Introduction

In this book you will study the fundamentals of cosmology, and how cosmological models are defined and used to describe the geometry and evolution of the Universe on the largest scales. There are five chapters in *Cosmology Part 1*, which develop the concepts and the mathematical tools of cosmological theory, and examine some of the observational evidence used to test it.

- Chapter 1 provides a general introduction to the science of cosmology.
- Chapter 2 develops the concepts, language and mathematics of special relativity and of curved geometries, which underpin cosmological theory.
- Chapter 3 explores the key ideas of general relativity and how they are used to build a mathematical description of the Universe.
- Chapter 4 introduces and examines the Friedmann equations, the basis of modern cosmological models, and defines the key cosmological parameters that can be used to test cosmological theory.
- Chapter 5 considers how measurements of distances to astronomical objects are used to measure cosmological parameters, and therefore test the theories introduced in previous chapters.

The next book you will study – *Cosmology Part 2* – continues by examining another main observational tool for cosmology, the cosmic microwave background (or CMB) radiation. It then provides an in-depth history of the Universe, from a 'soup' of energetic particles to the formation of the first stars and galaxies.

The exercises in each chapter are an important element of your learning. They are there to help develop and reinforce key ideas, and several important concepts in this book are developed through the exercises and nowhere else. Full solutions for the exercises are provided at the end of this book, but do try to complete them yourself before looking at the answers.

A table of physical constants is also given at the end of the book for use in your calculations. Be aware that, in some cases, your numerical answers to calculations may differ slightly from those in exercise solutions, depending on the precision used for constants and any intermediate rounding in the solution. Definitions for terms highlighted in **bold** may be found in the module glossary.

Throughout the text, coloured boxes are used to highlight particular types of information. Orange boxes highlight the most important equations and other key information. Turquoise boxes indicate additional information, such as reminders of concepts that you may have met in previous study, or ideas that are partly beyond the scope of the module but provide additional context. Blue boxes indicate where further, optional resources are available on the module website.

Finally, it is important to acknowledge some important influences on the writing of the S385 books. Some of the content presented here builds on books written for the previous Open University module S383 *The relativistic universe*, namely Robert Lambourne's *Relativity, gravitation and cosmology*, Stephen Serjeant's *Observational cosmology*, and Ulrich Kolb's *Extreme environment astrophysics* (all published in 2010). We are grateful to the module team who produced those texts.

Many other excellent textbooks on cosmology and extragalactic astrophysics have influenced the development of the books for S385. We would like in particular to acknowledge Barbara Ryden's *Introduction to cosmology* (2017) and Andrew Liddle's *An introduction to modern cosmology* (1999). We recommend these books as optional further reading, but please note that they may differ from S385 in their choice of notation, convention, and the mathematical approaches used to set out some key equations.

Chapter 1 Introduction to cosmology and the expanding Universe

Cosmology is the scientific study of the Universe as a whole. It involves taking a 'big picture' view of the contents, geometry, history and potential future of the Universe.

Over the last century or so, advances in theoretical physics, telescopes and other astronomical technology, and most recently the power of modern computation, have combined to provide powerful ways to mathematically describe and understand the Universe and its evolution, and to test the predictions of cosmological theory with precise observational measurements. We don't yet have all of the answers – there are some fundamental gaps in our knowledge – but modern cosmology provides a rich and powerful toolbox for trying to understand the Universe.

The main ingredients of modern cosmology are: a few basic assumptions about the nature of the Universe, Einstein's theory of general relativity, which relates space, time, matter and energy, and the physics of radiation and matter and the ways they interact. This introductory chapter sets out the context of modern cosmological theory and some important astronomical observations that have motivated the development of those concepts, and continue to provide ways to test our theories.

Objectives

Working through this chapter will enable you to:

- describe the key principles that underpin modern cosmological theory
- apply basic concepts of observational astronomy and the physics of matter and radiation relevant to cosmology
- identify the main stages in the history of the Universe and place particular events onto its timeline
- summarise the primary evidence for the expansion of the Universe and the hot big bang model
- solve numerical problems relating to the contents and temperature evolution of the Universe, and the interaction of matter and radiation.

1.1 Key ideas in cosmology

1.1.1 The cosmological principle

Cosmology is built on some fundamental assumptions, which make it possible to construct a (comparatively straightforward) mathematical description of the geometry of the Universe and how it changes with time. A key assumption is that there is no privileged location in the Universe: if the properties of space (or spacetime, the nature of which we will discuss later) are measured from any location in the Universe, then what we learn should be true of the Universe as a whole. This leads to the **cosmological principle**, an important starting point for all of cosmology.

The cosmological principle

On the largest scales, the Universe is assumed to be **isotropic** (i.e. the same in all directions) and **homogeneous** (i.e. the same in all locations).

You might immediately question this assumption. Our local environment in space does not look very homogeneous or isotropic: the Solar System contains one star and a variety of planets and other objects, all spread out in a very uneven way. But cosmology is the study of the Universe on the largest scales, and it is on much larger scales than the Solar System that we can test this principle.

Figure 1.1 shows a slice of the observed Universe on scales useful for constructing cosmological theories. You are seeing part of the **cosmic web**, a network of individual galaxies that form clusters and superclusters with hundreds to many thousands of galaxies in each. Superclusters span distances of hundreds of megaparsecs (see the box that follows shortly about distance units), and it is on these scales that the Universe is thought to be homogeneous.

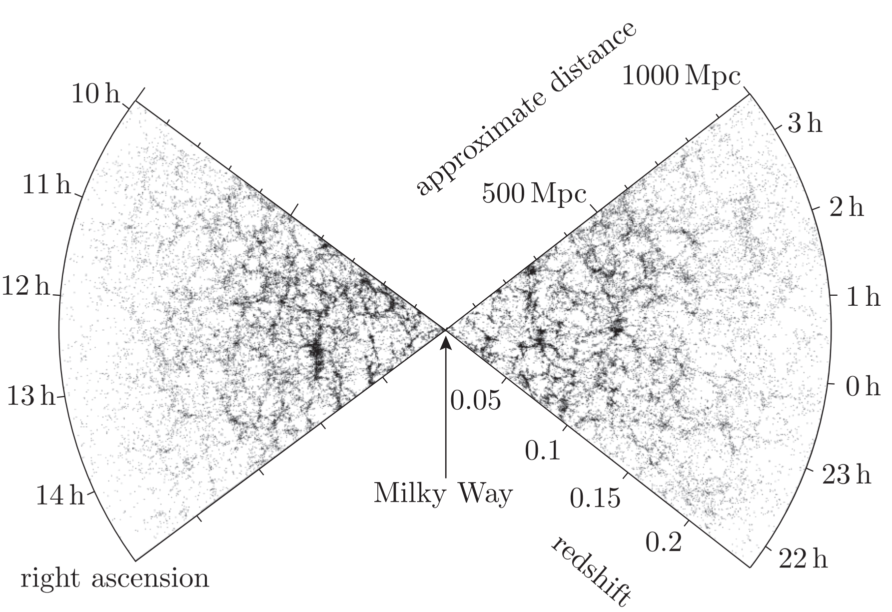

Figure 1.1 The large-scale structure of the Universe, mapped by the 2dF Galaxy Redshift Survey. Note that the apparent decrease in galaxy density at larger distances is due to it being harder to detect fainter objects, rather than there being fewer galaxies present.

Distance units in cosmology

The size scales considered in cosmology are vast, which means that ordinary SI units of metres and kilometres are impractical. Astronomers usually express distances beyond our Solar System in units of parsecs (pc), where $1\,\text{pc} = 3.086 \times 10^{16}\,\text{m}$, or roughly 3.3 light-years. The distances between individual galaxies in the Universe are typically thousands to millions of parsecs, and so cosmological calculations often involve working with distances in units of megaparsecs (Mpc).

- Why does assuming the Universe is homogeneous make it easier to construct mathematical models that describe it?

- Homogeneity means that we can define universal parameters that describe the contents or behaviour of the Universe at a particular time (e.g. the overall density of matter), and that these parameters will have the same value everywhere. If the Universe were inhomogeneous then its key properties would depend on location as well as time, which would make cosmological models both more mathematically complex and harder to test with observations.

1.1.2 The expanding Universe

One of the most widely known cosmological facts is that the Universe is expanding. This idea has an interesting history. When Einstein's theory of general relativity was first applied to the geometry of the Universe in 1917, the solution had a peculiar and initially undesirable feature: that the Universe tended to contract or expand rather than remaining static.

At that time it wasn't yet proven that the Universe extended beyond our own galaxy, the Milky Way. Over the next few years the properties of galaxies beyond our own became firmly established; by 1930 it had been clearly demonstrated that nearby spiral galaxies were not part of our Galaxy, and were all moving away from us.

Figure 1.2 shows the observed relationship between the distance of galaxies and the speed at which they are receding, based on the work of Edwin Hubble and collaborators, which was published in 1929. Note that the vertical axis units are labelled incorrectly in this original plot – they are in fact $\text{km}\,\text{s}^{-1}$. Systematic errors in the 1920s distance measurements mean that the value of the slope of the relationship is also now known to be different.

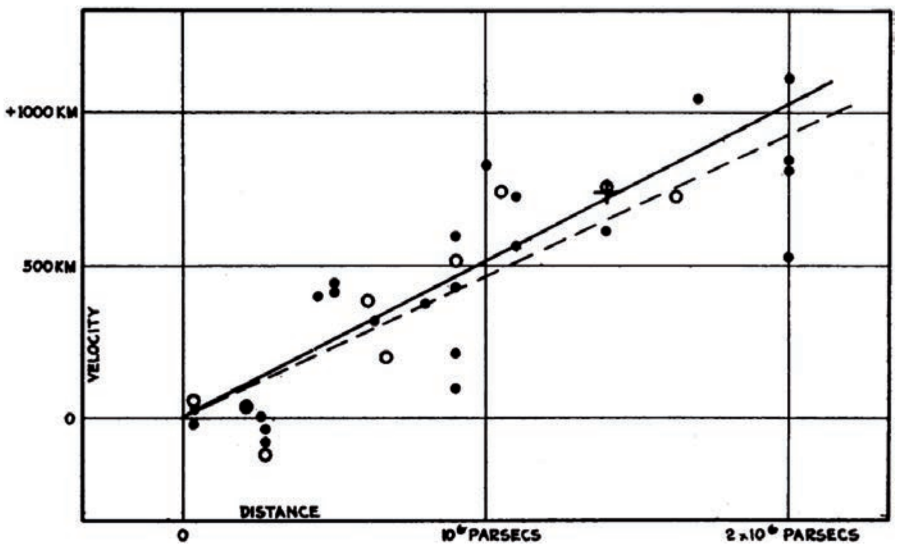

Figure 1.2 The famous result published by Hubble in 1929, showing a relationship between galaxy recession speed and distance. Such plots are now sometimes referred to as Hubble diagrams.

The galaxy recession speeds in Figure 1.2 were obtained by measuring redshifts of emission lines in the galaxies' spectra. A brief reminder of the concept of redshift is given in the following box.

Cosmological redshift

Redshift, z, measures the change in wavelength of radiation emitted from an astronomical body (such as a galaxy or quasar) caused by its movement away from the observer. It is defined as:

$$z = \frac{\lambda_{\rm obs} - \lambda_{\rm em}}{\lambda_{\rm em}} = \frac{\lambda_{\rm obs}}{\lambda_{\rm em}} - 1 \tag{1.1}$$

where $\lambda_{\rm em}$ and $\lambda_{\rm obs}$ are the wavelengths of a spectral feature (e.g. an emission line, such as from atomic transitions) as emitted from the source and as measured by the observer, respectively.

Spectral shifts are caused by motion towards the observer (blueshifting) or away from the observer (redshifting). For distant galaxies, measured redshifts are dominated by the effect of cosmological expansion, which causes the apparent recession of galaxies, and so can be used as a distance measure.

As Figure 1.2 indicates, in the relatively nearby Universe there is a linear relationship between recession velocity and distance. This is known as the **Hubble–Lemaître law**, and can be written as

$$z = \frac{H_0 D}{c} \tag{1.2}$$

where D is the galaxy distance, typically measured in units of Mpc, c is the speed of light, typically measured in units of $\rm km\,s^{-1}$, and H_0 is a quantity known as the **Hubble constant**.

Figure 1.3 shows a modern Hubble diagram, using redshift and distance measurements of supernova explosions in distant galaxies from several surveys. Note that the axes are swapped relative to Figure 1.2, with distance now on the vertical axis.

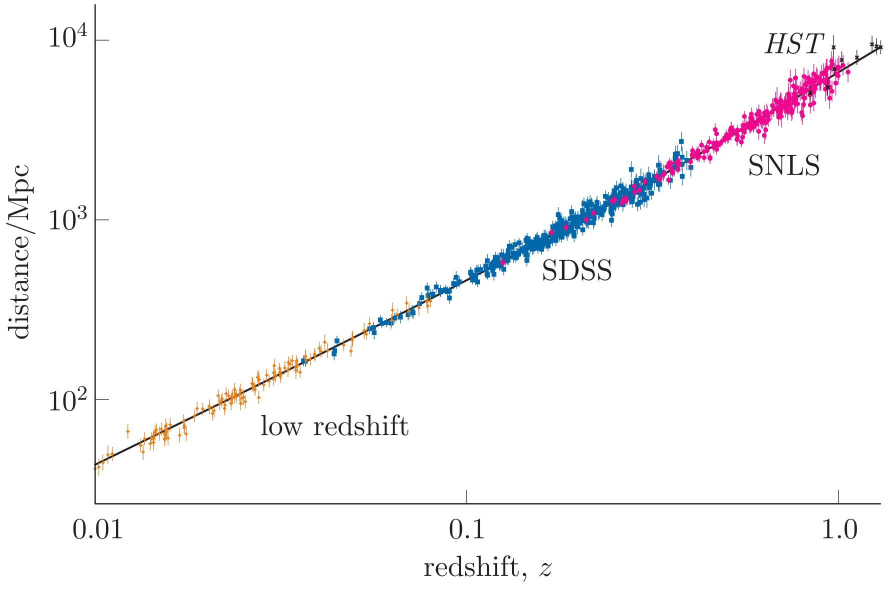

Figure 1.3 The relationship between distance and redshift from a compilation of supernova measurements (Betoule *et al.*, 2014), including a low-redshift sample, and samples from the Sloan Digital Sky Survey (SDSS), the Supernova Legacy Survey (SNLS) and *Hubble Space Telescope* (*HST*) observations.

If you look carefully at Figure 1.3 you will see that towards larger redshifts the relation begins to deviate from the straight line of the Hubble–Lemaître law. We will return to this observation later in the module.

- Based on Equation 1.2 and Figure 1.3, what is H_0 a measure of, and what units would you expect it to have?

- The Hubble constant is a measure of the rate of expansion of the Universe. Because redshift is a dimensionless quantity then, without applying any unit conversions, H_0 must have units of $\text{km s}^{-1}\,\text{Mpc}^{-1}$ to cancel out the units of D and c in Equation 1.2.

The units of the Hubble constant are a little unusual – kilometres per second makes sense for a rate of expansion speed, but what does the 'per megaparsec' part tell us? The answer lies in the relationship shown by Figure 1.2: we observe distant galaxies to recede more rapidly than nearer ones. In other words, the amount by which a region of space expands in a given time interval depends on the scale being considered. H_0 therefore measures the rate of expansion *over a fixed distance*: observers at any location at the current time will measure that, over a distance of 1 Mpc, the Universe expands at a rate of around $70\,\text{km s}^{-1}$.

What does it really mean for the entire Universe to be expanding? For this to be a well-defined concept, which can be rigorously tested with observations, it is necessary to set out a mathematical description of space and time.

The geometry of the Universe can be described by mapping out a four-dimensional geometry, with the usual three spatial dimensions plus the important dimension of time. This geometry is known as **spacetime**, and describes how the separation distance between locations in the Universe can be measured consistently by observers, irrespective of location. It is the expansion of this 'fabric' of spacetime that leads to the observation of the recession of distant galaxies. The concept of spacetime, and the mathematical framework to describe it, are not straightforward to grasp, and will be expanded on in the two chapters that follow.

In introducing the expansion of the Universe it is important to note that the term 'Hubble constant' is a misnomer, although it is widely used. Observations of the distant Universe demonstrate that the rate of expansion of the Universe is not constant with time. To avoid confusion, the subscript '0' in H_0 is used to indicate that this is the expansion rate that applies at the *current* time, i.e. at the point in the Universe's history at which we are measuring. When we treat the expansion of the Universe quantitatively in later chapters, you will instead be working with the Hubble parameter, which is the time-dependent generalisation of H_0 (the latter is therefore often referred to as 'H-nought', rather than as the Hubble constant).

The relationship described by the Hubble–Lemaître law allows cosmologists to draw conclusions about how the expansion of the Universe has changed with time over its history. Distance and time are intrinsically connected when studying the extragalactic Universe for a simple reason: the finite speed of light means that the light from distant galaxies has been travelling to us for billions of years, and so we are viewing them as they looked at a time when the Universe was very much younger. This link between distance and the time elapsed since light was emitted means that – because of the relationship shown in Figure 1.3 – astronomers can use redshift as a measure of cosmic time. High-redshift galaxies are effectively showing us the earlier history of the Universe, whereas low-redshift galaxies show us what is happening in the Universe closer to the present day.

● How does the cosmological principle help us to interpret any differences in the appearance of very distant galaxies compared to those nearby?

○ The cosmological principle tells us that the galaxies in any region of the Universe at the present day should, on average, have the same properties. Any systematic differences in the appearance of distant galaxies compared to nearby ones must therefore be caused by the fact that we are observing them at different cosmic times, i.e. they must be due to how the Universe has evolved with time.

1.1.3 The brightness of the night sky

The expansion of the Universe, as demonstrated by redshift observations, is one of the foundational ideas of modern cosmology. Another direct form of evidence that we do not live in an infinite, unchanging (non-expanding) Universe dates back to a conundrum that was first identified in the early nineteenth century.

One of the most basic observations about the Universe is that the night sky is mainly dark. But should we expect that to be the case if the Universe is infinite and full of stars? In this section we will investigate what we would predict the night sky to look like in a static, infinite Universe. We will make use of some basic quantities in observational astronomy in the process, some of which are outlined in the following box.

Flux and luminosity

Flux is defined as the energy per unit time per unit area passing through a surface, for example the rate of visible-light photons from a star passing through a telescope detector at the Earth (typically measured in units of $\mathrm{W\,m^{-2}}$). Flux, F, usually refers to the total energy across all relevant wavelengths. You may also meet the term **flux density**, which is flux per unit frequency or wavelength, measured in a narrower part of the spectrum.

Flux is related to **luminosity**, L, which is the total energy emitted per unit time by the system under consideration (for example, a star or a galaxy). The relationship between flux and luminosity is determined by geometry. If a distant object emits light isotropically, then the fraction of light that a detector of a particular size intercepts will depend on the distance from the object to the detector, leading to the following relation between flux and luminosity for an object at distance D from the Earth:

$$F = \frac{L}{4\pi D^2} \tag{1.3}$$

The **surface brightness**, Σ, of a square region of the sky with angular dimensions $\theta \times \theta$ is:

$$\Sigma = \frac{F}{\theta^2} \tag{1.4}$$

where F is the total flux of emission from that region of sky measured in units of $\mathrm{W\,m^{-2}}$, and Σ will have units of $\mathrm{W\,m^{-2}\,deg^{-2}}$, assuming that the angle θ is measured in units of degrees.

To explore what we can learn about the Universe from the observation that the night sky is dark, we will first consider, in the example that follows, what we would predict to be the surface brightness of a typical star. Before that, the next box provides a brief reminder of some common notation associated with cosmology.

Common notation conventions

The symbol $\odot$ is used in astronomy to indicate quantities associated with the Sun. Hence $L_\odot$ is the Sun's luminosity, $R_\odot$ is the solar radius, and so on. These quantities are also often used as units in astronomy, so that the luminosity of a galaxy, for example, can be expressed in units of solar luminosity ($\mathrm{L}_\odot$, typeset without italic formatting), rather than in SI units of watts.

In comparison, a subscript asterisk can be used to denote the properties of another star, so M_* would signify a given star's mass.

Example 1.1

Assuming that the luminosity and intrinsic radius of a typical star match those of the Sun ($L_\odot$ and $R_\odot$), calculate the ratio between the surface brightness, as observed at the Earth, of a typical star and that of the Sun.

Solution

Under this assumption, we can use the relationship between F and L to compare the surface brightness of the Sun, $\Sigma_\odot$, and that of a star, Σ_*, if the star is at a distance D_* from the Earth.

The ratio of the surface brightness of a star to that of the Sun is given by

$$\frac{\Sigma_*}{\Sigma_\odot} = \frac{(F_*/\theta_*^2)}{(F_\odot/\theta_\odot^2)} = \frac{F_*\theta_\odot^2}{F_\odot\theta_*^2}$$

Substituting in for luminosity using Equation 1.3 gives

$$\frac{\Sigma_*}{\Sigma_\odot} = \frac{(L_*/4\pi D_*^2)\theta_\odot^2}{(L_\odot/4\pi D_\odot^2)\theta_*^2} = \frac{L_* 4\pi D_\odot^2\theta_\odot^2}{L_\odot 4\pi D_*^2\theta_*^2} = \frac{D_\odot^2\theta_\odot^2}{D_*^2\theta_*^2}$$

where the (equivalent) luminosities have cancelled out, so that the ratio depends on the product of the distance and angular size of the two stars.

The final step is to relate the stars' angular sizes to their distances. The simple geometric considerations shown in Figure 1.4 illustrate that the angular diameter of a star is related to its size and distance by $\theta = 2R/D$, where we have used the small-angle approximation $\tan(\theta/2) \approx \theta/2 = R/D$.

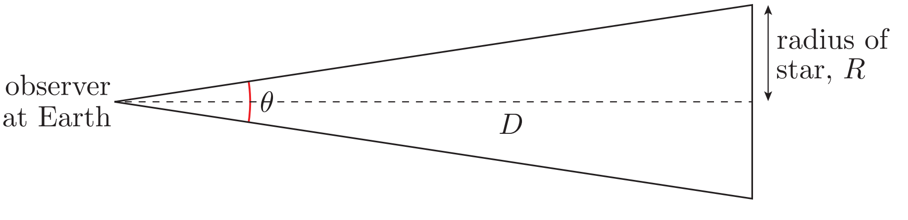

Figure 1.4 The relationship between distance D and physical radius R for an astronomical object spanning a small angle, θ, of the sky.

This means that

$$\frac{\Sigma_*}{\Sigma_\odot} = \frac{D_\odot^2(4R_\odot^2/D_\odot^2)}{D_*^2(4R_*^2/D_*^2)} = \frac{R_\odot^2}{R_*^2} = 1$$

where the last step follows because the star was assumed to be Sun-like (i.e. $R_* = R_\odot$). In other words, we conclude that the ratio between the surface brightness of a typical star and that of the Sun is 1.

We have just demonstrated that the surface brightness of a typical star of the same luminosity and radius as the Sun will be the same as that of the Sun. In other words, brightness per unit area of the sky is independent of distance for objects of the same intrinsic properties. Work through Exercise 1.1 to consider the implications of this conclusion further, by directly comparing the surface brightness of the Sun with the average surface brightness of the night sky.

Exercise 1.1

The Sun has an angular diameter of around 0.5 deg as seen from Earth, and a luminosity of $L_\odot = 3.8 \times 10^{26}$ W. Calculate its surface brightness, $\Sigma_\odot$, in units of $\mathrm{W\,m^{-2}\,deg^{-2}}$, and compare this value to the mean surface brightness of the night sky, which is $\sim 2 \times 10^{-13}\,\mathrm{W\,m^{-2}\,deg^{-2}}$. (*Hint*: the Sun's distance from Earth is 1.5×10^{11} m.)

Of course, it is not surprising to discover that the surface brightness of the Sun is very much higher than that of the night sky – our own eyes tell us this! But if the Universe is infinite, and all the stars in the Universe have (roughly) the same surface brightness, shouldn't we see stars in *every* direction we look at night, and the sky be much brighter than we observe it to be?

This conundrum is known as **Olbers' paradox**, after Heinrich Olbers, who considered it in the 1820s. You may have questioned some assumptions that were made, but (for example) assuming that stars are less similar to the Sun than we have assumed does not solve the problem.

One possible solution to the paradox is that the Universe is not infinite in space. We cannot directly test this with astronomical observations, but we do now have considerable evidence that the Universe is not infinite in *time*. The Universe has a finite age, thought to be $\sim$13.8 billion years, and the light from some stars has not yet had time to reach us.

- In a static (non-expanding) Universe with a finite age of $\sim$13.7 billion years, how far away is the most distant star from which light has had time to reach us?

- Using the familiar relationship between speed (here c), distance (D) and time (t), the distance light could have travelled is $\sim 1.3 \times 10^{26}$ m, or $\sim$4200 Mpc.

One of the main lines of evidence that the Universe is not infinitely old is that we observe it to be expanding, as discussed in the previous section. The expansion of the Universe also means that the most distant stars are receding further and further away. So Olbers' paradox is no longer a worry for modern astronomers, but it does provide clear evidence against the idea that we are living in a static, infinite Universe.

1.2 The contents of the Universe

Why does the Universe expand, and what controls its rate of expansion? The theory of general relativity provides a way of trying to answer these questions, and can be summed up in a famous quote from the physicist John Wheeler: 'Space tells matter how to move; matter tells space how to curve' (Misner, Thorne and Wheeler, 1973, p. 5). If we want to understand why and how space is expanding, it's necessary to think about the matter contained within it. In fact, it is not just 'matter' that matters: the behaviour of spacetime is influenced by both matter and energy, which are linked by Einstein's famous $E = mc^2$ equation.

Understanding the different forms of matter and energy present in today's Universe, and how they interact with each other, also underpins our ability to make astronomical observations. Our main tools for measuring the Universe are electromagnetic radiation, particles such as cosmic rays and neutrinos, and, most recently, gravitational waves. All of these cosmic messengers are produced by the matter in stars and galaxies and, at the earliest times, by a 'soup' of particles filling the Universe. In the following sections you will review what we know about the contents of the Universe.

1.2.1 Particles and interactions

Modern particle physics is based on quantum field theory, which can successfully predict the properties and interactions of the families of particles that form the **Standard Model**. Most of the particles that are important for understanding the history of the Universe are part of the Standard Model.

Figure 1.5 shows the two major particle families: the **fermions**, which include three generations of matter particles, and the **bosons**, which carry force or mediate interactions. These families have different quantum mechanical properties. For example, fermions are subject to the **Pauli exclusion principle**, but bosons are not (a difference that is important for the structure of stars). In this module, the only boson that features heavily is the **photon**, the particle that mediates the electromagnetic force. Photons strongly influenced how the early Universe evolved, are crucial for the physics of stars and galaxies, and result in the images and spectra we measure with telescopes here at the Earth.

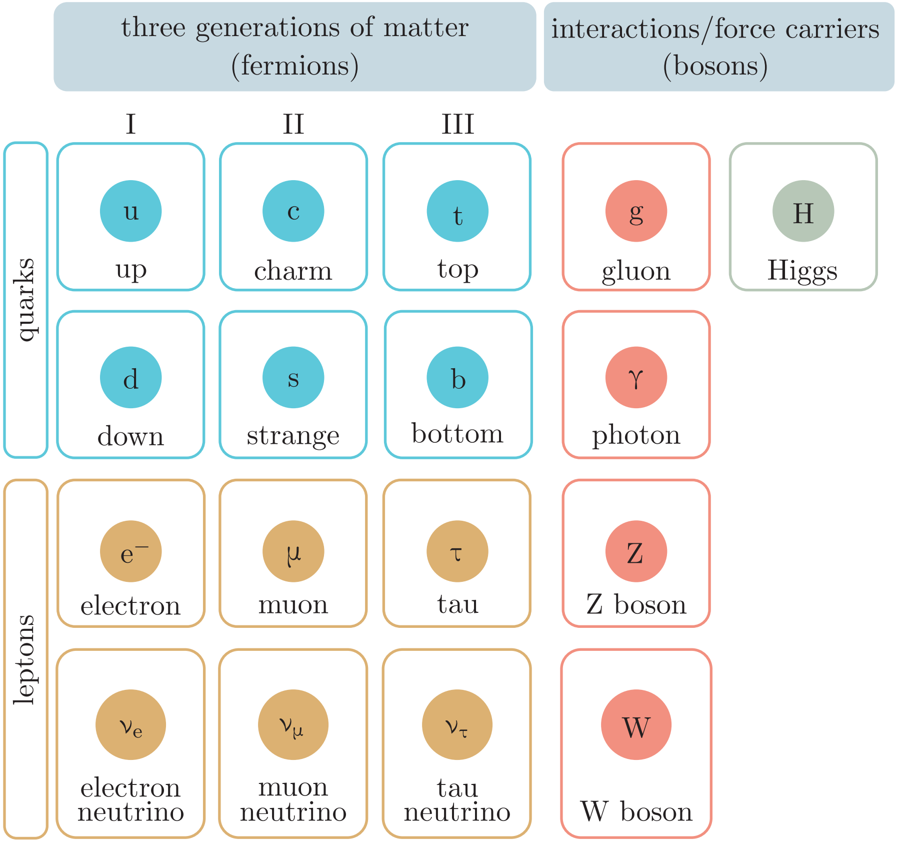

Figure 1.5 Particles in the Standard Model.

There are two families of fermions: the **quarks** and the **leptons**. These particles are distinguished by the types of fundamental interaction in which they participate. Leptons are subject to the **electromagnetic interaction** and the **weak interaction** (mediated by photons and the Z and W bosons respectively), while quarks also interact via the **strong interaction**, mediated by bosons known as gluons.

Considering the leptons first: although there are six types, as shown in Figure 1.5, in most cosmological contexts it is only necessary to consider the **electron** and its important 'partner', the electron **neutrino**. Electrons are, of course, an important constituent of atoms. There are also many situations in astrophysics where matter exists in the form of an ionised gas or **plasma**, in which ions and electrons are separated (and magnetic field effects can become more important than for neutral gases). Neutrinos are important by-products of a variety of weak interactions relevant to nuclear reactions occurring in the early Universe, as well as in stars.

Quarks also have six types, but as for the leptons it is only the first 'generation' that have much importance for the content of this module. Up and down quarks combine via the strong interaction to form the more familiar **proton** and **neutron**, each consisting of three quarks in a combination that yields electric charges of +1 and 0, respectively.

Each fermion also has an antiparticle equivalent, which is not shown in Figure 1.5. For example, the **positron** is the antiparticle counterpart of the electron, having the same mass but positive charge.

Protons and neutrons are types of **hadron**, a term that encompasses all particles made by assembling two or more quarks via the strong force. **Baryons** are hadrons that contain an odd number of quarks, and so protons and neutrons are included in this category. Protons and neutrons make up most of the visible mass in the Universe, and so it is common in cosmology to refer to the **baryonic matter** of the Universe, meaning the ordinary matter that makes up stars, galaxies and interstellar and intergalactic gas. You will encounter this terminology regularly in discussions of cosmological models in later chapters.

Although most particles of interest to cosmology are part of the Standard Model, there is one important exception. 'Invisible' dark matter, not explained by the Standard Model, is thought to also be an important component of the Universe – a fact that can cause some scepticism about the whole endeavour of cosmology!

1.2.2 Matter and energy

There are several reasons to consider the overall contents of the Universe at this point. Figure 1.6 summarises the matter and energy contents of the present-day Universe and of the early Universe, according to current best observational estimates and cosmological theory. You will see later that the energy densities (the energy per unit volume) of different constituents of the Universe are important parameters in cosmological models, because Einstein's field equations of general relativity tell us that the matter and energy content at any given epoch in the Universe's history determines its rate of expansion at that time.

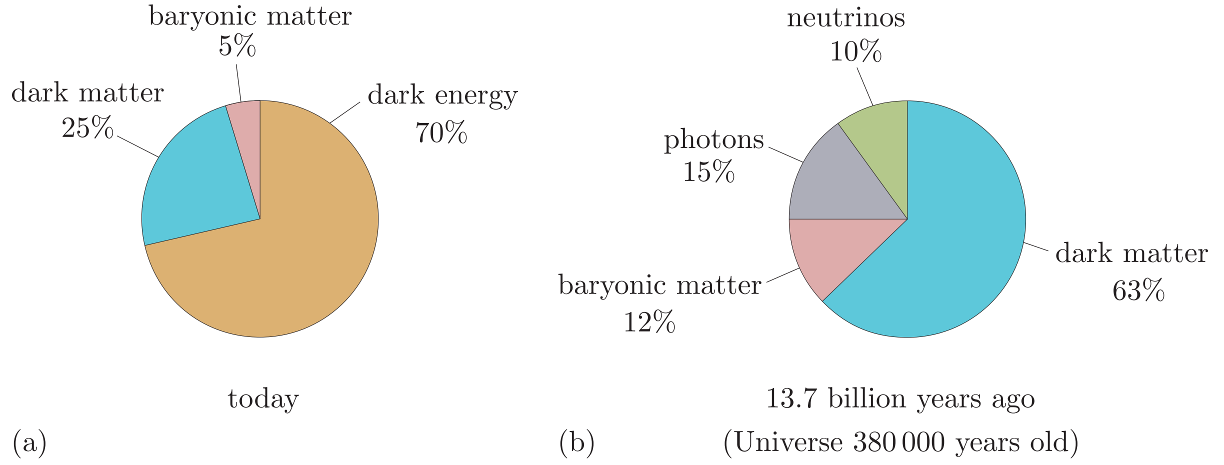

Figure 1.6 The matter and energy content of (a) the present-day Universe, and (b) the early Universe soon after the big bang.

Figure 1.6 also highlights the considerable limits of our current knowledge. All of the well-tested physics of the Standard Model discussed in the previous section apply only to $\sim$40 per cent of the matter and energy at early times, and an even smaller proportion of the overall matter and energy balance of the present-day Universe.

- What is the main difference between the make-up of the early Universe and that of the present day?
- The dominant energetic component of the present-day Universe is 'dark energy', but in the early Universe matter and radiation dominated, and a wider range of matter types contributed to its overall energy density.

Dark matter and **dark energy** are two of the biggest unknowns in cosmology. As mentioned in the previous section, it is currently thought that dark matter is a type of as-yet-undiscovered particle that is needed to explain the motions of stars in galaxies and the behaviour of galaxies and gas in clusters of galaxies, as well as several other important measurements. Dark energy is an as-yet-unexplained form of energy that was unimportant in the early Universe, but appears to be driving the rate at which the Universe is expanding in the present day. You will learn more about both of them in later chapters.

1.2.3 Properties of gases

The physics of matter and radiation in the Universe is primarily the physics of gases – or, more properly in many contexts, of plasmas. For most situations considered in this module this distinction can be ignored, and so for simplicity we will often use the term gas irrespective of whether or not the material is fully ionised.

One of the most important properties of gases for understanding the evolution of the Universe, and for many physical processes occurring in stars and galaxies, is density. Commonly, we will use n in this module to refer to particle number density (the number of particles per unit volume) and ρ to refer to mass density (the mass per unit volume).

The other fundamental property of gases to consider is temperature, T, which is a measure of the thermal or kinetic energy associated with the motions of individual particles of the gas. In a hotter gas the typical speeds of particles are higher than those in a gas at lower temperature. The populations of baryons and leptons within a gas will have a range of different speeds. When those speeds are sub-relativistic (i.e. well below the speed of light), interactions between the particles will result in a stable distribution of particle kinetic energies over time, or a situation of **thermal equilibrium**. Under these conditions the particles' speeds, v, follow a **Maxwell–Boltzmann distribution**, as shown in Figure 1.7, and are described by

$$f(v)\,\mathrm{d}v = \left(\frac{m}{2\pi k_\mathrm{B}T}\right)^{3/2} 4\pi v^2 \mathrm{e}^{-mv^2/(2k_\mathrm{B}T)}\,\mathrm{d}v \qquad (1.5)$$

where $f(v)\,\mathrm{d}v$ is the fraction of particles (of mass m) whose speed (magnitude of velocity vector) has a value between v and $v + \mathrm{d}v$, and k_B is the Boltzmann constant.

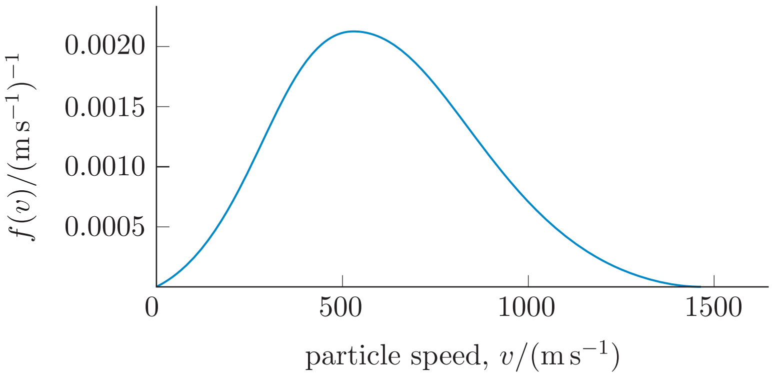

Figure 1.7 Example of a Maxwell–Boltzmann velocity distribution.

- Why is it very improbable for gas particles to have a speed at or close to zero?

- In a gas where most particles are moving at high speeds, it is very likely that any stationary particle will undergo a collision with a moving particle, which will cause the stationary particle to speed up.

It's important to be aware of this distribution of particle energies, because although the majority of interactions will involve particles with properties at the peak of the distribution, in some situations the presence of a 'tail' of higher-energy (faster-moving) particles is important. Equation 1.5 also explains why some other equations you will meet later in the module include an exponential factor that depends on particle kinetic energy (the mv^2 term) and temperature.

- Would you expect photons to follow a similar distribution? Briefly explain your reasoning.

- Photons always travel at the speed of light, so cannot have a distribution of speeds.

Although photons cannot have a distribution of speeds, those produced in a particular environment will have a distribution of *energies* (corresponding to the frequency or wavelength of light), and in situations of thermal equilibrium this can be used to assign an effective temperature to the radiation.

For matter particles with the most extreme energies, where their speeds approach the speed of light, particle velocity distributions no longer take a Maxwell–Boltzmann form; the somewhat different behaviour of such relativistic particles will be important in some situations considered in this module. Particle energies also depart from the Maxwell–Boltzmann distribution at high densities, where quantum mechanical effects, and the differing behaviours of fermions and bosons, become important.

One relationship that is different for (relativistic) gases made of photons rather than matter particles is that between pressure, P, and temperature, T. For a gas of particles in many astrophysical situations we can use the **ideal gas law**:

$$P = nk_{\mathrm{B}}T \tag{1.6}$$

In contrast, a gas made of photons has the following relationship between pressure and temperature, with no density dependence:

$$P = \frac{1}{3}aT^4 \qquad (1.7)$$

where a is the radiation constant.

1.2.4 Interaction of matter and radiation

Electromagnetic radiation was present in very large quantities in the early Universe, but the light that is observed at the Earth – and that allows us to study cosmology – has all been produced or affected by interactions involving baryons, electrons and/or ions. (Note that in this module we will use 'light' to refer to electromagnetic radiation across its whole spectrum, and will refer to 'optical' or 'visible' light at particular points if it is necessary to be more specific.)

- What are some of the processes by which light is produced in astrophysical environments?

○ Light can be produced by particle–antiparticle annihilation and by any process in which particles undergo acceleration, including electromagnetic interactions between ionised particles in a gas. In an atomic gas, light of particular wavelengths is produced by atomic transitions in which electrons transition to a lower energy state, releasing photons in the process.

- What gas properties influence the production of radiation?

○ Typically, gas density and temperature are the controlling parameters. Properties such as magnetic field strength can also be relevant.

Propagation of radiation in a gas

As well as determining the properties of the radiation we see, gas density and temperature also determine in what situations radiation escapes from its environment to travel to our telescopes, and thus what information the radiation is carrying about that environment.

A key concept here is **opacity** – the extent to which gas properties allow photons to travel through the medium without being absorbed or scattered by gas particles. If we know which absorption or scattering process is the dominant type of interaction in a particular environment, then it is possible to calculate the **mean free path**, which is the typical distance a photon can travel before interacting with a gas particle. For example, in a fully ionised gas – where ions and electrons are not combined into atoms – the most likely type of interaction for a photon under a wide range of conditions is **Thomson scattering**. Here, a photon's energy and direction are altered by interaction with an electron.

The mean free path, λ, for electron scattering in a gas is

$$\lambda = \frac{1}{n_\mathrm{e}\sigma_\mathrm{T}} \tag{1.8}$$

where n_e is the electron number density and σ_T is the Thomson cross-section ($6.652 \times 10^{-29}\,\mathrm{m}^2$), which is analogous to a cross-sectional 'target area' for the photon to hit.

Exercise 1.2

The present-day mean electron number density in the halo (outer regions) of the Milky Way is $\sim 100\,\mathrm{m}^{-3}$, while at a particular epoch in the early Universe it was $\sim 5.0 \times 10^9\,\mathrm{m}^{-3}$. Calculate the mean free path of a photon in each of these environments in units of kiloparsecs, assuming electron scattering is the dominant interaction process for photons. Comment on how these distances compare to the size of a typical present-day galaxy (e.g. the Milky Way has a diameter of $\sim 27\,\mathrm{kpc}$).

In the early Universe, photons could only travel on scales much smaller than individual galaxies before scattering off an electron, whereas photons produced in the halos of galaxies today can easily travel much larger distances without such interactions. This example highlights the importance of interactions between photons and gas particles for how light travels in the Universe, which is primarily controlled by gas density.

Ionisation of gas

Another very important interaction between radiation and matter concerns the **ionisation** of gas, which is dependent on gas temperature. The level of ionisation can affect the opacity of a gas, as well as determine which types of particle are available for other interactions.

The ionisation state of a gas of hydrogen in thermal equilibrium is described by the **Saha equation**. This equation compares the ratio of the number density of hydrogen atoms (n_H) to those of dissociated protons and electrons (n_p and n_e, respectively):

$$\frac{n_\mathrm{H}}{n_\mathrm{p}n_\mathrm{e}} = \left(\frac{m_\mathrm{e}k_\mathrm{B}T}{2\pi\hbar^2}\right)^{-3/2} \exp\left(\frac{Q}{k_\mathrm{B}T}\right) \tag{1.9}$$

Here, m_e is the electron mass, k_B the Boltzmann constant, T the gas temperature, $\hbar$ (h-bar, known as the reduced Planck constant) is equal to $h/2\pi$ where h is the Planck constant, and Q is the binding energy of a hydrogen atom, i.e. the energy an absorbed photon needs to have in order to ionise an atom of hydrogen ($13.6\,\mathrm{eV} = 2.18 \times 10^{-18}\,\mathrm{J}$).

The ionisation fraction of a gas is defined as $X = n_\mathrm{p}/n_\mathrm{b}$, which is the ratio of the number density of free protons to that of all baryons (which, for a pure hydrogen gas is given by $n_\mathrm{b} = n_\mathrm{p} + n_\mathrm{H}$). The next example explores this relationship.

Example 1.2

Show that for a pure hydrogen gas, in which $n_\mathrm{e} \approx n_\mathrm{p}$, the Saha equation can be rewritten in terms of the ionisation fraction as

$$\frac{1-X}{X} = n_\mathrm{p} \left(\frac{m_\mathrm{e} k_\mathrm{B} T}{2\pi\hbar^2}\right)^{-3/2} \exp\left(\frac{Q}{k_\mathrm{B} T}\right) \tag{1.10}$$

Solution

Examining the Saha equation (Equation 1.9), we can see that the right-hand side of the equation is completely determined by temperature, because all other terms are constants. We can therefore simplify the Saha equation temporarily by replacing the right-hand side with a constant depending only on T, which we'll call $C(T)$. Using the suggested assumption that $n_\mathrm{e} \approx n_\mathrm{p}$, Equation 1.9 becomes:

$$\frac{n_\mathrm{H}}{n_\mathrm{p}^2} = C(T)$$

The final expression we are aiming to derive involves only X on the left-hand side, while the right-hand side is equivalent to $n_\mathrm{p} C(T)$. We can therefore rearrange our already simplified Saha equation to produce this same right-hand side:

$$\frac{n_\mathrm{H}}{n_\mathrm{p}} = n_\mathrm{p} C(T)$$

We now need to express the left-hand side in terms of X. We start by expanding the expression for X in terms of the quantities in the Saha equation, n_H and n_p:

$$X = \frac{n_\mathrm{p}}{n_\mathrm{b}} = \frac{n_\mathrm{p}}{n_\mathrm{p} + n_\mathrm{H}}$$

There are various ways we could manipulate this to find a relation between $n_\mathrm{H}/n_\mathrm{p}$ and X. One is to take the reciprocal of both sides of the expression and then rearrange:

$$\frac{1}{X} = \frac{n_\mathrm{p} + n_\mathrm{H}}{n_\mathrm{p}} = 1 + \frac{n_\mathrm{H}}{n_\mathrm{p}}$$

Hence

$$\frac{n_\mathrm{H}}{n_\mathrm{p}} = \frac{1}{X} - 1$$

and multiplying both sides by X and rearranging again gives

$$\frac{n_\mathrm{H}}{n_\mathrm{p}} = \frac{1-X}{X}$$

which can be substituted into the rearranged and simplified Saha equation above to give the required expression:

$$\frac{1-X}{X} = n_\mathrm{p} C(T) = n_\mathrm{p} \left(\frac{m_\mathrm{e} k_\mathrm{B} T}{2\pi\hbar^2}\right)^{-3/2} \exp\left(\frac{Q}{k_\mathrm{B} T}\right)$$

Now try the following exercise to investigate ionisation fractions in different astrophysical situations.

Exercise 1.3

Calculate the ionisation fraction X for gas conditions corresponding to:

(a) the Sun's outer regions (corona), where $T \approx 10^6\,\mathrm{K}$ and $n_\mathrm{p} = 10^{14}\,\mathrm{m}^{-3}$

(b) a temperature of $T = 4000\,\mathrm{K}$ and $n_\mathrm{p} = 5 \times 10^9\,\mathrm{m}^{-3}$

(c) a temperature of $T = 3000\,\mathrm{K}$ and $n_\mathrm{p} = 5 \times 10^9\,\mathrm{m}^{-3}$.

In all cases you may make the simplifying assumption that the gas is composed of pure hydrogen.

The preceding exercise shows that the ionisation fraction is very sensitive to the gas temperature, because of the exponential term on the right-hand side of the Saha equation. More generally, in this section you have seen that gas density and temperature have a strong influence on how radiation and matter behave. The next section sets out a broader overview of how the changing density and temperature of the early Universe have driven the evolution of matter, leading to the eventual formation of stars and galaxies.

1.3 The big bang model

The inevitable consequence of the Universe's expansion is that any region of space must have been smaller at earlier times than at present, so that matter and energy must have been more concentrated. The structures that we see in the present-day Universe – stars and galaxies and so on – must have formed from material that was previously concentrated into a very small volume. This scenario is known as the **big bang model**, in which everything that we observe today evolved from an initial state that was much smaller, hotter and denser than today's Universe.

Imagining an early Universe so different from our own environment is hard, and it is perhaps also hard to conceive that we could have got from there to here. But there is a wealth of evidence from different types of astronomical observations in support of the big bang model. This evidence includes a powerful, direct way of measuring the properties of the early Universe (e.g. its temperature and density): the cosmic microwave background (CMB) radiation.

In the remainder of this chapter we will sketch out the timeline of the history of the Universe to give you a 'big picture' outline, which you will be able to supplement with additional knowledge as you work through the rest of the module. You will also be introduced to the basic properties of the CMB and why it provides such a powerful tool for understanding the history of the Universe.

1.3.1 A brief history of the Universe

The big bang model does not attempt to explain *how* or *why* the Universe came into being. It only seeks to construct a history of how the Universe has evolved by extrapolating details from what we can observe (including the laws of physics operating in the present-day Universe). This story must be self-consistent: the laws of physics should not need to change, and the conditions in the early Universe, however extreme, must evolve naturally to produce the stars, planets and galaxies we see today, over 13.5 billion years later.

Figure 1.8 sets out the key evolutionary changes between the big bang and the present day. A lot of important action occurred in the first half-million years after the big bang, as spacetime expanded and matter cooled. The seeds of today's large structures – galaxies and groups and clusters of galaxies – were present in the cosmic soup of particles at early times. Then, under the influence of gravity, these small variations in density grew over billions of years to form the stars and galaxies we see in the night sky now.

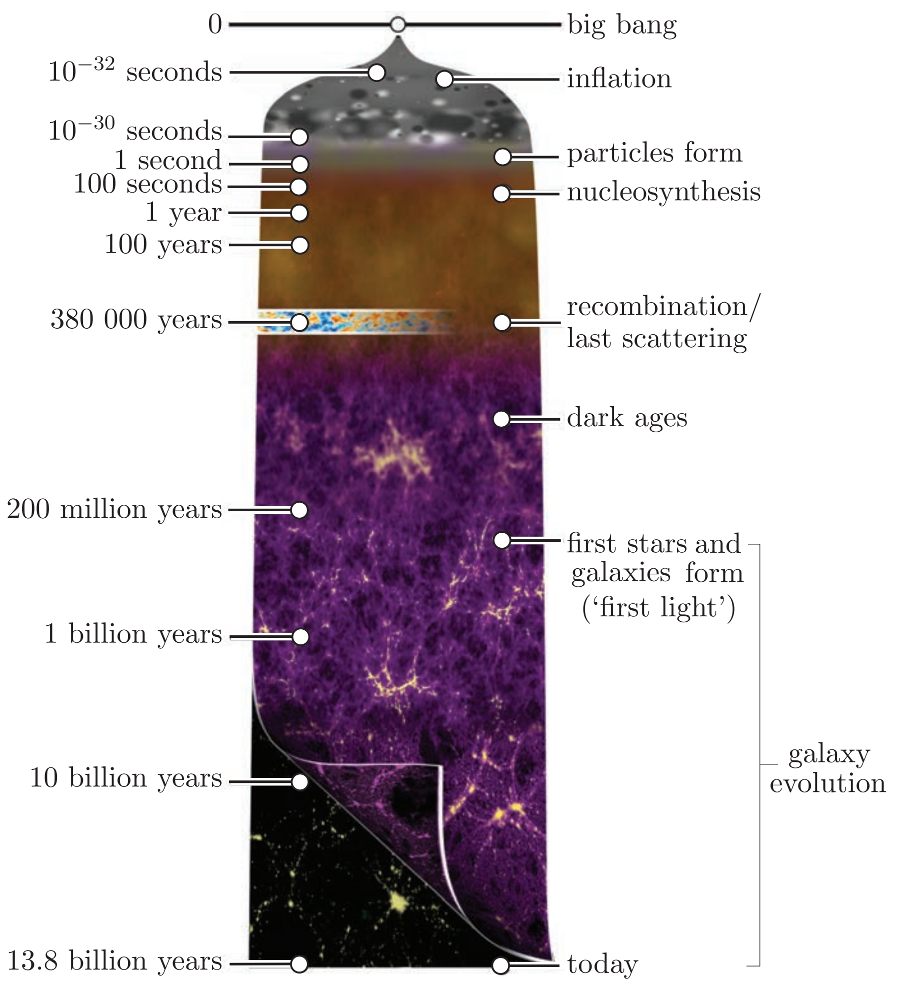

Figure 1.8 A simple timeline of the main stages in the history of the Universe. Note that this isn't a graph – the timeline is stretched to best enable all the important stages to be visualised.

A very brief description of the key periods in the timeline of the Universe is given below. As we move down the list the density of the Universe is decreasing, which reduces its energy per unit volume. Thus, temperature decreases with time too.

- **Inflation**: An almost incomprehensibly brief period, during which it is believed spacetime expanded by a very large factor, before slowing to a gentler expansion.
- **Particles form**: An initial cosmic soup of leptons, quarks, neutrinos and radiation cool, thereby allowing protons and neutrons to form (**baryogenesis**).
- **Nucleosynthesis**: When the Universe cooled sufficiently, it became possible for atomic nuclei to form and synthesise a small number of chemical elements (H, He, Li).
- **Recombination**: While high temperatures persist, gas remains ionised, because there are always many energetic photons to cause atoms to dissociate. Recombination describes the point at which the temperature became low enough to allow atoms to form.
- **Last scattering**: At high densities, ions, electrons and photons are constantly interacting, resulting in thermal equilibrium. By around the time of recombination, the density had dropped sufficiently for photons to escape from this gas, undergoing their 'last scattering' with the gas particles, after which the Universe became transparent to photons. The spherical 'surface' of this last scattering, as perceived from the Earth, delineates the earliest directly observable extent of the Universe, beyond which it was/is opaque.
- **Dark ages**: The early stages of **structure formation** – in which matter clumped together under gravity to form the seeds of stars and galaxies – are not visible to our telescopes, because the optical light produced in those environments was immediately absorbed by the gas.
- **First light**: The first stars and galaxies formed several hundred million years after the big bang, and their light ionised the surrounding gas. The Universe gradually became transparent to optical light.
- **Galaxy evolution**: Galaxies and their wider environments continued to evolve in complex ways over billions of years until the present day.

This timeline is the result of the combination of theoretical physics with testing via powerful observations. The concept of **lookback time**, an extremely useful consequence of the finite speed of light, is what allows such observational tests of cosmology. As we discussed earlier, the Universe is sufficiently large that we can observe objects and processes taking place at distances such that the light we measure left its originating location billions of years ago, when the Universe was a fraction of its current age. It is a remarkable property of the Universe that by looking as deeply as we can with our telescopes we are able to see how the Universe looked at an earlier point in time. Perhaps the most incredible such example is the subject of the next section.

1.3.2 The cosmic microwave background

Extraordinary claims require extraordinary evidence, and certainly some conclusions of cosmology are pretty extraordinary. But scientists have amassed rich and varied evidence in support of the currently favoured cosmological model. The **cosmic microwave background** (CMB) radiation is such a crucial part of this observational evidence that it deserves an introduction here. (In later chapters you will consider in detail the range of inferences that can be made from these observations.)

The CMB is a radio signal that pervades the Universe. It was first detected in 1965 by Penzias and Wilson, having previously been predicted as a consequence of theories of the early Universe. The history of its discovery is an interesting story, but one we don't have space to include here – it is well documented in science history books and online resources if you wish to learn more about it.

The two fundamental measurable properties of the CMB are its spectrum and its spatial distribution across the sky. Since its discovery, the CMB has been mapped in increasingly exquisite detail, most recently with all-sky maps obtained by two space missions: the *Wilkinson Microwave Anisotropy Probe* (*WMAP*) launched by NASA in the early 2000s and the European Space Agency's *Planck* mission a decade or so later.

Figure 1.9 shows the most detailed ever all-sky map of the CMB, made by *Planck*. The image uses a projection that maps the celestial sphere in a way that preserves the relative areas of different regions while showing the full sky. The colours indicate small shifts in the measured spectrum at different locations on the sky, which trace differences in the temperature (and hence density) of the material that produced the radiation. Structures are present both on large and small scales, and in a later chapter you will explore how these structures encode information about the early Universe.

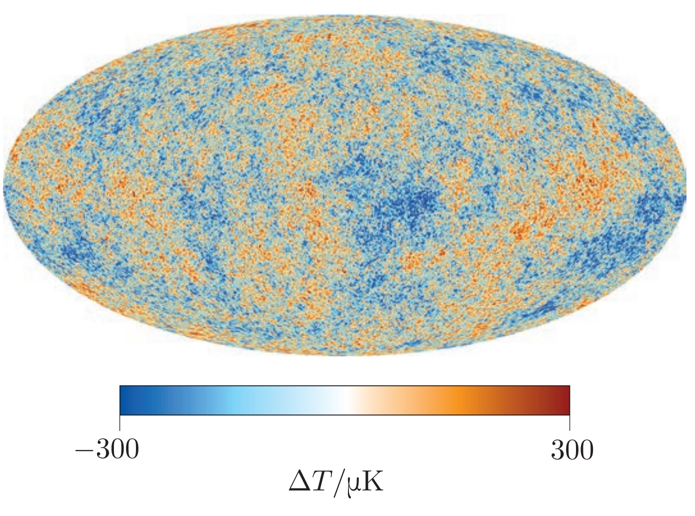

Figure 1.9 An all-sky map of the CMB radiation, as mapped by ESA's *Planck* mission. Colour indicates the deviation of temperature from the mean, ΔT, at each position on the sky, showing detailed structure that encodes information about the early Universe.

One of the most remarkable features of the CMB is how uniform it is when we measure it in all directions in space. The fluctuations shown in Figure 1.9 are variations of ~ 1 part in $>10\,000$, and so while the variations are very scientifically important (as you'll see later), the CMB has a uniform appearance on the sky in all directions when considered to an accuracy of much better than 1 per cent. This has profound implications, which we'll come back to in a moment.

A second important observation about the CMB is that it has a **black-body spectrum**: a characteristic dependence of brightness on wavelength, produced from regions that are opaque to radiation and in thermal equilibrium. The CMB's observed black-body spectrum reveals that the radiation comes from a region that is likely to be quite dense, so prior to escaping to travel towards us, the photons had a small mean free path and were constantly being scattered and/or absorbed and re-emitted.

Putting together these two features of the CMB – its high degree of uniformity and its black-body spectrum – leads to the conclusion that *the entire currently observable Universe* must have been in thermal equilibrium at the time the radiation was emitted. For this to be the case, the radiation must have been produced when regions of space were in close enough proximity to come into equilibrium. This makes it clear that the radiation is not being produced in the present-day Universe, when most of space is transparent to photons over cosmological distances, and the far reaches of the Universe are much too widely separated to come into equilibrium with each other.

Therefore, the CMB could only have been produced in a much denser early Universe. It is a key prediction of the big bang model that the photons produced when the Universe was last fully opaque should still exist, and so the CMB is one of three such predictions that are beautifully matched by observations. The other two predictions are the observed expansion of the Universe, and the abundances of key elements, both of which will be discussed in a later chapter.

Black-body radiation has the special property that the luminosity produced depends only on the temperature and surface area of the region from which it was emitted (i.e. unlike other processes, the amount of radiation produced doesn't depend on the quantity or density of matter present). The temperature, T, of a black body is related to the wavelength at which its brightness peaks. The observed wavelength peak of the CMB, at $\sim 280\,\text{GHz}$, indicates a temperature of $2.7\,\text{K}$.

- ● If the CMB was produced in the very hot early Universe, then why is the temperature we measure so low (close to absolute zero)?
- ○ The peak wavelength we measure here at the Earth doesn't tell us the temperature at which the radiation was produced. It is necessary to account for redshifting of the spectrum, i.e. the 'stretching' out of distances between wavefronts caused by the expansion of spacetime since the light was emitted.

Last scattering and the production of the CMB took place around 400 000 years after the big bang. According to the current cosmological model, the Universe has expanded by a very large factor since then, so that (although CMB photons were emitted throughout the entire volume of the Universe at the time) the CMB photons we observe here on Earth have travelled across a vast distance.

Exercise 1.4

The temperature of the CMB decreased as the Universe evolved according to $T \propto 1 + z$. If the ambient gas temperature at the time the CMB was emitted was $T \approx 3000\,\mathrm{K}$, calculate the redshift z corresponding to the time of CMB production.

The redshift of the CMB corresponds to a time when the Universe was around 0.4 million years old, or only around 3 per cent of its current age, so the CMB was, indeed, produced at a very early time in the Universe's history. Observations of the CMB therefore provide us with a powerful, direct way to measure properties of the early Universe.

1.4 Summary of Chapter 1

- The science of cosmology relies on the **cosmological principle**: the assumption that on the largest scales (hundreds of Mpc) the Universe is **homogeneous** and **isotropic**, i.e. it appears the same for all present-day observers at any location.
- The Universe is expanding, which is best described as an expansion of the spacetime intervals between fixed points. The Hubble parameter, H, measures the rate of expansion per unit distance. H changes with time as the Universe evolves, with its current value represented as H_0, termed the **Hubble constant**.
- The main evidence for the expansion of the Universe comes from observing the **redshifts** of distant galaxies, which increase with distance, as captured in the **Hubble–Lemaître law**, demonstrating that all distant galaxies are receding.
- It is the finite speed of light that allows us to see how the Universe looked at an earlier point in time and study its evolution. The **lookback time** is that elapsed between the emission of light at its source and our detection of it here at the Earth.
- The present-day Universe contains a variety of forms of matter and energy, which have evolved over time, but its energy density is currently dominated by two components whose physics are not fully understood: **dark matter** and **dark energy**.

- Many types of particle in the **Standard Model** are relevant to the science of cosmology. These include up and down **quarks** that interact via the **strong interaction** to form **protons** and **neutrons**, which are types of **baryon**. Baryons can combine via nucleosynthesis to form atomic nuclei, and with **electrons** to form atoms.
- Gas density and temperature influence the **ionisation** state of matter, the production of radiation, and the ability of radiation to escape from astrophysical environments (i.e. **opacity**).
- The observed expansion of the Universe motivates the **big bang model** in which the Universe began in a state of extremely high density and temperature, from which it has subsequently expanded in a process that continues to the present.
- Key periods in the history of the Universe include **inflation**, **baryogenesis**, **nucleosynthesis**, **recombination**, **last scattering**, the **dark ages**, **structure formation**, **first light** and **galaxy evolution**.
- The **cosmic microwave background** (CMB) radiation provides evidence for the hot big bang model.
- The CMB has a **black-body spectrum** and is spatially uniform, with low-level fluctuations that enable precision measurements of parameters that describe the Universe's subsequent evolution.