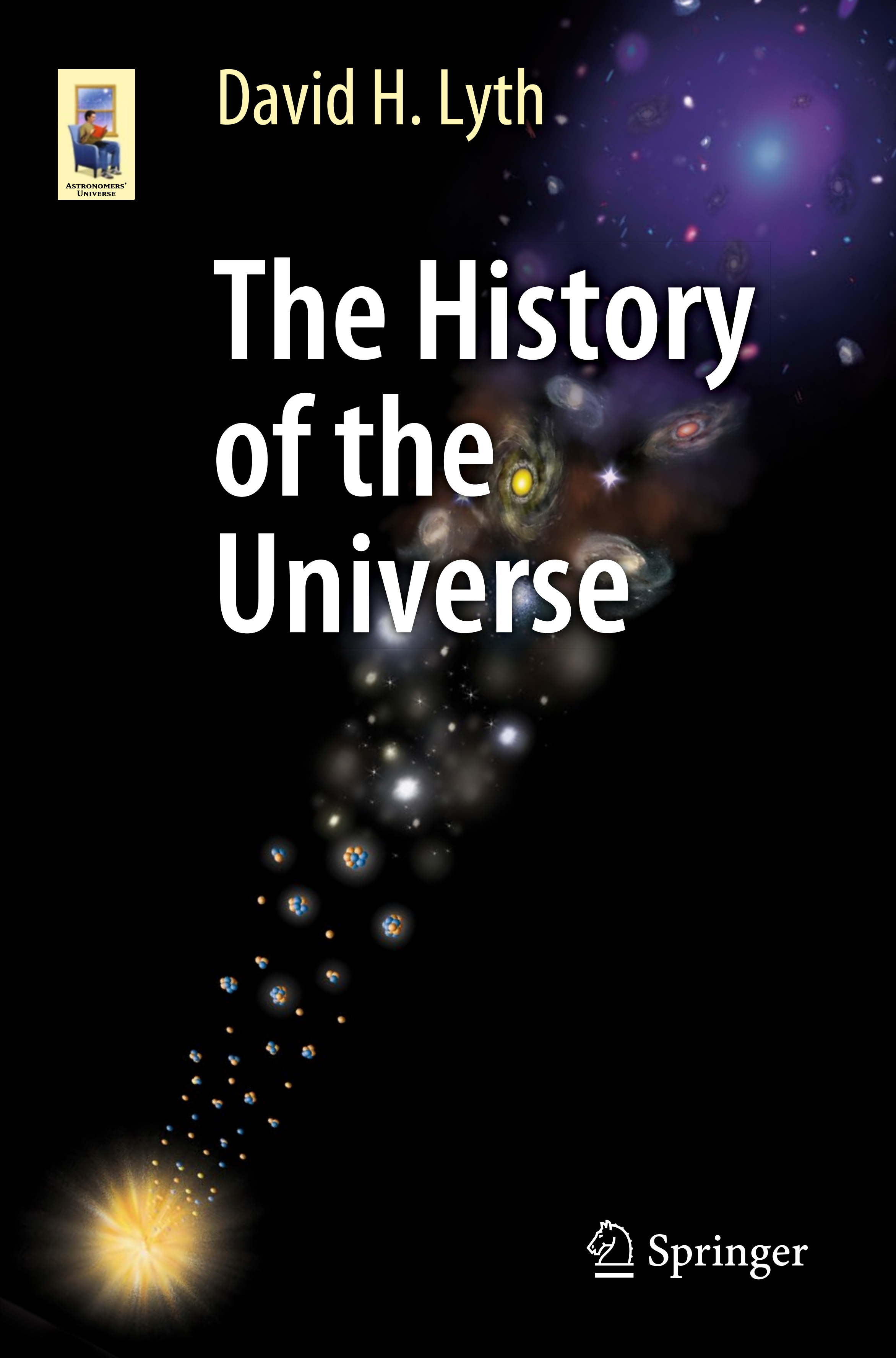
Astronomers' Universe

David H. Lyth

The History of the Universe



Springer

Astronomers' Universe

More information about this series at http://www.springer.com/series/6960

David H. Lyth

The History of the Universe

Springer

David H. Lyth
Physics
Lancaster University
Lancaster, United Kingdom

ISSN 1614-659X ISSN 2197-6651 (electronic)
Astronomers' Universe
ISBN 978-3-319-22743-6 ISBN 978-3-319-22744-3 (eBook)
DOI 10.1007/978-3-319-22744-3

Library of Congress Control Number: 2015950857

Springer Cham Heidelberg New York Dordrecht London

Cover illustration: Cosmic Timeline Illustration.
Image credit: NASA/CXC/M.Weiss. Used with kind permission.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Foreword

The way we view our world is changing dramatically. In the beginning of the previous century, we did not even know that other galaxies exist; some prominent astronomers thought that the Milky Way galaxy is the whole Universe. Indeed, our galaxy is incredibly large. It consists of about a hundred billion stars, many of which have planetary systems. Can we imagine anything bigger than that? Well, now we know that there are about a hundred billion of galaxies in the observable part of the Universe.

For a long time people believed that the Universe is static. Indeed, what else could it be? When we study physics at school, our teachers show us a three-dimensional reference frame with axes labeled by x, y, and z and explain how one can describe motion of particles with respect to it. This suggests that space and time are just a static set of coordinates. This picture was challenged by Albert Einstein in his general theory of relativity, but the consequences of this change were so dramatic that even Einstein could not fully appreciate it. When a mathematician Alexander Friedmann solved Einstein's equations and found that they describe an expanding Universe, Einstein at first thought that this was a mistake, and then he attempted to modify his own theory in an effort to make the Universe static.

Later on, thanks to the observations made by Edwin Hubble, we learned that galaxies move away from each other, and the theory of an expanding Universe became universally accepted. Gradually it evolved into the theory of the hot Big Bang, describing the creation of the Universe as an explosion from the cosmological singularity. The Universe was supposed to behave like a huge ball of fire expanding in all directions and cooling down. The echo of this primeval explosion was captured in 1965 and called the Cosmic Microwave Background Radiation (CMB). But where was the Universe expanding from? The theory of the hot Big Bang assumed that the total number of elementary particles, which constitute all matter in the Universe, did not change much since the moment of its creation. But how could it happen that in the beginning there was no Universe and then suddenly all particles necessary for creation of billions of galaxies emerged from nowhere? And not only that, but this was done in a rather "orderly" way: Density of matter in far removed parts of the world is practically the same, as if the newborn Universe was polished with an incredible accuracy before the Big Bang explosion.

In the 1980s, the theory of the hot Big Bang was replaced by inflationary theory. In the beginning, it looked like a science fiction story. According to this theory, the newborn Universe was in an energetic vacuum-like state. Most of the energy was contained in a special kind of matter called "scalar field." Less than a milligram of such matter is sufficient to create all matter in our Universe after an explosive exponentially rapid expansion of space called inflation. This rapid expansion stretched all previously existing inhomogeneities and made the Universe huge and nearly exactly uniform.

This could seem obviously wrong: How could it be possible to create everything from practically (or literally) nothing? What's about energy conservation? Is it possible for the Universe to expand so fast that its distant parts move away from each other with the speed faster than the speed of light? As if to add insult to injury, cosmologists proposed that galaxies were created due to amplification of tiny quantum fluctuations produced during inflation. Many people objected, saying that nothing classical could be created from quantum. And yet, during the last 30 years this theory has passed numerous theoretical consistency checks, and many of its predictions have been already confirmed by cosmological observations. It is gradually becoming the leading cosmological paradigm for the description of creation of the Universe and formation of its large-scale structure.

One of the most unexpected consequences of inflationary theory is that quantum fluctuations responsible for galaxy formations sometimes may be powerful enough to create new parts of inflationary Universe, rendering the Universe as a huge eternally growing self-reproducing fractal. Inflationary Universe becomes a multiverse, consisting of many exponentially large parts with different properties.

And each of these parts has many secrets which we are just beginning to uncover. Previously we thought that the Universe is guided by the famous principle "what you see is what you get." But cosmological observations have shown that we see only about 5 % of all matter in our part of the Universe. Approximately 27 % of all matter is in an invisible state called dark matter, and 68 % is in yet another invisible state, called dark energy.

Recent progress in this area is so significant that following it is quite difficult. Most of the new information is spread among thousands of scientific papers. It is summarized in some technical books written for cosmologists actively working in the field. Two of such books, *Cosmological Inflation and Large-Scale Structure* and *The Primordial Density Perturbation: Cosmology, Inflation and the Origin of Structure* by David Lyth and Andrew Liddle, achieved their goal beautifully, helping to educate a new generation of physicists in all matters related to cosmology. The success of these books was not surprising: Both of their authors are prominent scientists. David Lyth is one of the leading authorities in inflationary cosmology and the theory of inflationary perturbations responsible for creation of galaxies. His works created a new language which is often used as a bridge between theorists and observers.

The goal of the book by David Lyth *The History of the Universe* is to describe the new vision of the world in a simple, accessible, and reliable way. This is not an easy task. One cannot achieve it without making many simplifications which can easily produce a nice but blurry and imprecise image. That is why it is especially important to have a real expert to do this work. As the author says in the Preface, “the account of physics is meant to be accessible to anybody.” However, deep knowledge and understanding of the material as well as the experience with presenting this material in his earlier books has allowed David Lyth to make just the right amount of simplification to keep the book *The History of the Universe* readable and understandable by everyone. And yet the book is very accurate, rigorous, and informative; it provides a unique perspective, which, I believe, will make it very useful even for those who know cosmology well.

Andrei Linde

Preface

This book is about the history of the Universe, not only what happened but why it happened. In other words, it is about cosmology which is actually a branch of physics.

Ever since I first met them, I have been fascinated by physics and cosmology. With physics that was when I was 11 years old and with cosmology when I was 43 in 1983. Before encountering cosmology I worked on the theory of elementary particles, whose collisions are observed at machines like the one at CERN in Geneva. That turned out to be useful, because the collisions now observed in the laboratory happened also in the early Universe.

When I first encountered cosmology, it had recently been suggested that all of the structure in the Universe originated as a random quantum process, taking place at some very early time. The idea was both simple and remarkable, and I quickly published some research on it. That was easier in those days than it is now because the subject had hardly any previous history and there were only a few relevant papers.

More research followed, and along the way a couple of textbooks with my colleague Andrew Liddle. Those are academic books aimed mostly at other researchers. This book, by contrast, is intended to be accessible to everyone. As we need physics for cosmology, a few chapters of the book are devoted to it, and more physics is introduced as we go along. If you know physics already, you will be able to skip some or all of those bits.

I have many people to thank for questions that prompted the writing of this book. For invaluable comments on drafts of the book at various stages, I thank Phil Furneaux, Brian Martin, Vince Higgs, my wife Margaret, and my brother Peter. Finally, I am indebted to people in Lancaster for help with the figures, in particular Lingfei Wang whose continuous attention has been invaluable.

Lancaster, UK David H. Lyth

Contents

Chapter 1
The Universe

> From our home on the Earth, we look out into the distances and strive to imagine the sort of world into which we were born. . . .. The search will continue. . . .. It is not satisfied and it will not be suppressed.
> (Edwin P. Hubble in 'Realm of the Nebulae', 1936)

I suppose that we humans have always tried to understand what's around us. Until recently though, we didn't achieved much. As far as can be told from the written record, our ideas about the nature of physical reality were mostly way off course, and ideas about the Universe beyond the Earth were completely wrong.

The first step to understanding the Universe was not taken until the early sixteenth century, when Copernicus realized that the passage of the planets across the sky implies that both they and the Earth are going around the Sun in roughly circular orbits. Now, five centuries later, we know much more.

We know the geography of the Universe. We are on the edge of a galaxy, which contains billions of stars more or less like the Sun. The stars you see in the sky are a small sample of those stars. There are billions of other galaxies, more or less like ours, separated from each other by vast distances.

We know the history of the Universe, starting from a simple state of affairs when it was about a hundredth of a second old. I'll be calling that the **known history**. The history is a long one, because the Universe is now about 13.8 billion years old. Throughout the known history, the Universe has been expanding. At very early times, there were no galaxies, but just a nearly homogeneous gas.[1] It wasn't quite homogeneous though, and that is what led to the formation of galaxies. Each galaxy began as a slightly over-dense region, which attracted more and more matter towards it through the force of gravity until it finally collapsed under its own weight.

---

[1]In physics 'homogeneous' means 'the same everywhere', which is one of its ordinary English meanings.

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_1

When a galaxy first forms it's just a cloud of gas. Then over-dense regions within it collapse, to form stars. There are many different types of star, a veritable zoo. The study of individual objects like stars, or even galaxies, is called astrophysics and is not regarded as part of cosmology which studies the Universe as a whole. Just to satisfy your curiosity though, I'll describe the zoo briefly in Chap. 5.

We know, with one proviso, what the Universe is made of. There is electromagnetic radiation, which can be regarded as consisting of particles called photons, and there are also some other particles called neutrinos. There's also ordinary matter, that you and I are made of. Finally, there's what is called Cold Dark Matter or **CDM**. The CDM is called Cold because its particles have hardly any random motion, corresponding to a very low temperature. It's called Dark because it doesn't emit any radiation. But it doesn't absorb any either, so it should really be called Cold Transparent Matter. Even more remarkably, there is no discernable effect when a CDM particle encounters a particle of ordinary matter, even though such encounters must be taking place all the time. As a result of all this, we know of the CDM only through its gravitational effect, and we don't know anything about the nature of the particles that it's made of. (That's the proviso.)

We know what fraction of the Universe's energy is carried by each of those components, and now comes a surprise. The fractions don't add up to 1, because the vacuum itself has energy. The energy density of the vacuum is called the **Cosmological Constant**. As with the CDM, we know that the Cosmological Constant is there because of its gravitational effect. It's called a constant because it doesn't change with time. This is in contrast to the contribution of particles to the energy density. That contribution decreases with time as the particles move apart with the expansion.

Our biggest single source of information about cosmology came from what's called the **Cosmic Microwave Background** (CMB), shown in Fig. 1.1. This is electromagnetic radiation, emitted when the Universe was only about a third of a million years old long before galaxy formation. The CMB is almost isotropic,[2] varying only be about one part in ten thousand. That is because the cosmic gas was homogeneous to the same accuracy when the CMB was emitted.

More information comes from the observation of galaxies. The stars we see in the sky are in our galaxy, but hundreds of thousands of other galaxies have been observed through telescopes. As shown in Fig. 1.2, their distribution averaged over a large volume is nearly uniform.

The distribution of galaxies can be analyzed statistically, and so can the anisotropy of the CMB.[3] In both cases, one finds that the statistical properties are the same in every region, and that they don't pick out any particular direction. We say that the Universe is statistically homogeneous and statistically isotropic.

We don't know for sure what happened before the beginning of the known history, but it was presumably preceded by some era when the Universe was still an

[2] 'Isotropic' means the same in all directions'.

[3] 'Anisotropy' is noun corresponding to the adjective 'anisotropic', which means 'not isotropic'.

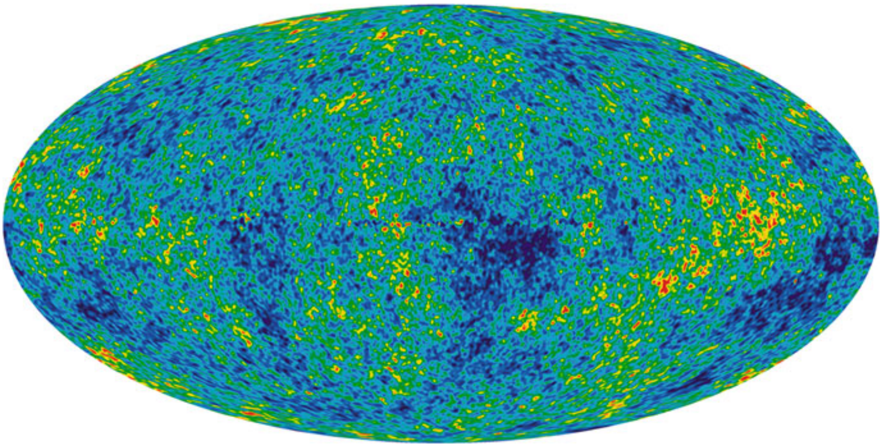


Fig. 1.1 The image is a map of the whole sky, and is analogous to the map of the whole Earth that one sometimes sees. The boundary of the map corresponds to a single direction in the sky, and its centre corresponds to the opposite direction. The map shows the brightness of the CMB, the *dark blue* regions being the brightest and the *red regions* being least bright. The variation in brightness is only about one part in 10,000 (Source: NASA [www.nasa.gov])

expanding gas of particles. The entire era when there was an expanding gas is called the **Big Bang**. The Big Bang was probably preceded by what is called **inflation**. Conditions at the end of a long era of inflation are more or less independent of what they were at the beginning. As a result, a suitably-chosen inflation scenario can provide a complete explanation of what we observe. This is both a blessing and a curse. It's a curse because it means that observation can't distinguish between different scenarios for what happened before inflation. It's a blessing, because it removes the need to think about such scenarios; for practical purposes, the history of the Universe begins with inflation!

Alternatives to inflation have been proposed, in which the very early Universe is contracting, the contraction only later giving way to expansion. There are also 'cyclic' proposals, in which that process is repeated. The alternatives are more complicated than inflation, and less popular.

In all of this, I've been writing 'Universe', but of course we don't really know about the whole Universe. We only know about a spherical region around us, called the **Observable Universe**. The boundary of this region is called the present horizon.[4] The distance to the present horizon is the biggest distance that any particle can have travelled, starting at the beginning of the Big Bang and ending at the present epoch. It really does mark the limit of the observable Universe, because all of our knowledge about the Universe is obtained by observing particles that were emitted during the Big Bang.

[4]One is also interested in the size of the region that could have been observed at an earlier epoch. Its boundary is called simply the horizon.

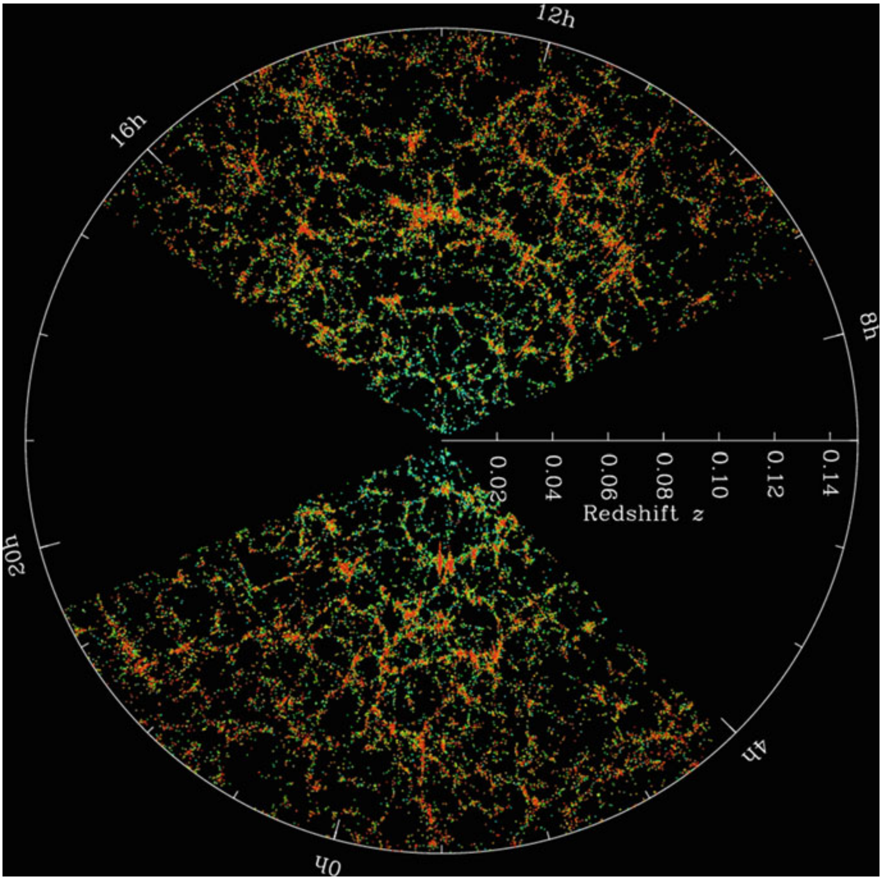


Fig. 1.2 A survey of the large galaxies, that are within two opposite wedges of the Universe around us. The survey finds more than 220,000 galaxies. The Earth is at the centre of the image, and the circle is about 2×10^9 light years away. Each *dot* is a large galaxy. The redshift of a galaxy is defined as its distance divided by what is called the Hubble distance, and the latter is approximately the distance to the present horizon (Image courtesy of M. Blanton and the Sloan Digital Sky Survey)

Nobody knows what's beyond the present horizon, and nobody ever will. There are ideas of course. One idea is that there are many regions, more or less like the one that we occupy, which constitute islands in a rather chaotic sea. That is one version of what is called the multiverse, which is suggested by some models of inflation.[5] A different idea is that the whole Universe has a finite volume, and is everywhere like the observable Universe. In that case, a straight line moving outwards from our

[5]In another version, the multiverse is the set of all possible Universes that might have existed (but actually do not) as a result of the quantum theory of inflation. In yet another, it is the set of all Universes that are being continuously created according to the 'many worlds' interpretation of quantum physics.

location would end up as a straight line moving towards from the opposite direction! In this book I'll stick to the observable Universe. To avoid cluttering the text though, I will usually just write 'Universe' instead of 'observable Universe'.

Chapter 2
Before We Begin

> Unthinking respect for authority is the greatest enemy of truth.
> (Albert Einstein in a letter to Jost Winteler, c. 1901)

Before we get to physics and cosmology, there are some things I need to deal with. I'll begin with the most important thing.

Why Should You Believe Me?

Why should you believe what you read? Why should anyone believe what a scientist is saying?

Here's an analogy that might be helpful. Suppose you've completed a crossword, and somebody suggests that your solution may be wrong even though it fits the clues. From that question, you know that you're dealing with someone whose never done a crossword. You will have to explain that although it's logically possible for your solution to be wrong, the fact that the words mesh with each other makes it very unlikely.

A piece of established science is a bit like the solution to a crossword. It starts with observations which are a bit like the clues. Then it makes sense of the observations through a theory which is a bit like the meshing of the answers. It's logically possible that the result is incorrect, but so unlikely that it's not worth thinking about.

Of course there is always a scientific frontier, where research is still in progress and we're not sure what is correct. That's as if you had only completed the top half of your crossword when someone suggested that the answers may be wrong. You would have to admit the that some of the answers near the bottom might be wrong because they wouldn't be as firmly embedded with each other as those higher up, which you can be sure about.

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_2

Something that the crossword analogy doesn't capture, is the issue of accuracy. A measurement can always be made more accurate, and a theory sometimes requires modification in the face of more accurate observations. As we'll see, Newton's theory of gravity was modified by Einstein, and Maxwell's theory of electromagnetism was modified when quantum physics came along. Currently accepted theories will no doubt require modification in the future. But the key word here is modification; the old theory will still be valid within some accuracy. Most importantly, the central facts are not going to change. Atoms definitely exist and the Earth definitely goes round the Sun.

New Words

There are new words in this book, as well as words whose meaning in physics is different from the usual one. There are also some abbreviations. New words and abbreviations are essential to produce something readable. Without them, the sentences would become too long to read. It would be like doing without the words 'cow', 'grazing' and 'field', to write 'some adult female cattle are biting off and chewing long green leaves, in an enclosed area of the countryside that contains no bare ground'.

The meaning of each word or abbreviation is explained the first time it appears, and is summarised in the Glossary. If it's going to be used a lot it appears in boldface **like this**.

As the book contains so many new ideas, you will need to take it quite slowly, with regular breaks to give yourself time to think. You can't digest all of the information in an afternoon, any more than you could digest several meals. You may also want to take a break to access a website, for further enlightenment on some particular topic.

About Mathematics

The body of the book is mostly descriptive, with hardly any mathematics. Appendix A amplifies what is said in the body of the book, and some of it uses mathematics.

Although the body of the book is free of mathematics, it does contain numbers. I need to explain three things about the numbers. The first is a piece of notation that allows us to write down very small and very big numbers, without using a lot of zeros. You may know already that 2^3, say, means $2 \times 2 \times 2$ and that 2^{-3}, means $1/2^3$. Using this notation we write say 1.5×10^{15}, instead of 1,500,000,000,000,000, and 1.5×10^{-15} instead 0.000,000,000,000,001,5. This is much better than exhibiting the zeros and leaving the reader to count them.

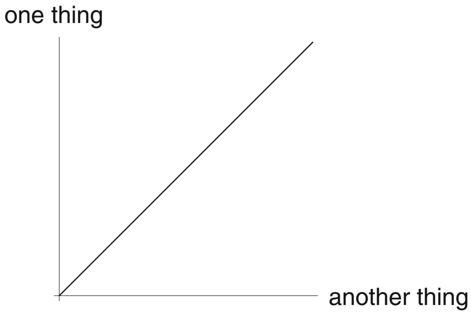


Fig. 2.1 One things is proportional to another

The second thing has to do with accuracy. I have stated, for instance, that the Universe is 13.8 billion years old but that's not exact. The age of the Universe is almost certainly between 13.0 and 14.6 billion years, but 13.8 billion is just a best guess. Given that situation, there would be no point in giving more than two figures. It would be nonsense to write that the age is 13.872 million years. On the other hand, I'll be writing later that the rest energy of the electron is 0.511 MeV. (Don't worry for now what 'rest energy' and 'MeV' are.) In fact, we know that it's 0.51099891 MeV with pretty high accuracy, but to write that would have cluttered the page to no purpose. I'll give the numerical values of things without comment in the rest of the book. In some cases the number I give is a bit uncertain. In other cases, I could have given it much more accurately.

The third thing is the meaning of the word 'magnitude'. If a quantity can be either positive or negative, its magnitude is its numerical value. So 8 and −8 both have magnitude 8.

Throughout the book, I will use the phrase '**proportional to**'. When we say that one thing is proportional to another, we mean that the two things depend on each other in such a way that doubling one means doubling the other, tripling one means tripling the other, halving one means halving the other and so on. As shown in Fig. 2.1, this means that a graph of one thing against the other is a straight line passing through the point where both quantities are zero. 'One thing' might be the distance travelled by something moving in a straight line with constant speed with 'another thing' being the time taken to travel that distance.

Chapter 3
The Fundamental Theories

> Nature and nature's laws lay hid in night:
> God said, Let Newton be! and all was light.
> (Epitaph by Alexander Pope on the death of Isaac Newton in 1727.)

In this chapter I give a birds-eye view of twenty-first century physics. Then I give a (very) brief history of physics.

The Twin Pillars of Physics

At the present time, almost all of physics rests on two fundamental theories. These are **General Relativity** which describes gravity, and the **Standard Model** which describes almost everything else.

General Relativity was worked out by Einstein and published in 1916. The Standard Model was the work of several people, and was completed in the 1970s after a series of dramatic breakthroughs. As the name suggests, the Standard Model was initially intended as a working hypothesis and its continuing success has been something of a surprise.

The Standard Model deals with what physicists call **particles**. As is the case with several other words the physics meaning of 'particle' is a bit different from the usual one. It does mean something very small, as in ordinary English, but there is more. The 'particles' of physics come in different varieties, called **species**, and all examples of a given species are identical.

The first species to be discovered, by J. J. Thomson in 1897, was the **electron**. An electric current within say a wire is simply a flow of electrons. Many other species have been discovered since, all described by the Standard Model. They are of two kinds; elementary particles that are not made of other particles, and composite particles that are made out of elementary particles.

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_3

Most particle species are usually denoted by a letter (Roman or Greek) rather than by a name. The elementary particles of the Standard Model are listed in Appendix B. They can be divided into three categories, called quarks, leptons and gauge bosons, plus an odd man out called the Higgs boson.[1] The composite particles of the Standard Model are called hadrons and are made out of quarks.

The Standard Model tells us what happens when particles **collide**. It also tells us about the process called **decay**, whereby a particle is replaced by two or more other particles. These two processes are collectively called the **interactions** of the particles.

Besides particles, the Standard Model deals with what physicists call **fields**. A field is not localised in space, but rather is spread out over a whole region or even the entire Universe—the very opposite of a particle. A familiar example is the magnetic field that is detected by our compasses.

Although the Standard Model accounts for almost all observations it's not perfect. It doesn't account for the mass of the three elementary particles, known collectively as **neutrinos**. Also, it doesn't include any particle species that could be the Cold Dark Matter. For these and other reasons, various extensions of the Standard Model have been proposed. The extensions use the same general framework as the Standard Model, called quantum field theory. The only difference is that there are additional particle species, which have specified interactions with each other and with the Standard Model particles.

General Relativity and quantum field theory work well in practice, and they can be combined to some extent. But, as will be seen in Chap. 16, one can envisage extreme situations in which they fail to work. In those situations one needs a better theory, which gives General Relativity and quantum field theory as approximations. The favoured candidate for such a theory is called string theory. The word 'string' refers to the fundamental objects of the theory, which are one-dimensional. According to string theory, each particle is a very small loop of string. The possibility of observing the strings themselves is mentioned in Chap. 16.

The History of Physics

Now I'll describe some of the most important steps, by which our present knowledge was achieved. At every stage, observation was crucial in pointing the way forward but to keep things simple I will focus on the theory. Along the way, I will introduce three numbers called fundamental constants,[2] denoted by the symbols G, c and h.

Physics, as we now know it, might be said to have started with the publication in 1687 of Isaac Newton's 'Mathematical Principles of Natural Philosophy'. In it, Newton dealt with what is called mechanics; the description of the motion of objects

[1]Boson is pronounced boze-on after the Indian physicist Satyendra Nath Bose.

[2]In physics, a constant is a quantity which doesn't change with time, i.e. is constant.

under the influence of the forces acting on them. As shown in Fig. 3.1, Newton's theory was the starting point for both General Relativity and the Standard Model.

It was the starting point for General Relativity because it gave the first theory of gravity. Indeed, it was Newton who coined the name 'gravity'. Newton's theory of gravity involves G, which specifies the strength of gravity. That theory was accepted as correct, right up until the publication in 1916 of Einstein's General Relativity. Einstein's theory involves, in addition to G, the **speed of light** which is denoted by c. Newton's theory of gravity is valid only as an approximation, for objects moving with speed much less than c. General Relativity works perfectly in all situations, within the accuracy of present observations.

So much for gravity. The road from Newton's mechanics to the Standard Model was far less direct. At every stage, gravity was excluded from the discussion. The first step was the formulation of what is called is called classical electrodynamics, that was brought into its final form by James Clerk Maxwell in the 1860s. Classical electrodynamics deals with magnetic and electric fields, and with **electromagnetic radiation** of which light is an example. Electromagnetic radiation consists of rapidly varying electric and magnetic fields, and it always travels through the vacuum with speed c. Classical electromagnetism also deals with charged objects, that generate electric and magnetic fields and are affected by them.

The next development came with the publication of Einstein's theory of **Special Relativity** in 1905. Special Relativity is a description of space and time that is quite different from the commonsense description, that we have on the basis of everyday experience with speeds much less than c. A parallel development was begun by Planck in 1900, who invoked the third fundamental constant h. It is called **Planck's constant**. Effects that depend on h are called quantum effects, and when one is considering them one is said to be doing **quantum physics**. In any situation where quantum effects are small enough to be ignored, we recover what is called classical physics. Roughly speaking, quantum physics is needed only for the description of very small objects like particles and atoms.

Although Planck introduced h in 1900, the subsequent development of quantum physics was quite slow. A significant step was taken in the 1920s with the formulation of quantum mechanics. Quantum mechanics is based on Newtonian mechanics, and gives a good description of atoms and molecules. It doesn't involve any fundamental constant except h. Then, in the 1930s and 1940s, Quantum Electrodynamics was developed, which involves h and c because it uses both Classical Electrodyamics and Special Relativity.

According to Quantum Electrodynamics, a beam of light can be regarded as a beam of particles, namely **photons**. In its original form, Quantum Electrodynamics worked with just the photon and the electron. Many particle species have been found since. Most of them have been found building what are called accelerators. An accelerator causes particles to collide violently with each other, or with atomic nuclei. This may be done by accelerating a single beam of particles, which collide against a stationary target. It can also be done by accelerating two beams, that collide with each other. Discoveries at accelerators led to the formulation of the Standard Model in the 1970s, and the construction of accelerators has continued, to produce beams of ever higher energy that can create ever more massive particles, and probe the nature of the interactions at ever shorter distances.

The LHC at the CERN laboratory in Geneva is at present the accelerator with the highest energy. It was designed to search for evidence of some extension of the Standard Model. None has been found at the time of writing, the results being completely consistent with the Standard Model. The cost of the accelerators has increased with their energy, and there is at present no firm plan to build a machine with more energy than the LHC.

> To specify the fundamental constants we need to choose units for length, mass and time. The standard choice for science are the meter, kilogram and second, defining what are called **MKS units**. Using these units, $c = 3.00 \times 10^5$ km seconds^{-1}, $G = 6.67 \times 10^{-11}$ metres^3 kg^{-1} seconds^{-2} and $h = 6.63 \times 10^{-34}$ metres^2 kg seconds^{-1}. The units are inserted to show what happens if they are changed; multiplying a unit by some factor changes the value by the indicated power of that factor. In the case of c, displaying the units has a direct physical interpretation, which is that light travels at $c = 3.00 \times 10^8$ kilometres per second.

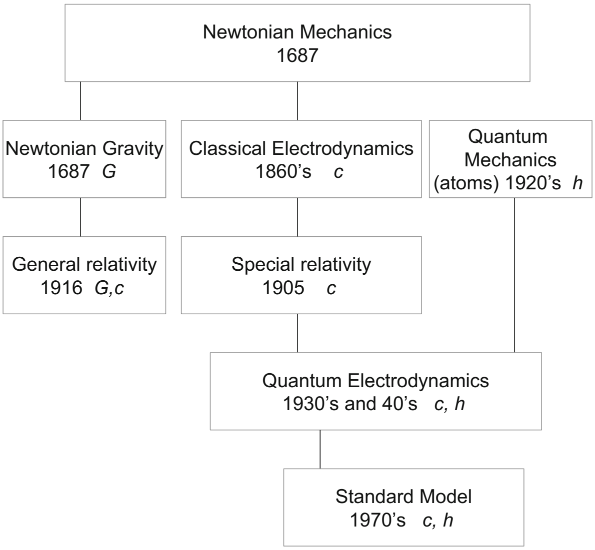


Fig. 3.1 The history of physics

Standard Model Parameters In addition to the fundamental constants, a theory usually involves more numbers. They are called **parameters** of the theory, and their values are determined by requiring that the theory agrees with observation. The Standard Model involves 18 parameters; 12 of these can be taken to be the masses of the elementary particles (excluding the neutrinos), and the other 6 describe the interactions of the particles. To fully describe the neutrinos, we need 7 more parameters; 3 of these can be taken to be the masses of the neutrinos, and the other 4 are needed to account for a phenomenon called neutrino oscillation. The masses are given in Table A.4 of Appendix B.

Chapter 4
Atoms and Electromagnetism

Matter, though divisible in an extreme degree, is nevertheless not infinitely divisible. . . . I have chosen the word atom to signify these ultimate particles.
(John Dalton in notes for a Royal Institution Lecture, 1810)

This chapter deals with atoms and electromagnetism. You may well be familiar with those topics, in which case please skip the chapter.

Atoms

The Earth and everything on it consists of atoms. An atom is made of just three particle species, namely the proton, the neutron and the electron. The proton and neutron are made out of quarks. As shown in Fig. 4.1, they are in a tightly packed ball at the centre of the atom which is surrounded by the electrons. The ball is called the **nucleus** of the atom.

The properties of an atom are almost entirely determined by the number of protons in its nucleus. The simplest atom is hydrogen with one proton. Hydrogen is denoted by the letter H. Next is helium with two protons, denoted by He. Then we have lithium (Li) with three protons, beryllium (Be) with four protons and so on. These are collectively called elements. The elements are listed in Fig. 4.2, arranged in what is called the Periodic Table.

Except for the simplest version of hydrogen, every nucleus contains some neutrons. The number of neutrons per proton cannot be too big or too small, or the nucleus would fall apart, but it is not completely fixed. When we want to distinguish between atoms with different numbers of neutrons, we attach a number to its letter to specify the total number of particles in the nucleus. Hydrogen comes in three versions, ^{1}H, ^{2}H and ^{3}H. Helium comes in two versions, ^{3}He and ^{4}He, and so on. The versions of an atom are called isotopes.

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_4

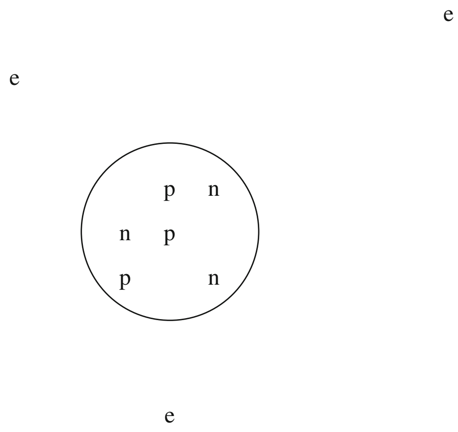


Fig. 4.1 Picture of a typical atom, namely ^{6}Li which consists of a nucleus containing three protons and three neutrons, and a cloud of electrons containing as many electrons as there are protons. The picture is not to scale because the size of a nucleus is only about 10^{-15} metres, whereas the size of an atom (i.e. of the electron cloud) is about 10^{-10} metres

Like many particles, protons and electrons have what is called **electric charge**, or for short just charge. This is a number that can be either positive or negative. Particles with charge of opposite sign are pulled together by what is called the electric force; we say that the force is attractive. Particles with charges of the same sign are pushed apart and we say that the force is repulsive.

I will specify charges as multiples of the proton charge. With that convention the proton has charge 1 and the electron has charge -1. The neutron has zero electric charge. An object with zero electric charge is said to be **neutral**, otherwise it is said to be **charged**.

The electrons in an atom are moving, but they don't fly away because the electric force pulls them toward the positively charged nucleus. You might wonder why the protons in the nucleus are not pushed apart by the electric force between them. The answer is that there is another force, called the strong force. The strong force doesn't distinguish between protons and neutrons, and pulls them together so as to overcome the electric repulsion between the protons.

An atom can exist on its own, or be part of a molecule which is a bundle of atoms. For instance, a water molecule contains two hydrogen atoms and one oxygen atom. Molecules usually contain only a few atoms, but molecules containing carbon atoms (called organic molecules) can contain an enormous number. The biggest that can be made in the laboratory contain around 10^8 atoms. Beyond that we come to molecules that form the basis of life such as DNA which typically contains around 10^{11} atoms.

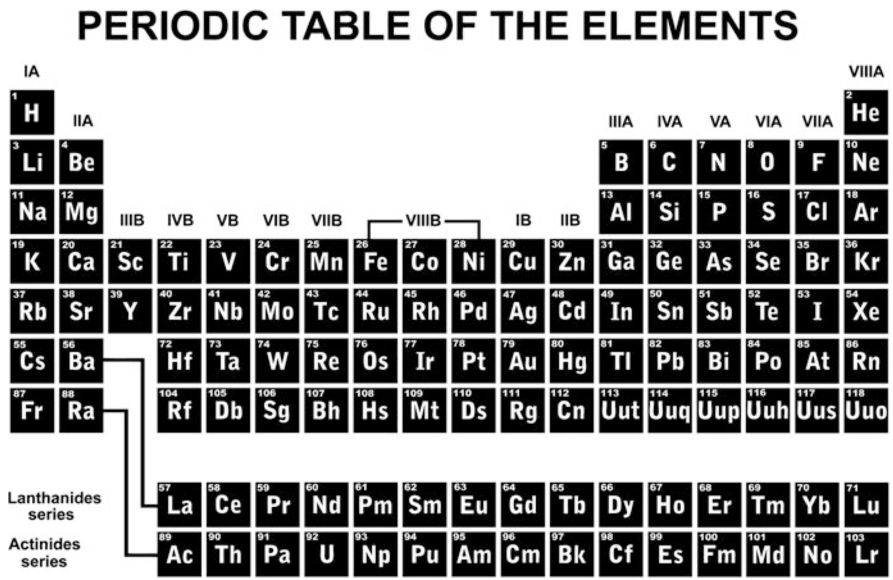


Fig. 4.2 In each box, the letters give the symbol which is often used instead of the atom's name, and the number gives the number of protons in the nucleus

Neutrality

A normal atom is neutral, being composed of equal numbers of protons and electrons. There are also atoms in which the number of protons is different from the number of electrons. The most usual situation is to have, within some region, a number of positively charged atoms that have lost one electron, with the lost electrons moving freely between them. In that case the region as a whole is neutral.

It can happen that a region has more protons than electrons or vice versa, but that is usually a temporary situation that is ended when electrons move to or from the region to make it electrically neutral again. Static electricity is an example. Combing your hair can move electrons from your hair to the comb, making both of them charged. Your hairs, if dry, will be pushed apart by the electric force. As time goes on though, your hair will pick up electrons from its surroundings and the comb will lose electrons, so that they both recover their neutrality.

Astronomical objects like planets and stars are neutral to very high accuracy. As a result, the only significant force between them is the force of gravity.

Electric and Magnetic Fields

The magnetic field exists at some level throughout the Universe. To specify the magnetic field, we have to give two pieces of information; we have to specify the direction of the field, and its strength. On the Earth, the magnetic field is detected by compasses. The compass needle points along the direction of the magnetic field, and it is pulled more or less strongly according to the strength of the field. The Earth's magnetic field is illustrated in Fig. 4.3. It is generated by an electric current flowing within the Earth's iron core.

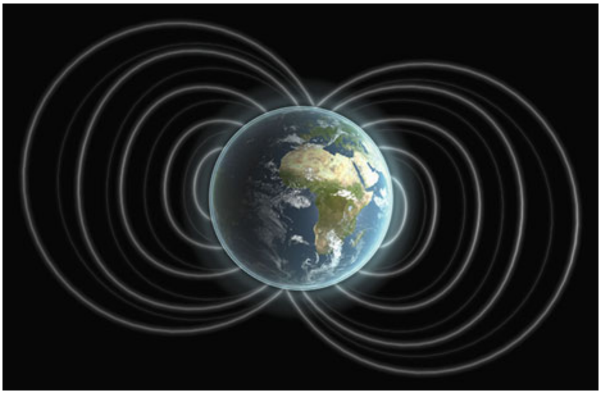


Fig. 4.3 This shows the Earth's magnetic field. The *lines* indicate the direction of the field. We can regard the space around the Earth as being filled by such lines (Source: NASA [www.nasa.gov])

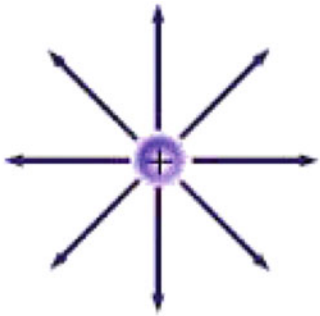


Fig. 4.4 This shows the electric field generated by a stationary positive charge. The *lines* indicate the direction of the field. The strength of the field is proportional to $1/r^2$, where r is the distance from the charge. The field generated by a negative charge is the same, except that it points inwards

The electric field also has both a strength and a direction. It exists near any charged object, as shown in Fig. 4.4. The electric field pulls a positively charged object in its direction, and it pulls a negatively charged object in the opposite direction.

The electric and magnetic fields are collectively called the electromagnetic field. The electromagnetic field exerts a force on every charged object. The electric force between two stationary charged objects occurs because each object generates an electric field, which is felt by the other object.

Electromagnetic Radiation

Now we come to electromagnetic radiation. In the simplest kind of electromagnetic radiation, the strengths of the electric and magnetic fields are oscillating as shown in Fig. 4.5. This is called an electromagnetic wave and it travels with the speed of light. As we move along the direction of the wave, the strengths of the fields vary as in Fig. 4.6. The most general kind of electromagnetic radiation consists of electromagnetic waves travelling in different directions, and oscillating with different frequencies.

Each colour of the rainbow corresponds to electromagnetic waves of a definite frequency. Violet has the highest frequency and red the lowest. Going down in frequency from red, we have electromagnetic radiation that is not visible to the

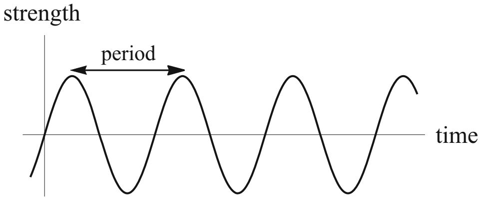


Fig. 4.5 This shows how the strengths of the electric and magnetic in an electromagnetic wave vary with time at a fixed position. The period is the time taken for one oscillations, measured usually in seconds. The frequency is 1 divided by the period, or in other words the number of oscillations per second

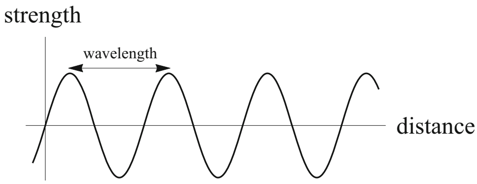


Fig. 4.6 This shows how the strengths of the electric and magnetic in an electromagnetic wave vary as one goes along the wave at a fixed time. The wavelength is the distance between points of maximum strength. An electromagnetic wave is like an ocean wave, with the strength of the field corresponding to the height of the water. When we stand in the water as the waves pass by, the waver level changes as in Fig. 4.5. If we look across the water, we see the pattern in Fig. 4.6

eye; first infra-red radiation, then microwaves and finally radio waves. Going up from violet we have ultra-violet radiation, X-rays and finally gamma rays. This is illustrated in Fig. 4.7.

According to Quantum Electrodynamics, electromagnetic radiation consists of photons. If the radiation has a certain frequency, the energy of each photon is equal to the frequency times Planck's constant h. Figure 4.7 shows the relation between energy and frequency.

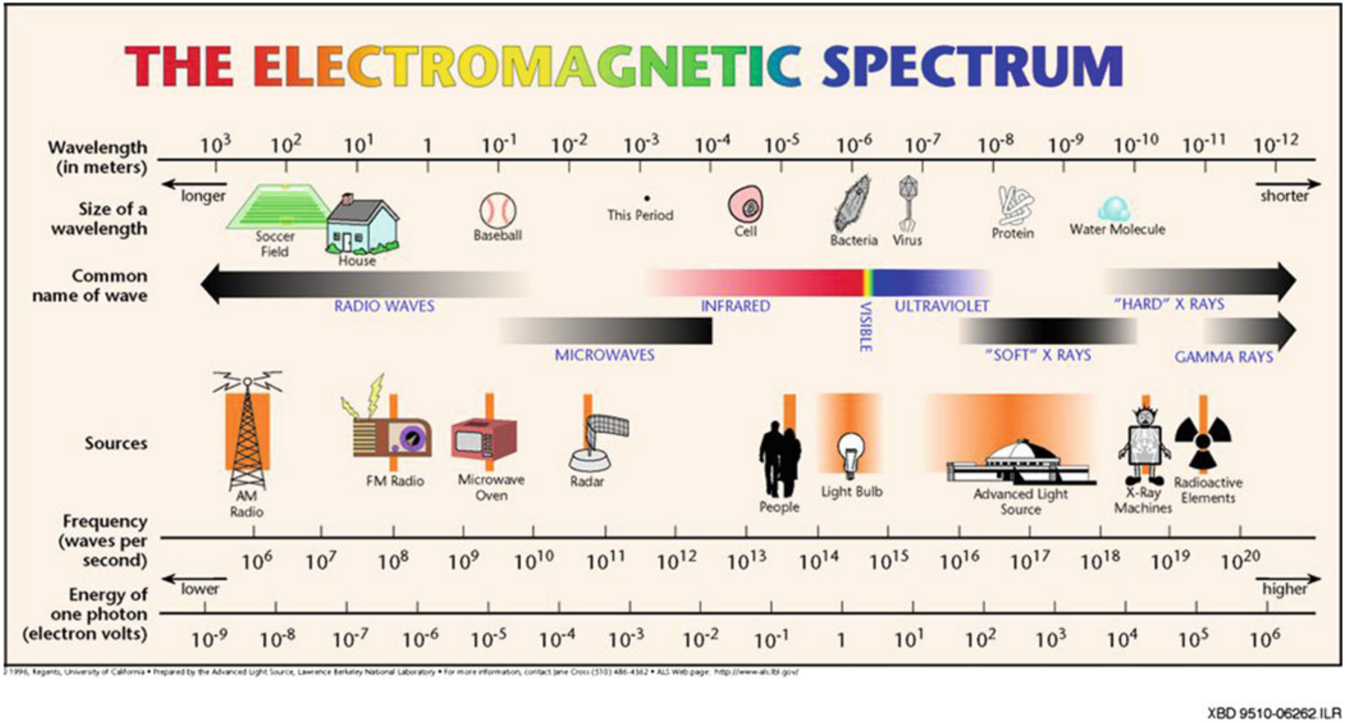


Fig. 4.7 Courtesy of the Advanced Light Source, Lawrence Berkeley National Laboratory. In the bottom line, the energy unit is the eV. I am using the MeV, equal to 10^6 eV

Magnets

A magnet is a piece of metal that generates a magnetic field. To understand a magnet we have to go to the level of individual atoms. Around each atom there is a magnetic field, which is usually a miniature version of the field generated by the Earth. In a normal material, the atoms are randomly orientated so that their magnetic fields cancel. In a magnetic though, the atoms are at least partially aligned so that they generate a net magnetic field. Magnets are usually made of iron. The magnetic field produced by a magnet has the same form as the Earth's magnetic field. If a freely moving magnet is placed in a magnetic field, its 'north pole' points in the direction of the field. A compass needle is such a magnet, and it points in the direction of the Earth's magnetic field.

Chapter 5
The Present Universe

There are countless Suns and countless Earths . . .
(Giordan Bruno in 'De L'Infinito Universo E Mondi (1584).

In this chapter I describe the objects, that exist in the Universe at the present time. Then I explain how we know about these objects, by observing the particles that they emit.[1] Finally I describe the 'cosmic backgrounds', consisting of particles that were emitted in the early Universe before there were any individual objects.

The Objects

Let's begin with the Sun. It's a dense ball of hot gas, containing atomic nuclei and freely moving electrons. The nuclei are undergoing nuclear fusion, which generates very energetic photons and neutrinos. The neutrinos escape promptly, but the photons bounce off the electrons. It takes about a million years for a photon to escape from the centre of the Sun.

The Universe is populated with **stars** more or less like the Sun, that are powered by nuclear fusion. Some stars are known to have planets circling round them, and there are presumably planets that resemble the Earth in their temperature and the mix of land, water and atmosphere. According to recent research, it seems likely that life developed on such planets.

When a star has used up all of its nuclear fuel, it collapses. If the star is not too heavy, the collapse is halted by a quantum effect called electron degeneracy. We then have a white dwarf consisting of highly compressed ordinary matter.

The mass of a white dwarf cannot be more than 1.44 times the mass of the Sun, because electron degeneracy cannot resist the pull of gravity for a heavier

[1]This includes electromagnetic radiation, which consists of photons.

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_5

star. A heavier star will become so highly compressed that each proton combines with an electron, to form a neutron and a neutrino. The neutrinos fly away but the neutrons remain, and for all except the very heaviest stars, the collapse is finally halted by what is called neutron degeneracy. We then have a neutron star consisting of neutrons. The radius of a neutron star is only around 10 km, and it is so dense that a teaspoon of a neutron star is about a thousand times as heavy as the Great Pyramid of Giza. Some neutron stars are emitting a strong beam of electromagnetic radiation which rotates like a lighthouse beam so that we see it as a series of pulses. Those neutron stars are called pulsars.

A neutron star cannot be more than about three times heavier than the Sun, because even neutron degeneracy cannot resist the pull of gravity for a heavier star. A heavier star will keep on collapsing, to form a black hole. The gravity of a black hole is so strong that nothing emerges from it, not even light.

The violent collapse leading to the formation of a neutron star or black hole is called a supernova. A supernova first emits an intense burst of neutrinos, lasting for less than a minute. Then there is an intense burst of photons lasting a few days. If a supernova occurs in our galaxy, and is not obscured by dust, it can outshine all other stars and be visible during the day. Those naked-eye supernovas are very rare and only a few have been described by past generations, the last one in 1604.

Supernovas are thought to account for what are called gamma ray bursts, which are observed at the rate of around one a day. The burst of gamma rays is apparently released as two oppositely directed jets, just after a star collapses to form a black hole. The amount of energy released is huge, around one percent of the star's rest mass.

Supernovas are the most violent events in the Universe. They are vital to our existence, and even to the existence of the Earth itself. That is because they are responsible for the creation of practically all nuclei and therefore, ultimately, the creation of practically all atoms. Without supernovas, the present Universe would contain only the hydrogen and helium, plus tiny amounts of the lithium and beryllium. (We will see how *those* originate in Chap. 7.)

Many stars are in pairs that orbit around each other. These are called binaries. If a binary contains a white dwarf and an ordinary star, matter can flow from the star to the white dwarf. If the mass of the white dwarf becomes bigger than $1.44 M_\odot$, it collapses to become a neutron star. The collapse is a supernova, just like the collapse of an isolated star.

The stars are gathered into **galaxies**. We are near the edge of a large galaxy, called the Milky Way galaxy. Looking at the sky with the naked eye, one sees stars in the Milky Way galaxy, and the Milky Way itself which consists of stars in our galaxy that can't be seen individually with the naked eye.[2] Most galaxies are gathered into

[2]The stars in the milky way were first seen by Galileo in 1610 using his telescope. Only three galaxies are visible to the naked eye; these are the Andromeda Nebula seen in the Northern Hemisphere, and the Small and Large Magellanic Clouds visible in the Souther Hemisphere.

galaxy clusters. This is illustrated schematically in Fig. 5.1, and a real image is shown in Fig. 1.2.

At least the larger galaxies have a black hole at their centre. The black hole accounts for only a tiny fraction of the galaxy's mass, but it is still far heavier than the black holes formed by the collapse of a star. Sometimes, a large amount of matter is falling violently into the black hole generating intense electromagnetic radiation. We then have what is called an active galactic nucleus.

Each star is hot ball of gas, consisting of the nuclei and electrons that make up ordinary matter. There is also a cooler gas, that pervades the whole Universe and contains both ordinary matter and Cold Dark Matter (CDM). As shown in Fig. 5.1, this gas becomes more dense as we go inside a galaxy cluster, and denser still as we go inside an individual galaxy.

The gas between the clusters is almost homogeneous and I will call it the **cosmic gas**. The distribution of the galaxy clusters is also almost homogeneous; in other words, the number of galaxy clusters within a region of given volume that contains many clusters, is almost independent of the location of the region. We can summarise this situation by saying that *the present Universe is almost homogeneous*.

Fig. 5.1 Stars are made of ordinary matter and are gathered into galaxies, the galaxies in turn being gathered into galaxy clusters. A galaxy contains 10^5 to 10^{12} stars, and a galaxy cluster can contain anything up to several thousand galaxies. There is also a gas that pervades the Universe, which consists of Cold Dark Matter (CDM) as well as ordinary matter. In the Figure the *dots* represent stars, the *small disks* represent galaxies and the *big disk* represents a single galaxy cluster. The gas becomes denser when we go inside a galaxy cluster, and denser still when we go inside a galaxy. The gas between the clusters is practically homogeneous and I am calling it the cosmic gas

Observing the Universe

Electromagnetic radiation is nowadays observed over the whole range of frequencies, from radio waves to very high-energy gamma rays. The ‘telescopes’ are of many different kinds, and are mounted in balloons and satellites as well as on the Earth. The radiation comes from all of the objects mentioned, including supernovas in distant galaxies.

Neutrinos are also observed. This is difficult, because a typical neutrino passes right through the Earth without interacting. A neutrino detector consists of a large tank of liquid, around which are devices for detecting the photons that are emitted on the rare occasions when a neutrino interacts with an atom in the tank. To shield it from cosmic rays, a neutrino detector has to be placed underground, or under a deep body of water.

Neutrinos created in accelerators are detected in this way. Neutrinos from the Sun are also detected, and in 1987 neutrinos from a nearby supernova were detected. Starting in 2013, neutrinos from sources outside the solar system are being observed routinely, at the IceCube neutrino detector under the Antarctic ice. This marks the beginning of neutrino astronomy.

Some objects in the Universe emit, in addition to photons and neutrinos, particles and nuclei that arrive at the Earth constituting what are called primary cosmic rays.[3] It’s not known for sure what those objects are, but likely candidates include supernovas, quasars and active galactic nuclei. The same objects are presumably responsible for the very high energy gamma rays, that like the cosmic rays come to us from all directions rather than from identifiable sources. Primary cosmic rays consist of electrons and nuclei (mostly hydrogen and helium) along with a tiny fraction of positrons and anti-protons. When they hit the atmosphere the primary cosmic rays produce more particles, called secondary cosmic rays. Cosmic rays are detected by a variety of methods.

That’s all for the moment, but within the next few years there will be another window on the Universe; we will detect gravitational radiation. Gravitational radiation is similar to electromagnetic radiation, except that the quantity which varies is the acceleration of gravity instead of the electromagnetic field. According to General Relativity it is generated in principle by any moving object, that is not travelling in a straight line with constant speed. So, if you move your arm as you come towards me, you emit a pulse of gravitational radiation which pulls me from side to side when it arrives. The effect though, is far too weak to be detectable. For gravitational radiation to be detectable, it must come from places in the Universe where huge amounts of matter are undergoing huge acceleration. The generation and detection of gravitational radiation is summarised in Fig. 5.2 and one of the boxes.

[3]The particles and nuclei interact with the cosmic gas on their journey towards us. The products of such interactions, arriving at the Earth, are included under the heading ‘primary cosmic rays’.

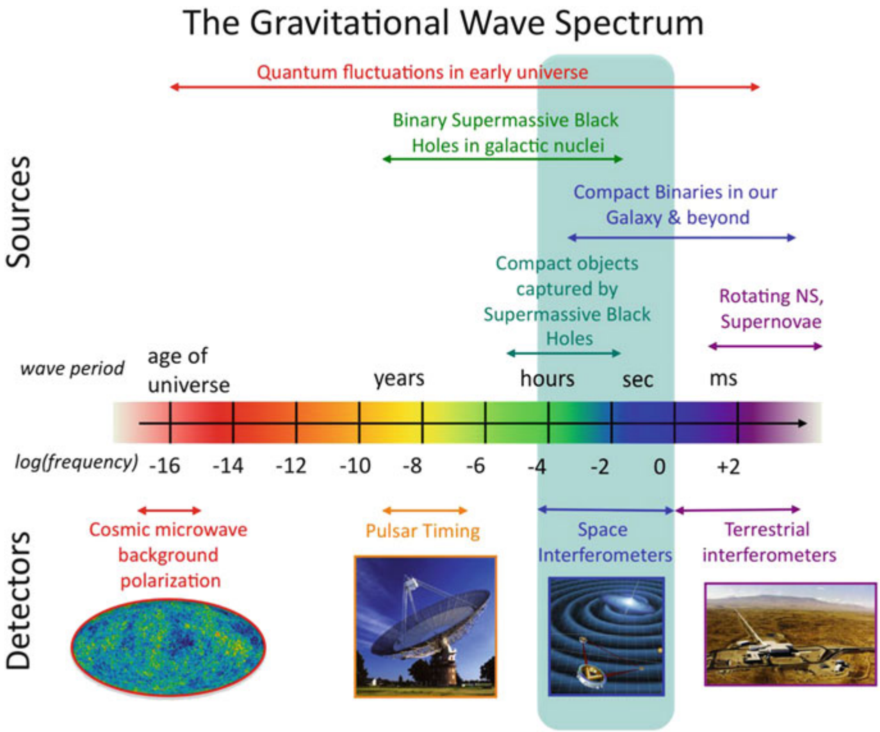


Fig. 5.2 Gravitational radiation (Source: NASA [www.nasa.gov])

The Cosmic Backgrounds

The photons and neutrinos that I've mentioned so far, were emitted quite recently, by stars and the gas between them. There are also photons and neutrinos that were created in the early Universe, before there were any stars or galaxies. The photons constitute the Cosmic Microwave Background (CMB) that was mentioned in Chap. 1. The neutrinos constitute what I'll call the **Cosmic Neutrino Background**. The CMB and the Cosmic Neutrino Background are practically homogeneous, and I'll take the term 'cosmic gas' to include them.

The CMB was first observed in 1964.[4] The slight variation in its brightness with direction (CMB anisotropy) was first observed in 1992. The Cosmic Neutrino Background has not been observed directly, but its existence can be deduced by comparing the CMB anisotropy with the one that would exist if it were absent.

[4]To be precise that is when it was first recognised as such. An earlier observation in 1957 was not recognised at the time.

Table 5.1 Cosmic sizes and masses The masses are in units of the Sun's mass, which is 1.99×10^{30} kg. The 'size' given for the Sun is its radius and the other sizes can be taken as radii as well. (Galaxies and galaxy clusters are not really spherical but that doesn't matter because the sizes given are only approximate.) The range of star sizes excludes neutron stars, whose radius is only about 10 km. A lightyear is the distance travelled by light in a year which is 9.46×10^{12} km

	Size	Mass
Sun	7×10^5 km	1
Star	10^4 km to 10^8 km	10^{-1} to 10^2
Stellar black hole	Roughly 10 km	Roughly 10
Galactic black hole	Roughly 10^6 km	Roughly 10^6
Galaxy	10^3 to 10^6 lightyears	10^4 to 10^{13}
Galaxy cluster	10^6 to 10^7 lightyears	10^{12} to 10^{16}

These CMB and the Cosmic Neutrino Background are generated during the known history of the Universe and certainly exist. If the Big Bang is preceded by an era of inflation, that will generate another background, consisting of gravitational radiation. If this is sufficiently intense, it will in the future be detected through its effect on the CMB anisotropy. I will have much more to say about these backgrounds in later chapters (Table 5.1).

Generating gravitational radiation

The most powerful sources of gravitational radiation are pairs of very nearby objects that are rotating around each other. Each object must be either a black hole or a neutron star, that being the only way in which they can be close enough to be a powerful source. The emission of the gravitational waves causes the distance between the pair to decrease so that their orbit spirals inwards. This has been observed in cases where one of the objects is a pulsar, and the rate of spiralling is in accordance with General Relativity. Eventually the pair will come together, leading to a final intense burst of gravitational radiation called a chirp.

Other sources of potential observable gravitational radiation include (i) a white dwarf, neutron star or stellar black hole which is orbiting around a galactic black hole, (ii) a pair of galactic black holes orbiting around each other, (iii) a rotating neutron star and (iv) a supernova. Gravitational radiation from a pair of galactic black holes will be detected through its distortion of the otherwise regular pulses of the electromagnetic radiation observed from a pulsar. Gravitational radiation from a variety of sources will be detected by devices called laser interferometers, either in space or on the Earth.

Chapter 6
Collisions and Decays

> The unleashed power of the atom has changed everything save our modes of thinking and we thus drift toward unparalleled catastrophe. [translated from the German]
> (Albert Einstein, telegram sent to prominent Americans, 24 May 1946)

The collision of two particles, and the decay of a single particle, are important in many different situations. So too are collisions and decays that involve atoms and nuclei. The aim of this chapter is to describe some collisions and decays that are important in the early Universe. Collisions and decays are described by the Standard Model, which works within the framework of Einstein's Special Relativity and ignores gravity.

Energy of an Object

Before coming to collisions and decays, I need to discuss energy. Particles have energy and so do fields. One of energy's most important properties is that it cannot be created or destroyed. Particles can exchange energy with each other and with fields, but the total amount of energy remains the same. A quantity with that property is said in physics to be **conserved**, so in physics parlance 'energy is conserved'.

Please be clear though, that the physics usage of the phrase 'energy conservation' has nothing to do with the ordinary English use of that phrase. In ordinary English, 'energy conservation' means ensuring that our houses, cars and so on don't use too much energy. In physics terminology, energy is conserved no matter how much energy is used by a house or a car, because that energy will come from somewhere else.

According to Special Relativity, the energy of a stationary object is equal to its mass, times c^2. Written as an equation, this is $E = mc^2$, where E is the energy and

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_6

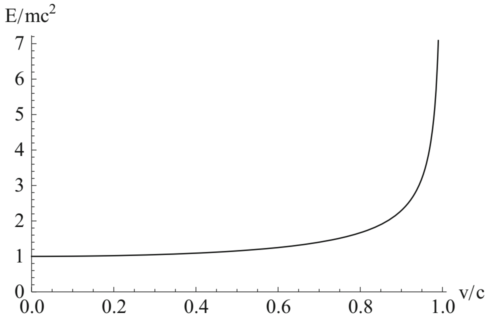


Fig. 6.1 The energy of a moving object. If the object moves with speed v, its energy is its rest energy divided by $\sqrt{1 - v^2/c^2}$. It increases without limit as the speed approaches c

m is the mass.[1] Surely the world's most famous equation! The energy of a stationary object is called its **rest energy** because in physics a stationary object is said to be at rest. The object might be a particle, or something of ordinary size like a rock, or something of astronomical size like a planet or a galaxy.

If the object is moving, its energy is bigger as shown in Fig. 6.1. The difference between the energy of a moving object and its rest energy is called the **kinetic energy** of the object. As the speed of the object approaches c, its energy increases with limit. When we increase the speed of an object, we give it energy which has to come from somewhere. One can never accelerate an object to speed c, because only a finite amount of energy is available. Photons are able to travel with speed c, only because that is *always* their speed. A particle with that property is said to have zero mass.

Contrary to what one might think, the mass of an atom or nucleus is not equal to the sum of the masses of its constituents. Equivalently, its rest energy is not the sum of the constituents' rest energies. It is smaller, by an amount called the **binding energy**. The binding energy is the amount of energy that is needed to break up the atom or nucleus into its constituents.[2]

Collision Processes

Let us now see what may happen when two particles collide.[3] One possibility is for nothing to happen; each particle continues as if the other were not there. Another

[1]If we were dealing with numbers we would write $E = m \times c^2$, but multiplication signs are not shown when the numbers are represented by symbols.

[2]I am here regarding the constituents of an atom is its electrons and its nucleus, not as its electrons and its individual protons and neutrons.

[3]By 'collision' I actually mean just a close encounter, because particles can affect each other without actual contact. Actual contact would make no sense anyway for elementary particles because they can be regarded as occupying points in space.

possibility is for the particles to simply bounce off each other. The most interesting possibility, though, is for one or more particles to be created or destroyed. In that case we have what is called a **collision process**. The Standard Model tells us the probability for each of these outcomes.

At the beginning of the known history, three collision processes are occurring. They are listed in Table 6.1. For each process, the reverse is also possible.[4] In the first process, two photons collide, to create an electron and a positron. The positron has the same mass as the electron, but the opposite charge.[5] Because of energy conservation, this electron-positron annihilation process can occur only if the energy of the photons is at least twice the rest energy of an electron.

In the second process, an electron-positron pair is created when a neutrino collides with an anti-neutrino. As seen in Table A.4, there are actually three neutrino species, all with very small masses. Neutrino physics involves various complications, that are mostly irrelevant to cosmology. To keep things simple, I will usually pretend that there is only one species of neutrino, with zero mass. The probability for *any* collision process involving a neutrino is very low, and a typical neutrino can pass right through the Earth without undergoing any such interaction. The second process occurs only in the very early Universe, when the number of neutrinos per unit volume is extraordinarily high. This is true also for the third process, in which an electron and proton annihilate to create a neutrino and a neutron. When a neutrino is created, its kinetic energy is always far bigger than its rest energy because the latter is so small. As a result, neutrinos always travel with speeds very close to c.[6]

Collision and decay processes may involve atoms and nuclei as well as particles. Four that are important for cosmology are shown in Table 6.2. In each case, the

Table 6.1 Particle collision processes. The plus signs don't mean that numbers are being added, they simple stand for 'and'. The forward arrow means that the particles on the left can disappear, to produce the particles on the right. The backward arrow means that the reverse process can occur

photon + photon $\rightleftharpoons$	electron + positron
neutrino + anti-neutrino $\rightleftharpoons$	electron + positron
proton + electron $\rightleftharpoons$	neutron + neutrino

Table 6.2 Nuclear and atomic collision processes

^{2}H + photon $\rightleftharpoons$	proton + neutron
^{3}He + photon $\rightleftharpoons$	proton + ^{2}H
^{4}He + proton $\rightleftharpoons$	^{2}H + ^{3}He
hydrogen atom + photon $\rightleftharpoons$	proton + electron

[4]That is true for any process; if it can occur, so can the reverse process.

[5]The positron is the anti-particle of the electron. As seen in Table A.4, many other particle species come with an antiparticle which always the same mass and (if charged) the opposite electric charge.

[6]As we shall see, the Cosmic Neutrino Background is an exception to that statement.

Table 6.3 These three quantities are conserved in every collision and decay process. The proton and neutron carry baryon number and are called baryons. The electron, positron, neutrino and anti-neutrino carry lepton number and are called leptons

	Baryon number	Charge	Lepton number
Proton	1	1	0
Neutron	1	0	0
Electron	0	−1	1
Positron	0	1	−1
Neutrino	0	0	1
Anti-neutrino	0	0	−1

produced particles have more mass than the colliding particles and the colliding particles must have enough kinetic energy to create the extra mass. The reverse of each process can also occur, which releases kinetic energy. The reverses of the second and third processes in Table 6.2 are examples of what is called **nuclear fusion**, in which two nuclei combine to form a heavier nucleus. They are what make the Sun and stars hot, and are used in H-bombs.

There are also processes, which I won't deal with, that produce kinetic energy by splitting a nucleus into lighter nuclei. These are called **nuclear fission**, and are used in nuclear reactors and ordinary atomic bombs. Einstein's telegram quoted at the head of this chapter was referring to the existence of atomic bombs.

Other Conserved Quantities

Energy is not the only quantity that is conserved. Another conserved quantity, that we already met, is electric charge and according to the Standard Model there are two more called **baryon number** and **lepton number**. The baryon number and lepton number of each species is listed in Table 6.3, along with its electric charge. It's easy to check that all three quantities are conserved by each of the processes in Tables 6.1 and 6.2.

Baryon number and lepton number are just book-keeping devices. Their conservation tell us that some processes don't happen, and that's all. For instance, the collision of a proton with an electron can't produce a pair of photons, because that would fail to conserve both baryon number and lepton number even though it conserves electric charge. In contrast, electric charge has significance going beyond mere conservation. The electric charge of an object determines the strength of the electromagnetic field that it generates, and the effect on its motion of that field.

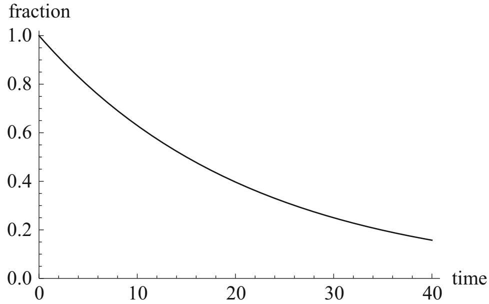


Fig. 6.2 Neutron decay This graph shows the fraction of neutrons, starting with a large sample, that will be present after a given number of minutes. After 15 minutes, half of the neutrons have decayed. That number will in turn be halved in another 15 minutes, and so on. For a single neutron, the graph gives the probability that an existing neutron will still be present after a given time interval. All of this applies to any decay process, only the half-life being different

Decay Processes

Now I come to decays. To be specific I'll describe neutron decay. When a neutron decays it is replaced by a proton, an electron and an anti-neutrino. These three particles move outwards from the place where the neutron decayed. For a single particle one can't say when the decay will occur, but for a large sample of the particles one can say what fraction of them will remain after a given length of time. This is shown in Fig. 6.2 for the neutron. After 15 minutes, only half of the neutrons will remain and one says that 15 minutes is the 'half-life' of the neutron. We can also view Fig. 6.2 in a different way. We can take it to refer to a single neutron. Then, it gives the probability that the neutron existing at some time, will still exist later.

Almost all other particle species decay, with a half-life much shorter than that of the neutron. Some nuclei also decay, but the half-life of a naturally occurring nucleus is usually much longer than that of the neutron and it can be bigger than the age of the Universe.

When discussing a decay process, it is convenient to suppose that the observer is moving with the decaying particle.[7] For that observer, the energy of the decaying particle is its rest energy. Energy conservation therefore requires that *the mass of the decaying particle is bigger than the total mass of the decay products*. That, together with the conservation of the other three conserved quantities, explains why the electron and proton are stable. The electron is stable because there is no charged particle with smaller mass. The proton is stable because no particle with smaller mass has baryon number. If baryon number conservation is not exact, the proton might decay with a very long lifetime. Observation at accelerators requires that the

---

[7]The half-life is defined as the one found by such an observer. According to Special Relativity, a differently moving observer will find a longer half-life.

lifetime is bigger than about 10^{29} years. That's much bigger than the age of the Universe, which is as well or we wouldn't exist!

Speed and velocity. In Physics, 'speed' and 'velocity' mean different things. **Speed** means the same thing as in ordinary English, so it's specified by a number. To specify the **velocity** of an object though, you have to give its speed *and* the direction of its motion. If the direction changes, the velocity changes even if the speed remains the same. Speed in physics is usually denoted by v, with I suppose the word velocity in mind.

The energy unit

The MKS unit of of energy is called the **Joule**. The rest energy of a one-kilogram weight is therefore c^2 Joules, ie. 9×10^{16} Joules. Several other energy units are in use. Your electricity bill will use the kilowatt hour which is 3.6 million Joules. Your guide to nutrition will use the food calorie which is 4,200 Joules. Particle physicists use a unit called the eV (electron-volt) which is equal to 1.602×10^{-19} Joules. They also use multiples of the eV, and I'm using the **MeV** (pronounced em-ee-vee) which is a million eV.

Chapter 7
A First Look at the History

... nuclear species must have originated ... as a consequence of a continuous building-up process arrested by a rapid expansion and cooling of the primordial matter.
(R. A. Alpher and G. Gamov, in the journal 'Physical Review Letters', 1948)

We have some certain knowledge of the observable Universe, only after it is about a hundredth of a second old. I am calling that the known history and in this chapter I give a brief account of it.

Expansion of the Universe

The observable Universe is expanding. To understand what that means, we can begin with the Cosmic Microwave Background (CMB). It is almost homogeneous at present, and remains so as we go back in time. The expansion means that any two pieces of the CMB are moving apart.

The expansion is **isotropic** (the same in all directions). To see what that means, please look at Fig. 7.1. Each dot represents the CMB that is within some small region, and the dots are arranged to make a square grid. The isotropic expansion scales up the grid with the passage of time, but it remains square because the rate of expansion is isotropic.

As a result of the expansion, the distance between any two pieces of the CMB is proportional to a **scale factor**, which is the same for any two pieces. The scale factor is denoted by a, and its present value is chosen to be 1. At an earlier epoch, when the scale factor was say $1/2$, two pieces of the CMB were only half as far apart as they are now.

Each object is moving through the isotropically expanding CMB, with some velocity called its peculiar velocity. In particular, each galaxy has some peculiar velocity. If every galaxy had zero peculiar velocity, the distance between any two of them would be proportional to the scale factor. Taking account of the peculiar

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_7

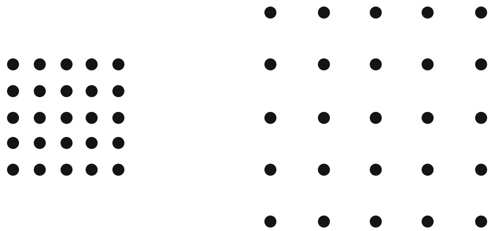


Fig. 7.1 Each *dot* represents a small piece of the cosmic gas. The *right hand grid* corresponds to the present epoch. The *left hand grid* corresponds to an earlier epoch, when the spacing of the points was only half as big as it is now. The scale factor a is equal to 1 at the present epoch, and to $1/2$ at the earlier epoch

velocity, it's still the case that the distance between a pair of galaxies is almost exactly proportional to the scale factor if they are far apart. That is because galaxies far apart are moving away from each other so quickly that their peculiar velocities make hardly any difference.

To describe the expansion, I focussed on the CMB. I could as well have focussed on the gas between the galaxy clusters, which I am calling the cosmic gas. It too, is almost homogeneous and is expanding almost isotropically just like the CMB. As we will see, the galaxies didn't always exist. Before they formed, the whole Universe was just a nearly homogeneous gas, that was expanding almost isotropically.

Abolishing Gravity

We are going to see how the collision and decay processes, described in Chap. 6, play out during the known history of the early Universe.[1] At first sight there seems to be a problem with doing that, because these processes are described by the Standard Model which ignores gravity. It turns out that gravity has a big effect on the expansion of the Universe, which seems to make the problem fatal.

Fortunately, it is always possible to abolish gravity within a small region for a short time. To understand how, we can think about the Earth's gravity. We don't actually have to abolish the Earth's gravity to use the Standard Model on the Earth's surface because the gravity is far too weak to matter. But suppose that were not the case; how then, could we use the Standard Model? The answer is simple. We would need to go to a space station orbiting the Earth. In that environment, gravity is abolished as we know from watching movies of the astronauts. Gravity is abolished because it has the same effect on every object within the space station, and on the

[1]By 'early Universe', I mean the Universe before there were any galaxies or stars.

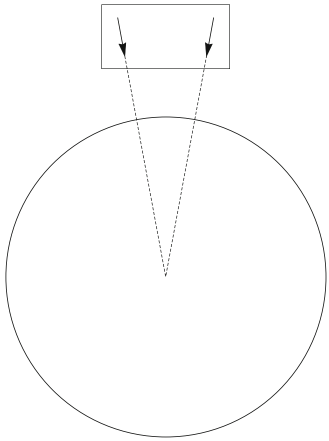


Fig. 7.2 The *circle* represents the Earth, and the *box* represents a ridiculously large space station. It is so big that the direction of the Earth's gravity varies significantly as we go across it. In such a space station, the Earth's gravity would pull together objects on opposite sides of the station, and gravity would no longer be abolished within it

space station itself.[2] There's no need for the space station to be in orbit. It could be moving away from the Earth with it's motors switched off, or falling toward it, and gravity would still be abolished. With the latter possibility in mind one says that the space station, or any object subject only to the force of gravity, is free-falling.

For gravity to be abolished, it's important that the space station is much smaller than the Earth. If it were not, then as illustrated in Fig. 7.2 the force of gravity would change direction as one moved across the space station. As a result, objects at opposite sides of the station would be pulled toward each other by the Earth's gravity. It's also important that we don't consider a long time interval, because no matter how small is the space station the direction of gravity will vary slightly within it, giving a significant effect if we wait long enough.

Although I focussed on a space station, I would as well have focussed on any free-falling region that is sufficiently small, and is considered for a sufficiently short time. This, then, is how we must apply the Standard Model to the early Universe. We must focus on a piece of the cosmic gas that is sufficiently small, for a sufficiently short time.[3] It turns out that 'sufficiently small' means 'much less than the size of the observable Universe' and that 'sufficiently short' means 'short enough that there is hardly any expansion'. To deal with the *whole* observable Universe and the *whole* known history, we have to do it in a sort of piecemeal way, by splitting up space and

[2]Objects within the space station are of course still pulled toward each other by the force of gravity, but that is a tiny effect because we deal with objects of ordinary size. There are no planets within the space station, and no moon!

[3]In a homogeneous Universe we don't have to worry about the requirement of free fall, because gravity is the only thing that affects the expansion rate of the Universe which means that each piece of the cosmic gas is indeed free-falling.

time into small chunks. In the following account of the known history I will take that for granted.

What Happens During the Known History

The known history is summarised in Table 7.1, and I'll go through it stage by stage. At the first stage, corresponding to the first line of the Table, the Universe contains several particle species. There are[4] protons, neutrons, photons, electrons, positrons, neutrinos and antineutrinos and the CDM. All of the collision processes listed in Table 6.1 are frequently occurring. The inverse of each process also occurs at the same rate, corresponding to what is called thermal equilibrium. This determines the number of protons per neutron; they are about equal at early times but by the end of the first stage there are about seven protons for every neutron. Thermal equilibrium also determines the number densities and the energy densities of the other species. They turn out to be about equal to each other, and much bigger than those of the protons and neutrons.

Table 7.1 Known history of the Universe. The first column shows the age of the Universe. The second column shows the scale factor *a*, which is the size of the region that will become the observable Universe, divided by its present size. In the third column, *E* is the average energy of a photon. In the first line, $<$ means less than and $>$ means more than. We don't know when this first stage begins

Age	a	E	What happens
<1 second 1 second	$<2.0 \times 10^{-10}$ 2.0×10^{-10}	>4 MeV 4 MeV	Frequent interactions, Neutrino interactions cease, CNB released
30 seconds	1.1×10^{-9}	0.5 MeV	Electron-positron annihilation
300 seconds 380,000 years	3.2×10^{-9} 1/1100	0.17 MeV 7.0×10^{-7} MeV	Big Bang Nucleosynthesis Epoch of last scattering CMB released
1.9×10^{8} years	0.05	1.3×10^{-8} MeV	First galaxies form
5.8×10^{9} years	0.5	1.3×10^{-9} MeV	Galaxy clusters form
9.3×10^{9} years	0.7	9.2×10^{-10} MeV	Sun forms
13.8×10^{9} years	1	6.5×10^{-10} MeV	Present time

[4] I'm dealing here with the situation towards the end of the first stage, which is all we know about for sure. We will see in Chap. 16 what might have happened earlier.

Densities

In physics, the density of any quantity is the amount of it in a small region, divided by the volume of the region. It's practically independent of the size of the region if it is small enough, unless we go down to the level where the region contains only a few particles. 'Density' on its own means the density of mass (mass density). In cosmology we are concerned with the density of energy (energy density) and the density of the number of particles of a given species (number density) and the densities of conserved quantities.

As the Universe expands, the average distance between the particles increases and their average energy decreases. As a result of these things, collision processes that involve neutrinos become less likely, and when the Universe is about a second old they practically cease marking the end of the first stage. The neutrinos then existing, exist still as the Cosmic Neutrino Background. About half a minute later, the energy of a typical photon becomes less than the rest energy of an electron. As a result, the process in the first line of Table 6.1 ceases. The reverse of the process removes all of the positrons and nearly all of the electrons, leaving just one electron per proton. We call this electron-positron annihilation.

After about five minutes, the neutrons combine with protons to form nuclei. It did not happen earlier, because the nuclear collision process in Table 6.2 broke up any nucleus that managed to form. Almost all of the neutrons end up inside ^{4}He nuclei, which make up about 25 % of the mass of ordinary matter. Almost all of the remaining 75 % comes from single protons (ordinary hydrogen nuclei). Some other light nuclei are created at the same time, and the other nuclei are created much later in supernovas, but these nuclei account for only a tiny fraction of the mass. As a result, about 25 % of the mass of ordinary matter is still ^{4}He and 75 % is still ordinary hydrogen.

After the formation of light nuclei, nothing happens for about half a million years. Then, almost every electron combines with a nucleus to form a helium or hydrogen atom. This didn't happen earlier, because there were many photons with energy bigger than the atomic binding energies. Those photons would have broken up any atom that managed to form.

After atom formation, the photons no longer bounce off electrons as they did earlier. Instead, each photon moves in a straight line. The epoch of atom formation is therefore called **the epoch of last scattering**. ('Scattering' is the physics term for the bouncing of particles off each other.) The photons existing just after last scattering are now seen as the Cosmic Microwave Background (CMB).

While all of these things are happening, the Universe is nearly homogenous. It's not completely homogeneous though, and some regions are slightly more dense then their surroundings. Additional matter is attracted towards such regions through the force of gravity, and after about 100 million years some of the regions collapse under their own weight, to form the first galaxies. Within each galaxy, over-dense regions in turn collapse to form stars leading to the Universe that we now observe.

Formation of light nuclei

The process starts with the formation of ^{2}H nuclei, which is the reverse of the process shown in the first line of Table 6.2. The formation becomes effective only when the average photon energy becomes so small that there are hardly any photons with the enough energy to undergo the direct process. More nuclear reactions quickly occur after the formation of ^{2}H nuclei, and Fig. 7.3 shows how the calculated abundance of each nucleus (and of the neutron) changes with time. To do this calculation, the rate at which each nuclear reaction occurs has to be determined by laboratory experiments. (The rates can in principle be calculated from the Standard Model, but the calculation is too difficult to perform.)

Almost all of the neutrons end up inside ^{4}He nuclei, but a tiny fraction of the neutrons end up inside other light nuclei. These are ^{2}H and ^{3}H, ^{3}He, and various lithium and beryllium nuclei. One would like to compare the calculated abundances with observation. What we observe though, are the abundances within stars of various kinds. During the process of star formation, further nuclear reactions take place which affect the abundances. To make the comparison one has to allow for those reactions, which can be done with sufficient accuracy only for ^{4}He, ^{2}H, and ^{7}Li. The reactions hardly affect the abundance of ^{4}He, but they affect the other abundances quite a bit. Allowing for the reactions, the calculated abundances agree with observation, except for the ^{7}Li abundance which seems to be too high.

The lithium discrepancy might be due to an error in the abundance of lithium in stars, that is deduced from observation. It might also be due to an error in the calculation of the change in the lithium abundance, caused by nuclear reactions within those stars. Otherwise, the lithium discrepancy means that there is something wrong with the calculation of the original lithium abundance. There could be an error in the laboratory determinations of the relevant nuclear reaction rates. Instead though, there could be an error in what has been assumed about the early Universe. One possibility is that an unstable particle is present, which decays at about the same time as the lithium is formed. Another is that the density of the protons and neutrons depends strongly on position. A third is that one or more of the fundamental constants is actually varying with time so that it is different when the lithium is formed. Time will tell which explanation is correct so—watch this space!

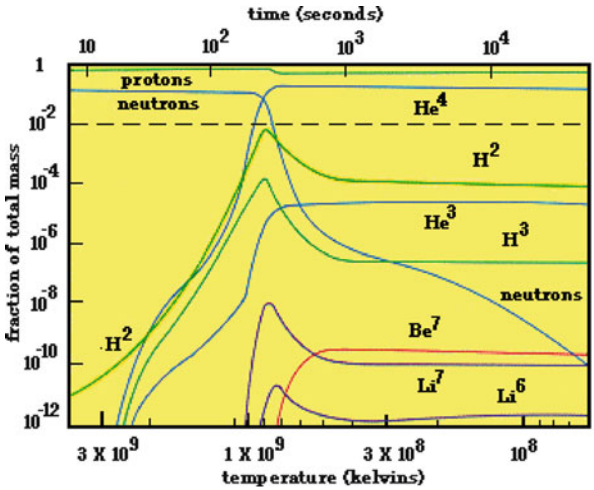


Fig. 7.3 Formation of light nuclei The upper horizontal scale is the age of the Universe in seconds. The lower horizontal scale is the temperature. The vertical scale gives the mass fraction of each nucleus. When the age is less than a few seconds, the number of neutrons per proton is determined by thermal equilibrium. After that, thermal equilibrium fails for the neutrons and protons and the neutron starts to decay with its half-life of 15 minutes. When the age is about 300 seconds, almost all of the neutrons bind into ^{4}He nuclei, but there is a small fraction of other nuclei (Source: Ned Wright [www.astro.ucla.edu/~wright/BBNS.html])

Why atom formation is the epoch of last scattering

To understand why the photons don't bounce off the electrons after the formation of atoms, it's useful to think in terms of electromagnetic radiation. After atom formation, each electron is inside an atom and the wavelength of the CMB is much bigger than the size of an atom. Since the atom has zero total charge, the electromagnetic radiation behaves as if it were encountering an object with zero charge which means that nothing happens to it.

Chapter 8
Energy Density of the Universe

> The introduction of such a [cosmological] constant implies a considerable renunciation of the logical simplicity . . . I am unable to believe that such an ugly thing should be realized in nature.
> (Albert Einstein in a letter to Georges Lemaitre, 1947)

One of the most important properties of the Universe is its energy density. By energy density, I mean the energy per unit volume. In other words, the energy within a region, divided by the volume of that region. As the Universe is homogeneous, it doesn't matter where the region is located. It must not be too big or too small though. It must be small compared with the size of the Universe, so that gravity can be ignored as described in Chap. 7. But it must be big enough to constitute a fair sample of the Universe.

How big that is, depends on the epoch under consideration. After the formation of galaxies the region must contain many galaxies, and after the formation of galaxy clusters it must contain many galaxy clusters. In the early Universe though, before galaxy formation gets underway, all we need is for the region to contain many particles.

About the Cosmological Constant

Table 8.1 shows the contribution to the present energy density, coming from each component of the Universe. The biggest contribution comes from the Cosmological Constant. As already mentioned in Chap. 1, the Cosmological Constant is the energy density of the vacuum.

How do we know that the vacuum energy exists? As we will see in Chap. 9, we know it exists through its gravitational effect. This is the same as the situation for the Cold Dark Matter (CDM). For that reason, the vacuum energy is sometimes called dark energy. The idea that the vacuum can have energy density is, of course

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_8

Table 8.1 The energy density of the Universe At the present epoch, the total energy density is 4,900 MeV per cubic meter, and the fraction contributed by each component is shown in the Table. Photons not in the CMB give a negligible contribution, as do neutrinos not in the Cosmic Neutrino Background. I am ignoring the neutrino mass which will be dealt with in Chap. 9

Cosmological Constant	CDM	Ordinary matter	CMB	Cosmic Neutrino Background
0.69	0.26	0.048	5.4×10^{-5}	3.7×10^{-5}

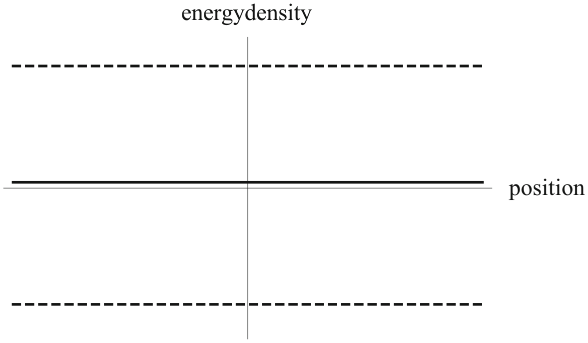


Fig. 8.1 The Cosmological Constant The *upper dashed line* is the quantum contribution and the *lower dashed line* is the classical contribution. The *full line* is the sum of the two, that is the observed value of the Cosmological Constant. The diagram is not to scale, because the quantum contribution is at least 10^{60} times as big as the observed value

outrageous. Einstein realised almost immediately that this is possible according to General Relativity, but the idea was not popular and a Cosmological Constant was generally regarded as unlikely.[1] Only in the late 1900s, was the existence of the cosmological constant finally established. Nature has evidently chosen to be outrageous!

Apart from its existence, the cosmological constant is puzzling in another way; a quantum effect, known as a vacuum fluctuation, would by itself make the cosmological constant far bigger than it actually is. To get the observed value of the cosmological constant, we must suppose that it also has a 'classical' contribution. As shown in Fig. 8.1, the two contributions must cancel to high accuracy.

There is no known reason why the cancellation should occur. On the other hand, if it did not occur, the cosmological constant would be much bigger that it is. As we will see in Chap. 12, that would have prevented the formation of galaxies, which would mean that there would be no stars, and no humans. Partly with that fact in mind, many physicists entertain the idea that the observable Universe may be just

[1]Einstein himself didn't regard the Cosmological Constant as the energy density of the vacuum, but instead as a fundamental constant possessed by General Relativity in addition to G and c. That makes no difference, because the effect of Einstein's Cosmological Constant is exactly the same as the effect of vacuum energy density.

one of many Universes, in which the cosmological constant takes on practically every value. Then we live in one of the Universes that corresponds to a tiny value simply because we could not live anywhere else.

The same line of reasoning has been invoked to explain some other puzzling features of the world we live in. It is called the anthropic principle and the physics community is polarized between those who love it and those who hate it. Those who hate it take the view that the smallness of the cosmological constant is an accident, or else that it can be explained by some as-yet undiscovered theory. That situation is never going to change, because the existence of a few or even a few dozen coincidences will not convince everybody that the anthropic principle is their explanation.

How the Energy Density Depends on the Scale Factor

Now we'll see how the energy density of each component of the cosmic gas changes with time. To be precise, we'll see how it depends on the scale factor a.

To do that, we have to think about the motion of the particles that make up the gaseous Universe.[2] To define the energy density we work within a region that is much smaller than the observable Universe. The edge of such a region is moving outwards with speed much less than c. But the particles have random motion, and for some species the typical speed is close to c (or, for photons, equal to c). Those species are called **radiation**. The other species, with speeds much less than c, are called **matter**. According to this terminology, the ordinary matter and the CDM are both matter, while the neutrinos and the photons are both radiation.

The matter is easy to deal with. because the energy of each particle is practically equal to its rest energy. As a result the energy of the matter within the expanding piece of gas is constant. Its volume on the other hand is proportional to a^3. (To see this, think about a cube of the expanding gas. The length of each side is proportional to a, and the volume is equal to the length cubed which indeed means that it is proportional to a^3.) The energy density of the matter is therefore proportional to $1/a^3$. It is shown in Fig. 8.2.

What about the radiation? The energy of the radiation within an expanding region decreases as the region expands, because energy has to be expended to push against the pressure of the surrounding radiation. It turns out that the energy of the region is proportional to $1/a$. As a result, the energy density is proportional to $1/a^4$. Assuming that the number of particles in the expanding region doesn't change,[3] their average energy is proportional to $1/a$. This was shown for the photon in Table 7.1.

[2] I take the constituents of the gas to be particles to keep the language simple. Depending on the epoch, the constituents may also include nuclei, atoms and molecules.

[3] That turns out to be true for the neutrinos, and also for the photons except at the epoch of electron-positron annihilation.

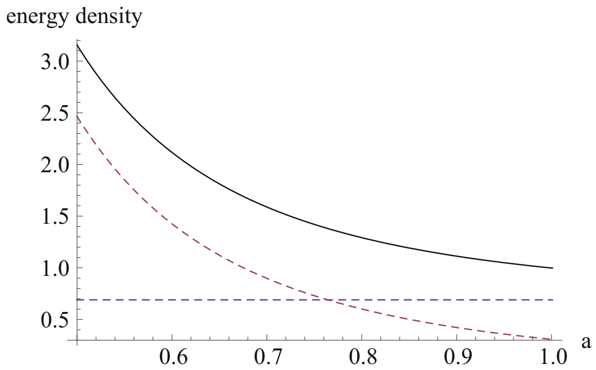


Fig. 8.2 The *full line* is the energy density of the Universe, divided by its present value. The *horizontal dashed line* is the contribution of the Cosmological Constant and the *other dashed line* is the contribution of the matter, while the contribution of the radiation is too small to show

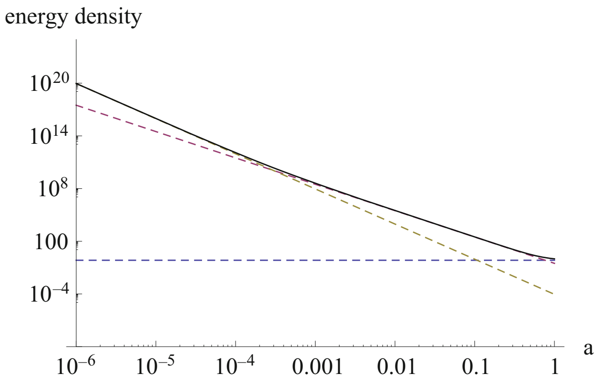


Fig. 8.3 The *full line* is again the energy density of the Universe divided by its present total energy density. The scales have been stretched so that 10^{-2}, 10^{-1}, 1, 10, 10^2 etc. are at equal distances on the paper. (Stretched scales allow large ranges to be equally visible, and I will often use them.) The *horizontal dashed line* is the contribution of the Cosmological Constant. The *dashed line* crossing it is the contribution of the matter and the *other dashed line* is the contribution of the radiation. The radiation dominates when a is less than 2.8×10^{-4}

All three contributions to the energy density are shown in Fig. 8.3. We see that the radiation gives the biggest contribution to the energy density at early times. The Universe is then said to be **radiation dominated**. For most of the subsequent history, matter gives the biggest contribution and we say that the Universe is **matter dominated**. Finally, at almost the present epoch, the cosmological constant starts to become significant.

Chapter 9
In the Beginning

> But although, as a matter of history, statistical mechanics owes its origins to investigations in thermodynamics, it seems eminently worthy of an independent development, both on account of the elegance and simplicity of its principles, and because it yields new results and places old truths in a new light in departments quite outside of thermodynamics.
> (J. Willard Gibbs (1839–1903), quoted in the textbook 'Thermal Physics' by C. Kittel and H. Kroemer)

The laws of physics determine the entire history of the Big Bang, as a consequence of the conditions that exists at its beginning. It's as if one had wound up an old-fashioned clock and then released the mechanism to do its own thing.

Instead of going to the very beginning, one can take any later epoch as the 'beginning'. Conditions existing then still determine the subsequent history. It's like winding up the clock only part of the way before releasing it. In this chapter, I'm going to take the beginning to be the first stage of the known history, corresponding to the first line of Table 7.1.

Thermal Equilibrium

To deal with the first stage, I need the concept of **thermal equilibrium**, which in turn involves the concept of **temperature**. When a region of space is in thermal equilibrium, it has a well-defined temperature. Given enough time, and left to itself, any region will come into thermal equilibrium and then things will stop changing, or only change slowly.

Temperature is of course a familiar concept, and for once everyday experience is a good guide to the physics. Suppose I get into my car one cold morning. The air outside the car, the car itself and the air within it are all at the same temperature; everything is in thermal equilibrium. Now I switch on the heating, which destroys

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_9

the thermal equilibrium by injecting hot air into the cold car. After a while, the air in the car is in thermal equilibrium at a higher temperature than before, and let's suppose that I now switch off the heating. The air cools down a bit, while remaining in thermal equilibrium, and it heats up things in the car like the steering wheel. After a while everything is in thermal equilibrium at the same temperature as the air. Everything, that is, except myself. I have my own heat supply, and my body core is in thermal equilibrium at a higher temperature. Heat is flowing from my body into the car.

The description that I've just given, which ignores the microscopic structure of things, corresponds to what is called thermodynamics. If we ask what is happening to the molecules of the air, the steering wheel and myself, that we arrive at statistical mechanics. From statistical mechanics, we learn that a gas can only be brought into a state of thermal equilibrium by collisions between its constituents. The gas may remain in a state of thermal equilibrium if the collisions cease, but the collisions must have happened in the first place to bring about that state. An example of this is provided by sunlight, which can be regarded as a gas of photons. Those photons were brought into a state of equilibrium by collisions within the Sun, and they are still approximately in that state even though they are undergoing no collisions. Another example is provided by the CMB photons. Before last scattering the CMB photons were brought into a state of equilibrium by collisions with other particles, and they are still in such a state now.

According to statistical mechanics, the temperature of a substance determines the average kinetic energy of its constituents. If we keep reducing the temperature the average kinetic energy will go down and it will eventually reach zero. The temperature at which that happens is the same for every substance, and it's called absolute zero. At absolute zero, every constituent has zero kinetic energy.[1] Nothing is moving! If we adopt the usual Celsius temperate scale, denoted by the letter C, absolute zero corresponds to $-273\,^{\circ}$C. In physics though, it's more convenient to take absolute zero to be zero temperature. We then arrive at the Kelvin temperature scale, denoted by the letter K. The freezing point of water corresponds to $0\,^{\circ}$C, or $273\,^{\circ}$K.

When a gas has two or more constituents, not all of the constituents need to be in thermal equilibrium. Also, different constituents may have different temperatures. As we'll see, that is important for cosmology.

Thermal Equilibrium During the First Stage

During the first stage of the history, all three of the processes in Table 6.1 are frequently occurring. These processes involve all of the particle species that are

[1] As kinetic energy is a positive quantity, its average can be zero only if the kinetic energy of every constituent is zero.

present, except the Cold Dark Matter (CDM) particles. As a result, all particle species except the CDM are in thermal equilibrium at the same temperature.

The particles of the CDM rarely if ever undergo collision processes with ordinary matter, so it will certainly not be in thermal equilibrium at the same temperature. The CDM particles might be bouncing off each other, which could put them into thermal equilibrium at a different temperature.[2] Such 'self-interacting CDM' is not thought to be likely though, which means that the CDM is probably not in thermal equilibrium even with itself. I'll ignore the CDM for the rest of this chapter.

The properties of a gas in thermal equilibrium are determined by the densities of the quantities that are conserved by the collision processes. One of these is always the energy, but with other conserved quantities in place its density is determined by the temperature so we don't need to consider it. In our case there are three other conserved quantities. As shown in Table 6.3, these are the electric charge, the lepton number and baryon number.

The electric charge density must be zero to very high accuracy, because there is no significant electric force in the Universe. In the simplest version of the history, that I am adopting, the lepton number density is also assumed to be zero, or anyhow so small that it doesn't matter. The baryon number density is the same thing as the number density of protons, plus the number density of neutrons. It's certainly not zero, and its value during the first stage has to be chosen so that the present baryon number density agrees with observation.

With the densities of the conserved quantities specified, thermal equilibrium determines the other properties of the cosmic gas during the first stage. They depend, not on the temperature itself but on the temperature times a number called the Boltzman constant. The Boltzmann constant is denoted by k, and is equal to 8.62×10^{-11} MeV per degree.

If we denote the temperature by T, the average photon energy is $2.70kT$. As shown in Table 7.1, the average photon energy is bigger than 4 MeV during the first stage, which means that the temperature during the first stage is bigger than 10^{11} °K. In contrast, the temperature in the Sun is only (!) about a million degrees.

What about the energies of the other species that are present? In the terminology of Chap. 7, all species except the proton and neutron turn out to be radiation. In the same terminology, the proton and neutron are matter at least by the end of the first stage.[3]

Thermal equilibrium determines the number of protons per neutron. At early times there is about one proton per neutron, but by the end of the first stage there are about seven protons per neutron. The number is still about the same when the first

[2] To agree with observation, the temperature would have to be low, which is why it's called *Cold* Dark Matter.

[3] We'll see in Chap. 10 what may have happened at much earlier times.

nuclei form because hardly any of the neutrons have decayed. This accounts for the fact that ^{4}He makes up about 25 % of the total mass.[4]

As the Universe expands, the temperature falls. One can show that it is proportional to $1/a$.

How thermal equilibrium works

Let's take as an example, the number of photons per electron. When that number corresponds to thermal equilibrium, the collision process in the first line of Table 6.1 (which replaces two photons by an electron-positron pair) takes place at the same rate per unit volume as its reverse. If the number were bigger, the forward process would occur at a faster rate than the reverse process, which would reduce the number. If it were smaller, the forward process would occur at a slower rate which would increase it. Either way, we would end up with the equilibrium number.

It's the same for any quantity; if there isn't thermal equilibrium, collisions between the particles change the quantity, and it stops changing only when thermal equilibrium is achieved.

One might think that the state of a gas in thermal equilibrium depends on the details of the collision processes, as specified by the Standard Model. That is not so, because the interactions have a property called time-reversal invariance, and it can be shown that this property alone determines the distribution functions at the thermal equilibrium. To understand time reversal invariance, imagine that a film is made of a pair of colliding billiard balls. If the film is shown run backwards, the viewer will have no way of knowing that. This is the simplest example of time-reversal invariance. For a collision process, we need to specify that each particle moves in a certain direction with a certain energy. Time reversal then says the probability for this process is the same as the probability for the reverse process.

Actually, time reversal invariance is not precisely correct; the fundamental laws as given by the Standard Model do involve the arrow of time if we work to a very high level of precision. That doesn't matter for our purpose. At the level of precision that we need, time reversal invariance can be regarded as correct.

[4] As the mass of the proton and neutron are almost the same, the proportion of ^{4}He by weight is practically equal to $2n/(n+p)$ where n is the number of neutrons in a region and p is the number of protons. This is equal to $2/(1+p/n)$, which means that indeed $p/n=7$ corresponds to 25 % by weight of ^{4}He.

Chapter 10
Friedmann Tells the Time

... my criticism ... was based on an error in my calculations. I consider that Mr. Friedmann's equations are correct and shed new light.
(Albert Einstein in the journal 'Zeitschrift Fur Physik', 1923)

The previous chapter makes no reference to time. It simply asserts that the Universe is in thermal equilibrium during the first stage of the history, and that the temperature is then proportional to $1/a$. To understand the rest of the history, as described in Chap. 7, we need the timescale shown in Table 7.1. Where does that come from? It comes from an equation derived by Alexander Friedmann in 1922, using General Relativity.

The Friedman equation tells us the effect of gravity on the rate at which the scale factor increases, or in other words on the expansion rate. According to the Friedmann equation, that effect depends on the scale factor itself, and the energy density. But we saw in Chap. 8 how the energy density depends on the scale factor. As a result, the Friedmann equation tells us the effect of the energy density on the expansion rate.

It turns out that the effect depends on where the energy density comes from. The part of the energy density that comes form particles has the effect of slowing down the expansion. In other words, the particles are being attracted towards each other by the force of gravity which is what one would expect. Strangely though, the part of the energy density that comes from the cosmological constant has the opposite effect; it increases the expansion rate. In other words, the cosmological constant is tending to push the particles apart.

According to observation, the energy density is related to the expansion rate in a very special way, which is described by saying that the energy density has the critical value. To define the critical value, we need to pretend that all of the energy density comes from particles. If that were the case, the expansion rate would always be decreasing and if the energy density were big enough it would eventually fall to zero and then become negative. In other words, the expansion would turn around, and become a contraction. On the other hand, if the energy density were small enough,

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_10

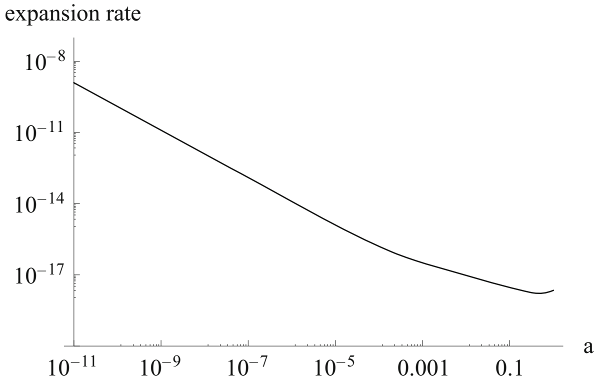


Fig. 10.1 The expansion rate is the rate of increase of the scale factor *a*, plotted here with the second as the time unit. Before the cosmological constant becomes significant, the expansion rate is decreasing because gravity is attracting the particles to each other. Recently, the expansion rate has started to increase because the cosmological constant corresponds to repulsive gravity which pushes things apart

it would have little effect and the expansion would continue for ever. The critical density is defined as the borderline between these possibilities; it is the one for which the expansion would eventually just grind to a halt, without actually turning round.

In summary then, the critical energy density is defined as the one that would just halt the expansion *if all of it came from the particles*. But part of the energy density actually comes from the cosmological contribution to the energy density, which as we saw in Chap. 8 is at present becoming the dominant contribution. As a result, the expansion rate is starting to increase and will continue to do so in the future, even though the energy density is critical. The dependence of the expansion rate on the scale factor is shown in Fig. 10.1. Using calculus, that dependence in turn determines the relation between the scale factor and the age of the Universe, displayed in the first two columns of Table 7.1.

The fact that the energy density is critical, relates to a story that begins around 300 BC. At about that time, Euclid of Alexandria produced a manuscript called ‘Elements’. The manuscript proves all sorts of statements about straight lines and circles, starting from a few assumptions which Euclid regarded as obviously correct. It proves, for instance, that the angles of a triangle add up to 180°. These statements, and their proofs, are called Euclidean geometry. Euclidean geometry agrees with everyday experience, and was accepted without question until the eighteenth century. Then, it was shown that by altering one of Euclid’s assumptions, logically possible alternatives could be constructed, called non-Euclidean geometry. In those alternatives, the sum of the angles of a triangle is different from 180°. It was found though, that even for a non-Euclidean geometry, Euclid’s geometry is still valid within any sufficiently small region. The fact that Euclid’s geometry agrees with everyday experience was therefore no guarantee that it would remain valid on larger scales.

There the matter rested, until Einstein came along with General Relativity. According to General Relativity, the geometry of the homogeneous Universe is Euclidean if the energy density is critical. That is the case according to observation, but it might have been otherwise and then Euclid would have been wrong.

All of this refers to the homogeneous Universe. The geometry of the Universe is different from Euclid's when we take account of the inhomogeneity. It is only slightly different in the early Universe, when the homogeneity is small. But it is significantly different near a very dense object like a neutron star or even a white dwarf, and near a black hole it becomes completely different.

How the Clockwork Universe Works

Now we are in a position to see how the laws of physics determine the known history of the Universe. The process starts with the first stage, which was described in the previous chapter. During the first stage, all three of the processes listed in Table 6.1 are frequent. To be precise, many of them can occur even in a time interval that is too short for the Universe to expand appreciably. The first stage ends when that ceases to be true, and the Standard Model tells us when that happens because it allows us to calculate the rate at which each collision process occurs.

After the first stage, the history is the one described in Chap. 7. It continues to be determined by the Standard Model, and the timescale provided by the Friedmann equation. The neutrinos remain in a state of thermal equilibrium, with temperature proportional to $1/a$. They exist now as the Cosmic Neutrino Background whose temperature is only $1.95\,^{\circ}$K. The photons also remain in a state of thermal equilibrium, with a temperature that is proportional to $1/a$ except for a slight boost around the time of electron-positron annihilation. They exist now as the Cosmic Microwave Background (CMB) with temperature $2.73\,^{\circ}$K.

To calculate the history of the Universe, we need of course the fundamental constants and some of the Standard Model parameters. We also need some numbers which specify the state of the Universe during the first stage. We can choose the numbers to refer to any epoch during the first stage. One of them is the temperature at the chosen epoch and another is the baryon number density. These are enough to determine the behaviour of the ordinary matter, but we need a third number to specify the energy density of the CDM, and a fourth to specify the Cosmological Constant.

Instead of actually considering the first stage, it's more convenient to specify these four quantities at the present epoch. That works just as well, because the present values are determined by the values during the first stage, so that knowing one set of values is the same as knowing the other. The present values are determined by the requirement that the calculation I have been describing agrees with observation, and you can find them in Table A.8 of Appendix B.

The Horizon Puzzle

The distance to the present horizon was defined in Chap. 1. It's the biggest distance that anything could have travelled since the beginning of the Big Bang. In other words, the distance that a photon would have travelled, had it been moving in a straight line.

One can also define the distance to the horizon at an earlier epoch of the Big Bang, as the biggest distance that anything could have travelled up to that epoch. It turns out that as we go back in time during the Big Bang, the region within the horizon represents an ever-decreasing fraction of the whole observable Universe.

That is a puzzle because it makes it difficult to understand why the entire observable Universe was almost homogeneous at early times. The homogeneity couldn't have been caused during the Big Bang, because there wasn't time for anything to move across the entire observable Universe. The homogeneity instead must have existed already, right at the beginning of the Big Bang. How did that happen?

As we will see, enough inflation can explain the homogeneity, because there is no limit to how far something can travel during inflation. As a bonus, inflation can also create the slight inhomogeneity, which we know existed and was the cause of galaxy formation.

Rate of change

If the expansion rate were constant we could define it as the change in a during any time interval, divided by that interval. In reality the expansion rate is changing with time, and we have to define it as the change in a during a *short* time interval divided by that interval. This definition is the starting point for calculus and it sounds sophisticated, but it's actually something that we take for granted in ordinary life. For instance, if I get into the car and drive down a straight road, my speed is the rate of change of my distance from the starting point. While the car is accelerating, the speed is increasing and it has to be defined as the change during a *short* time interval divided by that interval. If the speed at some instant is 50 kilometres per hour, that doesn't mean that I will travel 50 kilometres in the next hour. But 50 kilometres per hour is 14 metres per second, and I probably *will* travel about 14 metres in the next second. To even better accuracy, I will travel 1.4 metres in the next tenth of a second. We can define my speed using any time interval, so long as it's short enough that the speed hardly changes.

Chapter 11
Ripples in the Universe

> The COBE Differential Microwave Radiometer (DMR) instrument has observed ...structure ...primarily due to density variations, in the early universe. The results are supportive of gravitational instability theories of structure formation and inflation/quantum cosmology models.
> (George Smoot, in AIP Conference Proceedings, 1992)

I have so far taken the early Universe to be completely homogeneous. It's time now to deal with the slight inhomogeneity, which causes the CMB anisotropy displayed in Fig. 1.1 and leads eventually to the formation of galaxies.

The inhomogeneity of a quantity in the early Universe is called a **perturbation**. The most important perturbation is the energy density perturbation. At each epoch, it is convenient to define the **density contrast**, as the energy density perturbation divided by the average energy density. Figure 11.1 shows what the energy density perturbation might look like, and Fig. 11.2 shows the corresponding density contrast. As well as the perturbation in the total energy density, we are also interested in the perturbation in the energy densities of single components of the cosmic gas.

The term 'perturbation' indicates that we are dealing with a *small* variation with position. That corresponds to a density contrast which is much less than 1 in magnitude. As we will see, the density contrast of the CDM eventually becomes bigger than 1 in some places, leading to galaxy formation. At that stage though, one no longer deals with a perturbation. This chapter and the next deal only with perturbations.

Instead of the actual energy density, it is often convenient to consider the **smoothed** energy density. The smoothed energy density can be thought of as the one that an outsider would see, if they had poor eyesight. To be precise, the smoothed energy density at a given point in space is the average of the actual energy density, over a sphere centred at that point. As illustrated in Fig. 11.3, this indeed makes the density contrast smoother and it also makes it smaller.

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_11

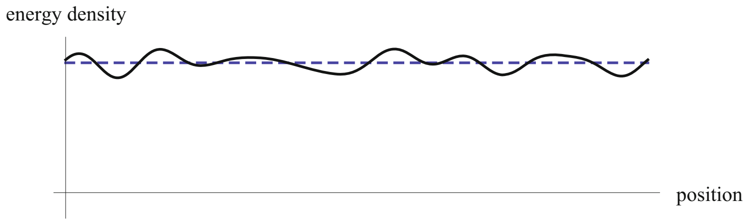


Fig. 11.1 Energy density The *full line* shows how the energy density might vary with position at some epoch. The *dashed line* is the average energy density, and the difference between the two is the energy density perturbation (edp)

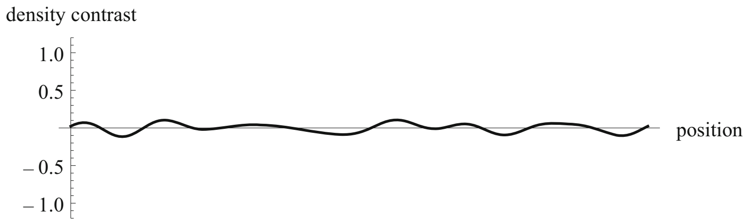


Fig. 11.2 Density contrast is defined as the edp divided by the average energy density. The *curve* shows the density contrast corresponding to the edp shown in Fig. 11.1. It is much less than 1 in magnitude and is independent of the energy unit

I will take the radius of the smoothing sphere to be proportional to the scale factor a, and will call its present value the **smoothing scale**. The smoothing scales that can be probed by observation range from the distance to the present horizon, which is 4.5×10^{10} lightyears, down to about 10^4 lightyears. I will call them **cosmological scales**.

The Initial State

The known history of the Universe is determined by the state of affairs at the beginning. We have seen how that works for the homogeneous Universe, and now we are going to see how it works for the perturbations.

At the beginning of the known history, the smoothing radius is bigger than the distance to the horizon on all cosmological scales. The perturbation in the smoothed energy density existing then, is called the primordial energy density perturbation. One usually assumes that at this stage, the energy densities of the separate components are completely determined by the total energy density. This is called the **adiabatic condition**, and it is consistent with current observation. One can show that in consequence of the adiabatic condition, each component of the cosmic gas starts out with about the same density contrast.

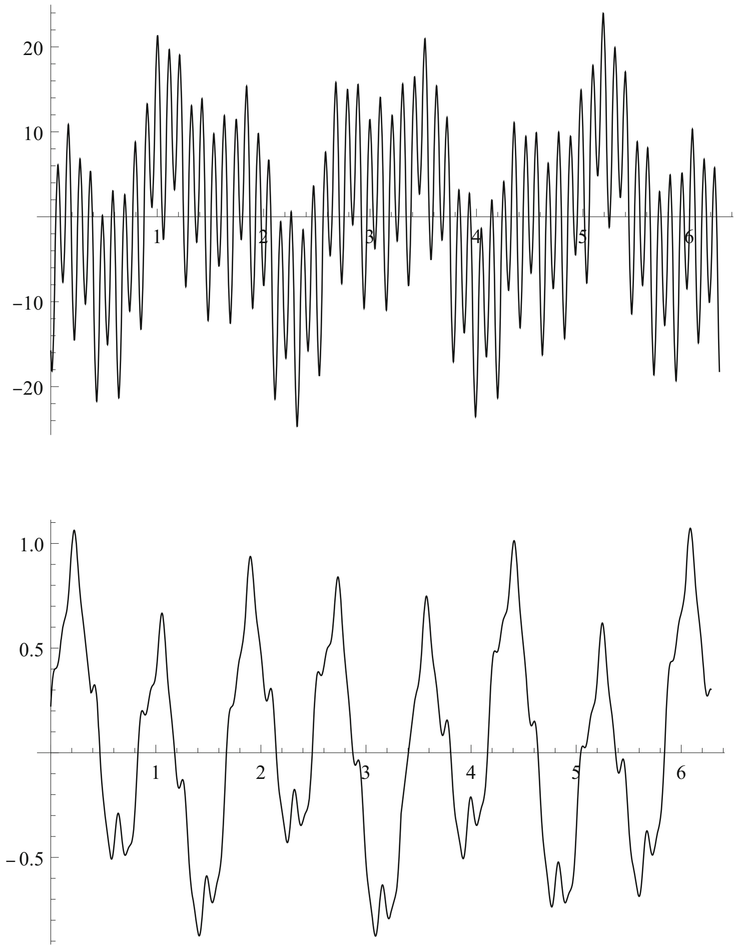


Fig. 11.3 Smoothing The *first graph* shows what the density contrast might look like, along some line in space. The *second graph* shows the effect of smoothing; at each point, the density contrast has been average over an interval which is centred on that point and whose length is 2 (The distance unit in these plots is not specified)

We are not interested in the details of the density contrasts. In particular, we are not interested in the locations of the individual peaks of the CDM's smoothed density contrast, that will become galaxies. The locations of individual galaxies are accidents, just as the distance of the Earth from the Sun is an accident. We interested instead, in the *statistical properties* of the density contrasts.

To understand what that means, let's imagine that a coin is tossed many times, and the results are written down. Then we could highlight all the 'heads' entries, and see how they are distributed. On average there would be as many heads as tails, but the distribution would be uneven and we would sometimes find two heads together. We might find even three or more but that becomes increasingly unlikely. We could work out of the probability for each of these possibilities, and of any other possibility regarding the distribution of ‘heads’. Then we would be studying the statistical properties of the distribution of ‘heads’.

The statistical properties of the density contrasts refer to their form at a fixed epoch. They are described in some detail in Appendix A. It turns out that the statistical properties at the beginning of the known history are specified by two parameters, whose values are shown in Table A.8 of Appendix B. Here in the text, I will only be dealing with the most important statistical property, which is the average of the square of the CDM matter density contrast.[1] This is called the **mean-square** because mean is another word for average.

The Acoustic Oscillation

Until the epoch of last scattering, the photons are bouncing off the electrons, and electrical neutrality ensures that the number densities of the electrons and nuclei are everywhere the same. As a result, the ordinary matter and photons can be regarded as a single constituent of the the cosmic gas which is called the baryon-photon fluid.

After the initial epoch, the smoothing radius becomes smaller than the distance to the horizon. That happens on successively smaller cosmological scales, starting with the biggest one. Except on rather small cosmological scales, it happens before the epoch of last scattering. We are then dealing with the baryon-photon fluid, whose density contrast starts to change rapidly. This is called the **acoustic oscillation** of the baryon-photon fluid because it resembles the rapid changes in the density of air that correspond to sound. It’s not like an ordinary sound wave though, which travels through the air from the place where it was generated. Instead, it’s like what is called a standing wave. As illustrated in Fig. 11.4, a standing wave oscillates without going anywhere. A standing sound wave is generated inside a wind instrument, which then travels out as an ordinary sound wave.

The CMB Anisotropy

Now we come to the CMB anisotropy, which was shown in Fig. 1.1. The CMB originates at the epoch of last scattering, on a sphere around us which is called the surface of last scattering The CMB therefore provides an approximate image of the energy density of the baryon-photon fluid on the surface of last scattering. The image is not exact though, because it is distorted by the motion of the baryon-photon

[1]The density contrast itself is positive in some regions and negative in others, and its average is zero.

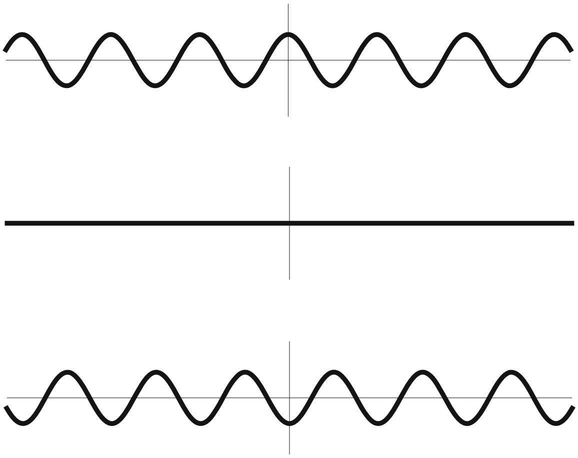


Fig. 11.4 Snapshots of the acoustic oscillation They refer to a single wave, the complete oscillation consisting of many waves with different wavelengths. The vertical axis is the density contrast during the acoustic oscillation of the baryon-photon fluid, and the horizontal axis is distance along the direction of the component. The changes take place over half a period. After another half period, the wave recovers its original form

fluid, and by the inhomogeneity that the CMB encounters on its journey from the last scattering surface.

Just as we are not interested in the detail of the density contrasts, we are not interested either in the detail of the CMB anisotropy. We are interested instead in its statistical properties. These are described in Appendix A.

The statistical properties of the CMB are determined by those of the primordial density perturbation. They depend upon all six of the cosmological parameters shown in Table A.8 of Appendix B. As a result, those parameters are determined quite accurately, just be the requirement that the statistical properties of the CMB agrees with observation. Additional observations, notably galaxy surveys, only slightly alter the estimated values of these parameters, and only slightly increase their accuracy.

Chapter 12
Galaxy Formation

> The large-scale anisotropy of the microwave background and the large-scale fluctuations of in the mass distribution are discussed under the assumptions that the universe is dominated by a massive, weakly interacting particle and that the primeval density fluctuations were adiabatic with the scale-invariant spectrum . . .
> (P. J. E. Peebles in The Astrophysical Journal, 1982)

In this chapter I describe galaxy formation. As already mentioned in Chap. 7, galaxies form when slightly over-dense regions of the Universe attract more matter and collapse under their own weight. What I want to emphasise here is the inevitability of the process. That it happens, and when it happens, is a consequence of the state of affairs in the early Universe, that I described in Chap. 11.

For both the CDM and ordinary matter, the energy density is just c^2 times the mass density and it makes more sense to deal with the latter. The density contrast, previously defined as the energy density perturbation divided by the average energy density, can as well be defined as the mass density perturbation divided by the average mass density. As in the previous chapter, I work with the smoothed density contrast.

As we saw in Chap. 8, the CDM outweighs the ordinary matter. As a result, the CDM smoothed density contrast evolves almost as if there were no ordinary matter. It grows with time because over-dense regions attract more matter towards them. The growth begins when the smoothing radius becomes less than the distance to the horizon. During matter domination, the CDM density contrast is proportional to the scale factor a. The smoothed density contrast has over-dense and under-dense regions, and as illustrated in Fig. 12.1 each region is roughly as big as the smoothing radius.

This description of the CDM's density contrast applies while it is much less than 1. When the CDM's density contrast becomes roughly equal to 1 in some region, the CDM in that region collapses under its own weight. As the collapse proceeds, the inward motion of the CDM particles is randomised by the effect of gravity, which

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_12

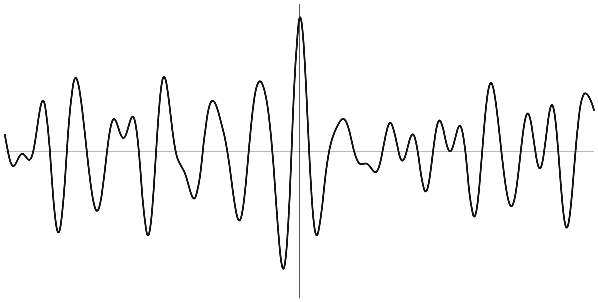


Fig. 12.1 This is what the smoothed density contrast might look like, in some region of space at an epoch before the formation of the first objects (galaxies or galaxy clusters, depending on the smoothing scale). The horizontal axis represents a line in space, centred on a region of exceptionally high density. There are other over-dense regions as well as under-dense regions, all with roughly the same size which is about the same as the smoothing radius. Regions that are sufficiently over-dense will collapse to form a galaxy or galaxy cluster, with the most over-dense regions collapsing first. The distance along the horizontal axis represents the distance from the exceptionally over-dense region, divided by a. A fixed point on the horizontal axis is moving with the expansion

halts the collapse when the collapsing region is about half as big as it was when the collapse started. We then have an object of fixed size, which is called a CDM halo.

At first, the collapse occurs only at places that represent exceptionally high peaks of the density contrast, like the one in the centre of Fig. 12.1. That doesn't significantly affect the density contrast at other locations, which continues to be proportional to the scale factor a. Eventually though, the collapse occurs even at peaks that are not exceptionally high. A large fraction of the matter in the Universe then collapses and we can no longer talk about a perturbation in the density of the CDM.

On a given scale, the final collapse occurs when the mean-square of the smoothed CDM density contrast is roughly equal to 1. Before it occurs the mean-square is proportional to a^2, because the density contrast itself is proportional to a. The mean-square is biggest on the smallest scales, because its growth started first on those scales. The lightest CDM halos form when the scale factor a is around $1/10$. After that, successively heavier CDM halos form.

So much for the CDM. What about the ordinary matter? At the epoch of last-scattering its smoothed density contrast is much smaller than that of the CDM, because it has been undergoing the acoustic oscillation instead of growing. After last-scattering, gravity tries to pull the ordinary matter into regions where the CDM is over-dense. On the other hand, the random motion of the ordinary matter tries to prevent that from happening. Gravity wins only if the mass of the CDM halo is big enough. We then have a galaxy or galaxy cluster, whereas for smaller masses we have a pure CDM halo. Pure CDM halos have not been observed. That is probably because they collide with each other, to form halos which are big enough to become galaxies.

As we saw, the collapse of a CDM halo is halted when the random motion of the CDM particles becomes big enough to counteract the effect of gravity. The same is true for the collapse of the ordinary matter, but the ordinary matter collapses further because it emits electric radiation which tends to reduce the random motion. As a result, the ordinary matter in a galaxy is confined to a luminous central region, with the CDM halo around it. That's why we call it a CDM *halo*.

The smallest galaxies form first. Successively bigger galaxies form through the collapse of successively bigger CDM halos, and also through the merging of existing galaxies.

Once galaxy formation is under way, a collapsing CDM halo sometimes produces a galaxy cluster instead of an individual galaxy. Successively larger galaxy clusters form in this way, until we arrive at the epoch when the cosmological constant starts to becomes significant. The process of gravitational collapse then ceases, because the gravitational effect of the cosmological constant is repulsive rather than attractive, as we saw in Chap. 9. If the cosmological constant lives up to its name, and indeed remains constant in the future, there will never be any gravitationally bound objects bigger than the biggest galaxy clusters.

Chapter 13
The Higgs Field

. . . as a consequence of this coupling, . . . some of the gauge fields acquire mass . . .
(Peter W. Higgs, in the journal 'Physical Review Letters', 1964)

In 2012, the discovery of the Higgs boson was announced. It had been created by particle collisions at the LHC in CERN Geneva, which had at last achieved the necessary collision energy. The discovery of the Higgs boson provided the final confirmation that the Standard Model is essentially correct. It also confirmed, indirectly, the existence of the **Higgs field** because that is another prediction of the Standard Model.

In this chapter I will discuss the Higgs field, and its relation to the Higgs boson. Apart from being important in its own right, this will set the stage for the discussion of inflation in the following chapters.

Recall that fields and elementary particles are the two kinds of fundamental objects that are believed to exist. An elementary particle is located at a point in space, whereas a field exists in an extended region of space or even throughout all space. We have so far met just two fields, namely the electric field and the magnetic field. The electric field exists only near an electrically charged object as illustrated in Fig. 4.4. It doesn't exist in empty space, because matter in bulk is electrically neutral, and doesn't generate an electric field. The magnetic field on the other hand exists at some level throughout the Universe. The magnetic field near the Earth is illustrated in Fig. 4.3.

To specify a magnetic or electric field, we have to specify both its strength and its direction. The Higgs field instead is specified just by its strength. A field of that kind is called a **scalar field**. Throughout the known history of the Universe, the strength of the Higgs field is homogeneous and constant. This strength is called the vacuum strength of the Higgs field, to remind us that it exists even in the vacuum.

According to the Standard Model, the vacuum strength of the Higgs field determines the mass of every elementary particle except the neutrino. That such a field might exist was suggested by Peter Higgs and others in 1964, some years before the completion of the Standard Model. In this chapter I will describe the

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_13

Higgs field in the context of the Standard Model which means that I ignore gravity. As explained in Chap. 7, that means working within a small part of the observable Universe, for a time that is so short that the expansion can be neglected.

When the Higgs field has its vacuum strength, its energy density is zero. It's useful also to think about the energy density that the Higgs field would have, if its strength were homogeneous and constant but different from its vacuum value. This energy density is called the Higgs potential. A simplified version of the Higgs potential is shown in Fig. 13.1, and I will use it because the real thing would complicate the discussion without adding anything very useful.

We see that the Higgs field strength can be either positive or negative, and that the potential is unaltered if we reverse the sign of the strength. We see also that the energy density of the Higgs field is zero for two values of the field strength. We don't know which of these has been chosen by nature as the vacuum value, because the predictions of the Standard Model are the same for both choices.

If the strength of the Higgs field were homogeneous, but slightly different from its vacuum value, it would oscillate around that value as illustrated in Fig. 13.2.[1] Here is a mechanical analogy. Suppose that you carefully place a ball-bearing on the surface of the gutter surrounding the roof of your house. If you place it at the bottom it will stay there, but if you place it away from the bottom it will roll down towards

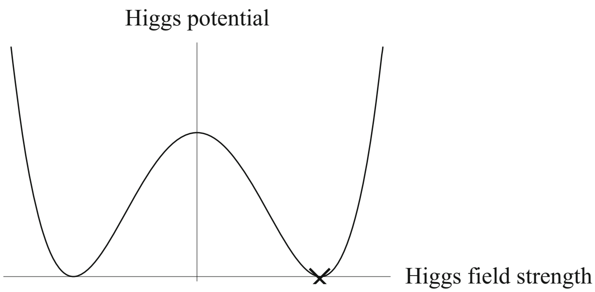


Fig. 13.1 Simplified version of the Higgs potential. I have taken the vacuum value of the field strength to be positive as indicated by the cross. The opposite choice would lead to the same physics

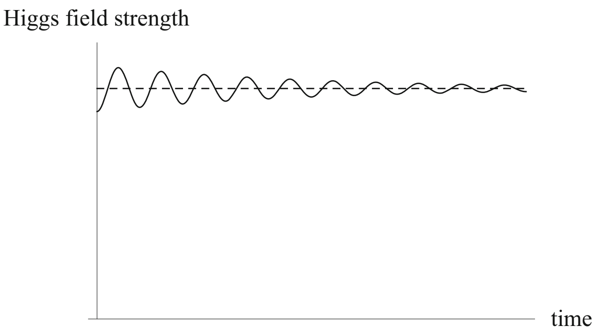


Fig. 13.2 If the Higgs field were slightly different from its vacuum value, while still being homogeneous, it would oscillate around the vacuum value. This would corresponds to the presence of Higgs bosons. Since the Higgs boson decays, its number density decreases with time and so does the amplitude of the oscillation

[1] That doesn't actually happen during the known history of the Universe, but it might have happened earlier. Whether it did or not, it's useful to discuss it as a hypothetical possibility.

the bottom. Then, if the gutter is smooth enough, the ball bearing will oscillate a few times before finally coming to rest.

That's enough for now about the Higgs field. What about the Higgs boson? The Higgs boson is related to the Higgs field in the same way that the photon is related to the electromagnetic field. Electromagnetic radiation, which consists of oscillating electric and magnetic fields, can be thought of as a collection of photons. In the same way, an oscillating Higgs field can be regarded as a collection of Higgs bosons. If the oscillating field is homogeneous, the Higgs bosons are at rest.

With the homogeneous oscillation in mind, we can understand how the vacuum value of the Higgs field strength generates masses for the other elementary particles. To do that, we have to realise that each elementary particle comes with a field. The fields of the other Standard Model particles are not scalar fields, and those of the quarks and leptons are essentially quantum objects that cannot be described using the language of classical physics, but they still exist. What the vacuum value of the Higgs field does, is to allow a *homogeneous* oscillation of the field of each particle. The homogeneous oscillation corresponds to a particle with mass. If the vacuum value of the Higgs field were zero, a homogeneous oscillation would be impossible. Instead, the oscillation would have to correspond to a wave travelling with the speed of light, and the particles would have zero mass.

In this discussion I have been assuming that the Standard Model is exactly correct. More probably, it will be some extension of the Standard Model that is correct. In some extensions, elementary particle masses depend on the vacuum values of two or more scalar fields and all of them are called Higgs fields. A Higgs boson has been observed at the LHC, which is referred to as the Higgs boson. In truth, we do not yet know whether it is *the* Higgs boson of the Standard Model or *a* Higgs boson of some extension. The matter is likely to be clarified in the near future, by more detailed observations.

Finally, I need to revisit the vacuum value of the Higgs field. When we talk about the vacuum value, we are ignoring quantum physics. When quantum physics is included, one finds that the strength of the Higgs field in the vacuum actually fluctuates around the vacuum value, varying with both time and position. This is called the **vacuum fluctuation** of the Higgs field. Every field has a vacuum fluctuation, and a similar fluctuation will be important when we come to inflation.

Chapter 14
Inflation

> . . ... hot big-bang cosmology requires initial conditions which are problematic . . .. These problems would disappear if, in its early history [the universe underwent] a period of exponential growth . . ..
> (A. Guth in 'Physical Review' 1980.)

Now it's time to deal with the era of inflation, which is believed to cause the Big Bang. In this chapter I explain how inflation can generate an almost homogeneous Universe. In the next I explain how it can generate the observed perturbations.

During the Big Bang, gravity is attractive; it pulls different parts of the Universe towards each other which slows down the expansion rate. During inflation, gravity is instead repulsive, speeding up the expansion rate. To achieve that, the energy density cannot be dominated by the contribution of particles. It must instead be dominated by the contribution of fields. The simplest assumption, which I will adopt, is that it is dominated by the potential of a single scalar field, called the inflaton field.

To achieve inflation, the inflaton field must vary slowly. That can happen if the potential has an almost flat region as shown in Fig. 14.1. Inflation ends when the inflaton field leaves the flat region, and starts to oscillate around its vacuum value. That corresponds to a Universe filled with 'inflaton' particles. The inflaton particles are supposed to decay, to eventually produce the particle species that are present at the beginning of the known history.

To do the job of creating the Universe that we observe, it's essential that inflation begins when the expansion rate $\dot{a}$ is much smaller than it is now. We have, then, two eras of inflation; an early era when the expansion rate is smaller than it is now, and a later era when it is bigger. The two eras do different jobs. The first era dilutes the energy density of any particles that might be present initially, so that the energy density of the inflaton field becomes completely dominant. This era also produces a practically homogeneous observable Universe, with the properties required by observation. The second era generates an energy density perturbation, that will become the one existing at the beginning of the known history.

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_14

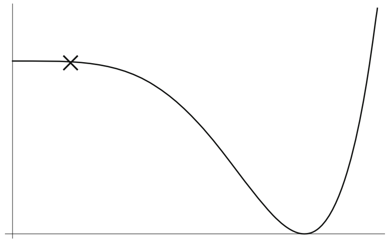


Fig. 14.1 A possible shape for the inflaton potential. Near the potential's maximum, there is a region within which it is much flatter than is the Higgs potential. If the inflaton is within that region, as indicated by the *cross*, it can change very slowly and cause inflation. Inflation ends when the inflaton field leaves the flat region, and starts to oscillate around its vacuum value

To do its part of the job, the first era must last for a sufficiently long time. Given that, let us see how the job is done. The first era makes the energy density homogeneous by driving energy away from regions of higher energy density into regions with lower energy density.[1] As a result, the inflaton field also becomes practically homogeneous and so does the rate of expansion. The first era makes the expansion practically isotropic simply because, according to General Relativity, any anisotropy of the expansion always decreases with time. Finally, as we see in Appendix A the fact that the expansion rate increases with time means that it will give critical energy density.

The way in which the second era can generate the energy density perturbation will be described in the next chapter. To get the statistical properties required by observation, the energy density during the second era must have some roughly constant value. This value is not known, but observation does tell us something about it. This is because the second era does something else in addition to generating the energy density perturbation; it generates a gravitational radiation background. The intensity of the gravitational radiation background depends on the value of the energy density during the second era, and the fact that the gravitational radiation background has not so far been observed tells us that the energy density during the second era of inflation can be at most about 10^{64} times its value at the beginning of the known history. If the energy density is not much below that limit, the gravitational radiation background will eventually be detected. It remains to be seen whether Nature has been so kind as to provide such a large energy density.

[1]Energy is driven away from over-dense regions because gravity is repulsive. This is in contrast with the situation after inflation, when attractive gravity pulls matter (and its rest energy) towards regions where there is already more matter leading eventually to the formation of galaxies.

Some history

An era of inflation was suggested around 1980 by several people, including Alan Guth who gave inflation its name. None of them explained how inflation would end. The first viable proposals for ending inflation were made in 1982, in a paper by Andrei Linde and another by Andreas Albrecht and Paul Steinhardt. These proposals invoked an inflaton potential with the shape shown in Fig. 14.1, that includes a nearly flat region. Linde in 1983 pointed out that the expansion of the Universe can make the inflaton field vary slowly even if its potential isn't particularly flat. He also pointed out that there is no need for inflation to be preceded by another Hot Big Bang, as had previously been assumed.

Who is the inflaton?

Only one scalar field is part of the Standard Model, namely the Higgs field. The Higgs field can be the inflaton if we don't insist on General Relativity during inflation, but we have to make sure that General Relativity still holds during the known history. Alternatively, the Higgs field can be the inflaton if we alter the Standard Model's prediction for the effect of the Higgs potential on the evolution of the Higgs field, but we have to make sure that the small oscillation of the Higgs field around its vacuum value (corresponding to Higgs bosons) is not significantly affected.

As both of these modifications are rather drastic they are usually ruled out. That has the unfortunate effect of ruling out, not just the Higgs field but any field that is part of a an extension of the Standard Model that has been proposed for reasons not having to do with inflation.

A varying dark energy?

As we noticed in Chap. 5, the energy density of the present Universe comes only partly from the particles that it contains. There is an additional contribution which, as far as we can tell, is constant. Taking that to be the case, the additional contribution is called the cosmological constant and is supposed to be a property of the vacuum.

It may be though, that the additional contribution varies with time. If that is the case, we had better call it just dark energy, rather than a cosmological constant. A time-varying dark energy cannot belong to the vacuum, because the vacuum is something that doesn't change with time. It might correspond instead come from the potential of some scalar field. Adopting that hypothesis, the dark energy is called quintessence, and the field responsible for it is called the quintessence field. The quintessence field cannot be the Higgs field, because a displacement of the Higgs field from its vacuum value would cause oscillations corresponding to the existence of Higgs bosons. To avoid such an oscillation, the quintessence field presumably needs to be on a nearly flat part of its potential. In that respect, the quintessence potential would be like the inflaton potential illustrated in Fig. 14.1, only much smaller because the vacuum energy density is much smaller than the energy density during inflation.

The idea of quintessence, is that the dark energy density would be zero if the quintessence field had its vacuum value. That sounds reasonable, but the problem is that there is no known reason why the energy density of the vacuum should be zero. In particular, there is no known symmetry of quantum field that would ensure that. Nevertheless, the quintessence idea is taken very seriously by the cosmology community and there is intense interest in future observations that may show a change in the dark energy density as we go back in time.

Chapter 15
Generating the Perturbation

. . .. models in which the de Sitter stage exists . . . are attractive because fluctuations of the metric sufficient for galaxy formation can occur.
(G. Chibisov and V. Mukhanov, in 'Soviet Physics', 1982)

In this chapter we will see how the second era of inflation can generate the energy density perturbation, that exists at the beginning of known history. The job might be done by a perturbation in the inflaton field, or else through a perturbation in some other scalar field.

To describe the perturbation in a scalar field, it's useful to smooth the field just as we smoothed the energy density. As in that case, we take the smoothing radius to be proportional to a, and call the present smoothing radius the smoothing scale.

To understand the behaviour of the smoothed inflaton field, we need a couple of new quantities.[1] We need the **Hubble time** which is the time taken for the Universe to expand appreciably. We also need the **Hubble distance** which is c times the Hubble time (in other words, the distance that a photon would travel in a Hubble time).

At the beginning of the second era, the smoothing radius on cosmological scales is smaller than the Hubble distance. When quantum physics is taken on board, that has an important implication. It means that smoothed inflaton field doesn't actually have the the precise value that I talked about in Chap. 14. Instead, it is fluctuating around that value. This quantum fluctuation of the inflaton field is analogous to the fluctuation of the Higgs field about its vacuum value.

As the second era proceeds, the smoothing radius becomes bigger than the Hubble distance. That happens on successively smaller cosmological scales, starting with the biggest one. When it happens, the smoothed inflaton field stops fluctuating and only changes slowly. Its perturbation generates a perturbation in the smoothed energy density, which in the simplest scenario becomes the primordial energy

[1]These are approximate definitions, see Appendix 2 for the exact definitions.

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_15

density perturbation described in Chap. 11. In that scenario, the adiabatic condition is automatically satisfied because the smoothed inflaton field at a given location determines all of the future properties of the smoothed energy density.

In alternative scenarios, the primordial energy density perturbation instead originates from the perturbation of a scalar field different from the inflaton field. Those scenarios don't automatically lead to the adiabatic condition.

Schrödinger's Cat in the Sky

The idea that the energy density perturbation originated as a quantum fluctuation during inflation is widely accepted. It raises though a serious problem, which also arises when we try to use quantum physics to describe collision and decay processes.

The problem arises, because quantum physics really describes only the outcome of measurements. It tells us the probability for the various outcomes of the measurement, but it doesn't tell us what was there before the measurement was performed.

To understand the problem, we can focus on a decay process. In Chap. 6, I said that Fig. 6.2 shows the probability that a neutron will have decayed after a certain length of time. According to that interpretation, if a neutron exists at some time then there is 50 % probability that it will have decayed after 15 minutes. A curve of the same shape applies to the decay of any particle, or nucleus.

Now, according to quantum physics the interpretation that I gave the curve is not actually correct. According to quantum physics, the curve only gives the probability that *a measurement will find that* the neutron has decayed. But it seems silly to insert the words in italics! Surely, the neutron would have decayed even if we had not made the measurement? The problem is, that according to quantum physics that is not the case. Instead, according to quantum physics, the neutron before the measurement was in what is called a 'superposition' of decayed and not-decayed states. It doesn't matter what 'superposition' means. What matters, is what it doesn't mean. It doesn't mean that the neutron had either decayed or not decayed and that the measurement will reveal which is the case.

The superposition of states is ubiquitous in quantum physics, and is certainly required for the correct description of very small things. But it doesn't seem to correspond with reality when applied to things of ordinary size. This was highlighted in 1935 by Erwin Schrödinger, one of the founders of quantum physics, who invited us to consider the following situation. A cat is put into a closed box, in which there is a substance containing atoms, whose nuclei decay with an extremely long lifetime. The lifetime is so long that there is only a 50 % probability that just *one* nucleus decays within say an hour. Within the box there is decay-detecting apparatus. If the apparatus finds that a decay has occurred, it opens a phial of poison that kills the cat.

According to quantum physics, the whole content of the box is in a superposition of decayed and not-decayed states, which means that the cat is in a superposition of alive and dead states. Now we open the box after an hour and of course find the cat either alive or dead. If it's alive we can be sure it's been alive all the time, and if its dead the time of its death could be determined by examining the body. Either way, we know that it was never in the superposition of alive and dead states that quantum physics requires.

With the generation of the curvature perturbation, we are in the same situation. According to quantum physics, the Universe is in a superposition of states. The curvature perturbation is different for each state, and each state corresponds to a Universe that might exist. But only one of the superposition corresponds to the Universe that actually does exist. This is ridiculous, because according to quantum physics the Universe will be in a definite only after a measurement is made. So a galaxy, for instance will not exist before it is observed. This is even more nonsensical than Schrödinger's cat!

The obvious solution, for the cat and the Universe, is to suppose that quantum physics isn't correct for big things. For the decay, to suppose that the decay actually occurs when it is detected by the apparatus, and for the curvature perturbation to suppose that it acquires a definite value before it has any observable effect. That solution is not as straightforward as it looks and is not popular. At the moment, the usual attitude among physicists is that quantum physics is absolutely correct and it's our job to interpret it. Unfortunately, there is no consensus about how to do that.

I am assuming that there are no particles during the second stage of inflation. A scenario there are particles has been proposed, called Warm Inflation. The particles are being generated continuously and are supposed to be in thermal equilibrium. Their contribution to the energy will be determined by thermal equilibrium and it will vary with position and time. This 'thermal fluctuation' could be responsible for the generation of ζ, instead of the quantum fluctuation of a scalar field.

Historical note In 1982, several papers proposed that the primordial energy density perturbation originates from the inflaton field perturbation. In 2002 it was proposed instead that the it originates from the perturbation of a field different from the inflaton, which was called the curvaton field. The proposal was made in a paper by T. Moroi, and in another paper by the present author and D. Wands. Several similar proposals followed.

Chapter 16
Prehistory of the Big Bang

> It is shown that at a temperature higher than about 10^2 GeV the symmetry broken at lower temperature is re-established. . . .
> D. A. Kirzhnits and A. D. Linde in the journal Physics Letters B (1972).

It is natural to ask what happened, between the end of inflation and the beginning of the known history. To answer that question, we have to make assumptions which may or may not be correct. Most importantly, we have to make an assumption about the energy density at the end of inflation. As we saw in Chap. 14, the energy density at the end of inflation might be anything up to 10^{64} times as big as at the beginning of the known history.

When inflation ends, the inflaton field begins to oscillate corresponding to the presence of inflaton particles. This marks the beginning of the Big Bang. As the inflaton field is nearly homogenous, the particles have negligible random motion. In the terminology of Chap. 8, we are dealing with a matter-dominated era.[1] The matter-dominated era ends when the matter particles decay, to produce a radiation-dominated Universe.[2] The radiation particles presumably undergo collision processes, which quickly establish thermal equilibrium at a temperature that is determined by the energy density. This marks the beginning of what is called the Hot Big Bang.

To discuss the Hot Big Bang, we have to make an assumption about the temperature when it begins. Equivalently, we have to make an assumption about kT, the temperature times the Boltzmann constant. As we saw in Chap. 9, it is kT

[1]If inflation involves more than one field, more than one field might be oscillating corresponding to the presence of more than one particle species. They would still constitute matter.

[2]To be precise, that is when matter-domination finally ends. Before the particles decay, the oscillating field can lose energy by a process called preheating, which involves the field itself and has nothing to do with particles. The particles created by preheating can constitute radiation, but the density of such radiation will fall more quickly than the energy density of the oscillating field and the latter is expected to dominate by the time that the inflaton particles decay.

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_16

that determines the properties of a cosmic gas. In particular, the average energy of the photons is $2.7kT$, and the average kinetic energies of the other species are about the same.

If the Hot Big Bang begins just before the known history, kT will only be a few MeV as shown in Table 7.1. But as we saw in Chap. 14, it might instead begin when the energy density is up to 10^{64} times bigger. If the energy density is that high, kT will be about 10^{16} MeV.[3]

Prehistory According to the Standard Model

I will first describe the prehistory assuming that the Standard Model accounts for everything, even though we know that this cannot be completely correct. Let us suppose first that it begins when kT is bigger than 10^5 MeV.[4]

In this case, the collision processes occur with enough kinetic energy to create all of the elementary particles described by the Standard Model.[5] There are no hadrons, nuclei or atoms because if any of these objects managed to form they would quickly be blown apart by collisions.

The state of thermal equilibrium is determined by the temperature, and the densities of the three conserved quantities that we encountered in Chap. 9. These are electric charge, baryon number and lepton. The amount of each of these carried by each particle is shown in Table A.6 of Appendix B.

As the amount of each conserved quantity within an expanding region is constant, its density is proportional to $1/a^3$. The density of each conserved quantity is therefore determined by its density at the beginning of the known history. This is zero for the charge and is assumed to be zero for the lepton number, but it is nonzero and positive for the baryon number. The positive baryon number implies that there are slightly more quarks than anti-quarks.

As the Universe expands, it cools and the distance between the particles increases. To understand what happens next, one needs to know that a quark or gluon can exist, only if there are other quarks and gluons within a distance of about 10^{-15} metres. This ceases to be true when kT falls to about 100 MeV. The quarks and gluons are then replaced by hadrons, which are made out of quarks. Of the hadrons, the proton is stable, and the neutron has a longish half-life (15 minutes).

[3]As we saw in Chap. 8, kT is proportional to $1/a$ while the energy density during radiation domination is proportional to $1/a^4$.

[4]Using the GeV unit, that is bigger than 10^2 GeV which means that I am describing the scenario first considered by Kirzhnits and Linde in the paper cited at the beginning of this chapter.

[5]The quarks and leptons actually have zero mass when kT is bigger than 10^5 MeV, because high temperature changes the Higgs potential, so that its minimum value corresponds to zero Higgs field. When kT falls below 10^5 MeV the Higgs potential reverts to its usual form and the quarks and leptons acquire their mass.

But all other hadrons have very short lifetimes; they quickly decay, and are not re-created because their rest energies are bigger than the kinetic energies of the photons and leptons that might have done the job. At about the same time, the Higgs bosons decay for the same reason, as do Z and W bosons and the μ and τ leptons listed in Table A.4. We are left with just the particle species that are present at the beginning of the known history of the Universe.

If kT is less than 100 MeV at the start of the Hot Big Bang, the history is the same except that there are never any isolated quarks and gluons, only hadrons.

Creating the CDM

The prehistory that I have given assumes that there are more u and d quarks than anti-u and anti-d quarks. If the numbers were equal, the quarks would produce equal numbers of protons and anti-protons (and equal numbers of neutrons and anti-neutrons) which would annihilate. As a result, the Universe would contain no ordinary matter. It may be that the excess of quarks over anti-quarks is produced by the decay of the inflaton, in which case the excess would not require any modification of the prehistory. More usually, it is supposed that the excess is created later through some modification of the prehistory. To keep things simple though, I won't discuss the latter possibility, since it is after all optional.

What I do need to discuss, is the creation of CDM. That's compulsory because the Standard Model contains no species that could be the CDM. In contrast with the situation for the inflaton, there *are* extensions of the Standard Model that include a CDM candidate, and yet are motivated by considerations that have nothing to do with CDM. There are two such extensions and I discuss them in turn.

According to one extension, the CDM is a particle species called the axion, which corresponds to the oscillation of a scalar field called the axion field. This extension is motivated by the fact that, according to the Standard Model, the neutron should be surrounded by an electric field even though it is on average electrically neutral.[6] The strength of the electric field is determined by a parameter of the Standard Model, and the fact that it has not been observed requires the parameter to be less than 10^{-12}. The axion extension of the Standard Model explains why the parameter is so small.

There are two possible scenarios for producing axion CDM. In one scenario, the axion field is homogeneous during inflation, except for a small perturbation generated from the vacuum fluctuation. After inflation, the axion field oscillates around its vacuum value corresponding to the presence of axions. In this scenario the adiabatic condition is not exactly satisfied, because the axion field has a perturbation on cosmological scales that is different from the inflaton field's perturbation, but

[6]The field arises because, according to the Standard Model, there is a small positive charge within one hemisphere of the neutron, and an opposite negative charge within the opposite hemisphere.

the departure from the adiabatic condition can be too small to observe. In the other scenario, axion field appears only after inflation, and the adiabatic condition is almost exactly satisfied on cosmological scales. On very small scales though, the CDM density perturbation is much bigger than the one corresponding to the adiabatic condition. This scenario predicts the existence of very small pure CDM halos. They have not been observed, but that may be because they have merged like the bigger pure CDM halos mentioned in Chap. 12. Both scenarios can therefore be in accordance with observation.

The interaction of axions with Standard Model particles is rather weak, in accordance with the fact that the axions are CDM, but it exists and it may allow the detection of axion CDM. Also, the interaction means that axions are emitted by the Sun whether or not they are the CDM. Experiments are underway to detect both CDM axions and solar axions.

So much for the axion extension of the Standard Model. The other extension, invoking what is called supersymmetry, is motivated by what is called the hierarchy problem. The hierarchy problem is the fact that the Standard Model gives a quantum contribution to the mass of the Higgs boson that is far bigger than the observed quantity. That is similar to the situation for the cosmological constant depicted in Fig. 8.1 and it has the same implication; there must be another 'classical' contribution to the Higgs mass, which almost exactly cancels the quantum contribution. The supersymmetry scenario seeks to avoid the hierarchy problem, by giving each Standard Model species a 'superpartner'. The quantum contribution to the Higgs boson mass then becomes about the same as the mass of a typical superpartner. The hierarchy problem would be completely removed if the superpartners were only about as heavy as the Higgs boson. That is ruled out by the fact that superpartners have not been produced at the LHC. There remains the possibility though, that the superpartners exist and are only say ten or a hundred times heavier than the Higgs boson. Then the difference between the classical and quantum contributions would be 10 % or 1 % which might not be regarded as a problem.

With supersymmetry, the prehistory of the Hot Big Bang is the same until after the electroweak transition, except for the presence of the superpartners. After the electroweak transition though, the superpartners decay except for the lightest superpartner which is stable and is the CDM candidate.

The generic name for *any* CDM particle with properties similar to the lightest superpartner is Weakly Interacting Massive Particle, abbreviated to WIMP. As with the axion, experiments are underway to detect WIMP CDM through its interaction with Standard Model particles. The interaction might allow the direct detection of WIMPs, or their indirect detection through electromagnetic radiation emitted when WIMPs in our galaxy collide.[7] It might also allow WIMPs to be produced at the LHC. If the WIMPs are light, they might have significant random motion corresponding to what is called Warm Dark Matter. That might be detected by

[7]Radiation is observed coming from the centre of our galaxy, for which WIMP collisions provide one of the most plausible explanations—but unfortunately not the only one.

comparing the structure of galaxies with the outcome of computer simulations of their formation.

Is that All?

There might be other species in addition to the ones that I have described. One possibility is to have a species that is abundant at early times and decays around the time that light nuclei form. As was mentioned in Chap. 7, that could affect the ^{7}Li abundance.

There are also much wider possibilities. To take an extreme example, there are plausible extensions of the Standard Model that would allow a ‘mirror world’ consisting of particle species exactly like ours with the same interactions. Then there would be mirror-atoms made out of mirror-protons, neutrons and electrons, a mirror-photon, a mirror electromagnetic field and so on. Even mirror-planets and mirror-people! The interaction of the mirror world with Standard Model particles could be too tiny to have any significant effect, and if there isn’t too much mirror stuff it wouldn’t have a significant gravitational effect either. Such things are interesting to contemplate, but we’re never going to know whether they exist or not.

What Might Future Observations Find?

I’ve now finished my account of cosmology, as it stands today. It remains to ask what the future may hold. The most dramatic cosmological discovery would be to detect the gravitational wave background. That would be a very strong indication that inflation occurred, because the gravitational wave background includes wavelengths not much less than the size of the observable Universe that could hardly have been generated by any other mechanism. Moreover, it would tell us that inflation occurred at a very early stage, leaving plenty of time for things to happen between the end of inflation and the beginning of the known history.

A less dramatic discovery would be a departure from the adiabatic condition, which relates the primordial energy density perturbations to the primordial curvature perturbation. As we saw in Chap. 16, such a departure is possible if, for example, the CDM consists of axions. Unfortunately though, there is no particular reason to think that the departure will be at a level that will be observable in the near future. A discovery that the dark energy varies with time would also be very interesting. It would mean that it it doesn’t come from the vacuum but instead, presumably, from some ‘quintessence’ field.

There is the possibility of finding out more about the CDM as mentioned earlier. It could be that the CDM comes from two or more particle species. Conceivably, it could be that the CDM doesn’t consist of particles at all, but of sizeable objects.

These couldn't be made out of ordinary matter. They might be black holes formed before the beginning of the known history though the mass of such black holes must lie within a very narrow range.

Cosmic strings; an idea that didn't work

In this book I have focussed on ideas about the early Universe that are known to be correct, or at least are not ruled out by observation. Of course there have been plenty of other ideas, that seemed promising at first but were later ruled out by observation. Here is the story of a famous example.

It starts in the late 1970s, when an extension of the Standard Model called Grand Unified Theory (GUT) was proposed. Around 1980, it was realised that such an extension would allow the creation of what are called cosmic strings. A cosmic string is a concentration of energy within a tube, which is so narrow so that it can for most purposes be regarded as a line. The energy within the string comes from fields that are far from their vacuum values.

The GUT cosmic strings would form a network, consisting of both long lines and small loops. It was proposed that galaxies formed, not in regions where the cosmic gas was denser than average, but instead around cosmic string loops. That idea seemed reasonable because a loop attracts matter towards it by the force of gravity, just like an over-dense region of the gas. Unfortunately, it failed to account for observation (in particular for the CMB anisotropy) and was abandoned in the early 1990s.

There remained the possibility that the network of GUT cosmic strings could exist. If so, a pair of strings would occasionally intersect and it was shown that this would lead to the emission of a gravitational waves. It was also shown that it would cause a variation in the observed period of pulsars, whose magnitude was calculated. The failure to detect a variation of the required magnitude in the 2000s ruled out the GUT cosmic strings.

It is still possible that a network of cosmic strings exist, similar to the GUT strings but with a lower mass per unit length. In particular, there could be a network consisting of the strings of string theory. Such a network might be detected in the future, but it would have nothing to with galaxy formation.

Chapter 17
What About Us?

> We must therefore not be discouraged by the difficulty of interpreting life by the ordinary laws of physics. . . .. We must also be prepared to find a new type of physical law prevailing in it.
> (Erwin Schrödinger in 'What is life' (1944).)

Physics explains how a rather ordinary sequence of events led to the present Universe. Wouldn't you agree, now that you've read the book? If you would, you are forgetting about ourselves. We are part of the Universe but we somehow don't fit into the general scheme of things. How has it come about, that a few thousand of the species of ape that we call human have managed to figure out how their Universe began? Has the same thing been done elsewhere? If so, by what kind of creature?

Leaving aside these probably for ever unanswerable questions, one might hope that physics will some day describe living things. At first sight, that looks promising. The first step was taken in 1859 by Charles Darwin, who declared that the species are not immutable. One species can change into another and in particular humans are descended from apes. The idea was greeted with outrage at the time. Now, except by a lunatic fringe, it is regarded as obvious. How else could the diversity of species have come about?

How did life itself arise? The issue wasn't mentioned by Darwin in any publication. I suppose he felt that he was in enough trouble already. But we know from his correspondence that he was aware of the obvious answer. Living things evolved from non-living molecules. This too is now accepted as obvious by practically everybody. How else could life have originated?

The life sciences have yet to identify for sure the sequence of events leading from molecules to life, though recent research is pointing to a plausible scenario. They have though, established the the molecular structure of the basic components of each living thing, and how those components are put together.

So it looks as if physics ought to be able to describe life. There is a fundamental problem though, which is the extreme complexity of even the simplest living thing. It's one thing to understand the basic components, but quite another to understand

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3_17

how they are put together. That problem will never go away. A more particular problem, at least for now, is that we don't have a theory that would do the job. The quantum physics that seems to work for molecules and other small collections of particles, seems not to work for large objects. Even if the basic equations are accepted as correct, their interpretation in terms of reality is quite obscure.

If we do manage to get a physics description of a living thing, its extreme complexity will surely require concepts of which we at present have no idea. That will be analogous to the situation of statistical mechanics, that we encountered in Chap. 9. It involves concepts like temperature, that don't apply to individual collisions within a gas, but they do apply to the gas as a whole.

After the physics description of a living thing, there remains its consciousness. I don't think the physics discussion has much to say about that, except for one thing. That's the issue of whether your consciousness can, by itself, tell you anything about the outside world. It certainly tells you about the experiences of other people, and to some extent about the experiences of other living things, because they are similar to yourself. But does it tell you anything else? People through the ages have thought so, but there's no basis in physics for that belief. We are made of atoms and molecules, which have significant communication with the outside world only through the limited means that we call the five senses. Only by using those can we, according to physics, learn anything about the outside world.

Appendix A
More Detail with Some Mathematics

Chapter 3 The Fundamental Theories

Gravity

According to both Newton and Einstein, gravity at a given location gives every object the same acceleration no matter how light or heavy it is. According to Newton, the acceleration caused by an object with mass M at a distance r from the object is GM/r^2. The acceleration due to a one-kilogram weight at a distance of one metre is therefore 6.67×10^{-11} metres per second per second. That's not much acceleration, but it can be measured in an experiment that physics students sometimes do in the laboratory.

The acceleration of gravity at the Earth's surface is much stronger. It comes from every piece of the Earth, but Newton showed that it's the same as if the entire mass of the Earth were located at its centre. The mass of the Earth is 5.97×10^{24} kg and its radius is about 6.4×10^6 metres, which gives 9.9 metres per second per second for the acceleration.

By the 'weight' of an object, we mean the force that gravity exerts on the object on the surface of the Earth. According to Newton, this force is equal to the mass of the object times the acceleration of gravity. When we specify the weight of an object in, say, kilograms we are implicitly choosing the distance and time units so that the acceleration of gravity is equal to 1.

The full theory of gravity is provided by Einstein's General Relativity. If gravity is weak we can describe gravity by introducing a gravitational field. If the gravitational field is slowly varying, and we only consider the motion of objects moving with speed much less than c, we arrive at Newton's theory of gravity in which the gravitational field is similar to the electric field, being specified by its strength and direction.

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3

The Principle of Relativity

Special and General Relativity are so-called because they invoke the Principle of Relativity, that was first described by Galileo in 1632. (The other word, Special, means that gravity is being ignored in contrast with General Relativity which includes gravity.) The Principle of Relativity says that the laws of physics don't depend on the velocity of the observer that is invoking them.[1] As Galileo pointed out, that implies that it's meaningless to talk about *the* velocity of an object. We can only talk about the velocity of an object, as measured by some observer. We could talk about *the* velocity of an object, only if we could talk about *the* observer. But according to the Principle Relativity, all observers moving with constant velocity have the same status and none of them can be regarded as *the* observer.

Maxwell's discovery of Classical Electrodynamics raised a serious problem for the Principle of Relativity. The problem was that Classical Electrodynamics requires electromagnetic radiation (in particular light) to always move with the same speed c. But the Principle of Relativity requires that Classical Electrodynamics is valid for every observer (moving with constant velocity). Therefore, Classical Electrodynamics and the Principle of Relativity, taken together, require the speed of light to be independent of the observer.

On the basis of our everyday experience, that doesn't seem to make sense. Suppose I measure the speed of light passing by me, and someone also measures that speed in a car which is travelling at 50 kilometres per hour in the same direction as the light. We would expect the other person to find a speed which was 50 kilometres per hour smaller than the speed that I find. In fact, they would find the same speed.

Special and General Relativity provide new descriptions of space and time, that are different from the one suggested by everyday experience and which do make the speed of light the same for all observers. Everyday experience is misleading here, because it only involves speeds much less than c. If, instead of light, some object passes by me and I find its speed to be say 1000 kilometres per hour, then the person in the car *will* find to high accuracy that the object's speed is only 950 kilometres per hour. That's because 1000 kilometres per hour is a lot less than the speed of light.

By virtue of the Principle of Relativity, the Standard Model description of collision and decay processes is valid for any observer moving with constant velocity. I assumed that implicitly in Chap. 6, by choosing the observer of a decaying particle to be moving with the particle.

[1]The observer is not supposed to feel any acceleration. In the context of Special Relativity, which ignores gravity, we can simply say that the observer is not accelerating. Including gravity, we should say that the observer is in free fall, as defined in Chap. 7.

Chapter 8 Energy Density of the Universe

Here is why the energy of radiation within an expanding region of the Universe is is proportional to $1/a$. Pressure is force per unit area, and when a force pushes something a certain distance the energy expended is equal to the force times the distance. From this, it follows that when the volume of an expanding region increases by an infinitesimal amount dV, the energy changes by an amount $dE = -p\,dV$ where p is the pressure.[2]

The energy within a region with volume V is ρV where ρ is the energy density. As we are dealing with radiation the constituents of the gas move with speeds close to c. The kinetic theory of gases shows that $p = \rho/3$ in that case. It follows that $dE/E = -\frac{1}{3}dV/V$. Since V is proportional to a^3, $dE/E = -da/a$. This implies that $E = b + d/a$ where b and d are constants, but $b = 0$ because $E = 0$ when $V = 0$ so indeed E is proportional to $1/a$. From that, it follows that the energy density of radiation is proportional to $1/a^4$.

This calculation assumes that the radiation does not significantly gain or lose energy to the matter. Also, it applies to each individual component of the radiation, if there is no significant exchange of energy between the components. Both of those things are true in the expanding Universe, except around the epoch of electron-positron annihilation.

Chapter 9 In the Beginning

The properties of a homogeneous gas are completely specified by giving the **distribution function** of each constituent. Figure A.1 explains how that works. The constituents of the early Universe are particles, and its properties are specified by giving the distribution function of each species. From the distribution function one can deduce the total energy density of the species, and its number density. We are also interested in the average energy of a particle, which is the energy density divided by the number density.

Figure A.1 shows the photon distribution function, that is established by thermal equilibrium during the first stage of the known history of the Universe. This is called the blackbody distribution.[3] Figure A.2 shows the same thing for the neutrinos, electrons and positrons. Table A.2 show the energy density, number density and average particle energy for these particles. All of them constitute radiation. The protons and neutrons constitute matter. The energy density of the radiation is

[2]This calculation ignores heat flow, which indeed is absent since we are considering the homogeneous Universe.

[3]The blackbody distribution applies whenever photons are being freely created and destroyed by processes in thermal equilibrium.

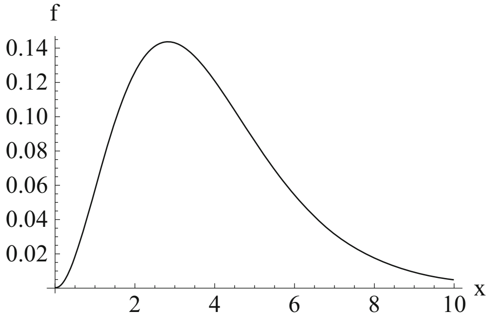


Fig. A.1 The blackbody distribution The distribution function of photons in thermal equilibrium is called the blackbody distribution. The vertical axis shows the blackbody distribution function in units of $(kT)^4/(\hbar c)^3$, which is denoted by f. The horizontal axis is the photon energy in units of kT, which is denoted by x. The energy density of those photons with x in a given infinitesimal range, is equal to f times the range

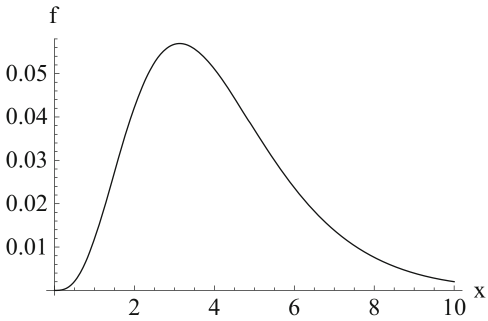


Fig. A.2 This gives the distribution function of each neutrino (and anti-neutrino) species. The notation is the same as in Fig. A.1 with now T the neutrino temperature. Before electron-positron annihilation, the electron and positron distribution functions have the same shape but but are twice as big

proportional to T^4. As we saw earlier, it is also proportional to $1/a^4$ which means that T is proportional to $1/a$.

These results hold until the first stage ends, at the epoch of electron-positron annihilation. Using the fact that entropy is conserved during thermal equilibrium, one can show that electron-positron annihilation boosts the photon temperature by a factor $(11/4)^{1/3}$, relative to the neutrino temperature which continues to be proportional to $1/a$. Table A.3 shows the resulting number and energy densities, for the photons and for the neutrinos. The photon energy density is 1.46 times the neutrino energy density, which is in agreement with the observed values shown in Table 8.1. There is now just one electron per proton, and the electrons are matter as opposed to radiation.

Instead of specifying the baryon number density during the first stage of the history, it's more convenient to specify the baryon number per photon. This is $4/11$ times its present value, because the baryon number density is proportional to $1/a^3$, but the photon number density is boosted by a factor $11/4$ at electron-positron annihilation. From observation, we know that the present baryon number per photon is 6.1×10^{-10}. During the first stage, the baryon number per photon is therefore 1.68×10^{-9}.

Chapter 10 Friedmann Tells the Time

I give here some consequences of the Friedmann equation, which is

$$\dot{a}^2 = \frac{8\pi G\rho a^2}{3c^2} + K$$

Critical density corresponds to $K = 0$. According to observation, $|K|$ is less than one percent of $\dot{a}^2$ at present, and much less at earlier times. I therefore set $K = 0$.

The Timescale

As well as the scale factor a and the expansion rate $\dot{a}$, it's useful to consider the Hubble parameter $H = \dot{a}/a$ and the Hubble time $1/H$. If gravity had no effect, $\dot{a}$ would be constant and the Hubble time would be the age of the Universe. We shall see that this remains approximately true during the Big Bang even with gravity taken into account.

The Friedmann equation gives

$$\frac{da}{dt} \propto a\rho^{1/2}(a)$$

and therefore

$$t(a) \propto \int_0^a \left[a\rho^{1/2}(a)\right]^{-1}$$

If $\rho \propto 1/a^4$ corresponding to radiation domination this gives $a \propto t^{1/2}$and $H = 1/2t$. If $\rho \propto 1/a^3$ corresponding to matter domination, it gives $a \propto t^{2/3}$ and $H = 2/3t$. After the cosmological constant has become significant, we can ignore the radiation. Keeping only the matter and cosmological contributions to ρ, one finds the result shown in Fig. A.3. The present Hubble time is 1.44×10^{10} years to be compared with the age 1.38×10^{10} years.

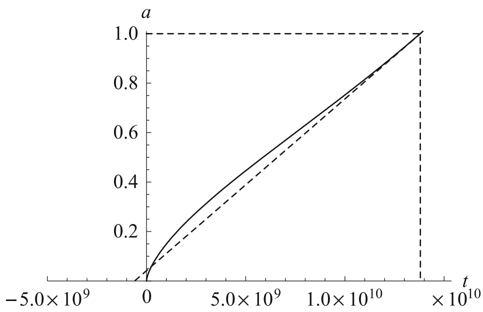


Fig. A.3 The scale factor with t in years. The *sloped dashed line* is what the scale factor would be, if $\dot{a}$ had its present value at all times. The age of the Universe would then be equal to the present Hubble time $1/H_0$ which is 1.44×10^{10} years. This is almost the same as actual age $t_0 = 1.38 \times 10^{10}$ years

The Friedmann equation also gives the timescale for the temperature during radiation domination. Before electron-positron annihilation, the energy density is given by Table A.2 with $A = 3.54$. Remembering that T is proportional to $1/a$, one finds with t in seconds and kT in MeV

$$t = \frac{1.39}{\sqrt{A}} \times \frac{1}{(kT)^2}$$

After electron-positron annihilation, the same expression holds with $A = 1.11$.

Distance to the Horizon

If we ignore the expansion of the Universe, the distance travelled by light in a Hubble time is c/H, which is called the Hubble distance. Since the Hubble time is about equal to the age of the Universe during the Big Bang, we may guess that the distance to the horizon is about equal to the Hubble distance. As we now see, that turns out to be correct.

To calculate the distance to the horizon, one uses comoving coordinates $\mathbf{x} \equiv \mathbf{r}/a(t)$, instead of the physical coordinates $\mathbf{r}$.[4] At any epoch during the Big Bang, the horizon is defined as the distance that a particle with speed c would have travelled since the Big Bang began. Suppose that a photon arrives at the origin $\mathbf{x} = 0$ at some time t, and was at a very short distance $a(t)dx$ a very short time dt earlier. Then $a(t)dx = cdt$ is a good approximation because a hardly changes during the time dt. As there choice of origin is arbitrary, this relation is valid without qualification.

[4]In this context, $\mathbf{r}$ can be taken to mean the three Cartesian coordinates (x, y, z). Comoving means 'moving with' (pronounced cohmoving), the idea being that a point with fixed $\mathbf{x}$ is moving with the expansion.

Going to infinitesimals we get an exact result, and integrating gives

$$x(t) = c\int_{t_{bb}}^{t} dt'/a(t') = c\int_{a_{bb}}^{a(t)} da/a^2H(a)$$
$$= \frac{c}{H(t)}\int_{a_{bb}}^{a(t)} \left[\frac{\rho(a(t))}{\rho(a)}\right]^{1/2} \frac{da}{a^2} \quad \text{(A.1)}$$

where bb denotes the start of the Big Bang and I have used $\rho \propto H^2$.

Let us first suppose that the Big Bang starts only at the beginning of the known history. During radiation domination when ρ is proportional to $1/a^4$, we can to a good approximation set a_{bb} to zero (except during the first few Hubble times) which gives $r(t) = c/H(t)$. During matter domination, we can in can ignore the radiation dominated era (except during the first few Hubble times) which gives to a good approximation which gives $r(t) = 2c/H(t)$. At the present epoch we need to keep both the cosmological constant and the matter, but we can still drop the radiation and set a_{bb} to zero. This gives $r_0 = 3.17c/H_0$.

These results are practically unchanged if the Big Bang started earlier, because during the Big Bang the Universe contains matter, radiation or a mixture of the two. Therefore, the integral converges rapidly as we make the start of the Big Bang earlier, giving indeed a negligible change to the result. Also, the result $r = c/H$ holds during any era of radiation domination and the result $r = 2c/H$ holds during any era of matter domination after a few Hubble times have elapsed. We see that in all cases, the distance to the horizon is at least roughly equal to the Hubble distance.

This explains why the distance to the horizon represents an ever-decreasing fraction of the distance to the edge of the observable Universe, as we go back in time. The former is approximately proportional to $1H$ while the latter is proportional to a. The former divided by the latter is proportional to $1/aH$, which is in turn equal to $1/\dot{a}$. But gravity is slowing down the expansion rate $\dot{a}$ which means that indeed $1/aH$ decreases as we go back in time.

It also explains why the distance to the horizon represents an ever-decreasing fraction of the smoothing radius, as was stated in Chap. 11. That is simply because the smoothing radius is proportional to a.

Chapter 11 Ripples in the Universe

To describe the perturbations we identify each point in the actual (perturbed) Universe with a point in the homogeneous (unperturbed) Universe. Then each point in the actual Universe is assigned a coordinate **x** in the homogeneous Universe. Since the homogeneous Universe can be obtained by smoothing on a scale much bigger than the observable Universe, this justifies the identification of the unperturbed energy density with the average energy density in the actual Universe.

Fourier Decomposition

It is useful to invoke the Fourier decomposition for each perturbation. At a given time, a perturbation $g(\mathbf{x})$ is written as[5]

$$g(\mathbf{x}) = \int d^3k\,[g_1 \sin(\mathbf{k}\cdot\mathbf{x}) + g_2 \cos(\mathbf{k}\cdot\mathbf{x})] \tag{A.2}$$

In this expression, $\mathbf{x}$ is the vector with components $(x/a, y/a, z/a)$. The magnitude of $\mathbf{k}$ is denoted by k and is called the comoving wavenumber. If we take the vector $\mathbf{k}$ to point along the x direction, $\mathbf{k}\cdot\mathbf{x} = kx$ and the each term looks like a wave with wavelength $2\pi/k$. In general though, we are not dealing with a wave because the amplitudes g_1 and g_2 are not in general oscillating.

To a good approximation, the amplitudes $g_1(\mathbf{k})$ and $g_2(\mathbf{k})$ for a given $\mathbf{k}$ evolve independently of each other, and independently of those quantities for different $\mathbf{k}$. To be precise, it we specify $g_1(\mathbf{k})$ and $\dot{g}_1(\mathbf{k})$ at some time for every perturbation but for a single $\mathbf{k}$, that determines $g_1(\mathbf{k})$ for every perturbation at future times, for the same $\mathbf{k}$. The same is true for $g_2(\mathbf{k})$. This is called the linear approximation.

The mean-square σ_g^2 of a perturbation can be written

$$\sigma_g^2 = \int_0^\infty \mathcal{P}_g(k)dk/k$$

The contribution to σ_g^2 from comoving wavenumbers in an interval dk is $\mathcal{P}_g(k)dk/k$, and $\mathcal{P}_g(k)$ is called the spectrum of g.

The comoving wavelength $2\pi/k$ is equal to the physical wavelength $2\pi a(t)/k$ at the present epoch. I will call $1/k$ the scale of the Fourier component. In Chap. 11 I defined a smoothed quantity. If the smoothing radius is denoted by $a(t)R$, smoothing is about the same as dropping Fourier components with $1/k < R$. Therefore, the 'smoothing scale' defined in Chap. 11 is about the same as the biggest surviving 'scale of the Fourier component' as defined here. We are interested only in the cosmological smoothing scales, shown in Table A.1.

Curvature Perturbation

In the perturbed Universe, Euclid's geometry is no longer valid which means that we can't choose Cartesian coordinates $\mathbf{x} = (x, y, z)$, such that the distance ds between

[5]Because the perturbations don't go to zero at large distances, one actually has to write a Fourier *series* within a box much bigger than the presently observable Universe, but I'll skip that complication. Assuming statistical homogeneity, we can take the box size to infinity for the purpose of evaluating the spectrum.

Table A.1 Cosmological scales The smoothing radius is proportional to a. The first column shows its present value, called the smoothing scale. The second column shows the value of the scale factor a when the smoothing radius becomes equal to the distance to the horizon. This happens during matter domination for entries below the line, and during radiation domination for entries above it. Last scattering occurs at the epoch $a = 10^{-3}$. The quantity ℓ refers to the CMB anisotropy and is defined in this Appendix. The fourth column gives the mass of the object formed when a region within the smoothing radius collapses, calculated in this Appendix

Smoothing scale (lightyears)	a when smoothing radius equals horizon radius	CMB ℓ	Mass of collapsed object
3×10^3	3×10^{-9}		$10^3 M_\odot$
3×10^4	3×10^{-8}		$10^6 M_\odot$
3×10^5	3×10^{-7}		$10^9 M_\odot$
3×10^6	3×10^{-6}		$10^{12} M_\odot$
3×10^7	3×10^{-5}	1400	$10^{15} M_\odot$
3×10^8	1×10^{-3}	63	
1.5×10^9	1×10^{-2}	14	
5×10^9	1×10^{-1}	2	

two points is given by $ds^2 = a^2(t)(dx^2 + dy^2 + dz^2)$. It can be shown though, that after smoothing with a smoothing radius much bigger than the horizon we *can* find coordinates such that[6]

$$ds^2 = a^2(x, y, z, t)(dx^2 + dy^2 + dz^2)$$

Also, we can choose the coordinates so that the energy density evaluated at fixed t is homogeneous, and so that an observer with fixed (x, y, z) moves with the expansion. Then, we can write for the locally defined scale factor

$$a(\mathbf{x}, t) = a(t)e^{\zeta(\mathbf{x},t)}$$

The quantity ζ is called the curvature perturbation. It is defined only on scales bigger than the horizon, but that includes all cosmological scales when the known history begins. At that era (which I will call the initial era) we are supposing that the adiabatic condition applies, according to which the pressure depends only on the energy density. As I now show, that makes ζ constant.

If we took the smoothing scale to be much bigger than the distance to the present horizon, we would arrive at the practically homogeneous Universe that was the subject of earlier chapters. But there's nothing special about the present—the Universe doesn't know we're here! If we go back to an earlier epoch, the smoothing

[6]To be precise, that is possible if we ignore the gravitational radiation, which at some level is generated by inflation. That can be included by inserting an extra factor, which distorts the shape of an expanding region while leaving its volume unchanged.

radius can be bigger than the distance to the horizon even if it is not so at present.[7] Then, the Universe at each location looks like some homogeneous Universe. An observer at that epoch could discover that it was actually inhomogeneous, only by moving more than a Hubble distance from their current location. As a result, *the smoothed energy density evolves at each location, just as it would in some homogeneous Universe*. In effect, the whole Universe becomes a collection of separate homogeneous Universes.

I will choose the spatial average of ζ to be zero, which is always possible because a spatially homogeneous ζ could be regarded as belonging to $a(t)$. According to the adiabatic condition, the pressure p depends only on the total energy density ρ. Therefore, since ρ is homogeneous at fixed t, so is p. Using the relation $dE = -pdV$ and $E \propto a^3\rho$, we have $V\dot{\rho} = -\dot{V}(\rho + p)$ which is equivalent to

$$\dot{\rho}(t) = -3\left[\frac{\dot{a}(t)}{a(t)} + \dot{\zeta}(\mathbf{x}, t)\right](\rho(t) + p(t))$$

This makes the rate of change of ζ independent of position, which means that it actually vanishes because the spatial average of ζ is zero.

The spectrum of ζ is determined by requiring that the spectrum of the CMB anisotropy agrees with observation.[8] The spectrum is almost independent of k, and observation is consistent with the form

$$\mathcal{P}_\zeta(k) = A(k/k_*)^{n-1}$$

with $k_*^{-1} = 6.5 \times 10^7$ lightyears, $A^{1/2} = 6.9 \times 10^{-5}$ and $n - 1 = -0.04$. The number n is called the spectral index.

One generally assumes that there is no correlation between the amplitudes of different Fourier components of ζ. In other words, that a measurement of one amplitude tells us nothing about the probable outcome of a measurement of a different amplitude. This property is called Gaussianity and it is consistent with present observation.

The Density Contrasts

Now let us see how ζ is related to the density contrasts. To do that, we use the approximation $e^\zeta = 1 + \zeta$, which implies that $a(\mathbf{x}, t) = a(t)(1 + \zeta)$. With that approximation, $\zeta = \delta a/a$.

[7] That is possible, because the scale factor decreases more slowly than the distance to the horizon as we go back in time as we saw at the end of Chap. 10.

[8] The galaxy distribution is also used, but the CMB alone gives almost the same result.

Let us denote the energy density of a matter component (CDM or ordinary matter) by $\rho_{\rm m}$. According to the adiabatic condition it is homogeneous at fixed t, because ρ is. We can choose a new time coordinate though, $t'(x, y, z, t)$, and then the energy density will not be homogeneous. So that we deal with perturbations, we assume that t' is almost the same as t so that

$$t'(x, y, z, t) = t + \delta t(x, y, z, t)$$

The change in time coordinate will change the locally defined scale factor by some amount δa. Using $\rho_{\rm m} \propto a^{-3}$ for the matter density contrast this gives

$$\frac{\delta\rho_{\rm m}}{\rho_{\rm m}} = \frac{d\rho_{\rm m}/da}{\rho_{\rm m}}\delta a = -3\frac{\delta a}{a} = -3\zeta$$

For a radiation component, the energy density is proportional to a^{-4} and the density contrast is -4ζ. These results hold for each Fourier component, well before the wavelength enters the horizon.

Acoustic Oscillation

The acoustic oscillation is an oscillation in the amplitude of each Fourier component.The amplitude of the acoustic oscillation decreases with time, because the photons diffuse out of regions of high density into regions of lower density. The decrease is called Silk damping after Joseph Silk who first pointed to its existence. The frequency of the acoustic oscillation is approximately equal to c divided by its wavelength. The acoustic oscillation ends at the epoch of last scattering.

The spectrum of the photon density contrast at the epoch of last scattering is shown in Fig. A.4. It exhibits a series of peaks and troughs, caused by the acoustic oscillation. For the wavelength corresponding to the first trough, there has been time for one quarter of an oscillation; for the wavelength corresponding to the first peak there has been time for half an oscillation, and so on.

The CMB Anisotropy

According to the linear approximation, which is more than enough to handle current observation, the CMB anisotropy corresponds simply to a shift in the temperature of the blackbody radiation. I will denote the shift by ΔT, and the average temperature over the sky by T. According to observation, $T = 2.73\,^\circ$K.

To extract most of the information from the CMB anisotropy, it is enough to focus on a small patch of sky. Within such a patch, we can define (almost) Cartesian coordinates, corresponding to a grid of ‘straight’ lines which is (almost) rectangu-

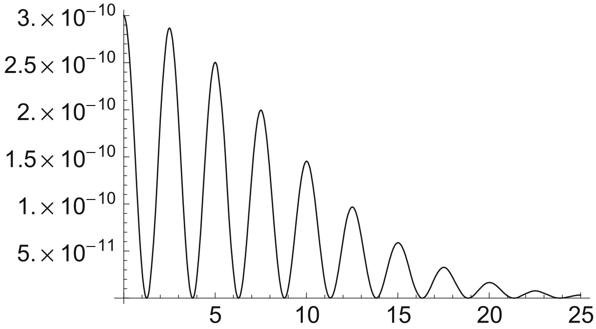


Fig. A.4 Spectrum of the photon density contrast This plot is an approximate representation of the spectrum of the photon density contrast at the epoch of last scattering. (The spectrum of the ordinary matter density contrast has the same shape but is a factor $(3/4)^2$ smaller.) The number k specifies the wavelength, as in Table A.1. The horizontal axis is $k/a_{\rm ls}H_{\rm ls}$, where $a_{\rm ls}$ is a at last scattering and similarly for $H_{\rm ls}$. For $k/a_{\rm ls}H_{\rm ls}$ much less than 1, the wavelength at last scattering is much bigger than the Hubble distance. Then, by virtue of the adiabatic condition, the spectrum is approximately $1/16$ times the spectrum of ζ. The troughs and peaks are caused by the acoustic oscillation. The decrease in the height of the peaks is caused by Silk damping

lar.[9] Then I can define Fourier components of the anisotropy. The wavenumber of a Fourier component, measured in radians, is denoted by ℓ. The spectrum of the CMB anisotropy is shown in Fig. A.5, it exhibits a series of peaks, with decreasing amplitude. Since the CMB is a (distorted) snapshot of the photon density contrast at last scattering, one might hope that the peaks in Fig. A.5 are related to the peaks in Fig. A.4. There is indeed a relation, but it's not very close and I won't go into it.

To describe the CMB anisotropy precisely, one must use the spherical harmonics to write $\Delta T = \sum a_{\ell m} Y_{\ell m}(\mathbf{e})$ where $\mathbf{e}$ is the direction in the sky. The spectrum of the CMB is then defined as the average of $(a_{\ell m})^2$ over m at fixed ℓ. For $\ell \gg 1$ this becomes the same as the spectrum defined earlier.

To have a prediction that is exact within the linear approximation, one must do two things. First, one must take account of the motion of photon gas at the epoch of last scattering, and also the effect on the CMB temperature of the variations in mass density that are encountered on its journey towards us. Second, the implicit assumption that last-scattering takes place instantly must be abandoned. Instead of assuming that all of the CMB photons come from a single sphere around us, one must accept that some of them were emitted earlier than others.

The line in Fig. A.5 is the calculated spectrum, using the cosmological parameters shown in Table A.8. The data points are observed values, which except for for very low ℓ have been averaged over a range of ℓ (binned). The prediction shown is averaged over the position of the observer. One can also calculate the mean-square

[9] I define a 'straight' line in the sky as one which doesn't change direction. It is the equivalent in the sky, of a Great Circle on the Earth which represents the shortest distance travelled on Earth between adjacent points.

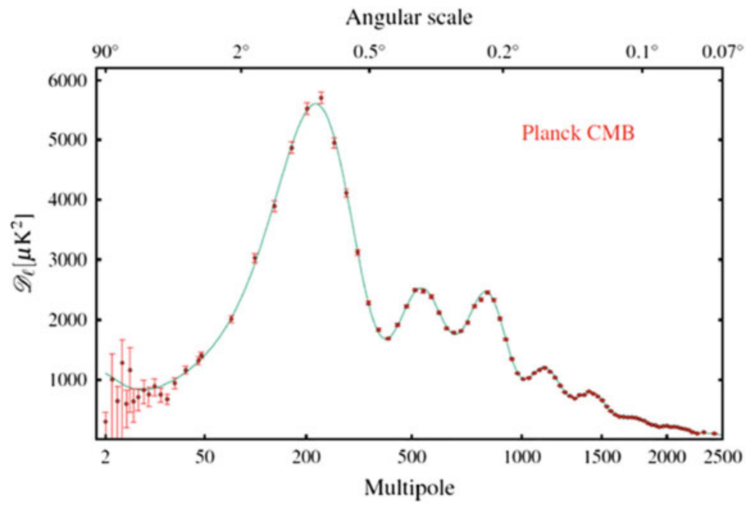


Fig. A.5 The spectrum of the CMB anisotropy The vertical axis is the spectrum of ΔT, in units of $(10^{-6}\ {}^\circ K)^2$. The lower horizontal axis gives the ℓ defined in the text. The upper horizontal axis gives the corresponding angle in the sky (Image: ESA [www.esa.int])

variation with position, which is called cosmic variance. This is included in the error bars, and dominates them at low ℓ.

Chapter 12 Formation of Galaxies and Clusters

Estimating the Mass

Let us see why the over- and under-dense regions of the CDM are roughly as big as the smoothing scale. The smoothed density contrast includes only Fourier components whose wavelength is bigger than the smoothing radius. Of these, the dominant ones have wavelength not *much* bigger than the smoothing radius, because they have spend the most time with wavelength less than the Hubble distance, allowing them to grow the most. Therefore, the dominant wavelengths of the smoothed density contrast are about equal to the smoothing radius which means that indeed the over- and under-dense regions are typically of that size.

Now let us see how to estimate the mass M of a galaxy or galaxy cluster, that results from the collapse of an over-dense region of the density contrast smoothed on a scale R. A region collapses if the smoothed density contrast becomes roughly equal to 1. Before that, the region is expanding with the Universe and its matter density is almost the same as the unperturbed matter density ρ_m. The size of the region is typically about the same as the smoothing radius aR. This gives the estimate

$$M = \frac{4}{3}\pi R^3 a^3 \rho_{\rm m}$$

Since $\rho_{\rm m}$ is proportional to a^{-3}, the quantity $a^3\rho_{\rm m}$ is independent of a and is equal to the present average matter density of the Universe. That is known from observation as shown in Table 8.1. Taking the unit of R to be the lightyear, and the unit of mass to be the Sun's mass, one finds $M = 6 \times 10^{-8} R^3$.

Neutrino Density Contrast

In this chapter I have been ignoring the neutrino density contrast, but it's easily dealt with. When the wavelength of a Fourier component becomes less than the Hubble distance, the neutrino density contrast decreases. That is because the neutrinos are moving freely with speed c, which causes a net flow of neutrinos out of regions with positive density contrast (where there is an excess of neutrinos) into regions with negative density contrast (where there are fewer neutrinos). That could not have happened earlier, because the Hubble distance at any epoch is about the same as the horizon, and the horizon is defined as the biggest distance that a photon could have travelled since the beginning of the Big Bang. The decrease in the neutrino density contrast must be included in the exact calculation of the CMB anisotropy.

Chapter 13 The Higgs Field

Planck Scale

In the text, I noticed that a measurement of the Higgs field strength will not yield precisely the vacuum value. Similarly, if we measure the average energy of the Higgs field within some region, we won't get precisely zero. The typical energy that we get, increases as the size of the region is decreased. If we make the region small enough, the typical value becomes big enough to make the region a black hole. The size needed for that to happen is approximately equal to the distance $\sqrt{\hbar G/c^3}$ which is 1.62×10^{-35} metres. That's called the Planck length and it's the only combination of the fundamental constants that corresponds to a distance. According to quantum field theory then, it makes no sense to talk about the energy within a region that is smaller than the Planck length. This places two important restrictions on the validity of quantum field theory.

The first restriction concerns the collision of two particles. According to quantum field theory, the typical distance probed by a collision becomes smaller as the energy is increased. That distance becomes the Planck length when the energy is about

$\sqrt{\hbar c^5/8\pi G}$, which is 2.43×10^{21} MeV. That's called the Planck energy and it's the only combination of the fundamental constants that corresponds to an energy.[10] Quantum field theory cannot describe the collision of particles with much bigger energy, which is a pity because some cosmic ray particles do have much bigger energy and they do occasionally collide. Nobody knows what happens when such a collision occurs. (The energies of the collisions at LHC are only about 10^6 MeV by the way and it would be completely impossible to build a collider with the Planck energy.)

The second restriction concerns the phenomenon of black hole evaporation. Assuming that the size of the a black hole is much bigger than the Planck length, one can use quantum field theory to demonstrate that the black hole would evaporate by emitting particles.[11] The evaporation rate is far too small to be of interest for known black holes. But tiny black holes might be formed in the early Universe, whose evaporation might be observable. As the evaporation proceeds, the black hole becomes smaller and eventually it becomes as small as the Planck length. Then the quantum field theory calculation no longer applies, and nobody knows what happens next. Some people thing that it will continue to evaporate till there's nothing left and others thing that it will stop evaporating, but nobody knows.

In this account of the vacuum fluctuation, I focussed on the average Higgs field within a given region. Instead, one can focus on the Fourier components of the Higgs field. If the amplitude of a Fourier component is re-measured after a time much less than the wavelength divided by c we will find almost the same result. If instead we re-measure it after a longer time, we get a new result which is unpredictable.[12]

Chapter 14 Inflation

As was noted in the text, the lack of observed gravitational radiation means that the energy density during the second stage of inflation cannot be more than about 10^{64} times as big as at the beginning of the known history. Using the Friedman equation, this means that the Hubble parameter H cannot exceed 10^{36} seconds^{-1}. The corresponding evolution of the Hubble parameter and the expansion rate are shown in Fig. A.6. Going the other way, the Hubble parameter during inflation must be bigger than the one at the beginning of the known history, corresponding to Fig. A.7.

To describe inflation in detail it is convenient to use what are called Natural Units. The energy unit is the MeV, and the units of time and length are chosen to make $\hbar$ and c equal to 1. The relation $E = hf$ determines the time unit to be

[10]The factor 8π is inserted according to a common though not universal convention.

[11]This was done in 1974 by Steven Hawking, building on previous work.

[12]This statement ignores the mass of the Higgs boson. It true to good accuracy if the wavelength divided by c is much less than the period of the homogeneous oscillation of the Higgs field.

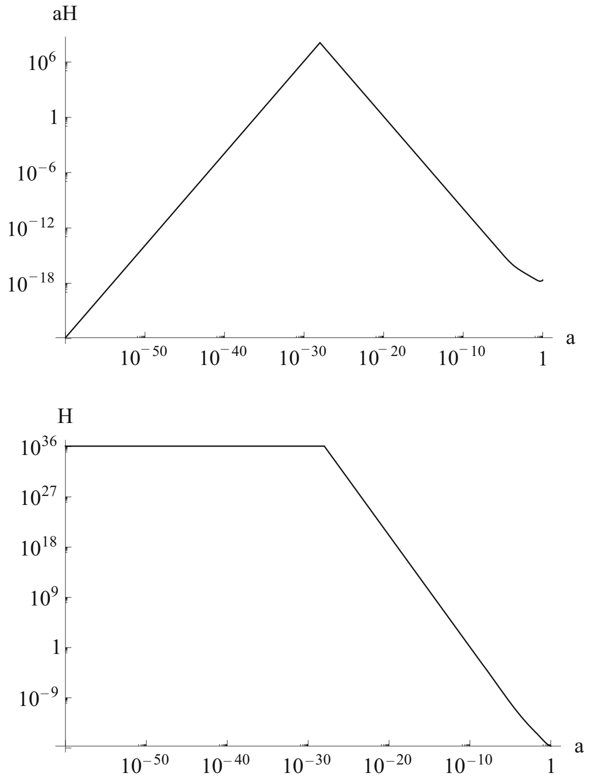


Fig. A.6 The earliest possible inflation era. The *first plot* shows the expansion rate *aH* for that case, while the *second plot* shows the Hubble parameter *H*, with the second as the unit of time. An earlier era would make *H* even bigger, generating gravitational waves at a level that would already have been observed

6.583×10^{-22} seconds, and the requirement $c = 1$ fixes the unit of length to be 1.973×10^{-13} metres. I will denote the Planck energy by $M_{\rm P}$. With the choice of Natural Units, the Planck energy is also called the Planck scale because it determines the Planck distance. During inflation, $\ddot{a} > 0$ which is equivalent to $|\dot{H}| < H^2$. To generate scalar field perturbations with the required nearly flat spectrum, we need $|\dot{H}| \ll H^2$ during the second era of inflation. The constant K in the Friedmann equation is then negligible so that $3M_{\rm P}^2H^2 = \rho$. In the expanding Universe a scalar field ϕ with potential $V(\phi)$ evolves according to the equation $-\ddot{\phi} = 3H\dot{\phi} + V'(\phi)$. To produce the required perturbations, we require the terms on the right hand side to be almost equal and opposite. Consistency with $|\dot{H}| \ll H^2$ then requires $\epsilon \ll 1$ where $\epsilon = M_{\rm P}^2(V'/V)^2/2$. The first derivative of that condition requires $|\eta| \ll 1$ where $\eta = M_{\rm P}^2 V''/V$. This setup is called slow roll. When it fails to apply, inflation ends and the inflaton potential oscillates around its vacuum value. In Fig. 14.1, the vacuum value is much smaller than the Planck scale, and inflation is ended by the failure of the condition on V''.

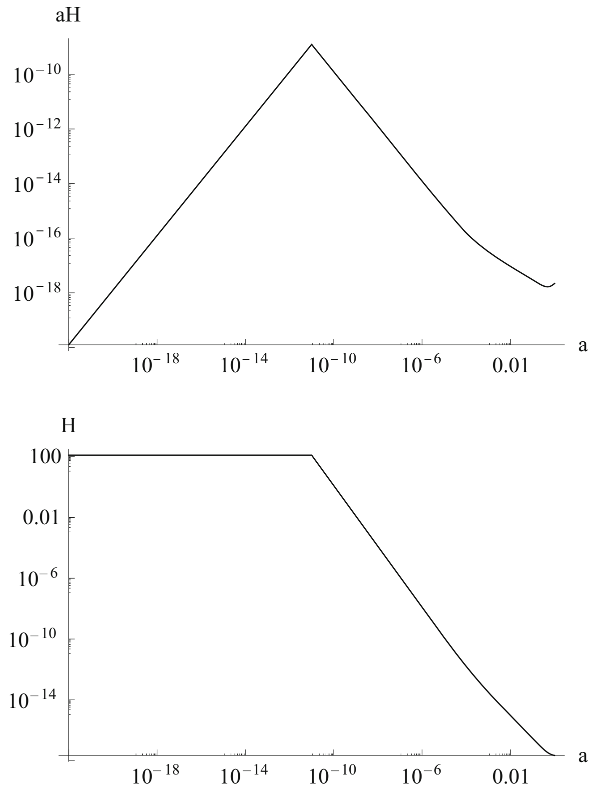


Fig. A.7 The latest possible inflation era is the one that would end just before the beginning of the known history of the Universe began

Chapter 15 Generating the Perturbations

Generating ζ from the Inflaton Field Perturbation

Chapter 15 describes the generation of the perturbations using a smoothed quantities. To have a more precise discussion, we need instead to use Fourier components, with each wavelength proportional to a. A Fourier component of the inflaton field perturbation stops fluctuating when its comoving wavenumber k is equal to aH. This generates a curvature perturbation ζ, given by

$$\zeta = \frac{1}{a}\frac{\delta\phi}{d\phi/da} = \frac{H\delta\phi}{\dot{\phi}}$$

If the inflaton field is the only one with a perturbation, its value at a given location determines the future evolution of both the energy density and the pressure. Then the pressure depends only on the energy density, and ζ is constant so that it is equal to the quantity existing at the beginning of the known history.

One can show that just after the inflaton field stops fluctuating, its spectrum is equal to $(H/2\pi)^2$.[13] Using the slow-roll result for $\dot{\phi}$, the spectrum $\mathcal{P}_\zeta$ of ζ is therefore

$$\mathcal{P}_\zeta(k) = \left(\frac{H}{2\pi}\right)^2 \frac{9H^4}{V'^2}$$

where the right hand side is evaluated when $aH = k$. Using again the slow-roll results, we can then calculate a quantity called the spectral index, which is denoted by n;

$$n(k) - 1 \equiv \frac{1}{\mathcal{P}_\zeta(k)} \frac{d\mathcal{P}_\zeta(k)}{dk} = 2\eta - 6\epsilon$$

The first expression is the definition of $n(k)$. In the second expression η and ϵ are evaluated at the epoch when $aH = k$, and the expression is obtained using the slow-roll results. For a typical potential, $n(k)$ doesn't vary much over cosmological scales, and taking it to be a constant it becomes the spectral index defined earlier. The form of the inflaton potential is restricted by requiring that the predicted $\mathcal{P}_\zeta(k)$ agrees with observation.

The correlation of the Fourier components of ζ in this scenario is almost certainly too small ever to observe. This is unfortunate, because the correlation has a distinctive form which is independent of the potential. If observed, it would have confirmed the inflaton scenario beyond all reasonable doubt.

Generating ζ from the Perturbation of a Different Field

If ζ is generated from the perturbation of some other field, one obtains quite different results. Regarding the spectrum, observation constrains both the inflaton potential and the potential of the other field. In particular, $n(k) - 1$ now becomes equal to $2\tilde{\eta} - 6\epsilon$, where $\tilde{\eta}$ is defined in the same way as η, but using the potential of the other field.

Regarding the correlation between Fourier components, ζ is typically of the form

$$\zeta = \zeta_{\mathrm{g}} + \frac{3}{5} f_{\mathrm{NL}} \zeta_{\mathrm{g}}^2$$

where ζ_{g} has uncorrelated Fourier components.[14] At present observation tells us only that f_{NL} is less than about 10 in magnitude. Future observation though, will

[13] I am using Natural Units.

[14] The factor 3/5 is inserted for a historical reason.

detect f_{NL} if it is bigger than about 1 in magnitude. That is quite interesting, because the simplest version of what is called the curvaton scenario predicts $f_{NL} = 5/4$ which is probably big enough to eventually detect.

Table A.2 The radiation before electron-positron annihilation. All species are in thermal equilibrium at the same temperature T. The energy densities are in units of $(kT)^4/(\hbar c)^3$ and the number densities are in units of $(kT)^3/(\hbar c)^3$. (The numbers in the third column are the totals adding all three species.) The average particle energies are in units of kT

	Photons	Electrons and positrons	Neutrinos and anti-neutrinos
Energy density	0.658	1.152	1.728
Number density	0.244	0.366	0.549
Particle energy	2.70	3.15	3.15

Table A.3 The radiation after electron-positron annihilation. The units are the same as in Fig. A.2, with T the photon temperature. At present, $T = 2.73\,^\circ$K which corresponds to $kT = 2.36 \times 10^{-10}$ MeV

	Photons	Neutrinos and anti-neutrinos
Energy density	0.658	0.450
Number density	0.244	0.200
Particle energy	2.70	2.26

Appendix B
Tables

Table A.4 The elementary particles The particles in the first table are called bosons, those in the second quarks and those in the third leptons. The rest energies are in MeV. The mass of a particle in kilograms is equal to 1.79×10^{-30} times its rest energy in MeV. The photon and gluon always move with speed *c* and are said to have zero mass. Most species come with an **antiparticle**. The antiparticle has the same rest energy as the particle. The ν_1, ν_2 and ν_3 are known collectively as **neutrinos**. The neutrino rest energies are known to be very small. Most probably, two of them have the values indicated and the third has a much smaller value. In any case, the spectrum of the CMB anisotropy require the sum of the masses to be less than 10^{-6} MeV

Particle	γ (photon)	Gluon	Z	W_+	Higgs boson
Antiparticle	–	–	–	W_-	–
Rest energy	–	–	91,190	80,420	125,000

Particle	u	d	c	s	t	b
Antiparticle	$\bar{u}$	$\bar{d}$	$\bar{c}$	$\bar{s}$	$\bar{t}$	$\bar{b}$
Rest energy	2.5	5.0	1270	95	173,000	4240

Particle	e (electron)	μ (muon)	τ	ν_1	ν_2	ν_3
Antiparticle	ē (positron)	$\overline{\mu}$	$\overline{\tau}$	$\overline{\nu}_1$	$\overline{\nu}_2$	$\overline{\nu}_3$
Rest energy	0.511	106	1780	5×10^{-8}	9×10^{-9}	?

Table A.5 The rest energies of atomic particles in MeV

Neutron	Proton	Electron
939.6	938.3	0.511

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3

Table A.6 The amount of each conserved quantity carried by the elementary particles. Each anti-particle carries the opposite amount. The electric charges are in units of the proton charge. I am ignoring another conserved quantity called colour (carried only by quarks and gluons) because its density is everywhere zero

	Electric charge	Baryon number	Lepton number
u, c & t quarks	2/3	1/3	0
d, s & b quarks	−1/3	1/3	0
Leptons	−1	0	1
Photon	0	0	0
Higgs boson	0	0	0
Gluon	0	0	0
Z	0	0	0
W_+	1	0	0

Table A.7 Fundamental constants in MKS units

G	c	$\hbar$
6.674×10^{-11}	3.000×10^{8}	1.055×10^{-34}

Table A.8 Fundamental cosmological parameters To calculate the known history of the Universe one needs (in addition to the fundamental constants and the relevant Standard Model parameters) six cosmological parameters. The first five are needed to describe the homogeneous Universe. These refer to the present Universe. The first is the Hubble time. The second is the temperature of the CMB, and third and fourth are the fractions of the present energy density provided by the CDM and the ordinary matter. The fifth is the constant K appearing in the Friedmann equation, which is zero or too small to measure. The other two parameters are needed to specify the spectrum of the primordial density contrast, as described in this Appendix. The values of the parameters are chosen so that calculations of the CMB anisotropy and the galaxy distribution agree with observation (the CMB anisotropy alone gives almost the same result, with almost the same accuracy). With that choice, everything else is in adequate agreement with observation, such as the abundance of light elements and the properties of galaxies and galaxy clusters

$1/H_0$	CMB temperature	CDM fraction	OM fraction	K	A	$n-1$
1.44×10^{10}	2.725^0 K	0.26	0.048	0	$(6.9 \times 10^{-5})^2$	−0.04

Table A.9 Derived cosmological parameters These can be derived from the fundamental cosmological parameters. The subscript eq denotes the epoch when the matter and radiation have equal energy density and ls denotes the epoch of last scattering. Baryon number is the number of protons plus the number of neutrons

Age of Universe	Distance to present horizon	a_{eq}	a_{ls}	Baryon number per photon
1.38×10^{10} years	4.56×10^{10} lightyears	$\frac{1}{3400}$	$\frac{1}{1100}$	6.1×10^{-10}

Fig. A.8 The Greek letters (Source [www.preceptaustin.org])

Form	Name	English	Pronunciation
Α α	Alpha	A	al-fah
Β β	Beta	B	bay-tah
Γ γ	Gamma	G	gam-ah
Δ δ	Delta	D	del-tah
Ε ε	Epsilon	E	ep-si-lon
Ζ ζ	Zeta	Z	zay-tah
Η η	Eta	E	ay-tay
Θ θ	Theta	Th	thay-tah
Ι ι	Iota	I	eye-o-tah
Κ κ	Kappa	K	cap-ah
Λ λ	Lambda	L	lamb-dah
Μ μ	Mu	M	mew
Ν ν	Nu	N	new
Ξ ξ	Xi	X	zzEye
Ο ο	Omicron	O	om-ah-cron
Π π	Pi	P	pie
Ρ ρ	Rho	R	row
Σ σ	Sigma	S	sig-ma
Τ τ	Tau	T	tawh
Υ υ	Upsilon	U	oop-si-lon
Φ φ	Phi	Ph	figh or fie
Χ χ	Chi	Ch	kigh
Ψ ψ	Psi	Ps	sigh
Ω ω	Omega	O	o-may-gah

Glossary

The meaning of each term is summarised, with more detail in the indicated chapters. If a term is used only at the point where it is explained, it is not listed.

Abolishing gravity (Chap. 7) – Working in a small free-falling region for short time.

Acceleration of gravity (Appendix A) – Gravity gives the same acceleration to every object, no matter what is its mass.

Accelerator (Chap. 2) – A machine creating one or two beams of particles with very high energy, which collide with a stationary target or with each other.

Acoustic oscillation (Chap. 11) – The rapid fluctuation the density contrast of the baryon-photon fluid, that exists before the epoch of last-scattering.

Amplitude (Chap. 4) – The maximum variation of a quantity which is oscillating.

Antiparticle (Appendix B) – The partner of a particle, which has the same mass but the opposite value for each conserved quantity.

Atom (Chap. 4) – An object consisting of a nucleus and one or more electrons.

Baryon number (Chap. 6 and Appendix B) – One of the conserved quantities.

Baryons (Chaps. 6 and Appendix B) – Particles that carry baryon number.

Baryon-photon fluid (Chap. 11) – The nuclei, electrons and photons before the epoch of last scattering.

Beryllium (Chap. 7) – The atom whose nucleus has four protons.

Binding energy (Chap. 6) – The energy required to break apart a nucleus, atom or molecule.

Big Bang (Chap. 1) – The era during which the Universe consists of particles.

Black hole (Chap. 5) – An object whose gravity is so strong that nothing can emerge from it.

Boltzmann constant (Chap. 9) – The physical properties of something with temperature T depend on kT where k is the Boltzmann constant.

CDM (Chap. 1) – Cold Dark Matter

CDM halo (Chap. 12) – An object formed when an over-dense region attracts more CDM and collapses under its own weight.

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3

Charge (Chap. 4) – Charge is carried by an object that generates an electric field when stationary, as shown in Fig. 4.4. Its magnitude determines the strength of the field and its sign determines the field's direction. The electron and proton carry opposite amounts of charge.
Classical electrodynamics (Chap. 3) – Describes charged objects and electric and magnetic fields, ignoring quantum effects.
Classical physics (Chap. 3) – Physics ignoring quantum effects.
CMB (Chaps. 1, 5 and 11) – Cosmic Microwave Background
Cold Dark Matter (Chap. 1) – Matter in the Universe which is invisible because it neither absorbs nor emits known particles.
Collision (Chap. 6) – The close encounter or actual contact, of two particles that are moving towards each other.
Collision process (Chap. 6) – A collision that produces one or more new particles.
Conserved quantity (Chap. 6) – A quantity whose total amount cannot be changed.
Constant – Not changing with time (adjective). A quantity that doesn't change with time (noun).
Cosmic Gas (Chap. 5) – The gas between galaxy clusters, which was all there was before galaxy formation.
Cosmic Microwave Background (Chaps. 1, 5 and 11) – The electromagnetic radiation made of the photons that existed just after the epoch of last scattering.
Cosmic Neutrino Background (Chaps. 1, 5 and 9) – The neutrinos that existed just after neutrino interactions cease.
Cosmological Constant (Chaps. 5 and 8) – The energy density of the vacuum, which is constant.
Cosmological scales (Chaps. 11 and Appendix A) – The scales of cosmological interest.
Curvaton scenario (Appendix A) – The scenario which generates ζ from the curvaton field.
Curvaton field (Chap. 15) – A hypothetical field whose perturbation generates ζ.
Curvature perturbation (Appendix A) – The initial perturbation ζ, which determines future perturbations.
Dark energy (Chap. 14) – Energy in the present Universe, that does not come from particles.
Decay (Chap. 6) – The disappearance of a particle, creating other particles.
Density (Chap. 9) – Amount of a quantity per unit volume. (Mass density if the quantity is not specified.)
Density contrast (Chap. 11) – The perturbation in the energy density, divided by the average energy density.
Distribution functions of a gas (Appendix A) – Specify the contribution to the energy density of each constituent, from particles with energy in a given range.
Early Universe (Chap. 7) – The Universe before galaxy formation.
Electric field (Chap. 4) – The field generated by a stationary charged object.
Electromagnetic field (Chap. 4) – The collective name for electric and magnetic fields.

Electromagnetic wave (Chap. 4) – A wave consisting of rapidly varying electric and magnetic fields.
Electron (Chap. 4) – One of the elementary particles.
Electron-positron annihilation (Chap. 6) – The collision of an electron with a positron causing both to disappear.
Electron-positron annihilation (epoch of) (Chap. 7) – The epoch when each positron in the Universe annihilates with some electron.
Elementary particle (Chap. 3) – A particle that is not known to consist of other particles.
Energy density (Chap. 8) – Energy per unit volume.
Expansion rate (Chap. 10) – The rate at which the scale factor increases, denoted by $\dot{a}$.
Extension of the Standard Model (Chap. 3) – A theory like the Standard Model but with more particles.
Field (Chap. 3) – A fundamental quantity occupying a whole region of space, as opposed to a particle.
Field (scalar) (Chap. 13) – A field that is specified just by its strength.
First stage (Chap. 7) – Beginning of the known history, when everything except CDM is in thermal equilibrium.
Fourier expansion (Appendix A) – Defined by Eq. (A.2) of Appendix A.
Fundamental constants (Chap. 3) – The numbers G, c and h.
Free-falling (Chap. 7) – Subject only to the force of gravity.
Frequency of an oscillation (Chap. 4) – Is 1 divided by its period.
Galaxy (Chap. 5) – An astronomical object held together by gravity, consisting of stars and gas.
Galaxy cluster (Chap. 5) – An astronomical object held together by gravity, consisting of galaxies and gas.
General Relativity (Chap. 3) – Einstein's theory of gravity.
Gravitational field (Appendix A) – The field which specifies the strength of gravity.
Gravitational radiation (Chap. 5) – An oscillating gravitational field.
Hadron (Chap. 3) – A particle made out of quarks.
Helium (Chap. 4) – The atom whose nucleus contains two protons.
Higgs boson (Chap. 13) – An elementary particle of the Standard Model.
Higgs field (Chap. 13) – The field whose oscillation would correspond to the presence of Higgs bosons.
Homogeneous – The same everywhere
Homogeneous Universe (Chap. 5) – The Universe, considered under the pretence that it is homogeneous.
Horizon (Chaps. 1 and 10) – Boundary of the biggest region that could be observed in principle.
Hot Big Bang (Chap. 16) – The Big Bang between the onset of thermal equilibrium and last scattering.
Hubble distance (Appendix A) – Defined as $ca/\dot{a}$.
Hubble parameter (Appendix A) – Defined as $\dot{a}/a$, denoted by H.

Hubble time (Appendix A) – Defined as $a/\dot{a}$.
Hydrogen (Chaps. 4 and 7) – The atom whose nucleus has one proton.
Inflation (Chaps. 1 and 14) – The hypothetical early era of expansion with repulsive gravity.
Inflaton (Chap. 14) – The particle corresponding to the inflaton field.
Inflaton field (Chap. 14) – The scalar field responsible for inflation.
Initial state (Chaps. 9 and 11) – The state of the Universe at the beginning of the known history.
Interactions (Chap. 3) – The collisions and decays of particles.
Isotopes (Chap. 4) – Atoms whose nuclei have the same number of protons but different numbers of neutrons.
Isotropic – The same in all directions.
Kinetic energy (Chap. 6) – The difference between the energy of a particle, and its rest energy.
Known history (Chap. 7) – The history after the first hundredth of a second.
Last scattering (epoch of) (Chap. 7) – The epoch after which photons cease to bounce off electrons.
Lepton number (Chap. 9 and Appendix B) – One of the conserved quantities.
Leptons (Appendix B) – Particles that carry lepton number.
Lightyear – The distance travelled by light in one year, 9.46×10^{12}] km.
Lithium (Chaps. 4 and 7) – The atom whose nucleus has three protons.
Magnetic field (Chap. 4) – The field that is detected by a compass needle.
Magnitude (Chap. 2) – A number ignoring its sign.
Matter (Chap. 8) – Particles in the Universe moving with speed much less than c.
Matter-dominated Universe (Chap. 8) – The Universe while most of its energy density comes from matter.
Mean-square (Chap. 11) – The average of the square of a quantity.
MeV (Chap. 6) – The energy unit used in this book.
MKS units (Chap. 3) – The system of units employing the metre, kilogram and second.
Natural Units (Appendix B) – The system of units in which $\hbar = c = 1$.
Neutral (Chap. 4) – Carrying zero charge.
neutrino (Chaps. 3, 5, 6, 7, 8, 9 and 11) – Three elementary particle species have this name.
Nucleus (Chap. 4) – The object at the centre of an atom, consisting of protons and neutrons.
Number density of a thing (Chap. 9) – The number of the thing per unit volume.
Observable Universe (Chaps. 1 and 10) – The Universe within a horizon-sized sphere around us.
Ordinary hydrogen (Chap. 4) – The atom whose nucleus is a single proton.
Ordinary matter – Consists of protons, neutrons and electrons.
Oscillating – Undergoing an oscillation.
Oscillation of a quantity – A periodic switch back and forth, described by the Sine function.

Parameters of the Standard Model (Chap. 3) – Numbers that specify the Standard Model, in addition to the relevant fundamental constants.

Parameters (cosmological) (Chaps. 10 and 11) – Numbers that define the evolution of the known Universe, in addition to the fundamental constants and the relevant Standard Model parameters.

Particle (Chap. 3) – A tiny object, of which many identical examples exist. The examples are called collectively a particle species.

Peculiar velocity (Chap. 7) – The velocity with which an object is moving through the Cosmic Microwave Background.

Period of an oscillation (Chap. 4) – The time interval between the maximum values of a quantity that is oscillating.

Perturbation (Chap. 11) – The small variation of a quantity with position, that exists in the early Universe.

Potential of a scalar field (Chap. 13) – The energy density that the field would have, if its strength were homogeneous and constant.

Photon (Chap. 4) – The elementary particle associated with an electromagnetic wave. Each electromagnetic wave can be regarded as a beam of photons.

Planck length (Appendix A) – $\sqrt{\hbar G/c^3} = 1.62 \times 10^{-35}]$ metres.

Planck energy (Appendix A) – $\sqrt{\hbar c^5/8\pi G} = 2.43 \times 10^{21}]$ MeV.

Planck scale (Appendix A) – When Natural Units are employed, the Planck energy and the Planck length.

Positron (Chap. 7) – The anti-electron.

Primordial energy density perturbation (Chap. 11) – The smoothed energy density perturbation, that exists on cosmological scales at the beginning of the known history.

Principe of Relativity (Appendix A) – The statement that the laws of physics are the same for every observer who is not accelerating.

Proportional to (Chap. 2) – Two things are proportional to each other if doubling one doubles the other, tripling one triples the other and so on.

Quantum effects (Chap. 3) – Physical phenomena that depend on the value of Planck's constant.

Quantum physics (Chap. 3) – Physics that includes quantum effects.

Radiation (Chap. 8) – Particles in the Universe moving with speed at or close to c.

Radiation-dominated Universe (Chap. 8) – The Universe while most of the energy density comes from radiation.

Rest energy (Chap. 6) – The energy of a stationary object.

Scale factor (Chap. 7) – The distance between any two parts of the Universe, divided by its present value. Denoted by a.

Scale (of smoothing) (Chap. 11) – Present value of the smoothing radius.

Scale (of a Fourier component) (Appendix A) – Present value of the inverse wavenumber of a Fourier component.

Scattering (Chap. 6) – The bouncing of particles off each other, which prevents them from travelling in straight lines.

Smoothing (Chap. 11) – Replacing the energy density at each point, by its average within a sphere.
Smoothing radius (Chap. 11) – The radius of the sphere used when smoothing.
Special Relativity (Chap. 3) – The description of space and time provided by Einstein, which ignores gravity.
Square of a number – The number times itself.
Standard Model (Chap. 3) – The currently accepted theory of particles and their interactions.
stretched scale – A scale in a graph, for which equal spacing corresponds to equal multiples.
Thermal equilibrium of a gas (Chap. 9) – The state of a gas which is brought about by frequent collision processes.
Vacuum value of a field (Chap. 13) – The value of a scalar field in the vacuum, ignoring quantum effects.
Vacuum fluctuation (Chap. 13) – The departure of a field from its vacuum value, due to quantum effects.

Symbols

a (Chap. 7) – The scale factor of the Universe.
$\dot{a}$ (Chap. 10) – The rate of change of a.
Be (Chap. 4) – Beryllium
c (Chap. 3) – The speed of light, 3.00×10^5 kilometres per second.
G (Chap. 3) – Newton gravitational constant,
h (Chap. 3) – Planck constant
$\hbar$ – $h/2\pi$.
H (Chap. 4) – Hydrogen.
H (Appendix A) – Hubble parameter $\dot{a}/a$.
He (Chap. 4) – Helium
^{2}**H** (Chap. 4) – Isotope of hydrogen whose nuclei contains two particles. (Notation used for any number and any atom.)
Li (Chap. 4) – Lithium
k (Chap. 9) – The Boltzman constant, 2.7×10^{11} °K. (In Appendix A, k is the present wavenumber of a Fourier component.)
t – Time. During the Big Bang, $t = 0$ is its beginning.
T (Chap. 9) – Temperature.

Index

D.H. Lyth, *The History of the Universe*, Astronomers' Universe,
DOI 10.1007/978-3-319-22744-3