From the first equation of Eq. (12.108), we have

$$\begin{aligned} (\partial_t + 3H)\delta \rho_{Dq} &= -\bar{\rho}_D \psi_q, \\ (\partial_t + 3H)\delta \rho_{Bq} - \frac{q^2}{a^2} \bar{\rho}_B \delta u_\gamma^{(S)} &= -\bar{\rho}_B \psi_q, \\ (\partial_t + 4H)\delta \rho_{\gamma q} - \frac{4q^2}{3a^2} \bar{\rho}_\gamma \delta u_\gamma^{(S)} &= -\frac{4}{3} \bar{\rho}_\gamma \psi_q, \\ (\partial_t + 4H)\delta \rho_{\nu q} - \frac{4q^2}{3a^2} \bar{\rho}_\nu \delta u_\nu^{(S)} &= -\frac{4}{3} \bar{\rho}_\nu \psi_q, \end{aligned} \quad (12.151)$$

where scalar anisotropic inertias of all cosmic components, and the scalar velocity potential of the dark matter as well, are set to zero. The photon and baryon form a tightly coupled plasma, so $\delta u_\gamma^{(S)} = \delta u_B^{(S)}$. The velocity potential of the dark matter vanishes, $\delta u_D^{(S)} = 0$.

12.9.2.1.3 Momentum conservation equations

The momentum of the cold dark matter is neglected. The baryonic matter and photon are coupled in their momenta so they enter jointly in the momentum conservation equation. The neutrino has its separate momentum conservation equation. From the second expression of Eq. (12.108), we obtain the equations for the perturbation functions of the velocity potentials,

$$\begin{aligned} (\partial_t + 3H) \left[\left(\bar{\rho}_B + \frac{4}{3} \bar{\rho}_\gamma \right) \delta u_\gamma^{(S)} \right] &= -\frac{1}{3} \delta \rho_{\gamma q}, \\ (\partial_t + 3H) \left[\left(\frac{4}{3} \bar{\rho}_\nu \right) \delta u_\nu^{(S)} \right] &= -\frac{1}{3} \delta \rho_{\nu q}, \end{aligned} \quad (12.152)$$

where we have set $\bar{\mathcal{P}}_B = 0$, $\bar{\mathcal{P}}_\gamma = \frac{1}{3} \bar{\rho}_\gamma$, and $\bar{\mathcal{P}}_\nu = \frac{1}{3} \bar{\rho}_\nu$. The dark matter has a vanishing velocity potential $\delta u_D^{(S)} = 0$ and (12.108) for the dark matter is satisfied identically.

Following [221], we can simplify the above equations significantly by defining the *scaled density perturbations*

$$\delta \beta_q \equiv \frac{\delta \rho_{\beta q}}{\bar{\rho}_\beta + \mathcal{P}_\beta}, \quad (12.153)$$

where β denotes the energy-matter species. The background energy densities have well-defined behavior in terms of the scale factor a , i.e., $\rho_B \sim a^{-3}$, $\rho_D \sim a^{-3}$,

$\rho_\gamma \sim a^{-4}$, and $\rho_\nu \sim a^{-4}$, so $\dot{\rho}_\beta/\bar{\rho}_\beta = -3H$ for the non-relativistic dark matter and baryonic matter, and $\dot{\rho}_\beta/\bar{\rho}_\beta = -4H$ for the photon and neutrino. In particular, the second and third equations of Eq. (12.151) are

$$\begin{aligned}\partial_t \delta_{Bq} - \frac{q^2}{a^2} \delta u_\gamma^{(S)} &= -\psi_q, \\ \partial_t \delta_{\gamma q} - \frac{q^2}{a^2} \delta u_\gamma^{(S)} &= -\psi_q,\end{aligned}\quad (12.154)$$

which lead to

$$\partial_t (\delta_{Bq} - \delta_{\gamma q}) = 0. \quad (12.155)$$

Then if the relation $\delta_{Bq} = \delta_{\gamma q}$ holds at some cosmic time, it will be true at all cosmic time. Since we are interested in the adiabatic condition $\delta_{Bq} = \delta_{\gamma q}$ at early time outside the horizon, we can take

$$\delta_{Bq} = \delta_{\gamma q} \quad (12.156)$$

in the hydrodynamic equations, Eqs. (12.150)–(12.152). So there are now six independent quantities to be determined by six equations.

12.9.2.1.4 Collecting governing differential equations for adiabatic solutions

We rewrite the time evolution equations in terms of $\delta_{\beta q}$ and set $\delta_{Bq} = \delta_{\gamma q}$. The gravitational field equation Eq. (12.150) becomes¹⁵⁷⁾

$$\frac{1}{a^2} \partial_t (a^2 \psi_q) = -4\pi G_N \left[\bar{\rho}_D \delta_{Dq} + \left(\bar{\rho}_B + \frac{8}{3} \bar{\rho}_\gamma \right) \delta_{\gamma q} + \frac{8}{3} \bar{\rho}_\nu \delta_{\nu q} \right]. \quad (12.157)$$

The energy conservation equations Eq. (12.151) becomes,

$$\begin{aligned}\partial_t \delta_{Dq} &= -\psi_q, \\ \partial_t \delta_{\gamma q} - \frac{q^2}{a^2} \delta u_\gamma^{(S)} &= -\psi_q, \\ \partial_t \delta_{\nu q} - \frac{q^2}{a^2} \delta u_{\nu q}^{(S)} &= -\psi_q.\end{aligned}\quad (12.158)$$

The momentum conservation equations Eq. (12.152),

$$\begin{aligned}\partial_t \left(\frac{1+R_B}{a} \delta u_{\gamma q}^{(S)} \right) &= -\frac{1}{3a} \delta_{\gamma q}, \\ \partial_t \left(\frac{1}{a} \delta u_{\nu q}^{(S)} \right) &= -\frac{1}{3a} \delta_{\nu q},\end{aligned}\quad (12.159)$$

where¹⁵⁸⁾

$$R_B \equiv \frac{3\bar{\rho}_B}{4\bar{\rho}_R}. \quad (12.160)$$

¹⁵⁷⁾Equations below (12.157)–(12.159) are Eqs. (6.2.9)–(6.2.14), p. 276 in [221].

¹⁵⁸⁾ R_B here is R in [221], see p. 276. We use a different notation to avoid confusion, since R has been used earlier to denote something else.

These simplified equations, Eqs. (12.157), (12.158), and (12.159), together with further approximations according to the physical situations under consideration, will be used to obtain the required solutions. There are six equations in Eqs. (12.157)–(12.159), involving six unknowns: ψ_q , $\delta_{\gamma q} = \delta_{Bq}$, δ_D , $\delta_{\nu q}$, $\delta u_{\gamma q}^{(S)}$, and $\delta u_{\nu q}^{(S)}$. The parameters entering the differential equations explicitly include q and R_B , and the scale factor $a(t)$.

12.9.2.2 Initial conditions

We first determine the initial conditions that the solutions of the evolution equations have to have. The epoch of the universe from which we choose to begin our calculation is relatively simple, consisting of well-defined energy-matter components. The perturbation functions are still small and can be treated linearly. We begin at the cosmic temperature of $T = 10^9$ K, when the universe is highly radiation-dominated¹⁵⁹⁾ and the parameters that describe the background universe are straightforward:

$$\begin{aligned}\bar{\rho} \approx \bar{\rho}_R \equiv \bar{\rho}_\gamma + \bar{\rho}_\nu &\gg \bar{\rho}_M \equiv \bar{\rho}_B + \bar{\rho}_D, \quad R_B \ll 1, \\ a(t) \propto \sqrt{t}, \quad H = \frac{\dot{a}}{a} &= \frac{1}{2t},\end{aligned}\quad (12.161)$$

With these inputs and some further suitable simplifications, the six equations given in Eqs. (12.157)–(12.159) will be simplified enough so that single-term solutions of all six quantities can be obtained. They will be taken as the leading terms of their respective more complete solutions valid for an extended time region, which includes $T = 10^9$ K as its earliest edge in time. Therefore, these initial time single-term solutions will play the role of initial conditions of the system of differential equations.

Since adiabatic solutions are used to study the CMB power spectrum, we restrict ourselves to the adiabatic perturbation only. We will see below that there are two independent solutions of the adiabatic nature. But only one of them is suitable for structure formation. To proceed we further assume the following:

- For radiation dominance, we can drop the background matter densities $\bar{\rho}_M$ in comparison with that of the radiation $\bar{\rho}_R$. Then R_q can be set to zero.
- At high enough temperature we can take the adiabatic perturbation for all components of the energy matter densities.¹⁶⁰⁾

¹⁵⁹⁾The ratio of radiation density over matter density at the cosmic temperature T is T/T_{EQ} . For T_{EQ} of the order of 10^4 , the radiation energy density is 5 orders of magnitude greater than that of the matter.

¹⁶⁰⁾As stated in [221], p. 277 above Eq. (6.2.23): “... only these modes are present in inflationary theories with a single scalar field, or if the universe is ever earlier in a state of complete local thermal equilibrium with no non-conserved quantities.”

$$\begin{aligned} \delta_{Dq} &= \delta_{Bq} = \delta_{\gamma q} = \delta_{\nu q} \equiv \delta_q, \\ \delta u_{\gamma q}^{(S)} &= \delta u_{\nu q}^{(S)} \equiv \delta u_q^{(S)}. \end{aligned} \tag{12.162}$$

We note that, although $\bar{\rho}_D$ and $\bar{\rho}_B$ are much smaller than $\bar{\rho}_\gamma$ and $\bar{\rho}_\nu$ and therefore negligible, the perturbation functions δ_D and δ_B , which are fractional with small denominators as defined in Eq. (12.153), may not be small in comparison with δ_γ and δ_ν . Here we make the ansatz that they are comparable in magnitude to satisfy the adiabatic condition.

- Drop all terms proportional to q^2/a^2 for being outside the horizon and early enough in the cosmic time so that the radiation dominates the universe.¹⁶¹ We will justify this approximation after we obtain the solution.

Under these additional approximations, Eqs. (12.157), (12.158), and (12.159) are reduced to three unknowns, δ_q , $\delta u_q^{(S)}$, and ψ_q , and three equations:

$$\begin{aligned} \frac{1}{t} \frac{\partial}{\partial t} (t\psi_q(t)) &= -4\pi G_N \left(\frac{8}{3} (\bar{\rho}_\gamma + \bar{\rho}_\nu) \right) \delta_q(t) \\ &= -4H^2 \delta_q(t) = -\frac{1}{t^2} \delta_q(t), \\ \partial_t \delta_q(t) &= -\psi_q(t), \\ \frac{\partial}{\partial t} \left(\frac{\delta u_q^{(S)}(t)}{\sqrt{t}} \right) &= -\frac{1}{3\sqrt{t}} \delta_q(t), \end{aligned} \tag{12.163}$$

where Eq. (12.161) $H = 1/(2t)$ has been used. The first two equations lead to a second order homogeneous equation for δ_q :

$$\frac{\partial}{\partial t} (t\partial_t \delta_q(t)) - \frac{1}{t} \delta_q(t) = 0, \tag{12.164}$$

which has two simple solutions: $\delta_q(t) \sim t$ and $\delta_q(t) \sim t^{-1}$. However, these are not the solutions of the evolution equations, of course. They will be taken as the leading terms to which the actual solutions should be matched with, which will be made at the beginning time when the cosmos evolution is calculated. Let us discuss these leading term solutions in some detail.

12.9.2.2.1 Mode 1: The leading term solution

$$\delta_q(t) \sim t$$

For $\delta_q(t) \sim t$, from the second expression of Eq. (12.163) we have the cosmic time behavior $\psi_q(t) \sim$ independent of t and $\delta u_q^{(S)}(t) \sim t^2$. From the differential equations in Eq. (12.163) we can write down the leading term solutions

$$\psi_q(t) \equiv \hat{\psi}_q, \quad \delta_q(t) = -\hat{\psi}_q t, \quad \delta u_q^{(S)}(t) = (2\hat{\psi}_q/9)t^2. \tag{12.165}$$

Then, $\hat{\psi}_q$, which is independent of t , can be determined by making a connection with the much earlier universe, in particular, the conserved quantity \mathcal{R} and ζ in the synchronous gauge given in Eq. (12.132). In the wave-number space, using Eq. (12.165), we have

$$\begin{aligned} q^2 \mathcal{R}_q &= 4\pi G_N a^2 \delta \rho_q(t) - a^2 H \psi_q(t) + q^2 H \delta u_q^{(S)}(t) \\ &= -2a^2 H^2 t \hat{\psi}_q - a^2 H \hat{\psi}_q + \left(\frac{q^2}{a^2 H^2} \right) \frac{2}{9} a^2 H^3 t^2 \hat{\psi}_q. \end{aligned} \tag{12.166}$$

To obtain the first term on the right-handed side of the second line above, we have rewritten the corresponding term in the first line, in the radiation-dominated regime and using Eq. (12.153), as

$$\begin{aligned} 4\pi G_N \delta \rho_q(t) &= 4\pi G_N ((\bar{\rho}_\gamma + \bar{P}_\nu) \delta_{\gamma q}(t) + (\bar{\rho}_\nu + \bar{P}_\nu) \delta_{\nu q}(t)) \\ &= 4\pi G_N \left(\frac{4}{3} \right) (\bar{\rho}_\gamma + \bar{\rho}_\nu) \delta_q \\ &= 2H^2 \delta_q = -2H^2 t \hat{\psi}_q, \end{aligned} \tag{12.167}$$

where Eq. (12.162), the equation of state for radiation components $\mathcal{P}_\gamma = (1/3)\rho_\gamma$ and $\mathcal{P}_\nu = (1/3)\rho_\nu$, and for radiation dominance $H^2 = (8\pi N_G/3)(\rho_\gamma + \rho_\nu)$ are used. Furthermore, in the radiation-dominated regime $a \sim \sqrt{t}$ and $H = 1/(2t)$, then the factor $a^2 H^3 t^2$ is time independent. Outside the horizon, the last term of the second line on the right-handed side of Eq. (12.166), which is proportional to $q^2/(a^2 H^2)$, can be dropped. Then Eq. (12.166) gives

$$\hat{\psi}_q = -\frac{t q^2 \mathcal{R}_q^{(o)}}{a^2}, \tag{12.168}$$

where $\mathcal{R}_q^{(o)}$ denotes that \mathcal{R}_q is evaluated outside the horizon.

We can now write down the solutions for the perturbation functions in Mode 1:

$$\begin{aligned} \psi_q^{(M1)}(t) &= -\frac{q^2 \mathcal{R}_q^{(o)} t}{a^2}, \\ \delta_q^{(M1)}(t) &= \delta_{\gamma q}(t) = \delta_{\nu q}(t) = \delta_{Bq}(t) = \delta_{Dq}(t) = \frac{q^2 \mathcal{R}_q^{(o)} t^2}{a^2}, \\ \delta u_q^{(SM1)}(t) &= \delta u_{\gamma q}^{(S)}(t) = \delta u_{\nu q}^{(S)}(t) = -\frac{2}{9} \frac{q^2 \mathcal{R}_q^{(o)} t^3}{a^2}, \end{aligned} \tag{12.169}$$

which will serve as the initial conditions, i.e., the leading term, for the perturbation functions outside the horizon. The solutions to be obtained later that are outside the horizon and valid in the time regime that includes this initial time at the temperature $T = 10^9$ K will have to be matched with Eqs. (12.169) in their leading terms in the cosmic time t .

Let us now demonstrate that the $(q^2/a^2)\delta u_{\beta q}^{(S)}$ terms in expressions of Eq. (12.158) are negligible. We can write, using the Model 1 expressions

¹⁶¹See [221], p. 277 above Eq. (6.2.23).

$$\begin{aligned} \frac{q^2}{a^2} \delta u_q^{(SM1)} &= -\frac{2 q^2 q^2 \mathcal{R}_q^{(o)} t^3}{9 a^2 a^2} \\ &= \frac{1}{18} \frac{q^2}{a^2 H^2} \psi_q^{(M1)}(t), \end{aligned} \quad (12.170)$$

where we have used the relation $H = 1/(2t)$. Indeed, the terms proportional to q^2/a^2 are negligible outside the horizon.

The other solution $\delta_q \sim t^{-1}$, which gives $\hat{\psi} \sim t^{-2}$ and $\delta u^{(S)} \sim \text{const}$, is a decaying solution for which the perturbations decrease in time. So they are not relevant to structure formation in cosmic evolution. Hence it does not concern us here. There are other modes of solution which we refer to [221] for more discussions.¹⁶²⁾

12.9.3 The approach to analytic solutions

12.9.3.1 Strategy for analytic solutions

The parameter space that describes the perturbation functions, besides the various energy components, involves the cosmic time t and the physical wave-number q/a corresponding to the wave-length $\lambda_q = 2\pi a/q$. The physical wave-length covers the whole range from 0 to ∞ , and the cosmic time extends almost the whole cosmic history. For the present treatment, we consider the cosmic time to start from a few hundreds seconds after the Big Bang, corresponding to the cosmic temperature $T \sim 10^9$ K when the physics of the universe has become well-known and clearly formulated. This is a vast, multi-dimensional space with immensely complicated dynamic processes involved. Not surprisingly, to obtain an analytic solution is a serious challenge. It requires judicious division of the parameter space into subregions in time and wavelength, and making approximations to simplify the dynamic equations that drive the evolution of the universe.

The cosmic time under consideration can be divided into two large periods. The first period is from this initial time to the LSS of about 380 000 years when the cosmic temperature is $T \sim 3000$ K at the red shift and $z_L = 1100$. The second period is from LSS to the present at $\tau_U = 13.8$ billion years. In the first period the universe went through important milestones which cause the universe to change its properties significantly. Containing the period of the matter-radiation equality, the first time period can be further divided into a radiation-dominated followed by a matter-dominated period. We will come back to the matter-radiation equality below and more details can be found in Section 9.5.4. In dealing with property changes of the universe, significant simplifications have to be made at times, however, with the essential physical properties of the system considered not being altered. The wavelength space is divided into a

long wavelength region and a short wavelength one. Long wavelengths contribute to temperature coefficients of low multipoles, while short wavelengths to those of high multipoles.

As stated earlier, we follow closely the treatment given in [221], which will be frequently referred to, in footnotes mostly, for details. We will not reproduce the whole argument that leads to the analytic solutions. We will present the details in the initial stages of the consideration when the situation is still relatively simple. In later stages, relevant results given in [221] will be quoted, and details will be referred to the pertinent pages and/or equation numbers.

Let us first briefly describe the milestone in the first period of the cosmic time that is critical for obtaining the analytic solutions: In the early times the universe is weighted heavily in favor of radiation. The density ratio of matter to radiation $\bar{\rho}_M/\bar{\rho}_R \sim a$ is very small. However, the ratio increases with time like \sqrt{t} . At LSS, matter and radiation contribute equally, known as the *matter-energy equality*. Afterward, the weight of matter vs radiation is reversed and the universe becomes matter-dominated. Let us denote the time of matter-energy equality by t_{EQ} ,

$$\left. \frac{\bar{\rho}_M}{\bar{\rho}_R} \right|_{t_{\text{EQ}}} = 1. \quad (12.171)$$

As can be seen in Table 9.6, $t_{\text{EQ}} \approx 50\,000$ yrs, and the temperature and redshift are $T_{\text{EQ}} \approx 10^4$ K and $z_{\text{EQ}} \approx 3500$. We will see below that the matter-radiation equality plays an important role in the present consideration.

Let us call the Hubble length defined by $D_H = H^{-1}$ ($c = 1$) the *horizon*. The scale of the physical wavelength is given by the ratio of the horizon to the physical wavelength

$$\frac{D_H}{\lambda_q} \equiv \frac{H^{-1}}{\lambda_q} = \frac{1}{2\pi} \frac{q}{aH}. \quad (12.172)$$

We refer to the situation that the horizon is smaller than the wavelength $H^{-1}/\lambda < 1$ as *outside the horizon*, and the situation that the horizon is greater than the wavelength $H^{-1}/\lambda > 1$ as *inside the horizon*. We summarize this, dropping the numerical factor $1/(2\pi)$:

$$\frac{q}{aH} = \frac{q}{a} \begin{cases} < 1, & \text{outside the horizon,} \\ > 1, & \text{inside the horizon.} \end{cases} \quad (12.173)$$

In the very early times of the inflationary epoch and also a certain period afterward, because of the huge exponential expansion while the Hubble expansion rate stays constant, so except for very short wavelengths or very large wave-numbers, the majority of the waves, which is represented by a very large range in the q value, are outside

¹⁶²⁾pp 278–279, [221].

the horizon. After this very brief period of time, when the universe further evolves, $q/(aH)$ grows. For instance in the radiation-dominated period $a \sim t^{1/2}$ and $H \sim t^{-1}$, then $q/(aH) \sim t^{-1/2}$. And in the matter-dominated period $a \sim t^{2/3}$ and $H \sim t^{-1}$, then $q/(aH) \sim t^{-1/3}$. Eventually, most wavelengths or most wave numbers, except for the very long ones of very small q values, get inside the horizon, i.e., $q/(aH) > 1$. A special wave-number called q_{EQ} is defined at t_{EQ} for the ratio $q/(aH)$ to be unity,

$$q_{\text{EQ}} \equiv aH|_{t=t_{\text{EQ}}} \equiv (aH)_{\text{EQ}} = a_0 H_0 \frac{\sqrt{2}\Omega_M}{\sqrt{\Omega_R}}, \quad (12.174)$$

where a_0 is the FLRW scale factor at the present epoch, H_0 the corresponding Hubble expansion rate. We have also used the identity $a_{\text{EQ}}/a_0 = \Omega_R/\Omega_M$. The details can be found in Section 13.4.3.¹⁶³⁾

Because of these very different time and wavelength behaviors, there does not exist a single set of analytic expressions for the perturbation functions to be valid for all wavelengths and in the whole time period under consideration. So we have to suitably divide the $q-t$ space into regions such that the properties of the perturbation functions are sufficiently uniform in a given region to allow, under suitable approximations, analytic expressions for the perturbation functions can be obtained. It is also necessary to require neighboring regions to overlap so that leading terms of solutions in different region of the same perturbation function are matched in overlapping regions.

The $q-t$ space is separated into four regions divided

by the lines $q/q_{\text{EQ}} = 1$ and $t/t_{\text{EQ}} = 1$. Each of the four regions have their own analytic solutions for the perturbation functions which are valid in a significant part of the neighboring regions, so that in their overlapping region, the two solutions have the common leading term. The solutions for the regions containing the early time $T = 10^9$ K, will be required to satisfy the every early time initial conditions given in Eq. (12.169) as discussed in Section 12.9.2.

12.9.3.2 Division of the $t-q$ space

It is convenient to define the cosmic time t and the wave number q in units of their values at the matter-radiation equality, i.e., using the variables t/t_{EQ} and q/q_{EQ} . Define the following scaled variables,

$$\bar{t} \equiv \frac{t}{t_{\text{EQ}}}, \quad \bar{q} \equiv \frac{q}{q_{\text{EQ}}}, \quad \bar{a} \equiv \frac{a}{a_{\text{EQ}}} = \frac{\bar{\rho}_M}{\bar{\rho}_R}, \quad (12.175)$$

where the last identity can be verified by noting that $\bar{\rho}_M \sim a^{-4}$, $\bar{\rho}_R \sim a^{-3}$, and $\bar{\rho}_{REQ} = \bar{\rho}_{MEQ}$. The $t-q$ space can be conveniently divided into regions according to lines of constant \bar{t} and \bar{q} .

We will first explore some of the useful relationship in the scaled variables \bar{t} , \bar{q} , and \bar{a} . The ratio $aH/(aH)_{\text{EQ}}$ has been investigated in Section 13 and an analytic expression valid for the cosmic period in which the vacuum energy is negligible in comparison with matter and radiation is given in Eq. (13.53), Section 13.4.3, to which we refer for more details. First, we have

$$\frac{aH}{(aH)_{\text{EQ}}} = \frac{1}{\sqrt{2}} \frac{\sqrt{\bar{a}(\bar{t})+1}}{\bar{a}(\bar{t})} \approx \begin{cases} \frac{1}{\sqrt{2\bar{a}(\bar{t})}} \sim (\bar{t})^{-1/2}, & \bar{t} < 1 (\bar{a}(\bar{t}) < 1) \text{ radiation-dominated,} \\ \frac{1}{\sqrt{2\bar{a}(\bar{t})}} \sim (\bar{t})^{-1/3}, & \bar{t} > 1 (\bar{a}(\bar{t}) > 1) \text{ matter-dominated.} \end{cases} \quad (12.176)$$

The horizon to wavelength ratio can be written in various forms which will be useful below,

$$\begin{aligned} \frac{q}{aH} &= \frac{q/q_{\text{EQ}}}{(aH)/(aH)_{\text{EQ}}} = \bar{q} \frac{\sqrt{2\bar{a}}}{\sqrt{\bar{a}+1}} = \bar{q} \frac{\sqrt{2}}{\sqrt{\bar{a}+1}} \frac{\bar{\rho}_M}{\bar{\rho}_R} = \bar{q} \frac{\sqrt{2\bar{a}}}{\sqrt{\bar{a}+1}} \sqrt{\frac{\bar{\rho}_M}{\bar{\rho}_R}} \\ &\approx \bar{q} \begin{cases} \sqrt{2} \frac{\bar{\rho}_M}{\bar{\rho}_R} \sim (\bar{t})^{1/2}, & t < 1 (\bar{a}(\bar{t}) < 1) \text{ radiation-dominated,} \\ \sqrt{2} \sqrt{\frac{\bar{\rho}_M}{\bar{\rho}_R}} \sim (\bar{t})^{1/3}, & t > 1 (\bar{a}(\bar{t}) > 1) \text{ matter-dominated,} \end{cases} \end{aligned} \quad (12.177)$$

¹⁶³⁾At the matter-radiation equality, $\rho_{MEQ} = \rho_{REQ}$ gives $a_{\text{EQ}}/a_0 = \Omega_R/\Omega_M$, where the subscript 0 denote quantities at the present epoch and $\Omega_R = \rho_R/\rho_c$, etc., where ρ_c is the critical density. Neglecting the contribution of the dark energy which is a tiny fraction of the cosmic energy at this early universe, we have the Hubble expansion rate at the matter-radiation equality $H_{\text{EQ}}^2 = (8\pi G_N/3)(2\rho_{MEQ})$. The Hubble expansion rate at the present time is $H_0^2 = (8\pi G_N/3)\rho_0$. For a flat universe, $\rho_0 = \rho_c$. Then $H_{\text{EQ}}/H_0 = \sqrt{2\rho_{MEQ}/\rho_0} = \sqrt{2\Omega_M(a_0/a_{\text{EQ}})^3} =$

where we have used the last identity given in Eq. (12.175). In comparison with the exact expressions, the approximate expressions in Eqs. (12.176) and (12.177) are good within 20%.

In the radiation dominated region $\bar{t} < 1$ we have $\bar{a}(\bar{t}) < 1$, and in the mass dominated region $\bar{t} > 1$ and $\bar{a}(\bar{t}) > 1$. This suggests that we can construct approximately $(a_0/a_{\text{EQ}})\sqrt{2}\Omega_M/\sqrt{\Omega_R}$.

invariant expressions from $q/(aH)$ and $\bar{\rho}_M/\bar{\rho}_R = \bar{a}(\bar{t})$ in the matter- and radiation-dominated regions as can be seen in Eq. (12.177):

- $(q/(aH))(\bar{\rho}_R/\bar{\rho}_M)$ is approximately constant in time in the radiation dominated region of $\bar{t} < 1$, varying by a factor of $\sqrt{2}$ having the value between $\sqrt{2}\bar{q}$ and \bar{q} .
- $(q/(aH))^2(\bar{\rho}_M/\bar{\rho}_R)$ is close to being a constant in the matter-dominated region of $\bar{t} > 1$, varying also by a factor $\sqrt{2}$ having also the value between \bar{q} and $\sqrt{2}\bar{q}$.

In Fig. 12.7 we plot several relevant quantities as functions of \bar{t} . This is a busy figure containing 6 curves representing 6 functions labeled along the left- and right-handed vertical axis of the same scale. The curves are tagged numerically 1 to 6. Some details of these curves are given below:

- Curves 1 and 2. The figures represent \bar{q} vs \bar{t} for a fixed value of $q/(aH)$ according to the expression given in Eq. (12.177)

$$\bar{q} = \frac{q}{aH} \left(\frac{\sqrt{2\bar{a}(\bar{t})}}{\sqrt{\bar{a}(\bar{t})} + 1} \right)^{-1}. \quad (12.178)$$

The solid blue curve 1 is for $q/(aH) = 1$ tracing $\sqrt{\bar{a}(\bar{t})} + 1/(\sqrt{2\bar{a}(\bar{t})})$, where $\bar{a}(\bar{t})$ is given in Eq. (13.63). The dotted blue curve 2 is for the ap-

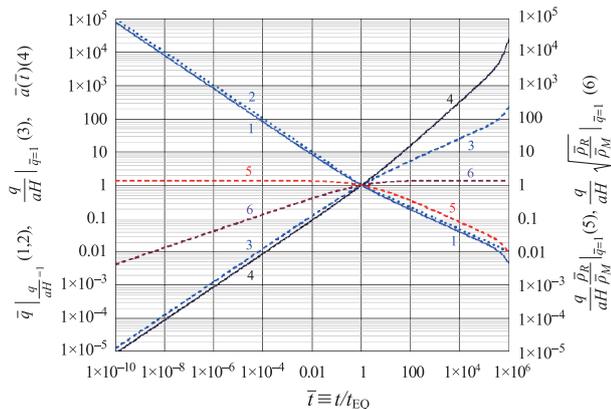


Fig. 12.7 The horizontal axis is $\bar{t} \equiv t/t_{\text{EQ}}$, i.e., cosmic time in units of the time of matter-radiation equality. The vertical axes, left- and right-handed sides having the same scale, represent six different functions. The left-handed side vertical axis is labeled for the curves 1 to 4, representing \bar{a} , and the relation between q/q_{EQ} and $q/(aH)$. The right-handed side vertical axis, labeled for curves 5 and 6, is for $(q/(aH))^n(\bar{\rho}_R/\bar{\rho}_M)$, $n = 1$ and 2 , with fixed $\bar{q} = q/q_{\text{EQ}} = 1$. Details of the various curves are given in the text.

proximated expression, which we can rewrite as,

$$\bar{q}|_{q/(aH)=1} = \left(\frac{\sqrt{2\bar{a}(\bar{t})}}{\sqrt{\bar{a}(\bar{t})} + 1} \right)^{-1} \approx (\bar{t})^{-1/2} \Theta(1 - \bar{t}) + (\bar{t})^{-1/3} \Theta(\bar{t} - 1). \quad (12.179)$$

Corresponding curves of different values of $q/(aH)$ are parallel to curves 1 and 2. Since $\bar{a}(\bar{t} = 1) = 1$, the value of $q/(aH)$ for a given curve can be simply read off from its intersect with the vertical line $\bar{t} = 1$.

- Dashed blue curve 3. The figure represents

$$\frac{q}{aH} = \bar{q} \frac{\sqrt{2\bar{a}}}{\sqrt{\bar{a}} + 1} \quad (12.180)$$

as given in Eq. (12.177) for $\bar{q} = 1$. Curves of different values of \bar{q} are parallel to curve 3 and the intersect of the curve with the vertical line $\bar{t} = 1$ gives the value of \bar{q} .

- Solid Black curve 4. This is \bar{a} vs \bar{t} according to Eq. (13.63). The range of application of the expression of \bar{a} is from the time of nucleosynthesis to the present.
- Dashed red curve 5. This curve plots

$$\frac{q}{aH} \frac{\bar{\rho}_R}{\bar{\rho}_M} = \bar{q} \sqrt{\frac{2}{\bar{a}(\bar{t})} + 1} \quad (12.181)$$

for $\bar{q} = 1$. The curve is flat for $\bar{t} < 1$, i.e., close to being constant in the $\bar{t} < 1$ region. Curves of different values of \bar{q} are parallel to curve 5, and the value of \bar{q} can be read off from the intersect of the curve with the vertical line $\bar{t} = 1$.

- Dashed dark brown curve 6. This represents

$$\frac{q}{aH} \left(\frac{\bar{\rho}_R}{\bar{\rho}_M} \right)^{1/2} = \bar{q} \sqrt{\frac{2\bar{a}(\bar{t})}{\bar{a}(\bar{t})} + 1} \quad (12.182)$$

for $\bar{q} = 1$. the curve is practically flat representing approximately a constant value for $\bar{t} > 1$. The value of \bar{q}^2 can also be read off from the intersect of the curve with the line $\bar{t} = 1$.

We make a few additional comments on Fig. 12.7 below:

- The time region under consideration extends from $\bar{t} \approx 10^{-10}$ to 10^6 . The initial time corresponds to $t \approx 200$. at the temperature $T = 10^9$ K and the final time to the present epoch of 13.8 Byrs, while the matter-radiation equality time is at $t \approx 49$ 000 yrs.

- The horizon to wavelength ratio, given by $q/(aH) = q/\dot{a}$ up to a factor $(2\pi)^{-1}$, classifies the wavelengths according to the curve $q/(aH) = 1$, represented by the blue solid curve 1. The region above $q/(aH) = 1$, i.e., $q/(aH) > 1$, which has wavelengths shorter than the horizon, is said to be inside the horizon, and the region below $q/(aH) = 1$ ($q/(aH) < 1$) is outside the horizon. In the very early times, most of the wavelengths, except for the very short ones or very large \bar{q} values, are outside the horizon. As can be seen from Fig. 12.7 at $\bar{t} = 10^{-8}$ all waves of $\bar{q} < 10^4$ are outside the horizon. At the initial time of $\bar{t} = 10^{-10}$ all waves of $\bar{q} < 10^5$ are outside the horizon.
- For a given wave of fixed q , as time increases, aH decreases and hence $q/(aH)$ increases, then more and more wavelengths outside the horizon will cross into the horizon. The shorter the wavelength or the larger the value of q , the earlier the crossing transition will occur. This is illustrated by $q/(aH)$ for fixed q in curves 3. All wavelengths of $\bar{q} > 1$ cross into the horizon before the matter-radiation equality time represented by the $\bar{t} = 1$ vertical line, while all those of $\bar{q} < 1$ make the crossing transition afterward. Hence for a given value of \bar{q} , the corresponding wavelength will sooner or latter cross to enter in the horizon, the larger the \bar{q} the sooner. In the radiation-dominated region, i.e., $\bar{t} < 1$, all wavelengths of $\bar{q} > 1$ will make the crossing transition in the radiation dominated era, while all wavelengths of $q/q_{\text{EQ}} < 1$ will make the transition in the matter-dominated era. For long wavelengths, say, $\bar{q} = 0.1$, the wave will stay outside the horizon even at the LSS.

12.9.3.3 Magnitudes and engineering dimensionalities of relevant quantities

Here we made an estimate of several relevant parameters in order to see the physical relevance of this division of the $t - q$ space.¹⁶⁴⁾ In matter-radiation dominated epoches various quantities can be expressed conveniently in terms of measurable cosmological parameters, such as the redshift and present day matter and energy density fractions. The basic tools needed for this task are the cosmic time behavior of the scale factor in the radiation and matter-dominated eras: $a_R \sim \sqrt{t}$ and $a_M \sim t^{2/3}$, and redshift relationship $a = (1 + z)^{-1}a_0$, where the subscript 0 denotes quantities at the present epoch. It should be kept in mind that numerical values of cosmological parameters are updated annually. But the variation of values of the parameters involved are generally smaller than the uncertainties introduced by the approx-

imations employed in the present calculation. However, keeping numerical values updated is a good practice.¹⁶⁵⁾

- The redshift at the matter-radiation equality, using the fact that $\rho_R \sim a^{-4}$ and $\rho_M \sim a^{-3}$ we can write

$$1 + z_{\text{EQ}} = \frac{a_0}{a_{\text{EQ}}} = \frac{\rho_{M0} \rho_{REQ}}{\rho_{R0} \rho_{MEQ}} = \frac{\Omega_M}{\Omega_R} = 2.273 \times 10^4 (\Omega_M h^2) \approx 3200. \quad (12.183)$$

The magnitude of the redshift at the last scattering surface, $1 + z_L = 1100$, is given in Table 9.6 and discussed in Section 9.5.6.

- The Hubble expansion rate in the radiation and matter-dominated eras:¹⁶⁶⁾ The radiation part of the contribution gives

$$H_R^2 = \frac{8\pi G_N}{3} \rho_R = H_0^2 \frac{\rho_R}{\rho_C} = H_0^2 \Omega_R \left(\frac{a_0}{a}\right)^4 = H_0^2 \Omega_R (1 + z)^4, \quad (12.184)$$

which gives

$$H_R = H_0 \sqrt{\Omega_R} (1 + z)^2 = 3.2409 \times 10^{-18} \sqrt{\Omega_R h^2} (1 + z)^2 \text{ s}^{-1} = 2.15 \times 10^{-20} (1 + z)^2 \text{ s}^{-1}. \quad (12.185)$$

Similarly the matter part of the Hubble expansion rate gives¹⁶⁷⁾

$$H_M = H_0 \sqrt{\Omega_M} (1 + z)^{3/2} = 3.2409 \times 10^{-18} \sqrt{\Omega_M h^2} (1 + z)^{3/2} \text{ s}^{-1} = 1.221 \times 10^{-18} (1 + z)^{3/2} \text{ s}^{-1}. \quad (12.186)$$

The vacuum energy has a constant contribution

$$H_\Lambda = H_0 \sqrt{\Omega_\Lambda} = 3.2409 \times 10^{-18} \sqrt{\Omega_\Lambda h^2} = 1.81 \times 10^{-18} \text{ s}^{-1}. \quad (12.187)$$

The total Hubble expansion rate is given by

$$H = \sqrt{H_\Lambda^2 + H_R^2 + H_M^2}. \quad (12.188)$$

The Λ -radiation equality occurs at redshift $z \approx 9$, and the Λ -matter equality at $z \approx 0.42$. So matter-radiation will dominate for all eras of the redshift

¹⁶⁵⁾We use the following cosmological parameters provided in the 2014 Particle Data Book: $H_0 = 3.240905h \times 10^{-18} \text{ s}^{-1}$, $\Omega_M = 0.142h^{-2}$, $\Omega_R = 4.4h^{-2} \times 10^{-5}$, $h = 0.673$, $\Omega_\Lambda = 0.685$, and $\Omega_M = 0.315$. The temperature of the present epoch is taken to be $T_0 = 2.7255 \text{ K}$. A useful conversion factor is $1 \text{ Mpc} = 1.029 \times 10^{14} \text{ s}$.

¹⁶⁶⁾Following [221], we treat neutrinos as massless and hence they are included in the radiation part of the energy density.

¹⁶⁷⁾For a consistency check, Eq. (12.183) can be obtained from Eqs. (12.185) and (12.186)

¹⁶⁴⁾This subsection corresponds to pp 280–282, [221].

higher than a few tens, from the every early universe down to the appearance of first stars.

Then the Hubble expansion rate at the matter-radiation equality is

$$H_{\text{EQ}} = \sqrt{2}H_{\text{M}}|_{\text{EQ}} = 1.77(\Omega_{\text{M}}h^2)^2 \times 10^{-11} \text{ s}^{-1} \approx 3.55 \times 10^{-13} \text{ s}^{-1}, \quad (12.189)$$

which is already a rather small expansion rate.

- The physical wavelength at the matter-radiation equality viewed in the present epoch:¹⁶⁸⁾

$$\begin{aligned} \lambda_{\text{EQ0}} &= \frac{2\pi}{q_{\text{EQ}}/a_0} = \frac{2\pi}{q_{\text{EQ}}/a_{\text{EQ}}} \frac{a_0}{a_{\text{EQ}}} = \frac{2\pi}{H_{\text{EQ}}} (1+z_{\text{EQ}}) \\ &= 8.069 \times 10^{15} (\Omega_{\text{M}}h^2)^{-1} \\ &= 78.4 (\Omega_{\text{M}}h^2)^{-1} \text{Mpc} \cdot \text{s} \\ &\approx 553 \text{ Mpc} = 1800 \text{ Myr}. \end{aligned} \quad (12.190)$$

- The order of the multipole moments at the matter-radiation equality viewed by an Earth observer, ℓ_{EQ} , can be estimated, using Eqs. (12.186) and (12.189),

$$\begin{aligned} \ell_{\text{EQ}} &= q_{\text{EQ}} \cdot r_{\text{L}} = \frac{q_{\text{EQ}}}{a_{\text{EQ}}} \frac{a_{\text{EQ}}}{a_0} \frac{a_0}{a_{\text{L}}} a_{\text{L}} \cdot r_{\text{L}} \\ &= H_{\text{EQ}} \frac{1+z_{\text{L}}}{1+z_{\text{EQ}}} a_{\text{L}} \cdot r_{\text{L}} \\ &= H_0 \sqrt{2\Omega_{\text{M}}(1+z_{\text{EQ}})} (1+z_{\text{L}}) d_{\text{A}}^{(LSS)} \\ &= 3.12 \sqrt{2\Omega_{\text{M}}(1+z_{\text{EQ}})} \approx 140, \end{aligned} \quad (12.191)$$

where a_{L} and r_{L} are the FLRW scale factor and comoving diameter at LSS, z_{L} is the redshift at LSS, and $d_{\text{A}}^{(LSS)} \equiv a_{\text{L}} r_{\text{L}}$ is the *angular diameter distance* at the last scattering surface as given in Eq. (13.79). We have used the value $\Omega_{\text{M}} = 0.315$. For a general wave number q at LSS, we have

$$\ell_q = q \cdot r_{\text{L}} = \ell_{\text{EQ}} \bar{q} \approx 140 \bar{q}. \quad (12.192)$$

Hence the long wavelength regime $\bar{q} < 1$ contributes to “low” multipoles $\ell < 140$, and the short wavelength regime $\bar{q} > 1$ to “high” multipoles $\ell > 140$.

It is interesting and sometime useful to know the engineering dimensionality of physical quantities involved in a discussion. In the natural unit there is only one engineering dimension. We take it to be the dimension of the cosmic time denoted as $[t]$. So all physical quantities have the dimension in powers of $[t]$. We list the dimensionality of various quantities below:

¹⁶⁸⁾For a comparison of the numerical values, see Eq. (6.2.45), p. 281, [221].

- $[t]^2$: G_{N} ; metric perturbation function B ; vector part of the three-scalar vector coordinate shift $\epsilon^{(S)}$.
- $[t]$: The comoving coordinates $x^0 = t, x^j$; perturbation functions F, C_j ; scalar velocity potential $\delta u^{(S)}$; three-scalar and vector coordinate shift ϵ_0 and $\epsilon_j^{(V)}$.
- $[t]^0$: velocity; FLRW scale factor a ; metric tensors $g_{\mu\nu}$; metric perturbations $h_{\mu\nu}$; metric perturbation functions E, G_j, A, D_{jk} ; velocity potentials u_μ ; perturbation functions of velocity potentials δu_0 and $\delta u_j^{(V)}$, the conserved quantity outside the horizon $\mathcal{R} = A/2 + H\delta u^{(S)}$.
- $[t]^{-1}$: energy; mass; H ; gravitational potential $\psi = (3\dot{A} + \nabla^2 \dot{B})/2$; wave number q .
- $[t]^{-2}$: scalar anisotropic inertia $\pi^{(S)}$.
- $[t]^{-3}$: vector anisotropic inertia $\pi_j^{(V)}$.
- $[t]^{-4}$: energy-mass densities and their perturbations ρ_j and $\delta\rho_j$; pressure and its perturbation \mathcal{P} and $\delta\mathcal{P}$; energy-momentum tensor $T_{\mu\nu}$; tensor anisotropic inertia π_{Jk} ,
- Fourier components of all types of functions have engineering dimension of their original quantity multiplied by $[t]^3$, For an examples, A_q has the dimension $[t]^3$, $\delta\rho_q$ has the dimension $[t]^{-1}$, etc.

12.9.4 Solutions in the long wavelength regime

The long wavelength regime is defined for the perturbation functions which have wavelengths sufficiently long so that they remain well outside the horizon at the matter-radiation equality. As shown in Fig. 12.7, this region lies well below curve 3, i.e., $q/(aH) = 1$. From Eq. (12.180) this means $\bar{q} \ll 1$. From curves 5 and 6 of Fig. 12.7, and Eqs. (12.181) and (12.182), we obtain the following constraints:

$$\frac{q}{aH} \frac{\bar{\rho}_R}{\bar{\rho}_M} = \bar{q} \frac{2}{\sqrt{\bar{a}+1}} \ll 1 \quad (12.193)$$

and

$$\left(\frac{q}{aH}\right)^2 \frac{\bar{\rho}_R}{\bar{\rho}_M} = \bar{q}^2 \frac{2\bar{a}}{\bar{a}+1} \ll 1. \quad (12.194)$$

The quantities $[q/(aH)](\bar{\rho}_R/\bar{\rho}_M)$ and $[q/(aH)]^2(\bar{\rho}_R/\bar{\rho}_M)$ are nearly constant respectively in the radiation- and matter-dominated regions.

Since $\bar{q} < 1$ throughout the long wavelength regime, the regime contributes only to multipoles not larger than 140, $\ell \leq 140$, according to Eq. (12.192).

For the six quantities that define the cosmic perturbations, which are related through Eqs. (12.157)–(12.159), it is not possible to have a single set of analytic solutions which is valid in the entire long wavelength regime which contains both radiation-dominated and matter dominated epochs. In order to obtain analytic solutions, we

further divide the long wavelength regime in cosmic time according to the approximate invariant constraints. The constraint of Eq. (12.193) in the radiation-dominated region, in which $[q/(aH)](\rho_R/\rho_M)$ is approximately constant in time, will be labeled as *Long wavelength outside horizon* (LOH). And the matter-dominated region with the constraint Eq. (12.194), in which $[q/(aH)]^2(\rho_R/\rho_M)$ is approximately constant, is labeled as *Long wavelength matter-dominated era* (LMD). With suitable additional approximations, sets of analytic solutions can be found separately for LOH and LMD. The two regions overlap in a small range in the beginning of the matter-dominated region. The two sets of solutions are matched in this overlapping range.

12.9.4.1 Long wavelength - Outside the horizon (LOH)

This region covers the radiation-dominated and an initial part of the matter-dominated eras, with $q/(aH) \ll 1$ or $q/a \ll H$ and $\bar{q} \ll 1$ according to Eq. (12.193) which involves the radiation-dominated regions and extends to the early part of a matter-dominated region for the cosmic time not too much later than t_{EQ} . So, in the defining differential equations Eqs. (12.157), (12.158), and (12.159) we can drop terms proportional to q^2/a^2 . From Eq. (12.158), all scaled density perturbations satisfy the same differential equation, i.e., $\dot{\delta}_{\beta q} = -\psi_q$. Then we have the adiabatic solution outside the horizon for all scaled perturbed density to be the same. Labeling all the relevant quantities with the superscript (LOH), we have

$$\delta_q^{(LOH)} \equiv \delta_{Dq}^{(LOH)} = \delta_{Bq}^{(LOH)} = \delta_{\gamma q}^{(LOH)} = \delta_{\nu q}^{(LOH)}. \tag{12.195}$$

The equations set Eqs. (12.157)–(12.159) is reduced to the following four equations:

$$\begin{aligned} \partial_t \left(a^2 \psi_q^{(LOH)} \right) &= -4\pi G_N a^2 \left(\bar{\rho}_M + \frac{8}{3} \bar{\rho}_R \right) \delta_q^{(LOH)}, \\ \partial_t \delta_q^{(LOH)} &= -\psi_q^{(LOH)}, \\ \partial_t \left(\frac{1 + R_B}{a} \delta u_{\gamma q}^{(LOH)} \right) &= -\frac{1}{3a} \delta_q^{(LOH)}, \\ \partial_t \left(\frac{1}{a} \delta u_{\nu q}^{(LOH)} \right) &= -\frac{1}{3a} \delta_q^{(LOH)}, \end{aligned} \tag{12.196}$$

where $R_B = 3\bar{\rho}_B/(4\bar{\rho}_R)$ as defined before in Eq. (12.160), $\bar{\rho}_R = \bar{\rho}_\gamma + \bar{\rho}_\nu$, and $\bar{\rho}_M = \bar{\rho}_B + \bar{\rho}_D$.

There are analytic solutions to the set of differential equations given in Eq. (12.196). We will state briefly how to obtain the solutions and refer to [221] for details.¹⁶⁹⁾ Substituting the expression of the second equation of Eq. (12.196) for $-\psi_q$ into the first equation, we

¹⁶⁹⁾pp 283–284, [221].

obtain a second order differential equation for δ_q . There are two particular solutions to this differential equation. The general solution is a linear combination of the two particular solutions. Imposing the boundary conditions given in Eq. (12.169), the following set of solutions is obtained,¹⁷⁰⁾

$$\begin{aligned} \psi_q^{(LOH)} &= \frac{2\sqrt{2}q^2\mathcal{R}_q^{(o)}}{5H_{EQ}a_{EQ}^2\bar{a}^4} \left(\sqrt{1+\bar{a}} (32+8\bar{a}-\bar{a}^3) - 32+24\bar{a} \right), \\ \delta_q^{(LOH)} &= \frac{4q^2\mathcal{R}_q^{(o)}}{5H_{EQ}^2a_{EQ}^2\bar{a}^2} \left(16+8\bar{a}-2\bar{a}^2+\bar{a}^3 - 16\sqrt{1+\bar{a}} \right), \\ \delta u_{\gamma q}^{(LOH)} &= -\frac{\sqrt{2}\bar{a}}{3H_{EQ}(1+R_B)} \int_0^{\bar{a}} \frac{\delta_q(\bar{a}')}{\sqrt{1+\bar{a}'}} d\bar{a}', \\ \delta u_{\nu q}^{(LOH)} &= -\frac{\sqrt{2}\bar{a}}{3H_{EQ}} \int_0^{\bar{a}} \frac{\delta_q(\bar{a}')}{\sqrt{1+\bar{a}'}} d\bar{a}', \end{aligned} \tag{12.197}$$

We can follow the treatment given in §12.9.2, in particular, for Eq. (12.170), to check that outside the horizon with $\bar{q}^2 \ll 1$, the terms $(q^2/a^2)\delta u_{\beta q}$, $\beta = \gamma, \nu$, are indeed much smaller than the other terms which appear in the same equations in Eq. (12.158). However, the present situation is more complicated because an explicit expression for $\delta u_{\beta q}$ is not available. We outline the proof in the footnote below.¹⁷¹⁾ Hence the omission of the term $(q^2/a^2)\delta u_{\beta q}$ is justified.

To prepare for the matching condition which is valid later in the matter-dominated era where $\bar{a} > 1$, we list

¹⁷⁰⁾These are Eqs. (6.3.13)–(6.3.16) of [221], in which $y = a/a_{EQ}$ is just \bar{a} . In checking the initial conditions, which is in the deep radiation-dominated regime, we make a leading order small \bar{a} expansion and notice the following relation: $H_{EQ}^2\bar{a}^{-4} = (8\pi G_N/3)2\bar{\rho}_{REQ}(a_{EQ}^4/a^4) = (8\pi G_N/3)2\bar{\rho}_{REQ}$, and $2H^2 = 1/(2t^2)$.

¹⁷¹⁾Referring to Eq. (12.197), we can write

$$\begin{aligned} \frac{q^2}{a^2} \delta u_{\nu q}^{(LOH)} &= \bar{q}^2 \frac{2\sqrt{2}q^2\mathcal{R}_q^{(o)}}{5H_{EQ}a_{EQ}^2\bar{a}^4} F_u(\bar{a}), \\ F_u(\bar{q}) &= -\frac{2}{3}\bar{a}^3 \int_0^{\bar{q}} \frac{d\bar{a}'}{\bar{a}'^2\sqrt{1+\bar{a}'}} \left(16+8\bar{a}'-2\bar{a}'^2+\bar{a}'^3 - 16\sqrt{1+\bar{a}'} \right), \end{aligned} \tag{12.198}$$

where the relation $q^2/a^2 = \bar{q}^2(H_{EQ}^2/\bar{a}^2)$ has been used. We also write

$$\begin{aligned} \psi_q^{(LOH)} &= \frac{2\sqrt{2}q^2\mathcal{R}_q^{(o)}}{5H_{EQ}a_{EQ}^2\bar{a}^4} F_\psi(\bar{a}), \\ F_\psi(\bar{a}) &= \sqrt{1+\bar{q}}(32+8\bar{a}-\bar{a}^3) - 32 - 24\bar{a}. \end{aligned} \tag{12.199}$$

We can numerically compare $F_u(\bar{a})$ and $F_\psi(\bar{a})$. The ratio $F_u(\bar{a})/F_\psi(\bar{a})$ begins at zero for $\bar{q} = 0$ and increases with \bar{a} . The ratio is 0.08 for $\bar{a} = 1$, approximately unity for $\bar{a} = 4.6$, and 2.9 for $\bar{a} = 10$.

the leading large $\bar{a} \gg 1$ expansion of above quantities:¹⁷²⁾

$$\begin{aligned} \psi_q^{(LOH)} \Big|_{\bar{a} \gg 1} &\rightarrow -\frac{2\sqrt{2}q^2\mathcal{R}_q^{(o)}}{5H_{EQ}a_{EQ}^2\sqrt{\bar{a}}} = -\frac{3q^2\mathcal{R}_q^{(o)}t}{5a^2}, \\ \rightarrow \delta_q^{(LOH)} \Big|_{\bar{a} \gg 1} &\rightarrow \frac{4q^2\mathcal{R}_q^{(o)}\bar{a}}{5H_{EQ}^2a_{EQ}^2} = \frac{9q^2\mathcal{R}_q^{(o)}t^2}{10a^2}, \\ \delta u_{\gamma q}^{(LOH)} \Big|_{\bar{a} \gg 1} &\rightarrow -\frac{8\sqrt{2}q^2\mathcal{R}_q^{(o)}\bar{a}^{5/2}}{45H_{EQ}^3a_{EQ}^2(1+R_B)} \\ &= -\frac{3q^2\mathcal{R}_q^{(o)}t^3}{10(1+R_B)a^2}, \\ \delta u_{\nu q}^{(LOH)} \Big|_{\bar{a} \gg 1} &\rightarrow -\frac{8\sqrt{2}q^2\mathcal{R}_q^{(o)}\bar{a}^{5/2}}{45H_{EQ}^3a_{EQ}^2} \\ &= -\frac{3q^2\mathcal{R}_q^{(o)}t^3}{10a^2}. \end{aligned} \quad (12.200)$$

Note that the final expressions are written to be proportional to a^{-2} in order to take care of the normalization factor in a_{EQ} .

12.9.4.2 Long wavelength – Matter-dominated era (LMD)

In the matter-dominated region $\bar{t} \gg 1$, we can drop the contribution of radiation in the equation of the gravitational potential Eq. (12.157). To further simplify the treatment, the baryonic matter contribution is also dropped in comparison to the dark matter.¹⁷³⁾ These approximations are to say that the dark matter largely determines the Hubble expansion rate and hence the perturbation to the gravitational potential. So the right-handed side of (12.157) can be written as $-(3/2)H^2\delta_{Dq}$. However, due to the much faster growing in time $\delta u_{jq} \sim t^3$ in comparison with $\psi_q \sim t$ as shown in Eq. (12.200), terms proportional to q^2/a^2 in expressions of Eq. (12.158) can not be neglected. Using the fact that in the matter-dominated regime $a \sim t^{2/3}$ and $H = 2/(3t)$, we can reduced Eqs. (12.157)–(12.159) to

$$\begin{aligned} \partial_t \left(t^{4/3} \psi_q^{(LMD)} \right) &= -\frac{2}{3t^{2/3}} \delta_{Dq}^{(LMD)}, \\ \partial_t \delta_{Dq}^{(LMD)} &= -\psi_q^{(LMD)}, \\ \partial_t \delta_{\gamma q}^{(LMD)} - \frac{q^2}{a^2} \delta u_{\gamma q}^{(LMD)} &= -\psi_q^{(LMD)}, \end{aligned}$$

¹⁷²⁾These include Eqs. (6.3.23) and (6.3.24) in [221]. To obtain large \bar{a} leading terms for expressions in Eq. (12.197), we have used the relation $H_{EQ} = \sqrt{2}H\bar{a}^{3/2}$ and $H = (2/3)t^{-1}$ in the matter-dominated regime. The second relation follows from $a \sim t^{2/3}$. The first relation can be proved as follows: $H^2 \approx (8\pi G_N/3)\bar{\rho}_M = (8\pi G_N/3)\bar{\rho}_{MEQ}(a_{EQ}^3/a^3) = H_{MEQ}^2\bar{a}^{-3}/2$.

¹⁷³⁾Note that $\Omega_B/\Omega_D = 0.02207/0.1198 = 18.4\%$ even as late as the present epoch. Hence neglecting the baryonic matter is a drastic approximation. But it enables analytic solutions to be obtained.

$$\begin{aligned} \partial_t \delta_{\nu q}^{(LMD)} - \frac{q^2}{a^2} \delta u_{\nu q}^{(LMD)} &= -\psi_q^{(LMD)}, \\ \partial_t \left(\frac{(1+R_B)}{t^{2/3}} \delta u_{\gamma q}^{(LMD)} \right) &= -\frac{1}{3t^{2/3}} \delta_{\gamma q}^{(LMD)}, \\ \partial_t \left(\frac{1}{t^{2/3}} \delta u_{\nu q}^{(LMD)} \right) &= -\frac{1}{3t^{2/3}} \delta_{\nu q}^{(LMD)}. \end{aligned} \quad (12.201)$$

As stated above the first two expressions indicate that the perturbation to the gravitational potential is determined by the dark matter density. The equations for the photon and neutrino are more complicated to solve. Below we sketch the solutions and refer the details to [221]. We discuss the different cosmic components below separately.

12.9.4.2.1 Perturbations in dark matter and gravitational potential

In this region under the approximation stated above, the perturbations to dark matter and gravitational potential are determined by themselves: Substituting the second expression of Eq. (12.201) into the first one, we obtain a second order equation for δ_{Dq} .

$$\partial_t \left(t^{4/3} \partial_t \delta_q^{(LMD)} \right) = \frac{2}{3t^{2/3}} \delta_q^{(LMD)}. \quad (12.202)$$

There are two particular solutions which are simple to obtained: One solution is $\delta_{Dq} \sim t^{2/3}$ and $\psi_q \sim t^{-1/3}$, and the other $\delta_{Dq} \sim t^{-1}$ and $\psi_q \sim t^{-2}$. These solutions have to be matched with the previous solution for long wavelengths outside the horizon in the large \bar{a} limit as given in Eq. (12.200). The first set of solutions can match with these initial conditions. The resultant functional forms of $\delta_{Dq}^{(LMD)}$ and $\psi_q^{(LMD)}$ are

$$\begin{aligned} \delta_{Dq}^{(LMD)} &= \frac{9q^2\mathcal{R}_q^{(o)}t^2}{10a^2}, \\ \psi_q^{(LMD)} &= -\frac{3q^2\mathcal{R}_q^{(o)}t}{5a^2}. \end{aligned} \quad (12.203)$$

We note that the second particular solution which is discarded in a decay solution.

12.9.4.2.2 Perturbations in photons and baryons

Perturbations of the photon and baryon fractional energy densities, which are identical $\delta_{Bq} = \delta_{\gamma q}$, and the photon velocity potential perturbation $\delta u_{\gamma q}$ are determined by the third and fifth equations of Eq. (12.201). One can obtain a second order inhomogeneous differential equation for either $\delta_{\gamma q}^{LMD}$ or $\delta u_{\gamma q}^{LMD}$. We follow [221] to work on the differential equation for $\delta_{\gamma q}^{LMD}$ then obtain $\delta u_{\gamma q}^{LMD}$ from the third expression of Eq. (12.201). The solution of an inhomogeneous differential equation is made of two parts: the solution of the inhomogeneous equation and that of the homogeneous equation. Let us

denote the two solutions for the fractional photon density perturbation function by, respectively, $\delta_{\gamma q}^{(iLMD)}$ and $\delta_{\gamma q}^{(hLMD)}$. We have

$$\begin{aligned} & \partial_t \left(\frac{a^2}{t^{2/3}} (1 + R_B) \partial_t \delta_{\gamma q}^{(iLMD)} \right) + \frac{q^2}{3t^{2/3}} \delta_{\gamma q}^{(iLMD)} \\ &= -\partial_t \left(\frac{a^2}{t^{2/3}} (1 + R_B) \psi_q^{(LMD)} \right), \\ & \partial_t \left(\frac{a^2}{t^{2/3}} (1 + R_B) \partial_t \delta_{\gamma q}^{(hLMD)} \right) + \frac{q^2}{3t^{2/3}} \delta_{\gamma q}^{(hLMD)} = 0, \end{aligned} \tag{12.204}$$

where $\psi_q^{(LMD)}$ is given by the second expression of Eq. (12.203) and $R_B \equiv 3\bar{\rho}_B/(4\bar{\rho}_R)$ is defined in Eq. (12.160). The complete solution to this differential equation is given by

$$\begin{aligned} \delta_{\gamma q}^{(LMD)} &= \delta_{\gamma q}^{(iLMD)} + C_h \delta_{\gamma q}^{(hLMD)}, \\ \delta u_{\gamma q}^{(LMD)} &= \delta u_{\gamma q}^{(iLMD)} + C_h \delta u_{\gamma q}^{(hLMD)}, \end{aligned} \tag{12.205}$$

where C_h is an arbitrary constant which is allowed by a homogeneous differential equation. We discuss separately the solutions of the homogeneous and inhomogeneous equations.

Solutions of the inhomogeneous equation

We start by considering the inhomogeneous differential equation Eq. (12.204). This complicated looking second order inhomogeneous differential equation has a set of closed-form analytic solutions¹⁷⁴⁾

$$\begin{aligned} \delta_{\gamma q}^{(i)} &= \frac{3q^2 \mathcal{R}_q^{(o)} (1 + 3R_B) t^2}{5a^2 (t^2 q^2/a^2 + 2R_B)}, \\ \delta u_{\gamma q}^{(i)} &= -\frac{3q^2 \mathcal{R}_q^{(o)} t^3}{5a^2 (t^2 q^2/a^2 + 2R_B)}. \end{aligned} \tag{12.206}$$

To proceed we look for regions in the $q - t$ space in which the above set of solutions to the inhomogeneous equations can be simplified. The simplified solutions together with analytic solutions of the homogeneous differential equations will form the required solutions of LMD. The simplification is to neglect either the term $t^2 q^2/a^2$ or $2R_B$ in comparison with the other in the denominators of the expressions in Eq. (12.206). We first consider the case $R_B \gg t^2 q^2/a^2$. Then $\delta_{\gamma q}^{(i)}$ and $\delta u_{\gamma q}^{(i)}$ above satisfy already their matching conditions given in Eq. (12.200) if $R_B \gg 1$. The condition $R_B \gg t^2 q^2/a^2$ can be satisfied for sufficiently small wavelengths because of the relation $q^2 t^2/a^2 = (q^2/(a^2 H^2)) t^2 H^2 = (4/9)(q^2/(a^2 H^2))$ in the matter-dominated region. However, the condition $R_B \gg 1$ can not be satisfied at LSS or earlier. This can

be seen by examining the magnitude of R_B :

$$\begin{aligned} R_B &= \frac{3\bar{\rho}_B}{4\bar{\rho}_R} = \frac{3\bar{\rho}_{B0}(a_0/a)^3}{4\bar{\rho}_{R0}(a_0/a)^4} \\ &= (1+z)^{-1} \frac{3\Omega_B}{4\Omega_R} = \frac{376.2}{1+z}, \end{aligned} \tag{12.207}$$

which varies from 0.11 at matter-radiation equality of $z_{EQ} = 3500$ to 0.34 at LSS $z_L = 1100$.¹⁷⁵⁾ So $R_B < 1$ in the range between matter-radiation equality and LSS.

So we make an additional constraint on the region of the wave number considered, i.e.,

$$R_B \equiv \frac{3\bar{\rho}_B}{4\bar{\rho}_R} \ll \frac{q^2 t^2}{a^2} \ll \frac{\bar{\rho}_M}{\bar{\rho}_R}. \tag{12.208}$$

This is consistent with the above assumption that $\bar{\rho}_B$ is much smaller than $\bar{\rho}_D$. This additional constraint also says that we restrict ourselves to moderately long wavelengths with the q value small enough so that $q/(aH) \ll 1$ at the matter-radiation equality. Since all terms in Eq. (12.208) have the same time behavior, i.e., proportional to $t^{2/3}$, the inequality can hold in the entire matter-dominated era. This moderate wavelength restriction enables us to establish a connection between the present solutions with those in the case of short wavelengths to be discussed in the next subsection. We note that short wavelength solutions are relevant to large values of ℓ . With the condition of moderately long wavelengths, Eq. (12.208), the inhomogeneous solution given in Eq. (12.206) can be approximated as

$$\begin{aligned} \delta_{\gamma q}^{(iLMD)} &= \frac{3}{5} (1 + 3R_B) \mathcal{R}_q^{(o)}, \\ \delta u_{\gamma q}^{(iLMD)} &= -\frac{3}{5} \mathcal{R}_q^{(o)} t. \end{aligned} \tag{12.209}$$

Solution of the homogeneous equation

The solution of the homogeneous differential equation given in Eq. (12.204) is more complicated. The exact solution is made of Gaussian hypergeometric functions. However, simpler analytic solutions can be obtained if additional approximations are made. For the purpose of imposing the matching conditions, we divide the region into two parts. In the first part, $R_B \ll 1$ is negligible, corresponding to the case of not so deeply into the matter-dominated era. In the second subregion R_B is not negligible compared to unity.

For $R_B \ll 1$, R_B can be dropped in comparison with unity in the second expression of Eq. (12.204), we can readily check that the differential equation has the particular analytic solutions $\sin(\sqrt{3}qt/a)$ and $\cos(\sqrt{3}qt/a)$.

¹⁷⁵⁾We use the value $\Omega_B = 0.02207h^{-2}$ and $\Omega_R = 4.4h^{-2} \times 10^{-5}$. The redshift at the matter-radiation equality is $z_{EQ} = 3500$ and at LSS $z_L = 1100$.

¹⁷⁴⁾See, Eqs. (6.3.25), p. 286, [221].

So the general homogeneous solutions for the photon density and velocity perturbations are¹⁷⁶⁾

$$\begin{aligned} \delta_{\gamma q}^{(h)} &= c_q \cos\left(\frac{\sqrt{3}q}{a}t\right) + d_q \sin\left(\frac{\sqrt{3}q}{a}t\right) \\ \delta u_{\gamma q}^{(h)} &= \frac{a}{\sqrt{3}q} \left(-c_q \sin\left(\frac{\sqrt{3}q}{a}t\right) + d_q \cos\left(\frac{\sqrt{3}q}{a}t\right)\right). \end{aligned} \quad (12.210)$$

These are not the required solutions yet, since the present region also includes the situation in which R_B becomes non-negligible in comparison with unity. So the homogeneous solution has to be modified too. In the presence of the R_B term in the homogeneous differential equation in Eq. (12.204), an approximated solution can be obtained by the WKB method, using Eq. (12.210) as guidance. The approximate solution for non-negligible R_B takes the form¹⁷⁷⁾

$$\begin{aligned} \delta_{\gamma q}^{(hLMD)} &= -\frac{3\mathcal{R}_q^{(o)}}{5} \frac{1}{(1+R_B)^{1/4}} \cos\varphi, \\ \delta u_{\gamma q}^{(hLMD)} &= \frac{3\mathcal{R}_q^{(o)}}{5} \frac{a}{\sqrt{3}q(1+R_B)^{3/4}t} \sin\varphi, \\ \varphi &\equiv \int_0^t \frac{qdt}{a\sqrt{3(1+R_B)}} \\ &= \frac{\sqrt{3}qt}{a\sqrt{R_B}} \ln\left(\sqrt{R_B} + \sqrt{1+R_B}\right). \end{aligned} \quad (12.211)$$

We note that the term $\ln(\sqrt{R_B} + \sqrt{1+R_B})/\sqrt{R_B}$ decreases monotonically as R_B increases, having the maximal value of unity at $R_B = 0$. Therefore, in the matter-dominated region and outside the horizon of $q^2t^2/a^2 \ll 1$, φ is small. It is straightforward to check that $\delta_{\gamma q}^{(hLMD)}$ and $\delta u_{\gamma q}^{(hLMD)}$ are reduced to $\delta_{\gamma q}^{(h)}$ and $\delta u_{\gamma q}^{(h)}$ when $R_B \ll 1$, with¹⁷⁸⁾ $c_q = -\mathcal{R}_q^{(o)}/5$ and $d_q = 0$ in Eq. (12.210).

Complete solutions for the photon and baryon

The photon solutions Eq. (12.205), which include both the inhomogeneous and homogeneous solutions given respectively in Eqs. (12.209) and (12.211), takes the form:

$$\begin{aligned} \delta_{\gamma q}^{(LMD)} &= \delta_{\gamma q}^{(iLMD)} + \delta_{\gamma q}^{(hLMD)}, \\ \delta u_{\gamma q}^{(LMD)} &= \delta u_{\gamma q}^{(iLMD)} + \delta u_{\gamma q}^{(hLMD)}, \end{aligned} \quad (12.212)$$

¹⁷⁶⁾See Eqs. (6.3.29) and (6.3.30) p. 289, [221]. We follow the same notation. In obtaining $\delta u_{\gamma q}^{(h)}$ from $\delta_{\gamma q}^{(h)}$, the term involving $\psi_q^{(LMD)}$ in the third expression of Eq. (12.201) has to be dropped as it is the inhomogeneous term.

¹⁷⁷⁾See pp 288–289, [221]. To verify the second equality in the expression for φ in Eq. (12.211) below, the following identifies in the matter dominated region can be used: $R_B = 3\bar{\rho}_B/(4\bar{\rho}_R) = (3/4)(\Omega_B/\Omega_R)(a/a_0)$ and $\dot{R}_B = (2/3)R_B t^{-1}$.

¹⁷⁸⁾See p. 288, [221].

which corresponds to $C_h = 1$ in Eq. (12.205).

Now we have the solutions for the photon and baryon system,

$$\begin{aligned} \delta_{\nu q}^{(LMD)} &= \delta_{Bq}^{(LMD)} \\ &= \frac{3\mathcal{R}_q^{(o)}}{5} \left[1 + 3R_B - \frac{1}{(1+R_B)^{1/4}} \cos\varphi\right], \\ \delta u_{\nu q}^{(LMD)} &= \frac{3\mathcal{R}_q^{(o)}t}{5} \left[-1 + \frac{a}{\sqrt{3}q(1+R_B)^{3/4}t} \sin\varphi\right], \end{aligned} \quad (12.213)$$

where φ is given in Eq. (12.211).

12.9.4.2.3 Comment on the neutrino contribution

For completeness let us comment on the solutions for the neutrino system, although the neutrino system does not contribute significantly. As shown in Eq. (12.201) the differential equations for the neutrino system are identical to those of the photon system by setting $R_B = 0$. So the solutions for the neutrino system can be obtained similarly, i.e., by setting $R_B = 0$ in Eq. (12.213):

$$\begin{aligned} \delta_{\nu q}^{(LMD)} &= \delta_{Bq}^{(LMD)} = \frac{3\mathcal{R}_q^{(o)}}{5} (1 - \cos\varphi), \\ \delta u_{\nu q}^{(LMD)} &= \frac{3\mathcal{R}_q^{(o)}t}{5} \left(-1 + \frac{a}{\sqrt{3}qt} \sin\varphi\right). \end{aligned} \quad (12.214)$$

However, we should note that the contribution of the neutrino system in this wavelength region is negligible.

12.9.4.2.4 List of LMD solutions

We summarize the solutions for the case of matter-dominated long wavelength by collecting the expressions of Eqs. (12.203), (12.213), and (12.214):

$$\begin{aligned} \delta_{Dq}^{(LMD)} &= \frac{9q^2\mathcal{R}_q^{(o)}t^2}{10a^2} \\ \psi_q^{(LMD)} &= -\frac{3q^2\mathcal{R}_q^{(o)}t}{5a^2}. \\ \delta_{\gamma q}^{(LMD)} &= \delta_{Bq}^{(LMD)} = \frac{3\mathcal{R}_q^{(o)}}{5} \left[1 + 3R_B - \frac{1}{(1+R_B)^{1/4}} \cos\varphi\right], \\ \delta u_{\gamma q}^{(LMD)} &= \frac{3\mathcal{R}_q^{(o)}t}{5} \left[-1 + \frac{a}{\sqrt{3}q(1+R_B)^{3/4}t} \sin\varphi\right]. \\ \delta_{\nu q}^{(LMD)} &= \frac{3\mathcal{R}_q^{(o)}}{5} (1 - \cos\varphi), \\ \delta u_{\nu q}^{(LMD)} &= \frac{3\mathcal{R}_q^{(o)}t}{5} \left(-1 + \frac{a}{\sqrt{3}qt} \sin\varphi\right), \\ \varphi &\equiv \frac{\sqrt{3}qt}{a\sqrt{R_B}} \ln\left(\sqrt{R_B} + \sqrt{R_B+1}\right), \\ R_B &\equiv \frac{3\bar{\rho}_B}{4\bar{\rho}_R}. \end{aligned} \quad (12.215)$$

It can be checked that in the case out the horizon and under the condition Eq. (12.208), i.e., $R_B \ll (q^2 t^2)/a^2 \ll 1$, the boundary conditions Eq. (12.200) are satisfied.

12.9.5 Solutions in the short wavelength regime

This regime, in contrast to that of long wavelengths, includes wavelengths which are sufficiently short so that they are well inside the horizon at the matter-radiation equality. As can be seen in Fig. 12.7 this requires $\bar{q} \gg 1$. Hence Eq. (12.192) shows that this is the wavelength regime which contributes to large multipole moments of the CMB anisotropy $\ell \gg 140$. Perturbations in this wavelength regions cause gravitational condensations which lead to the structure formation of galaxies and clusters.

Similar to the long wavelengths regime the cosmic time space in the present case is again divided into the radiation dominated region deep inside the horizon. In contrast to Eqs. (12.193) and (12.194), the conditions in the region of short wavelength is

$$\frac{q}{aH} \frac{\bar{\rho}_R}{\bar{\rho}_M} = \bar{q} \frac{\sqrt{2}}{\sqrt{\bar{a} + 1}} \gg 1 \quad (12.216)$$

for the radiation-dominated region deep inside the horizon, and

$$\left(\frac{q}{aH}\right)^2 \frac{\bar{\rho}_R}{\bar{\rho}_M} = \bar{q}^2 \frac{2\bar{a}}{\bar{a} + 1} \gg 1. \quad (12.217)$$

in the matter-dominated region deep inside the horizon. Since in this short wavelengths region $\bar{q} \gg 1$, they contribute to large multipoles moments of the order of hundreds or greater.

12.9.5.1 Short wavelengths in the radiation dominated region (SRD)

In the radiation dominated region we drop matter density in comparison with radiation density. The governing differential equations can be obtained from Eqs. (12.157), (12.158), and (12.159) by dropping the matter density function in the gravitational field equation and drop the term R_B in comparison with unity in the photon momentum equation: using $a \sim t^{1/2}$

$$\begin{aligned} \frac{1}{t} \partial_t (t\psi_q) &= -\frac{32\pi}{3} G_N (\bar{\rho}_\gamma \delta_{\gamma q} + \bar{\rho}_\nu \delta_{\nu q}), \\ \partial_t \delta_{Dq} &= -\psi_q, \\ \partial_t \delta_{\gamma q} - \frac{q^2}{a^2} \delta u_{\gamma q} &= -\psi_q, \\ \partial_t (t^{-1/2} \delta u_{\gamma q}) &= -\frac{1}{3} t^{-1/2} \delta_{\gamma q}, \\ \partial_t \delta_{\nu q} - \frac{q^2}{a^2} \delta u_{\nu q} &= -\psi_q, \\ \partial_t (t^{-1/2} \delta u_{\nu q}) &= -\frac{1}{3} t^{-1/2} \delta_{\nu q}. \end{aligned} \quad (12.218)$$

We briefly describe how to further simplify these differential equations, check the fact that they satisfy the initial conditions, and then write down the solutions. We refer to [221] for details.¹⁷⁹⁾

- We are interested in adiabatic solutions in which all reduced density functions δ_q and velocity potential δu_q are the same at the very early time. In particular, we see in Eq. (12.218) that the perturbation functions of the photon and neutrinos satisfy the same set of differential equations, so we can set

$$\delta_{\gamma q} = \delta_{\nu q}, \quad \delta u_{\gamma q} = \delta u_{\nu q}. \quad (12.219)$$

- In the highly radiation-dominated regime, the gravitational equation, i.e., the first expression in Eq. (12.218) can be further simplified for adiabatic solutions by writing

$$\frac{8\pi G_N}{3} (\bar{\rho}_\gamma + \bar{\rho}_\nu) \approx H^2 = \frac{1}{4t^2}, \quad (12.220)$$

where we have taken $a \sim t^{1/2}$ to calculate $H = \dot{a}/a = 1/(2t)$.

The solutions to Eq. (12.218) are listed below:¹⁸⁰⁾

$$\begin{aligned} \psi_q^{(SRD)} &= \frac{3\mathcal{R}_q^{(o)}}{t} \left[\frac{2}{\theta} \sin \theta + \frac{2}{\theta^2} (\cos \theta - 1) - 1 \right], \\ \delta_{Dq}^{(SRD)} &= -6\mathcal{R}_q^{(o)} \int_0^\theta \left[\frac{2}{\theta'^2} \sin \theta' + \frac{2}{\theta'^3} (\cos \theta' - 1) - \frac{1}{\theta'} \right] d\theta', \\ \delta_{\gamma q}^{(SRD)} &= \delta_{\nu q}^{(SRD)} = 3\mathcal{R}_q^{(o)} \left[\frac{2}{\theta} \sin \theta + \frac{2}{\theta^2} (\cos \theta - 1) - \cos \theta \right], \\ \delta u_{\gamma q}^{(SRD)} &= \delta u_{\nu q}^{(SRD)} = 4t\mathcal{R}_q^{(o)} \left[\frac{1}{2\theta} \sin \theta + \frac{1}{\theta^2} (\cos \theta - 1) \right], \end{aligned} \quad (12.221)$$

where

$$\theta \equiv \frac{2qt}{\sqrt{3}a} \approx \frac{1}{\sqrt{3}} \frac{q}{aH}. \quad (12.222)$$

The approximation follows from the fact that in the deep radiation-dominated regime the Hubble expansion parameter is given by $H = 1/(2t)$.

It is straightforward to check that, in the small θ limit, expressions listed in Eq. (12.221) satisfy the initial conditions given in Eq. (12.169). The expression for the reduce density perturbation of dark matter is different from that of other reduced density perturbations. Straightforward numerical comparison shows that they are in fact the same for $\theta < 1$ or $q/(aH)$ not much larger than unity which is the case under consideration. So the adiabatic condition is satisfied.

¹⁷⁹⁾ pp 290–292, [221].

¹⁸⁰⁾ See Eqs. (6.4.11)–(6.4.14), p. 291, [221].

12.9.5.2 Short wavelengths-deep inside the horizon (SDH)

The system of differential equations Eqs. (12.157)–(12.159) have analytic solutions for short wavelengths well inside the horizon $q/(aH) \gg 1$ independent of the relative size of the matter and radiation density. Referring to Fig. 12.7, this includes much of the region above curve 1. It is not so straightforward to derive the solutions to the various perturbation functions. We shall give a short description of how to obtain the solution and then list the solutions given in [221] to which we refer for details, with the page and equation numbers provided.

- For deep inside the horizon, $q/a \gg H$, we have q/a and H which have the dimension of inverse of time,

provide two scales of time rate of change. Terms of time derivatives of the order of q/a are the *fast modes*, while those of the order of H are *slow modes*.

- The rapid oscillation in the fast mode causes a damping effect which allows us to ignore the effect of the neutrino.
- There are analytic solution for the fast and slow modes for the perturbation functions. Further more there is an overlapping region for the two modes where the two sets of solution are matched.

The complete solution of the perturbation functions in this regime, up to the era of recombination, are given blow.¹⁸¹⁾ On the right-handed sides of the four expressions in Eq. (12.223), the first terms are the solutions for the slow mode and the second terms the fast mode.

$$\begin{aligned} \psi_q^{(SDH)} &= -\frac{6\mathcal{R}_q^{(o)}a}{a_{\text{EQ}}t}\mathcal{J}(\hat{\kappa}) + 16\sqrt{3}\pi G_N \bar{\rho}_\gamma(2 + R_B)(1 + R_B)^{1/4} \left(\frac{a}{q}\right) \mathcal{R}_q^{(o)}e^{-\hat{\Gamma}_q(t)} \sin \Theta_q(t), \\ \delta_{Dq}^{(SDH)} &= -\frac{9\mathcal{R}_q^{(o)}a}{a_{\text{EQ}}}\mathcal{J}(\hat{\kappa}) + 48\pi G_N \bar{\rho}_\gamma(2 + R_B)(1 + R_B)^{3/4} \left(\frac{a}{q}\right)^2 \mathcal{R}_q^{(o)}e^{-\hat{\Gamma}_q(t)} \cos \Theta_q(t), \\ \delta_{\gamma q}^{(SDH)} &= \delta_{Bq}^{(SDH)} = \frac{6\mathcal{R}_q^{(o)}a^3(1 + 3R_B)}{a_{\text{EQ}}q^2t^2}\mathcal{J}(\hat{\kappa}) - \frac{3\mathcal{R}_q^{(o)}}{(1 + R_B)^{1/4}}e^{-\hat{\Gamma}_q(t)} \cos \Theta_q(t), \\ \delta u_{\gamma q}^{(SDH)} &= -\frac{6\mathcal{R}_q^{(o)}a^3}{a_{\text{EQ}}q^2t}\mathcal{J}(\hat{\kappa}) + \frac{\sqrt{3}a\mathcal{R}_q^{(o)}}{q(1 + R_B)^{3/4}}e^{-\hat{\Gamma}_q(t)} \sin \Theta_q(t), \end{aligned} \tag{12.223}$$

where the first terms on the right-handed side are the slow mode and the second terms the fast mode, and

$$\begin{aligned} \mathcal{R}_q^{(o)} &= \frac{1}{2}\hat{A}_q \Big|_{\text{outside horizon}} = \frac{1}{2} \left(A_q + 2H\delta u^{(S)} \right) \Big|_{\text{outside horizon}}, \\ \mathcal{J}(\hat{\kappa}) &= \left(-\frac{7}{2} + \gamma + \ln \frac{4\hat{\kappa}}{\sqrt{3}} \right), \\ \hat{\kappa} &= \sqrt{2}\bar{q} = \sqrt{2}\frac{q}{q_{\text{EQ}}} = \frac{\sqrt{2}q}{a_{\text{EQ}}H_{\text{EQ}}} = \frac{\sqrt{\Omega_R}}{H_0\Omega_M} \frac{q}{a_0} = \frac{19.3}{\Omega_M h^2} \frac{q}{a_0}, \\ \Theta_q(t) &= \frac{q}{\sqrt{3}} \int_0^t \frac{dt}{a\sqrt{1 + R_B}} \equiv \int_0^t \omega_q(t')dt', \quad \omega_q(t) \equiv \frac{1}{\sqrt{3}} \frac{q}{a\sqrt{1 + R_B}}, \\ \hat{\Gamma}_q(t) &= \frac{q^2}{6} \int_0^t \frac{t_\gamma}{a^2(1 + R_B)} \left(\frac{16}{15} + \frac{R_B^2}{1 + R_B} \right) dt \equiv \int_0^t \Gamma_q dt, \\ \Gamma_q &\equiv \frac{q^2}{6} \frac{t_\gamma}{a^2(1 + R_B)} \left(\frac{16}{15} + \frac{R_B^2}{1 + R_B} \right), \quad R_B \equiv \frac{3\bar{\rho}_B}{4\bar{\rho}_R}, \quad t_\gamma \equiv \frac{1}{\sigma_T n_e}, \end{aligned} \tag{12.224}$$

where $\mathcal{R}_q^{(o)}$ is defined in Eq. (12.124) evaluated outside the horizon and $\gamma = 0.5772156649\dots$ is the Euler constant. t_γ is the photon mean free path or mean free time in a non-relativistic plasma of electron number density

n_e and the Thomson scattering cross section σ_T is calculated in the rest frame of the plasma electron.

Let us remark that the first, slow mode terms and the second, fast mode terms in the above perturbation functions in Eq. (12.223) terms have different orders of magnitude which can be sorted out straightforwardly. The perturbation functions ψ_q and δ_{Dq} are dominated by

¹⁸¹⁾The expressions are Eqs. (6.4.59)–(6.4.62), p. 302, [221], where details of the derivation are given.

their respective first, slow mode terms. However owing to their oscillatory nature, the second, fast mode terms can have significant effect in the baryonic acoustic oscillation. The functions $\delta_{\gamma q}$ and $\delta u_{\gamma q}$ are dominated, by their second, faster mode terms at the horizon entry. But the slow terms have noticeable effect in the study of the CMB anisotropy in multipole coefficients C_ℓ which we will take up later.

12.9.6 *Interpolating between long and short wavelengths (ILS) – Transfer functions*

We have obtained analytic solutions in two special regions of the wavelength: (1) very long wavelengths which enter the horizon very late in comparison with

the matter-radiation equality and they contribute to low multipoles moments of $\ell < 140$, and (2) very short wavelengths which enter the horizon quite early in comparison with the matter-radiation equality and they contribute to high multipoles moments of the order of several hundred. To compare the analytic result with observations we have to know the solution for intermediate wavelengths that enter the horizon around the matter-radiation equality and contribute to the multipoles in the range of $\ell \approx 200$ where lies the first acoustic peak in the CMB anisotropy. Analytic expressions of solution in this intermediate wavelength region can be constructed to interpolate between long and short wavelengths obtained above. Let us quote the final expressions¹⁸²⁾ in leading order

$$\begin{aligned} \psi_q^{(ILS)} &= -\frac{3\mathcal{R}_q^{(o)}q^2t\mathcal{T}(\hat{\kappa})}{5a^2}, \\ \delta_{Dq}^{(ILS)} &= \frac{9\mathcal{R}_q^{(o)}q^2t^2\mathcal{T}(\hat{\kappa})}{10a^2}, \\ \delta_{\gamma q}^{(ILS)} &= \delta_{Bq}^{(ILS)} = \frac{3\mathcal{R}_q^{(o)}}{5} \left[(1 + 3B_R)\mathcal{T}(\hat{\kappa}) - \frac{1}{(1 + R_B)^{1/4}} e^{-\hat{r}_q} \mathcal{S}(\hat{\kappa}) \cos(\Theta_q(t) + \Delta(\hat{\kappa})) \right], \\ \delta u_{\gamma q}^{(S)(ILS)} &= -\frac{3\mathcal{R}_q^{(o)}}{5} \left[t\mathcal{T}(\hat{\kappa}) - \frac{a}{\sqrt{3}q(1 + R_B)^{3/4}} e^{-\hat{r}_q} \mathcal{S}(\hat{\kappa}) \sin[\Theta_q(t) + \Delta(\hat{\kappa})] \right], \end{aligned} \tag{12.225}$$

where $\hat{\kappa}$, $\hat{r}_q(t)$, and $\Theta_q(t)$ are given in Eq. (12.224). The transfer functions $\mathcal{T}(\hat{\kappa})$, $\mathcal{S}(\hat{\kappa})$, and $\Delta(\hat{\kappa})$, are given in [221]¹⁸³⁾ and we quote them below:¹⁸⁴⁾

$$\begin{aligned} \mathcal{T}(\hat{\kappa}) &= \frac{\ln(1 + (0.124\hat{\kappa})^2)}{(0.124\hat{\kappa})^2} \left[\frac{1 + (1.257\hat{\kappa})^2 + (0.4452\hat{\kappa})^4 + (0.2197\hat{\kappa})^6}{1 + (1.606\hat{\kappa})^2 + (0.8568\hat{\kappa})^4 + (0.3927\hat{\kappa})^6} \right]^{1/2}, \\ \mathcal{S}(\hat{\kappa}) &= \left[\frac{1 + (1.209\hat{\kappa})^2 + (0.5116\hat{\kappa})^4 + \sqrt{5}(0.1657\hat{\kappa})^6}{1 + (1.9459\hat{\kappa})^2 + (0.4249\hat{\kappa})^4 + (0.1657\hat{\kappa})^6} \right]^2, \\ \Delta(\hat{\kappa}) &= \frac{(1.1547\hat{\kappa})^2 + (0.5986\hat{\kappa})^4 + (0.2578\hat{\kappa})^6}{1 + (1.723\hat{\kappa})^2 + (0.8707\hat{\kappa})^4 + (0.4581\hat{\kappa})^6 + (0.2204\hat{\kappa})^8}, \end{aligned} \tag{12.226}$$

where $\hat{\kappa}$, \hat{r}_q , and Θ are defined in Eq. (12.224). For long wavelengths, i.e., $\hat{\kappa} \ll 1$, the transfer functions take the limiting values $\mathcal{T}(\hat{\kappa}) \rightarrow 1$, $\mathcal{S}(\hat{\kappa}) \rightarrow 1$, and $\Delta(\hat{\kappa}) \rightarrow 0$. Then $\psi_q^{(ILS)}$ and $\delta_{Dq}^{(ILS)}$ are reduced to $\psi_q^{(LMD)}$ and $\delta_{Dq}^{(LMD)}$.

Plots of the transfer functions are given in Fig. 12.8.¹⁸⁵⁾

12.10 *Temperature anisotropy due to scalar perturbations*

In this section we shall derive an analytic expression for the CMB anisotropy due to perturbation functions in the scalar mode, through a series of approximations in treating the cosmic evolution after LSS. The final result so obtained can be studied with simple numerical inte-

¹⁸²⁾The corresponding expressions are given in Eqs. (6.5.15) and (6.5.16) p. 309, and Eqs. (6.5.17) and (6.5.18) p. 310, [221], where details of the derivation can be found.

¹⁸³⁾The transfer function are given in Eqs. (6.5.12)–(6.5.14), pp 307–308, [221].

¹⁸⁴⁾The expression for $\Delta(\hat{\kappa})$ given below is the corrected form which can be found in the errata of the First Print of [221]. See: <http://zippy.ph.utexas.edu/weinberg/swcorrections.pdf>.

¹⁸⁵⁾This plot is similar to that of Fig. 6.1 given on p. 309, [221], where the function $\Delta(\hat{\kappa})$ has the form before the correction as mentioned in Footnote ¹⁸⁴⁾.

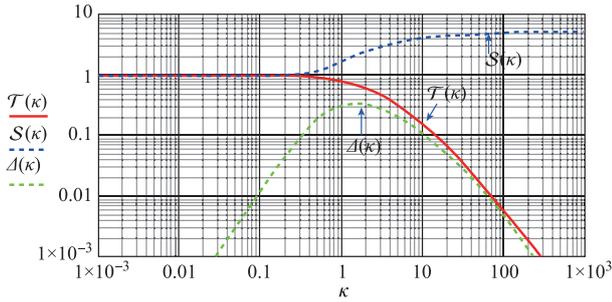


Fig. 12.8 Transfer functions defined in Eq. (12.226).

grations and readily compared with observational data. We will not concern ourselves with the tensor mode which is much smaller than the scalar mode. We will not concern ourselves with the CMB polarization either. Let us state again, we follow closely the approaches of [221] and frequently refer to its relevant expressions in footnotes.

In the preceding section we have obtained the expressions of the scalar perturbations associated with the gravitational potential due to matter and energy from very early times down to the last scattering surface. Evolutions of these expressions have to be worked out to the present time when the CMB anisotropy is observed. To enable an analytic calculation of this evolution, a key approximation will be made: a *sharp transition of photons from thermal equilibrium to free propagation*. In other words, the approximation says that the transition of photons from opacity, which means fully interacting with charged particles, to complete transparency, which means fully non-interacting free streaming, takes place at a certain time instantaneously. In this approximation, LSS is considered to be located at a sharp time, called say t_L , instead of being a process which happened during a significantly long time interval. So t_L is the time of the opacity to transparency transition. This approximation also neglects the scattering of CMB photons by matter that re-ionizes the baryon matter in the appearance of first stars, at around the red shift $z \approx 10$.¹⁸⁶⁾ However, certain relatively simple approaches can be used to make up for the approximation, including:

- The reionization corrections entailed in the temperature correlation can be simply taken into account for the case of multipoles $\ell > 20$. This will be included in the final result.
- The sharp transition can also be partially made up by including the effect of viscous damping, and by including effects of averaging over a finite time interval of the last scattering surface.

¹⁸⁶⁾See discussion in p. 329, [221].

12.10.1 Sudden transparency approximation – Following the photon trajectory

Under the assumption of sudden transparency at LSS, the expression of the CMB anisotropy can be derived by following the trajectory of a freely propagating CMB photon, in a perturbed metric, from the LSS to the present. The calculation can be carried out in a class of gauges in which the perturbation of the time-space components of the metric functions are set to vanish, i.e., $g_{0j} = 0$, such as the Newtonian and synchronous gauges. From Eq. (12.41), we have

$$\begin{aligned} g_{00} &= -1 + h_{00} \equiv -1 - E(\mathbf{r}, t), \\ g_{0j} &= h_{0j} \equiv a(\partial_j F + G_j) = 0, \\ g_{jk} &= a^2(t)\delta_{jk} + h_{jk}(\mathbf{r}, t) \\ &\equiv a^2[(1 + A)\delta_{jk} + \partial_{jk}B + \partial_j C_k + \partial_k C_j + D_{jk}]. \end{aligned} \quad (12.227)$$

Let us take the comoving frame with the observer at the origin and consider a pulse of photons traveling toward the observer along a radial direction \hat{n} . At the cosmic time t the coordinate of the photon is $r(t)\hat{n}$. Following a null path of its trajectory, the photon is on the geodesic

$$\begin{aligned} g_{jk}dx^j dx^k &= -[1 + E(r(t)\hat{n}, t)](dt)^2 \\ &\quad + [a^2(t) + h_{rr}] (dr)^2 = 0, \end{aligned} \quad (12.228)$$

which gives a time evolution equation for the photon radial coordinate, in an expanding universe,

$$\begin{aligned} \frac{d}{dt} r(t) &= - \left[\frac{1 + E(r(t)\hat{n}, t)}{a^2(t) + h_{rr}(r(t)\hat{n}, t)} \right]^{1/2} \\ &\equiv \frac{1}{a(t)} [-1 + N(\bar{r}(t)\hat{n}, t)], \end{aligned} \quad (12.229)$$

where $\bar{r}(t)$ is the radial coordinate of the background universe or the zeroth order form of the above expression. To the first order in the perturbation

$$N(\bar{r}(t)\hat{n}, t) = \frac{1}{2} \left[\frac{h_{rr}(\bar{r}(t)\hat{n}, t)}{a^2(t)} - E(\bar{r}(t)\hat{n}, t) \right]. \quad (12.230)$$

The negative sign on the right-handed side of the first line of Eq. (12.229) indicates that the photon pulse in question travels towards the origin. Note that since the perturbation functions E and h_{rr} are already first order, the coordinate argument in them can be replaced by the photon zeroth order radial coordinate $\bar{r}(t)$, i.e., that in the absence of metric perturbations. Dropping the perturbation functions, we obtain from Eq. (12.229) the zeroth order equation

$$\frac{d}{dt} \bar{r}(t) = - \frac{1}{a(t)}. \quad (12.231)$$

We can solve both Eqs. (12.229) and (12.231):

$$\bar{r}(t) = r_L - \int_{t_L}^t \frac{dt'}{a(t')}, \quad (12.232)$$

where t_L is the time of LSS and r_L the radial coordinate at the time of LSS. The radial coordinate as a function of the cosmic time is given by

$$\begin{aligned} r(t) &= \bar{r}(t) + \int_{t_L}^t \frac{1}{a(t')} N(\bar{r}(t')\hat{n}, t') dt' \\ &= r_L + \int_{t_L}^t \frac{1}{a(t')} [N(\bar{r}(t')\hat{n}, t') - 1] dt'. \end{aligned} \quad (12.233)$$

The initial conditions are satisfied:

$$r(t_L) = \bar{r}(t_L) = r_L \quad (12.234)$$

Assuming that this pulse of photon reaches the observer at the present epoch at t_0 , we have,

$$r(t_0) = 0 = r_L + \int_{t_L}^{t_0} \frac{1}{a(t)} [N(\bar{r}(t)\hat{n}, t) - 1] dt. \quad (12.235)$$

Up to now we have considered the trajectory of a fixed point, say a crest, of the wave of a photon pulse. Let us consider the trajectory of the next crest passing through the LSS to reach the observer at the origin. Suppose that the next crest passes through LSS at $t_L + \delta t_L$ and reaches the observer at $t_0 + \delta t_0$. This latter photon pulse satisfies an equation similar to Eq. (12.235), but with t_L and t_0 being replaced respectively by $t_L + \delta t_L$ and $t_0 + \delta t_0$. Furthermore, there is a variation in t_L due to the perturbation of the radial velocity of the photon gas or the photon-nucleon plasma. Combining the equations starting at t_L and those at $t_L + \delta t_L$, we obtain an equation which is the variation of the original Eq. (12.235). When δt_L and δt_0 are identified as the respective periods of the photon wave at LSS and at the present, denoted as τ_L and τ_0 , a relation between the two periods, or their frequencies denoted as ν_L and ν_0 , can be obtained. Note that the photon periods or their frequencies at t_L and t_0 are not the same because of the expansion of the universe, which affects the wavelength of the photon in question.

Let us illustrate this time variations below. First consider the variation in the zeroth order coordinate for $t < t_0$. From Eq. (12.232), we have

$$\begin{aligned} \delta\bar{r}(t) &= \delta r_L - \delta \int_{t_L}^t \frac{dt'}{a(t')} \\ &= \delta r_L - \left[\int_{t_L + \delta t_L}^t \frac{dt}{a(t)} - \int_{t_L}^t \frac{dt}{a(t)} \right] \\ &= \delta u_\gamma^{(r)}(r_L \hat{n}, t_L) \delta t_L + \frac{\delta t_L}{a(t_L)}. \end{aligned} \quad (12.236)$$

The first term on the right-handed side of the second line, $\delta r_L = \delta u_\gamma^{(r)}(r_L \hat{n}, t_L) \delta t_L$ arises from the perturbation of the radial velocity of the photon gas or the photon-nucleon plasma $\delta u_\gamma^{(r)}$, which introduces a non-vanishing variation to the zeroth order radial distance caused by the time elapse δt_L , even though r_L is a constant.

For the second photon wave crest, which leaves LSS at $t_L + \delta t_L$, its zeroth order radial coordinate is $\bar{r}(t) + \delta u_\gamma^{(r)} \delta t_L$. The variation of Eq. (12.235) is

$$\begin{aligned} 0 &= \delta r_L + \delta \int_{t_L}^{t_0} \frac{dt}{a(t)} [N(\bar{r}(t)\hat{n}, t) - 1] \\ &= \delta r_L + \left(\int_{t_L + \delta t_L}^{t_0 + \delta t_0} - \int_{t_L}^{t_0} \right) \frac{dt}{a(t)} [N(\bar{r}(t)\hat{n}, t) - 1] \\ &\quad + \int_{t_L}^{t_0} \frac{dt}{a(t)} \delta N(\bar{r}(t)\hat{n}, t), \end{aligned} \quad (12.237)$$

where

$$\begin{aligned} \delta N(\bar{r}(t)\hat{n}, t) &= N((\bar{r}(t) + \delta\bar{r}(t))\hat{n}, t) - N(\bar{r}(t)\hat{n}, t) \\ &= \frac{\partial}{\partial x} N(x\hat{n}, t) \Big|_{x=\bar{r}(t)} \frac{\delta t_L}{a(t_L)}. \end{aligned} \quad (12.238)$$

Note that in the second line of the above expression we have dropped a $\delta u_\gamma^{(r)}$ dependent term in $\delta\bar{r}(t)$ given by Eq. (12.236). The term makes a second order contribution.

We now have all the ingredients needed to calculate the variation of Eq. (12.235) from Eq. (12.237). It is

$$\begin{aligned} 0 &= \left[1 - N(r_L \hat{n}, t_L) + a(t_L) \delta u_\gamma^{(r)}(r_L \hat{n}, t_L) \right. \\ &\quad \left. + \int_{t_L}^{t_0} \frac{dt}{a(t)} \frac{\partial}{\partial x} N(x\hat{n}, t) \Big|_{x=\bar{r}(t)} \frac{\delta t_L}{a(t_L)} \right. \\ &\quad \left. - (1 - N(0, t_0)) \frac{\delta t_0}{a(t_0)} \right], \end{aligned} \quad (12.239)$$

where we have used the zeroth order result $\bar{r}(t_0) = 0$ in the first order term $N(\bar{r}(t_0), t_0)$. The above expression can be further simplified by the relation

$$\begin{aligned} &\frac{1}{a(t)} \frac{\partial}{\partial x} N(x\hat{n}, t) \Big|_{x=\bar{r}(t)} \\ &= -\frac{d}{dt} N(\bar{r}(t)\hat{n}, t) + \frac{\partial}{\partial t} N(x\hat{n}, t) \Big|_{x=\bar{r}(t)}. \end{aligned} \quad (12.240)$$

Substituting Eq. (12.240) into Eq. (12.239) and performing the doable integration, we obtain

$$\begin{aligned} 0 &= \left[1 - N(0, t_0) + a(t_L) \delta u_\gamma^{(r)}(r_L \hat{n}, t_L) \right. \\ &\quad \left. + \int_{t_L}^{t_0} dt \frac{\partial}{\partial t} N(x\hat{n}, t) \Big|_{x=\bar{r}(t)} \frac{\delta t_L}{a(t_L)} \right. \\ &\quad \left. - (1 - N(0, t_0)) \frac{\delta t_0}{a(t_0)} \right]. \end{aligned} \quad (12.241)$$

Now we can calculate the ratio of the two time intervals, to the first order in the perturbation,

$$\frac{\delta t_L}{\delta t_0} = \left[1 - \int_{t_L}^{t_0} dt \frac{\partial}{\partial x} N(x\hat{n}, t) \Big|_{x=\bar{r}(t)} - a(t_L) \delta u_\gamma^{(r)}(r_L \hat{n}, t_L) \right] \frac{a(t_L)}{a(t_0)}. \quad (12.242)$$

The periods of the photon at different time t are given by the proper time intervals which are

$$\begin{aligned} \delta \tau_L &= \sqrt{-g_{00}(r_L, t_L)} \delta t_L = \sqrt{1 + E(r_L, t_L)} \delta t_L, \\ \delta \tau_0 &= \sqrt{-g_{00}(0, t_0)} \delta t_0 = \sqrt{1 + E(0, t_0)} \delta t_0. \end{aligned} \quad (12.243)$$

We now obtain the ratio of the frequencies, to the first order in perturbation,

$$\begin{aligned} \frac{\nu_0}{\nu_L} &= \frac{\sqrt{1 + E(r_L, t_L)} \delta t_L}{\sqrt{1 + E(0, t_0)} \delta t_0} \\ &= \left[1 + \frac{1}{2} (E(r_L \hat{n}, t_L) - E(0, t_0)) - a(t_L) \delta u_\gamma^{(r)}(r_L \hat{n}, t_L) - \int_{t_L}^{t_0} dt \frac{\partial}{\partial t} N(x\hat{n}, t) \Big|_{x=\bar{r}(t)} \right] \frac{a(t_L)}{a(t_0)}. \end{aligned} \quad (12.244)$$

In the absence of the perturbation, this is just the discussion of the redshift shown in Eq. (9.60).

12.10.2 Temperature fluctuation due to scalar perturbations

We now calculate the temperature fluctuation at the present epoch by relating it to that at the LSS through the cosmic evolution. We observe that photons leaving LSS to propagate freely but under the influence of the universe will behave as if they are in equilibrium, having a number density $n_\gamma \sim T_\gamma^3$ and energy density $\rho_\gamma \sim T_\gamma^4$. The energy of individual photons are proportional to their frequencies $\epsilon_\gamma \sim \nu_\gamma$. Then the relation $\rho_\gamma = n_\gamma \epsilon_\gamma$ gives $T_\gamma^4 \sim T_\gamma^3 \nu_\gamma$ which leads, in turn, $\nu_\gamma \sim T_\gamma$. Hence the temperature T_0 at the present time t_0 is related to that at LSS $T_L \equiv T(t_L)$ at time t_L by the ratio of the frequencies of their photons. Let $T(\hat{n})$ be the temperature of the photon coming from the direction \hat{n} . We write

$$T(\hat{n}) = T(t_L) \frac{\nu_0}{\nu_L} \equiv \left[\bar{T}(t_L) + \delta T^{(i)}(r_L \hat{n}, t_L) \right] \frac{\nu_0}{\nu_L}, \quad (12.245)$$

where $\bar{T}(t_L)$ is the averaged temperature on LSS at t_L and $\delta T^{(i)}(r_L \hat{n}, t_L)$ the temperature fluctuation on LSS at t_L . The average temperature on LSS at t_L and that at the present time T_0 are related by

$$T_0 = \frac{a(t_L)}{a(t_0)} \bar{T}(t_L) \quad (12.246)$$

Now we can define the fractional temperature perturbation at the present time observed in the direction \hat{n} , to the first order in the perturbation,

$$\begin{aligned} \frac{\Delta T(\hat{n})}{T_0} &= \frac{T(\hat{n}) - T_0}{T_0} \\ &= \frac{\nu_0}{\nu_L} \frac{a(t_0)}{a(t_L)} \left[1 + \frac{\delta T^{(i)}(r_L \hat{n}, t_L)}{\bar{T}(t_L)} \right] - 1 \\ &= \frac{\nu_0}{\nu_L} \frac{a(t_0)}{a(t_L)} + \frac{\delta T^{(i)}(r_L \hat{n}, t_L)}{\bar{T}(t_L)} - 1. \end{aligned} \quad (12.247)$$

To obtain the last line we note that $\delta T^{(i)}(r_L \hat{n}, t_L)/\bar{T}(t_L)$ is of the first order in the perturbation, therefore we can take the zeroth order of the ratio $(\nu_0/\nu_L)(a(t_0)/a(t_L))$ which is unity.¹⁸⁷⁾ Substituting Eq. (12.244) into Eq. (12.247), we obtain

$$\begin{aligned} \frac{\Delta T(\hat{n})}{T_0} &= \frac{1}{2} (E(r_L \hat{n}, t_L) - E(0, t_0)) - a(t_L) \delta u_\gamma^{(r)}(r_L \hat{n}, t_L) \\ &\quad - \int_{t_L}^{t_0} dt \frac{\partial}{\partial t} N(x\hat{n}, t) \Big|_{x=\bar{r}(t)} + \frac{\delta T^{(i)}(r_L \hat{n}, t_L)}{\bar{T}(t_L)}. \end{aligned} \quad (12.248)$$

In the temperature perturbation there are both scalar and tensor contributions from the metric perturbations, while the metric vector perturbation is set to zero $h_{0j} = 0$ in the class of gauge under consideration. Since there is no correlation between the scalar and tensor perturbations in the temperature multipole coefficient C_ℓ , their contributions can be separately treated. As stated before we concern ourselves only with scalar perturbations which make the dominant contribution to the CMB anisotropy.

12.10.2.1 Dissection of scalar perturbations

The non-vanishing perturbation functions of the relevant scalar metric perturbation are given in Eqs. (12.38) and (12.40):

$$\begin{aligned} h_{00}^{(S)} &= -E, \quad h_{jk}^{(S)} = a^2 \left(A \delta_{jk} + \frac{\partial^2 B}{\partial x_j \partial x_k} \right), \\ N &= \frac{1}{2} \left(A + \frac{\partial^2 B}{\partial r^2} - E \right), \end{aligned} \quad (12.249)$$

where the superscript S denotes the scalar contribution. The scalar radial photon fluid velocity $\delta u_\gamma^{(r)}$ is obtained from the photon velocity potential $\delta u_\gamma^{(S)}$ as follows:¹⁸⁸⁾

$$\delta u_\gamma^{(r)} = \bar{g}^{r\lambda} \frac{\partial \delta u_\gamma^{(S)}}{\partial x^\lambda} = \frac{1}{a^2} \frac{\partial \delta u_\gamma^{(S)}}{\partial r}. \quad (12.250)$$

We have the scalar part of the fractional temperature fluctuation Eq. (12.248),

¹⁸⁷⁾In the background universe the product of the FLRW scale factor times the frequency of a photon is constant at an arbitrary cosmic time t , i.e., $a(t)\nu(t) = \text{const}$. See Eq. (9.60).

¹⁸⁸⁾See, p. 339, Eq. (7.1.32), [221].

$$\begin{aligned} \left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(S)} &= \frac{1}{2}(E(r_L \hat{n}, t_L) - E(0, t_0)) - \frac{1}{a(t_L)} \frac{\partial}{\partial x} \delta u_\gamma^{(S)}(x \hat{n}, t_L) \Big|_{x=\bar{r}(t_L)} \\ &\quad - \frac{1}{2} \int_{t_L}^{t_0} dt \left(\dot{A}(x \hat{n}, t) + \frac{\partial^2}{\partial r^2} \dot{B}(x \hat{n}, t) - \dot{E}(x \hat{n}, t) \right) \Big|_{x=\bar{r}(t)} + \frac{\delta T^{(i)}(r_L \hat{n}, t_L)}{\bar{T}(t_L)}, \end{aligned} \tag{12.251}$$

where we have used the notation for the partial derivative with respect to the cosmic time

$$\dot{A}(x \hat{n}, t) \equiv \frac{\partial}{\partial t} A(x \hat{n}, t), \quad \dot{E}(x \hat{n}, t) \equiv \frac{\partial}{\partial t} E(x \hat{n}, t), \text{ etc.} \tag{12.252}$$

To make the physics transparent, we can rewrite the integrand of the integral appearing in the second line of Eq. (12.251):¹⁸⁹⁾

$$\begin{aligned} \left(\dot{A}(x \hat{n}, t) + \frac{\partial^2}{\partial x^2} \dot{B}(x \hat{n}, t) - \dot{E}(x \hat{n}, t) \right) \Big|_{x=\bar{r}(t)} &= -\frac{d}{dt} \left(\left(a^2(t) \ddot{B}(x \hat{n}, t) + a(t) \dot{a}(t) \dot{B}(x \hat{n}, t) + a(t) \frac{\partial}{\partial x} \dot{B}(x \hat{n}, t) \right) \Big|_{x=\bar{r}(t)} \right) \\ &\quad + \frac{\partial}{\partial t} \left(a^2(t) \ddot{B}(x \hat{n}, t) + a(t) \dot{a}(t) \dot{B}(x \hat{n}, t) + A(x \hat{n}, t) - E(x \hat{n}, t) \right) \Big|_{x=\bar{r}(t)}. \end{aligned} \tag{12.253}$$

Substituting the result of Eq. (12.253) into Eq. (12.251), performing the straightforward time integration, and grouping the resulting terms according to their dependence on the cosmic time variables: time at the LSS t_L , time at the present epoch t_0 , and the time interval in between, i.e., integration between t_L and t_0 . They are named respectively as “early”, “late”, and “ISW”:¹⁹⁰⁾

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(S)} = \left(\frac{\Delta T(\hat{n})}{T_0}\right)_{\text{early}}^{(S)} + \left(\frac{\Delta T(\hat{n})}{T_0}\right)_{\text{late}}^{(S)} + \left(\frac{\Delta T(\hat{n})}{T_0}\right)_{\text{ISW}}^{(S)}. \tag{12.254}$$

The expressions of these terms and their physical meanings are discussed below.

The “early” term is given by:¹⁹¹⁾

$$\begin{aligned} \left(\frac{\Delta T(\hat{n})}{T_0}\right)_{\text{early}}^{(S)} &\equiv -\frac{1}{2} \left(a^2(t_L) \ddot{B}(r_L \hat{n}, t_L) + a(t_L) \dot{a}(t_L) \dot{B}(r_L \hat{n}, t_L) - E(r_L \hat{n}, t_L) \right) \\ &\quad + \frac{\delta T^{(i)}(r_L \hat{n}, t_L)}{\bar{T}(t_L)} - a(t_L) \frac{\partial}{\partial x} \left(\frac{1}{2} \dot{B}(x \hat{n}, t_L) + \frac{1}{a^2(t_L)} \delta u_\gamma^{(S)}(x \hat{n}, t_L) \right) \Big|_{x=r_L}. \end{aligned} \tag{12.255}$$

The name of the term is clear: All terms are functions of the LSS variables r_L and t_L . The second term of the second line, involving radial derivatives, comes from the Doppler effect. The first term of the second line, as remarked earlier, represents the intrinsic temperature fluctuation that occurs on LSS. This “early” term makes the largest contribution to multipole moment coefficients $\ell > 20$.

The “late” term is¹⁹²⁾

$$\begin{aligned} \left(\frac{\Delta T(\hat{n})}{T_0}\right)_{\text{late}}^{(S)} &\equiv \frac{1}{2} \left(a^2(t_0) \ddot{B}(0, t_0) + a(t_0) \dot{a}(t_0) \dot{B}(0, t_0) - E(0, t_0) \right) \\ &\quad + a(t_0) \frac{\partial}{\partial x} \left(\frac{1}{2} \dot{B}(x \hat{n}, t_0) + \frac{1}{a^2(t_0)} \delta u_\gamma^{(S)}(x \hat{n}, t_0) \right) \Big|_{x=0}. \end{aligned} \tag{12.256}$$

¹⁸⁹⁾The following identity has been used:

$$\begin{aligned} &\frac{\partial^2}{\partial x^2} \dot{B}(x \hat{n}, t) \Big|_{x=\bar{r}(t)} \\ &= -\frac{d}{dt} \left(\left(a^2(t) \ddot{B}(x \hat{n}, t) + a(t) \dot{a}(t) \dot{B}(x \hat{n}, t) + a(t) \frac{\partial}{\partial x} \dot{B}(x \hat{n}, t) \right) \Big|_{x=\bar{r}(t)} \right) \\ &\quad + \frac{\partial}{\partial x} \left(a^2(t) \ddot{B}(x \hat{n}, t) + a(t) \dot{a}(t) \dot{B}(x \hat{n}, t) \right) \Big|_{x=\bar{r}(t)}, \end{aligned}$$

which is given in [221] as the first expression given on p. 342.

¹⁹⁰⁾See, [221], Eq. (7.1.36).

The name of the term is also clear: All terms are functions of variables of the present epoch, $r_0 = 0$ and t_0 . The “late” term is similar to the “early” term, but as expected, without an intrinsic temperature fluctuation term. The first three terms on the right-handed side do not depend on the incoming photon direction \hat{n} . Hence it contributes to the monopole moment only. As will be

¹⁹¹⁾See, [221], Eq. (7.1.37).

¹⁹²⁾See, [221], Eq. (7.1.38).

demonstrated in the next subsection in the discussion of the Fourier decomposition of scalar perturbation functions, the last term depends linearly on \hat{n} . Therefore,

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)_{\text{ISW}}^{(S)} \equiv -\frac{1}{2} \int_{t_L}^{t_0} dt \left[\frac{\partial}{\partial t} \left(a^2(t) \ddot{B}(x\hat{n}, t) + a(t) \dot{a}(t) \dot{B}(x\hat{n}, t) + A(x\hat{n}, t) - E(x\hat{n}, t) \right) \right]_{x=\bar{r}(t)}, \quad (12.257)$$

which describes the *integrated Sachs–Wolfe effect* (ISW). It is a consequence of time-dependent fluctuation of gravitational fields integrated over the time from LSS t_L to the present t_0 . For time independent gravitational potentials that lie in the photon path from LSS to the present, the ISW effect will vanish. Hence the gravitational effect of the cold dark matter does not contribute to ISW. The ISW effect contributes to relatively small multipole coefficients, $\ell < 20$.

12.10.2.2 Fourier decomposition of scalar perturbation functions

Let us now work out the various terms that enter in “early” expression, etc., in terms of Fourier amplitudes of relevant perturbation functions. This is to rewrite the perturbation functions, $E(\mathbf{x}, t)$, etc. in their Fourier decomposition. See Section 12.8.2 for more details. We write

$$B(\mathbf{x}, t) \equiv \int d^3q \alpha(\mathbf{q}) \exp(i\mathbf{q} \cdot \mathbf{x}) B_q(t), \quad \text{etc.}, \quad (12.258)$$

where a single mode is assumed to dominate the scalar temperature fluctuation. To simplify the notation we write $B_q(t)$ mean a function of both q and t . So we use the notation $B(q, t)$ and $B_q(t)$ interchangeably. The stochastic variable $\alpha(\mathbf{q})$ is normalized as

$$\langle \alpha(\mathbf{q}) \alpha^*(\mathbf{q}') \rangle = \delta(\mathbf{q} - \mathbf{q}'). \quad (12.259)$$

We can rewrite the “early” terms, etc.,

$$\left(\frac{\Delta T(\mathbf{n})}{T_0}\right)_{\text{early}}^{(S)} = \int d^3q \alpha(\mathbf{q}) \exp(i\mathbf{q} \cdot \hat{n} t_L) [F_T(q, t_L) + i\hat{q} \cdot \hat{n} G_T(q, t_L)], \quad (12.260)$$

where¹⁹⁴⁾

$$F_T(q, t) \equiv F_{T1}(q, t) + \frac{\delta T_q^{(i)}(t_L)}{\bar{T}(t_L)},$$

¹⁹³⁾See, [221], Eq. (7.1.39).

¹⁹⁴⁾Here F_T and G_T are F and G of [221], Eqs. (7.1.44) and (7.1.45). We modify the notation slightly to avoid possible confusion with the metric perturbation functions F and G_j defined in Eq. (12.39). The metric perturbation function F will appear in our discussion of gauge invariance below.

this term contributes only to the dipole moment $\ell = 1$. Hence the “late” term can be ignored for $\ell > 1$.

The “ISW” term is¹⁹³⁾

$$F_{T1}(q, t) \equiv -\frac{1}{2} \left[a^2(t) \ddot{B}_q(t) + a(t) \dot{a}(t) \dot{B}_q(t) - E_q(t) \right],$$

$$G_T(q, t) = -q \left[\frac{a(t)}{2} \dot{B}_q(t) + \frac{1}{a(t)} \delta u_{\gamma q}^{(S)}(t) \right]. \quad (12.261)$$

The “late” term becomes

$$\left(\frac{\Delta T(\mathbf{n})}{T_0}\right)_{\text{late}}^{(S)} = \int d^3q \alpha(\mathbf{q}) [F_{T1}(q, t_0) + i\hat{q} \cdot \hat{n} G_T(q, t_0)], \quad (12.262)$$

which shows that the “late” term contributes to multipoles only up to $\ell = 1$.

The “ISW” term is

$$\left(\frac{\Delta T(\mathbf{n})}{T_0}\right)_{\text{ISW}}^{(S)} = \int_{t_L}^{t_0} dt \int d^3q \alpha(\mathbf{q}) \exp(i\mathbf{q} \cdot \hat{n} \bar{r}(t)) \times \frac{\partial}{\partial t} \left(F_{T1}(q, t) - \frac{A_q(t)}{2} \right). \quad (12.263)$$

12.10.2.3 Gauge invariance of scalar perturbations

For the separation of the various terms in Eq. (12.254)–(12.257) and the labeling of the Doppler term in Eq. (12.255) to be meaningful, they have to be individually gauge invariant for the class of gauges under consideration, $g_{0j} = 0$.¹⁹⁵⁾ Let us rewrite the relevant perturbation functions in gauge invariant form discussed in §12.6.3. The gauge invariant scalar perturbation functions are given in Eq. (12.84). The relevant ones are given below, suppressing the functional dependence:

$$\hat{E} \equiv E + 2\delta\dot{u}^{(S)}, \quad \hat{A} \equiv A + 2H\delta u^{(S)},$$

$$\hat{B}_F \equiv F - \frac{1}{2}a\dot{B} - \frac{1}{a}\delta u^{(S)},$$

$$\delta\hat{T} \equiv \delta T + \dot{\bar{T}}\delta u^{(S)} = \delta T - H\bar{T}\delta u^{(S)}, \quad (12.264)$$

where \hat{E} , \hat{A} , \hat{B}_F , and $\delta\hat{T}$ are gauge invariant under arbitrary gauge transformations. In the last line of the above equations, δT is a scalar and $\bar{T} \sim 1/a$, so $\dot{\bar{T}} = -H\bar{T}$. Let us recall that F is the scalar function appears in the perturbation of g_{0j} .

There are basically three groups of terms entering in Eqs. (12.255)–(12.257) which we rewrite in the following

¹⁹⁵⁾We can prove the gauge invariance directly by making gauge transformation on the various terms to show that they are invariant. This is the approach taken in [221], p. 341.

in terms of gauge invariant perturbation functions. Their functional arguments are suppressed. First, the group of terms entering in Eqs. (12.255) and (12.256),

$$\begin{aligned} \frac{1}{2}\dot{\hat{B}} + \frac{1}{a^2}\delta u_\gamma^{(S)} &= \frac{1}{a} \left[\frac{a}{2}\dot{\hat{B}} + \frac{1}{a}\delta u_\gamma^{(S)} \right] \\ &= -\frac{1}{a}\hat{B}_F + \frac{1}{a}F; \end{aligned} \quad (12.265)$$

the group entering in Eq. (12.255)

$$\begin{aligned} a^2\ddot{\hat{B}} + a\dot{\hat{B}} - E - 2\frac{\delta T^{(i)}}{\bar{T}} \\ &= a^2\partial_t \left[\frac{2}{a} \left(-\hat{B}_F + F - \frac{1}{a}\delta u_\gamma^{(S)} \right) \right] \\ &\quad + 2\dot{a} \left[-\hat{B}_F + F - \frac{1}{a}\delta u_\gamma^{(S)} \right] \\ &\quad - \left[\hat{E} - 2\delta\dot{u}_\gamma^{(S)} \right] - 2 \left[\frac{\delta\hat{T}^{(i)}}{\bar{T}} + H\delta u_\gamma^{(S)} \right] \\ &= -a^2\partial_t \left(\frac{2}{a}\hat{B}_F \right) - 2\dot{a}\hat{B}_F - \hat{E} - \frac{\delta\hat{T}^{(i)}}{\bar{T}} + 2a\dot{F}; \end{aligned} \quad (12.266)$$

and the group entering in Eq. (12.257)

$$\begin{aligned} a^2\ddot{\hat{B}} + a\dot{\hat{B}} + A - E \\ &= a^2\partial_t \left[\frac{2}{a} \left(-\hat{B}_F + F - \frac{1}{a}\delta u_\gamma^{(S)} \right) \right] \\ &\quad + 2\dot{a} \left[-\hat{B}_F + F - \frac{1}{a}\delta u_\gamma^{(S)} \right] \\ &\quad + \left[\hat{A} - 2H\delta u_\gamma^{(S)} \right] - \left[\hat{E} - 2\delta\dot{u}_\gamma^{(S)} \right] \\ &= -a^2\partial_t \left(\frac{2}{a}\hat{B}_F \right) - 2\dot{a}\hat{B}_F + \hat{A} - \hat{E} + 2a\dot{F}. \end{aligned} \quad (12.267)$$

These groups of terms, Eqs. (12.265)–(12.267) are gauge invariant in the gauge $g_{0j} = 0$ which maintains $F = 0$. This makes the “early” term Eq. (12.255) and the “ISW” term Eq. (12.257) gauge invariant. But the “late” term (12.256) is not gauge invariant. As shown

in Eq. (12.266), since the first term in the right-handed side of the “late” term does not have an intrinsic temperature fluctuation, it is not gauge invariant. However, as we have seen in Eq. (12.262), the “late” term contributes to multipole moments only up to the dipole $\ell = 1$, it will be ignored.

Let us take a brief look of the gauge transformation under the present requirement. As shown in Eq. (12.76), a general gauge transformation due to an infinitesimal coordinate shift $x_\mu \rightarrow x_\mu + \epsilon_\mu$ in the form of Eq. (12.66)

$$\epsilon_\mu \equiv (\epsilon_0, \partial_j \epsilon^{(S)} + \epsilon_j^{(V)}). \quad (12.268)$$

This introduces shifts in all scalar perturbation functions and some vector perturbation functions as well, as discussed in §12.6.3, listed in Eqs. (12.72), (12.77)–(12.79), (12.81), and (12.82). To maintain $g_{0j} = 0$, it requires, in particular, $\Delta F = 0$, which leads to a relation among the scalar parts of the gauge transformation parameters given in Eq. (12.78)

$$\epsilon_0 + a^2\partial_t \left(\frac{\epsilon^{(S)}}{a^2} \right) = 0. \quad (12.269)$$

Then all shifts of the perturbation functions ΔE , ΔA , etc. can be worked out in terms of the parameter, say ϵ_0 . One can verify directly that the “early” and “ISW” terms, Eqs. (12.255) and (12.257), are gauge invariant.¹⁹⁶⁾

12.10.2.4 Scalar mode temperature multipole coefficients and angular power spectra

We will now work out the coefficients of the temperature multipoles moments for perturbations in the scalar mode and the corresponding power spectra. For large multipoles of $\ell > 20$ we can ignore the “ISW” term. So we will focus on the “early” term given in Eq. (12.260) and write it in a multipole expansion

$$\begin{aligned} \left(\frac{\Delta T(\hat{n})}{T_0} \right)^{(S)} &\approx \left(\frac{\Delta T(\hat{n})}{T_0} \right)^{(S)}_{\text{early}} = \int d^3q \alpha(\mathbf{q}) \left[F_T(q, t_L) + G_T(q, t_L) \frac{\partial}{\partial x} \right] \exp(i\hat{q} \cdot \hat{n}x) |_{x=q r_L} \\ &\equiv \frac{1}{T_0} \sum_{\ell m} a_{T, \ell m}^{(S)} Y_{\ell m}(\hat{n}), \end{aligned} \quad (12.270)$$

where $F_T(q, t_L)$ and $G_T(q, t_L)$ are given in Eq. (12.261) and $a_{T, \ell m}^{(S)}$ are coefficients of the multipole moment expansion of scalar temperature fluctuations. To obtain these multipole moment coefficients we follow the de-

velopment presented in Section 12.4. We make a multipole expansion of the exponential function as given in Eq. (12.19), i.e.,

$$\exp(i\hat{q} \cdot \hat{n}x) = 4\pi \sum_{\ell m} i^\ell j_\ell(x) Y_{\ell m}(\hat{n}) Y_{\ell m}^*(\hat{q}). \quad (12.271)$$

The scalar temperature fluctuation multipole coefficient

¹⁹⁶⁾As already noted earlier, this approach of direct verification of gauge invariance is employed in [221], p. 341.

can be readily calculated,

$$\begin{aligned} \alpha_{T,\ell m}^{(S)} &= 4\pi i \ell T_0 \int d^3q \alpha(\mathbf{q}) (j_\ell(qr_L)F - T(q, t_L) \\ &\quad + j'_\ell(qr_L)G_T(q, t_L)) Y_{\ell m}(\hat{q}), \\ j'_\ell(qr_L) &= \frac{\partial}{\partial x} j_\ell(x) \Big|_{x=qr_L}. \end{aligned} \tag{12.272}$$

The angular power spectra is defined in Eq. (12.13)

$$\begin{aligned} C_{TT,\ell}^{(S)} &= \frac{1}{2\ell + 1} \sum_m \langle a_{\ell m} a_{\ell m}^* \rangle \\ &= (4\pi T_0)^2 \int_0^\infty q^2 dq \left| j_\ell(qr_L)F_T(q, t_L) \right. \\ &\quad \left. + j'_\ell(qr_L)G_T(q, t_L) \right|^2. \end{aligned} \tag{12.273}$$

Here we are interested in large order multipole moments. The expression of the angular power spectra Eq. (12.273) can be simplified using the large order approximation of the spherical Bessel function of the first kind discussed in §12.5.3. We note that for $\ell \gg 1$ $j_\ell(z)$ has substantial values only for $z \gtrsim \ell$ and for $z > \ell$ it oscillates rapidly. This behavior of the $\ell - z$ relationship can be seen clearly in Fig. 12.6, and it will be advantageous to factor ℓ out as much as possible. We rewrite the large order asymptotic expression Eqs. (12.27) and (12.28), by changing the variable to make the argument of the trigonometric function looks simpler:¹⁹⁷⁾

$$qr_L \equiv \ell\beta, \quad \text{for } \beta \geq 1, \tag{12.274}$$

The comoving wave number q can be scaled out by the order of the multipole, $q = (\ell/r_L)\beta$, for a given order of the multipole moment.

$$\begin{aligned} j_\ell(qr_L) &\equiv j_\ell(\ell\beta) \approx \frac{1}{\ell\sqrt{\beta(\beta^2 - 1)^{1/4}}} \cos\left(\ell\sqrt{\beta^2 - 1} - \ell\cos^{-1}\left(\frac{1}{\beta}\right) - \frac{\pi}{4}\right), \\ j'_\ell(qr_L) &\equiv j'_\ell(\ell\beta) \approx -\frac{1}{\ell\sqrt{\beta(\beta^2 - 1)^{1/4}}} \cdot \frac{\sqrt{\beta^2 - 1}}{\beta} \sin\left(\ell\sqrt{\beta^2 - 1} - \ell\cos^{-1}\left(\frac{1}{\beta}\right) - \frac{\pi}{4}\right). \end{aligned} \tag{12.275}$$

The angular power spectra become

$$\begin{aligned} C_{TT,\ell}^{(S)} &\approx (4\pi T_0)^2 \frac{\ell}{r_L^3} \int_1^\infty \frac{\beta d\beta}{\sqrt{\beta^2 - 1}} \left| F_T\left(\frac{\ell\beta}{r_L}, t_L\right) \cos\left(\ell\sqrt{\beta^2 - 1} - \ell\cos^{-1}\left(\frac{1}{\beta}\right) - \frac{\pi}{4}\right) \right. \\ &\quad \left. - \frac{\sqrt{\beta^2 - 1}}{\beta} G_T\left(\frac{\ell\beta}{r_L}, t_L\right) \sin\left(\ell\sqrt{\beta^2 - 1} - \ell\cos^{-1}\left(\frac{1}{\beta}\right) - \frac{\pi}{4}\right) \right|^2. \end{aligned} \tag{12.276}$$

The trigonometric functions entering in Eq. (12.276) include the following terms: $\cos^2(\eta) = \frac{1}{2}(1 + \cos(2\eta))$, $\sin^2(\eta) = \frac{1}{2}(1 - \cos(2\eta))$, and $\cos(\eta)\sin(\eta) = \frac{1}{2}\sin(2\eta)$, where $\eta \equiv \ell\sqrt{\beta^2 - 1} - \ell\cos^{-1}(1/\beta) - \pi/4$. For large ℓ they are rapidly oscillating functions of β . Assuming that $F_T(\ell\beta/r_L, t_L)$ and $G_T(\ell\beta/r_L, t_L)$ are smooth functions of β , we have, under the β integration, $\cos(2\eta)$ and $\sin(2\eta)$ are effectively zero, while $\cos^2(\eta)$ and $\sin^2(\eta)$ effectively $1/2$.¹⁹⁸⁾ Multiplied by the factor $\ell(\ell + 1) \approx \ell^2$

as usually done, the angular power spectra has a much simpler expression,

$$\begin{aligned} \ell(\ell + 1)C_{TT,\ell}^{(S)} &\approx 8\pi^2 T_0^2 \frac{\ell^3}{r_L^3} \int_1^\infty \frac{\beta d\beta}{\sqrt{\beta^2 - 1}} \left[F_T^2\left(\frac{\ell\beta}{r_L}, t_L\right) \right. \\ &\quad \left. + \frac{\beta^2 - 1}{\beta^2} G_T^2\left(\frac{\ell\beta}{r_L}, t_L\right) \right]. \end{aligned} \tag{12.277}$$

From the above functional form and the physics input in the formulation we can draw several conclusions even without knowing the explicit expressions of the form factors F_T and G_T .

- For multipole moments $\ell > 20$, the power spectra of the temperature anisotropy depend largely on the physics at the LSS.
- From the approximations adopted so far, we see that the form factors depend on the total matter density and the baryonic matter density at LSS. These density on LSS can be related to their present values, and therefore the density parameter Ω_{Bh} and Ω_{Mh} . The dark matter fractional density at the present epoch is $\Omega = \Omega_M - \Omega_B$.

¹⁹⁷⁾Stages of variable changes are employed in the calculation in [221]. Here we follow its second stage of variable change, see pp 347–348, [221].

¹⁹⁸⁾These results can be argued as follows. In a measurement, the value of a variable, such as q or β in the present case, can not be given in absolute precision. Each point can be regarded as being associated with a small interval. Hence the value of the function of the variable should be taken as the average of the values in the small interval of the variable. For the case of a rapid oscillating function, there would be several oscillations in the averaging interval. So the sine and cosine functions are averaged to 0, and their squares to $1/2$. This argument is essential in the formulation of long baseline neutrino oscillations. We have to be careful in the present case, because η is small when β is very close to 1. Then a reasonably valued percentage interval in $\beta \approx 1$ might not cover several oscillation cycles. So a small but finite interval of β has to be assigned there in making the average.

- For large ℓ the scaled power spectra $\ell(\ell+1)C_{\text{TT},\ell}^{(S)}$ depends on the ratio ℓ/r_L , not on ℓ and r_L separately, where r_L depends on H_0 as well as all density parameters, including the space curvature κ . Hence the temperature anisotropy can not be used to determine the cosmic parameters H_0 , Ω_A , and Ω_κ separately.
- In case of a flat universe, i.e., $\kappa = 0$, we can obtain $\Omega_A = 1 - \Omega_M$. In the accuracy of the present treatment, the radiation fractional density, $\Omega < 10^{-3}$ can be neglected.

We can simplify the expressions of the form factors F_T and G_T by the assumption that at LSS the gravitational potential ψ , given by Eq. (12.107) which has the momentum space expression

$$\psi_q = \frac{1}{2} \left(\dot{A} - q^2 \dot{B} \right), \tag{12.278}$$

is dominated by the cold dark matter.¹⁹⁹⁾ Then we can relate the form factors to three potential perturbation functions: the gravitational potential ψ , the photon-baryon velocity potential $\delta u_\gamma^{(S)}$, and the photon fractional density perturbation δ_γ . The key point is that LSS is dominated by matter and with the assumption of cold dark matter domination, we can show that the perturbation function A is time independent. Then we can express \dot{B} and \ddot{B} in terms of ψ .

From Eq. (12.109) we have

$$q^2 A_q = 8\pi G_N a^2 \bar{\rho}_D \delta_{Dq} - 2a^2 H \psi_q. \tag{12.279}$$

The time dependence of the various quantities are as follows: The following quantities are obvious $H \sim t^{-1}$,

$a^2 \sim t^{4/3}$ for matter domination, and $\bar{\rho}_D \sim a^{-3} \sim t^{-2}$. From Eq. (12.215) or (12.225), we have $\psi_q \sim t/a^2 \sim t^{-1/3}$ and $\delta_{Dq} \sim t^2/a^2 \sim t^{2/3}$. Hence the two terms on the right-handed side of the above equation are separately time-independent. So A_q is time independent. Then Eq. (12.278) gives

$$\dot{B}_q |_{\text{DMD}} = -\frac{2\psi_q}{q^2}, \tag{12.280}$$

where DMD stands for dark matter domination. Knowing the time dependence of $\psi_q \sim t^{-1/3}$ we can take the time derivative of the above expression to obtain

$$\ddot{B}_q |_{\text{DMD}} = \frac{2\psi_q}{3q^2 t}. \tag{12.281}$$

For the intrinsic temperature fluctuation on LSS we related it to the photon density perturbation on LSS²⁰⁰⁾

$$\frac{\delta T^{(i)}(r_L \hat{n}, t_L)}{\bar{T}(t_L)} = \frac{1}{3} \delta_\gamma(t_L). \tag{12.282}$$

In the synchronous gauge the time component of the metric perturbation function is set to vanish, $E = 0$. Then, from Eq. (12.261) we have a simplified expressions of the two form factors:²⁰¹⁾

$$F_T(q, t_L) = \frac{1}{3} \delta_{\gamma q}(t_L) + \frac{1}{3} \frac{a^2(t_L)}{q^2 t_L} \psi_q(t_L),$$

$$G_T(q, t_L) = -\frac{q}{a(t_L)} \delta u_{\gamma q}^{(S)}(t_L) + \frac{a(t_L)}{q} \psi_q(t_L), \tag{12.283}$$

where the perturbation functions are given by their interpolating form $\delta_{\gamma q}^{(ILS)}(t_L)$, $\psi_q^{(ILS)}(t_L)$, and $\delta u_{\gamma q}^{(S)(ILS)}(t_L)$ as given in Eqs. (12.225). We list the perturbation functions below again:

$$\psi_q(t_L) = -\frac{3\mathcal{R}_q^{(o)} q^2 t_L \mathcal{T}(\hat{\kappa})}{5a^2(t_L)},$$

$$\delta_{\gamma q}(t_L) = \frac{3\mathcal{R}_q^{(o)}}{5} \left[(1 + 3B_{RL}) \mathcal{T}(\hat{\kappa}) - \frac{1}{(1 + R_{BL})^{1/4}} e^{-\hat{r}_q(t_L)} \mathcal{S}(\hat{\kappa}) \cos(\Theta_q(t_L) + \Delta(\hat{\kappa})) \right],$$

$$\delta u_{\gamma q}^{(S)}(t_L) = -\frac{3\mathcal{R}_q^{(o)}}{5} \left[t_L \mathcal{T}(\hat{\kappa}) - \frac{a(t_L)}{\sqrt{3}q(1 + R_{BL})^{3/4}} e^{-\hat{r}_q(t_L)} \mathcal{S}(\hat{\kappa}) \sin(\Theta_q(t_L) + \Delta(\hat{\kappa})) \right], \tag{12.284}$$

where $\hat{\kappa}$, $\hat{r}_g(t)$, and $\Theta_g(t)$ are given in Eq. (12.224), and R_{BL} is R_B are evaluated at the LSS time t_L . The transfer functions $\mathcal{T}(\hat{\kappa})$, $\mathcal{S}(\hat{\kappa})$, and $\Delta(\hat{\kappa})$, are given in Eq. (12.226).

We now address two complications that have been ignored earlier in our approximation. One is the instant transparency approximation at the LSS, and the other the re-ionization effect due to the appearance of the first stars.

12.10.2.4.1 Average over LSS

To take into account of the feature that the photon transparency happens over a finite period of time instead of

¹⁹⁹⁾See Footnote 173) appearing in the paragraph above Eq. (12.201) for a comment on the assumption that the dark matter dominates the gravitational potential.

²⁰⁰⁾This can be demonstrated as follows: From $\rho_\gamma \sim T^4$, $\delta\rho_\gamma[\sim 4T^4(\delta T/T)] = 4\rho_\gamma(\delta T/T)$, and also $\delta\rho_\gamma = (\bar{\rho}_\gamma + \bar{P}_\gamma)\delta_\gamma = (4/3)\bar{\rho}_\gamma\delta_\gamma$, we obtain Eq. (12.282).

²⁰¹⁾These are Eqs. (7.2.18) and (7.2.19), p. 349 of [221].

instantaneously, we represent LSS by a peaked function of the cosmic time which can be parameterized as a Gaussian distribution²⁰²⁾

$$P_L(t) \equiv \frac{1}{\sqrt{2\pi\sigma_t}} e^{-(t-t_L)^2/(2\sigma_t^2)}, \quad (12.285)$$

where $2\sqrt{2 \cdot \ln(2)}\sigma_t$ is the full-width at half maximum. The integration of $P_L(t)$ over time from $-\infty$ to ∞ is unity. For the zero width Gaussian, i.e., in the limit $\sigma_t \rightarrow 0$, $P_L(t)$ becomes the delta function $\delta(t-t_L)$ which recovers the case of instantaneous transparency. We can replace the trigonometric functions by their finite width average,

$$\begin{aligned} \cos(\Theta_q(t_L) + \Delta(\hat{\kappa})) &\rightarrow \int_{-\infty}^{\infty} dt P_L(t) \cos(\Theta_q(t) + \Delta(\hat{\kappa})), \\ \sin(\Theta_q(t_L) + \Delta(\hat{\kappa})) &\rightarrow \int_{-\infty}^{\infty} dt P_L(t) \sin(\Theta_q(t) + \Delta(\hat{\kappa})). \end{aligned} \quad (12.286)$$

The original trigonometric functions are recovered in the limit of vanishing σ_t . The integrals in the above expressions can be performed if a further approximation is made by replacing the argument of the trigonometric function by the first two terms in its Taylor series expansion around the LSS time t_L :

$$\begin{aligned} \Theta_q(t) &= \Theta_q(t_L) + \frac{d}{dt}\Theta(t)|_{t_L} \cdot (t-t_L) + \mathcal{O}((t-t_L)^2) \\ &= \int_0^{t_L} \omega_q(t) dt + \omega_q(t_L)(t-t_L) + \mathcal{O}((t-t_L)^2), \end{aligned} \quad (12.287)$$

where $\Theta_q(t)$ is given in Eq. (12.224) and

$$\omega_q(t) \equiv \frac{q}{\sqrt{3} a \sqrt{1+R_B}}. \quad (12.288)$$

We can do the integrals in Eq. (12.286) by keeping the first two terms in the above Taylor series expansion. The first expression of Eq. (12.286) gives

$$\begin{aligned} &\cos(\Theta_q(t) + \Delta(\hat{\kappa})) \\ &\rightarrow \int_{-\infty}^{\infty} dt P_L(t) \cos(\Theta_q(t_L) + \Delta(\hat{\kappa}) + \omega_q(t_L)(t-t_L)) \\ &= \cos(\Theta_q(t_L) + \Delta(\hat{\kappa})) \int_{-\infty}^{\infty} dt P_L(t) \cos(\omega_q(t_L)(t-t_L)) \\ &\quad - \sin(\Theta_q(t_L) + \Delta(\hat{\kappa})) \int_{-\infty}^{\infty} dt P_L(t) \sin(\omega_q(t_L)(t-t_L)). \end{aligned} \quad (12.289)$$

The last line above proportional to the sine functions vanishes, because $P_L(t)$ as given in Eq. (12.285) is even

²⁰²⁾The argument follows that of [221], pp 350–352. See Eq. (7.2.25), [221].

in $t-t_L$ while $\sin(\omega_q(t_L)(t-t_L))$ is odd, so the integral vanishes. The remaining integral gives the result,²⁰³⁾

$$\int_{-\infty}^{\infty} dt P_L(t) \cos(\omega_q(t_L)(t-t_L)) = e^{-\omega_q^2(t_L)\sigma_t^2/2}. \quad (12.290)$$

The second expression in Eq. (12.286) can be treated similarly. So we have

$$\begin{aligned} \cos(\Theta_q(t_L) + \Delta(\hat{\kappa})) &\rightarrow \cos(\Theta_q(t_L) + \Delta(\hat{\kappa})) e^{-\omega_q^2(t_L)\sigma_t^2/2}, \\ \sin(\Theta_q(t_L) + \Delta(\hat{\kappa})) &\rightarrow \sin(\Theta_q(t_L) + \Delta(\hat{\kappa})) e^{-\omega_q^2(t_L)\sigma_t^2/2}. \end{aligned} \quad (12.291)$$

The average is to introduce a damping effect represented by $\exp(-\omega_q^2(t_L)\sigma_t^2/2)$.

Substituting Eq. (12.291) into the perturbation functions listed in Eq. (12.284) which enter the form factors Eq. (12.283), we see two damping factors appearing in the density perturbations $\delta_{\gamma q}(t_L)$ and $\delta u_{\gamma q}^{(S)}(t_L)$: $\hat{I}_q(t_L)$ and $\exp(-\omega_q^2(t_L)\sigma_t^2/2)$. Their explicit expressions can be found in Eqs. (12.224) and (12.288), which show that they are proportional to the square of the physical wave number on LSS, i.e., q^2/a_L^2 , and therefore inversely proportional to the physical wavelengths as defined in Eq. (12.113), where $a_L \equiv a(t_L)$. It is therefore convenient to define a damping length for each of the damping factor

$$\begin{aligned} \hat{I}_q(t_L) &\equiv \frac{q^2}{a_L^2} d_{\text{Silk}}^2, \\ \frac{1}{2}\omega_q^2(t_L)\sigma_t^2 &\equiv \frac{q^2}{a_L^2} d_{\text{Landau}}^2. \end{aligned} \quad (12.292)$$

From Eqs. (12.224) and (12.288), we have

$$\begin{aligned} d_{\text{Silk}}^2 &= \frac{a_L^2}{6} \int_0^{t_L} \frac{t_\gamma}{a^2(1+R_B)} \left(\frac{16}{15} + \frac{R_B^2}{1+R_B} \right) dt, \\ d_{\text{Landau}}^2 &= \frac{\sigma_t^2}{6[1+R_B(t_L)]}. \end{aligned} \quad (12.293)$$

²⁰³⁾The proof is the following:

$$\begin{aligned} &\int_{-\infty}^{\infty} dt P_L(t) \cos(\omega_q(t_L)(t-t_L)) \\ &= \frac{1}{2\sqrt{2\pi}\sigma_t} e^{-\omega_q^2(t_L)\sigma_t^2/2} \int_{-\infty}^{\infty} dt \left[e^{-(t-t_L-i\omega_q(t_L)\sigma_t)^2/(2\sigma_t^2)} \right. \\ &\quad \left. + e^{-(t-t_L+i\omega_q(t_L)\sigma_t)^2/(2\sigma_t^2)} \right] \\ &= e^{-\omega_q^2(t_L)\sigma_t^2/2}. \end{aligned}$$

The final result is obtained by dropping the imaginary parts in the exponents of the integrand by a change of variable of integration. A mathematically rigorous proof of this procedure can be made using the Cauchy theorem. The Cauchy theorem says that the integration of an analytic function over a closed path vanishes. In the present case the close path is the boundary of an infinite rectangle, including the real axis, plus a suitable straight line in the complex plan parallel to the real axis, and plus two line segments parallel to the imaginary axis but locating at $t \rightarrow \pm\infty$, where t is the real variable of the complex plan.

The overall damping length d_D is given by

$$d_D^2 \equiv d_{\text{Silk}}^2 + d_{\text{Landau}}^2. \quad (12.294)$$

We also express two other dimensionless quantities which enter the multipole coefficient, $\hat{\kappa}$ and $\Theta_q(t_L)$, in terms of products of the physical wave number on LSS q/a_L and suitably defined lengths. First, the argument of transfer functions,²⁰⁴⁾ we write

$$\begin{aligned} \hat{\kappa} &= \sqrt{2} \frac{q}{q_{EQ}} \equiv \frac{q}{a_L} d_T, \\ d_T &= \frac{\sqrt{2} a_L}{a_{EQ} H_{EQ}} = \frac{\sqrt{\Omega_R}}{(1+z_L)\Omega_M H_0}. \end{aligned} \quad (12.295)$$

Second, the argument of the trigonometric functions in Eq. (12.284) is rewritten as

$$\begin{aligned} \Theta_q(t_L) &= \frac{q}{\sqrt{3}} \int_0^{t_L} \frac{dt}{\sqrt{1+R_B}} \equiv \frac{q}{a_L} d_H, \\ d_H &= \frac{a_L}{\sqrt{3}} \int_0^{t_L} \frac{dt}{a\sqrt{1+R_B}}, \end{aligned} \quad (12.296)$$

where d_H is called the *acoustic horizon distance*. It turned out that the integral above has an analytic ex-

pression by rewriting the time integration in the integration in R_B .²⁰⁵⁾

$$dt = \frac{R_B dR_B}{H_0 \sqrt{\Omega_M R_{R0}^3} \sqrt{R_B + R_{BEQ}}}, \quad (12.297)$$

where $R_{B0} \equiv R_B(t_0) = 3\Omega_B/(4\Omega_\gamma)$ and $R_{BEQ} \equiv R_B(t_{EQ}) = (\Omega_R/\Omega_M)R_{B0}$. The integral in Eq. (12.296) can be performed, when we rewrite $a/a_0 = R_B/R_{R0}$. We obtain²⁰⁶⁾

$$\begin{aligned} d_H &= \frac{2}{\sqrt{3} R_{B0} \Omega_M H_0 (1+z_L)} \\ &\times \ln \left(\frac{\sqrt{1+R_{BL}} + \sqrt{R_{BEQ} + R_{BL}}}{1+R_{BEQ}} \right), \end{aligned} \quad (12.298)$$

where $R_{BL} \equiv R_B(t_L)$. We note that R_{B0} , R_{BEQ} , and R_{BL} are related by their redshifts: $R_{B0} = (1+z_{EQ})R_{BEQ} = (1+z_L)R_{BL}$.

We can now write the two form factors given in Eq. (12.283), using Eqs. (12.284), (12.291), (12.294), (12.295), (12.296), and (12.298)²⁰⁷⁾

$$\begin{aligned} F_T(q, t_L) &= \frac{\mathcal{R}_q^{(o)}}{5} \left[R_{BL} \mathcal{T} \left(\frac{q}{a_L} d_T \right) - \frac{e^{-(q^2/a_L^2)d_D^2}}{(1+R_{BL})^{1/4}} S \left(\frac{q}{a_L} d_T \right) \cos \left(\frac{q}{a_L} d_H + \Delta \left(\frac{q}{a_L} d_T \right) \right) \right], \\ G_T(q, t_L) &= -\frac{\sqrt{3}\mathcal{R}_q^{(o)}}{5} \frac{e^{-(q^2/a_L^2)d_D^2}}{(1+R_{BL})^{3/4}} S \left(\frac{q}{a_L} d_T \right) \sin \left(\frac{q}{a_L} d_H + \Delta \left(\frac{q}{a_L} d_T \right) \right). \end{aligned} \quad (12.299)$$

²⁰⁴⁾We have used the following identities: $a_{EQ}/a_0 = \Omega_R/\Omega_M$ and $H_{EQ} = \sqrt{2}H_0(a_0/a_{EQ})^2$.

²⁰⁵⁾This can be derived as follows: We write

$$R_B = \frac{3\bar{\rho}_B}{4\bar{\rho}_\gamma} = \frac{3\Omega_B}{4\Omega_\gamma} \frac{a}{a_0},$$

which allows us to write

$$dR_B = \frac{3\Omega_B}{4\Omega_\gamma} \frac{\dot{a}}{a_0} dt = R_B H dt.$$

Up to the time of the LSS, to a very good approximation the Hubble expansion rate can be approximated by matter and radiation densities. Since the dark energy is very small in those epoches and the curvature has a negligible contribution, we write, for $\rho_0 = \rho_c$,

$$\begin{aligned} \frac{H}{H_0} &= \sqrt{\frac{\bar{\rho}_M + \bar{\rho}_R}{\bar{\rho}_0}} = \sqrt{\Omega_M \left(\frac{a_0}{a} \right)^3 + \Omega_R \left(\frac{a_0}{a} \right)^4} \\ &= \sqrt{\Omega_M \left(\frac{3\bar{\rho}_{B0}}{4\bar{\rho}_{\gamma 0}} \frac{4\bar{\rho}_\gamma}{3\bar{\rho}_B} \right)^3 + \Omega_R \left(\frac{3\bar{\rho}_{B0}}{4\bar{\rho}_{\gamma 0}} \frac{4\bar{\rho}_\gamma}{3\bar{\rho}_B} \right)^4} \\ &= \sqrt{\Omega_M \left(\frac{R_{B0}}{R_B} \right)^3 + \Omega_R \left(\frac{R_{B0}}{R_B} \right)^4} \\ &= \frac{\sqrt{\Omega_M R_{B0}^3}}{R_B^2} \sqrt{R_B + R_{BEQ}}, \end{aligned}$$

where $R_{BEQ} = (\Omega_R/\Omega_M)R_{B0}$ and $R_{R0} = 3\Omega_B/(4\Omega_\gamma)$. Combining the results, we obtain Eq. (12.297).

12.10.2.4.2 Reionization by first stars

Let us first describe briefly the reionization effect.²⁰⁸⁾

Reionization is an epoch of the universe during which the first generation of stars and galaxies are formed at the beginning of the end of the cosmic dark ages. It took place at the cosmic age of around 400 million years, corresponding to the redshift of the order $z_{Re} \approx 10$. It is one of the critical phase changes in the evolution of the universe. Prior to this phase change, the universe was dark, permeated with a dense fog of primordial neutral hydrogen and helium gas. Briefly, the prior states of the universe were the followings: For the first 370 000 years, the universe was filled with a hot ionized gas. As the universe expanded, the thermal energy of the big bang, in the form of electromagnetic radiation of a black body dis-

²⁰⁶⁾This is Eq. (7.2.39) on p. 353, [221], but we rewrite some of the terms, i.e., $R_{BL} = R_{B0}/(1+z_L)$.

²⁰⁷⁾These are Eqs. (7.2.36) and (7.2.37), p. 352 in [221].

²⁰⁸⁾The 2015 Planck data on CMB polarization have pushed the epoch of the first star by 100 million years from 450 million years to 550 million years. See the Planck news release [304] and the list of publication [305]. The redshift of the reionization z_{reion} should be reduced. Tentatively our discussion uses the old value of z_{reion} .

tribution, continues to redshift to lower energies. This allows more and more of the gas of electrons, protons, and helium nuclei to combined to form neutral hydrogen and helium atoms. As the temperature of the universe dropped below 3000 K, almost all ions and electrons are combined into neutral atoms, so photons are decoupled from the baryon and the universe becomes transparent. This is the last scattering surface, of course, and here the photon left the imprint of the CMB. Although the universe is transparent to radiation, there were no light sources existed yet. The CMB imprints have the averaging temperature lower than 3000 K which are therefore mostly in the middle infrared range and redshifted further downward to lower frequencies. This hence begins an epoch of the *dark ages*.

The dark age ends when the universe is about 400 million years old with the appearance of first stars and galaxies which emit ultra violet ionizing radiations to ionize neutral atoms to initiate the epoch of reionization (EoR).²⁰⁹⁾ In the beginning of EoR, the intergalactic medium is neutral except in the immediate regions around the first stars. The fraction of the reionized gas increases with the increase of ionizing sources of more stars and galaxies. The universe was reverted from a neutral state to the state of an ionized plasma. This happens in the redshift interval of $6 < z < 15$ in the universe age interval of 400 million years to 1 billion years. At this point the universe had made sufficient expansions, matter had become diffused enough so that the photon-electric charge interactions, causing scatterings with electrons and, to a much smaller rate, protons, are sufficiently rare, the universe stays largely transparent.

Although for the majority of CMB photons the universe is entirely transparent, there is a small but finite probability that some of the photons are scattered by electrons in the reionized plasma.²¹⁰⁾ So the CMB we observed today consists of two components. One component involves photons which are unaffected by the reionization effect. They preserve the foot prints of the LSS, which took place at around $z_{LSS} = 1100$. The probability of this unaffected photon is $\exp(-\tau_{reion})$, where τ_{reion} is the optical depth of the reionized plasma and is one of the parameters defining the standard Λ CDM model. The other component involves CMB photons undergone scattering off electrons rather recently at much smaller redshifts around $z_{reion} \approx 10$. The probability of this component of photons is $\sim 1 - \exp(-\tau_{reion})$. The radius of this “reionization surface” r_{reion} is much smaller than that of LSS r_L . So for a given wave number q , it contributes to much smaller multipole moments, $\ell \sim qr_{reion}$ (instead of $\ell \sim qr_L$). So we can neglect the contribution from this

component of CMB photon when we are interested in the effect of high multipole moments. Then the effect of the reionization is to multiply the form factors F_T and G_T given in Eq. (12.299) by the factor $\exp(-\tau_{reion})$, and the multipole coefficients by the factor $\exp(-2\tau_{reion})$. The most recent PDG [13] gives the value $\tau_{reion} = 0.091$ which corresponds to $\exp(-2\tau_{reion}) = 0.83$.²¹¹⁾

12.10.3 Explicit expression for the temperature multipole coefficient

Most ingredients are now in place for computing the temperature multipole coefficient as a function of the multipole order ℓ . We first specify the overall scale factor $\mathcal{R}_q^{(o)}$ of the perturbation functions given in Eq. (12.284). Let us recall that $\mathcal{R}_q^{(o)}$ is a conserved quantity outside the horizon and it characterizes the primordial fluctuations in the very early universe. We have to express the form factors F_T and G_T Eq. (12.299) as functions of the multipole order ℓ , and clarify a point relevant to the expression of the temperature multipole coefficient. Because of their significant number, a summary of the parameters that appear in the expression of the temperature multipole coefficient, will be given.

12.10.3.1 Overall factor $\mathcal{R}_q^{(o)}$

The overall factor can be parameterized as

$$|\mathcal{R}_q^{(o)}|^2 = \frac{N^2}{q^3} \left(\frac{q}{a_0 k_R} \right)^{n_s - 1}, \quad (12.300)$$

where n_s is the *scalar spectral index* which is one of the cosmological parameters defining the standard Λ CDM model [307], N a dimensionless constant, and k_R a scale factor for the wave number usually in units of Mpc^{-1} .²¹²⁾ We will come back to them later again when we plot the temperature multipole coefficient. We remark that $\mathcal{R}_q^{(o)}$ is related to the *curvature fluctuation amplitude*, also one of the parameters of the standard Λ CDM model [307]. As can be seen from Eqs. (12.277) and (12.299), $|\mathcal{R}_q^{(o)}|^2$ goes with the factor ℓ^3/r_L^3 , then we have²¹³⁾

$$\begin{aligned} \frac{\ell^3}{r_L^3} |\mathcal{R}_q^{(o)}|^2 &= N^2 \frac{\ell^3}{q^3 r_L^3} \left(\frac{q r_L}{(a_0/a_L) a_L r_L k_R} \right)^{n_s - 1} \\ &= \frac{N^2}{\beta^3} \left(\frac{\ell \beta}{(1 + z_L) k_R d_A^{(L)}} \right)^{n_s - 1}, \end{aligned} \quad (12.301)$$

where $q \equiv \ell \beta / r_L$ defined in Eq. (12.274) has been used. $d_A^{(L)} \equiv a_L r_L$ is the angular diameter distance of

²⁰⁹⁾Many descriptions of the reionization epoch can be found in the literature. See, for an example, [306].

²¹⁰⁾The discussion follows that given on p. 353, [221].

²¹¹⁾[221] uses $\exp(-2\tau_{reion}) = 0.8$ which gives $\tau_{reion} \approx 0.117$.

²¹²⁾See a discussion given in [307], §27.3 *Cosmological Parameters*.

²¹³⁾This is the first equation on p. 354, [221].

LSS, which, in the case of a flat geometry, is given in Eq. (13.79),

$$d_A^{(L)} = a_L r_L = \frac{1}{(1+z_L)H_0} \int_0^{z_L} \frac{dz}{\sqrt{\Omega_\Lambda + \Omega_M(1+z)^3 + \Omega_R(1+z)^4}}. \tag{12.302}$$

We refer to §13.6 for a discussion of the angular diameter distance.

12.10.3.2 Relating distances with multipole orders

Using Eqs. (12.274), (12.294)–(12.296), and (12.302), we can convert the wave number q dependence into the dependence on the order of multipole moments ℓ . We write,

$$\frac{q}{a_L} d_D = \ell\beta \frac{d_D}{d_A^{(L)}} \equiv \beta \frac{\ell}{\ell_D}, \quad \ell_D \equiv \frac{d_A^{(L)}}{d_D},$$

$$\begin{aligned} \frac{q}{a_L} d_T &= \ell\beta \frac{d_T}{d_A^{(L)}} = \beta \frac{\ell}{\ell_T}, & \ell_T &\equiv \frac{d_A^{(L)}}{d_T}, \\ \frac{q}{a_L} d_H &= \ell\beta \frac{d_H}{d_A^{(L)}} = \beta \frac{\ell}{\ell_H}, & \ell_H &\equiv \frac{d_A^{(L)}}{d_H}, \end{aligned} \tag{12.303}$$

and

$$\begin{aligned} \frac{\ell\beta}{(1+z_L)d_A^{(L)}k_R} &= \beta \frac{\ell}{\ell_R}, \\ \ell_R &\equiv (1+z_L)k_R d_A^{(L)}. \end{aligned} \tag{12.304}$$

12.10.3.3 Explicit expression of the scalar multipole coefficient and some of its properties

With the above relations, substituting Eq. (12.299) into Eq. (12.277), and rescale the coefficient by the multiplying the factor $\ell(\ell+1) \approx \ell^2$ for large ℓ , we have²¹⁴⁾

$$\begin{aligned} \frac{\ell(\ell+1)}{2\pi} C_{TT,\ell}^{(S)} &= \frac{4\pi}{25} T_0^2 N^2 e^{-2\tau_{reion}} \int_1^\infty d\beta \frac{1}{\beta^2 \sqrt{\beta^2 - 1}} \left(\frac{\ell}{\ell_R} \beta\right)^{n_s - 1} \\ &\times \left\{ \left[3R_{BL} \mathcal{T} \left(\frac{\ell}{\ell_T} \beta\right) - \frac{e^{-2(\ell^2/\ell_D^2)\beta^2}}{(1+R_{BL})^{1/4}} S \left(\frac{\ell}{\ell_T} \beta\right) \cos \left(\frac{\ell}{\ell_H} \beta + \Delta \left(\frac{\ell}{\ell_T} \beta\right)\right) \right]^2 \right. \\ &\left. + \frac{3(\beta^2 - 1)}{\beta^2 (1+R_{BL})^{3/2}} e^{-2(\ell^2/\ell_D^2)\beta^2} S^2 \left(\frac{\ell}{\ell_T} \beta\right) \sin^2 \left(\frac{\ell}{\ell_H} \beta + \Delta \left(\frac{\ell}{\ell_T} \beta\right)\right) \right\}. \end{aligned} \tag{12.305}$$

The gross behavior of the temperature multipole coefficient can be seen from the integrand of the above expression. We outline it briefly below.

- It is an oscillating function of ℓ and the periodicity of the oscillation is largely determined by ℓ_T . It is suppressed exponentially in ℓ and β . Therefore, important contributions to the integral come from the lower limit of integration around $\beta \approx 1$.
- The $(n+1)$ th oscillation peak will be suppressed relative to the n th peak as the former occurs at a high ℓ value.
- The height of the curve is determined by ℓ_D and the transfer function S which depends on ℓ_T .
- The contribution from the form factor G_T , which contains the factor $(\beta^2 - 1)/\beta^2$, is suppressed relative to the form factor F_T . Hence the behavior of the multipole coefficient is largely determined by the form factor F_T . Hence the second line of the right-handed side of Eq. (12.305), which comes from F_T as can be seen in Eq. (12.299), is the controlling term.

- Let us denote the argument of the cosine term by $\vartheta = \ell\beta/\ell_H + \Delta(\ell\beta/\ell_T)$. Then the peaks occur at $\vartheta = n\pi$, $n = 1, 2, \dots$. The two terms in the controlling line are added constructively for odd $n = 2n' - 1$ but destructively for even $n = 2n'$, where $n' = 1, 2, \dots$. This implies that the peak at $2n' + 1$ may not be suppressed as strongly in comparison with that at $2n'$ as it would be otherwise. Since the transfer function Δ is small as we will see later, the first peak will occur at ℓ to be a few times ℓ_T . The second peak roughly doubles that of the value of ℓ , the third peak roughly triples, etc., unless ℓ_H varies significantly from the value determined by the relevant future cosmic parameters, which is unlikely.
- Strictly speaking there is a weak singularity at $\ell = 0$ in the form $(\ell)^{n_s - 1}$ for $n_s < 1$. Since the expression is only valid for ℓ much greater than 1, we can ignore this singularity.

12.10.3.4 Numerical results

Because of the complication due to the dependence on multiple parameters, detailed studies of the temperature multipole coefficient can best be made numerically. We

²¹⁴⁾This is Eq. (7.2.41), p. 354, [221].

take the needed parameters given in the most recent PDG [13] as the *baseline* and vary some of the parameters to see how $C_{TT,\ell}^{(S)}$ responses to changes, while hold some other parameters fixed.

- The fixed input parameters and their PDG2014 [13] values are: $T_0 = 2.7255$ K, $n_s = 0.958$, $\tau_{reion} = 0.091$, and $z_L = 1090 - 1$.²¹⁵⁾
- The scale factors $N^2 = 1.736 \times 10^{-10}$ and $k_R = 0.05$ Mpc^{-1} are given in [221].²¹⁶⁾
- With the input values given above we have the overall coefficient in front of the integral in Eq. (12.305),²¹⁷⁾

$$\frac{4\pi}{25} T_0^2 N^2 e^{-2\tau_{reion}} (10^6 \mu\text{K})^2 = 540.3 \mu\text{K}^2. \quad (12.306)$$

- ℓ_H , ℓ_L , and ℓ_H depend on the matter-energy density straightforwardly. They will be calculated with each set of parameters we choose to work with.
- $\ell_D = d_A^{(L)}/d_D$ [Eqs. (12.303) and (13.87)], where $d_D = \sqrt{d_{\text{Silk}}^2 + d_{\text{Landau}}^2}$ [Eq. (12.294)], is complicated. Although we have only glossed over them, d_{Silk} and d_{Landau} can be found in details in [221].²¹⁸⁾ With the expressions d_{Landau} given in [221],²¹⁹⁾

$$d_{\text{Landau}}^2 = \frac{3\sigma^2 t_L^2}{8T_L(1 + R_{BL})}, \quad (12.307)$$

where $t_L = 0.1134$ Mpc (370 000 years) and $T_L = (1 + z_L)T_0$ are the LSS time and temperature, and $\sigma = 262$ K. It can be readily calculated for a given set of matter-radiation densities, which we will do below. But d_{Silk} is much more involved. We will approximate it as fixed and take its value $d_{\text{Silk}} = 0.006555$ Mpc as given in [221].²²⁰⁾ This approximation will introduce some uncertainties in the results

²¹⁵⁾These values are not exactly those given in [221], but very close to. See p. 356, [221].

²¹⁶⁾See, p. 356, [221]. The same value for k_R is also given in [307] where it is denoted as k_0 .

²¹⁷⁾This is Eq. (7.2.51), p. 356 of [221] and the numerical values given is $519.7 \mu\text{K}^2$ which corresponds to a slightly larger value of $\tau_{reion} = 0.1103$.

²¹⁸⁾See Eqs. (7.3.34) and (7.3.35), p. 352, and the related discussions, [221].

²¹⁹⁾Eq. (7.3.35), p. 352, [221].

²²⁰⁾Eq. (7.2.49), p. 356, [221]. The expression for d_{Silk} is given in Eq. (7.2.34), p. 352 in [221]. For completeness we record the expression here. The integration in time t is converted into that in R_B by Eq. (??),

$$d_{\text{Silk}}^2 = \frac{R_{BL}^2}{6(1 - Y)n_{B0}\sigma_T cH_0\sqrt{\Omega_M}R_{B0}^{9/2}} \times \int_0^{R_{BL}} \frac{dR_B R_B^2}{X(R_B)(1 + R_B)\sqrt{R_B + R_{BEQ}}} \left(\frac{16}{15} + \frac{R_B^2}{1 + R_B} \right),$$

where $Y = 0.24$ is the helium fraction of nucleons, n_{B0} is the present baryon number density, σ_T is the Thomson cross section, and $X(R_B)$ is the proton density fraction as a function of R_B .

below. But we do not expect the variation in d_{Silk} to be anything drastic in the parameter ranges we will consider below. So we expect that the main characteristics of the parameter variation will be maintained.

Below we list 10 sets input data in Table 12.2 to be used to calculate the temperature multipole coefficient.

The 10 sets are made into 4 groups: The baseline group uses the observational data. The rest are divided into 3 groups, having 3 sets of data in each group with a definite characteristics of variations:

- For the baseline input we take the most recent data from [13]: $h = 0.673$, $\Omega_B = 0.02207h^{-2} = 0.0487$, $\Omega_M = 0.1419h^{-2} = 0.3132$, $\Omega_\Lambda = 1 - \Omega_M = 0.6868$.
- Data sets *A1*, *A2*, and *A3*: having the same values as the baseline set in the matter density and therefore dark energy density but varying baryon density, with $\Omega_B h^2 = 0.005, 0.08, 0.14$ respectively for *A1*, *A2*, and *A3*. So the dark matter density will vary accordingly.
- Data sets *B1*, *B2*, and *B3*, having the same values as the baseline set in the baryon density but varying matter density and therefore the dark energy and dark matter densities, taking $\Omega_M h^2 = 0.1, 0.25, 0.45$ respectively for *B1*, *B2*, and *B3*.
- Data sets *C1*, *C2*, and *C3*, with varying scale factor for the Hubble expansion rate: $h = 0.38, 0.5, 0.9$

For a flat universe the condition $\Omega_M + \Omega_\Lambda = 1$ is imposed in the accuracy range we are working with. Therefore we do not consider the case of varying the dark energy density while keeping all other densities fixed. The above variations of the input data considered are straightforward. There are more complicated ways to vary the input data. It belongs to the realm of data fitting which is not what we intend to do here.

We plot in Fig. 12.9 the temperature multipole coefficient for the three groups of input parameters. Each group is presented together with the baseline data for the purpose of contrast. The vertical axis is the conventional temperature multipole coefficient $\ell(\ell + 1)C_\ell^{(S)}/(2\pi)$ and the horizontal axis is the multipole order ℓ . The data set for each curve is indicated.

Let us briefly comment on the characteristics of the data sets and aspects of physics they are sensitive to.

- As shown in the top panel of Fig. 12.9 for the data set *A*, when the baryon density increases, peaks and valleys move to higher values of ℓ . The height of the first peak increases with Ω_B when Ω_B is small. But the height changes very little with Ω_B when $\Omega_B h^{-2}$ is greater than 0.02 or Ω_B greater than 0.05 as shown by curves *A3*, *A2* and *Base*.

Table 12.2 Values of parameters entering the temperature multipole coefficient in units of μK^2 .

	Base	A1	A2	A3	B1	B2	B3	C1	C2	C3
h	0.673	0.673	0.673	0.673	0.673	0.673	0.673	0.38	0.5	0.9
Ω_B	0.049	0.011	0.177	0.309	0.049	0.049	0.049	0.153	0.088	0.027
Ω_M	0.313	0.313	0.313	0.313	0.221	0.552	0.994	0.983	0.568	0.175
Ω_Λ	0.687	0.687	0.697	0.687	0.779	0.448	0.006	0.017	0.432	0.825
R_{B0}	670.1	151.8	2429	4251	670.1	760.1	670.1	670.1	670.1	670.1
R_{BL}	0.615	0.139	2.229	3.900	0.615	0.615	0.615	0.615	0.615	0.615
R_{BEQ}	0.196	0.044	0.711	1.244	0.278	0.111	0.062	0.196	0.196	0.196
$d_A^{(L)}$	12.83	12.83	12.83	12.83	14.76	10.18	7.95	14.14	13.54	12.09
d_H	0.133	0.145	0.111	0.097	0.144	0.114	0.094	0.133	0.133	0.133
d_T	0.125	0.125	0.125	0.125	0.177	0.071	0.039	0.125	0.125	0.125
$d_{\text{Landau}} \times 10^3$	4.820	5.739	3.409	2.767	4.820	4.820	4.820	4.820	4.820	4.820
$d_D \times 10^3$	8.137	8.712	7.388	7.115	8.137	8.137	8.137	8.137	8.137	8.137
ℓ_D	1577	1473	1736	1803	1814	1251	977	1738	1664	1486
ℓ_H	96.39	88.6	116.4	132.2	102.3	89.2	84.1	106.3	101.7	90.8
ℓ_R	699.2	699.2	699.2	699.2	804.3	554.6	433.2	770.7	737.9	658.0
ℓ_T	102.7	102.7	102.7	102.7	83.26	143.53	202.0	113.2	108.4	96.8

- When the matter density changes the trend is quite clear as shown in the middle panel of Fig. 12.9. The peaks and valleys move to lower values of ℓ and heights of all peaks become lower when Ω_M increases. The trend in decreasing heights of even peaks are particularly noticeable. As figure B3 shows, for very large matter density, e.g., $\Omega_M = 0.994$ or $\Omega_M h^{-2} = 0.45$, the even peaks disappear.
- When h changes, as shown in the bottom panel of Fig. 12.9, the difference among the four data sets is not very pronounced except for quite small h . As shown by curve C1, for $h = 0.38$ peaks are higher and at larger ℓ values.

Taking all curves in Fig. 12.9 together, We can conclude that there are lot degeneracies in the parameter sets. Therefore, good data together with detailed simulations are necessary to extra the cosmological parameters.

12.10.4 Multipole coefficients $C_{TT,\ell}$

12.10.4.1 Observations

The standard ΛCDM model is generally defined to involve 6 parameters: Ω_M , Ω_B , Ω_Λ , n_s , τ_{reion} , and the curvature fluctuation amplitude related to $\mathcal{R}_q^{(0)}$ Eq. (12.300). However, for the comparison of theory with observational data, the scope of the required parameters evolves and the number varies from 5 to 10.²²¹⁾

Determination of the temperature multipole coefficient is a fundamental task of cosmological observations. The most recent high statistics data came in 2013 from

²²¹⁾See the review article by O. Lahav and A.R. Liddle, 24. *The cosmological parameters* in the PDG [13].

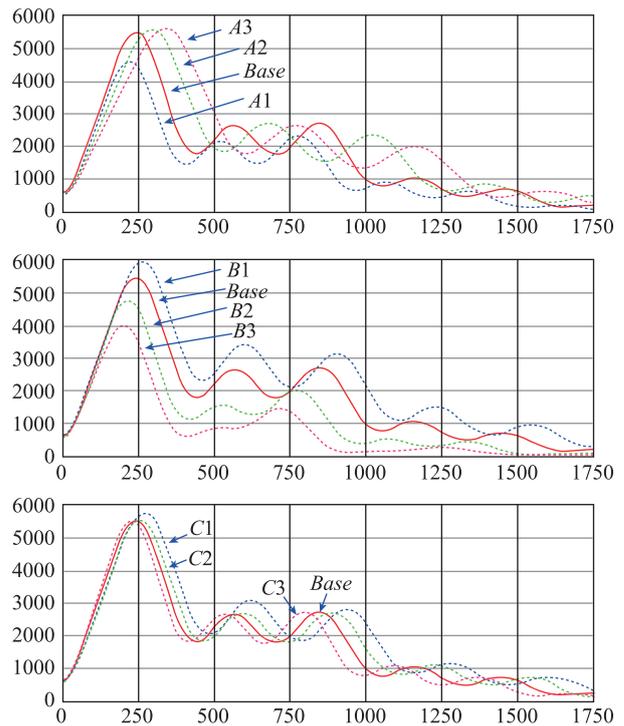


Fig. 12.9 A comparison of the three of input data. For details see the text.

the Planck collaboration [308, 309], and in 2012 from WMAP9 [310, 311]. The Planck 2013 results on the temperature multipole coefficient is reproduced here in Fig. 12.10. The vertical axis is the conventional quantity $D_\ell \equiv \ell(\ell + 1)C_\ell/(2\pi)$. The plot shows several acoustic peaks and valleys which can be well-fitted by the six-parameter ΛCDM model. The horizontal axis is in linear

scale for $\ell > 50$ and logarithmic scale for $\ell < 50$. The shaded area below $\ell < 50$ increases when ℓ decreases to include cosmic variance. We refer to the Planck publications cited, [308] and [309], for details of the plot and the cosmological parameters obtained. Since the corresponding plot of WMAP9 uses different scales, which do not facilitate an easy comparison, we will not show the WMAP9 results here. The WMAP9 results can be readily found in the two publications [310] and [311]. Scientific Publications of WMAP are listed in [312].

A comparison of the Planck and WMAP9 results has been made in [313]. The figure which compares temperature multipole coefficient of the two collaborations is reproduced in Fig. 12.11. In the whole range of the data sets, the two results agree within a few per cent, depending on the particular data sets included in the comparison. We note that “eCMB” means the extended WMAP data set. For an explanation see [310].

12.10.4.2 Analytic expression vs observational data

The temperature multipole coefficient Eq. (12.305) depends on nine quantities: T_0 , N , τ_{reion} , n_s , R_{BL} , ℓ_D , ℓ_H , ℓ_R , and ℓ_T . The last five R_{BL} , ℓ_D , etc., are functions of the cosmic matter-energy densities, where ℓ_D has a complicated dependence on some other quantities as well, as shown in Eqs. (12.293), (12.293), (12.304), and Footnote 220). To make a numerical representation let us examine these parameters in some detail.

- The cosmological quantities, such as T_0 , $\rho_{\gamma 0} = \Omega_\gamma h^{-2}$, z_L , z_{EQ} , are well-delineated, so they are parts of the fixed input. Hence the value of R_{B0} , which enters various parameters of Eq. (12.305), can be used to determine Ω_B .
- The baryon to photon density ratio R_B at the three temporal milestones, the matter-radiation equality R_{REQ} , the LSS R_{BL} , and the present epoch R_{B0} are related:

$$R_{B0} = \frac{3\Omega_B}{4\Omega_\gamma}, \quad R_{BL} = \frac{R_{B0}}{1+z_L}, \quad R_{B0} = \frac{R_{B0}}{1+z_{EQ}}. \quad (12.308)$$

- The multipole orders ℓ_T , ℓ_H , ℓ_R are given in terms of the density parameters as follows:

$$\begin{aligned} \ell_T &= \frac{d_A^{(L)}}{d_T} = \frac{\Omega_M}{\sqrt{\Omega_R}} I(z_L), \\ \ell_H &= \frac{d_A^{(L)}}{d_H} = \left(\frac{2}{\sqrt{3R_{B0}\Omega_M}} \right. \\ &\quad \left. \times \ln \frac{\sqrt{1+R_{BL}} + \sqrt{R_{BEQ}+R_{BL}}}{1+R_{BEQ}} \right)^{-1} I(z_L), \\ \ell_R &= \frac{k_R}{H_0} I(z_L), \end{aligned} \quad (12.309)$$

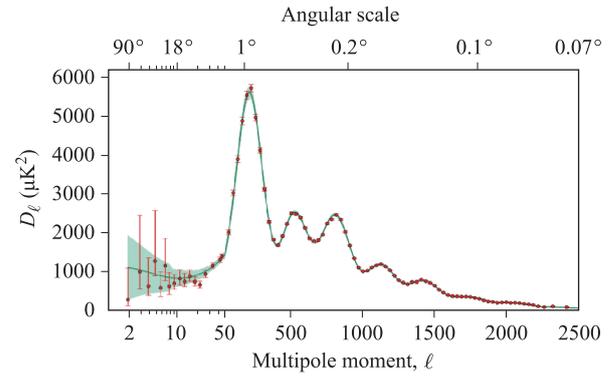


Fig. 12.10 CMB temperature power spectrum from Planck. See the related text for more explanation. More details are found in the relevant references, e.g., [308] and [309].

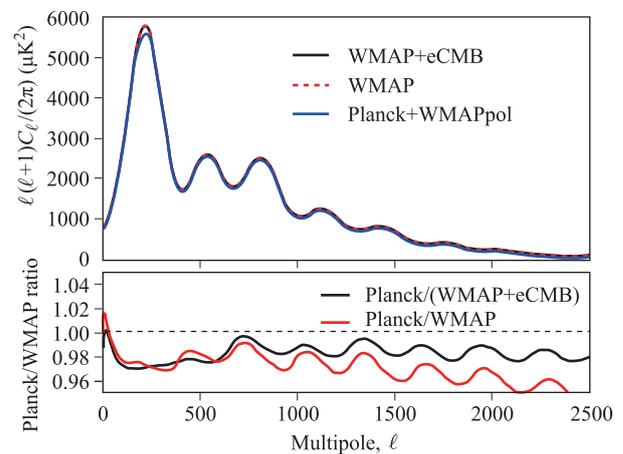


Fig. 12.11 A comparison of the Planck2013 and WMAP9 best fits of the multipole coefficient given in [313]. The bottom panel gives the ratio of Planck divided by WMAP. The agreement is generally around 2% for the most prominent part of the data sets.

where

$$I(z_L) \equiv \int_0^{z_L} \frac{dz}{\sqrt{\Omega_\Lambda + \Omega_M(1+z)^3 + \Omega_R(1+z)^4}}. \quad (12.310)$$

We assume a flat universe.

- ℓ_D has a complicated expression. It is composed of two components, the Silk and Landau damping lengths. We refer to the discussions given in [221].²²²⁾
- The dark energy enters the calculation in the angular diameter distance $d_A(L)$ through the integral defined in Eq. (12.310), together with the other matter-energy densities.

²²²⁾See Eqs. (7.2.34) and (7.2.35), p. 352, [221].

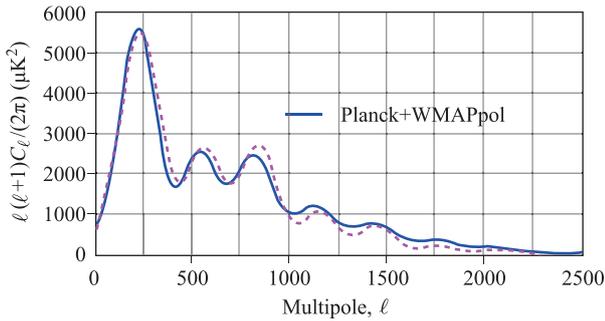


Fig. 12.12 A comparison of the analytic expression Eq. (12.305) with the Planck + WMAP polarization data set as plotted in [313]. The red dash-dotted curve represented the analytic expression and the blue solid curve the said data set.

In Fig. 12.12 we plot the analytic expression Eq. (12.305) using the baseline parameter together with the corresponding observational curve given by [313], i.e., the Planck+WP data set. The blue solid curve is the observation and the red dash-dotted curve represents the analytic expression (12.305). The overall physics is represented fairly well by the analytic expression. The mismatch of the details is also quite clear. The oscillation in the analytic expression is somewhat slower and the dissipation or radiation damping for large ℓ is too strong. However, in view of the complication of the physics involved that entails the approximation adopted and the multiple characters that enter into the play, the analytic expression derived in [221] is doing a very respectable job.

13 Cosmological distances and times

To make this section more or less independent, there are overlaps and repetitions with Section 9.

13.1 Introduction

Well-defined concept of distances is fundamental to the understanding of a dynamic system, such as the expanding universe of ours. However, distance measurements in cosmology²²³⁾ is complicated, since the distances between objects vary as time evolves due to the expansion of the universe. For an observer on Earth, distance measures out into distant galaxies and stars are also look back in time, as both measurements rely on light propagation in the null path, which is the trajectory of the electromagnetic radiation emitted from the object under observation and received by an Earth observer. Consequently, expressions of the various distance and time

²²³⁾Also known as cosmography. See [314] for a discussion on distance measurements.

measures are related. To stay general, we will derive distance and time expressions in the FLRW (Friedmann-Lemaître-Robertson-Walker) metric for the case of a curved space of finite curvature constant K and summarize them for the case of the flat Λ CDM universe.

There are various distance scales defined in cosmology, but most are not directly measurable, under the general situation of a curved space. In a curved space, some of the distance definitions depend on certain basic quantities, such as the scale factor $a(t)$ and the comoving coordinate²²⁴⁾ radial distance r which are not measurable experimentally. Most of the distance quantities to be discussed below are related essentially to one of the two distance definitions: the *proper distance* and the *angular diameter distance*. The latter is defined for distances perpendicular to the line of sight, while the former along the line of sight, of the observer in the comoving coordinates. Measurable quantities include the red shift, angular diameter distance, and the flux of a radiation source, etc. Hence some of the distance quantities are “derived concept”²²⁵⁾ which are related to measurable quantities to be discussed below.

An observable distance here refers to the situation, in which, e.g., photons emitted during an event can reach the observer in the present age of the universe, or eventually into the future. It does not necessarily mean that the event can be measured experimentally. Actual observable events in terms of photon observables can take place only after the last scattering surface (LSS) after the photon decouples from the matter plasma, when the universe is about 370 000 years of age. Such photons are said to be in the state of *free streaming* so that they suffer no interactions with charged particles in their path from their emitting sources to the observer. Much earlier measurable events can happen, in principle, in terms of neutrinos, which are decoupled when the universe is about one second old. Gravitational waves, which carry information of the cosmic birth, can serve to probe the every beginning of the universe.

13.2 The FLRW metric

To make the present discussion self-contained, some of the material discussed in Section 9, to which we refer for details, will be repeated here. We start with the FLRW metric.

$$ds^2 = -g_{\mu\nu}dx^\mu dx^\nu$$

²²⁴⁾Let us recall that the comoving coordinate frame is a coordinate system which expands in synchronization with the Hubble expansion of the universe. Cosmological objects, such as galaxies, when removed of their *peculiar velocities*, are fixed in the comoving frame. In the FLRW metric, the r , θ , and ϕ are the spherical coordinates of the comoving frame and $a(t)$ is the Hubble expansion factor.

²²⁵⁾See p. 74, [222].

$$= c^2 dt^2 - a^2(t) \left[\frac{dr^2}{1 - \kappa r^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \right], \tag{13.1}$$

where, here and below, we keep the velocity of the light c for clarity and for the proper engineering dimensionality of the distance. $\sqrt{ds^2}$ is called the *proper time interval* and t the *cosmic time*. $a(t)$ is the Hubble scale factor which determines the Hubble expansion rate. K is the curvature constant, with a suitably normalized scale factor, taken to be 0 for the flat space, +1 for the closed space, and -1 for the open space. r , θ , and ϕ are the *comoving spherical coordinates*, where r is referred to as the *comoving radial coordinate*. Another time variable, called the *conformal time* η , is defined by

$$d\eta \equiv \frac{dt}{a(t)}, \tag{13.2}$$

which is convenient in the discussions below.

We note that the comoving coordinates are time independent in the expanding homogeneous and isotropic background universe. They define the comoving positions of cosmic objects like galaxies when the objects' peculiar motion are neglected.²²⁶⁾

An equivalent definition of the comoving radial coordinate, denoted by χ , is related to r by

$$d\chi = \frac{dr}{\sqrt{1 - \kappa r^2}}. \tag{13.3}$$

Integrating the expression, with the condition $\chi = 0$ for $r = 0$, we can relate the two forms of the comoving radial coordinate:²²⁷⁾

$$\chi = \frac{1}{\sqrt{\kappa}} \arcsin(\sqrt{\kappa}r) = \begin{cases} \frac{1}{\sqrt{\kappa}} \arcsin(\sqrt{\kappa}r), & \kappa > 0, \\ r, & \kappa = 0, \\ \frac{1}{\sqrt{|\kappa|}} \operatorname{arcsinh}(\sqrt{|\kappa|}r), & \kappa < 0, \end{cases} \tag{13.4}$$

and inverting the above expression to give

$$r = \frac{1}{\sqrt{\kappa}} \sin(\sqrt{\kappa}\chi) = \begin{cases} \frac{1}{\sqrt{\kappa}} \sin(\sqrt{\kappa}\chi), & \kappa > 0, \\ \chi, & \kappa = 0, \\ \frac{1}{\sqrt{|\kappa|}} \sinh(\sqrt{|\kappa|}\chi), & \kappa < 0. \end{cases} \tag{13.5}$$

²²⁶⁾The peculiar motion, or peculiar velocity, refers to the part of the motion of a galaxy which cannot be explained by the Hubble flow of the expansion of the universe. The peculiar velocity of an object is generally due to gravitational interactions with nearby galaxies or over-densities. The survey of the peculiar velocity field is a useful tool served as independent probes of cosmological parameters.

²²⁷⁾The identity $\sin(i\theta) = i \sinh(\theta)$ is used below.

So the metric can also be written as

$$ds^2 = c^2 dt^2 - a^2(t) [d\chi^2 + S_\kappa(\chi)(d\theta^2 + \sin^2\theta d\phi^2)], \tag{13.6}$$

where

$$S_\kappa(\chi) = r^2 = \frac{1}{\kappa} \sin^2(\sqrt{\kappa}\chi). \tag{13.7}$$

In the flat space, $\kappa = 0$, $r = \chi$, the two radial coordinates are identical.

13.2.1 The past and future light-cones

Let an observer be located at the origin of the comoving coordinate system. Consider a light signal traveling between the origin and the a point with the comoving radial coordinate r_1 . The cosmic time of the signal happens at the origin at t_0 or η_0 , and at r_1 or χ_1 at t_1 or η_1 . The light-cones, past and future, with respect to the observer can be defined. For the past light-cone $t_1 < t_0$ the light signal travels from r_1 to the origin, and for the future light-cone $t_1 > t_0$ from the origin to r_1 . Taking the light signal traveling path along a great circle of constant angular coordinates, i.e., $ds^2 = 0$ and $d\Omega = 0$, we have

$$cd\eta = c \frac{dt}{a(t)} = \mp d\chi = \mp \frac{dr}{\sqrt{1 - \kappa r^2}}. \tag{13.8}$$

The “-” sign defines the past light-cone while the “+” sign the future light-cone. The integral relation is

$$\int_{t_1}^{t_0} d\eta = \mp \int_0^{\chi} d\chi, \tag{13.9}$$

$$\eta_1 = \eta_0 \mp \chi_1.$$

Again the - and + are respectively for the past and future light-cones.

13.2.2 The Λ CDM model

We define the Hubble expansion rate H in the Λ CDM model,

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G_N}{3} \sum_j \rho_j, \tag{13.10}$$

where G_N is the Newtonian gravitational constant. ρ_j is one of the four forms of energy densities: the cosmological constant ρ_Λ , non-relativistic matter ρ_M , radiation ρ_R , and curvature constant ρ_K , with

$$\rho_\Lambda \equiv \frac{1}{8\pi G_N} \Lambda$$

$$\rho_\kappa \equiv -\frac{3}{8\pi G_N} \frac{c^2 \kappa}{a^2}, \tag{13.11}$$

where Λ is the cosmological constant. The form of the energy density as a function of the cosmic scale factor is:

$$\rho_j \sim a^{-3(1+w_j)}, \tag{13.12}$$

where w_j denotes the equation of state of the various components of the energy forms: for the cosmological constant $w_\Lambda = -1$, non-relativistic matter $w_M = 0$, radiation $w_R = 1/3$, and curvature constant $w_K = -1/3$. A summary of the w_j can be found in Table 9.1.

Below the subscript “0” denotes quantities of the present epoch, e.g., t_0 is the cosmic time at the present epoch, H_0 the present expansion rate or the Hubble constant, a_0 the present scale factor, etc. The critical energy density is the total energy density of the universe at the present time defined in terms of the Hubble constant H_0 ,

$$\rho_c = \frac{3H_0^2}{8\pi G_N}. \tag{13.13}$$

Define the density ratios of the present epoch

$$\begin{aligned} \Omega_j &\equiv \frac{\rho_j}{\rho_c}, \\ \Omega_\Lambda + \Omega_M + \Omega_R + \Omega_K &= 1 \end{aligned} \tag{13.14}$$

In particular

$$\Omega_\kappa = \frac{\rho_\kappa}{\rho_c} = -\frac{c^2\kappa}{a_0^2 H_0^2}. \tag{13.15}$$

We also define

$$\begin{aligned} \Omega_0 &\equiv \Omega_\Lambda + \Omega_M + \Omega_R, \\ \Omega_\kappa &= 1 - \Omega_0. \end{aligned} \tag{13.16}$$

We can rewrite the curvature constant κ in terms of the effective curvature energy density ratio,

$$\sqrt{\kappa} = i \frac{a_0 H_0}{c} \sqrt{\Omega_\kappa} = i \frac{a_0}{D_H} \sqrt{\Omega_\kappa}, \tag{13.17}$$

where D_H is the *Hubble length*²²⁸⁾ at the present epoch defined by

$$D_H \equiv \frac{c}{H_0}, \tag{13.18}$$

which we will come back to later. We can rewrite Eq. (13.7) in different forms:

$$\begin{aligned} S_\kappa(\chi) &= \frac{1}{\kappa} \sin^2(\sqrt{\kappa}\chi) = \frac{D_H^2}{a_0^2 \Omega_\kappa} \sinh^2\left(\frac{a_0 \sqrt{\Omega_\kappa}}{D_H} \chi\right) \\ &= \frac{D_H^2}{a_0^2 (1 - \Omega_0)} \sinh^2\left(\frac{a_0 \sqrt{1 - \Omega_0}}{D_H} \chi\right), \end{aligned}$$

In the case of a flat universe $\Omega_\kappa = 0$, $\Omega_0 = 1$, $\chi = r$, and

$$S_\kappa(\chi)|_{\Omega_0=1} = \chi^2 = r^2 \tag{13.19}$$

²²⁸⁾It is also referred to as the *Hubble radius*.

13.3 Distance & time, comoving radial coordinate

Since distance measurements are carried out by the electromagnetic radiation, which travels along a null path, we can convert a distance measurement into a measurement of the cosmic time. We explore this below.

13.3.1 Light propagation and null path in the FLRW metric

From the FLRW metric, we have the null path of light propagation,

$$\begin{aligned} ds^2 &= c^2 dt^2 - a(t)^2 \left(\frac{dr^2}{1 - \kappa r^2} + r^2 d\Omega \right) \\ &= c^2 dt^2 - a(t)^2 [d\chi^2 + S_\kappa(\chi) d\Omega] = 0. \end{aligned} \tag{13.20}$$

In particular, for an observer located at the origin of the comoving frame, we can choose the comoving coordinates so that the null path has constant angular variables so that $d\Omega = 0$. Then

$$\frac{cdt}{a(t)} \equiv cd\eta = \frac{dr}{\sqrt{1 - \kappa r^2}} = d\chi. \tag{13.21}$$

Note that the left-handed side is a function of the cosmic time while terms after the second equality sign are functions of comoving radial coordinates which are time independent. So, the above relation characterizes the null path and allows us to relate distance and time, and express them in terms of the redshift z or the Hubble scale factor $a(t)$.

For $t \leq t_0$, where t_0 is the time of the present epoch, the expansion scale factor and the redshift are related by

$$a(t) = a(z) = (1 + z)^{-1} a_0, \tag{13.22}$$

where $a_0 \equiv a(t_0)$ and z is the redshift.

Time and distance expressions become explicit in Λ CDM and can be evaluated numerically. They may have analytical expressions in special cases, as the expression of the Hubble expansion rate becomes known. We can rewrite the Hubble expansion rate at arbitrary cosmic time as

$$\begin{aligned} H &= \frac{\dot{a}}{a} = H_0 \left(\frac{H}{H_0} \right) \\ &= H_0 \sqrt{\frac{\sum_j \rho_j}{\rho_c}} \equiv H_0 \tilde{E}, \end{aligned} \tag{13.23}$$

where \tilde{E} can be written explicitly as

$$\begin{aligned} \tilde{E} &\equiv \frac{H}{H_0} = \sqrt{\frac{\sum_j \rho_j}{\rho_c}} \\ &\equiv \tilde{E}_z = \sqrt{\Omega_\Lambda + \Omega_M (1+z)^3 + \Omega_R (1+z)^4 + \Omega_\kappa (1+z)^2}. \end{aligned} \tag{13.24}$$

We can rewrite the time derivative from $H = \dot{a}/a$,

$$dt = \frac{1}{H} \frac{da}{a} = \frac{1}{H_0 \tilde{E}} \frac{da}{a} = -\frac{1}{H_0} \frac{1}{1+z} \frac{dz}{\sqrt{\Omega_\Lambda + \Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_\kappa(1+z)^2}}. \quad (13.25)$$

13.3.2 Comoving Radial coordinate

Since the comoving coordinates are fixed, we can obtain an explicit expression for it in terms of the energy-matter density variables of the universe, up to a scale factor. Let us take an object located at the radial coordinate r_1 emit a light signal at the cosmic time t_1 which reaches an (Earth) observer at t_0 . The object has the observed redshift z_1 . From Eqs. (13.21) and (13.25) we can write the comoving radial distance in terms of the redshift, etc.,

$$d\chi = \frac{cdt}{a(t)} = cd\eta = -\frac{D_H}{a_0} \frac{dz}{\sqrt{\Omega_\Lambda + \Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_\kappa(1+z)^2}}, \quad (13.26)$$

where D_H is the Hubble length defined in Eq. (13.18).

We can relate the radial coordinate χ_1 or r_1 located at the redshift z_1

$$\begin{aligned} \chi_1 &= \chi(z_1) = c \int_{t_1}^{t_0} \frac{dt}{a(t)} \\ &= \frac{D_H}{a_0} \int_0^{z_1} \frac{dz}{\sqrt{\Omega_\Lambda + \Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_\kappa(1+z)^2}}, \\ r_1 &= r(z_1) = \frac{1}{\sqrt{\kappa}} \sin(\sqrt{\kappa}\chi_1). \end{aligned} \quad (13.27)$$

So the comoving radial coordinate is determined by the redshift up to an unknown constant normalization factor a_0 . Since \tilde{E}_z is positive definite, this expression shows clearly that the larger the redshift of an object has, the further it is from the observer.

The validity of the z integration lies in the range between 0 and ∞ in z . But what are their ranges of validity as the density ratios Ω_j are fixed in value? We make the following observation: Up to the present time the universe spends most of its time after the LSS. Starting from LSS, at $z_L \approx 1100$, all energy components of the universe are decoupled from one another and Ω_j 's are constant. However, when we go back in times earlier than LSS, there are transitions of non-relativistic to relativistic matter. So the values of Ω_j would be altered, and the radiation part will become larger. For quantities involving \tilde{E}_z^{-1} , since the radiation energy density is proportion to $(1+z)^4$, its value for larger z should be larger.

So applications of Eqs. (13.25) and (13.26) involved integrations of time periods earlier than LSS using constant Ω_j determined presently is an approximation. Since the dominant contribution comes from the period from LSS to the present, this approximation is valid.

Comparing the radial variables χ and r , We see from the above discussion that χ is related to the cosmic time and the redshift directly and can be evaluated straightforwardly. r is related to χ through the geometric or hyperbolic functions. In the case of a flat universe, which is the case of our universe, r and χ are identical, and the computation of distances are simplified.

13.3.3 Lookback time and light traveling time

$$\Delta t = t_0 - t_1$$

Considering a pulse of EM radiation emitted at time t_1 , from a source located at the comoving radial coordinate r_1 of redshift z_1 . The observer on Earth, located at $r_0 = 0$ receives the signal at time t_0 (the present epoch). The trajectory of the EM wave follows a null path along a fixed angular coordinates so that $d\Omega = 0$. The time interval $\Delta t = t_0 - t_1$, known as the *lookback time*, can be expressed in the FLRW metric as follows. According to Eq. (13.25), the *lookback time* as a function of the redshift z_1 is given by

$$\begin{aligned} \Delta t_{\text{lb}} = t_0 - t_1 &= \int_{t_1}^{t_0} dt = \tau_H \int_0^{z_1} \frac{1}{1+z} \frac{dz}{\tilde{E}_z} \\ &= \tau_H \int_0^{z_1} \frac{1}{1+z} \frac{dz}{\sqrt{\Omega_\Lambda + \Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_\kappa(1+z)^2}}, \end{aligned} \quad (13.28)$$

where we have defined the *Hubble time* τ_H by

$$\tau_H \equiv \frac{1}{H_0} = \frac{D_H}{c}. \quad (13.29)$$

For small z_1 , the lookback time Eq. (13.28) can be approximated as

$$\Delta t_{\text{lb}}|_{z_1 \ll 1} \approx \tau_H z_1. \quad (13.30)$$

13.4 Hubble scales, age of the universe, and explicit forms of scale factor

13.4.1 Hubble time τ_H and Hubble length D_H

The Hubble expansion rate at the present time is given by

$$\begin{aligned} H_0 &= 100h \text{ km} \cdot \text{s}^{-1} \cdot \text{Mpc}^{-1} \\ &= (9.777752 \text{ Gyr})^{-1} h \\ &= 3.2408 \times 10^{-18} h \text{ s}^{-1}. \end{aligned} \quad (13.31)$$

Then we have the Hubble time at the present epoch to

be²²⁹⁾

$$\begin{aligned} \tau_H &= \frac{1}{H_0} = 9.777752h^{-1} \times 10^9 \text{ yr} \\ &= \begin{cases} 1.377 \times 10^{10} \text{ yr} = 4.346 \times 10^{17} \text{ s}, & h = 0.71, \\ 1.453 \times 10^{10} \text{ yr} = 4.585 \times 10^{17} \text{ s}, & h = 0.673. \end{cases} \end{aligned} \quad (13.32)$$

The *Hubble length* is

$$\begin{aligned} D_H = c\tau_H &= 0.925063h^{-1} \times 10^{26} \text{ m} = 2.997925h^{-1} \text{ Gpc} \\ &= \begin{cases} 1.303 \times 10^{26} \text{ m} = 4.222 \text{ Gpc}, & \Omega_\Lambda = 0.73, \quad h = 0.71, \\ 1.374 \times 10^{26} \text{ m} = 4.455 \text{ Gpc}, & \Omega_\Lambda = 0.685, \quad h = 0.673. \end{cases} \end{aligned} \quad (13.33)$$

The Hubble length and Hubble time respectively provide scales for the cosmic length and time. The meaning of the Hubble length can be illustrated by considering the galaxy expansion velocity. As will be seen in Section 13.5.1, the Hubble length is the proper distance at the present epoch where the recession velocity equals to the velocity of light.

13.4.2 Ages of the universe τ_U and the matter-radiation epoch t_{EQ}

The age of the universe, $\tau_U = t_0$ is given by the above expression of lookback time by taking $z_1 \rightarrow \infty$ as discussed in Eq. (9.81).

$$\tau_U = \tau_H \int_0^\infty \frac{1}{1+z} \frac{dz}{\sqrt{\Omega_\Lambda + \Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_\kappa(1+z)^2}}. \quad (13.34)$$

Given the energy-matter density ratios the integral above can be performed readily numerically. We obtain the following values for a flat universe:²³⁰⁾

$$\tau_U = \begin{cases} 0.993\tau_H = 13.7 \text{ Byr}, & \Omega_\Lambda = 0.73, \quad h = 0.71, \\ 0.951\tau_H = 13.8 \text{ Byr}, & \Omega_\Lambda = 0.685, \quad h = 0.673, \end{cases} \quad (13.35)$$

where Byr is billion years.

The age of the epoch of the matter-radiation equality can also be calculated straightforward. This is to set the lower limit of the integration given above in Eq. (13.35) at the corresponding redshift z_{EQ} which can be calculated as follows: $\rho_R = (a_0/a)^3 \rho_{R0}$ and $\rho_M = (a_0/a)^3 \rho_{M0}$. At the matter-radiation equality $\rho_R = \rho_M$ which gives

$$z_{EQ} = \frac{a_0}{a_L} - 1 = \frac{\Omega_M}{\Omega_R} - 1 = 3430 \quad (13.36)$$

for $\Omega_M = 0.315$ and $\Omega_R = 9.18 \times 10^{-5}$. Then we have

$$t_{EQ} = 50 \text{ 300 yr}. \quad (13.37)$$

²²⁹⁾For numerical calculations, we use the central values given in the 2014 PDG: $\Omega_\Lambda = 0.685$, $\Omega_M = 0.315$, $\Omega_\gamma = 5.46 \times 10^{-5}$, and $h = 0.673$. The contribution of the radiation energy with Ω_γ of the order of 10^{-5} can be neglected in many situations. However if it is needed, say in cases of large redshift such as at the last scattering surface $z_L = 1100$, we take neutrinos as massless which give us $\Omega_R = 9.18 \times 10^{-5}$. For comparison we frequently also give results for the WMAP7 data set as given in the 2013 PDG: $\Omega_\Lambda = 0.73$, $\Omega_M = 0.27$, and $h = 0.71$.

²³⁰⁾The integral is not sensitive to the upper limit of integration when it is over 100. The integral is also not sensitive to the presence of Ω_R . We can take $\Omega_R = 0$ and hence $\Omega_\Lambda + \Omega_M = 1$ to integrate z from 0 to the redshift of the last scattering surface $Z_L \approx 1100$. Hence the complication of Ω_R and Ω_M above the LSS can be ignored in the accuracy that concerns us.

13.4.3 Functional form of the scale factor a

The Hubble expansion rate Eq. (13.23) defines the functional relationship between scale factor a and the cosmic time t as follows. Let us define the normalized scale factor

$$\hat{a}(t) \equiv \frac{a(t)}{a_0} = \frac{1}{z+1}, \quad (13.38)$$

where $a_0 \equiv a(t_0)$, t_0 is the time of the present epoch. We can write, setting curvature parameter to zero $K = 0$,

$$\begin{aligned} \frac{H^2}{H_0^2} \sum_j \frac{\rho_j}{\rho_c} &= \sum_j \Omega_j \left[\frac{a_0}{a(t)} \right]^{3(1+w_j)} \\ &= \frac{1}{\hat{a}^4(t)} [\Omega_\Lambda (\hat{a}(t))^4 + \Omega_M \hat{a}(t) + \Omega_R], \end{aligned} \quad (13.39)$$

where the curvature term has been dropped. Equation (13.23) can be rewritten as

$$dt = \frac{1}{H_0} \frac{da}{a \sqrt{\sum_j \frac{\rho_j}{\rho_c}}} = \frac{1}{H_0} \frac{\hat{a} d\hat{a}}{\sqrt{\Omega_\Lambda \hat{a}^4 + \Omega_M \hat{a} + \Omega_R}}. \quad (13.40)$$

This expression relates the cosmic time and the normalized scale factor \hat{a} with the initial condition $\hat{a}(t=0) = 0$. In general one can obtain \hat{a} as a function of t by numerical inversion of the above expression. However, if any one of the of the three densities ratios is set to zero, an analytic result can be obtained. Two interesting cases are $\Omega_R = 0$ and $\Omega_\Lambda = 0$. The latter has been used in the discussion of CMB anisotropy and the former will be used in this chapter in the discussion of cosmic event horizon.

13.4.3.1 Early time scale factor, vacuum energy negligible

Since the contribution of the vacuum energy is constant in the cosmic time given by Ω_Λ , while the contributions of radiation and matter are respectively $\Omega_R(1+z)^4$ and $\Omega_M(1+z)^3$, the effect of the vacuum energy is negligible in the early universe when the redshift is several tens or larger. This is the case in the epoches before the appearance of the first stars, at about $z = 70$. Then Eq. (13.40) can be written as

$$dt \simeq \tau_H \frac{\hat{a} d\hat{a}}{\sqrt{\Omega_M \hat{a} + \Omega_R}}, \tag{13.41}$$

where $\tau_H = H_0^{-1}$. The integration of the above equation is straightforward to give

$$t = \frac{\tau_H}{\sqrt{\Omega_M}} \left[\frac{2}{3} \sqrt{\hat{a} + \hat{a}_{\text{EQ}}} (\hat{a} - 2\hat{a}_{\text{EQ}}) + \frac{4}{3} \hat{a}_{\text{EQ}}^{3/2} \right], \tag{13.42}$$

where \hat{a}_{EQ} is the reduced scale at the matter-radiation equality $\rho_M = \rho_R$,

$$\hat{a}_{\text{EQ}} = \frac{\Omega_R}{\Omega_M}. \tag{13.43}$$

It is straightforward to see that in the radiation dominated region, i.e., $\hat{a} < \hat{a}_{\text{EQ}}$ we have $\hat{a} \sim t^{1/2}$. And in the matter dominated region, i.e., $\hat{a} > \hat{a}_{\text{EQ}}$, $\hat{a} \sim t^{2/3}$. These are the expected behavior of the scale factor in the cosmic time. We can calculate the time of the matter-radiation equality for $\hat{a} = \hat{a}_{\text{EQ}}$. We have

$$\begin{aligned} \bar{a}(\bar{t}) &= 1 - \frac{1}{2} \left[s_1(\bar{t}) + s_2(\bar{t}) - i\sqrt{3}(s_1(\bar{t}) - s_2(\bar{t})) \right] \Theta \left(2 \left(1 + \frac{1}{\sqrt{2}} \right) - \bar{t} \right) + [s_1(\bar{t}) + s_2(\bar{t})] \Theta \left(\bar{t} - 2 \left(1 + \frac{1}{\sqrt{2}} \right) \right), \\ \hat{a}^{(e)}(\bar{t}) &= \hat{a}_{\text{EQ}} \bar{a}(\bar{t}), \end{aligned} \tag{13.48}$$

where the superscript (e) denotes the early time, \hat{a}_{EQ} is given in Eq. (13.43), “ i ” is the usual imaginary symbol, Θ the Heaviside step function, and

$$\begin{aligned} s_1(\bar{t}) &= \sqrt[3]{2f_t^2(\bar{t}) - 1 + 2f_t(\bar{t})\sqrt{f_t^2(\bar{t}) - 1}}, \\ s_2(\bar{t}) &= \sqrt[3]{2f_t^2(\bar{t}) - 1 - 2f_t(\bar{t})\sqrt{f_t^2(\bar{t}) - 1}}, \\ f_t(\bar{t}) &\equiv \left(1 - \frac{1}{\sqrt{2}} \right) \bar{t} - 1. \end{aligned} \tag{13.49}$$

Needless to say, the expressions given in Eq. (13.48) are real. More details on the solution of the cubic equation is given in Section 13.7.5. The complexity in Eq. (13.48) in comparison with Eq. (13.46) is due to fact that in go-

²³¹⁾We have used $\Omega_M = 0.315$ and $\Omega_R = 9.18 \times 10^{-5}$ as before.

$$t_{\text{EQ}} = t(\hat{a}_{\text{EQ}}) = \frac{\tau_H}{\sqrt{\Omega_M}} \frac{4}{3} \left(1 - \frac{1}{\sqrt{2}} \right) \hat{a}_{\text{EQ}}^{3/2}, \tag{13.44}$$

which gives the numerically value $t_{\text{EQ}} \approx 5 \times 10^4$ yr for either set of the cosmological parameters,²³¹⁾ agreeing with the value listed in Table 9.6.

The time-Hubble scale factor relationship, Eq. (13.42) can be further simplified if we normalize the time in units of the time of the matter-radiation equality and the scale factor in units of the scale factor at the equality. Define

$$\begin{aligned} \bar{a} &\equiv \frac{\hat{a}}{\hat{a}_{\text{EQ}}} = \frac{a}{a_{\text{EQ}}}, \quad \bar{t} \equiv \frac{t}{t_{\text{EQ}}}, \\ \bar{a}_{\text{EQ}} &\equiv \bar{a}(t_{\text{EQ}}) = 1. \end{aligned} \tag{13.45}$$

The convenience in defining \bar{a} as a function of \bar{t} is that $\bar{a}(\bar{t}) < 1$ for $\bar{t} < 1$, and $\bar{a}(\bar{t}) > 1$ for $\bar{t} > 1$. Equation (13.42) can be rewritten as

$$\bar{t} = \left(1 + \frac{1}{\sqrt{2}} \right) \left(\sqrt{\bar{a} + 1} (\bar{a} - 2) + 2 \right). \tag{13.46}$$

We can also express \bar{a} as a function of \bar{t} . Rearranging the above expression Eq. (13.46) as a cubic equation of \bar{a} :

$$\bar{a}^3 - 3\bar{a}^2 + 4 - 4 \left[\left(1 - \frac{1}{\sqrt{2}} \right) \bar{t} - 1 \right]^2 = 0. \tag{13.47}$$

A linear combination of two of the three roots of this cubic equation gives the needed expression for \bar{a} as a function of \bar{t} . We have the early time solution

ing from the latter to Eq. (13.47) it involve the square of the terms appearing in both sides of the equality symbol. This introduces spurious solutions in Eq. (13.47). Therefore an appropriate linear combinations of the cubic roots is necessary in order to recover the starting information of Eq. (13.46). Some more details about the three roots and the selection of them in obtaining the correct solution for \bar{a} can be found in Section 13.7.5.

Let us rewrite the Hubble expansion rate in a form which finds usages in the discussion of CMB anisotropy in Section 12.9.3. We write

$$\begin{aligned} H &= H_0 \frac{H}{H_0} = H_0 \frac{1}{\hat{a}^2} \sqrt{\Omega_M \hat{a} + \Omega_R} \\ &= H_0 \sqrt{\Omega_M} \left(\frac{1}{\hat{a}_{\text{EQ}}} \right)^{3/2} \left(\frac{1}{\bar{a}} \right)^2 \sqrt{\bar{a} + 1}, \end{aligned} \tag{13.50}$$

where $\bar{a} \equiv \bar{a}(\bar{t})$ is given in Eq. (13.48). And from Eq. (13.50) and recalling $\bar{a}_{\text{EQ}} = 1$, we have

$$aH = a_0 \hat{a}_{\text{EQ}} \bar{a} H = a_0 H_0 \sqrt{\frac{\Omega_M}{\hat{a}_{\text{EQ}}}} \frac{\sqrt{\bar{a} + 1}}{\bar{a}},$$

$$(aH)_{\text{EQ}} = a_0 H_0 \sqrt{\frac{2\Omega_M}{\hat{a}_{\text{EQ}}}}. \tag{13.51}$$

Then

$$\frac{aH}{(aH)_{\text{EQ}}} = \frac{\sqrt{\bar{a}(\bar{t}) + 1}}{\sqrt{2\bar{a}(\bar{t})}}. \tag{13.52}$$

We note that this function behaves like $\bar{t}^{-1/2}$ for $\bar{t} \ll 1$ and like $\bar{t}^{-1/3}$ for $\bar{t} \gg 1$ as expected. The coefficients in front of the power terms are not exactly constant but close to unity. The coefficient in front of $\bar{t}^{-1/2}$ varies monotonically from 0.80 to 1 as \bar{t} varies from 0 to 1, and the coefficient in front of $\bar{t}^{-1/3}$ varies also monotonically from 1 to 0.85 as \bar{t} varies from 1 to ∞ .²³²⁾ So, this function takes approximately the following form, with an error no more than 20% within the whole range of the cosmic time,

$$\frac{aH}{(aH)_{\text{EQ}}} \approx \bar{t}^{-1/2} \Theta(1 - \bar{t}) + \bar{t}^{-1/3} \Theta(\bar{t} - 1), \tag{13.53}$$

The range of validity of the given form of the scale factor from the time of nucleosynthesis to the photon last scattering. We have used Eq. (13.53) in Section 12.9.3 in the discussion of CMB.

13.4.3.2 Late time scale factor, radiation energy negligible

This is the case of late universe when the redshift is a few or less, say, after the the start of the structure formation

at roughly $z = 7$. In this late period $\Omega_R(1+z)^4 \ll 1$ the radiation contribution can be neglected. The vacuum energy becomes important. we an write,

$$dt = \tau_H \frac{\sqrt{\hat{a}} d\hat{a}}{\sqrt{\Omega_\Lambda \hat{a}^3 + \Omega_M}}. \tag{13.54}$$

Here we can set $\Omega_\Lambda + \Omega_M = 1$. The integral can be performed straightforwardly by a variable change²³³⁾ so as to obtain

$$\hat{a}^{(l)}(t) \equiv \frac{a(t)}{a(t_0)} = \left(\frac{1 - \Omega_\Lambda}{\Omega_\Lambda}\right)^{1/3} \sinh^{2/3} \left(\frac{3\sqrt{\Omega_\Lambda}}{2} \frac{t}{\tau_H}\right), \tag{13.55}$$

where $\tau_H = 1/H_0$ and the superscript (l) means late time to distinguish it from the early time expression of Eq. (13.48). It is clear that the first derivative of \hat{a} is positive, $\hat{a}^{(l)} > 0$, representing an expanding universe. The second derivative of $\hat{a}(t)$ can be shown to vanish at

$$t_{\text{AV}} = \frac{2}{3\sqrt{\Omega_\Lambda}} \ln \left(\frac{1 + \sqrt{3}}{\sqrt{2}}\right) \tau_H. \tag{13.56}$$

For $t < t_{\text{AV}}$, $\ddot{a} < 0$ the expansion of the universe decelerates and for $t > t_{\text{AV}}$, $\ddot{a} > 0$ the universe expansion accelerates. Roughly, $t_{\text{AV}} \approx \tau_H/2$.²³⁴⁾ It is also straightforward to see that for $\Omega_\Lambda = 0$, i.e., for a matter dominated universe, Eq. (13.55) leads to $\hat{a}^{(l)}(t) \sim t^{2/3}$. This is the expected behaviors of the Hubble scale factor $a(t)$ for a matter dominated universe.

The scale factor normalized at the matter-radiation equality as a function of the normalized time $\bar{t} = t/t_{\text{EQ}}$ is given by

$$\bar{a}^{(l)}(\bar{t}) = \frac{\hat{a}(t)}{\hat{a}_{\text{EQ}}} = \frac{\Omega_M^{4/3}}{\Omega_R(1 - \Omega_M)^{1/3}} \sinh^{2/3} \left[(2 - \sqrt{2}) \frac{\Omega_R^{3/2} \sqrt{1 - \Omega_M} \bar{t}}{\Omega_M^2} \right] \tag{13.57}$$

Inverting the above expression Eq. (13.55), we have

$$t = \frac{2\tau_H}{3\sqrt{\Omega_\Lambda}} \operatorname{arcsinh} \left(\sqrt{\frac{\Omega_\Lambda}{1 - \Omega_\Lambda}} \hat{a}^3 \right). \tag{13.58}$$

²³²⁾Equation (13.46) gives $\bar{a} = (4/\sqrt{6})(1 - 1/\sqrt{2})^{1/2} \bar{t}^{1/2}$ and $aH/(aH)_{\text{EQ}} = (\sqrt{6}/4)(1 + 1/\sqrt{2})^{1/2} \bar{t}^{-1/2}$ for $\bar{t} \ll 1$, and $\bar{a} = (2(1 - 1/\sqrt{2})^{2/3} \bar{t}^{2/3})$ and $aH/(aH)_{\text{EQ}} = (1/\sqrt{2})(1 + 1/\sqrt{2})^{1/3} \bar{t}^{-1/3}$ for $\bar{t} \gg 1$.

²³³⁾A change of variable $y = (\hat{a})^{3/2}$ gives $dt = (2/(3\sqrt{H_0})) dy / \sqrt{y^2 + (1 - \Omega_\Lambda)/\Omega_\Lambda}$ which allows one to obtain easily the result given in Eq. (13.55).

²³⁴⁾ $t_{\text{AV}} = 0.530\tau_H$ for $\Omega_\Lambda = 0.685$, and $t_{\text{AV}} = 0.514\tau_H$ for $\Omega_\Lambda = 0.73$.

For $t \leq t_0$, we can set $\hat{a} = (1+z)^{-1}$ to write

$$t(z) = \frac{2\tau_H}{3\sqrt{\Omega_\Lambda}} \operatorname{arcsinh} \left(\sqrt{\frac{\Omega_\Lambda}{1 - \Omega_\Lambda}} (1+z)^{-3} \right). \tag{13.59}$$

From $\hat{a}(t_0) = 1$ or setting $z = 0$, we have

$$t_0 \equiv t(0) = \frac{2\tau_H}{3\sqrt{\Omega_\Lambda}} \operatorname{arcsinh} \left(\sqrt{\frac{\Omega_\Lambda}{1 - \Omega_\Lambda}} \right), \tag{13.60}$$

which is just the age of the universe. The numerical values of Eq. (13.60) reproduce τ_U of Eq. (13.34) for both sets of the cosmological parameters of $\Omega_\Lambda = 0.73$ and $h = 0.71$, and $\Omega_\Lambda = 0.685$ and $h = 0.673$.

13.4.3.3 Connecting the early time and late time scale factors

The expressions of the scale function $\hat{a}^{(e)}$ in Eqs. (13.48) and $\hat{a}^{(l)}$ in (13.55) are valid in different time regimes, the latter near and includes the present epoch, and the former around the matter-radiation equality. The two differ significantly at very early time $t \ll t_{\text{EQ}}$ and very late time $t \gg t_{\text{EQ}}$. However, there is a sizable time region in between, in which matter density dominates over both radiation and vacuum energy. There the two scale factors are practically the same. This can be verified by a straightforward numerical comparison of the two. So we can smoothly joint the two expressions to obtain an analytic expression for the normalized scale function which is valid over the cosmic time, practically the entire homogeneous and isotropic background universe, except for the every early time period, from the time of nucleosynthesis of $t \approx 200$ s. to the present of $t = 13.8$ Byr.

It is convenient to express the normalized scale factors in the scaled time $\bar{t} = t/t_{\text{EQ}}$. Let us rewrite $\hat{a}^{(l)}$ of Eq. (13.55) as a function of the scaled time:

$$\begin{aligned} \hat{a}^{(l)}(\bar{t}) &= \left(\frac{\Omega_M}{1-\Omega_M}\right)^{1/3} \sinh^{2/3}\left(\frac{3\sqrt{1-\Omega_M} t_{\text{EQ}}}{2} \frac{t}{\tau_H t_{\text{EQ}}}\right) \\ &= \left(\frac{\Omega_M}{1-\Omega_M}\right)^{1/3} \sinh^{2/3}\left((2-\sqrt{2})\frac{\Omega_R^{3/2}\sqrt{1-\Omega_M}}{\Omega_M^2}\bar{t}\right). \end{aligned} \quad (13.61)$$

We take the merging point of the two normalized scale factors at the time of vacuum energy-radiation equality when the universe is about $t_{\text{AR}} \approx 540$ Myr. The time of matter-radiation equality is $t_{\text{EQ}} \approx 50,000$ yr. The scaled time at the merging point is

$$\bar{t}_{\text{MG}} = t_{\text{AR}}/t_{\text{EQ}} \approx 1.1 \times 10^4. \quad (13.62)$$

We write the overall normalized scale factor as

$$\begin{aligned} \hat{a}(\bar{t}) &= \hat{a}^{(e)}(\bar{t})\Theta(\bar{t}_{\text{MG}} - \bar{t}) + \hat{a}^{(l)}(\bar{t})\Theta(\bar{t} - \bar{t}_{\text{MG}}) \\ \bar{a}(\bar{t}) &= \bar{a}^{(e)}(\bar{t})\Theta(\bar{t}_{\text{MG}} - \bar{t}) + \bar{a}^{(l)}(\bar{t})\Theta(\bar{t} - \bar{t}_{\text{MG}}). \end{aligned} \quad (13.63)$$

The range of validity is from the nucleosynthesis at the cosmic time of about 200 s. to the present of 13.8 Byr.

13.5 Proper distance and related length measures

13.5.1 Proper distance, d_p , comoving distance d_C , and the Hubble expansion

The *proper distance*²³⁵⁾ is the distance between two points measured on the spatial geodesic along a hypersurface of constant cosmic time. It is the distance from

²³⁵⁾See, [218], p. 415; [221], p. 4; [74], p. 36; and [225], p. 100. We note that in some of the newer texts, proper distance is no long defined, or used differently. For the latter see [224], and for the former see, e.g., [223].

an observer to the redshift z_1 measured at a cosmic time t . In the FLRW metric, the proper distance between an observer and a distant galaxy is defined on a spatial surface of $dt = 0$. In particular, when the observer is defined to be the origin of the comoving frame, the angular variables of the path of the measurement are fixed, so that $d\Omega = 0$, we have²³⁶⁾

$$ds_{//} = \sqrt{g_{rr}(t)}dr = \frac{a(t)}{\sqrt{1-\kappa r^2}}dr = a(t)d\chi, \quad (13.64)$$

where g_{rr} is given in Eq. (9.22) or can be read off from Eq. (13.1). For an observer at the origin, the *proper distance* at some time t of a galaxy located at the time-independent comoving radial coordinates r_1 of redshift z_1 is given by

$$\begin{aligned} d_p(t, r_1) &\equiv a(t) \int_0^{r_1} \frac{dr}{\sqrt{1-\kappa r^2}} = a(t) \int_0^{\chi_1} d\chi \\ &= a(t)\chi_1 = a(t) \frac{1}{\sqrt{\kappa}} \arcsin(\sqrt{\kappa}r_1). \end{aligned} \quad (13.65)$$

We should emphasize that no physical measurement can be carried out for the proper distance. In order to make the measurement so defined it has to be carried out infinitely fast or to freeze the expansion of the universe at the time of the measurement, so that the cosmic scale factor stays at a constant value during the measurement.²³⁷⁾

A related distance is the *comoving distance*, which is the proper distance with t being evaluated at the time when the light signal reaches the observer, which is the present time $t = t_0$. From Eq. (13.67), putting $z = 0$, we have²³⁸⁾

$$d_c(r_1) = d_p(t_0, r_1) = a_0\chi_1 = \frac{a_0}{\sqrt{\kappa}} \arcsin(\sqrt{\kappa}r_1). \quad (13.66)$$

In the case that the Hubble expansion rate is known, such as the Λ CDM, the proper distance and the comoving distance can be evaluated using the null path of light propagation using Eqs. (13.21) and (13.25). Let a light beam be emitted at t_1 from the galaxy with the redshift z_1 reaches the observer at t_0 , and the scale factor is $a(t) = a_0/(1+z)$, we can rewrite Eq. (13.65) as

²³⁶⁾This is the line element on the space-like surface, not to be confused with the FLRW metric line element Eq. (A17) which defines the proper time.

²³⁷⁾A way to imagine the measurement of the proper distance is given in [218], p. 415. Let us reproduce the argument below: Let us set up a chain of measuring stations closely spaced along the line of sight between the two galaxies in question. Observers at each station will measure the distance to their respective next stations at the same time t , say, by measuring the traveling times of their respective light signals. The proper distance in question is obtained by summing up all these small distance measurements, with the small space between neighboring measuring stations going to zero. Such a measurement cannot be carried out in a real experiment.

²³⁸⁾The comoving distance is not uniquely defined. In some of the literature, it is defined simply as χ_1 of Eq. (13.4).

$$\begin{aligned}
 d_p(t, r_1) &= a(t)c \int_{t_1}^{t_0} \frac{dt'}{a(t')} = a(t)c \int_{a_1}^{a_0} \frac{da}{\dot{a}} = \frac{a(t)c}{H_0} \int_{a_1}^{a_0} \frac{1}{\tilde{E}} \frac{da}{a^2} \\
 &= \frac{D_H}{1+z(t)} \int_0^{z_1} \frac{dz}{\sqrt{\Omega_\Lambda + \Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_\kappa(1+z)^2}}, \\
 d_c(r_1) &= D_H \int_0^{z_1} \frac{dz}{\sqrt{\Omega_\Lambda + \Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_\kappa(1+z)^2}}, \tag{13.67}
 \end{aligned}$$

where $z(t_0) = 0$.

Although in the general setting of arbitrary t the proper distance has no real relevance to observes, it can be used for the discussion of issues related to the expanding universe [315]. We can use the proper distance to derive the Hubble expansion law as follows. Let us take the time derivative of the proper distance Eq. (13.65), the recession velocity at a general comoving radial distance and an arbitrary time is given by

$$\dot{d}_p(t, r_1) = \dot{a}(t)\chi_1(r_1) = H(t)d_p(t, r_1), \tag{13.68}$$

i.e., the Hubble recession velocity for the coordinate point r_1 at any time t is proportional to the corresponding proper distance times the Hubble expansion rate at the given time. In particular, at the present epoch for small $z_1 \ll 1$, we have, from Eq. (13.68),

$$\dot{d}_p(t_0, r_1)|_{z_1 \ll 1} \approx H_0 D_H z_1 = cz_1. \tag{13.69}$$

As stated earlier, the meaning of the Hubble length can be further illustrated by considering the galaxy expansion velocity. Let us take the Hubble recession velocity Eq. (13.68) for the coordinate point r_1 at the present epoch and setting the recession velocity at the light velocity $\dot{d}_p(t_0, r_1) = c$. Then

$$\begin{aligned}
 d_p(t_0, r_1)|_{\dot{d}(t_0, r_1)=c} &= d_c(r_1)|_{\dot{d}_c(r_1)=c} = \frac{c}{H_0} = D_H. \\
 \end{aligned} \tag{13.70}$$

Hence the Hubble length is the comoving distance which has the recession speed equal to the speed of light.

Within the Hubble length all the recession velocity is subluminal and beyond the Hubble length all superluminal. Having superluminal recession velocity for galaxies does not violate the fundamental principle of special relativity.²³⁹⁾ In the Λ CDM model, as can be calculated from Eq. (13.67), the redshift value for the subluminal and superluminal boundary happens at $z = 1.41$ for $\Omega_\Lambda = 0.73$ and $h = 0.71$, and $z = 1.48$ for $\Omega_\Lambda = 0.685$ and $h = 0.673$.

13.5.2 Particle horizon d_{ph}

The *particle horizon*²⁴⁰⁾ is the maximal distance from which an observer, on Earth for instance, can receive a signal emitted from an object in the past. It is also referred to simply as the *horizon*.²⁴¹⁾ Because of the finiteness of the speed of light and the age of the universe, the particle horizon of an observer is finite and will increase as the cosmic time increases. The particle horizon at the present epoch is determined by the upper limit of the allowed comoving distance. Hence the particle horizon of the present epoch is given by

$$\begin{aligned}
 d_{ph} \equiv d_c(r_{\max}) &= a_0 \int_0^{r_{\max}} \frac{dr}{\sqrt{1 - \kappa r^2}} = a_0 c \int_0^{t_0} \frac{dt}{a(t)} \\
 &= D_H \int_0^{z_{\max}} \frac{dz}{\tilde{E}_z}, \tag{13.71}
 \end{aligned}$$

where r_{\max} is the maximal value of r_1 and $z_{\max} \rightarrow \infty$ the value of the redshift in the past corresponding to r_{\max} . We have

$$\begin{aligned}
 d_{ph} &= D_H \int_0^\infty \frac{dz}{\sqrt{\Omega_\Lambda + \Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_\kappa(1+z)^2}} \\
 &= c\tau_U = ct_0 = \begin{cases} 3.45D_H = 4.49 \times 10^{26} \text{ m,} & \text{for } \Omega_\Lambda = 0.73, \quad h = 0.71, \\ 3.24D_H = 4.45 \times 10^{26} \text{ m,} & \text{for } \Omega_\Lambda = 0.685, \quad h = 0.673, \end{cases} \tag{13.72}
 \end{aligned}$$

where Ω_R and Ω_κ are ignored and Ω_M is set to be $1 - \Omega_\Lambda$. The difference of the two set of data is about 5%.

The particle horizon will increase with time. For observers on earth, more and more portion of the universe becomes visible as the age of the universe increases, because longer time allows further light to reach earth.

²³⁹⁾For a clear accessible discussion, see [315] and [316].

²⁴⁰⁾[218], p. 489; [74], p. 36. The name *particle horizon* together with the term *event horizon* was coined by Rindler in 1956, [317]. See [218], p. 490.

²⁴¹⁾See [240], p. 93. It is also referred to as the cosmological horizon or the light horizon.

13.5.3 Cosmic event horizon

The *cosmic event horizon* or simply the *event horizon* defines the largest distance from which a light signal is emitted and can eventually reach an observer in the distant future. This distance may be finite or infinite, depending on the cosmological models that determine the large time behavior of the scale factor $a(t)$. In some models, including the present Λ CDM, there exists a finite radial distance, denoted as χ_{eh} or r_{eh} , called *cosmic event horizon* or *event horizon*, beyond which no object, such as a galaxy, can ever be visible to the observer.²⁴²⁾

Consider a photon source which emits a signal at t_{em} which is received by an observer at t_{obs} . Another signal is emitted at t'_{em} and received by the same observer at t'_{obs} . In the comoving frame in which the observer is at the coordinate origin, the time independent comoving radial distance of the photon source, denoted as χ_1 or r_1 , is given by

$$\begin{aligned} \chi_1 &= \frac{1}{\sqrt{\kappa}} \arcsin(\sqrt{\kappa}r_1) \\ &= c \int_{t_{\text{em}}}^{t_{\text{obs}}} \frac{dt}{a(t)} \\ &= c \int_{t'_{\text{em}}}^{t'_{\text{obs}}} \frac{dt}{a(t)}. \end{aligned} \tag{13.73}$$

We shall assume that the source and observer can exist forever, so that their relationship in the cosmic time evolution can be explored, not being intruded by the complications of the finite lifetime of either the source or the observer.

Since $a(t)$ is non-negative, χ_1 increases with t_{obs} for a given t_{em} . χ_1 also increases for a given t_{obs} when t_{em} decreases. These state that the observer's visible universe increases with the time of observation, and with the signal emission going back to an earlier time. For a given t_{em} , the maximal range of observation is reached when $t_{\text{obs}} \rightarrow \infty$. This gives rise to the quantity called event horizon. If the integral in (13.73) diverges for $t_{\text{obs}} \rightarrow \infty$, then all parts of the universe will eventually be visible to the observer. So there does not exist an event horizon. However, if the integral converges for $t_{\text{obs}} \rightarrow \infty$, there is a maximal $\chi_1(t_{\text{em}})$, which is the *event horizon* denoted as $\chi_{\text{eh}1}$, beyond which no object can ever be in causal contact with the said observer. In plain words, what the event horizon means is that for all light signals emitted at the time t_1 , the furthestmost source is located at the comoving radial coordinate $r_{\text{eh}}(t_{\text{em}})$, beyond which an electromagnetic signal emitted at t_{em} will never be able to reach the observer. The criterion for the existence of an event horizon is quite straightforward. From

²⁴²⁾The cosmic event horizon should not be confused with the black hole event horizon, although the two have some similarities.

Eq. (13.73) we have:

$$\frac{a(t)}{t} \xrightarrow{t \rightarrow \infty} \begin{cases} 0, & \text{no finite event horizon,} \\ \neq 0, & \text{existence of a finite event horizon.} \end{cases} \tag{13.74}$$

We will see later that in the cosmological standard model of Λ CDM the cosmic event horizon exists. For a given t_{em} it is,

$$\chi_{\text{eh}}(t_{\text{em}}) = c \int_{t_{\text{em}}}^{\infty} \frac{dt}{a(t)}. \tag{13.75}$$

There is also an overall event horizon beyond which no light signal will be able to reach an observer. The overall event horizon is obtained by setting $t_{\text{em}} = 0$ in the expression of the event horizon

$$\chi_{\text{eh}}^{(\text{OA})} = c \int_0^{\infty} \frac{dt}{a(t)}. \tag{13.76}$$

Further discussions and a concrete evaluation of the event horizon in the current Λ CMD is given in Section 13.7.3.

13.6 Physical distance measures²⁴³⁾

13.6.1 Angular diameter distance d_A and comoving angular diameter distance d_{cA} ²⁴⁴⁾

In analogues to distances along the line of sight which can be represented by the radial variable r or χ as given in Eq. (13.64), we can also define distances or sizes transversal to the line of sight of an (Earth) observer. Let an object of redshift z_1 be located at the comoving coordinate r_1 and ϕ_1 and having an finite extension in the θ variable $d\theta$. The transverse distance element is defined by

$$\begin{aligned} ds_{\perp} &= \sqrt{g_{\theta\theta}}d\theta = a(t)r_1d\theta = a(t)\frac{1}{\sqrt{\kappa}}\sin(\sqrt{\kappa}\chi_1)d\theta \\ &= \frac{a(t)}{\sqrt{\kappa}}\sin\left(\sqrt{\kappa}\frac{D_{\text{H}}}{a_0}\right) \\ &\times \int_0^{z_1} \frac{dz}{\sqrt{\Omega_{\Lambda} + \Omega_{\text{M}}(1+z)^3 + \Omega_{\text{R}}(1+z)^4 + \Omega_{\kappa}(1+z)^2}} d\theta, \end{aligned} \tag{13.77}$$

where $t_{\theta\theta}$ is given in Eq. (9.22).

²⁴³⁾Presently there are four approaches for distance measures. For near objects there are the parallax and proper motion approaches. For objects outside the Milky Way there are measurements of their apparent vs absolute luminosities, and measurements of their angular vs true diameter. These four distance measurements give similar results for objects that are no farther than 10^9 light years away which correspond to redshifts of $z < 2$. Beyond this range they differ from one another and also from the "proper distance" discussed above.

²⁴⁴⁾See, [218], p. 422, and [101], p. 35

Suppose that an object has a transversal size d_{tr} . Light signals are emitted from the opposite ends of the object

at time t_1 and reach an observer at time t_0 , forming an angle $\delta\theta$. We define the *angular diameter distance* d_A ,

$$d_A(r_1) = \frac{ds_{\perp}}{\delta\theta} \Big|_{t_1} = a(t_1)r_1 = a(t_1) \frac{1}{\sqrt{\kappa}} \sin(\sqrt{\kappa}\chi_1),$$

$$= \frac{a(t_1)}{\sqrt{\kappa}} \sin \left(\sqrt{\kappa} \frac{D_H}{a_0} \int_0^{z_1} \frac{dz}{\sqrt{\Omega_{\Lambda} + \Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_{\kappa}(1+z)^2}} \right), \tag{13.78}$$

where the time variable is evaluated at t_1 .

In the discussion of CMB anisotropy we need the angular diameter distance at the last scattering surface $d_A^{(L)}$. It is given by setting $t_1 = t_L$ in Eq. (13.78). Taking $\kappa = 0$ and using the fact $a_L/a_0 = (1+z_L)^{-1}$, we have, for $z_L = 1100$,

$$d_A^{(L)} = a_L r_L = \frac{D_H}{1+z_L} \int_0^{z_L} \frac{dz}{\sqrt{\Omega_{\Lambda} + \Omega_M(1+z)^3 + \Omega_R(1+z)^4}}$$

$$= \frac{H_0^{-1}}{1+z_L} \begin{cases} 3.32, & \Omega_{\Lambda} = 0.73, & \Omega_M = 0.27, \\ 3.12, & \Omega_{\Lambda} = 0.685, & \Omega_M = 0.315. \end{cases} \tag{13.79}$$

Another situation is the distance between two objects, say galaxies, located on a line transverse to the earth line of sight. The above discussion also applies. In this case the distance between the two galaxies expands with the cosmic flow and it is appropriate to evaluate the distance at the time t_0 , i.e., the cosmic time of the observer. This defines the *comoving angular diameter distance*,

$$d_{cA} = a_0 r_1 = \frac{a_0}{a(t_1)} a(t_1) r_1 = (1+z_1) d_A$$

$$= \frac{a_0}{\sqrt{\kappa}} \sin \left(\sqrt{\kappa} \frac{D_H}{a_0} \int_0^{z_1} \frac{dz}{\sqrt{\Omega_{\Lambda} + \Omega_M(1+z)^3 + \Omega_R(1+z)^4 + \Omega_{\kappa}(1+z)^2}} \right). \tag{13.80}$$

This is also known as the *proper-motion distance* or the *transverse comoving distance* [293]. It is similar to the comoving distance defined in Eq. (13.66). In a flat universe the transverse comoving distance is identical to the comoving distance Eq. (13.67) and its numerical evaluation can be done straightforwardly.

13.6.2 The luminosity distance²⁴⁵⁾ d_L

The normal way to calculate the rate of energy received by an observer from a light source is the of total luminosity L divided by $4\pi d^2$, and multiplying the effective area of the detector δA , where d is the distance between the light source and the observer, i.e., $L\delta A/(4\pi d^2)$. In the expanding universe, however, this is more complicated. From the point of view of the light source the radiation fronts form various spherical shells around it. The distance to be calculated will involve the metric $g_{\theta\theta}$ and the distance will be related to the angular diameter distance. However, for an ever expanding universe, complications are introduced by the expansion taking place during the light traveling between the light source and the observer: One complication is the effect of the expansion of the distance and cosmic time between the light

source and the observer. The other is the shift of the frequency of the photon emitted by the distant source to a lower frequency, and hence to a lower energy, when it arrives at the observer. Taking these complications into consideration, we define the apparent flux measured by an observe f_A and a *luminosity distance* d_L ,

$$f_A \equiv \frac{L}{4\pi d_L^2}, \tag{13.81}$$

where L is the true luminosity of the galaxy at the time the photon signal is emitted. Below we relate the luminosity distance with the angular diameter distance.

Let the observer be the origin of the comoving coordinate and the comoving radial coordinate of the light emitter be r_1 of redshift z_1 . To obtain the luminosity distance we have to consider three factors due to Hubble expansion: (1) When the light signal reaches the observer, the Hubble expansion stretches the wavelength by a factor $1+z_1$ and hence lowers the measured energy by the same factor. (2) Since the luminosity and flux related to energy density per unity time, we have to take into account of the effect of the Hubble expansion on the time. The time interval measured at r_1 is stretched by a factor $1+z_1$. The same time stretch factor has already been considered in Section 9.1.4, Eq. (9.60). So

²⁴⁵⁾See, [218], pp 418–421 and Kolb and Turner, p. 41.

the flux will be reduced by another factor $1 + z_1$. (3) Furthermore, the (angular diameter) distance from the galaxy to the observer is expanded from $d_A = a(t_1)r_1$ to $d_{cA} = a(t_0)r_1 = (1 + z)d_A$, which introduces another suppression factor $(1 + z_1)^2$ in the measured radiation flux. Then in relation to the angular diameter distance which is defined as the distance at the time of emission of the light signal, the luminosity is suppressed by a factor $(1 + z_1)^4$. Hence we have the *luminosity distance*, which relates the true luminosity of the source at its emission and the measured flux by the observer, to be

$$\begin{aligned} d_L &= (1 + z_1)d_{cA} = (1 + z_1)^2 d_A \\ &= \frac{a_0}{a(t_1)} d_{cA} = \left(\frac{a_0}{a(t_1)} \right)^2 d_A = \frac{a_0^2}{a(t_1)} r_1, \end{aligned} \quad (13.82)$$

where r_1 is given in Eq. (13.27), d_{cA} in Eq. (13.80), and d_A in Eq. (13.78).

13.7 Recapitulating results in a flat universe $\kappa = 0$

There are strong evidences that the universe is flat so that the space-time curvature constant $K = 0$ and therefore the vanishing of the curvature energy density $\Omega_K = 0$, which gives rise to simplifications to most expressions related to universe expansion.

13.7.1 Summary of results

Line element:

$$ds^2 = c^2 dt^2 - a^2(t)(dr^2 + r^2 d\Omega^2). \quad (13.83)$$

Null path:

$$\frac{cdt}{a(t)} = dr = d\chi. \quad (13.84)$$

Comoving radial coordinate:

$$\chi_1 = r_1 = \frac{D_H}{a_0} \int_0^{z_1} \left(\frac{1}{\tilde{E}_z} \right)_{K=0} dz. \quad (13.85)$$

Distance measures:

$$\begin{aligned} d_p(t_1, r_1) &= d_A(r_1) = a(t_1)r_1 \\ &= \frac{D_H}{1 + z} \int_0^{z_1} \left(\frac{1}{\tilde{E}_z} \right)_{K=0} dz, \\ d_c(r_1) &= d_{cA}(r_1) = a_0 r_1 = a_0 \chi_1 \\ &= D_H \int_0^{z_1} \left(\frac{1}{\tilde{E}_z} \right)_{K=0} dz, \\ d_L(r_1) &= (1 + z_1)d_{cA} = (1 + z_1)^2 d_A \\ &= (1 + z_1)D_H \int_0^{z_1} \left(\frac{1}{\tilde{E}_z} \right)_{K=0} dz, \\ d_{ph} &= D_H \int_0^\infty \left(\frac{1}{\tilde{E}_z} \right)_{K=0} dz = 3.502 D_H, \end{aligned} \quad (13.86)$$

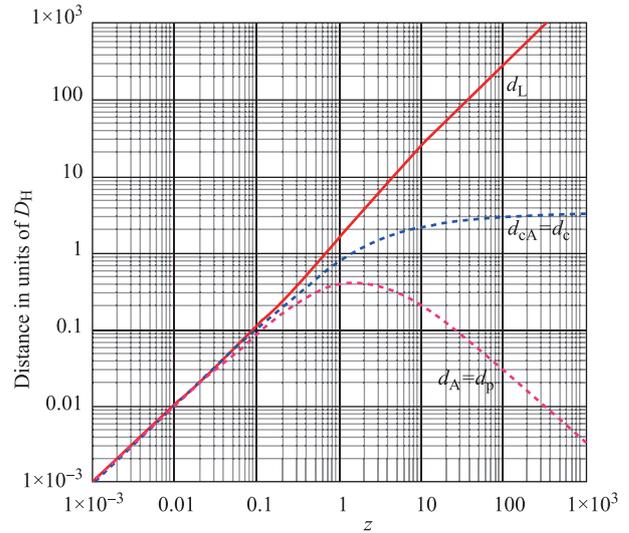


Fig. 13.1 Distances in units of the Hubble length D_H as functions of the redshift z .

where \tilde{E}_z is given in Eq. (13.24).

We plot in Fig. 13.1 the three distances in the flat space $d_p(z_1, r_1) = d_A(r_1)$, $d_c(r_1) = d_{cA}(r_1)$, and $d_L(r_1)$ as functions of the redshift for the case $\Omega = 0.685$. For small redshift all the distances are about the same. They deviate from one another for $z > 0.1$. The luminosity distance increases linearly with the redshift. The comoving and comoving angular diameter distance flattens out for $z > 100$. The proper distance and angular diameter distance reaches a maximum at about $z \approx 1.6$ then decreases as z increases.

13.7.2 Explicit examples: LSS and later epoches

LSS (the last scattering surface)

At the last scattering surface of $z_1 = 1090$ we have

$$\begin{aligned} d_A^{(L)} &= d_p(z_L, r_L) = \frac{d_{cA}}{1 + z} = 13.4 \text{ Mpc}, \\ d_{cA} &= d_c = a_0 r_1 = a_0 \chi_1 = 14.580 \text{ Gpc}, \\ d_L &= (1 + z_L)d_{cA} = 1.59 \times 10^4 \text{ Gpc}. \end{aligned} \quad (13.87)$$

13.7.3 Cosmic event horizon and a sketch of the longterm cosmic future²⁴⁶⁾

To calculate the event horizon we need to know the behavior of the integral in Eq. (13.73) when $t_{\text{obs}} \rightarrow \infty$. We rewrite the expression of Eq. (13.25) by setting $K = 0$ and dropping the contribution of the radiation energy. So we assume that the universe is saturated by matter and the dark energy which says $\Omega_\Lambda + \Omega_M = 1$. This is a good approximation for epoches after the LSS. and the

²⁴⁶⁾Some of the discussion presented here below follows that given in [318] and to which we refer for more details.

case discussed in Section 13.4.3. The reduced scale factor can be expressed in analytic form in the cosmic time as given in Eq. (13.55).

For an exponentially expanding universe at large t as indicated in Eq. (13.61), the integral in Eq. (13.73) converges as $t_{\text{obs}} \rightarrow \infty$, hence there exists a finite cosmic event horizon for a given t_{em} . All presently observable galaxies, except for those gravitationally bound to that of the observer, will sooner or later move out of the event horizon, after which their electromagnetic signals will no longer be able to reach the said observer. So any of those galaxies can only be monitored for a finite period in their lifetime. Let us calculate the event horizon for of a source object sending out a signal at t_{em} which reaches an Earth observer at $t_{\text{ob}} \rightarrow \infty$. We denote the comoving radial coordinate of the source object by d_{eh} . For the case that the emission time lies in the past, i.e., $t_{\text{em}} < t_0$ we can express the cosmic time in terms of the redshift. Denoting the redshift of the source object by z_{em} , we have

$$\begin{aligned} d_{\text{eh}}(t_{\text{em}}) &= a_0 r_{\text{eh}}(t_{\text{em}}) = c \int_{t_{\text{em}}}^{\infty} \frac{dt}{\hat{a}(t)} \\ &= \frac{2}{3} D_{\text{H}} \left[\frac{1}{\Omega_{\Lambda}(1 - \Omega_{\Lambda})^2} \right]^{1/6} \int_{y_{\text{em}}}^{\infty} \frac{dy}{[\sinh(y)]^{2/3}}, \\ y_{\text{em}} &= \frac{3}{2} \sqrt{\Omega_{\Lambda}} \frac{t_{\text{em}}}{\tau_{\text{H}}} \\ &= \operatorname{arcsinh} \left(\sqrt{\frac{\Omega_{\Lambda}}{1 - \Omega_{\Lambda}}} (1 + z_{\text{em}})^{-3} \right) \Big|_{t_{\text{em}} < t_0}, \end{aligned} \quad (13.88)$$

where $\hat{a}(t)$ is given in Eq. (13.61) and a change of variable of integration has been made.

There are two notable cosmic event horizons. One is the event horizon of the present epoch and the other is the overall event horizon of our standard Λ CDM world. The event horizon of the present epoch, i.e., the remotest comoving radial distance that we are able to reach, which we denote by $d_{\text{eh}}^{(0)}$, is obtained by setting the emission time at the present time, $t_{\text{em}} = t_0$ or $z_{\text{em}} = 0$,

$$\begin{aligned} d_{\text{eh}}^{(0)} &= \frac{2}{3} D_{\text{H}} \left[\frac{1}{\Omega_{\Lambda}(1 - \Omega_{\Lambda})^2} \right]^{1/6} \int_{y_0}^{\infty} \frac{dy}{[\sinh(y)]^{2/3}}, \\ &= \begin{cases} 1.12 D_{\text{H}} = 4.73 \text{ Gpc} = 15.4 \text{ Gyr}, & \text{for } \Omega_{\Lambda} = 0.73, \quad h = 0.71, \\ 1.15 D_{\text{H}} = 5.12 \text{ Gpc} = 16.8 \text{ Gyr}, & \text{for } \Omega_{\Lambda} = 0.685, \quad h = 0.673. \end{cases} \end{aligned} \quad (13.89)$$

The overall event horizon of the Λ CDM cosmos is to set $t_{\text{em}} = 0$ or $z_{\text{em}} \rightarrow \infty$, we have

$$\begin{aligned} d_{\text{eh}}^{(\text{Oa})} &= \frac{2}{3} D_{\text{H}} \left[\frac{1}{\Omega_{\Lambda}(1 - \Omega_{\Lambda})^2} \right]^{1/6} \int_0^{\infty} \frac{dy}{[\sinh(y)]^{2/3}} \\ &= \begin{cases} 4.62 D_{\text{H}} = 19.5 \text{ Gpc} = 63.6 \text{ Gyr}, & \text{for } \Omega_{\Lambda} = 0.73, \quad h = 0.71, \\ 4.38 D_{\text{H}} = 20.0 \text{ Gpc} = 63.9 \text{ Gyr}, & \text{for } \Omega_{\Lambda} = 0.685, \quad h = 0.673. \end{cases} \end{aligned} \quad (13.90)$$

So the cosmic event horizon is finite of limited size. The overall event horizon is about $d_{\text{eh}}^{(\text{Oa})} \approx 64$ Gyr. We should remark that this result is an approximation, because at large redshift z , i.e., small y the integrand is not valid. But we expect the effect to be small.

Let us examine the fate to all the galaxies that are visible to us at the present epoch. Each visible galaxy has a redshift z_1 , where $z_1 \leq z_{\text{L}}$, and the radiation signals we are receiving from it are emitted when the universe has the age τ_{z_1} or the cosmic time t_{z_1} given in Eq. (13.59)

$$\tau_{z_1} = \frac{2}{3\sqrt{\Omega_{\Lambda}}} \operatorname{arcsinh} \left(\left(\frac{1 - \Omega_{\Lambda}}{\Omega_{\Lambda}} (1 + z_1)^3 \right)^{-1/2} \right) \tau_{\text{H}}. \quad (13.91)$$

We denote the comoving radial coordinate of this galaxy by $\chi_1 = r_1$. Let a light pulse be emitted by the source

of redshift z_1 at t_1 and reach the observer at the present epoch t_0 . Consider another pulse from the source which is emitted at $t_{\text{em}} > t_1$ and reaches the observer at $t_{\text{ob}} > t_0$. Both allow us to calculate the same comoving radial coordinate of the source. From Eq. (13.26), we have

$$\begin{aligned} \chi_1 &= \chi(z_1) = c \int_{t_1}^{t_0} \frac{dt}{a(t)} \\ &= \frac{c}{a_0 H_0} \int_0^{z_1} \frac{dz}{\sqrt{\Omega_{\Lambda} + \Omega_{\text{M}}(1+z)^3 + \Omega_{\text{R}}(1+z)^4 + \Omega_{\kappa}(1+z)^2}} \\ &= c \int_{t_{\text{em}}}^{t_{\text{obs}}} \frac{dt}{a(t)}, \end{aligned} \quad (13.92)$$

where t_{obs} varies with t_{em} . The second line above allows the determination of $\chi(z_1)$, and the third line provides the constraint on the time interval during which a source and observer can make causal contact, which lie in the future time for $t_{\text{em}} > t_1$ and $t_{\text{obs}} > t_0$.

For a given χ_1 , t_{obs} increases with t_{em} . Since the integral is convergent as $a(t)$ increase exponentially, there exists a value of t_{em} , called the latest emission time denoted as t_{Lem} , which makes $t_{\text{obs}} \rightarrow \infty$. Then Eq. (13.92) can no longer be satisfied for $t_{\text{em}} > t_{\text{Lem}}$. This is the physical situation which says that the light emitted by the source galaxy after the latest emission time will not be able to reach the observe. For a given z_1 there is an unique t_{Lem} , which is obtained as the solution of the following equation, obtained from the second and third line of Eq. (13.92). Defining $\tilde{t} \equiv t/\tau_{\text{H}}$ and neglecting both Ω_K and Ω_R , we have

$$\int_0^{z_1} \frac{dz}{\sqrt{\Omega_\Lambda + (1 - \Omega_\Lambda)(1+z)^3}} = \int_{\tilde{t}_{\text{Lem}}}^\infty \frac{d\tilde{t}}{\hat{a}(\tilde{t})}$$

$$= \left(\frac{1 - \Omega_\Lambda}{\Omega_\Lambda}\right)^{-1/3} \int_{\tilde{t}_{\text{Lem}}}^\infty \frac{d\tilde{t}}{\left[\sinh\left(\frac{3\sqrt{\Omega_\Lambda}}{2}\tilde{t}\right)\right]^{2/3}}, \quad (13.93)$$

where $\hat{a}(\tilde{t})$ can be obtained from Eq. (13.55). We plot the results in Fig. 13.2, as functions of the redshift of the source object z_1 . The red solid curve is the cosmic time of the last emission t_{Lem} which the observer will receive in the infinite future. We also plot by the dashed cyan curve the age of the universe at the redshift z_1 . Note that the age of the light emission, i.e., the cyan curve determines the comoving radial coordinates of various objects.

Let us elaborate some more, referring to Fig. 13.2. For a radiation source of a given redshift, the light signal emitted at this last emission time will take an infinite amount of time to reach an Earth observer. For example, for a galaxy of $z_1 \neq 0$, an Earth observer can only monitor it for a finite period of time of its evolution. For example, for a galaxy of $z = 5$, the light we receive from it presently was emitted at the cosmic time of $0.082\tau_{\text{H}}$ or 1.1 Gly which is in the early period of the formation of

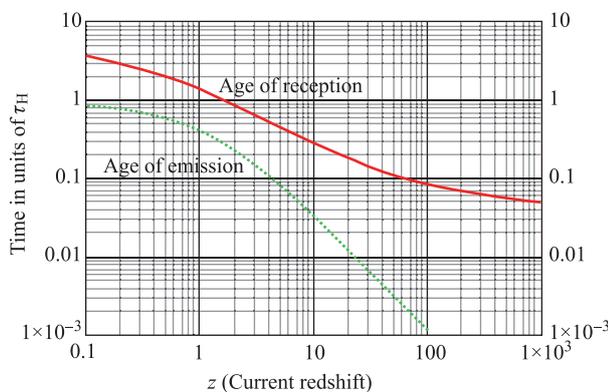


Fig. 13.2 The red solid curve gives the latest emission time of a galaxy at a give redshift z at the present epoch. The cyan dashed curve shows the age of the galaxy when light signals are emitted that are presently reaching the observer.

galaxies.²⁴⁷⁾ The last light we can receive from it is emitted at the time when the universe is $0.46\tau_{\text{H}} = 6.4$ Gly old, and after which we lose causal contact with it and will no long be able to monitor its evolution beyond the last light. Another example is the presently oldest galaxy observed which has a redshift 11.9 [319].²⁴⁸⁾ The light was emitted at about $0.026\tau_{\text{H}} \approx 340$ Mly and the last light was emitted at $0.25\tau_{\text{H}} \approx 3.5$ Gly.

13.7.4 Results for low redshifts

For low redshift $z < 1$, we have, independent of the values of the various energy densities,

$$r_1 \approx \chi_1 \approx z_1 \frac{D_{\text{H}}}{a_0},$$

$$d_A \approx d_L \approx d_P \approx d_{\text{pr}} \approx d_A^{(0)} = a_0 r_1 \approx D_{\text{H}} z_1. \quad (13.94)$$

13.7.5 Remarks on the cubic equation relating time and the Hubble scale factor

The roots of a cubic equation as a textbook material is well known. For a summary, see [291].²⁴⁹⁾ Since there are subtle features in the present case concerning Eqs. (13.46) and (13.47), here we make some remarks on them.

In the notation of [291], the cubic equation is written in the form

$$x^3 + b_2 x^2 + b_1 x + b_0 = 0. \quad (13.95)$$

Referring to Eq. (13.47), we have

$$b_2 \equiv -3, \quad b_1 = 0, \quad b_0 \equiv 4 - 4 \left[\left(1 - \frac{1}{\sqrt{2}}\right) \bar{t} - 1 \right]^2. \quad (13.96)$$

We define

$$q \equiv -\frac{b_2^2}{9} = -1,$$

$$r \equiv -\frac{b_0}{2} - \frac{b_2^3}{27} = 2 \left[\left(1 - \frac{1}{\sqrt{2}}\right) \bar{t} - 1 \right]^2 - 1,$$

$$q^3 + r^2 = -1 + \left\{ 2 \left[\left(1 - \frac{1}{\sqrt{2}}\right) \bar{t} - 1 \right]^2 - 1 \right\}^2,$$

$$s_1 \equiv \sqrt[3]{r + \sqrt{q^3 + r^2}},$$

$$s_2 \equiv \sqrt[3]{r - \sqrt{q^3 + r^2}}. \quad (13.97)$$

²⁴⁷⁾For this calculation we take $\Omega_\Lambda = 0.7$, in between the values we used earlier. We also take $h = 0.7$.

²⁴⁸⁾NASA News Release, Dec. 12, 2012 [319].

²⁴⁹⁾The search for the general solution of the roots a cubic equation has a rich history. A web search provides some interesting readings.

The three roots are given by

$$\begin{aligned} \bar{a}_1 &= s_1 + s_2 - \frac{b_2}{3}, \\ \bar{a}_2 &= -\frac{1}{2}(s_1 + s_2) - \frac{b_2}{3} + \frac{i\sqrt{3}}{2}(s_1 - s_2), \\ \bar{a}_3 &= -\frac{1}{2}(s_1 + s_2) - \frac{b_2}{3} - \frac{i\sqrt{3}}{2}(s_1 - s_2). \end{aligned} \tag{13.98}$$

The sign of $q^3 + r^2$ is a discriminator of the solutions.²⁵⁰⁾ First we need to know the sign and zeros of $q^3 + r^2$:

$$q^3 + r^2 \begin{cases} < 0, & \bar{t} < 4 \left(1 + \frac{1}{\sqrt{2}}\right), \\ = 0, & \bar{t} = 0, \quad 2 \left(1 + \frac{1}{\sqrt{2}}\right), \quad 4 \left(1 + \frac{1}{\sqrt{2}}\right), \\ > 0, & \bar{t} > 4 \left(1 + \frac{1}{\sqrt{2}}\right). \end{cases} \tag{13.99}$$

The root \bar{a}_1 is real. The roots \bar{a}_2 and \bar{a}_3 are real for $\bar{t} < 4(1 + 1/\sqrt{2})$, but complex for $\bar{t} > 4(1 + 1/\sqrt{2})$. Hence the initial condition $\bar{a}(\bar{t} = 0) = 0$ and the defining condition $\bar{t} = 1$ and $\bar{a}(\bar{t} = 1) = 1$ can be used to discriminate among the solutions. It can be checked that these conditions are satisfied by \bar{a}_3 only. So \bar{a}_3 is the possible solution for $\bar{t} < 4(1 + 1/\sqrt{2})$. However, since \bar{a}_3 joins smoothly with the always real root \bar{a}_1 at $\bar{t} < 2(1 + 1/\sqrt{2})$, the Hubble scalar factor takes the expression

$$\begin{aligned} \bar{a}(\bar{t}) &= \bar{a}_3(\bar{t})\Theta \left(2 \left(1 + \frac{1}{\sqrt{2}}\right) - \bar{t} \right) \\ &+ \bar{a}_1(\bar{t})\Theta \left(\bar{t} - 2 \left(1 + \frac{1}{\sqrt{2}}\right) \right). \end{aligned} \tag{13.100}$$

We note several specific points in the cosmic time (in units of t_{EQ}): $\bar{a} = 0$ at $\bar{t} = 0$ is the initial condition; $\bar{a} = 1$ at $\bar{t} = 1$ is the matter-radiation equality; $\bar{a} = 2$ at $\bar{t} = 2 \left(1 + \frac{1}{\sqrt{2}}\right)$, where \bar{a}_1 and \bar{a}_3 join to form the solution, is when the Hubble scale factor is twice of that of the matter-radiation equality; $\bar{a} = 3$ at $\bar{t} = 4 \left(1 + \frac{1}{\sqrt{2}}\right)$, where both \bar{a}_2 and \bar{a}_3 become complex, is when the scale factor is three times as large as that of the matter-radiation equality. All these features can be shown in Fig. 13.3. Note that the complex solutions for \bar{a}_2 and \bar{a}_3 are not shown in the figure. The reduced Hubble scale factor $\bar{a}(\bar{t})$, a monotonic increasing function of the cosmic time, is represented by the pink dotted curve marked as

²⁵⁰⁾The sign of $q^3 + r^2$ constrains on the solutions. For $q^3 + r^2 > 0$ there are one real root and two complex roots which are complex conjugate to each other. For $q^3 + r^2 = 0$ $s_1 = s_2$, all three roots are real with one degenerated pair. For $q^3 + r^2 < 0$ all three roots are real.

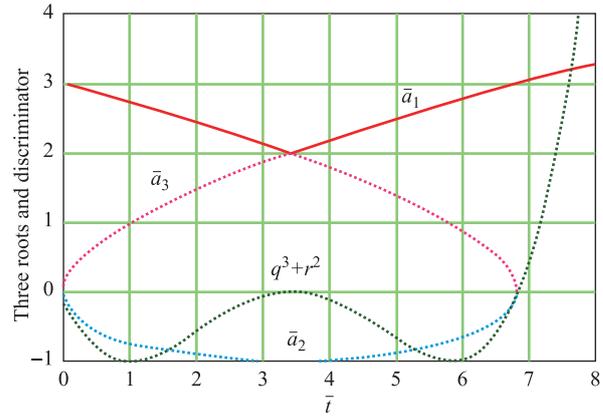


Fig. 13.3 The three roots \bar{a}_1 red solid, \bar{a}_2 blue dotted, and \bar{a}_3 pink dashed as marked by the curves, and the root discriminator $q^3 + r^2$ cyan dash-dotted. The root \bar{a}_1 stays real for all values of \bar{t} , while roots \bar{a}_2 and \bar{a}_3 become complex for $\bar{t} < 0$ and $\bar{t} > 4(1 + 1/\sqrt{2}) \approx 6.828$. The present plot shows only the real part of the roots. The reduced Hubble scale factor $\bar{a}(\bar{t})$ is represented by the pink dotted curve $\bar{a}_3(\bar{t})$ for $\bar{t} \leq (1/2)(1 + 1/\sqrt{2}) \approx 3.414$ and the red solid curve $\bar{a}_1(\bar{t})$ for $\bar{t} \geq (1/2)(1 + 1/\sqrt{2})$.

\bar{a}_3 for $\bar{t} \leq (1/2)(1 + 1/\sqrt{2})$ and the red solid curve marked as \bar{a}_1 for $\bar{t} \geq (1/2)(1 + 1/\sqrt{2})$.

13.8 Summary table

We summarize the present distance-time discussion in Table 13.1.

Appendix A: Phase space, invariants, and the Liouville's theorem

In this Appendix we first demonstrate that the element of the phase space volume

$$d^3x d^3p = dx_1 dx_2 dx_3 dp_1 dp_2 dp_3 \tag{A.1}$$

is invariant under Lorentz transformations. Then we prove one form of the Liouville theorem that says that the number density functions of the phase space is also Lorentz invariant. The thermal distribution function given in (9.99) is an example of this class of Lorentz invariant phase space distribution functions, as has been noted at the end of Section 9.2.1.

The Liouville theorem is an important tool in many branches of physics. It is a fundamental theorem in statistical mechanics and classical mechanics, It has also been said that Liouville theorem underlines all of astronomy.²⁵¹⁾

²⁵¹⁾See, p. 2, [320].

Table 13.1 A summary of the various length and time measures. As noted the quantities listed in the last column are all evaluated by setting $\kappa = 0$.

Distance/Time		Expression	Flat universe $K = 0$
		(Current value) ^①	
Hubble time	$\tau_H = \frac{1}{H_0}$	(13.77 Byr/14.53 Byr) ^②	
Hubble length	$D_H = \frac{c}{H_0} = c\tau_H$	(4.222 Gpc/4.455 Gpc) ^②	
Hubble expansion rate	$H(z) = H_0 \tilde{E}_z$	$H_0 \sqrt{\Omega_\Lambda + \Omega_R z_+^4 + \Omega_M z_+^3 + \Omega_\kappa z_+^2}$ ^③	$H_0 \tilde{E}_z$
Comoving radial coordinate	$\chi_1 = \frac{1}{\sqrt{\kappa}} \arcsin(\sqrt{\kappa} r_1)$ ^④	$\int_0^{r_1} (dr/(1 - \kappa r^2)) = c \int_0^{t_1} (dt/a(t))$ $= \frac{D_H}{a_0} \int_0^{z_1} (dz'/\tilde{E}'_z)$	$\frac{D_H}{a_0} \int_0^{z_1} (dz'/\tilde{E}'_z)$
	$r_1 = \frac{1}{\sqrt{\kappa}} \sin(\sqrt{\kappa} \chi_1)$	$\frac{1}{\sqrt{\kappa}} \sin\left(\sqrt{\kappa} \frac{D_H}{a_0} \int_0^{z_1} (dz'/\tilde{E}'_z)\right)$	$\frac{D_H}{a_0} \int_0^{z_1} (dz'/\tilde{E}'_z)$
Lookback time	$\Delta t_{lb} = \int_{t_1}^{t_0} dt$	$\tau_H \int_0^{z_1} (dz'/((1+z')\tilde{E}'_z))$	$\tau_H \int_0^{z_1} (dz'/((1+z')\tilde{E}'_z))$
Age of universe	$\tau_U = \Delta t_{lb} _{t_1=0}$	$\tau_H \int_0^\infty (dz'/((1+z')\tilde{E}'_z))$	(0.992 τ_H /0.948 τ_H) (13.66 Byr/13.77 Byr)
Proper distance	$d_p(t, r_1) = a(t)\chi_1$	$\frac{D_H}{1+z_1} \int_0^{z_1} (dz'/\tilde{E}'_z)$	$\frac{D_H}{1+z_1} \int_0^{z_1} (dz'/\tilde{E}'_z)$
Comoving distance	$d_c(r_1) = a(t_0)\chi_1 = d_p(t_0, r_1)$	$D_H \int_0^{z_1} (dz'/\tilde{E}'_z)$	$D_H \int_0^{z_1} (dz'/\tilde{E}'_z)$
Particle horizon	$d_{ph}(t_1) = d_c(r_{1max})$ $= a(t_1) \int_0^{t_1} (dt'/a(t'))$	$D_H \int_{z_1=0}^\infty (dz'/\tilde{E}'_z)$	(3.383 D_H /3.171 D_H) (14.28/14.13) Gpc
Angular diameter distance	$d_A(r_1) \equiv \frac{d_{tr}}{\delta\theta} = a(t_1)r_1$	$\frac{a_0}{1+z_1} \frac{1}{\sqrt{\kappa}} \sin\left(\sqrt{\kappa} \frac{D_H}{a_0} \int_0^{z_1} (dz'/\tilde{E}'_z)\right)$	$\frac{D_H}{1+z_1} \int_0^{z_1} (dz'/\tilde{E}'_z)$
Comoving angular diameter	$d_{cA}(r_1) = a_0 r_1$	$(1+z_1)d_A(r_1)$	$D_H \int_0^{z_1} (dz'/\tilde{E}'_z)$
Luminosity distance	$d_L(r_1) = (1+z_1)d_{cA}(r_1)$	$(1+z_1)d_{cA}(r_1) = (1+z_1)^2 d_A(r_1)$	$(1+z_1)D_H \int_0^{z_1} (dz'/\tilde{E}'_z)$
Cosmic event horizon			

^①Two data sets are used. Set 1 is given by WMAP in the 2013 PDG: $\Omega_\Lambda = 0.73$, $\Omega_M = 0.27$ and $h = 0.71$. Set 2 is given by the best fit of Planck 2013 plus WMAP [236]: $\Omega_\Lambda = 0.685$, $\Omega_M = 0.315$ and $h = 0.673$.

^②Conversion of different unites: 1 pc = 3.262. 1y = 1.02927 $\times 10^8$ s = 3.08568 $\times 10^{16}$ m, 1 ly = 0.96405 $\times 10^{16}$ m.

^③We used the notation $z_+ \equiv 1 + z$.

^④ r_1 and χ_1 refer to a fixed comoving radial coordinates corresponding to the redshift z_1 and when the cosmic age is t_1 .

A.1 Lorentz invariant volume element of the momentum space

Let us first consider the Lorentz invariant element of the momentum space. This is well-known and relatively easily to obtain. Here we revert to the usual convention of positive metric used in particle physics, i.e., $p_\mu p^\mu = (p^0)^2 - (\mathbf{p})^2$, instead of the negative metric given in Eq. (9.16), which is more convenient in general relativity. The element of the energy-momentum $d^4p = dp^0 dp^1 dp^2 dp^3$ and the mass-shell condition $\delta_+(p_\mu p^\mu - m^2)$ are Lorentz invariant, where the subscript “+” in the δ -function means taking the positive solution for p_0

in $p_\mu p^\mu - m^2 = 0$. Then their product

$$d^4p \delta_+(p_\mu p^\mu - m^2) = \frac{1}{2E} d^3p \tag{A.2}$$

is clearly Lorentz invariant. This is a typical example of an elegant or slick argument which uses the combination of Lorentz invariant quantities to obtain another Lorentz invariant result.

For readers who are suspicious of any slick arguments and would prefer to prove results using the very fundamental relationship that can be employed, Eq. (A.2) can be proved as follows: Let p^μ and p'^μ be the 4-momentum

of a particle in two parallel coordinate systems²⁵²⁾ related by a Lorentz transformation with the relative velocity $\vec{\beta}$ in an arbitrary direction. We have²⁵³⁾

$$\begin{aligned} p'^0 &= \gamma(p^0 - \boldsymbol{\beta} \cdot \mathbf{p}), \\ \mathbf{p}' &= \mathbf{p} + \frac{\gamma - 1}{\beta^2}(\boldsymbol{\beta} \cdot \mathbf{p})\boldsymbol{\beta} - \gamma p^0 \boldsymbol{\beta}, \end{aligned} \tag{A.3}$$

where $\boldsymbol{\beta}$ is the relative velocity of the two inertial frames, $\gamma = 1/\sqrt{1 - \beta^2}$ is the regular Lorentz boosting factor, $p^0 = \sqrt{m^2 + \mathbf{p}^2}$, and $p'^0 = \sqrt{m^2 + \mathbf{p}'^2}$. Recall that we work in the natural units, setting the speed of the light $c = 1$. We can rewrite the relationship in Eq. (A.3) in the matrix form,

$$p' = \tilde{\Lambda} p, \tag{A.4}$$

where p' and p are respectively the column matrices made of the components of the momenta p'^μ and p^μ , and the elements of the matrix $\tilde{\Lambda}$ are

$$\begin{aligned} \tilde{\Lambda}_0^0 &= \gamma, \\ \tilde{\Lambda}_j^0 &= \tilde{\Lambda}_0^j = -\gamma\beta_j, \\ \tilde{\Lambda}_k^j &= \delta_{jk} + \frac{\gamma - 1}{\beta^2}\beta_j\beta_k. \end{aligned} \tag{A.5}$$

It is straightforward to check that $\tilde{\Lambda}$ has a unity determinant $|\tilde{\Lambda}| = 1$, and the inverse of $\tilde{\Lambda}$ is obtained by changing the sign of the vector $\boldsymbol{\beta}$ in $\tilde{\Lambda}$, i.e.,

$$\tilde{\Lambda}^{-1}(\boldsymbol{\beta}) = \tilde{\Lambda}(-\boldsymbol{\beta}). \tag{A.6}$$

The Lorentz transformation leaves the metric of special relativity invariant $G = (2\delta_{\mu 0}\delta_{\nu 0} - g_{\mu\nu})$, i.e., in the 4×4 form

$$\tilde{\Lambda}^T G \tilde{\Lambda} = G, \quad G \equiv \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} \tag{A.7}$$

Now we can calculate directly the transformation of the volume element of the 3-momentum space from the second expression of Eq. (A.3) by taking into account of the mass-shell condition $p^0 = \sqrt{\mathbf{p}^2 + m^2}$

$$dp^1 dp^2 dp^3 = \left| \frac{\partial p'^j}{\partial p^k} \right| dp^1 dp^2 dp^3, \tag{A.8}$$

²⁵²⁾What we mean by parallel coordinate systems is that their three coordinate-axes are parallel. This is just a technic assumption to make the formulae involved simpler. The case of two coordinate axes arbitrarily oriented with respect to each other can be obtained from the case of parallel coordinate-axes by a rotation. Since rotation is a unitary operation. The argument below still hold.

²⁵³⁾The expressions given in Eq. (A.3) can be found in, e.g., [321], Eq. (11.19), p. 517.

where the determinant is

$$\begin{aligned} \left| \frac{\partial p'^j}{\partial p^k} \right| &= \left| \delta_{jk} + \frac{\gamma - 1}{\beta^2}\beta_j\beta_k - \frac{\gamma}{p^0}\beta_j p_k \right| \\ &= \frac{\gamma(p^0 - \boldsymbol{\beta} \cdot \mathbf{p})}{p^0} = \frac{p'^0}{p^0}, \end{aligned} \tag{A.9}$$

which gives the Lorentz invariance equation given in Eq. (A.2)

$$\frac{1}{p'^0} d^3 p' = \frac{1}{p^0} d^3 p. \tag{A.10}$$

A.2 Lorentz invariant volume element of the configuration space

The Lorentz invariance volume element of the configuration space, which is not commonly discussed explicitly, can be shown similarly, but with a slight twist. The answer can be expected from one's first contact with special relativity, i.e., the topic of Lorentz contraction. The calculation below is just a fancy version of the topic of Lorentz contraction. Let us examine a massive particle in different coordinate frames. First consider the reference frame X in which the particle has the 3-momentum \mathbf{p} and energy p^0 , together with the rest frame of the particle which we call X_R . The two reference frames are taken to be parallel and related by a Lorentz transformation in which the X frame moves relative to the X_R with the velocity $-\mathbf{p}/p^0 = -\boldsymbol{\beta}$. The γ factor is given by $\gamma = p^0/\sqrt{p^{02} - \mathbf{p}^2}$. The coordinates of the particle in the two reference frames are related according to Eq. (A.3)

$$\begin{aligned} t &= \gamma(t_R + \boldsymbol{\beta} \cdot \mathbf{x}_R), \\ \mathbf{x} &= \mathbf{x}_R + \frac{\gamma - 1}{\beta^2}(\vec{\beta} \cdot \mathbf{x}_R)\boldsymbol{\beta} + \gamma t_R \vec{\beta}. \end{aligned} \tag{A.11}$$

We take small variations of the coordinates:

$$\begin{aligned} \Delta t &= \gamma(\Delta t_R + \boldsymbol{\beta} \cdot \Delta \mathbf{x}_R), \\ \Delta \mathbf{x} &= \Delta \mathbf{x}_R + \frac{\gamma - 1}{\beta^2}(\vec{\beta} \cdot \Delta \mathbf{x}_R)\boldsymbol{\beta} + \gamma \Delta t_R \vec{\beta}. \end{aligned} \tag{A.12}$$

The length measurement in the X -frame, expressed in terms of that of the X_R -frame, is an simultaneous event in the former, so we have to take $\Delta t = 0$. Then we have

$$\Delta \mathbf{x} = \Delta \mathbf{x}_R + \frac{\gamma - 1}{\beta^2}(\vec{\beta} \cdot \Delta \mathbf{x}_R)\boldsymbol{\beta} - \gamma(\vec{\beta} \cdot \Delta \mathbf{x}_R)\vec{\beta}. \tag{A.13}$$

In vector components we can write

$$\Delta x_j = \left(\delta_{jk} - \frac{\gamma - 1}{\gamma\beta^2}\beta_j\beta_k \right) \Delta x_{Rk}. \tag{A.14}$$

The infinitesimal volume elements in the two frames of reference are related by

$$\Delta x_1 \Delta x_2 \Delta x_3 = \left| \frac{\partial \Delta x_j}{\partial \Delta x_{Rk}} \right| \Delta x_{R1} \Delta x_{R2} \Delta x_{R3}. \tag{A.15}$$

The determinant is very simple

$$\left| \frac{\partial \Delta x_j}{\partial \Delta x_{Rk}} \right| = \left| \delta_{jk} - \frac{\gamma - 1}{\gamma \beta^2} \beta_j \beta_k \right| = \frac{1}{\gamma}, \quad (\text{A.16})$$

which gives us the expected result

$$\Delta x_{R1} \Delta x_{R2} \Delta x_{R3} = \gamma \Delta x_1 \Delta x_2 \Delta x_3. \quad (\text{A.17})$$

If the particle is considered in another coordinate frame X' of energy p^0 or the γ factor γ' , a relation similar to Eq. (A.17) holds,

$$\Delta x_{R1} \Delta x_{R2} \Delta x_{R3} = \gamma' \Delta x'_1 \Delta x'_2 \Delta x'_3. \quad (\text{A.18})$$

Then we obtain the invariant volume elements in the configuration space

$$\begin{aligned} \gamma' \Delta x'_1 \Delta x'_2 \Delta x'_3 &= \gamma \Delta x_1 \Delta x_2 \Delta x_3, \\ p'^0 d^3 x' &= p^0 d^3 x, \end{aligned} \quad (\text{A.19})$$

where $p^0/p'^0 = \gamma'/\gamma$. This is the relationship we are looking for.

A.3 Lorentz invariant phase space elements

Combining Eqs. (A.10) and (A.19), we have the Lorentz invariant phase space element as the product of the two Lorentz invariant elements:

$$d^3 p' d^3 x' = \left(\frac{p^0}{p'^0} d^3 p \right) \left(\frac{p'^0}{p^0} d^3 x \right) = d^3 p d^3 x. \quad (\text{A.20})$$

An intuitive quick check of this result can be made by considering a massive particle, starting being stationary in the x -frame. It is boosted by the γ factor along, say, the z -direction in the x' -frame. We have, then, $\Delta x' = \Delta x$, $\Delta y' = \Delta y$, $\Delta p'_x = \Delta p_x$, and $\Delta p'_y = \Delta p_y$, while momentum in the z -direction is boosted by a factor γ , $\Delta p'_z = \gamma \Delta p_z$, and the length in the z -direction is suffered a contraction by the same factor, $\Delta z' = \gamma^{-1} \Delta z$. We have

$$d^3 p' d^3 x' = (\gamma d^3 p) \left(\frac{1}{\gamma} d^3 x \right) = d^3 p d^3 x. \quad (\text{A.21})$$

A.4 Liouville's theorem

In a simple form, the often stated Liouville theorem is the assertion that the phase-space distribution function is constant along trajectories of a system, under quite broad constraining conditions. So the theorem deals with the conservation of the phase-space distribution function. One can find many discussions of the theorem in the literature and on the web. Let us first clarify what we mean by the phase-space distribution function. We use a mechanical system as an illustration. A state of

a mechanical system is determined by its coordinates: x_j and momentum p_j , $j = 1, 2, 3$. Denote the density of the number of states in the phase space volume $dV_{ps} = d^3 x d^3 p$ by $f(t, x_j, p_j)$, where t is the time. The total number of states in this volume element is

$$\Delta N = f(t, x_j, p_j) dV_{ps}. \quad (\text{A.22})$$

Under the assumption that the system is under the influence of a smooth, conserved force, which means no collision or frictional forces, the phase space elements dV_{ps} and the total number of states ΔN are constant in time as the system moves along its trajectory. Hence the state distribution function is also a constant. This means

$$\begin{aligned} \frac{d}{dt} f(t, x_j, p_j) &= \frac{\partial}{\partial t} f(t, x_j, p_j) + \left(\frac{\partial}{\partial x_j} f(t, x_k, p_k) \right) \dot{x}_j \\ &\quad + \left(\frac{\partial}{\partial p_j} f(t, x_k, p_k) \right) \dot{p}_j \\ &= 0. \end{aligned} \quad (\text{A.23})$$

A proof of the above expression can be found in [322]. A detailed exposition of the Liouville's theorem, including some useful elementary material can be found in [320]. A discussion of the more complicated cases of the presence of abrupt forces, such as collisions, can be found in [237]²⁵⁴⁾

To conclude: we have shown that the phase-space volume element dV_{ps} is invariant in the preceding subsection. The total number of states in an infinitesimal phase-space volume element, obviously, cannot be changed by a Lorentz transformation, so the phase-space distribution function has to be Lorentz invariant. This says, from Eq. (9.99), that the energy of the photon concerned and its temperature track each other under the Lorentz transformation.

Appendix B: Useful cross section formulae

In this Appendix we list some of the relevant low energy cross sections.

B.1 Electromagnetic interaction cross sections

B.1.1 Klein–Nishina formula

The photon electron/positron scattering, $\gamma + e^\pm \rightarrow \gamma + e^\pm$ in the laboratory frame with an incident photon of three-momentum \mathbf{k} and out going photon of momentum \mathbf{k}'

$$|\mathbf{k}'| = \frac{m_e |\mathbf{k}|}{m_e + |\mathbf{k}|(1 - \cos \theta)}, \quad (\text{B.1})$$

²⁵⁴⁾See, pp 408–413, [237].

where θ is the angle of the final photon momentum with respect to the direction of the momentum of the incident photon. Then the differential cross section is given by

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{2m_e^2} \frac{|\mathbf{k}|^2}{|\mathbf{k}'|^2} \left(\frac{|\mathbf{k}'|}{|\mathbf{k}|} + \frac{|\mathbf{k}|}{|\mathbf{k}'|} - \sin^2 \theta \right), \quad (\text{B.2})$$

where $\alpha = 1/137$ is the fine structure constant at vanishing momentum.

B.1.2 Thomson scattering cross section

The Thomson cross section is obtained by taking the limit of zero energy for the incident photon and integrating over the directions of the outgoing photons. In the limit of vanishing energy of the photon, $|\mathbf{k}'|/|\mathbf{k}| \rightarrow 1$, we have

$$\left. \frac{d\sigma}{d\Omega} \right|_{|\mathbf{k}'| \rightarrow |\mathbf{k}|} = \frac{\alpha^2}{2m_e^2} (1 + \cos^2 \theta). \quad (\text{B.3})$$

The integration over the solid angle is straightforward, we obtain

$$\begin{aligned} \sigma_{\text{Th}} &= \frac{8\pi}{3} r_e^2, \\ r_e &= \frac{\alpha}{m_e}. \end{aligned} \quad (\text{B.4})$$

r_e is the classical radius of the electron. Numerically we have

$$\begin{aligned} r_e &= 2.81794 \times 10^{-13} \text{ cm}, \\ \sigma_{\text{Th}} &= 6.65256 \times 10^{-25} \text{ cm}^2. \end{aligned} \quad (\text{B.5})$$

Because of the inverse mass-square dependence of the Thomson cross section, the scattering cross section of massive fermion will be suppressed by about a factor $(m_e/M)^{-2}$, where M is the mass of the heavier fermion in question. In particular, the low energy photon-proton scattering cross section, much below that of the one pion production process which has the photon threshold energy $m_\pi[1 + m_\pi/(2m_p)] \approx m_\pi$ in the laboratory frame (i.e., the initial proton at rest), the cross section involves a Thomson-cross-section term plus the contribution from the proton anomalous magnetic moment. So the cross section is smaller roughly by a factor $(m_e/m_p)^2 \approx 3 \times 10^{-7}$. Hence in the early universe the equilibrium of photon with the charged plasma of electrons and protons is mainly due to the photon-electron scattering and then the electron-proton scattering.

B.2 Weak interactions cross sections

B.2.1 Muon lifetime

The lowest order calculation of the decay width of the muon (and antimuon), $\mu^- \rightarrow \nu_\mu + e^- + \bar{\nu}_2$, is

$$\Gamma_\mu = \frac{G_F^2 m_\mu^5}{192\pi} + \mathcal{O}\left(\frac{m_e}{m_\mu}\right), \quad (\text{B.6})$$

which gives the mean lifetime of the muon, with the Fermi constant $G_F = 1.1664 \times 10^{-5} \text{ GeV}^{-2}$,

$$\tau_\mu = \frac{1}{\Gamma_\mu} \approx 2.187 \times 10^{-6} \text{ s}. \quad (\text{B.7})$$

The experimental value is $\tau_\mu^{\text{exp}} = (2.197034 \pm 0.000021) \times 10^{-6} \text{ s}$ [11].

B.2.2 Leptonic elastic cross sections

These reactions involve the elastic scattering of the three flavors of neutrinos against the electron:

$$\begin{aligned} \sigma(\nu_e e^- \rightarrow \nu_e e^-) &= \frac{G_F^2 S}{\pi} \left(\left(\frac{1}{2} + X_W \right)^2 + \frac{1}{3} X_W^2 \right) \\ &\xrightarrow{\text{Lab frame}} 9.5 \times 10^{-45} \left(\frac{E_\nu + m_e/2}{1 \text{ MeV}} \right)^2 \text{ cm}^2, \\ \sigma(\bar{\nu}_e e^- \rightarrow \bar{\nu}_e e^-) &= \frac{G_F^2 S}{\pi} \left(\frac{1}{3} \left(\frac{1}{2} + X_W \right)^2 + X_W^2 \right) \\ &\xrightarrow{\text{Lab frame}} 4.0 \times 10^{-45} \left(\frac{E_\nu + m_e/2}{1 \text{ MeV}} \right)^2 \text{ cm}^2, \\ \sigma(\nu_\mu e^- \rightarrow \nu_\mu e^-) &= \frac{G_F^2 S}{\pi} \left(\left(\frac{1}{2} - X_W \right)^2 + \frac{1}{3} X_W^2 \right) \\ &\xrightarrow{\text{Lab frame}} 1.6 \times 10^{-45} \left(\frac{E_\nu + m_e/2}{1 \text{ MeV}} \right)^2 \text{ cm}^2, \\ \sigma(\bar{\nu}_\mu e^- \rightarrow \bar{\nu}_\mu e^-) &= \frac{G_F^2 S}{\pi} \left(\frac{1}{3} \left(\frac{1}{2} - X_W \right)^2 + X_W^2 \right) \\ &\xrightarrow{\text{Lab frame}} 1.3 \times 10^{-45} \left(\frac{E_\nu + m_e/2}{1 \text{ MeV}} \right)^2 \text{ cm}^2, \end{aligned} \quad (\text{B.8})$$

where $X_W \equiv \sin^2 \theta_W \approx 0.23$ [11], θ_W is the weak mixing or Weinberg angle, and $S = (p_e + p_\nu)^2$, i.e., the square of the sum of the 4-momenta of the incoming electron and neutrino. The second expression of each of the above cross section formulae is given in the laboratory frame.

B.2.3 Leptonic inelastic cross sections

The followings are muon production cross section:

$$\begin{aligned} \sigma(\nu_\mu e^- \rightarrow \mu^- \nu_e) &= \frac{G_F^2}{\pi} \frac{(S - m_\mu^2)^2}{S} \Theta(S - m_\mu^2), \\ \sigma(\bar{\nu}_\mu e^- \rightarrow \mu^- \bar{\nu}_e) &= \frac{G_F^2}{3\pi} \frac{(S - m_\mu^2)^2}{S} \Theta(S - m_\mu^2), \end{aligned} \quad (\text{B.9})$$

where the Θ -function denotes the threshold requirement. We do not give the results in the laboratory frame because it requires rather high energies to produce a muon from the electron of the order of 20 GeV.

B.2.4 More neutrino leptonic and annihilation cross sections

Formulae of neutrino-neutrino elastic scatter and annihilation, and neutrino charged lepton two-body reaction for the spin sum of squared matrix elements can be found in [324].

B.2.5 Neutrino nucleon elastic cross sections

Below are neutrino-nucleon elastic scattering cross section at low energies in the laboratory frame. The formula applies to all three flavors of neutrinos and antineutrinos:

$$\begin{aligned} \sigma(\nu n \rightarrow \nu n) &= \frac{G_F^2 E_\nu^2}{\pi} (g_V^2 + 3g_A^2) \\ &= 9.3 \times 10^{-44} \left(\frac{E_\nu}{1 \text{ MeV}} \right)^2 \text{ cm}^2, \\ \sigma(\nu p \rightarrow \nu p) &= \frac{G_F^2 E_\nu^2}{4\pi} (g_V^2 + 3g_A^2) (1 - 4X_W)^2 \\ &= 6.0 \times 10^{-46} \left(\frac{E_\nu}{1 \text{ MeV}} \right)^2, \end{aligned} \tag{B.10}$$

where E_ν is the incident neutrino energy in the laboratory frame and

$$g_V = 1, \quad g_A = 1.257. \tag{B.11}$$

B.2.6 Two-body neutrino nucleon inelastic cross sections

The low energy neutrino nucleon scattering cross sections are directly relevant to the present discussion.²⁵⁵⁾ We consider the case that the nucleons involved in the reaction are practically at rest.

$$\begin{aligned} \sigma(\nu_e n \rightarrow e^- p) &= \frac{G_F^2 E_\nu^2}{\pi} (g_V^2 + 3g_A^2) \left(1 + \frac{Q}{E_\nu} \right) \sqrt{1 + 2\frac{Q}{E_\nu} + \frac{Q^2 - m_e^2}{E_\nu^2}} \\ &= \frac{G_F^2}{\pi} (g_V^2 + 3g_A^2) E_{e^-} p_{e^-} \\ &= \frac{G_F^2}{\pi} (g_V^2 + 3g_A^2) \beta_{e^-} E_{e^-}^2, \\ \sigma(\bar{\nu}_e p \rightarrow e^+ n) &= \frac{G_F^2 E_\nu^2}{\pi} (g_V^2 + 3g_A^2) \left(1 - \frac{Q}{E_\nu} \right) \sqrt{1 - 2\frac{Q}{E_\nu} + \frac{Q^2 - m_e^2}{E_\nu^2}} \\ &= \frac{G_F^2}{\pi} (g_V^2 + 3g_A^2) E_{e^+} p_{e^+} \\ &= \frac{G_F^2}{\pi} (g_V^2 + 3g_A^2) \beta_{e^+} E_{e^+}^2, \end{aligned} \tag{B.12}$$

where E_ν is the incident neutrino energy, $Q = m_n - m_p = 1.2933 \text{ MeV}$ is the neutron and proton mass difference,

²⁵⁵⁾ A derivation of the cross section formula below can be found in [253].

and β_{e^\mp} are the relativistic beta factor for the electron and positron, i.e., their velocity in the natural units. In the approximation that the nucleons are at rest, we have $E_{e^-} = E_\nu + Q$ and $E_{e^+} = E_\nu - Q$, and $p_{e^\mp} = \sqrt{E_{e^\mp}^2 - m_e^2}$, which are the relationships that allow us to write the second expressions in each of the two cross section formulae. Furthermore we note that the above two expressions are related by crossing the two nucleons, which changes the sign of Q , and then applies the charge conjugation and parity (CP) operation.

Using crossing by interchanging the initial and final leptons in the two reactions of Eq. (B.12), we obtain the cross sections of two more lepton nucleon reactions:

$$\begin{aligned} \sigma(e^+ n \rightarrow \bar{\nu}_e p) &= \frac{G_F^2}{\pi} (g_V^2 + 3g_A^2) E_{\bar{\nu}}^2, \\ \sigma(e^- p \rightarrow \nu_e n) &= \frac{G_F^2}{\pi} (g_V^2 + 3g_A^2) E_\nu^2, \end{aligned} \tag{B.13}$$

where the mass of the neutrino is neglected. Let us note that the above two sets of reactions Eqs. (B.12) and (B.13) enter the discussion of cosmological nucleosynthesis.²⁵⁶⁾

B.2.7 Orders of magnitude of typical cross sections

Interaction	Typical cross section	Decay, Life-time
Strong	$10^{-26} \text{ cm}^{-2} = 10^2 \text{ barn} = 10 \text{ mb}$	10^{-24} s
Electromagnetic	$10^{-32} \text{ cm}^{-2} = 10^{-8} \text{ barn} = 10 \text{ nb}$	10^{-16} s
Weak	$10^{-38} \text{ cm}^{-2} = 10^{-14} \text{ barn} = 10 \text{ fb}$	10^{-8} s

Appendix C: Useful constants and units conversion

In this Appendix we list a few important physical constants which are meant for quick references. For details and more complete listings, we refer to the appropriate tables in [13].

C.1 Natural units and units conversion

In the extended natural units there is only one unit, usually taken as the energy. Proper units of physical quantities can be recovered by insert enough powers of \hbar , c , and k_B , or converting the energy units to another units with fix numerical values: $\hbar = 1$ relates energy to time, $\hbar c = 1$ energy to length, and $k_B = 1$ energy to temperature. First we list the three well-known fundamental

²⁵⁶⁾ See [218], pp 547–548.

Table C.1 The three fundamental constants.

Fundamental constant	value	Proper units	Natural units
\hbar	6.582118×10^{-22}	MeV·s	1
c	2.997825×10^{10}	cm·s ⁻¹	1
k_B	8.617343×10^{-11}	MeV·K ⁻¹	1
G_N	6.67428×10^{-14}	m ³ ·g ⁻¹ ·s ⁻²	
	6.70881×10^{-39}	$\hbar c(\text{GeV}/c^2)^{-2}$	GeV ⁻²
G_F	1.16637×10^{-5}	$(\hbar c)^3 \text{GeV}^{-2}$	GeV ⁻²

constants \hbar , c , k_B , and G_N and G_F . To our adopted accuracy the errors can be omitted.

C.1.1 Units conversion

In the natural units the particle number density has the dimension of $(\text{energy})^3$, the energy density $(\text{energy})^4$, etc. To convert the energy units into a length units, one needs to multiply it by the factor

$$\begin{aligned}
 (\hbar c)^{-1} &= (1.97327 \times 10^{-11} \text{ MeV} \cdot \text{cm})^{-1} \\
 &= 5.06773 \times 10^{10} \text{ MeV}^{-1} \text{ cm}^{-1}.
 \end{aligned}
 \tag{C.1}$$

The following table provides this kind of conversion factors.

As an example, to write the particle number density, in the units of MeV³, to the proper units cm⁻³, the factor is multiplied:

$$\begin{aligned}
 1 \text{ MeV}^3 &= (5.06773 \times 10^{10} \text{ cm}^{-1})^3 \\
 &= 1.30149 \times 10^{32} \text{ cm}^{-3}.
 \end{aligned}
 \tag{C.2}$$

C.2 Masses and binding energies

We list a few useful mass values and nuclear binding energies in Table C.3.

Table C.2 Units conversion.

Energy	Temperature	Length	Time
1 MeV	1.16045×10^{10} K	5.06773×10^{10} cm ⁻¹	1.51927×10^{21} s ⁻¹
1.60218×10^{-6} erg			
1.78266×10^{-27} g			
8.61734×10^{-11} MeV	1 K	4.36704 cm ⁻¹	1.30920×10^{11} s ⁻¹
5.06773×10^{10} MeV ⁻¹	4.36704 K ⁻¹	1 cm	3.33564×10^{-11} s
1.51927×10^{21} MeV ⁻¹	1.30920×10^{11} K ⁻¹	2.99783×10^{10} cm	1 s
1.97327×10^{-11} MeV	2.28988×10^{-1} K	1 cm ⁻¹	2.99792×10^{10} s ⁻¹
6.58212×10^{-22} MeV	7.63822×10^{-12} K	3.33564×10^{-11} cm ⁻¹	1 s ⁻¹

Binding energy, half-life, and spin of the light nuclei in Table C.4.

Table C.3 The nucleon mass values and difference.

Nucleon	Mass value (MeV)
Proton m_p	938.2720
Neutron m_n	939.5653
Electron m_e	0.510999
$Q = m_n - m_p$	1.2933

Table C.4 Binding energies of selected light nuclei. Most of the values are obtained from [325]. Some of the values may not be exact. Depending on the use of the information, it is advisable that independent sources should be checked for accuracy.

Nucleus	Binding energy (MeV)	spin	half-life
Deuteron D	2.23452	1	-
Tritium T	8.4818	1/2	12.32 yrs
Helium-3 ${}^3\text{He}$	7.7180	1/2	-
Helium-4 ${}^4\text{He}$	28.302	0	-
Lithium-4 ${}^4\text{Li}$	1.15/nucleon	2	7.58×10^{-23} s
Lithium-6 ${}^6\text{Li}$	5.33/nucleon	1	-
Lithium-7 ${}^7\text{Li}$	5.606/nucleon	3/2	-
Lithium-8 ${}^8\text{Li}$	5.160	2	839 ms
Lithium-9 ${}^9\text{Li}$			178.3 ms
Beryllium-7 ${}^7\text{Be}$	5.371/nucleon	3/2	53 days
Beryllium-8 ${}^8\text{Be}$	7.062	0	7×10^{-17} s
Beryllium-9 ${}^9\text{Be}$	6.463/nucleon	3/2	-
Beryllium-10 ${}^{10}\text{Be}$	6.498/nucleon	0	2.18×10^6 yrs

C.3 Useful constants

Astrophysical constants & parameters in Table C.5.

C.4 Planck quantities

Planck quantities involve the gravitational constant G_N are given in Table C.6.

Note that all Planck quantities in the table can be obtained from the Planck mass M_P by applying units conversion factors given in Table C.2 which come from the definition of natural units. The numerical value of $M_P = 1.22 \times 10^{19}$ GeV in nature units can be directly obtained from Table C.1. The Planck mass and energy densities are a matter of definition as defined in Table C.2. But their form can also be argued as follows, using the Planck mass density as an illustration. First, we notice that the expression $M_P = \sqrt{\hbar c/G_N}$ has indeed the proper mass units, although the Planck mass is usually expressed in the energy units, i.e., GeV. To obtain the Planck mass density, we observe that in natural units the mass density has the dimension of energy to the 4th power. Since the only relevant mass in the Planck scale is the Planck mass, we start with M_P^4 and convert three powers of M_P each into units of the length inverse, i.e., $M_P \rightarrow M_P(c/\hbar)$. Hence we have

$$\begin{aligned} \rho_P &\equiv M_P^4 \left(\frac{c}{\hbar}\right)^3 = \left(\frac{\hbar c}{G_N}\right)^2 \left(\frac{c}{\hbar}\right)^3 \\ &= \frac{c^5}{\hbar G_N^2}, \end{aligned} \tag{C.3}$$

which is just M_P/ℓ_P^3 . To obtain its numerical value in the proper units, using Table C.2 we convert one power of M_P into units of mass in grams, and three powers of M_P each into the inverse of the length in cm^{-1} ; or, we simply use the numerical values of the quantities involved as given in Table C.6.

Table C.5 Some useful astrophysical constants and parameters. All relevant numbers are referred to the present epoch.

Constant/Parameter	Value	Units
c (speed of light)	2.99792×10^{10}	cm
Yr	3.15569×10^7	sec
ly (light year)	9.46053×10^{15}	m
pc (parsec)	3.08568×10^{16}	m
	3.262	ly
n_γ	$410.5(T/2.725)^3$	cm^{-3}
n_b	$(2.482 \pm 0.032) \times 10^{-7}$	cm^{-3}
	0.256 protons	m^{-3}
$\eta = n_b/n_\gamma$	6.05×10^{-10}	
h	0.673 ± 0.012	
H_0	$(9.777752)^{-1}h$	Gyr^{-1}
	$3.24091 \times 10^{-18}h$	sec^{-1}
$\tau_H^{(0)} = 1/H_0$ (Hubble time)	$3.08556 \times 10^{17}h^{-1}$	sec
	$9.77775h^{-1}$	Gyr
$D_H^{(0)} = c/H_0$ (Hubble length)	$9.25025 \times 10^{25}h^{-1}$	m
τ_U (age of universe)	13.6	Gyr
$\rho_c = 3H_0^2/8\pi G_N$	$1.05368 \times 10^{-2}h^2$	$(\text{MeV}/c^2)/\text{cm}^3$
	$1.87835 \times 10^{-29}h^2$	g/cm^3
ρ_γ	0.2604	eV/cm^3
ρ_b	240	eV/cm^3
ε_γ (average γ energy)	6.34×10^{-4}	eV
$\Omega_\Lambda = \rho_\Lambda/\rho_c$	0.685	
$\Omega_m = \rho_m/\rho_c$	0.315	
$\Omega_b = \rho_b/\rho_c$	$0.02207h^{-2}$	
$\Omega_\gamma = \rho_\gamma/\rho_c$	2.471×10^{-5} $(T/2.725)^4h^{-2}$	

Table C.6 The Planck quantities.

Quantity	Defined	value	units
Planck mass	$M_P = \sqrt{\frac{\hbar c}{G_N}}$	1.22089×10^{19}	GeV
		2.17651×10^{-5}	g
Reduced Planck mass	$\mu_P = \sqrt{\frac{\hbar c}{8\pi G_N}}$	2.43532×10^{18}	GeV
Planck length	$\ell_P = \sqrt{\frac{\hbar G_N}{c^3}}$	1.61620×10^{-33}	cm
Planck time	$t_P = \sqrt{\frac{\hbar G_N}{c^5}} = \frac{\ell_P}{c}$	5.39123×10^{-44}	sec
Planck temperature	$T_P = \sqrt{\frac{\hbar c^5}{G_N k_B^2}} = \frac{1}{k_B} M_P c^2$	1.41585×10^{32}	K
Planck mass density	$\rho_P = \frac{M_P}{\ell_P^3} = \frac{c^5}{\hbar G_N^2}$	5.15555×10^{93}	g/cm^3
Planck energy density	$\varepsilon_P = \rho_P c^2 = \frac{c^7}{\hbar G_N^2}$	4.63298×10^{114}	erg/cm^3

Acknowledgements Most of the materials contained in the first eight chapters of Part I have been presented at summer schools and special graduate lecture series organized in several institutions in the past several years: Institute of Theoretical Physics, Chinese Academy of Sciences and Kavli Institute of Theoretical Physics China, Beijing; National Tsinghua University, Hsinchu; Shanghai Jiaotong University, Shanghai; Weihai High Energy Physics School (WHEPS)-2015; and Shandong University, Jinan. I would like to thank the following colleagues at these institutions for providing me with the opportunity to lecture on dark matter and for their kind hospitality and support: Prof. Yue-Liang Wu, Prof. Chao-Qiang Geng, Prof. Xiang-Dong Ji, Prof. Meng Wang, Prof. Rong-Gen Cai, and Prof. Zuo-Tang Liang. With them and many other colleagues in China I have had the privilege of years of scientific interactions.

Remarks Some of the figures and cartoon images were downloaded from websites over the past several years. Their original sources of some of the figures and images are lost to me and therefore no suitable credits are attributed. Some are credited in footnotes or figure captions. Whether attributed or not, I want to express my appreciation to the magnanimous authors of these figures and images. However, figures adopted from published articles have their sources cited in the bibliography.

Being a review in nature, this report have borrowed many arguments from the literature and the sources are cited at the appropriate point. However, I would like to mention the sources of the approaches which I follow very closely in two chapters: Section 10 follows the treatment of [251], and Section 12 follows that of [221]. In particular for [221], I have tried to cited in detail down to the level of specific equations. These chapters present analytic derivations, under appropriate approximation schemes, of important topics in astrophysics and cosmology.

The lengthiness of this report is partly due to the detailed treatment of most of the topics, in particular, the derivations of various expressions. In Part II, each section is intended to be independent. The lengthy Table of Contents is trying to list in detail the topics covered.

The writing of the report has been extended for a few years and a significant part written early was based on data of the 2012 edition of the Particle Data Group's (PDG's) *Review of Particle Physics* (RPP). Some portions written more recently were based on the 2014 edition of RPP. The differences of most of the needed data, such as the matter-energy densities, etc. in the two editions are small, mostly agreeing within $1 - \sigma$. When actual data are involved, the version of RPP used are generally noted. Hence, even though updating of older data to the newest available is desirable in principle, I felt that it is not necessary in the present case. So updating the data used for the whole report has not been made.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References and notes

1. S. Weinberg, The cosmological constant problem, *Rev. Mod. Phys.* 61, 1 (1989)

2. D. J. Gross, The frontier physicist, *Nature* 467(7317), S8 (2010)
3. E. W. Kolb, Particle Physics and Cosmology, in: K. L. Peach and L. L. J. Vick (Eds.), St. Andrews, 1993, Proceedings, High Energy Phenomenology, arXiv: astro-ph/9403007
4. G. Aad, et al. (ATLAS Collaboration), Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, *Phys. Lett. B* 716, 1 (2012), arXiv: 1207.7214 [hep-ex]
5. S. Chatrchyan, et al. (CMS Collaboration), Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, *Phys. Lett. B* 716, 30 (2012), arXiv: 1207.7235 [hep-ex]
6. S. Chatrchyan, et al. (CMS Collaboration), Study of the mass and spin-parity of the Higgs boson candidate via its decays to Z boson pairs, *Phys. Rev. Lett.* 110(8), 081803 (2013)
7. G. Aad, et al. (Atlas Collaboration), Evidence for the spin-0 nature of the Higgs boson using ATLAS data, *Phys. Lett. B* 726, 120 (3013), arXiv: 1307.1432 [hep-ex]
8. Y. Baryshev, Paradoxes of cosmological physics in the beginning of the 21st century, in: Proceedings of the XXX-th International Workshop on High Energy Physics - Particle and Astroparticle Physics, Gravitation and Cosmology - Predictions, Observations and New Projects, June 23-27, 2014, in Protvino, Moscow region, Russia; arXiv: 1501.01919 [physics.gen-ph]
9. D. N. Spergel, et al., Wilkinson Microwave Anisotropy Probe (WMAP) Three Year Results: Implications for Cosmology, *Astrophys. J. Suppl.* 170, 377 (2007), arXiv: astro-ph/0603449
10. This is the WMAP third data release made in March 2006. See The WMAP homepage including the WMAP 9-year results can be found at.
11. J. Beringer, et al. (Particle Data Group), The review of particle physics, *Phys. Rev. D* 86, 010001 (2012) (<http://pdg.lbl.gov>)
12. A list of publications on various aspects of the Planck data can be found in
13. K. A. Olive, et al. (Particle Data Group), The review of particle physics, *Chin. Phys. C* 38, 090001 (2014) (Updated version can be found at the PDG website: <http://pdg.lbl.gov/>)
14. http://chandra.harvard.edu/press/10_releases/press_051110.html
15. L. Zappacosta, et al., Studying the WHIM content of the galaxy large-scale structures along the line of sight to H 2356-309, arXiv: 1004.5359 [astro-ph.CO]
16. <http://map.gsfc.nasa.gov/news/index.html>
17. E. Komatsu, et al., Five-year *Wilkinson Microwave Anisotropy Probe* Observations: Cosmological interpretation, *Astrophys. J. Suppl.* 180(2), 330 (2009), arXiv: 0803.0547

18. K. A. Olive, TASI lecture on dark matter, arXiv: astro-ph/0301505
19. H. Murayama, Physics beyond the standard model and dark matter, Lecture given in Les Houches 2006, arXiv: 0704.2276 [hep-ph]
20. F. Zwicky, Die Rotverschiebung von extragalaktischen Nebeln, *Helvetica Physica Acta* 6, 110U127 (1933); See also: F. Zwicky, On the masses of nebulae and of clusters of nebulae, *Astrophys. J.* 86, 217 (1937)
21. J. Einasto, Dark Matter, Astronomy and Astrophysics 2010, Eds. O. Engvold, R. Stabell, B. Czerny, and J. Lattanzio, in: Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford, UK; arXiv: 0901.0632 [astro-ph.CO]
22. K. Freeman and G. McNamara, In Search of Dark Matter, Springer, 2006
23. H. Zinkernagel, High-energy physics and reality-some philosophical aspects of a science, Ph.D. thesis, 1998, Niels Boho Institute, pp 4–5, <http://www.nbi.dk/zink/HEPthesis.pdf>
24. G. Bertone, D. Hooper, and J. Silk, Particle dark matter: Evidence, candidates and constraints, *Phys. Rep.* 405, 279 (2005), arXiv: hep-ph/0404175
25. G. B. Gelmini, TASI 2014 Lectures: The hunt for dark matter, arXiv: 1502.01320 [hep-ph]
26. Powerpoint presentations of talks given at the 42nd SLAC Summer Institute (2014) can be found at <https://indico.cern.ch/event/297618/other-view?view=standard>
27. J. Primack, A brief history of dark matter, <http://physics.ucsc.edu/joel/Ay/214/Jan12-Primack-DM-History.pdf>
28. V. Springel, S. D. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce, Simulations of the formation, evolution and clustering of galaxies and quasars, *Nature* 435(7042), 629 (2005), arXiv: astro-ph/0504097
29. NASA images from Large Synoptic Survey Telescope (LSST): <http://www.lsst.org/lsst/public>
30. M. Bartelmann and P. Schneider, Weak gravitational lensing, *Phys. Rep.* 340(4-5), 291 (2001)
31. R. Massey, J. Rhodes, R. Ellis, N. Scoville, A. Leauthaud, et al., Dark matter maps reveal cosmic scaffolding, *Nature* 445 (7125), 286 (2007)
32. O. Goske, B. Moore, J. Kneib, and G. Soucail, A wide-field spectroscopic survey of the cluster of galaxies Cl0024+ 1654-II. A high-speed collision? *Astron. Astrophys.* 386, 31 (2002)
33. More can be found in <http://www.spacetelescope.org/news/html/heic0709.html>.
34. F. Kahlhoefer, K. Schmidt-Hoberg, M.T. Frandsen, and S. Sarkar, Colliding clusters and dark matter self-interactions, *Mon. Not. R. Astron. Soc.* 437, 2865 (2014), arXiv: 1308.3419 [astro-ph.CO]
35. <http://chandra.harvard.edu/photo/2006/1e0657/http://home.slac.stanford.edu/pressreleases/2006/20060821.htm>
36. D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, A direct empirical proof of the existence of dark matter, *Astrophys. J.* 648, L109 (2006), arXiv: astro-ph/0608407
37. http://www.astro.umd.edu/ssm/mond/moti_bullet.html
38. G. W. Angus, B. Famaey, and H. S. Zhao, Can MOND take a bullet? Analytical comparisons of three versions of MOND beyond spherical symmetry, *Mon. Not. R. Astron. Soc.* 371, 138 (2006), arXiv: astro-ph/0606216
39. M. Bradač, S. W. Allen, T. Treu, H. Ebeling, R. Massey, R. G. Morris, A. von der Linden, and D. Applegate, Revealing the properties of dark matter in the merging cluster MACSJ0025.4-1222, arXiv: 0806.2320 [astro-ph]; a NASA news release at: http://www.nasa.gov/mission_pages/chandra/news/08-111.html. A short video can be found at
40. D. Harvey, R. Massey, T. Kitching, A. Taylor, and E. Tottley, The non-gravitational interactions of dark matter in colliding galaxy clusters, *Science* 347, 1462 (2015), arXiv: 1503.07675 [astro-ph.CO]
41. R. Massey, et al., The behaviour of dark matter associated with 4 bright cluster galaxies in the 10kpc core of Abell 3827, *Mon. Not. R. Astron. Soc.* 449, 3393 (2015), arXiv: 1504.03388 [astro-ph.CO]
42. J. Navarro, C. S. Frenk, and S. D. White, The structure of cold Dark Matter Halos, *Astrophys. J.* 462, 563 (1986), arXiv: astro-ph/9508025
43. J. Navarro, C. S. Frenk, and S. D. White, Universal density profile from hierarchical clustering, *Astrophys. J.* 490, 493 (1997), arXiv: astro-ph/9611107
44. J. Einasto, On the construction of a composite model for the galaxy and on the determination of the system of galactic parameters, *Trudy Astrofizicheskogo Instituta Alma-Ata* 5, 87 (1965)
45. D. Merritt, A. W. Graham, B. Moore, J. Diemand, and B. Terzic, Empirical models for Dark Matter Halos. I. Nonparametric construction of density profiles and comparison with parametric models, *Astrophys. J.* 132, 2685 (2006), arXiv: astro-ph/0509417
46. A. A. Dutton and A. V. Macciò, Cold dark matter haloes in the Planck era: Evolution of structural parameters for Einasto and NFW profiles, *Mon. Not. R. Astron. Soc.* 441, 3359 (2014), arXiv: 1402.7073 [Astroph.CO]
47. A. Burkert, The structure of dark matter haloes in dwarf galaxies, *Astrophys. J.* 447, L25 (1995), arXiv: astro-ph/9504041

48. M. Pierre, J. M. Siegal-Gaskins, and P. Scott, Sensitivity of CTA to dark matter signals from the galactic center, *JCAP* 1406, 024 (2014), arXiv: 1401.7330 [astro-ph.HE]; *Erratum: JCAP* 1410, E01 (2014)
49. V. Vikram, et al. (DES Collaboration), Wide-Field lensing mass Maps from DES science verification data, arXiv: 1504.03002 [astro-ph.CO]
50. D. Scott and G. F. Smoot, Cosmic Microwave Background, given in Ref. [13]
51. D. H. Weinberg, J. S. Bullock, F. Governato, R. K. de Naray, and A. H. G. Peter, Cold dark matter: Controversies on small scales, Proceedings of the National Academy of Sciences of the USA (PNAS), approved Dec. 2, 2014; arXiv: 1306.0913 [astro-ph.CO]
52. M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, Too big to fail? The puzzling darkness of massive Milky Way subhaloes, *Mon. Not. R. Astron. Soc.* 415, L40 (2011), arXiv: 1103.0007 [astro-ph.CO]
53. E. Papastergis, R. Giovanelli, M. P. Haynes, and F. Shankar, Is there a “too big to fail” problem in the field? *Astron. Astrophys.* 574, A113 (2015), arXiv: 1407.4665 [astro-ph/GA]
54. J. R. Primack, Cosmological structure formation, arXiv: 1505.02821 [astro-ph.GA]
55. A. Schneider, D. Amderjadem, A. V. Maccio, and J. Diemand, Warm dark matter does not do better than cold dark matter in solving small-scale inconsistencies, *Mon. Not. R. Astron. Soc.* 441, 6 (2014), arXiv: 1309.5960 [astro-ph.CO]
56. D. N. Spergel and P. J. Steinhardt, Observational evidence for self-interacting cold dark matter, *Phys. Rev. Lett.* 84(17), 3760 (2000), arXiv: astro-ph/9909386
57. M. Milgrom, A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis, *Astrophys. J.* 270, 365 (1983) (listed in HEP-INSPIRE <http://inspirehep.net/> but not in the arXiv e-Print archive)
58. J. D. Bekenstein, Relativistic gravitation theory for the MOND paradigm, *Phys. Rev. D* 70, 083509 (2004), arXiv: astro-ph/0403694
59. B. Famaey and S. McGaugh, Modified Newtonian Dynamics (MOND): Observational phenomenology and relativistic extension, arXiv: 1112.3960 [astro-ph.CO]
60. I. Ferreras, N. Mavromatos, M. Sakellariadou, and M. F. Yusaf, Confronting MOND and TeVeS with strong gravitational lensing over galactic scales: An extended survey, *Phys. Rev. D* 86, 083507 (2012), arXiv: 1205.4880 [astro-ph.CO]
61. J. W. Moffat, Scalar-tensor-vector gravity theory, *JCAP* 0603, 004 (2006), arXiv: gr-qc/0506021
62. C. Tao, Astrophysical constraints on dark Matter, to appear in the proceedings of CYGNUS 2011: 3rd Workshop on directional detection of dark matter (conference: C11-06-08), arXiv: 1110.0298 [astro-ph.CO]
63. <http://zebu.uoregon.edu/~soper/Mass/WIMPS.html>
64. C. Munoz, Direct WIMP search and theoretical scenario, TAUP 2011, <http://taup2011.mpp.mpg.de/?pg=Agenda>
65. M. Drees and G. Gerbier, Dark Matter, review article given in Ref. [13]
66. F. Iocco, M. Pato, and G. Bertone, Evidence for dark matter in the inner Milky Way, *Nat. Phys.* 11, 245 (2015), arXiv: 1502.03821 [astroph.GA]
67. M. Pato and F. Iocco, The dark matter profile of the Milky Way: A non-parametric reconstruction, *Astrophys. J.* 803, L3 (2015), arXiv: 1504.03317 [astrph.GA]
68. J. Silk, The Big Bang, Freeman, 1988
69. J. R. Bond, J. Centgrella, and A. S. Wilson, Dark matter and shocked pancakes, in: Proceedings of the Third Moriond Astrophysics Meeting, Formation and evolution of galaxies and large structures in the universe, edited by J. Audouze and J. Tran Thanh Van, Reidel, Dordrecht, 1984, pp 87–99
70. J. R. Primack and G. R. Blumenthal, What is the Dark Matter, in: Proceedings of the third Moriond Astrophysics Meeting, Formation and evolution of galaxies and large structures in the universe, edited by J. Audouze and J. Tran Thanh Van, Reidel, Dordrecht, 1984, pp 162–183
71. Daniel Chalonge Workshop CIAS Meudon 2010-2014, see: http://chalonge.obspm.fr/Cias_Meudon201X.html, X=0,1,2,3,4
72. G. Gelmini and P. Gondolo, DM production mechanisms, Ch. 7 of Particle Dark Matter: Observations, Models and Searches, edited by G. Bertone, Cambridge University Press, 2010, arXiv: 1009.3690 [astro-ph.CO]
73. H. Baer, K.Y. Choi, E. Kim, and L. Roszkowski, Dark matter production in the early universe: Beyond the thermal WIMP paradigm, *Phys. Rep.* 555, 1 (2014), arXiv: 1407.0017 [hep-ph]
74. E. W. Kolb and M. S. Turner, The Early Universe, Addison-Wesley, 1989
75. G. L. Kane, P. Kumar, B. D. Nelson, and B. Zhang, Dark matter production mechanisms with a non-thermal cosmological history - A classification, arXiv: 1502.05406 [hep-ph]
76. L. D. Duffy and K. Van Bibber, Axions as dark matter particles, *New J. Phys.* 11, 105008 (2009), arXiv: 0904.3346 [hep-ph]
77. P. Sikivie, Dark matter axions, *Int. J. Mod. Phys. A* 25, 554 (2010), arXiv: 0909.0949 [hep-ph]
78. D. Hooper, Kaluza-Klein dark matter, in Proceeding of the Workshop on Exotic Physics with Neutrino Telescopes, 2006, available at: www.physics.uu.se/files/hooper_epnt.pdf
79. K. Griest and M. Kamionkowski, Unitarity limits on the mass and radius of dark-matter particles, *Phys. Rev. Lett.* 64(6), 615 (1990)

80. D. J. H. Chung, E. W. Kolb, and A. Riotto, Non-thermal Supermassive Dark Matter, *Phys. Rev. Lett.* 81, 4048 (1998), arXiv: hep-ph/9805473; WIMPZILLAS! Proceedings of the 2nd International Conference on dark matter in astro and particle physics, arXiv: hep-ph/9810361
81. V. Kuzmin and T. I. Tkachev, Matter creation via vacuum fluctuations in the early universe and observed ultra-high energy cosmic ray events, *Phys. Rev. D* 59, 123006 (1999), arXiv: hep-ph/9809547
82. J. A. Frieman, G. B. Gelmini, M. Gleiser, and E. W. Kolb, Primordial origin of nontopological solitons, *Phys. Rev. Lett.* 60(21), 2101 (1988)
83. A. L. Macpherson and B. A. Campbell, Biased discrete symmetry breaking and Fermi balls, *Phys. Lett. B* 347, 205 (1995), arXiv: hep-ph/9408387
84. R. B. Metcalf and J. Silk, New constraints on macroscopic compact objects as dark matter candidates from gravitational lensing of type Ia supernovae, *Phys. Rev. Lett.* 98(7), 071302 (2007)
85. Goddard Space Flight Center, Dark Matter may be Black Hole Pinpoints. NASA's Imagine the Universe, <http://imagine.gsfc.nasa.gov/docs/features/news/14may04.html>. Retrieved 2008-09-13
86. M. Kesden and S. Hanasoge, Transient solar oscillation driven by primordial black holes, *Phys. Rev. Lett.* 107, 111101 (2011), arXiv: 1106.0011 [astro-ph.CO]
87. J. L. Feng and J. Kumar, Dark-matter particles without weak-scale masses or weak interactions, *Phys. Rev. Lett.* 101(23), 231301 (2008)
88. K. K. Boddy, J. L. Feng, M. Kaplinghat, Y. Shadmi, and T. M. P. Tait, Strongly interaction dark: Self-interactions and keV lines, *Phys. Rev. D* 90, 095016 (2014), arXiv: 1408.6532 [hep-ph]
89. R. Essig, et al., Working Group Report: Dark sectors and new, light, weakly-coupled particles, arXiv: 1311.0029 [hep-ph]
90. G. Steigman and M. S. Turner, Cosmological constraints on the properties of weakly interacting massive particles, *Nucl. Phys. B* 253, 375 (1985)
91. N. Daci, I. De Bruyn, S. Lowette, M. H. G. Tytgat, and B. Zaldivar, Simplified SIMOs and the LHC, arXiv: 1503.05505 [hep-ph]
92. Y. Hochberg, E. Kufflik, T. Volansky, and J. G. Wacker, The SIMP miracle, *Phys. Rev. Lett.* 113, 171301 (2014)
93. N. Bernal, C. Garcia-Cely, and R. Rosenfeld, WIMP and SIMP dark matter from the spontaneous breaking of a globle group, arXiv: 1501.0197 [hep-ph]
94. J. L. Feng, Dark matter candidates from particle physics and methods of detection, *Ann. Rev. Astron. Astrophys.* 48, 495 (2010), arXiv: 1003.0904 [astro-ph.CO]
95. L. Roszkowski, Particle dark matter — A theorist's perspective, *Pramana* 62, 389 (2004)
96. Report on the Direct Detection and Study of Dark Matter, The Dark Matter Scientific Assessment Group, A Joint Sub-panel of HEPAP and AAAC, p. 59, https://www.nsf.gov/mps/ast/aaac/dark_matter_scientific_assessment_group/dmsag/_final_report.pdf
97. S. Tremaine and J. E. Gunn, Dynamical role of light neutral leptons in cosmology, *Phys. Rev. Lett.* 42, 407 (1979)
98. J. Madsen, Phase-space constraints on bosonic and fermionic dark matter, *Phys. Rev. Lett.* 64(23), 2744 (1990) J. Madsen, Generalized Tremaine-Gunn limits for bosons and fermions, *Phys. Rev. D* 44(4), 999 (1991)
99. F. Wilczek, Asymptotic Freedom: From Paradox to Paradigm, Nobel Lecture, December 8, 2004, http://www.nobelprize.org/nobel_prizes/physics/laureates/2004/wilczek-lecture.pdf
100. J. Ellis and K. A. Olive, in: Particle Dark Matter, Observations, Models and Searches, Cambridge University Press, 2010, Ch. 8, Supersymmetric dark matter candidates, arXiv: 1001.3651 [astro-ph.CO]
101. S. Dodelson, Modern Cosmology, Academic Press, 2003; Erratum: http://home.fnal.gov/_dodelson/errata.html
102. F. Bezrukov, Light sterile neutrino dark matter in extensions of the standard model, talk given at the Workshop CIAS Neudon 2011, Warm Dark matter in the galaxies: Theoretical and observational progress, June 8-10, 2011, http://chalonge.obspm.fr/Cias_Meudon2011.html
103. M. R. Lovell, V. Eke, C. S. Frenk, L. Gao, A. Jenkins, T. Theuns, J. Wang, S. D. M. White, A. Boyarsky, and O. Ruchayskiy, The haloes of bright satellite galaxies in a warm dark matter universe, *Mon. Not. R. Astron. Soc.* 420, 2318 (2012), arXiv: 1104.2929 [astro-ph.CO]
104. N. Smith, Status update on deep underground facilities, talk given in TAUP 2011, <http://taup2011.mpp.mpg.de/?pg=Agenda>
105. H. Chen, Underground laboratory in China, *Eur. Phys. J. Plus* 127, 105 (2012)
106. Focus point on Deep Underground Science Laboratories and Projects, edited by A. Bettini, *Eur. Phys. J. Plus* 127, Sep. 2012
107. J. D. Lewin and P. F. Smith, Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil, *Astropart. Phys.* 6, 87 (1996)
108. E. Armengaud, Gif Lectures on direct detection of dark matter, arXiv: 1003.2380 [hep-ph]
109. G. Jungman, M. Kamionkowski, and K. Griest, Supersymmetric dark matter, *Phys. Rep.* 267, 195 (1996), arXiv: hep-ph/9506380
110. K. Freese, J. Frieman, and A. Gould, Signal modulation in cold-dark-matter detection, *Phys. Rev. D* 37(12), 3388 (1988)

111. D. R. Tovey, R. J. Gaitskell, P. Gondolo, Y. A. Rasmachera, and L. Roszkowski. A new model-independent method for extracting spin-dependent cross section limits from dark matter searches, *Phys. Lett. B* 488, 17 (2000), arXiv: hep-ph/0005041
112. Dark Matter Portal, <http://lpsc.in2p3.fr/mayet/dm.php>
113. Dark Matter Hub, <http://www.interactions.org/cms/?pid=1034004>
114. S. C. Kim, et al. (KIMS Collaboration), New limits on interactions between weakly interacting massive particles and nucleons obtained with CsI(Tl) Crystal Detectors, *Phys. Rev. Lett.* 108, 181301 (2012), arXiv: 1204.2646 [astro-ph.CO]
115. 9th International Conference: Identification of Dark Matter, July 23-27, 2012, Chicago. PDF files of the presentations can be found at: <http://kicp-workshops.uchicago.edu/IDM2012/program.php>
116. T. Saab, An introduction to dark matter direct detection searches & techniques, arXiv: 1203.2566 [physics.ins-det]
117. M. Boezio, et al. (PAMELA Collaboration), PAMELA and indirect dark matter searches, *New J. Phys.* 11, 102053 (2009)
118. D. Hooper, D. P. Finkbeiner, and G. Dobler, Evidence of dark matter annihilations in the WMAP haze, *Phys. Rev. D* 76, 083012 (2007), arXiv: 0705.3655 [Astro-ph]
119. M. Cirelli, Indirect search for dark matter: A status review, *Pramana* 79, 1021 (2012), arXiv: 1202.1454 [hep-ph]
120. P. Conversers, P. Nath, and B. Nelson, The hunt for New physics at the Large Hadron Collider, Ch. 5, Connecting Dark Matter to the LHC, arXiv: 1001.2693 [hep-ph]
121. <http://project-slhc.web.cern.ch/project-slhc/about/>
122. A. Ringwald, L. J. Rosenberg, and G. Rybka, Axions and other similar particles, a mini review given in Ref. [13]
123. G. G. Raffelt, Astrophysical axion bounds, *Lect. Notes Phys.* 741, 51 (2008), arXiv: hep-ph/0611350
124. A. Friedland, M. Giannotti, and M. Wise, Constraining the axion-photon coupling with massive stars, *Phys. Rev. Lett.* 110(6), 061101 (2013)
125. G. Raffelt, Viewpoint: Particle physics in the sky, *Physics* 6, 14 (2013)
126. I. N. T. Workshop, 12-50W Vistas in Axion Physics: A Roadmap for theoretical and Experimental Axion Physics through 2050, April 23-26, 2012, Seattle. http://www.int.washington.edu/talks/WorkShops/int_12_50W/
127. C. Athanassopoulos, L. B. Auerbach, D. A. Bauer, R. D. Bolton, B. Boyd, et al., Candidate Events in a Search for Muon Antineutrino to Electron Antineutrino Oscillations, *Phys. Rev. Lett.* 75(14), 2650 (1995), arXiv: nucl-ex/9504002
128. This website entitled LSND is a collection of information related to LSND: <http://www.nu.to.infn.it/exp/all/lsnd/#1>
129. A. A. Aguilar-Arevalo, et al. (MiniBooNE Collaboration), Improved Search for $\nu^- \mu > \nu^- e$ Oscillations in the MiniBooNE Experiment, *Phys. Rev. Lett.* 110, 161801 (2013)
130. K. B. M. Mahn, et al. (MiniBooNE Collaboration), Dual baseline search for muon neutrino disappearance at $0.5 \text{ eV}^2 < \Delta m^2 < 40 \text{ eV}^2$, *Phys. Rev. D* 85, 032007 (2012), arXiv: 1106.5685 [hep-ex]
131. This website contains various information on the sterile neutrino: http://www.nu.to.infn.it/Sterile_Neutrinos/
132. K. N. Abazajian, et al., Light sterile neutrinos: A white paper, arXiv: 1204.5379 [hep-ph]
133. T. Asaka, S. Blanchet, and M. Shaposhnikov, The ν MSM, dark matter and neutrino masses, *Phys. Lett. B* 631, 151 (2005), arXiv: hep-ph/0503065
134. T. Asaka and M. Shaposhnikov, The ν MSM, dark matter and baryon asymmetry of the universe, *Phys. Lett. B* 620, 17 (2005), arXiv: hep-ph/0505013
135. A. Kusenko, Sterile neutrinos: The dark side of the light fermions, *Phys. Rep.* 481, 1 (2009), arXiv: 0906.2968 [hep-ph]
136. A. Boyarsky, D. Iakubovskiy, and O. Ruchayskiy, Next decade of sterile neutrino studies, *Phys. Dark Univ.* 1, 136 (2012), arXiv: 1306.4954 [astro-ph.CO]
137. The Sterile Neutrinos references and useful link website: http://www.nu.to.infn.it/Sterile_Neutrinos/
138. E. Bulbul, et al., Detection of an unidentified emission line in the stacked X-ray spectrum of galaxy clusters, *Astrophys. J.* 789, 13 (2014), arXiv: 1402.2301 [astroph.CO]
139. A. Boyarsky, O. Ruchayskiy, D. Iakubovskiy, and J. Franse, An unidentified line in X-ray of the Andromeda and Spersesus galaxy cluster, *Phys. Rev. Lett.* 113, 251301 (2014), arXiv: 1402.4119 [astro-ph.CO]
140. K. N. Abazajian, Resonantly-produced 7 KeV sterile neutrino dark matte models and the properties of Milky Way satellites, *Phys. Rev. Lett.* 112, 161303 (2014), arXiv: 1403.0954 [astro-ph.CO]
141. R. Bernabei, et al. (DAMA/LIBRA Collaboration), New results from DAMA/LIBRA, *Eur. Phys. J. C* 67, 39 (2010), arXiv: 1002.1028 [astro-ph.GA]
142. R. Angnese, et al. (CDMS Collaboration), Silicon Detector Dark Matter Results from the Final Exposure of CDMS II, *Phys. Rev. Lett.* 111, 251401 (2013), arXiv: 1304.4279 [hep-ex]
143. C. E. Aaseth, et al. (CoGeNT Collaboration), CoGeNT: A Search for Low-Mass Dark Matter using p-type Point Contact Germanium Detectors, *Phys. Rev. D* 88, 012002 (2013), arXiv: 1208.5737 [astro-ph.CO]

144. G. Angloher, et al. (CRESST Collaboration), Results from 730 kg days of the CRESST-II Dark Matter Search, *Eur. Phys. J. C* 72, 1971 (2012), arXiv: 1109.0702 [astro-ph.CO]
145. P. Belli, Results from DAMA/LIBRA and perspectives of phase 2, talk given at Aspen 2013-Closing in on Dark Matter, January 28-February 3, 2013
146. P. Belli, Results and strategies for dark matter investigations, talk at NDM 2015, Jyväskylä, Finland, June 1-5, 2015, http://people.roma2.infn.it/belli/belli_NDM15_jun15.pdf
147. DAMA Collaboration homepage: <http://people.roma2.infn.it/dama/web/home.html>
148. M. Drees and G. Gerbier, Dark matter, a mini review in Ref. [11]
149. K. Blum, DAMA vs. the annually modulated muon background, arXiv: 1110.0857 [astro-ph.HE]
150. J. Klinger and V. A. Kudryavtsev, Muon-induced neutrons do not explain the DAMA data, *Phys. Rev. Lett.* 114, 151301 (2015), arXiv: 1503.07225 [hep-ph]
151. C. Arina, E. Del Nobile, and P. Panci, Dark matter with pseudoscalar-mediated interactions explains the DAMA signal and the galactic center excess, *Phys. Rev. Lett.* 114(1), 011301 (2015)
152. J. Cherwinka, et al. (DM-Ice Collaboration), First data from DM-Ice17, *Phys. Rev. D* 90, 092005 (2014), arXiv: 1401.4804 [astro-ph.IM]
153. R. Agnese, et al. (The SuperCDMS Collaboration), Search for low-mass weakly interacting massive particles with SuperCDMS, *Phys. Rev. Lett.* 112(24), 241302 (2014)
154. P. Cushman, Lecture given at the 2014 SLAC Summer Institute [26], entitled WIMP Direct Detection Searches: Solid State Technologies
155. B. Angloher, et al. (The EURECA Collaboration), EURECA conceptual design report, Physics of the Dark Universe 3, 41-74 (2014), available at: <http://www.sciencedirect.com/science/article/pii/S2212686414000090>
156. C. E. Aalseth, et al. (CoGeNT Collaboration), Results from a Search for Light-Mass Dark Matter with a p -type Point Contact Germanium Detector, *Phys. Rev. Lett.* 106, 131301 (2011), arXiv: 1002.4703 [astro-ph.CO]
157. C. E. Aalseth, et al. (CoGeNT Collaboration), Search for an Annual Modulation in a p -type Point Contact Germanium Dark Matter Detector, *Phys. Rev. Lett.* 107, 141301 (2011), arXiv: 1106.0650 [astro-ph.CO]
158. P. J. Fox, J. Kopp, M. Lisanti, and N. Weiner, A CoGeNT modulation analysis, *Phys. Rev. D* 85, 036008 (2012), arXiv: 1107.0717 [hep-ph]
159. C. McCabe, DAMA and CoGeNT without astrophysical uncertainties, arXiv: 1107.0741 [hep-ph]
160. J. Herrero-Garcia, T. Schwetz, and J. Zupan, Astrophysics independent bounds on the annual modulation of dark matter signals, arXiv: 1205.0134 [hep-ph]
161. E. Aprile, et al. (XENON100 Collaboration), First dark matter result from the XENON100 experiment, arXiv: 1005.0389 [astro-ph.CO]
162. C. C. Aalseth, et al., Search for an annual modulation in three years of CoGeNT dark matter detector data, arXiv: 1401.3295 [astro-ph.CO]
163. J. H. Davis, C. McCabe, and C. Boehm, Quantifying the evidence for dark matter in CoGeNT data, *JCAP* 1408, 014 (2014), arXiv: 1405.0495 [hep-ph]
164. G. Angloher, et al. (CRESST Collaboration), Results from 730 kg days of the CRESST-II Dark Matter Search, *Eur. Phys. J. C* 72, 1971 (2012), arXiv: 1109.0702 [astro-ph.CO]
165. G. Angloher, et al. (CRESST Collaboration), Results on low mass WIMPs using an upgraded CRESST-II detector, *Eur. Phys. J. C* 74 3184 (2014), arXiv: 1407.3146 [astro-ph.CO]
166. X.-J. Bi, P.-F. Yin, and Q. Yuan, Status of dark matter detection, *Front. Phys.* 8, 794 (2013), arXiv: 1409.4590 [hep-ph]
167. M. Boudaud, et al., A new look at the cosmic ray positron fraction, *Astron. Astrophys.* 575, A67 (2015)
168. S.-J. Lin, Q. Yuan, and X.-J. Bi, Quantitative study of the AMS-02 electron/positron spectra: Implications for pulsars and dark matter properties, *Phys. Rev. D* 91, 063508 (2015), arXiv: 1409.6248 [astro-ph.HE]
169. J. Feng and H. H. Zhang, Pulsar interpretation of the lepton spectra measured by AMS-02, arXiv: 1504.03312 [hep-ph]
170. S. Ting, The AMS Experiment, talk given at the AMS Day at CERN, April 15-17, 2015, <http://indico.cern.ch/event/381134/other-view?view=standard>
171. G. Giesen, M. Boudaud, Y. Génolini, V. Poulin, M. Cirelli, P. Salati, and P. D. Serpico, AMS-02 Antiprotons, at last! Secondary astrophysical component and immediate implications for Dark Matter, arXiv: 1504.04276 [astro-ph.HE]
172. K. Hamaguchi, T. Moroi, and K. Nakayama, AMS-02 Antiprotons from Annihilating or Decaying Dark Matter, *Phys. Lett. B* 747, 523 (2015), arXiv: 1504.05937 [hep-ph]
173. L. Bergström, Dark matter evidence, particle physics candidates and detection methods, arXiv: 1205.4882
174. C. Weniger, A tentative Gamma-ray line from dark matter annihilation at the Fermi Large Area Telescope, arXiv: 1204.2797 [hep-ph]
175. M. Ackermann, et al. (Fermi-LAT Collaboration), Constraining Dark Matter Models from a Combined Analysis of Milky Way Satellites with the Fermi Large Area Telescope, *Phys. Rev. Lett.* 107, 241302 (2011), arXiv: 1108.3546 [astro-ph.HE]
176. A. Geringer-Smith and S. M. Koushiappas, Exclusion of canonical WIMPs by the joint analysis of Milky Way dwarfs with Fermi, *Phys. Rev. Lett.* 107, 241303 (2011), arXiv: 1108.2914 [astro-ph.CO]

177. Y-L. S. Tsai, Q. Yuan, and X. Huang, A generic method to constrain the dark matter model parameter from Fermi observations of dwarf spheroids, *JCAP* 1303, 018 (2013), arXiv: 1212.3990 [astro-ph.HE]
178. S. Ando and D. Nagai, Fermi-LAT constraints on dark matter annihilation cross section from observations of the Fornax cluster, *JCAP* 1207, 017 (2012), arXiv: 1201.0753 [astro-ph.HE]
179. Jianxin Han, C. S. Frenk, V. R. Eke, and Liang Cao, Constraining Extended Gamma-ray Emission from Galaxy Clusters, *Mon. Not. R. Astron. Soc.* 427, 1651 (2012), arXiv: 1207.6749 [astro-ph.CO]
180. S. Ando and E. Komatsu, Constraints on the annihilation cross section of dark matter particles from anisotropies in the diffuse gamma-ray background measured with Fermi-LAT, arXiv: 1301.5901 [astro-ph.CO]
181. M. Ackermann, et al. (Fermi-LAT Collaboration), The spectrum of isotropic diffuse gamma ray emission between 100 MeV and 820 GeV, *Astrophys. J.* 799, 86 (2015), arXiv: 1410.3696 [astro-ph.DE]
182. M. Fornasa and M. A. Sanchez-Conde, The nature of the diffuse gamma-ray background, arXiv: 1502.02866 [astro-ph.CO]
183. D. Hooper and T. Linden, On the origin of the gamma rays from the galactic center, *Phys. Rev. D* 84, 123005 (2011), arXiv: 1110.0006 [astro-ph.HE]
184. A. Boyarsky, D. Malyshev, and D. Ruchayskiy, A comment on the emission from the Galactic Center as seen by the Fermi telescope, *Phys. Lett. B* 705, 165 (2011), arXiv: 1012.5839 [hep-ph]
185. K. N. Abazajian and M. Kaplinghat, Detection of a gamma-ray source in the galactic center consistent with extended emission from dark matter annihilation and concentrated astrophysical emission, *Phys. Rev. D* 86, 083511 (2012), Erratum: *Phys. Rev. D* 87, 129902 (2013), arXiv: 1207.6047 [astro-ph.HE]
186. T. Daylan, D. P. Finkbeiner, D. Hooper, T. Linden, S. K. N. Portillo, N. L. Rodd, and T. R. Slatyer, The characterization of the gamma-ray signal from the central Milky Way: A compelling case for annihilating dark matter, arXiv: 1402.6703 [astro-ph.HE]
187. T. Bringmann, X. Huang, A. Ibarra, S. Vogl, and C. Weniger, Fermi-LAT search for internal bremsstrahlung signatures from dark matter annihilation, *JCAP* 1207, 054 (2012), arXiv: 1203.1312 [hep-ph]
188. E. Tempel, A. Hektor, and M. Raidal, Fermi 130 GeV gamma-ray excess and dark matter annihilation in subhaloes and in the galactic center, *JCAP* 1209, 032 (2012), arXiv: 1205.1045 [hep-ph]
189. M. Ackermann, et al., Updated search for spectral lines from galactic dark matter interactions with pass 8 data from the Fermi Large Area Telescope, *Phys. Rev. D* 91, 122002 (2015)
190. N. Prantzos, et al., The 511 KeV emission from positron annihilation in the Galaxy, *Rev. Mod. Phys.* 83, 1001 (2011), arXiv: 1009.4620 [astro-ph.HE]
191. K. Helbing, et al. (The IceCube Collaboration), IceCube as a discovery observatory for physics beyond the standard model, arXiv: 1107.5227 [hep-ex]
192. R. Kappl and M. W. Winkler, New limits on dark matter from Super-Kamiokande, *Nucl. Phys. B* 850, 505 (2011) arXiv: 1104.0679 [hep-ph]
193. M. G. Aartsen, et al., Search for dark matter annihilations in the Sun with the 79-string IceCube detector, *Phys. Rev. Lett.* 110, 131302 (2013), arXiv: 1212.4097 [astro-ph.HE]
194. K. Choi, et al. (Super-K Collaboration), Search for neutrinos from annihilation of captured low-mass dark matter particles in the Sun by Super-Kamiokande, arXiv: 1503.04858 [hep-ex]
195. M. G. Aartsen, et al. (IceCube Collaboration), Search for dark matter annihilation in the galactic center with IceCube-79, arXiv: 1505.07259 [astro-ph.HE]
196. F. Donato, N. Fornengo, and P. Salati, Antideuterons as a signature of supersymmetric dark matter, *Phys. Rev. D* 62, 043003 (2000), arXiv: hep-ph/9904481
197. M. Kadastik, M. Raidal, and A. Strumia, Enhanced anti-deuteron Dark Matter signal and the implications of PAMELA, *Phys. Lett. B* 683, 248 (2010), arXiv: 0908.1578 [hep-ph]
198. T. Aramaki, et al., Review of the theoretical and experimental status of dark matter identification with cosmic-ray antideuterons, arXiv: 1505.07785 [hep-ph]
199. J. B. Billard and E. Figueroa-Feliciano, Implication of neutrino backgrounds on the reach of next generation dark matter direct detection experiments, *Phys. Rev. D* 89, 023524 (2024), arXiv: 1307.5458 [hep-ph]
200. M. Schumann, Dark Matter 2014, *EPJ Web Conf.* 96, 01027 (2015), arXiv: 1501.01200 [astro-ph.CO]
201. E. Aprile, et al. (XENON Collaboration), Physics reach of the XENON1T dark matter experiment, *JCAP* 04, 027 (2016), arXiv: 1512.07501 [physics.ins-det]
202. L. Baudis, et al., Neutrino physics with multi-ton scale liquid xenon detector, *JCAP* 1401, 044 (2014), arXiv: 1309.7024 [physics.ins-det]
203. A. Kish, Direct Dark Matter Detection with Xenon and DARWIN Experiment, *PoS TIPP* 2014, 164 (2014), C14-06-02 Proceedings
204. D. S. Akerib, et al. (LZ Collaboration), LUX-ZEPLIN (LZ) Conceptual Design Report, LBNL-190005, arXiv: 1509.02910 [physics.ins-det]
205. D. Bauer, et al., Snowmass CF1 Summary: WIMP Dark Matter Direct Detection, arXiv: 1310.8327 [hep-ex]
206. L. Hsu, Direct searches for dark matter, plenary talk given at the ICHEP 2012, <https://indico.cern.ch/conferenceTimeTable.py?confId=181298#20120705>. detailed
207. R. Aaij, et al., First evidence for the decay $B_s^0 \rightarrow \mu^+ \mu^-$, *Phys. Rev. Lett.* 110(2), 021801 (2013)

208. K. Hara, et al. (Belle Collaboration), Evidence for $B^- \rightarrow \tau^- \bar{\nu}_\tau$ with a Hadronic Tagging Method Using the Full Data Sample of Belle, *Phys. Rev. Lett.* 110, 131801 (2013), arXiv: 1208.4678 [hep-ex]
209. A. Dighe, D. Ghosh, K. M. Patel, and S. Raychaudhuri, Testing Times for Supersymmetry: Looking under the Lamp Post, arXiv: 1303.0721 [hep-ph]
210. G. Rolandi, LHC Results – Highlights, Lecture given at the European School of High Energy Physics (ES-HEP2012), June 2012, Anjou, France, arXiv: 1211.3718 [hep-ex]
211. A. V. Gladyshev and D. I. Kazakov, Is (low Energy) SUSY still alive? Lecture given at the European School of High Energy Physics (ESHEP2012), June 2012, Anjou, France, arXiv: 1212.2548 [hep-ex]
212. P. Bectle, T. Plehn, and C. Sander, The Status of Supersymmetry after the LHCC Run 1, arXiv: 1506.03091 [hep-ex]
213. B. W. Lee and S. Weinberg, Cosmological Lower Bound on Heavy-Neutrino Masses, *Phys. Rev. Lett.* 39, 165 (1977)
214. D. E. Kaplan, M. A. Luty, and K. M. Zurek, Asymmetric Dark Matter, *Phys. Rev. D* 79, 115016 (2009), arXiv: 0901.4117 [hep-ph]
215. H. Davoudiasl and R. N. Mohapatra, On Relating the Genesis of Cosmic Baryons and Dark Matter, *New J. Phys.* 14, 095011 (2012), arXiv: 1203.1247 [hep-ph]
216. P. Gondolo, Theory of low mass WIMPs, talk given at UCLA Dark Matter 2012 <https://hepconf.physics.ucla.edu/dm12/agenda.html>
217. N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, A Theory of Dark Matter, *Phys. Rev. D* 79, 015014 (2009), arXiv: 0810.0713 [hep-ph]
218. S. Weinberg, Gravitation and Cosmology, Principles and Applications of the General Theory of Relativity, John Wiley & Sons, 1972
219. P. J. E. Peebles, Principles of Physical Cosmology, Princeton University Press, 1993
220. A. Linde, Particle Physics and inflationary Cosmology, Hardwood, Chur, Switzerland, 1990
221. S. Weinberg, Cosmology, Oxford University Press, 2008. Erratum: http://zippy.ph.utexas.edu/_weinberg/corrections.html
222. L. Bergström and A. Goobar, Cosmology and Particle Astrophysics, Second Edition, Springer, 2003
223. V. Mukhanov, Physical Foundations of Cosmology, Cambridge University Press, 2005
224. D. H. Lyth and A. R. Liddle, The Primordial Density Perturbation, Cosmology, Inflation, and the Origin of Structure, Cambridge University Press, 2009
225. A. Liddle and J. Loveday, Oxford Companion to Cosmology, Oxford University Press, 2008
226. K. W. Ford and J. A. Wheeler, Geons, Black Holes, and Quantum Foam: A Life in Physics, W.W. Norton & Company, Inc., 1998, p. 235
227. E. Hubble, A relation between distance and radial velocity among extra-galactic nebulae, *Proc. Natl. Acad. Sci. USA* 15(3), 168 (1929)
228. P. P. Penzias and R. W. Wilson, A measurement of excess antenna temperature at 4048-Mc/s, *Astrophys. J.* 142, 419 (1965)
229. J. C. Mather, et al., A preliminary measurement of the cosmic microwave background spectrum by the Cosmic Background Explorer (COBE) satellite, *Astrophys. J.* 354, 237 (1990); J. C. Mather, et al., Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument, *Astrophys. J.* 420, 439 (1994)
230. S. Weinberg, The First Three Minutes, Basic Books, 1993
231. A. G. Riess, et al., Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant, *Astron. J.* 116, 1009 (1989)
232. S. Perlmutter, et al., Measurements of Ω and Λ from 42 High-Redshift Supernovae, *Astrophys. J.* 517, 565 (1999)
233. http://en.wikipedia.org/wiki/Stress-energy_tensor
234. S. M. Carroll, Lecture Notes on General Relativity, Ch. 8, arXiv: gr-qc/9712019, and <http://preposterousuniverse.com/grnotes/>
235. V. Eric, Linder, First Principles of Cosmology, Addison-Wesley, 1997
236. P. A. R. Ade, et al. (Planck Collaboration), Planck 2013 results. XVI. Cosmological parameters, arXiv: 1303.5076 [astro-ph.CO]
237. C. Kittel and H. Kroemer, Thermal Physics, W.H. Freeman and Company, 1998
238. K. A. Olive, The violent Universe: The Big Bang, lectures given at the 2009 European School of High-Energy Physics, Bautzen, Germany, June 2009, arXiv: 1005.3955 [hep-ph]
239. R. J. Scherer and M. S. Turner, On the relic, cosmic abundance of stable, weakly interacting massive particles, *Phys. Rev. D* 33, 1585 (1986); Erratum: *Phys. Rev. D* 34, 3263 (1986)
240. B. S. Ryden, Introduction to Cosmology, January 13, 2006. <http://www.if.ufrgs.br/oei/santiago/fis02012/Introduction-Cosmology-Ryden.pdf>. This is presumably the web form of the same book published in 2003.
241. L. Verde, et al. (WMAP Collaboration), First Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters, *Astrophys. J. Suppl.* 148, 175 (2003), arXiv: astro-ph/0302209
242. J. R. III Gott, M. Jurič, D. Schtegel, F. Hoyle, M. Vogele, M. Tegmark, N. Bachall, and J. Brinkmann, A Map of the Universe, *Astrophys. J.* 624, 463 (2005), arXiv: astro-ph/0310571

243. R. E. Alpher, H. Bethe, and G. Gamov, The origin of chemical elements, *J. Wash. Acad. Sci.* 38(8), 288 (1948)
244. K. Jedamzik and M. Pospelov, Big bang nucleosynthesis and particle dark matter, *New J. Phys.* 11, 105028 (2009), arXiv: 0906.2087 [hep-ph]
245. K. Jedamzik and M. Pospelov, Big bang nucleosynthesis as a probe of new physics, *Ann. Rev. Nucl. Sci.* 60, 539 (2010), arXiv: 1011.1054 [hep-ph]
246. C. L. Bennett, et al. (WMAP Collaboration), First year Wilkinson Microwave Anisotropy probe (WMAP) observations: Preliminary maps and basic results, *Astrophys. J. Suppl.* 148, 1 (2003), arXiv: astro-ph/0302207
247. The following articles are given in Footnote 1, pp 159–160, [221] to which we refer for more details: V. F. Mukhanov, arXiv: astrp-ph/0303073, G. Steigman, arXiv: astro-ph/0307244, arXiv: astro-ph/0308511, arXiv: astro-ph/0501591, and arXiv: astro-ph/0511534
248. K. A. Olive, G. Steigman, and T. P. Walker, Primordial Nucleosynthesis: Theory and Observations, *Phys. Rep.* 333, 389 (2000), arXiv: astro-ph/9905320
249. D. N. Schramm and M. S. Turner, Big-bang nucleosynthesis enters the precision era, *Rev. Mod. Phys.* 70, 303 (1998), arXiv: astro-ph/9706069
250. S. Sarkar, Big Bang nucleosynthesis and physics beyond the Standard Model, *Rep. Prog. Phys.* 59, 1493 (1996), arXiv: hep-ph/9602260
251. J. Bernstein, L. S. Brown, and G. Feinberg, Cosmological helium production simplified, *Rev. Mod. Phys.* 61, 25 (1989)
252. R. Esmailzaadeh, G. D. Starkman, and S. Dimopoulos, Primordial nucleosynthesis without a computer, *Astrophys. J.* 387, 504 (1991)
253. V. Mukhanov, Nucleosynthesis without a computer, *Int. J. Theor. Phys.* 43, 669 (2003), arXiv: astro-ph/0303073
254. C. Hayashi, Proton-neutron concentration ratio in the expanding universe at the stages preceding the formation of the elements, *Prog. Theor. Phys.* 5, 224 (1950)
255. G. Steigman, D. N. Schramm, and J. Gunn, Cosmological limits to the number of massive leptons, *Phys. Lett. B* 66, 202 (1977)
256. K. A. Olive and G. Steigman, A new look at neutrino limits from big bang nucleosynthesis, *Phys. Lett. B* 354, 357 (1995)
257. B. W. Carroll and D. A. Ostlie, An Introduction to Modern Astrophysics, Pearson Education, 2007, Ch. 29
258. J. Bernstein, L. S. Brown, and G. Feinberg, Cosmological heavy-neutrino problem, *Phys. Rev. D* 32(12), 3261 (1985)
259. D. A. Dicus, E. W. Kolb, and V. L. Teplitz, cosmological upper bound on heavy-neutrino lifetimes, *Phys. Rev. Lett.* 39, 168 (1977)
260. E. W. Kolb and K. A. Olive, Lee-Weinberg bound re-examined, *Phys. Rev. D* 33(4), 1202 (1986) (in INSPIRE Search, fulltext available at the Fermilab Library Server)
261. M. T. Ressell and M. S. Turner, *Comments Astrophys.* 14, 323 (1990), *Bull. Am. Astron. Soc.* 22, 753 (1990) [Fermilab-pub-89/214-A, Oct. 1989]
262. A. Lasenby, Physics of Primary CMB Anisotropy, <http://www.iac.es/congreso/isapp2012/media/Lasenby-lectures/Lasenby.pdf>
263. J. C. Mather, et al., A preliminary measurement of the cosmic microwave background spectrum by the Cosmic Background Explorer (COBE) Satellite, *Astrophys. J.* 354, L37 (1990)
264. G. F. Smoot, et al., Structure in the COBE Differential Microwave Radiometer First-year Maps, *Astrophys. J.* 396, L1 (1992)
265. R. Adam, et al. (Planck Collaboration), Planck 2015 results. I. Overview of products and Scientific results, arXiv: 1502.01582 [astro-ph.CO]
266. P. A. R. Ade, et al. (Planck Collaboration), Planck 2015 results. XIII. Cosmological parameters, arXiv: 1502.01589 [astro-ph.CO]
267. <http://lambda.gsfc.nasa.gov>
268. See: http://www.astro.ucla.edu/_wright/CMB-dipole-history.html
269. Y. Itoh, K. Yahata, and M. Takada, A dipole anisotropy of galaxy distribution: Does the CMB rest-frame exist in the local universe? *Phys. Rev. D* 82, 043530 (2010), arXiv: 0912.1460 [astro-ph. CO]
270. G. Hinshaw, et al. (WMAP Collaboration), Five-Year Wilkinson Microwave Anisotropy Probe Observations: Data processing, sky maps, and basic results, *Astrophys. J. Suppl.* 180, 225 (2009)
271. Ya. B. Zel'dovich and R. A. Sunyaev, The Interaction of Matter and Radiation in a Hot-Model Universe, *Astrophys. Space Sci.* 4, 301 (1969); R. A. Sunyaev and Ya. B. Zel'dovich, The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies, *Comments Astrophys. Space Phys.* 2, 173 (1972)
272. J. E. Carlstrom, G. P. Hoder, and E. D. Reese, Cosmology with the Sunyaev–Zel'dovich Effect, *Annu. Rev. Astron. Astrophys.* 40, 643 (2002), arXiv: astro-ph/0208192
273. R. K. Sachs and A. M. Wolfe, Perturbations of a cosmological model and angular variations of the microwave background, *Astrophys. J.* 147, 73 (1987)
274. R. K. Sachs and A. M. Wolfe, Perturbations of a cosmological model and angular variations of the microwave background, *Gen. Relativ. Gravit.* 39, 1929 (2007)
275. A. J. Nishizawa, Integrated Sachs Wolfe Effect and Rees Sciama Effect, *Prog. Theor. Exp. Phys.* 2014, 06B110 (2014), arXiv: 1404.5102 [astro-ph.CO]
276. W. Hu, CMB anisotropy tutorial: <http://background.uchicago.edu/whu/intermediate/>
277. D. Langlois, Isocurvature cosmology perturbation and the CMB, *C. R. Phys.* 4, 953 (2003)

278. C. Gordon, Adiabatic and entropy perturbations in cosmology, arXiv: astro-ph/0112523
279. W. Hu and M. White, Acoustic signatures in the cosmic microwave background, *Astrophys. J.* 471, 30 (1996), arXiv: astro-ph/9602019
280. J. A. Peacock, Large-scale surveys and cosmic structure, 7. Anisotropies in the CMB <http://ned.ipac.caltech.edu/level5/Sept03/Peacock/Peacock7.html>
281. M. White, Big Bang Acoustics: Sound for the new born univers, 8. Removing Distortion http://www.astro.virginia.edu/dmw8f/BBA_web/index_frames.html
282. H. Kurki-Suonio, Cosmology I & II. See Chapters 12 and 13, http://theory.physics.helsinki.fi/_cpt/
283. U. Seljak, A Two-Fluid Approximation for Calculating the Cosmic Microwave Background Anisotropies, *Astrophys. J.* 435, L87 (1994), arXiv: astro-ph/9406050
284. W. Hu and N. Sugiyama, Anisotropies in the Cosmic Microwave Background: An Analytic Approach, *Astrophys. J.* 444, 489 (1995), arXiv: astro-ph/9407093
285. M. White and J. D. Cohn, TACMB-1: The Theory of Anisotropies in the Cosmic Microwave Background (Bibliographic Resource Letter), *Am. J. Phys.* 70, 106 (2002), arXiv: astro-ph/0203120
286. See: http://en.wikipedia.org/wiki/Table_of_spherical_harmonics
287. A wealth information on the Cosmic Background Explorer (CORE) is accessible on the web. For instance a quick description can be found in http://en.wikipedia.org/wiki/Cosmic_Background_Explorer
288. Information on the Wilkinson Microwave Anisotropy probe (WMAP) can be found on its official site, <http://map.gsfc.nasa.gov/>. A description of the angular resolutions at different frequency ranges: http://map.gsfc.nasa.gov/mission/observatory_res.html or a brief complete description of the experiment: http://en.wikipedia.org/wiki/Wilkinson_Microwave_Anisotropy_Probe
289. The Planck Spacecraft is an European Space Agency (ESA) experiment for the observation of the CMB anisotropy, launched in May 2009. For a general description, see official sites: <http://www.rssd.esa.int/index.php?project=Planck> http://www.esa.int/esaSC/120398_index_0_m.html. The angular resolution varies with the frequency. For the high frequency instrument the angular resolution varies from 10 to 5 arcmin. The low frequency instrument has lower resolution. For a general description, see: [http://en.wikipedia.org/wiki/Planck_\(spacecraft\)](http://en.wikipedia.org/wiki/Planck_(spacecraft))
290. I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and Products, Corrected and Enlarged Edition, Academic Press, 1980
291. M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, Dover Pub., 1970
292. D. Baumann Cosmology part III Mathematical Tripos, pp 80–81, <http://www.damtp.cam.ac.uk/user/db275/Cosmology/Lectures.pdf>
293. H. Kurki-Suonio, Cosmological Perturbation Theory, Sept. 30, 2012, <http://www.helsinki.fi/hkurkis/CosPer.pdf/>
294. H. Kodama and M. Sasaki, Cosmological Perturbation Theory, *Prog. Theor. Phys. Suppl.* 78, 1 (1984)
295. J. Fritz, An introduction to the theory of hydrodynamic limits, <http://www.math.bme.hu/jofri/JOFRI/PUBLI/je.pdf>
296. U. Seljak and M. Zaldarriaga, A line of sight approach to cosmic microwave background anisotropies, *Astrophys. J.* 469, 437 (1996), arXiv: astro-ph/9603033
297. A. Lewis, A. Challinor, and A. Lasenby, Efficient Computation of CMB anisotropies in closed FRW models, *Astrophys. J.* 538, 473 (2000), arXiv: astro-ph/9911177 10.1086/309179
298. M. Doran, CMBEASY: an Object Oriented Code for the Cosmic Microwave Background, *JCAP* 0510, 011 (2005), arXiv: astro-ph/0302138
299. J. Lesgourgues, The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview, arXiv: 1104.2932 [astro-ph.CO]
300. D. Blas, J. Lesgourgues, and T. Tram, The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes, arXiv: 1104.2933 [astro-ph.CO]
301. J. Lesgourgues, The Cosmic Linear Anisotropy Solving System (CLASS) III: Comparison with CAMB for LambdaCDM, arXiv: 1104.2934 [astro-ph.CO]
302. J. Lesgourgues and T. Tram, The Cosmic Linear Anisotropy Solving System (CLASS) IV: Efficient implementation of non-cold relics, arXiv: 1104.2935 [astro-ph.CO]
303. J. Lesgourgues, The Cosmic Linear Anisotropy Solving System (CLASS) III: Comparison with CAMB for LambdaCDM, arXiv: 1104.2934 [astro-ph.CO]
304. The Planck news release on the epoch of the first stars can be found at http://www.esa.int/Our_Activities/Space_Science/Planck/Planck_reveals_first_stars_were_born_late. See also the news report from the University of Cambridge: <http://www.cam.ac.uk/research/news/planck-reveals-first-stars-were-born-late>
305. The list of Planck publications is at <http://www.cosmos.esa.int/web/planck/publications>.
306. S. Zaroubi, The Epoch of Reionization, arXiv: 1206.0267 [astro-ph.CO]
307. D. Scott and G. F. Smoot, Cosmic Microwave Background, PDB [13]
308. P. A. R. Ade, et al. (Planck Collaboration), Planck 2013 results. I. Overview of products and scientific results, *Astron. Astrophys.* 571, A1 (2014), arXiv: 1303.5062 [astro-ph.CO]
309. P. A. R. Ade, et al. (Planck Collaboration), Planck 2013 results. XV. CMB power spectra and likelihood, *Astron. Astrophys.* 571, A15 (2014), arXiv: 1303.5075 [astro-ph.CO]

310. G. Hinshaw, et al. (WMAP Collaboration), Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results, *Astrophys. J. Suppl.* 208, 19 (2013), arXiv: 1212.5226 [astro-ph.CO]
311. C. L. Bennett, et al. (WMAP Collaboration), Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results, *Astrophys. J. Suppl.* 208, 20 (2013), arXiv: 1212.5225 [astro-ph.CO]
312. For a list of WMAP scientific publications, see: http://lambda.gsfc.nasa.gov/product/map/current/map_bibliography.cfm
313. D. Larson, J. L. Weiland, G. Hinshaw, and C. L. Bennett, Comparing Planck and WMAP9: Maps, Spectra, and Parameters, arXiv: 1409.7718 [astro-ph.CO]
314. D. W. Hogg, Distance measures in cosmology, arXiv: astro-ph/9905116
315. T. M. Davis and C. H. Lineweaver, Expanding Confusion: common misconceptions of cosmological horizons and the superluminal expansion of the Universe, *PASA* 21, 97 (2004), arXiv: astro-ph/0310808
316. C. H. Lineweaver and T. M. Davis, Misconceptions about the big bang, *Sci. Am.* 292(3), 24 (2005)
317. W. Rindler, Visual Horizons in World models, *Mon. Not. Roy. Ast. Soc.* 116, 662 (1956)
318. A. Loeb, The Long-Term Future of Extragalactic Astronomy, *Phys. Rev. D* 65, 047301 (2002), arXiv: astro-ph/0107568
319. <http://www.nasa.gov/news/releases/archives/index.html#.UexHS237tuD>
320. H. Bradt and S. Olbert, Liouville's Theorem, Suppl. to Ch. 3 of *Astrophysical Processes* by the same authors. The article can be found at https://www.cfa.harvard.edu/scranner/Ay201a_2014/LecNotes/supp_liouville_Bradt.pdf
321. J. D. Jackson, *Classical Electrodynamics*, Second Edition, John Wiley & Sons, Inc., 1975
322. H. Goldstein, *Classical Mechanics*, Addison-Wesley, 1950, pp 266–268
323. C. Kittel and H. Kroemer, *Thermal Physics*, W.H. Freeman and Company, 1998
324. S. Dodelson and M. S. Turner, Nonequilibrium neutrino statistical mechanics in the expanding Universe, *Phys. Rev. D* 46(8), 3372 (1992)
325. Wolfram Alpha, Wolfram Research Company: <http://www.wolframalpha.com>