Chapter 9 Early Universe physics meets observations

The previous chapter described a timeline extending from the creation of the very first particles to the production of the first elements via nucleosynthesis. This timeline is controlled by the expansion of spacetime and the resulting decrease in temperature and particle density with time.

As the Universe continued to cool and expand, it became possible for nuclei to combine with free electrons to form the first atoms. This is a crucial point in the history of the Universe, because it is around this time that the cosmic microwave background (CMB) was produced, enabling us to observe directly the conditions of the early Universe. In this chapter we will consider how a range of observations provide tests of the theory of how the early Universe evolved, and of the relative contributions of different types of matter, including dark matter.

Objectives

Working through this chapter will enable you to:

- describe, and mathematically investigate, the interactions of matter and radiation at the time when the CMB was produced
- discuss how the observed abundances of chemical elements in the Milky Way and other galaxies provide evidence for the big bang model
- explain how the cosmic abundances of helium, deuterium and lithium can be measured, and the challenges and limitations in making these measurements
- explain the importance of baryon density, Ω_{b}, as a cosmological parameter
- describe and critically discuss the key evidence for a dominant quantity of non-baryonic (i.e. dark) matter.

9.1 From nuclei to atoms

9.1.1 Recombination

After the formation of baryons, the Universe remained ionised for around 250 000 years. While the temperature remained high enough for photoionisation to be a dominant process, nuclei and electrons could not remain stably combined.

Eventually, as the temperature dropped, the ionised plasma was able to transform into a neutral gas of atoms – this transition is known as recombination. As noted in Chapter 8, in the context of the early Universe this term is misleading, because it is the *first* time that atoms form.

- By analogy with the process of deuterium synthesis, what quantity other than the mean temperature will influence when recombination takes place?

- The baryon-to-photon ratio, η. If there are large quantities of photons in a particular environment then the high-energy tail of the photon distribution can still photoionise atoms. The importance of these rarer high-energy photons for the recombination process depends on the overall number of photons relative to baryons.

The balance between ionised and neutral gas is described by the Saha equation:

$$\frac{n_{\mathrm{H}}}{n_{\mathrm{p}} n_{\mathrm{e}}} = \left(\frac{m_{\mathrm{e}} k_{\mathrm{B}} T}{2\pi\hbar^2}\right)^{-3/2} \exp\left(\frac{Q}{k_{\mathrm{B}} T}\right) \tag{9.1}$$

where Q is the ionisation energy for hydrogen.

- What simplifying assumption does Equation 9.1 make about the composition of the gas?

- That the gas is purely composed of hydrogen.

While we know from the previous chapter that deuterium, helium and lithium were also be present at this point in the history of the Universe, it is sufficient to make the simplifying assumption of a hydrogen-only gas in order to get a first estimate of recombination conditions.

- In a hypothetical universe with a significantly lower value of η, would recombination occur at an earlier or later point in time?

- If there were fewer baryons per photon, then the high-energy photons in the tail of the CMB distribution would have a proportionally greater ionising effect at the same temperature. Recombination would therefore be delayed until the mean temperature was lower.

The following example investigates this dependence in more detail.

Example 9.1

Rearrange the Saha equation to show that the ionisation fraction of the Universe

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

depends on η as well as temperature.

Solution

We start by using the result of Example 1.2 from Chapter 1:

$$\frac{1 - X}{X} = n_{\mathrm{p}} \left(\frac{m_{\mathrm{e}} k_{\mathrm{B}} T}{2\pi\hbar^2}\right)^{-3/2} \exp\left(\frac{Q}{k_{\mathrm{B}} T}\right) \tag{9.2}$$

Other than temperature, the unknown quantity in this equation is the proton number density, n_{p}. We need to eliminate n_{p} by relating it to the quantity of interest, η.

Since $\eta = n_\mathrm{b}/n_\gamma$ and $X = n_\mathrm{p}/n_\mathrm{b}$, it follows that:

$$\eta = \frac{1}{X}\frac{n_\mathrm{p}}{n_\gamma} \tag{9.3}$$

Rearranging for n_p gives:

$$n_\mathrm{p} = \eta X n_\gamma \tag{9.4}$$

and substituting this into Equation 9.2 gives:

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

and so:

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

We now have an expression that relates ionisaton fraction to η, T and n_γ.

But, as you saw in Example 8.2 in Chapter 8, the photon number density depends only on T – it is the photon energy density $\epsilon = aT^4$ divided by the mean photon energy ($\approx 3k_\mathrm{B}T$) – and so the ionisation fraction expression can be written as:

$$\frac{1-X}{X^2} = \eta\left(\frac{aT^3}{3k_\mathrm{B}}\right)\left(\frac{m_\mathrm{e} k_\mathrm{B} T}{2\pi\hbar^2}\right)^{-3/2} \exp\left(\frac{Q}{k_\mathrm{B}T}\right) \tag{9.5}$$

Note that in this context the symbol a denotes the radiation constant rather than the scale factor.

We have therefore shown that the ionisation fraction depends only on η and T (albeit in an algebraically complicated way!)

Recombination does not take place instantly, but a simple estimate for the time of recombination is to take the point at which half of the baryons are contained within atoms (i.e. $X = 0.5$).

Exercise 9.1

Show that if the baryon-to-photon ratio $\eta = 10^{-9}$, then recombination can take place when the temperature is around 3795 K. (*Hint*: you do not need to solve for T to demonstrate this.)

It is possible to estimate the redshift of recombination by assuming a relation between T and the scale factor, a. T is proportional to a^{-1} and so $T \propto 1 + z$.

- Given the recombination temperature of $T_\mathrm{rec} \approx 3795\,\mathrm{K}$, estimate the redshift at which recombination occurred. (You can assume that the present-day CMB temperature is $T_0 \approx 2.7\,\mathrm{K}$.)
- Using the same approach as Exercise 1.4 in Chapter 1 we can substitute values into

  $$\frac{T_\mathrm{rec}}{T_0} = \frac{1+z_\mathrm{rec}}{1+z_0}$$

  to find that $z_\mathrm{rec} \approx 1400$.

More sophisticated approaches to this calculation account for the additional ionisation caused by photons released in the recombination process, and for recombination to excited states of hydrogen. These lead to slightly lower estimates of $z_{\rm rec}$.

9.1.2 Decoupling of matter and radiation

We have now reached the point in the timeline of the early Universe where theory starts to connect directly to observations. Soon after recombination comes the point at which photons decouple from the baryonic gas and are able to travel unimpeded to our telescopes, i.e. the time of last scattering, when the Universe becomes transparent to radiation.

- In Chapter 6 you read about the concept of last scattering: the point at which a typical photon last undergoes Thomson scattering (i.e. interacts with a free electron). How would you expect the process of recombination to influence the scattering of photons?
- Recombination binds free electrons to atoms, and so greatly reduces the number of free electrons available to scatter photons.

Recombination therefore dramatically reduced the rate of Thomson scattering, so decoupling of matter and radiation occurred fairly soon after this point in time. Figure 9.1 shows these final two stages of early Universe particle evolution.

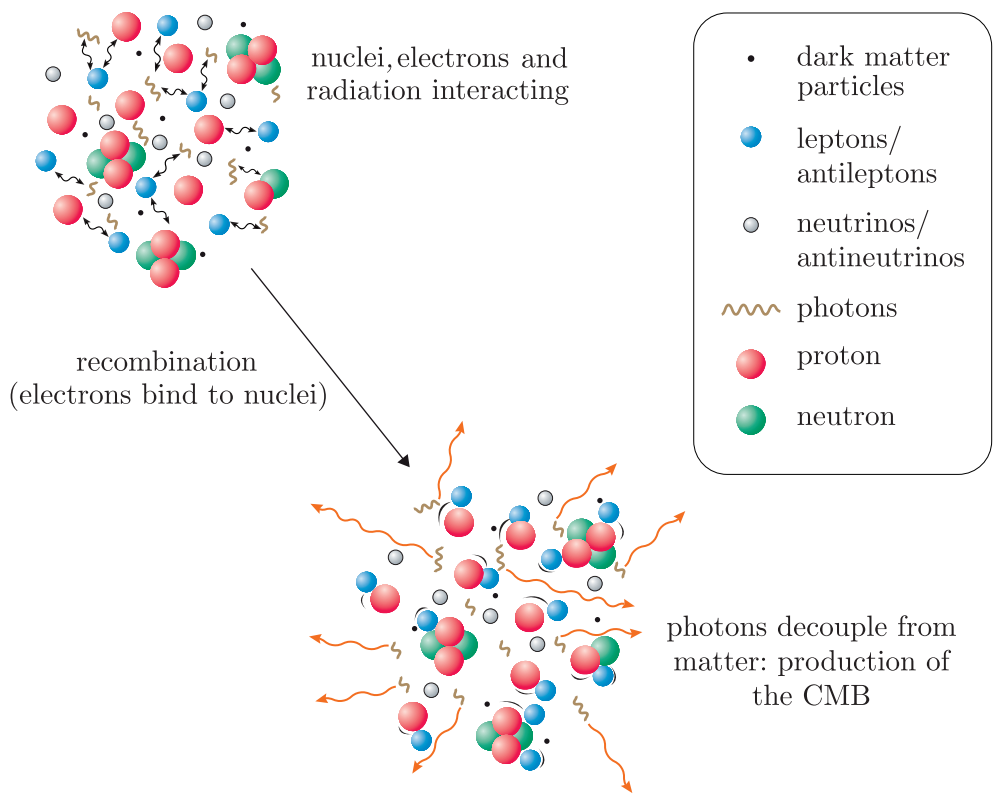


Figure 9.1 The evolution of particles and their interactions around the time of CMB production.

Decoupling is defined to be the point when the rate of photon–electron interactions drops below the rate at which the Universe expands, and is effectively the same point in cosmic history as the epoch of last scattering ($z_{\rm ls}$). The situation is similar to the neutrino decoupling we considered in Chapter 8, and we can use the same approach of comparing the interaction rate, Γ, with the Hubble parameter, $H(t)$, explained in Section 8.1.1.

The interaction rate for Thomson scattering is given by:

$$\Gamma_{\rm e} = n_{\rm e}\sigma_{\rm T}c \tag{9.6}$$

where $\sigma_{\rm T}$ is the Thomson cross-section.

The electron number density is highly dependent on the ionisation fraction, and so it changes with redshift. $\Gamma_{\rm e}$ can be related to the present-day baryon number density $n_{\rm b,0}$ via:

$$\Gamma_{\rm e} = X(z)(1+z)^3 n_{\rm b,0}\sigma_{\rm T}c \tag{9.7}$$

The Hubble parameter can also be expressed as a function of redshift instead of time, as follows:

$$H(z) = H_0\sqrt{\Omega_{\rm m,0}(1+z)^3} \tag{9.8}$$

and so equating $\Gamma_{\rm e}(z)$ and $H(z)$ and rearranging leads to the following expression for the redshift of decoupling (or last scattering):

$$z_{\rm ls} = \left[\frac{H_0\sqrt{\Omega_{\rm m,0}}}{X(z)n_{\rm b,0}\sigma_{\rm T}c}\right]^{2/3} - 1 \tag{9.9}$$

Equation 9.9 shows that it is necessary to know how the ionisation fraction depends on z to estimate the redshift at which decoupling takes place. (This makes sense, because interactions of photons with electrons are the most important interaction for how the radiation propagates.) The example below uses tabulated values of $X(z)$ to estimate $z_{\rm ls}$.

Example 9.2

Table 9.1 lists the ionisation fraction $X(z)$ for different redshift values following the time of recombination, obtained from a modern recombination modelling code that accounts for the additional corrections mentioned at the end of Section 9.1.1.

Table 9.1 Tabulated values of hydrogen ionisation fraction $X(z)$.

z	$X(z)$
900	6.3×10^{-4}
1000	2.5×10^{-3}
1100	6.8×10^{-3}
1200	3.5×10^{-2}
1300	7.9×10^{-2}
1400	2.2×10^{-1}

Use this table, and Equation 9.9, to estimate the redshift of decoupling. Assume the present-day baryon density is $0.249\,{\rm m}^{-3}$.

Solution

All of the parameters in Equation 9.9 apart from $X(z)$ are known, and redshift-independent. Since we only have tabulated values and not an expression to describe $X(z)$ we can't solve algebraically for z.

Instead we must substitute values of X into the equation, and see if there is a value that gives a self-consistent result. In other words, for each row in the table we can evaluate the right-hand side of Equation 9.9 and compare the resulting redshift to the corresponding value in that row of the table.

Using the *Planck* 2018 parameter values of $H_0 = 67.7\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (which can be expressed as $2.19 \times 10^{-18}\,\mathrm{s^{-1}}$ in SI units) and $\Omega_{\mathrm{m},0} = 0.3097$, and the provided value of $n_{\mathrm{b},0}$, we can now evaluate z_{ls} for each row of the table and compare it to the input value.

It may be most efficient to write a short Python script to calculate the results, but this could also be done by hand. The results are shown in Table 9.2, rounded to two significant figures.

Table 9.2 z_{ls} calculation results.

z	$X(z)$	z_{dec} (output)
900	6.3×10^{-4}	5300
1000	2.5×10^{-3}	2100
1100	6.8×10^{-3}	1100
1200	3.5×10^{-2}	370
1300	7.9×10^{-2}	210
1400	2.2×10^{-1}	110

It is clear from Table 9.2 that the calculation does not give self-consistent output at low and high redshifts. But a redshift of $\sim$1100 *does* give a self-consistent result: the input and output redshifts agree to a precision of two significant figures, and so this is the best estimate for z_{ls}.

The relatively crude approach we have taken here has shown that decoupling takes place at a redshift of $\sim$1100 – a result very close to the accepted value for the redshift of CMB production, as quoted in Chapter 6, of $z_{\mathrm{ls}} \sim 1090$. The example has also shown that the physics of decoupling, and therefore the precise timing of the production of the CMB, depends strongly on how recombination proceeded. This brings us back once again to the crucial importance of the baryon-to-photon ratio (which helps determine the time of recombination) for cosmology.

9.2 Measuring primordial abundances

Elemental abundances are a fundamental tool in astronomy. All of the elements in the present-day Universe were either synthesised in the big bang or are produced by the evolution of stars.

In the previous chapter you saw that a small number of elements were synthesised in the very early Universe, before the temperature dropped below that needed for further fusion. You also saw that the fusion processes were highly sensitive to the baryon-to-photon ratio, η, and the neutron-to-proton ratio at early times. If the theory we have presented is correct, the start and end times for big bang nucleosynthesis – and so the yields for different elements – depend mainly on the value of η. Hence measuring elemental abundances in astrophysical environments provides a useful cosmological test, which is very complementary to other measurements, such as CMB anisotropies.

- Which cosmological density parameter is closely related to the baryon-to-photon ratio?
- The baryon density parameter, Ω_b, is closely related to η.

In this section we will explore how observations can be used to measure the abundances of the elements produced in the early Universe. We will also examine how these measurements can be used to confirm the theories of early-Universe physics presented in the last few chapters.

9.2.1 Abundances and metallicity

The majority of measurements of elemental abundances come from the use of **spectroscopy**, whereby the light from stars, gas or other sources of emission is measured as a function of its wavelength or frequency. This method allows us to observe emission or absorption lines caused by the presence of particular elements, as shown in Figure 9.2 for a relatively nearby galaxy.

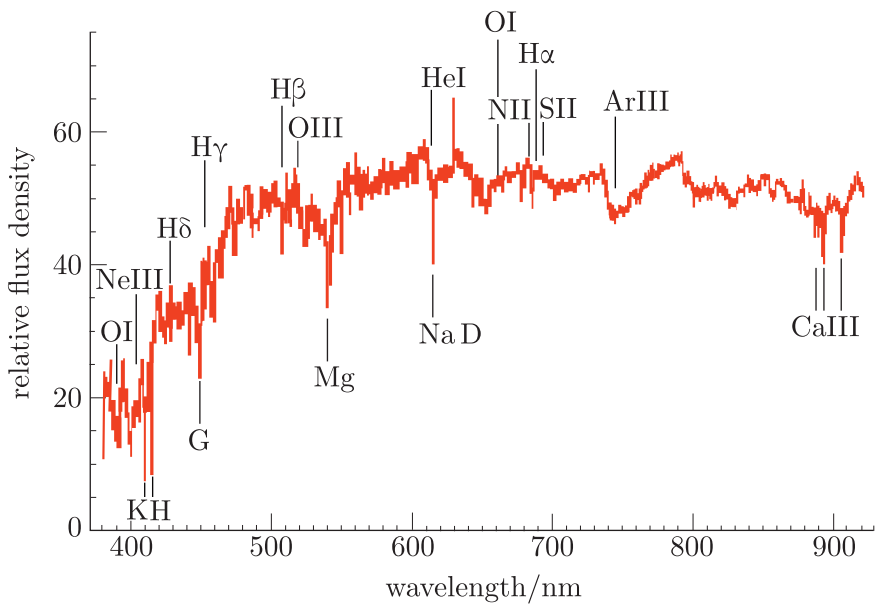


Figure 9.2 An optical spectrum of a galaxy, observed as part of the Sloan Digital Sky Survey. Several emission and absorption features are labelled.

Measuring the strength of spectral line features for galaxies and gas clouds enables the **relative abundance** of specific elements to be determined. For example, the abundance of oxygen relative to hydrogen is commonly used as a proxy for the total amount of **chemical enrichment** (the increase in proportion of elements heavier than helium) that has taken place due to the synthesis of new elements in stars.

Elemental abundances can be reported in several different ways. The simplest abundance measure is the mass of a given element relative to the mass of hydrogen within the same sample (e.g. a gas cloud from which a spectrum has been measured). In this book we will write abundance ratios of this sort as a ratio of the two species, so ^{4}He/H is the mass ratio of helium-4 relative to hydrogen.

It is also common to make use of abundances defined relative to those observed in the spectrum of the Sun because these are very well determined. This relationship is defined as

$$[\mathrm{O/H}] = \log_{10}\left(\frac{n_\mathrm{O}}{n_\mathrm{H}}\right) - \log_{10}\left(\frac{n_{\mathrm{O},\odot}}{n_{\mathrm{H},\odot}}\right) \tag{9.10}$$

where $n_\mathrm{O}/n_\mathrm{H}$ is the ratio of the oxygen and hydrogen number densities measured in the same object of interest, and the right-hand log term is the same ratio measured for the Sun. Oxygen can be replaced by any other element in this expression. The square brackets indicate that we are describing a logarithmic abundance ratio by number density.

Abundance ratios of heavy elements, such as oxygen and iron, are related to a quantity known as **metallicity**, Z. This describes the proportion of baryonic matter (by mass) that is made up of elements other than hydrogen and helium. Such elements are referred to as 'metals' in the context of astronomy, and changing metallicity in stars and galaxies is used to track star formation throughout cosmic history. Metallicity is defined as:

$$Z = 1 - (X + Y) \tag{9.11}$$

where X and Y are the mass fractions of hydrogen and helium, respectively, for a given environment.

The abundance ratios [O/H] and [Fe/H] are often used to measure Z, so metallicity is sometimes used loosely to refer to one of these observed abundance ratios. A quantity such as [O/H] can be related to Z by assuming that oxygen makes up a specific fraction of the metals, for example the same fraction that it does in the Sun.

The use of abundances and metallicity is widespread in extragalactic astronomy, where it is a powerful way to investigate the nuclear reactions that have historically taken place as a result of star formation in a particular region of the Universe (e.g. within the Milky Way). But we are not able to directly observe and measure spectral features for regions in the period immediately after the end of nucleosynthesis, and abundance measurements for a present-day galaxy may not reflect the primordial abundances created immediately after the big bang. So how can abundance measurements be used in cosmology?

The answer is that to infer early-Universe abundances, great care needs to be taken to choose the best locations to make measurements, and then considerable thought is needed to interpret them using all of the information available. In the next sections you will meet some specific examples of how this is done.

9.2.2 Deuterium abundances

Deuterium was the first element to be synthesised in the early Universe. Its abundance provides an important cosmological test, because its production was so sensitive to the relative proportion of energetic photons available to dissociate nuclei as they formed (i.e. η). Figure 9.3 shows a prediction for how the cosmic abundance of deuterium (D/H) depends on η.

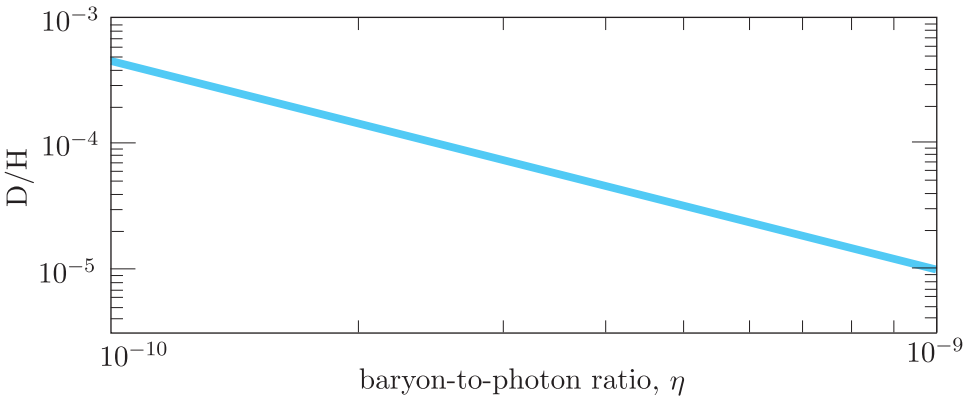


Figure 9.3 The dependence of deuterium abundance on η (after Cyburt *et al.*, 2016).

Deuterium is the primordial element whose abundance can be measured most reliably because it was *not* produced in significant quantities after the big bang nucleosynthesis (BBN) period. It can only decrease via **astration**, where it is destroyed via processes occurring inside stars. The quantity present in the Universe today must therefore be a lower limit on the amount that was produced during the BBN period.

Drawing conclusions about early-Universe abundance is harder for other elements, where both production and destruction may have taken place at later times.

Deuterium has been measured in several different astrophysical locations over many decades. The two main environments in which it is measured are the local interstellar medium (i.e. regions within the Milky Way) and the environments of very distant quasars. The latter are bright, distant galaxies whose central black hole is rapidly accreting gas, making them very luminous and so visible out to large distances. Absorption features in their spectra reveal the abundance of intervening intergalactic gas.

Figure 9.4, taken from a study by Cooke *et al.* (2014), plots a compilation of deuterium abundance measurements obtained from spectra of high-redshift quasars. It shows that the values along the paths travelled by light from different quasars cluster quite closely around an abundance value of $\log_{10}(\mathrm{D/H}) = -4.6$. The coloured band indicates the predicted deuterium abundance based on the *Planck* measurement of $\Omega_{\mathrm{b},0}$.

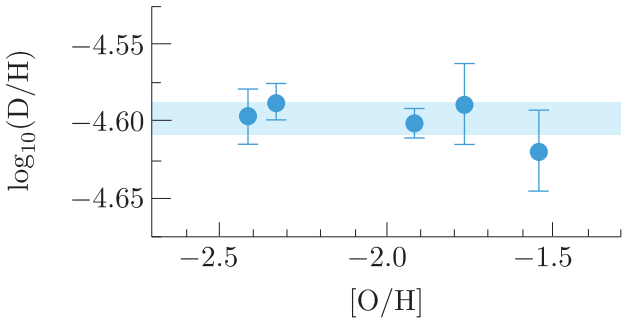


Figure 9.4 Observational measurements of deuterium abundance from the spectra of high-redshift quasars (blue data points) relative to the predicted abundance based on *Planck* measurements (blue band).

The next example considers another set of abundance measurements determined by observing objects within the Milky Way.

Example 9.3

Table 9.3 lists a set of D/H ratios measured from absorption features in the ultraviolet (UV) spectra of stars from different regions of the Milky Way.

Table 9.3 Deuterium abundance measurements from UV sightlines to selected Milky Way stars, with uncertainties.

Stellar sightline	**D/H ($\times 10^{-6}$)**
BD +28 4211	13.9 ± 1.0
BD +39 3226	11.7 ± 3.1
δ Ori	7.4 ± 1.0
Feige 110	21.4 ± 4.8
G191−B2B	16.6 ± 4.1
GD 246	15.1 ± 1.9
γ Cas	9.8 ± 2.5
HD 191877	7.8 ± 2.0
HD 195965	8.5 ± 1.6
HZ 43	16.6 ± 1.4
ι Ori	14.1 ± 2.8
LSS 1274	7.6 ± 1.9
TD1 32709	18.6 ± 5.3
WD 1034+001	21.4 ± 5.3
WD 1634−573	15.8 ± 2.5

(Adapted from Linsky *et al.*, 2006, p. 1118.)

Use these data, and the additional mathematical information provided below, to answer the following questions.

(a) Calculate the mean and median values of D/H.

(b) Next, determine the uncertainty on the mean.

(c) Finally, comment on how these measurements compare with the quasar absorption measurements shown in Figure 9.4, for which the mean value is $\mathrm{D/H} = (2.52 \pm 0.02) \times 10^{-5}$.

Recall that the uncertainty in the mean value for a set of measurements is given by the **standard error on the mean**, $\sigma_{\rm m}$, which is related to the **standard deviation**, σ, via:

$$\sigma_{\rm m} = \frac{\sigma}{\sqrt{N}} \tag{9.12}$$

where N is the number of measurements.

The standard deviation, σ reflects the spread of values measured, and is defined as:

$$\sigma^2 = \frac{1}{N}\sum_i (x_i - \langle x \rangle)^2 \tag{9.13}$$

where x_i are the measured quantities and $\langle x \rangle$ is the mean of the x_i values.

Solution

One of the simplest ways to do these calculations is to write a short Python script that reads the list of numbers into an array, and then uses functions in the numpy library (e.g. `np.mean`, `np.median`, `np.std`) to do the calculation. However, we will work through the calculations by hand here.

(a) Summing the values – ignoring uncertainties – and diving by the number of measurements gives a mean value of 1.38×10^{-5}. The median is 1.41×10^{-5}. The fact that the mean and median are very similar provides useful reassurance that the mean is a good approximation to the centre of the distribution of values.

(b) To determine the uncertainty we must first calculate σ, and then $\sigma_{\rm m}$.

Calculating the standard deviation manually requires determining $x_i - \langle x \rangle$ for each row in Table 9.3 (where $\langle x \rangle = 1.38 \times 10^{-5}$ as calculated above), and then applying Equation 9.13. Whether you do this by hand or use `numpy.std`, you should obtain a value of $\sigma = 4.6 \times 10^{-6}$.

Since there are 15 measurements, this means that $\sigma_{\rm m} = 1.2 \times 10^{-6}$.

(c) We can now make a useful comparison with the quasar measurements in Figure 9.4. The mean value and uncertainty are given in the question as 2.52×10^{-5} and 2×10^{-7}, respectively.

Therefore, accounting for the uncertainty on the two mean values, the mean abundance estimate from the Milky Way (1.38×10^{-5}) is inconsistent with the quasar absorption measurements. The Milky Way deuterium content therefore appears to be significantly lower than that in the environments of high-redshift quasars.

- Which location – the Milky Way or the environments of quasars – is likely to contain gas whose composition better reflects the primordial Universe?

- Distant galaxies are being observed at an earlier time in the history of the Universe, because of the time taken for their light to reach us. It would therefore be expected that their environments will have been less affected by processes that alter the gas composition, since they will have operated over a much shorter time period.

Exercise 9.2

Use Figure 9.3 to find the values of η corresponding to the measured deuterium abundance determined via: (i) the high-redshift quasar measurements that you saw in Figure 9.4; (ii) the analysis of the Milky Way measurements in Example 9.3.

Calculate the corresponding value of $\Omega_{\mathrm{b},0}$ for each one, assuming the mean particle mass is m_{p} and the present-day photon number density is $n_{\gamma,0} = 4.0 \times 10^8\,\mathrm{m}^{-3}$.

As the previous example and exercise show, the quasar measurements of D/H provide a tighter constraint than measurements of Milky Way regions, and they also result in a value of η that is in better agreement with the baryon density measured from the CMB. Since we already have grounds to expect the high-redshift measurements to be more reliable, the consistency between the D/H abundances and CMB measurements is a very impressive success of big bang theory. It seems likely that the Milky Way deuterium abundance has been depleted by processes occurring at more recent times.

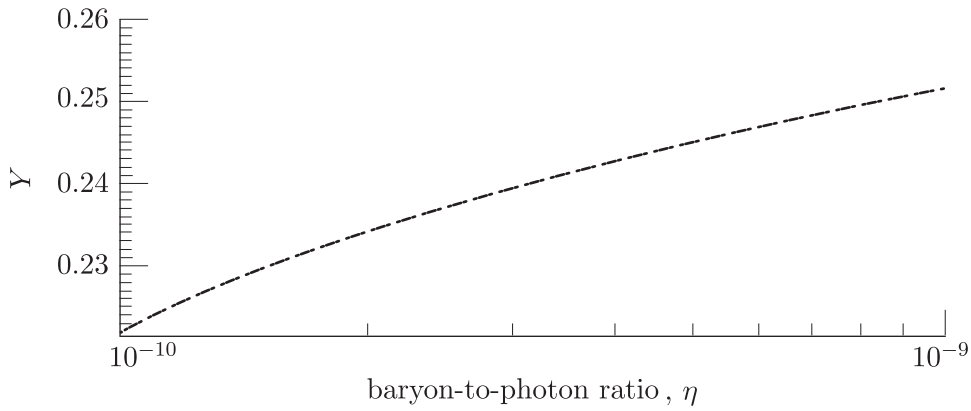
Another interesting route to measuring primordial abundances is to look within the Solar System. The abundance of deuterium in the atmospheres of Jupiter and Saturn has been measured to have a value of $\mathrm{D/H} \sim 2 \times 10^{-5}$ (Lellouch *et al.*, 2001). Since no nuclear synthesis takes place in planetary atmospheres, these values should reflect the elemental abundances in the Universe at the time the Solar System formed, $\sim 4.5 \times 10^9$ years (or 4.5 gigayears, Gy) ago. The lookback time to galaxies at a redshift of ~ 3 (assuming *Planck*-measured cosmological parameters) is $\sim 12\,\mathrm{Gy}$, so the Solar System and quasar deuterium abundance measurements also appear consistent, assuming a fraction of the deuterium has been astrated in the time prior to Solar System formation.

9.2.3 Helium abundances

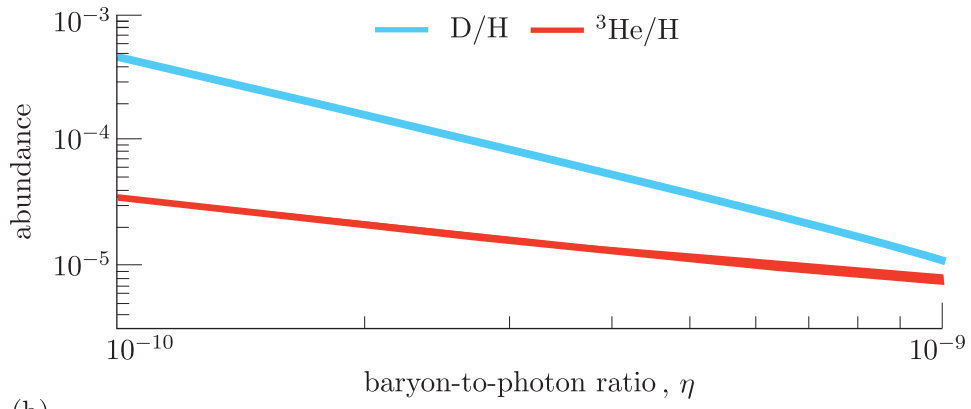
Helium fusion can only take place once the deuterium bottleneck is overcome, so the abundance of helium can also be used to measure η and Ω_{b}.

- Which two forms of helium are produced via BBN, and how do they differ?
- Both ^{3}He (helium-3) and ^{4}He (helium-4) are created.

Figure 9.5 shows the dependence of both of these helium isotopes on η, with D/H shown for comparison. Both ^{4}He (shown as Y) and D/H have a strong dependence on the baryon-to-photon ratio, whereas ^{3}He is less sensitive to it.



(a)



(b)

Figure 9.5 The dependence of (a) 4-helium abundance (expressed as the mass fraction, Y) and (b) of 3-helium abundance on η, with deuterium abundance (D/H) shown for comparison (after Cyburt *et al.*, 2016).

The next exercise invites you to consider further the interpretation of Figure 9.5.

Exercise 9.3

Explain why the abundance of helium-4 is expected to *increase* for higher values of η, whereas the abundance of deuterium *decreases*, as shown in Figure 9.5.

Measuring helium abundances is therefore, in principle, an independent cosmological test that can provide different information from the

You may wonder why ^{4}He is expressed as a mass fraction, but the other isotopes as abundance ratios. Y is an intuitive quantity, as it tells us that $\sim$25% of matter is in the form of ^{4}He. The mass fractions for deuterium and ^{3}He are so much lower that it makes more sense to think about ratios of numbers of particles.

deuterium measurement. Helium abundance is also important for the conversion between η and $\Omega_{b,0}$, because this calculation depends on the mean particle mass at present times. It is harder to obtain reliable measurements of the primordial abundance for helium than for deuterium, because the importance of helium production within stars means that its abundance has changed a lot over the history of the Universe. However, it is still possible to achieve interesting and useful results that are helpful for cosmology by identifying regions to study where relatively little star formation has taken place.

- How can regions with a history of low star formation be identified?
- The metallicity of galaxies and extragalactic gas clouds traces the amount of star formation that has taken place. Low-metallicity regions can be identified by their spectra, and should correspond to locations where the gas has been least processed by stars.

Figure 9.6 shows a compilation of helium abundance measurements taken from low-metallicity extragalactic **HII regions** – these are regions of ionised hydrogen gas in distant galaxies.

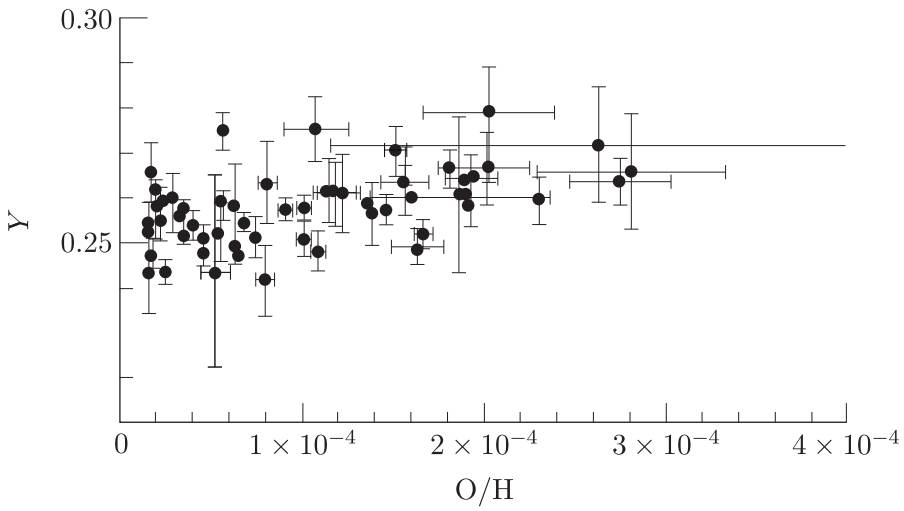


Figure 9.6 Measurements of helium abundance (Y) from extragalactic HII regions plotted against relative oxygen abundance for those regions, which is a measure of their metallicity.

You might notice that the amount of scatter in the helium abundances is much larger than for the quasar deuterium measurements in the previous section. This is not surprising given the effect of stellar processing on helium. However, it is possible to use the slight dependence of abundance on metallicity to extrapolate a value for a metallicity of zero, which would correspond to the primordial helium abundance. Extrapolation of these data suggests an Y abundance of ~ 0.25 at that time.

9.2.4 Lithium abundance

Measurements of cosmic lithium abundance have presented a challenge for cosmologists for a number of years. Lithium is the only element in the Universe that was produced in the big bang, but is also produced in stars *and* via interactions with cosmic rays within the interstellar medium. This makes it challenging to interpret observational measurements.

A number of attempts have been made to relate lithium abundance to metallicity, in order to be able to extrapolate back to a primordial value. There have been conflicting results on how this trend behaves, but one example from Sbordone *et al.* (2010) is shown in Figure 9.7.

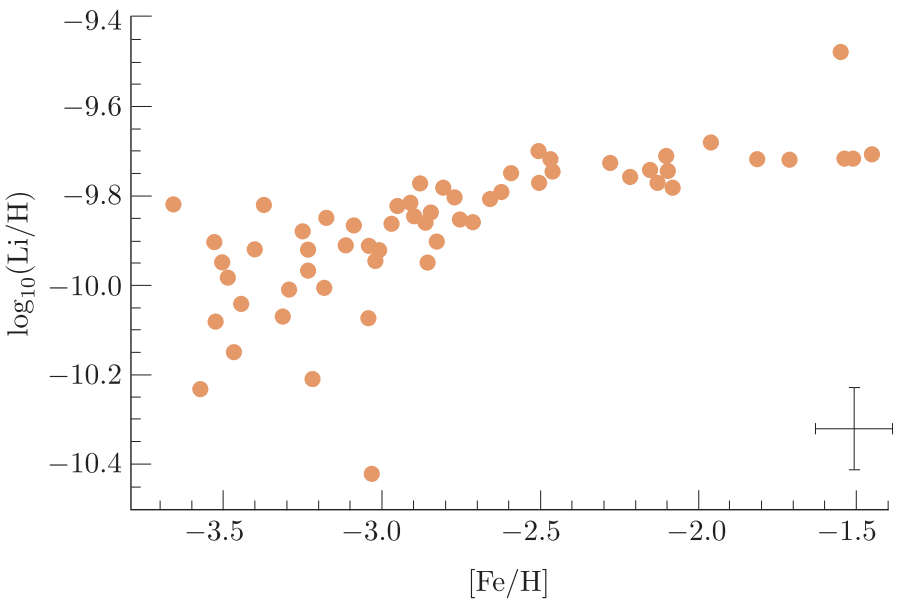


Figure 9.7 Observational measurements of lithium-7 abundance (plotted as a logarithmic abundance) compared to metallicity, in this case measured via iron abundance. The cross indicates the typical size of measurement uncertainties.

Unfortunately there is a lot of scatter in lithium abundance measurements at low metallicity compared with high metallicity. But a wide range of studies have reached the conclusion that – unlike the well-determined results for deuterium and helium – the primordial abundance of lithium-7 appears to be somewhat *lower* than that predicted by BBN, which is plotted in Figure 9.8.

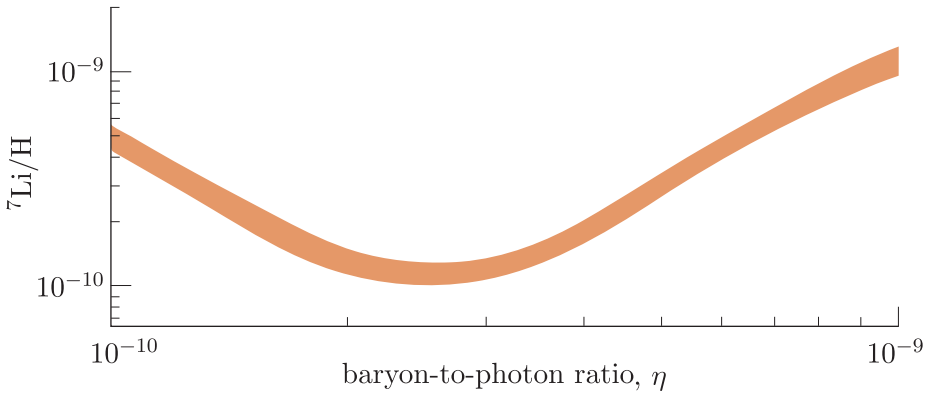


Figure 9.8 The dependence of lithium-7 abundance on η (after Cyburt *et al.*, 2016).

- What is the predicted value of the lithium-7 abundance for the *Planck*-measured value of $\eta = 6.1 \times 10^{-10}$, and how does this compare to the measurements in Figure 9.7?
- For $\eta = 6.1 \times 10^{-10}$, the predicted abundance inferred from the plot is $(^7\text{Li/H}) \approx 4.5 \times 10^{-10}$. To compare with the observed values shown in Figure 9.7 we need to take the logarithm to obtain $\log_{10}(^7\text{Li/H}) = -9.3$. The prediction is therefore considerably higher than the observed measurements.

The most widely accepted explanation for this result is that there is a process taking place in which stars can destroy lithium in their outer atmospheres. An alternative is that some aspect of early Universe physics affecting only the final stage of nucleosynthesis is not yet understood. Although this is a long-standing uncertainty, the potential explanations are not thought to suggest any major problem in our overall understanding of early Universe physics.

9.2.5 The cosmological significance of Ω_b

The previous sections have demonstrated that, although they can be challenging to interpret, observed measurements of elemental abundances provide important confirmation of our understanding of early Universe physics, and specifically the baryon-to-photon ratio. You have seen that observational determinations of $\Omega_{\text{b},0}$ from elemental abundances conclude that its value is ~ 0.05, with a similar value now resulting from CMB observations.

- What do the observational estimates of $\Omega_{\text{b},0}$ imply for the influence of (ordinary) baryonic matter on the expansion of the Universe, if we assume that the Universe is spatially flat (i.e. $k = 0$)?
- A flat Universe has $\Omega = 1$, so if $\Omega_{\text{b},0} \ll \Omega$ then this implies that other elements, such as dark matter and/or dark energy, must play the dominant role in controlling the expansion of spacetime in the present-day Universe.

Elemental abundance measurements have been very important in the historical development of cosmology, because they provide a key piece of evidence that there is not enough baryonic matter in the Universe to explain its rate of expansion. The next section therefore considers the need for non-baryonic matter.

9.3 Measuring non-baryonic matter

The primordial abundances of helium and deuterium discussed in the previous section are part of a range of important clues that point to the existence of some form of dark matter. We were able to neglect the role of dark matter in our investigation of how nuclei and atoms form, but you have already seen that it is part of the story of how the CMB originated, and may have noticed its inclusion in the pictorial summary of early-Universe evolution in Figure 8.5. Its role becomes especially important in the next part of the history of the Universe, in which matter gradually assembles to form stars and galaxies.

In this section we will consider the variety of evidence that has led to the current consensus that dark matter must be an important constituent of the Universe. We will start by examining how observations of nearby galaxies and galaxy clusters provided the first evidence for dark matter, and now provide very tight constraints on how any form of dark matter (or alternative theory) must behave. We will then return to the topic of the CMB and discuss how its anisotropies, together with modern theories of structure formation, point firmly to the existence of dark matter as an explanation for how galaxies assemble.

9.3.1 Weighing galaxy clusters

Galaxy clusters are groups of large numbers of galaxies that comprise the largest gravitationally bound structures in the Universe. Both their brightness and the fact that they contain a representative sample of the matter distribution throughout the Universe mean that they provide ideal 'laboratories' to understand the composition of the Universe.

Figure 9.9 shows two of the main constituents of galaxy clusters for the same object, Abell 1689. Panel (a) is an optical image of the individual galaxies that are members of this cluster, with all their constituent stars, whereas panel (b) is an X-ray image of a large quantity of hot, diffuse gas spread out in between the galaxies, which has been heated to X-ray-emitting temperatures through falling into the gravitational potential well of the cluster itself.

One of the first astronomers to estimate the total mass contained within a galaxy cluster was Fritz Zwicky, who, in 1933, measured the velocities of the galaxies in the Coma cluster. The galaxies in a cluster are expected to orbit the cluster's centre of mass with velocities that depend on their mass (due to mutual gravitational attraction). However, Zwicky found that the

(a)

(b)

Figure 9.9 The galaxy cluster Abell 1689 viewed at (a) optical and (b) X-ray wavelengths.

range of velocities of the galaxies was considerably larger than expected. He used the **virial theorem** – a fundamental theorem, of wide use in astronomy – to calculate the total mass in the galaxy cluster and compared this with the mass contained in the individual galaxies and their stars.

Virial theorem

For an assembly of particles in stable equilibrium, the total gravitational potential energy (E_g) due to mutual attraction is related to the total kinetic energy (E_k) according to:

$$2E_k + E_g = 0 \tag{9.14}$$

In other words, the magnitude of the total kinetic energy is half that of the gravitational potential energy.

To apply the virial theorem to the comparison of total gravitational potential and kinetic energies in a galaxy cluster we need to be able to sum over particular properties of the cluster at different distances from the centre. This analysis requires the use of a volume integral, which you can read more about in the box that follows.

Volume integrals

It is common in astrophysics to encounter situations where we want to evaluate the sum of a quantity that varies over a volume or a surface. If we can model a star, a star cluster or a galaxy cluster as spherical, this means we can integrate that quantity over a range of radii, from $r = 0$ (the centre of the sphere) to $r = R$ (the outer radius).

Consider a gas cloud whose density ρ depends on radius, according to $\rho = \rho_0\left[1 + (r/r_0)^{-2}\right]$, where r_0 is a constant representing a characteristic radius. To determine the total mass of that gas cloud we need to account for the variation of density with radius. We can therefore divide the sphere into a series of concentric spherical shells, each with its own density $\rho(r)$. The total mass of the gas cloud is the sum of all of the shells, which – if we make the shell width infinitesimally small – is given by:

$$M = \int_0^R \rho(r)\,\mathrm{d}V$$

where $\mathrm{d}V$ is the volume of a shell, calculated as a spherical surface area $4\pi r^2$ multiplied by a shell thickness $\mathrm{d}r$, so that:

$$M = \int_0^R 4\pi r^2 \rho(r)\,\mathrm{d}r$$

Therefore, for the particular density distribution given,

$$M = \int_0^R 4\pi r^2 \rho_0 \left[1 + (r/r_0)^{-2}\right] \, \mathrm{d}r = 4\pi\rho_0 r_0^2 \int_0^R \left[(r/r_0)^2 + 1\right] \, \mathrm{d}r$$

which evaluates to $M = (4/3)\pi\rho_0 r_0^3 \left[(R/r_0)^3 + 3R/r_0\right]$.

We can use the virial theorem to estimate the total mass of a galaxy cluster, because both the gravitational potential energy and the kinetic energy depend on the mass. If m_i is the mass of an individual galaxy (i) within the cluster, and v_i is its time-averaged velocity, then the total time-averaged kinetic energy of the galaxy is given by:

$$E_\mathrm{k} = \sum_i \frac{1}{2} m_i v_i^2$$

Of course the sum of all of the m_i values is the total mass, M, and so:

$$E_\mathrm{k} = \frac{1}{2} M \langle v^2 \rangle$$

where $\langle v^2 \rangle$ is the average of the square of the velocities of the individual galaxies. We can't directly measure the overall galaxy velocities, but we can relate it to the **velocity dispersion**, σ_v, which is defined as the typical galaxy speed (magnitude of the velocity) in the radial direction. (We are able to use the Doppler shifting of emission lines in spectra to determine this component of velocity.) The velocity dispersion only measures the velocity in one dimension, but we can assume the distributions are similar in the other two dimensions, which means that $\langle v^2 \rangle \sim 3\sigma_\mathrm{v}^2$. Therefore:

$$E_\mathrm{k} = \frac{3}{2} M \sigma_\mathrm{v}^2 \tag{9.15}$$

To obtain an expression for the total gravitational potential energy we need to consider the distribution of galaxies with radius in the cluster. The simplest assumption to make is that the galaxies are distributed uniformly within a spherical region of radius R, which means that we can define a galaxy density, ρ, that is independent of radius.

We can now use the volume integral approach explained above to sum the gravitational potential energy of a series of spherical shells, each of which spans a narrow range of radius from r to $\mathrm{d}r$. The gravitational potential energy of a spherical shell at radius r is:

$$\mathrm{d}E_\mathrm{g} = -\frac{GM(r)\,\mathrm{d}m}{r}$$

where $M(r)$ is the total mass contained at radii smaller than r and $\mathrm{d}m$ is the mass contained within the particular shell.

We can use the assumption of uniform density to find expressions for both $M(r)$ and $\mathrm{d}m$. The mass within a spherical shell is then given by $4\pi r^2 \mathrm{d}r\rho$, namely the volume of an infinitesimally thin spherical shell multiplied by the density of that region.

Substituting for $M(r)$ and $\mathrm{d}m$ and then integrating over all radii gives:

$$E_{\mathrm{g}} = -\int_0^R \frac{G(4/3)\pi\rho r^3 4\pi r^2 \rho \,\mathrm{d}r}{r} = -\frac{16}{3}\pi^2 G\rho^2 \int_0^R r^4 \,\mathrm{d}r$$

which evaluates to:

$$E_{\mathrm{g}} = -\frac{16\pi^2 G\rho^2 R^5}{15}$$

We require an expression in terms of the total mass M, which is $M(R)$, and so we can substitute back the relationship between ρ and M to find the expression for the total gravitational potential energy of a uniform density sphere:

$$E_{\mathrm{g}} = -\frac{3GM^2}{5R} \tag{9.16}$$

We could now apply the virial theorem to relate our two total energies. However, we started with the simplifying assumption of uniform density, ρ. In reality, we know that galaxy clusters (like stars and galaxies) are centrally concentrated. This does not significantly change the analysis above, but the magnitude of the gravitational potential energy for a centrally concentrated mass distribution is higher by a small factor, so that it is common to neglect the factor of 3/5 in Equation 9.16. In other words, the gravitational potential energy of a more centrally concentrated sphere is taken to be:

$$E_{\mathrm{g}} = -\frac{GM^2}{R} \tag{9.17}$$

Now substituting our expressions for the two total energies into Equation 9.14 and rearranging for M gives a relationship between total mass, velocity dispersion and cluster radius:

$$M = \frac{3\sigma_{\mathrm{v}}^2 R}{G} \tag{9.18}$$

Exercise 9.4

The Coma cluster has a velocity dispersion of $\sigma_{\mathrm{v}} \sim 998\,\mathrm{km\,s^{-1}}$, an estimated radius of $R \sim 3.5\,\mathrm{Mpc}$ and an optical luminosity $L = 5.0 \times 10^{12}\,\mathrm{L_\odot}$.

(a) Estimate the total mass, M, of the Coma cluster.

(b) Calculate the mass-to-light ratio, M/L, of the Coma cluster in units of solar mass divided by solar luminosity.

(c) Comment on what the calculations you've made tell us about the nature of the mass content in Coma. Could all of the mass be in the form of stars?

Exercise 9.4 demonstrates why Zwicky concluded that there must be something else present to explain the motions of the galaxies – clusters appear to have a significant quantity of 'missing mass'.

Further corroboration of the need for dark matter in galaxy clusters comes from the X-ray-emitting hot-gas component shown in Figure 9.9b. The virial theorem does not just apply to galaxies: the hot gas should also obey it, with the kinetic energy of the gas related to the gravitational potential in the same way. Using Equation 9.16 and taking the total kinetic energy of a hot gas to be

$$E_{\mathrm{k}} = \frac{3}{2}\frac{M}{\langle m \rangle} k_{\mathrm{B}} T \tag{9.19}$$

where $\langle m \rangle$ is the mean mass of particles in the gas, the temperature of the intracluster gas should be related to the total mass of a galaxy cluster by:

$$T = \frac{GM\langle m \rangle}{5 k_{\mathrm{B}} R} \tag{9.20}$$

Exercise 9.5

X-ray spectroscopy of galaxy clusters shows that the temperature of the intracluster gas typically ranges from 10^7–10^8 K. Assuming a typical cluster radius of 1 Mpc and a typical particle mass of $\langle m \rangle = 0.6 m_{\mathrm{p}}$, calculate the range of total cluster mass (in units of solar mass) to which this corresponds.

You may wonder whether the mass of the X-ray-emitting hot gas, when added to that of the individual galaxies, could explain the missing mass. Unfortunately this is not the case. Very accurate estimates of the total gas mass in galaxy clusters can be made from X-ray observations, and although it makes up a larger proportion of the cluster mass than the stars within galaxies, the intracluster gas typically only accounts for around 10% of the total mass required by the virial theorem.

Perhaps one of the most compelling arguments for the existence of dark matter comes from multi-wavelength observations of a galaxy cluster known as the Bullet cluster. The Bullet cluster is a system in which two smaller groups of galaxies are merging together to form a single galaxy cluster. (This type of merger process is part of how present-day galaxy clusters assembled.)

Figure 9.10 shows the location of the two groups of galaxies (the circular and ellipse regions embedded in the blue regions) together with the location of the X-ray emission (shown in red). The emission forms two bright regions that are somewhat displaced from the groups of galaxies. The blue regions show the overall distribution of mass in the system, measured by a technique called gravitational lensing, in which the light from background galaxies is distorted as it passes near to the cluster.

Figure 9.10 Optical-light image of galaxies in the Bullet cluster. An overlaid X-ray image (red) and gravitational lensing map (blue) have been used to measure the hot gas and total mass distributions of the cluster, respectively.

The crucial point for interpreting these images is to understand why the X-ray emission is displaced. When galaxy clusters merge, the individual galaxies don't collide with each other, whereas the gas particles in the intracluster medium do. Hence the gas clouds are slowed down as they pass through each other, so that they have been 'left behind' by the galaxies.

If there was no dark matter then the total mass of the cluster would be dominated by the intracluster gas, which weighs much more than the galaxies. But the total mass is located in the same place as the galaxies, *and doesn't coincide with the gas.* This suggests the mass is dominated by dark matter, which – unlike the gas particles, but similarly to the galaxies – is expected to be 'collisionless', because dark matter particles interact only through gravity. Dark matter clumps would be expected to pass through each other and remain with the galaxies. There are no satisfactory explanations for the Bullet cluster observations that don't require the existence of dark matter.

9.3.2 Galaxy rotation curves

Zwicky's investigation of the motions of galaxies in clusters was the first observed evidence for the existence of some form of dark matter. But it was not until new evidence emerged in the 1970s from the rotation curves of spiral galaxies that the idea started to be taken more seriously within the astronomy community. The US astronomers Vera Rubin and Kent Ford made careful observations of the rotation speeds of stars at different locations within spiral galaxies. They found that – as for the orbits of entire galaxies within clusters – stellar orbits do not behave as theory would predict if only the visible stars and gas are present.

If stars are assumed to follow a circular orbit within the disc of the Milky Way, then their motion at a particular location should be described by **Keplerian rotation**, with the magnitude of the velocity given by:

$$v = \sqrt{\frac{GM(<r)}{r}} \tag{9.21}$$

where $M(<r)$ is the mass enclosed within a radius r.

- If most of the mass of a galaxy is concentrated close to the centre, how would you expect the velocity to change in the outer parts of the galaxy?
- If $M(<r)$ tends towards a constant value in the outer regions, then the velocity will decrease as r increases because of the inverse dependence on r.

Figure 9.11 shows an example of a spiral galaxy rotation curve, where the data points in panel (b) are the observed velocities. If there is no dark matter present, then the observed velocities should agree with those predicted from summing the mass of stars and gas (the dotted and dashed lines lines, respectively, in the plot) according to Equation 9.21. This is clearly not the case: the velocity continues to increase in the outer regions.

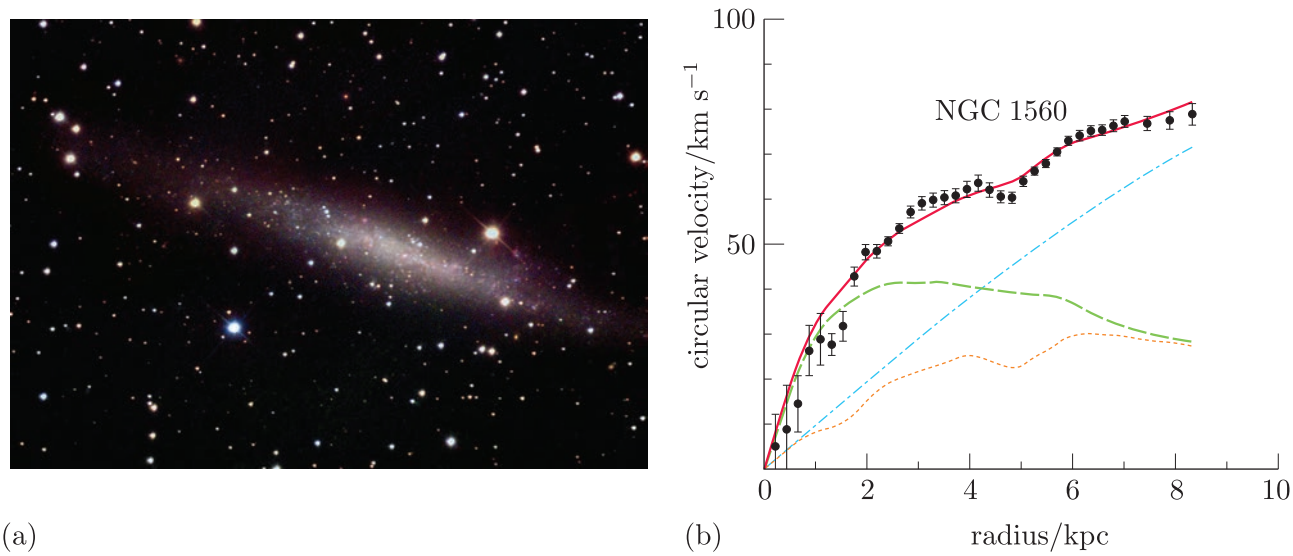


Figure 9.11 (a) An optical image of the approximately edge-on spiral galaxy NGC 1560 and (b) the galaxy's rotation curve. The green long-dashed curve shows the predicted contribution to the velocities from the inferred mass of stars, while the orange short-dashed line shows the contribution from the inferred mass of the gas. These are not enough on their own to explain the observed velocities. The blue dash-dotted line is the dark matter distribution required to make up the remaining mass.

The rotation curves of large samples of spiral galaxies are found to be either flat or increasing at radii where Keplerian theory suggests they should be decreasing. This has even been shown to be true of our own Galaxy. Figure 9.12 shows the Milky Way rotation curve, which is flat to distances of tens of kpc.

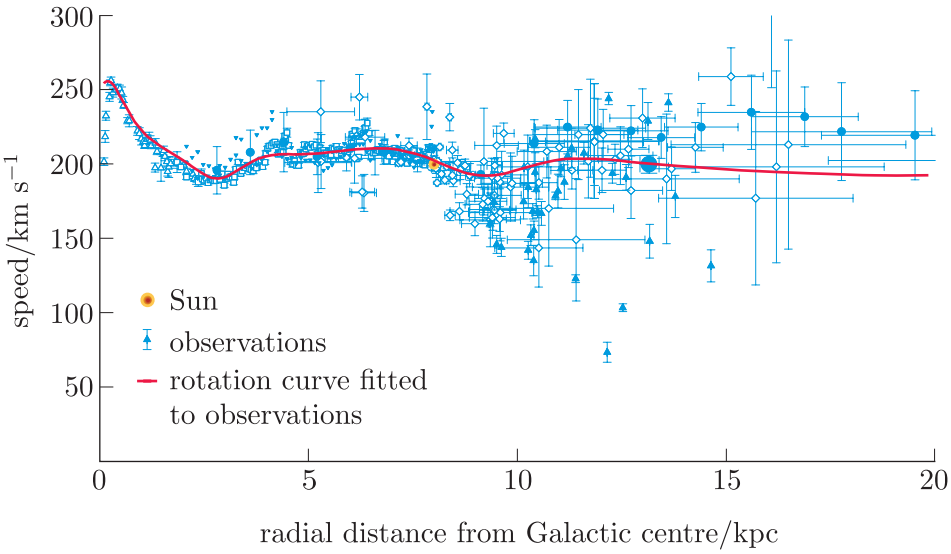


Figure 9.12 Circular velocity of the Milky Way plotted against radial distance from the galaxy centre (different symbols indicate several different sources of observational data).

The total mass of the Milky Way has been measured using the orbits of a large number of different types of tracers: stars, gas clouds, planetary nebulae, star clusters and small satellite galaxies at large distances. These measurements are generally in good agreement, and show that the Milky Way has a total mass of $\sim 10^{12}\,\mathrm{M}_\odot$. This figure is around ten times the mass of the visible constituents of the Milky Way (stars, gas and dust), so it is believed that the Milky Way, like other galaxies, has a vast dark matter 'halo' that pervades the entire galaxy and extends to large distances.

9.3.3 Dark matter, the CMB and structure formation

In Chapter 7 you saw that CMB anisotropies are understood to originate in the oscillations of a fluid of photons and baryons in a 'landscape' of non-uniform density caused by perturbations in the initial distribution of matter. You also saw how the angular power spectrum can be used to measure Ω_b and Ω_m, which provides strong evidence for dark matter as the relative heights and locations of the power spectrum peaks require $\Omega_\mathrm{b} \ll \Omega_\mathrm{m}$.

But it is important to consider whether the entire framework we used to interpret the angular power spectrum could be incorrect. How would density perturbations evolve if there was *only* baryonic matter and radiation, with no dark matter valleys and hills in which to oscillate?

In fact, the overall uniformity of the CMB provides very strong evidence for the existence of dark matter. If there is no dark matter then all of the structure in the Universe – the galaxies and galaxies clusters we observe

today – must have grown from the gas density perturbations present at the time the CMB was emitted. The tiny amplitude of those observed fluctuations ($\sim$1 part in 10^5) means that the baryonic density perturbations at that time were too small to grow quickly enough via gravity to explain the structure that we observe today.

The model described in Chapter 7, of photon–baryon oscillations driven by the presence of perturbations in an underlying distribution of dark matter, allows for the presence of larger variations in density at early times: large enough to grow relatively quickly to form the cosmic web of structure (galaxies and galaxy clusters) that we see today.

All modern cosmological simulations start by modelling the growth of dark-matter perturbations. Structure formation theory also places an important requirement on the *nature* of dark matter: the majority of successful structure formation simulations assume **cold dark matter**. By 'cold', we mean that whatever type of unknown particle is involved must travel slowly, in contrast to 'hot' dark matter, which would be made up of relativistic particles. Cold dark matter is able to clump together, whereas other types are not expected to form stable structure so easily.

You will learn more about dark matter as the module continues. In the next two chapters we will finish exploring the history of the Universe by discussing how structure formation from the time of the CMB onwards led to the presence of stars and galaxies.

9.4 Summary of Chapter 9

- Recombination is the period in the early Universe when nuclei and electrons combined to form atoms. The time at which this happened depends on temperature and (less strongly) on the baryon-to-photon ratio, η.
- The process and timing of recombination influences when decoupling of matter and radiation takes place, because the main interaction process is Thomson scattering by free electrons, whose density becomes very low once atoms form at recombination.
- Decoupling occurs around the time of last scattering, so corresponds to the point at which the CMB is produced. The redshift of decoupling is related to the ionisation fraction X and cosmological parameters via:

$$z_{\rm ls} = \left[\frac{H_0\sqrt{\Omega_{\rm m,0}}}{X(z)n_{\rm b,0}\sigma_{\rm T}c}\right]^{2/3} - 1 \qquad \text{(Eqn 9.9)}$$

  and current best estimates give $z_{\rm ls} \approx 1090$.
- **Elemental abundances** can be measured observationally, providing an important test for big bang nucleosynthesis (BBN) and for our understanding of early Universe physics. They provide a way to measure the cosmic baryon density, $\Omega_{\rm b,0}$, because of the dependence of nucleosynthesis on η.

- Abundances are typically measured relative to the abundance of hydrogen, either as the relative proportion of a given element by mass or by particle number.
- Primordial abundances are difficult to measure, because many elements are also produced or destroyed in stars and galaxy environments. The best measurements come from regions of low **metallicity**, where the least processing by stars has taken place.
- Deuterium abundance has been measured most reliably via **spectroscopy** of high-redshift quasars. Their spectra contain absorption features caused by the presence of deuterium in the quasar's environment. The relative abundance of deuterium is found to be $\mathrm{D/H} \sim 2.5 \times 10^{-5}$, which corresponds to a baryon density in good agreement with CMB measurements.
- Helium abundance can be measured from ionised gas regions outside the Milky Way: estimates from the lowest metallicity regions also show good agreement with BBN predictions.
- The inferred baryon density from cosmic abundances provides one strong piece of evidence for the existence of dark matter: there cannot be enough baryons present to explain the evolution of the Universe's expansion.
- The total gravitating mass of **galaxy clusters** can be measured in multiple ways, all of which demonstrate the need for dark matter.
- The **virial theorem** relates cluster mass to the **velocity dispersion** of the constituent galaxies:

  $$M = \frac{3\sigma_{\mathrm{v}}^2 R}{G} \qquad \text{(Eqn 9.18)}$$

  while the hot gas temperature measured via X-rays also provides a mass estimate:

  $$M = \frac{5k_{\mathrm{B}}RT}{G\langle m \rangle}$$

- The rotation curves of spiral galaxies, including the Milky Way, also provide strong evidence for the existence of dark matter, because velocities of stars, gas clouds and satellite galaxies orbiting in their outer regions are higher than otherwise expected.
- Finally, the CMB anisotropies provide some of the strongest evidence for non-baryonic **cold dark matter**, because the perturbations we observe are too small to explain the growth of galaxies if only baryons were present, and because dark matter particles need to be relatively slow moving ('cold') to allow structure to form.