Chapter 11 Formation of stars and galaxies

In Chapter 10 we examined the collapse of dark-matter overdensities, which form halos over a very wide mass range ($\sim 10^5$–$10^{16}\,\mathrm{M}_\odot$), explored simple models for dark-matter overdensity collapse and compared sophisticated numerical simulations with observations. We briefly considered the behaviour of ordinary (baryonic) gas in the early Universe and concluded that the gravitational collapse of gas must have taken place after dark-matter overdensities were able to grow substantially.

In this chapter we will explore what happens to the baryonic matter in more detail. We will investigate the processes that follow the collapse of dark-matter halos and how these processes can explain the present-day distribution of stars and galaxies in the local Universe. The aim of this chapter is to provide a relatively high-level overview of how the observable structure in the present-day Universe is understood to have evolved. We note that the physics of star formation is a complex topic in its own right and a full treatment is beyond the scope of this module.

Objectives

Working through this chapter will enable you to:

- explain how baryonic material evolves in the early Universe, and the differences and similarities in the evolution of baryonic and dark matter
- describe the cooling processes in the Universe and their implications for how stars and galaxies formed
- summarise how matter is cycled between stars and the interstellar medium
- discuss the main processes affecting galaxy evolution, including mergers, stellar and black-hole feedback and chemical changes
- summarise how large catalogues of galaxies are used to test models of structure formation and galaxy evolution, and how galaxy populations are used as cosmological probes.

Online resources: properties of galaxies

This chapter assumes a basic understanding of galaxies, including the structure of galaxies like the Milky Way, the main galaxy types, and a basic understanding of the lives of stars. If you have not previously studied the properties of galaxies, you may find the online module resources on this topic helpful, which are taken from our Stage 2 curriculum.

11.1 From baryonic gas to stars

In previous chapters we have seen how primordial fluctuations in the early Universe imprinted structure in the CMB radiation, and led to the growth of a cosmic web of dark-matter halos. In this section we will examine how the baryonic gas collapses into the dark-matter halos, and then cools and compresses to reach densities capable of forming the first stars.

11.1.1 Baryons in dark-matter halos

We have also seen that the mass range of dark-matter halos has some correspondence with the mass range we observe for galaxies in the present-day Universe. This suggests that the collapse of ordinary matter (baryonic gas) proceeds in a similar way to that of the dark matter.

- Explain why the baryonic gas will not collapse in *exactly* the same way as the dark matter.
- The baryonic gas will be attracted to the growing dark-matter overdensities through gravity. However, as we saw in Chapter 7, at early times baryons are coupled to radiation to form a photon–baryon fluid, which oscillates in a landscape of dark-matter overdensities and underdensities. This prevents the baryons from immediately collapsing in the same way as the dark matter.

Figure 11.1 shows how overdensities of dark matter and baryons evolve with time. (Recall from Section 10.1.2 that δ is defined as the local density relative to the background mean density.)

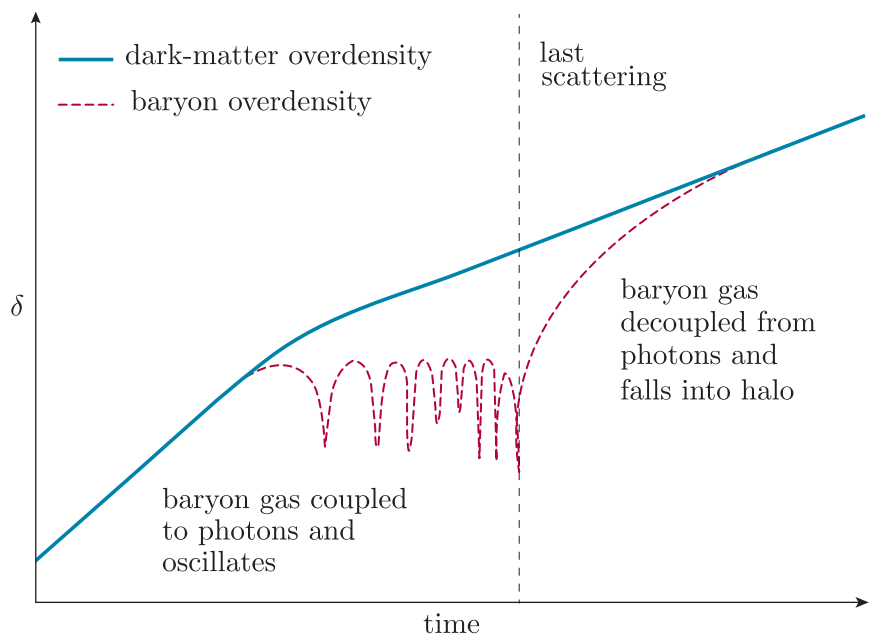


Figure 11.1 A schematic illustration of the evolution of an overdensity of dark matter and an overdensity of baryonic gas at early times.

The dark-matter overdensity grows continuously with time as gravity pulls more dark matter together. In contrast, after a short period of initial growth, the baryonic overdensity starts to oscillate as part of the baryon–photon fluid due to the opposing effects of radiation pressure and gravity. However, the decoupling of photons from baryons at the epoch of last scattering (the vertical dashed line in Figure 11.1) means that the supporting pressure of radiation is removed from the baryonic gas. This enables the baryons to stop oscillating and collapse further into the dark-matter overdensities. The baryonic overdensities are now able to grow much more rapidly, and they quickly evolve to track the dark-matter overdensities more closely.

The baryonic gas within dark-matter halos is the material from which stars and galaxies will form. However, although the gravitational attraction of the dark matter is strong, star formation does not take place immediately. Structures continued to grow in the early Universe through the period sometimes called the dark ages, during which no stars were present. This was a comparatively long period in the Universe's early history, beginning at the epoch of last scattering ($z \approx 1100$, when the Universe was around 380 000 years old) and lasting for about 200 million years (until $z \sim 15$–20).

Figure 11.2 shows part of a simulation of the density structure in dark matter and in gas for a short period following the end of the dark ages. The bright regions indicate the highest densities. Similar structures are visible in the two panels, demonstrating that the baryons and dark matter trace the same structures across the redshift range shown.

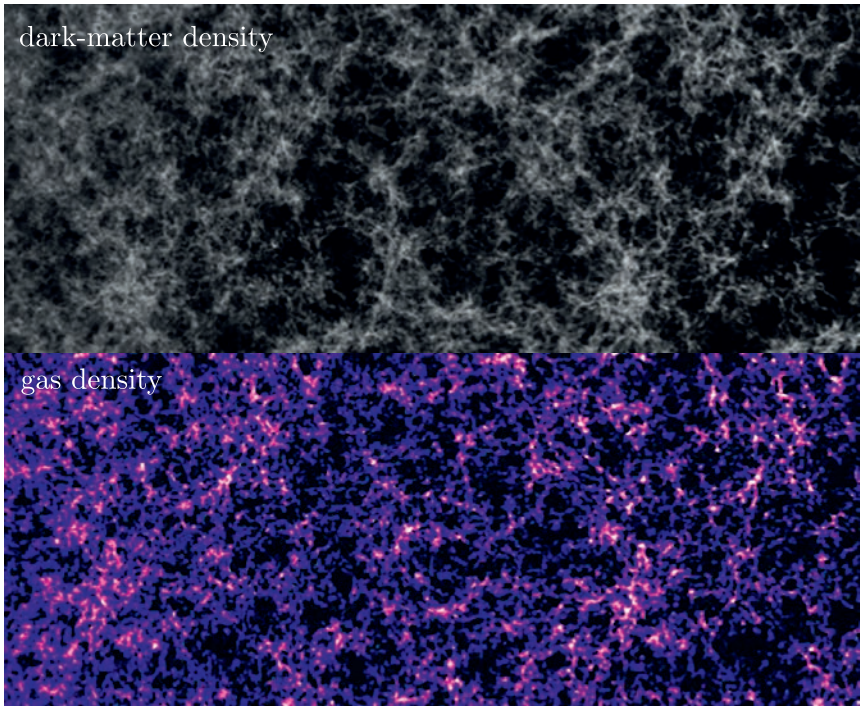


Figure 11.2 The structure of dark matter (top) and baryonic gas (bottom) in the early Universe towards the end of the dark ages, with redshift decreasing from $z = 16$ at the left to $z = 5$ at the right.

11.1.2 Forming structure on galaxy scales

To understand why baryonic collapse does not immediately proceed to turn all of the gas into stars, recall the virial theorem (Equation 9.14). A virialised system is one where the total average kinetic energy and total average gravitational potential energy are balanced such that $\langle E_k \rangle = \frac{1}{2}\langle E_p \rangle$, but where energy is constantly exchanged between the two as particles move towards or away from the centre of mass.

The constant exchange of kinetic and gravitational potential energy can be understood in terms of kinetic theory. Clumps of matter fall into the gravitational potential well and collide, causing the macroscopic kinetic energy of the cloud to be distributed among its individual particles, increasing their random motions – what we would measure as an increased temperature. As gravitational collapse proceeds, smaller clumps formed within the baryonic gas will move together with high bulk velocities and cause shock waves and thus heating when they collide. Section 9.3.1 explained that the gas temperatures in these clumps can be as high as 10^8 K. It is the pressure from this hot gas that prevents localised collapse of baryons into galaxies and stars. The gas needs to cool so that galaxy-scale and stellar-mass clumps are below the Jeans mass (Section 10.1.2).

- What mechanism can act to reduce the temperature of baryonic matter to allow stars and galaxies to form?
- Baryonic matter differs from dark matter in that it interacts with itself and with photons, which means that it can radiate heat. Therefore, if the environment is transparent, then radiation can carry away energy from the collapsing cloud. This cooling process reduces the gas pressure and allows the cloud to reach densities where stars can form.

11.1.3 Gas cooling

Section 10.2.1 stated that dark-matter halo collapse can occur on gigayear timescales. During the dark ages, from recombination ($z \approx 1100$) to $z \approx 20$ (the exact endpoint is highly uncertain), baryon collapse proceeded slowly and did not reach the densities required to form stars. Although matter and radiation were decoupled at recombination, the CMB radiation continued to bathe matter in heat, which dictated the lowest temperature at which matter could easily exist. However, by $z \approx 17$, the CMB temperature had dropped from 3000 K (when it was emitted) to 50 K, meaning gas in the Universe could potentially cool to these temperatures.

Before considering the potential gas-cooling mechanisms that were present in the early Universe, it is instructive to review the main cooling mechanisms at work in the Universe today. Table 11.1 lists a range of processes that include interactions of electrons and ions, electron transitions within atoms (including **forbidden lines**, which only occur in low-density space plasmas) and vibrations and rotations of molecules. Each process occurs when gas is in a particular, restricted energy range, and each involves the release of energy as radiation in a particular part of the electromagnetic spectrum, as shown in the last column.

Table 11.1 Primary present-day cooling mechanisms for gas in the Milky Way.

Temperature	Main cooling mechanism	Emission band
$>10^7$ K	Free–free (bremsstrahlung)	X-ray
10^7–10^8 K	Iron resonance lines	X-ray
10^5–10^7 K	Other metal resonance lines	ultraviolet (UV), X-ray
8000–10^5 K	Metal forbidden lines (especially C, N, O, Ne)	infrared (IR), optical
1000–8000 K	Atomic transitions (e.g. H, O)	IR, optical
100–1000 K	Molecular vibrational and atomic transitions (e.g. O, C^+, H_2)	far-IR
10–100 K	Molecular rotational transitions (especially CO)	sub-millimetre

- How do you think the gas-cooling processes in the early Universe might have differed from the present-day mechanisms given in Table 11.1?
- Baryonic matter in the early Universe was comprised essentially of hydrogen and helium, without astronomical 'metals'. This means none of the cooling mechanisms in Table 11.1 that rely on metals could operate, (i.e. metal resonance and forbidden lines, atomic and molecular transitions other that hydrogen and H_2. Hence, gas in the early Universe cooled much more slowly.

Figure 11.3 plots **cooling rate**, $\Lambda_{\rm cool}$, as a function of temperature, T, for interstellar gas of different compositions.

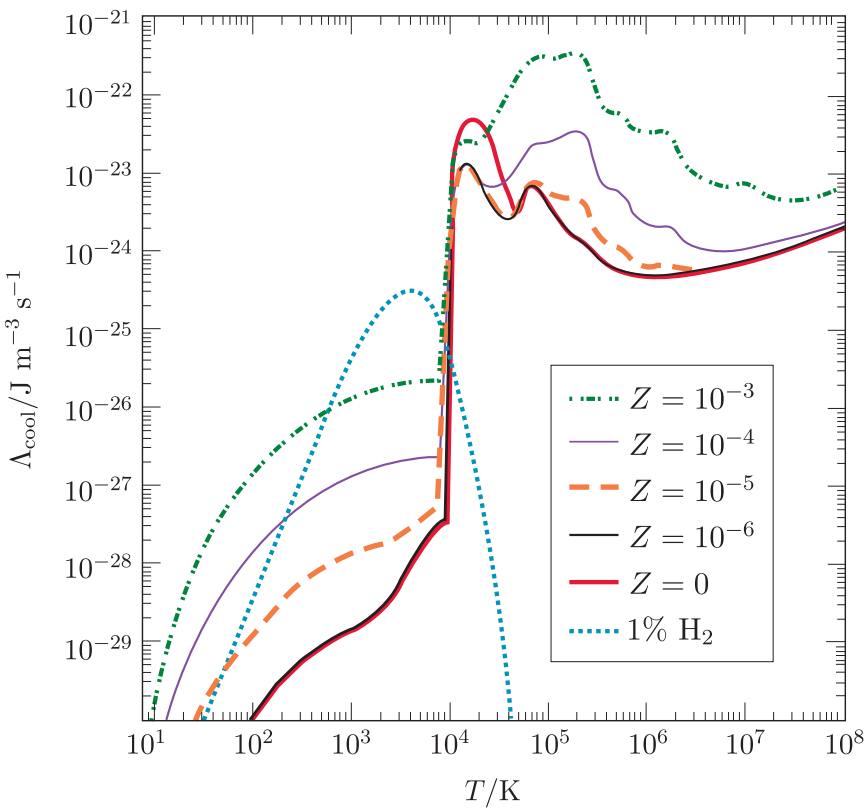


Figure 11.3 Cooling rate, $\Lambda_{\rm cool}$, of interstellar gas for different metallicities, Z. Note that the five curves labelled with their metallicity values do not include the effect of cooling by H_2. The cooling rate for interstellar gas containing an H_2 fraction of 1% is shown separately.

- Consider the units of $\Lambda_{\rm cool}$ in Figure 11.3. What does this suggest about how $\Lambda_{\rm cool}$ relates to the luminosity of radiation produced?
- The units of $\Lambda_{\rm cool}$ are $\mathrm{J\,m^{-3}\,s^{-1}}$, equivalent to $\mathrm{W\,m^{-3}}$. This is the rate at which a cubic metre of material will lose heat energy by radiation, so this defines not only how quickly the material will cool down, but also the luminosity of radiation it will produce.

The lack of effective cooling mechanisms in the early Universe meant the gas pressure remained high. This kept the Jeans mass high also (Section 10.1.2) and meant gas clouds could more easily heat up as they collapsed and dissipated again. An important part of the gas collapse process was the formation of molecular hydrogen, H_2. In today's Universe, H_2 mostly forms on interstellar dust grains, but in the dust-free primordial Universe it could only happen in a comparatively cool, dense environment where hydrogen atoms would collide frequently at slow enough speeds so that they stuck together. Forming H_2 in this manner released chemical binding energy. This process, **H_2 formation heating**, can heat up gas clouds, helping them dissipate. The ability of gas to cool therefore depends both on its temperature and density.

Figure 11.4 show the relationship between temperature and number density as typical gas clouds cool, for two different chemical compositions: the primordial composition of zero metallicity Z defined in Equation 9.11 and solar composition. A gravitationally contracting cloud will increase in density, moving from left to right on the diagram, following a track that is determined by the interplay of cooling and heating as collapse progresses.

If clumps of gas can cool as they increase in density, then they will fragment into smaller clumps that can go on to form stars. If they cannot cool as they become denser, then they will expand and not go on to form stars. When the gas becomes opaque at the highest densities, radiation can no longer escape and so gas heats much more rapidly.

Exercise 11.1

Examine the two cooling tracks shown in Figure 11.4.

(a) Over which number density range on the cooling track for solar composition is the gas cooling, and over which range is it heating?

(b) What is the Jeans mass corresponding to the conditions where stars of solar composition are most likely to form? How does this compare to typical present-day stars?

(c) Using similar considerations, what would you expect to be the typical mass of the first stars?

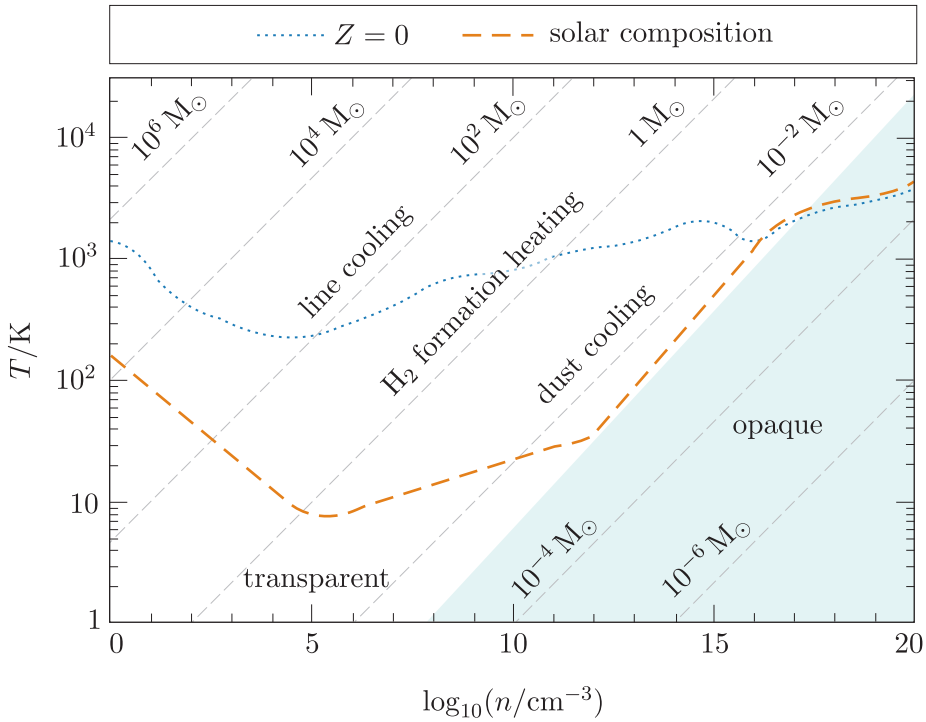


Figure 11.4 Cooling tracks of interstellar gas of primordial ($Z = 0$) and solar compositions. Diagonal lines show the Jeans mass (Equation 10.1) of clumps at that number density, n, and temperature, T. The main cooling and heating mechanisms at each mass are indicated. The blue shaded region shows where the gas becomes opaque.

We can work out an approximate timescale on which gas can cool enough so that stars can form. Consider the kinetic energy contained within atomic hydrogen gas, $E_\mathrm{k} = \frac{\gamma}{2} k_\mathrm{B} T$ per particle, where k_B is the Boltzmann constant, T is temperature and γ is the number of degrees of freedom. ($\gamma = 3$ for atomic gas and $\gamma = 6$ for molecular hydrogen, H_2. Note that γ here is unrelated to the Lorentz factor, for which the same symbol is used.) The total kinetic energy per unit volume is nE_k where n is the number density of particles. Then, assuming a constant cooling rate, the **cooling timescale** (which can be considered equivalent to the minimum timescale on which stars could form) is given as:

Cooling timescale

$$t_\mathrm{cool} = \frac{nE_\mathrm{k}}{\Lambda_\mathrm{cool}} = \frac{\gamma n k_\mathrm{B} T}{2\Lambda_\mathrm{cool}} \tag{11.1}$$

The following exercise explores further how the concept of cooling timescales can be used to compare star formation today to formation of the first stars.

Exercise 11.2

At recombination ($\approx$380 000 years after the big bang), the baryonic number density in the Universe was $n \approx 3 \times 10^8\,\mathrm{m}^{-3}$ and the temperature was $T \approx 3000\,\mathrm{K}$. Use this information together with Figure 11.3 and Equation 11.1 to answer the following questions about how different assumptions affect the cooling timescales.

(a) First make the assumption that the gas at recombination has a metallicity of zero ($Z = 0$) and contains no molecular hydrogen. Estimate the time taken for stars to form, and therefore the age of the Universe at which the first stars could appear.

(b) Now assume instead that the metal-free gas does contain molecular hydrogen at the level of 1% H_2 and carry out a similar calculation to estimate a timescale on which stars form.

(c) Comment on what the results of parts (a) and (b) suggest about the role of H_2 in how gas cools to form the first stars.

(d) By considering the time-evolution of both the Universe and a collapsing cloud of baryons within it, state some of the reasons why the ages you have calculated are estimates.

Exercises 11.1 and 11.2 suggest that the first stars took millions of years to form (detailed calculations give the 200 million years mentioned earlier in the chapter); the exercises also imply that the first stars were extremely massive. The more massive a star, the shorter it lives: if the first stars were very massive, they would have burned brightly but would have exploded very quickly (in less than 10^6 years) as supernovae. We wouldn't see any of these stars at low or medium redshifts today. Searches for these first stars must therefore look to the highest-redshift galaxies in the Universe: we cannot observe individual stars at such colossal distances, but must rely on indirect tracers. The study of these first stars is therefore a challenging and active topic of research. We call these still-hypothetical massive first stars **Population III** stars to differentiate them from the two populations of stars we already observe in the nearby Universe: older **Population II** and younger **Population I** stars.

11.1.4 Reionisation and Strömgren spheres

Massive (blue) stars generate a lot of UV radiation at wavelengths shorter than the **Lyman limit** (91.2 nm). Photons at wavelengths below the Lyman limit can ionise hydrogen atoms directly from their ground state, so the gas in the interstellar medium (ISM) around the first stars became ionised. This was a crucial point in the history of the Universe, known as **reionisation**, and represented the point at which the complex physics of baryonic matter became very important for the observable behaviour of the Universe.

The ionised bubble of plasma surrounding an individual star is known as a **Strömgren sphere**, and will grow outwards from the star. We can model the growth of a Strömgren sphere by considering the rate at which ionising photons are produced by a source, Q, and the number density of material the source is ionising, n. For simplicity, we will assume a cosmos of pure neutral hydrogen ($n = n_{\mathrm{H}}$).

In a relatively diffuse medium, most photons should reach the edge of the ionised region and cause the ionisation of a new atom; the radius R of the Strömgren sphere will expand at a rate of

$$\frac{\mathrm{d}R}{\mathrm{d}t} \approx \frac{Q}{4\pi R^2 n_{\mathrm{H}}} \tag{11.2}$$

However, in dense regions, some of the ionised atoms will recombine with free electrons. These recombined atoms will then need to be ionised again to maintain the Strömgren sphere, or it will collapse back to neutrality. Mathematically, this effectively decreases Q.

The recombination rate depends on the probability of ion–electron interactions, which is proportional to the ion and electron densities (n_{i} and n_{e}, respectively) and the probability of the two recombining when they meet – the **recombination coefficient**, $\alpha(T)$. Note that $\alpha(T)$ is temperature-dependent, since it is more difficult for ions to capture faster electrons. Note also that at this later cosmological epoch the term 'recombination' is more apt than the use of the same term to describe the formation of the first atoms: during reionisation the atoms are indeed recombining having previously been ionised.

The average time it takes an ion to capture an electron, t_{r}, is given by:

$$t_{\mathrm{r}} = \left[n_{\mathrm{e}}\alpha(T)\right]^{-1} \tag{11.3}$$

Including the recombination coefficient in Equation 11.2 gives:

$$\begin{aligned}\frac{\mathrm{d}R}{\mathrm{d}t} &\approx \left[Q - \frac{4}{3}\pi R^3 n_{\mathrm{i}} n_{\mathrm{e}} \alpha(T)\right] \frac{1}{4\pi R^2 n_{\mathrm{H}}} \\ &\approx \frac{Q}{4\pi R^2 n_{\mathrm{H}}} - \frac{n_{\mathrm{i}} n_{\mathrm{e}} \alpha(T) R}{3 n_{\mathrm{H}}}\end{aligned} \tag{11.4}$$

By setting $\mathrm{d}R/\mathrm{d}t = 0$ and rearranging Equation 11.4, we can determine the radius R_{max} where recombination balances out ionisation. Hence, the final (i.e. maximum) radius of the Strömgren sphere is

$$R_{\mathrm{max}} \approx \left[\frac{3Q}{4\pi n_{\mathrm{i}} n_{\mathrm{e}} \alpha(T)}\right]^{1/3} \tag{11.5}$$

In general, as the sphere expands, recombination increases and the growth slows down. By integrating Equation 11.4 we see the radius changes as:

Radius of a Strömgren sphere

$$R(t) \approx R_{\mathrm{max}}\left[1 - \exp(-t/t_{\mathrm{r}})\right]^{1/3} \tag{11.6}$$

Exercise 11.3

Consider the following scenarios:

(i) a very massive, young star embedded in its host galaxy's ISM where $Q = 10^{50}\,\mathrm{s}^{-1}$ and $n_\mathrm{H} = n_\mathrm{i} = n_\mathrm{e} = 10\,\mathrm{cm}^{-3}$

(ii) a galaxy of 10^5 such stars surrounded by a neutral medium at $z \approx 17$ where $Q = 10^{55}\,\mathrm{s}^{-1}$ and $n_\mathrm{H} = n_\mathrm{i} = n_\mathrm{e} = 10^{-5}\,\mathrm{cm}^{-3}$.

Assuming $\alpha(T) \approx 4 \times 10^{-19}\,\mathrm{m}^3\,\mathrm{s}^{-1}$ in each case, for both scenarios calculate:

(a) the maximum radius of the Strömgren sphere

(b) the timescale for atoms recombining

(c) the size of the Strömgren sphere after 10^6 years.

Exercise 11.3 shows that even the largest, brightest stars can only ionise a small part of their galaxy's ISM. In contrast, a galaxy of large, bright stars can completely ionise the **intergalactic medium** (IGM) over a much larger region than the size of the galaxy itself. The primary reason is that recombination timescales inside galaxies are short (much shorter than the timescales of stellar evolution), whereas in the IGM between galaxies, an ionised atom is unlikely to ever meet an electron and become neutral again. Consequently, the galactic ISM will remain mostly neutral, with pockets of ionisation generated by localised star formation. Meanwhile, the ionisation front from the first galaxies may have travelled through intergalactic space at close to the speed of light, and the IGM remains almost completely ionised to this day.

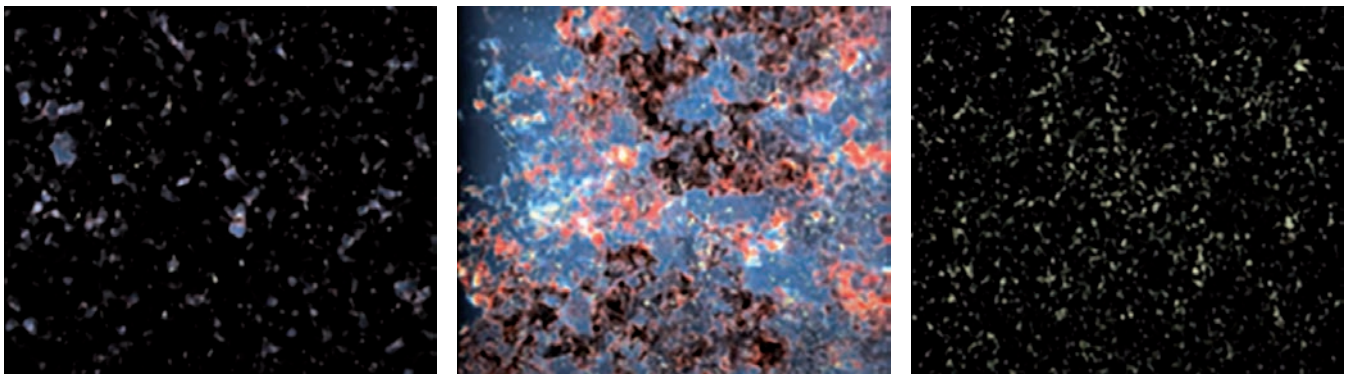
The calculations of recombination timescales and the evolution of Strömgren spheres outlined in this section are approximations for several reasons. These include:

- although most of the matter in the Universe is hydrogen, other elements contribute too
- recombination releases photons, some of which will contribute to maintaining ionisation
- if ionisation occurs sufficiently quickly, then lower-energy photons can significantly contribute to it
- the hot, ionised gas will be over-pressured compared to the surrounding neutral gas, so will expand of its own accord
- the edge of the Strömgren sphere can become poorly defined, because the mean free path of particles is long
- Q and T are liable to change as stars evolve
- the expansion of a Strömgren sphere is fundamentally limited by the speed of light.

The period when the Universe was filled with intense radiation and returned to a largely ionised environment is referred to as the **epoch of reionisation**. It lasted from (very approximately) $z \sim 20$ to $z \sim 6$, and transformed the baryonic Universe from a cold, neutral gas (containing density fluctuations) into today's Universe, with dense, neutral interstellar media within galaxies, but hot, ionised intergalactic media outside of galaxies. Studying the epoch of reionisation and determining the redshifts at which it occurred place important constraints on the evolution of the Universe: this is a major goal of telescopes such as *JWST*.

Online resources: reionisation simulations

The snapshots in Figure 11.5 were taken from a video rendering of the full numerical simulation. A link to the simulation can be found in the online module resources.



(a) (b) (c)

Figure 11.5 Snapshots from a simulation of reionisation, showing its approximate appearance in UV light at three different redshifts from (a) $z \sim 15$ to (b) $z \sim 10$ to (c) $z \sim 6$ (note that the exact redshifts it reaches each stage are not well determined). The Universe is initially opaque to UV with isolated Strömgren spheres. These spheres then merge together and the partially ionised gas becomes transparent in some directions. Finally, the gas becomes diffuse and fully ionised: the Universe is then fully transparent to UV, and we can see galaxies throughout the simulation.

11.2 Key processes in galaxy evolution

This section discusses the main factors that affect the structure and chemistry of galaxies. In particular, we will look at the conversion from the hydrogen–helium Universe of the 'dark ages' to the bright, galaxy-filled, chemically rich Universe of today.

11.2.1 Galaxy assembly

Figure 11.6 shows the major processes involved in building a galaxy, and how it evolves with time.

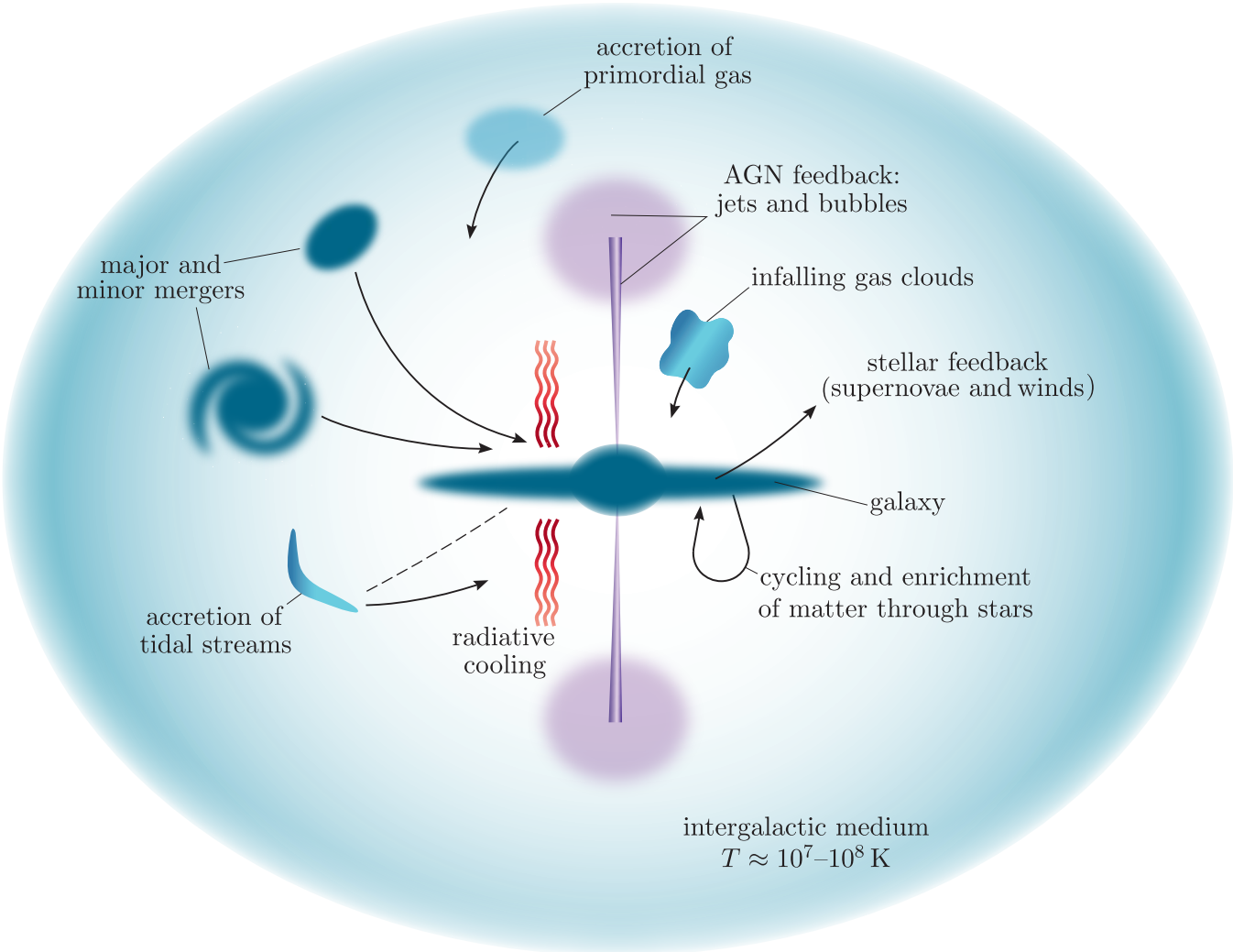


Figure 11.6 The primary methods by which a galaxy interacts with its environment. Note that the infall processes can occur from any direction, but radiation, AGN, stellar and supernova feedback occur primarily in the directions perpendicular to the galaxy disc.

Galaxy formation begins with the accretion of primordial gas clumps, which can cool enough to form stars. These are the first **protogalaxies**, which evolved to become the galaxies we observe at the highest redshifts. Observations of high-redshift galaxies show these to be mostly irregular in shape and typically much smaller in size than nearer (low-redshift) galaxies that are being observed at a later stage in their evolution.

As galaxies evolve, they continue to fall towards the centres of dark-matter halos, forming groups and clusters of galaxies. Collisions between galaxies can lead to **galaxy mergers** – either a major merger between two galaxies of roughly equal size, or a minor merger if a small galaxy falls into a large one. Galaxies can also be shredded into **tidal streams** if they pass too close to a larger galaxy: a process that we observe in the Milky Way today. Stars from these streams can be accreted onto the larger galaxy indirectly.

As early galaxies evolved, they increased in mass through the accretion of gas, while collapsing via gravity as they cooled. A collapsing galaxy is likely to be irregular in shape, which leads to net rotation because collapse is not completely symmetric. Conservation of angular momentum increases the rate of rotation as the collapse progresses. While gravitational collapse still occurs along the axis of rotation, the effect of rotation is to counteract gravitational collapse in directions perpendicular to the axis. Hence, the matter in rotating galaxies tends to form spirals constrained to a disc.

Major galaxy mergers occur in denser environments, and tend to disrupt the rotation of galaxies and scatter stars, leading to the formation of elliptical galaxies. The most dynamic part of a dark-matter halo is its centre: galaxies here accrete the most gas and undergo the most mergers, so the centre of a galaxy cluster tends to host large elliptical galaxies.

If a galaxy has a massive central black hole and the galaxy centre is rich in gas, it is likely to go through phases of gas accretion onto the central black hole, resulting in an **active galactic nucleus** (AGN). The AGN can also produce jets that push through the surrounding gas and terminate in bubbles called radio lobes. Jets, radio lobes and the processes of stellar feedback annotated on the right-hand side of Figure 11.6 are discussed in the remainder of this section.

11.2.2 The cosmic cycle of matter

Stars play a significant role in shaping the chemical composition of a galaxy. Consider the first stars to form in a galaxy comprised mainly of hydrogen and helium, which are expected to have high masses (Section 11.1.3). The rate at which stars consume fuel increases with stellar mass. Hence, the massive first-generation stars will reach the end of their lives quickly. This allows an appreciable amount of material to be ejected from these stars into the ISM via stellar winds and supernovae remnants.

Crucially, this ejected material is chemically enriched with the products of the stars' nuclear fusion, and so astronomical metals begin to diffuse through the ISM. The next generation of stars will form from this chemically enriched ISM, thus creating a cosmic cycle of matter between the ISM and stars (see Figure 11.7).

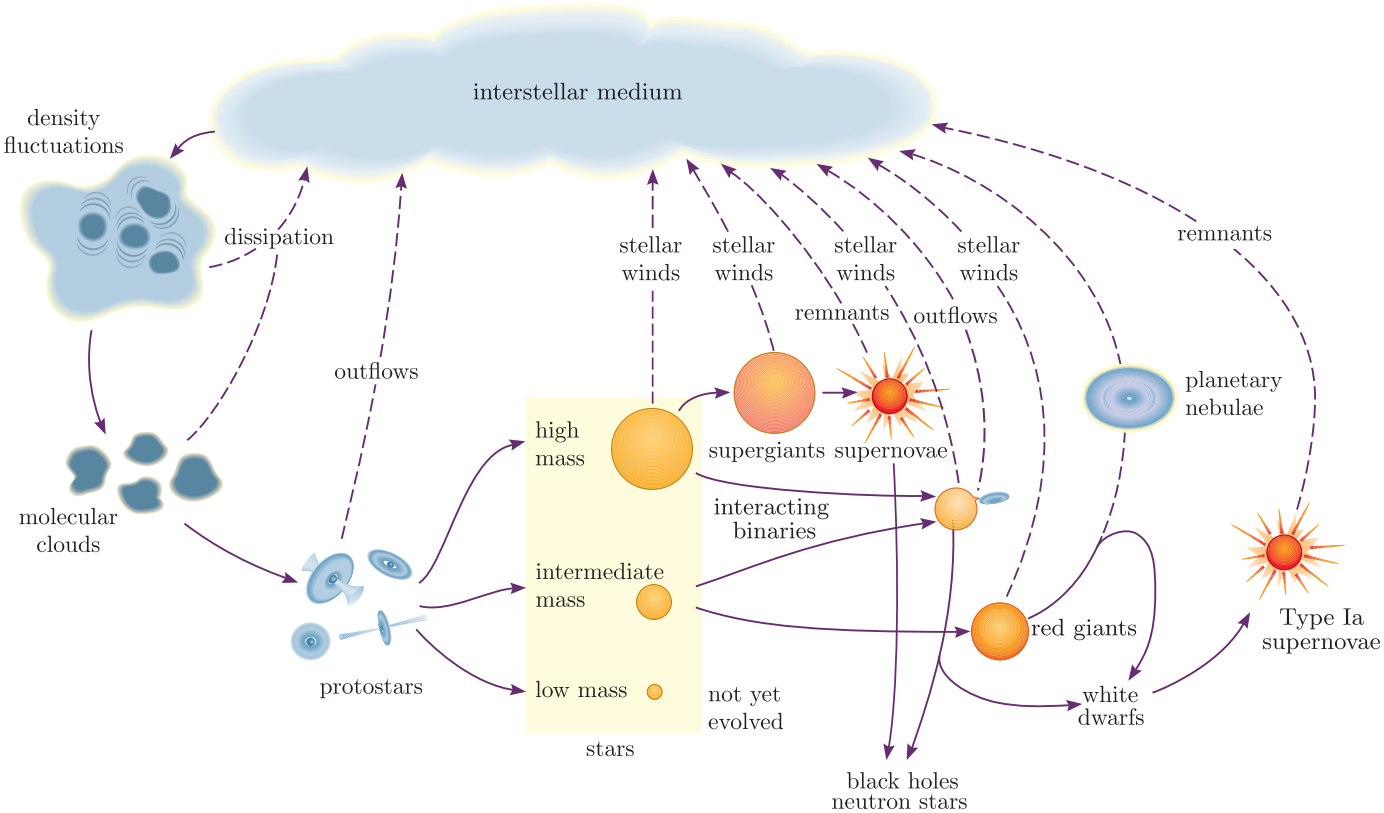


Figure 11.7 The cosmic cycle of matter, showing the main pathways by which matter moves between the ISM and stars and back again. Processes at the left began when stars first formed; processes at the right only began later in the timeline of the Universe.

The cosmic cycle of matter allows us to infer the nature of stars in the early Universe. The oldest stars we see today are Population II stars, formed from gas already enriched by the first generation of Population III stars. The chemical make-up of these early Population II stars reveals the elements ejected by Population III stars, from which we can infer what the Population III stars were like.

The chemical enrichment of the ISM changes with time. As the Universe ages, progressively lower-mass stars start to die. After $\sim$35 million years, stars with masses a little less than $8\,M_\odot$ evolve to become red giants, with increasingly lower-mass stars gradually following them into this stage over time. Towards the end of the lives of the more massive red giants, violent pulsations essentially shake stellar atmospheres into space to form planetary nebulae. During this shedding process, further nuclear reactions take place, involving the capture of neutrons, which enable heavier elements to be synthesised.

Figure 11.8 illustrates, for each element of the Periodic Table, the various – mainly astrophysical – origins for the majority of that element present in the Universe. It shows that both low- and high-mass stars contribute to producing the elements of most importance for everyday life. For example,

carbon and nitrogen are primarily produced in low-mass stars, whereas the most-massive stars are essential to produce a range of other important elements, including oxygen, aluminium, sodium and magnesium.

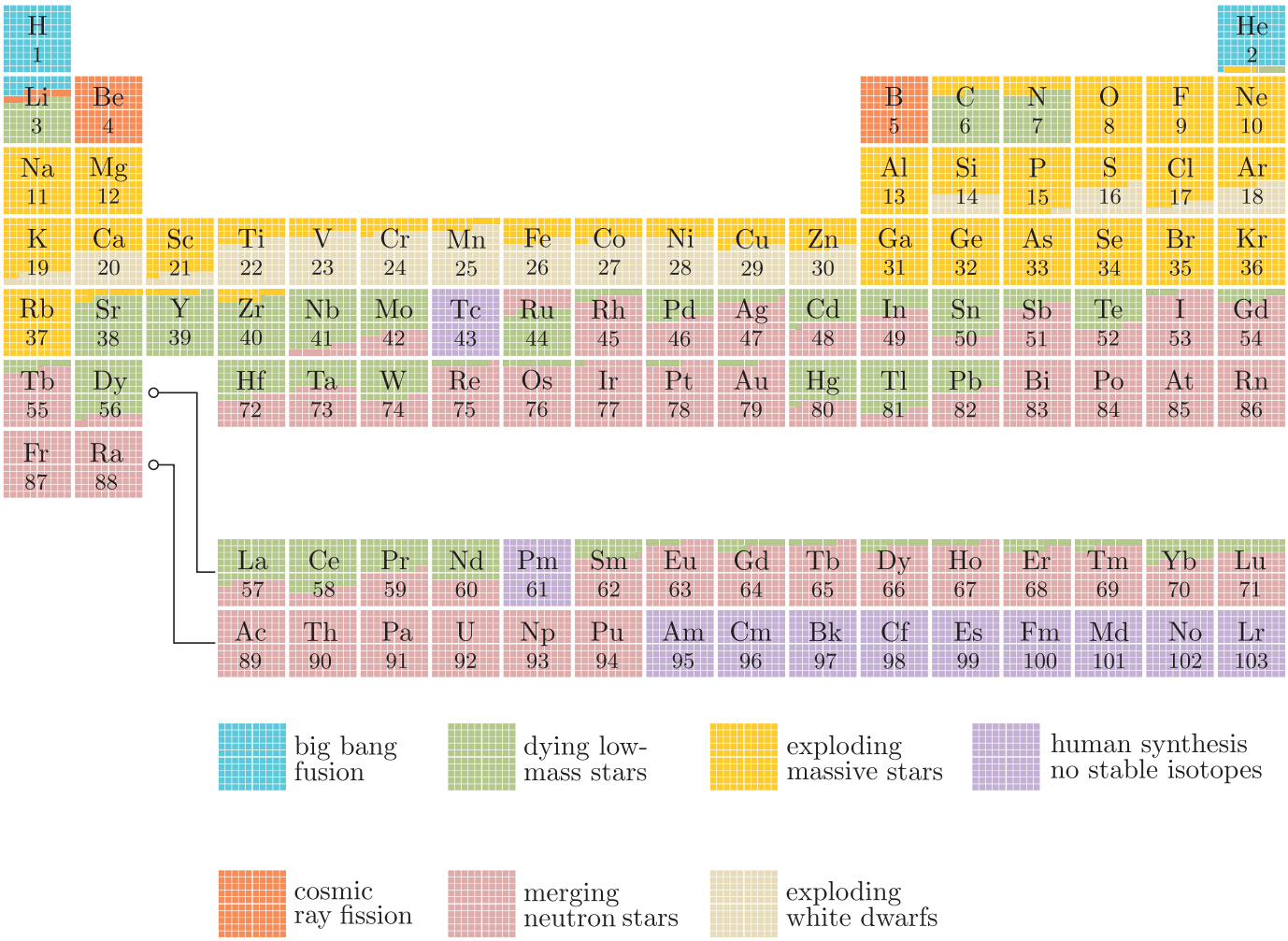


Figure 11.8 The Periodic Table, showing the sources of each element.

As well as enriching the ISM with astronomical metals, stars also provide radiation and kinetic energy to the ISM. This **stellar feedback** (Figure 11.6) can promote or inhibit further star formation. Ejecta from the winds of hot stars, and from supernovae and other outflows, can have enough kinetic energy to escape the gravitational potential of their galaxies. Such ejecta can form bubbles and, especially in disc galaxies, we see these bubbles expanding out of the galaxy as **galactic outflows**, as shown for the galaxy Messier 82 in Figure 11.9. Material from bubbles can subsequently fall back down into the galaxy in different places, replenishing the galaxy's gas reservoir.

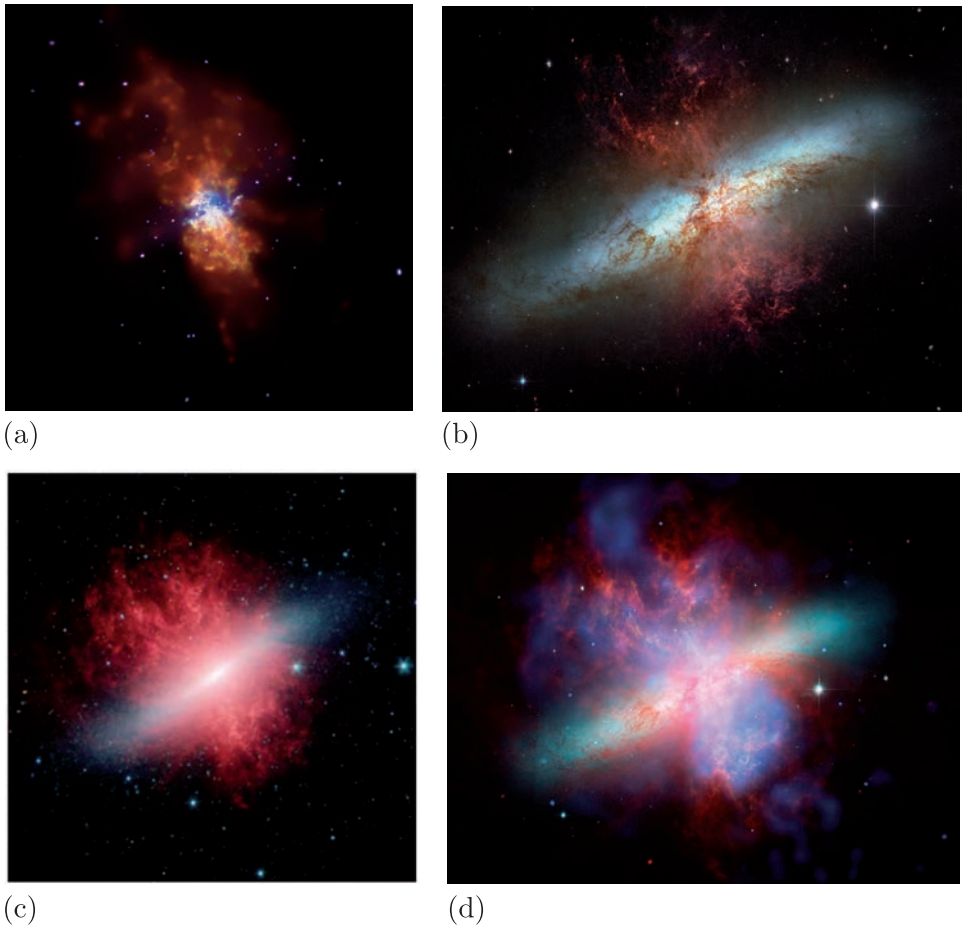


Figure 11.9 Messier 82, as seen in (a) X-ray, (b) optical and (c) infrared, with (d) a colour composite including all three wavebands. The cigar-shaped disc of the galaxy extends from bottom-left to top-right, while the galactic outflow formed by its ongoing episode of rapid star formation is perpendicular to this.

The next exercise involves investigating these feedback processes further.

Exercise 11.4

The escape speed (see Section 3.2.2) for gas at a particular radius R from the centre of a region of mass M is given by

$$v_{\text{esc}} = \sqrt{\frac{2GM}{R}} \tag{11.7}$$

Consider a point $R = 8\,\text{kpc}$ from a galaxy's centre, and assume that the galaxy has a mass of $M = 10^{12}\,\text{M}_\odot$ within this radius.

(a) What is the escape speed from this point?

(b) How does the escape speed compare with:

(i) a $800\,\text{km}\,\text{s}^{-1}$ supernova outflow

(ii) a $20\,\text{km}\,\text{s}^{-1}$ outflow from a dying low-mass star?

(c) Using Figures 11.6 and/or 11.7, suggest the stellar feedback mechanisms in this galaxy that processes (i) and (ii) contribute to.

(d) How would your answers to (b) and (c) differ if processes (i) and (ii) were located in a smaller galaxy of $M = 10^8\,\mathrm{M}_\odot$ and $R = 800\,\mathrm{pc}$. Describe the implications for chemical enrichment in the smaller galaxy.

The information about a galaxy's evolutionary history encoded in the chemical composition of its stars can be used to place limits on how the mass and radius of galaxies have changed over time. This helps astronomers distinguish dark-matter dominated galaxies, which are both massive and compact enough to retain ejecta from supernovae, from smaller systems like globular clusters, which don't contain dark matter and only show small amounts of elements produced in supernovae. The minimum mass of a galaxy, including dark matter, is therefore of order of the most-massive globular cluster: $\sim 10^7\,\mathrm{M}_\odot$.

11.2.3 Feedback from stellar remnants and supermassive black holes

We have seen from the cosmic cycle of matter that stellar winds and supernovae can return a lot of mass from stars to the ISM, but a fraction of matter will remain locked in compact **stellar remnants** (namely black holes, neutron stars or white dwarfs). Since not all mass is recycled, the amount of gas in a galaxy will decline over time. Therefore, unless the gas-depleted galaxy can accrete more gas by merging with a gas-rich galaxy, the galaxy will eventually stop forming any young, hot, blue stars entirely and become 'red and dead'. When we examine high-redshift galaxies, we find that star formation peaked around 10 billion years ago ($z \approx 2$) and has been gradually declining since.

As galaxies age, the number of remnants increases. These stellar remnants continue to play an important role in shaping galaxies. Over half of stars (especially high-mass stars) are found in binary or higher-order systems, and sometimes two stars can interact. **Interacting binary stars** affect each others' nuclear and chemical evolution. If an evolving star transfers mass onto a compact object, it can trigger outbursts called novae, which add a distinct chemical fingerprint to galaxies.

In dense environments, including the centres of galaxies and stellar clusters, stellar encounters are common. This shares out the kinetic energy of the stars and stellar systems: small stars become kinematically hot (i.e. they have high velocities), while massive stars become kinematically cold (i.e. they have small velocities, but similar amounts of orbital momentum). The relevant mass here is the total mass of the stellar system (i.e. binary stars are slowed like one single star of the same mass). This redistribution of kinetic energy leads to **mass segregation**: massive stars

and binary stars – and any remnants they create – will tend to sink to the centre of their host clusters or galaxies. This encourages massive remnants like black holes to congregate towards the galactic centre.

At the centre of all large galaxies are **supermassive black holes** (SMBHs). The origins of SMBHs are not well known, but they attained masses of 10^6–$10^8\,M_\odot$ within a fraction of a billion years of the big bang. One hypothesis is that they built up via a very rapid version of mass segregation from the remnants of Population III stars. An alternative theory is that supermassive black holes originated in massive 'seed' black holes formed by direct collapse of large amounts of matter in the very early Universe.

SMBHs are often surrounded by chaotic environments, including active sites of star formation and orbiting, infalling material. In these active galactic nuclei (AGN), material is actively being accreted by the black hole. The mass energy of this accreted material can power **jets** that channel energy, and energetic particles, outwards from very close to the central black hole, at speeds close to the speed of light. The jets produce synchrotron radiation, seen at radio wavelengths, and the large bubbles they create are visible as **radio lobes**, as shown in Figure 11.10. Jets from supermassive black holes transfer substantial amounts of energy into the outer regions of the host galaxies and the wider environment, affecting subsequent star formation.

The energy deposited by an AGN into its host galaxy affects galaxy evolution. This **AGN feedback** can be 'positive', promoting star formation by triggering the collapse of gas clouds; or 'negative', **quenching** (halting) star formation by either preventing radiative cooling of infalling gas, or ejecting too much gas from the galaxy.

(a) (b) (c)

Figure 11.10 (a) Radio jets and lobes of the radio source Centaurus A, which originate in the AGN at the centre of the it hosting galaxy NGC 5128, shown in an optical image in (b), which reveals strong absorption from a dust lane across the centre of the galaxy. Panel (c) shows a super-position of NGC 5128's optical and radio structure.

11.2.4 Chemical evolution of galaxies

Stellar winds, novae, supernovae, outflows, accretion of infalling gas, and AGN feedback all combine to change the chemical properties of individual galaxies. When galaxies merge, their chemical properties become entwined. However, their individual stellar populations can sometimes remain distinguishable using chemical tracers, such as the ratios of individual elements.

Galactic archaeology describes the process of using chemical tracers to understand the merger histories of our Galaxy and others, and/or to explore how the Universe as a whole became chemically enriched. Chemical tracers can be found either in the spectra of individual stars or those of the interstellar gases in galaxies. Common tracers include oxygen and iron abundances: oxygen is relatively easy to measure in the gas phase, while iron is relatively easy to measure in stars. (Use of these tracers is similar to the methods for investigating primordial elements discussed in Chapter 9.)

Star formation is strongly suppressed when negative AGN feedback is strong, the gas in a galaxy is ejected by outflows, or too much material in a galaxy is locked up in stars or remnants. When star formation is quenched, galaxies will change in appearance from being dominated by hot, young, blue stars, to being dominated by cooler, old, red stars. Hence the colours of galaxies give clues about their evolutionary histories.

In the next exercise you will explore how elemental abundances can be used to investigate galaxy evolutionary histories.

Exercise 11.5

Figure 11.11 (overleaf) shows the relative abundances of oxygen and iron for four galaxies. (Recall the nomenclature [O/Fe] and [Fe/H] described in Section 9.2.1. Values of [Fe/H] = 0 or [O/Fe] = 0 correspond to solar abundance ratios.)

Use Figure 11.11 together with Figures 11.7 and 11.8 to answer the following questions.

(a) What are the processes that produce most of the hydrogen, oxygen and iron?

(b) In what order do these processes occur throughout the Universe?

(c) Using your answers to parts (a) and (b), explain the common trends in Figure 11.11.

(d) Describe and explain the differences between the four galaxies shown in Figure 11.11.

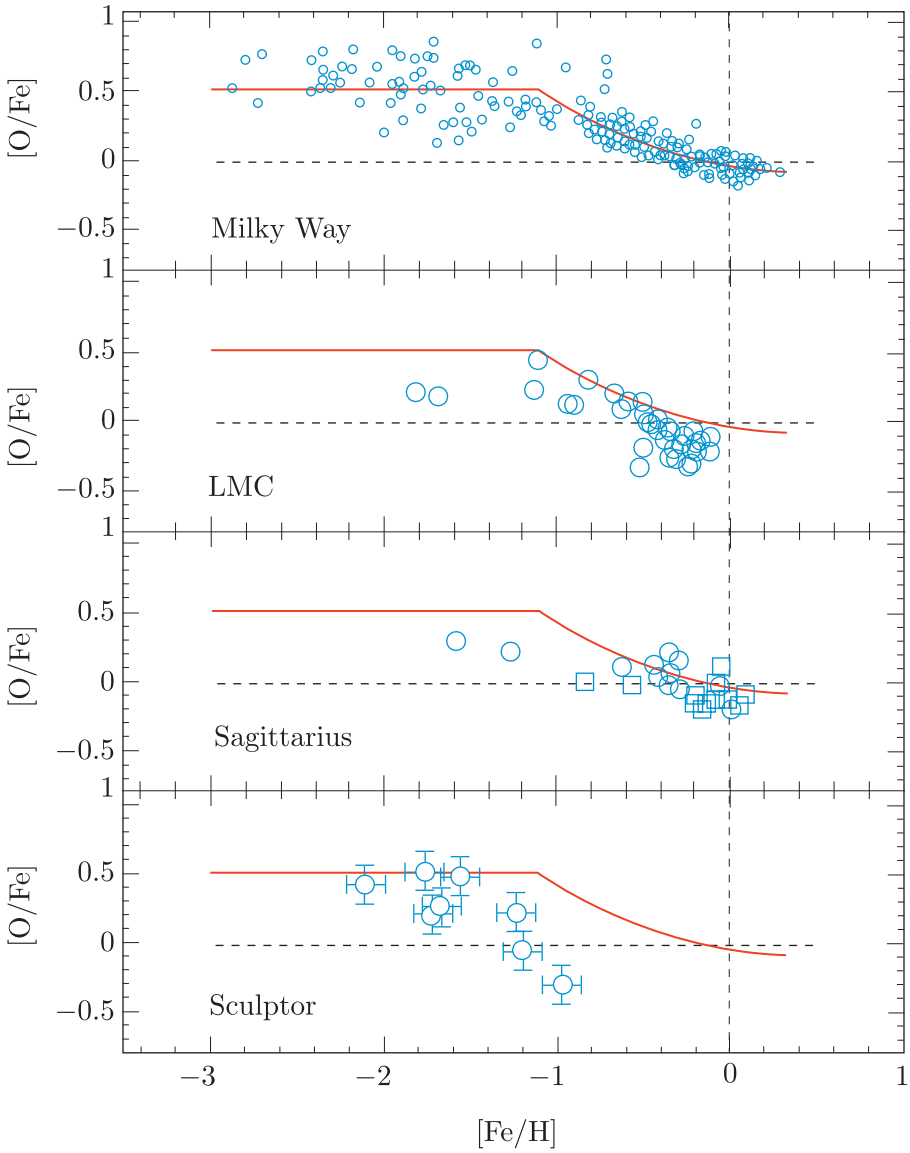


Figure 11.11 Oxygen-to-iron abundance versus iron abundance for stars in the Milky Way and three of its satellite galaxies, the Large Magellanic Cloud (LMC), Sagittarius and Sculptor. The red line shows a model fit to the Milky Way and is identical in all four plots. The vertical dashed line shows the Sun's iron abundance.

11.3 Galaxies as cosmological probes

As we cannot directly observe dark matter, we must rely on baryonic matter in stars, galaxies and galaxy clusters to act as a tracer for all of the mass in the Universe. Knowledge of the processes of galaxy formation and evolution that we discussed in the previous sections is essential to enable reliable conclusions to be drawn about the overall matter distribution from observations of galaxies.

A wide range of observations can be used to trace the matter distribution in the Universe. In earlier chapters you learned about the use of Type Ia supernovae to trace the overall expansion of the Universe, and of galaxy clusters to measure the halo mass function. In this section you will learn about some of the ways that surveys of very large numbers of galaxies can be used to learn about galaxy and structure formation.

11.3.1 Galaxy number counts

Galaxies (and clusters of galaxies) signify localised peaks in the matter distribution of the Universe. The number of these peaks and spacing between them can be used to test whether our understanding of structure formation and galaxy evolution is correct, and to test cosmological models.

In Section 10.3.3 we saw that measurements of the halo mass function could be used to test the predictions of numerical simulations. These types of comparison also allow us to determine which cosmological parameters (e.g. density parameters, neutrino masses, the cosmological constant's equation of state, and many others) are most consistent with the observed properties of the Universe.

The simplest statistical tool that can be used to compare observations with theoretical predictions is that of **number counts**. Here, a histogram is made of the number N of galaxies, or galaxy clusters, in a particular magnitude or flux range. Alternatively, a cumulative histogram can be constructed to show the number of galaxies or clusters brighter than a limiting flux or magnitude.

If the Universe was static and Euclidean (flat), with a relatively uniform distribution of galaxies, we would expect the number of galaxies of a given intrinsic brightness (luminosity) and within a given radius R to be $N(<R) \propto R^3$. Since flux is proportional to R^{-2}, this means that the number of galaxies above a flux F should be $N(>F) \propto F^{-3/2}$ or

$$\log\big[N(>F)\big] = -\tfrac{3}{2}\log(F) + \text{constant} \tag{11.8}$$

Deviations from this expression tell us about the expansion of the Universe and the evolution of structure formation. Since the Universe appears flat, deviations from Equation 11.8 were used as evidence against a steady-state Universe. Today, we use number counts of galaxies to identify how galaxies

have evolved, and number counts of clusters to investigate the formation and evolution of dark-matter halos.

Modern galaxy redshift surveys enable sophisticated tests of cosmological models using measurements of properties such as luminosity and **stellar mass** (total mass contained in stars) for large samples of galaxies across a wide range of redshifts.

11.3.2 Galaxy luminosity and stellar mass functions

The galaxy **luminosity function** and **stellar mass function** are defined similarly to the halo mass function we met in Chapter 10, but they are plotted as a *differential* quantity Φ. If n is the number density of galaxies per cubic megaparsec with luminosities greater than L, then:

Luminosity function

$$\Phi_{\mathrm{L}} = \frac{\mathrm{d}n}{\mathrm{d}\log_{10} L} \tag{11.9}$$

Similarly, if we instead consider the number density of galaxies above a stellar mass M, then:

Stellar mass function

$$\Phi_{\mathrm{M}} = \frac{\mathrm{d}n}{\mathrm{d}\log_{10} M} \tag{11.10}$$

Note that $\mathrm{d}\log_{10} L$ represents an increment in luminosity on a logarithmic scale, e.g. from 10^9 to $10^{10}\,\mathrm{L}_\odot$. Similarly, $\mathrm{d}\log_{10} M$ represents an increment in mass on a logarithmic scale. Hence Φ_{L} and Φ_{M} can be thought of as the number of galaxies per cubic megaparsec, per factor of ten in luminosity and mass, respectively.

It is worth noting that, somewhat confusingly, the luminosity function can also be reported in the form of optical magnitudes, which also has the symbol M. Usually it is fairly easy to tell which M is intended as in the case of magnitudes it will have a subscript to indicate a particular telescope filter, whereas a mass function will be reported using units of mass.

Figure 11.12 shows the observed galaxy luminosity function measured in the UV part of the spectrum and the corresponding stellar mass function for a range of redshifts from $z = 0.05$ (shown in dark brown) to $z = 4.0$ (shown in purple).

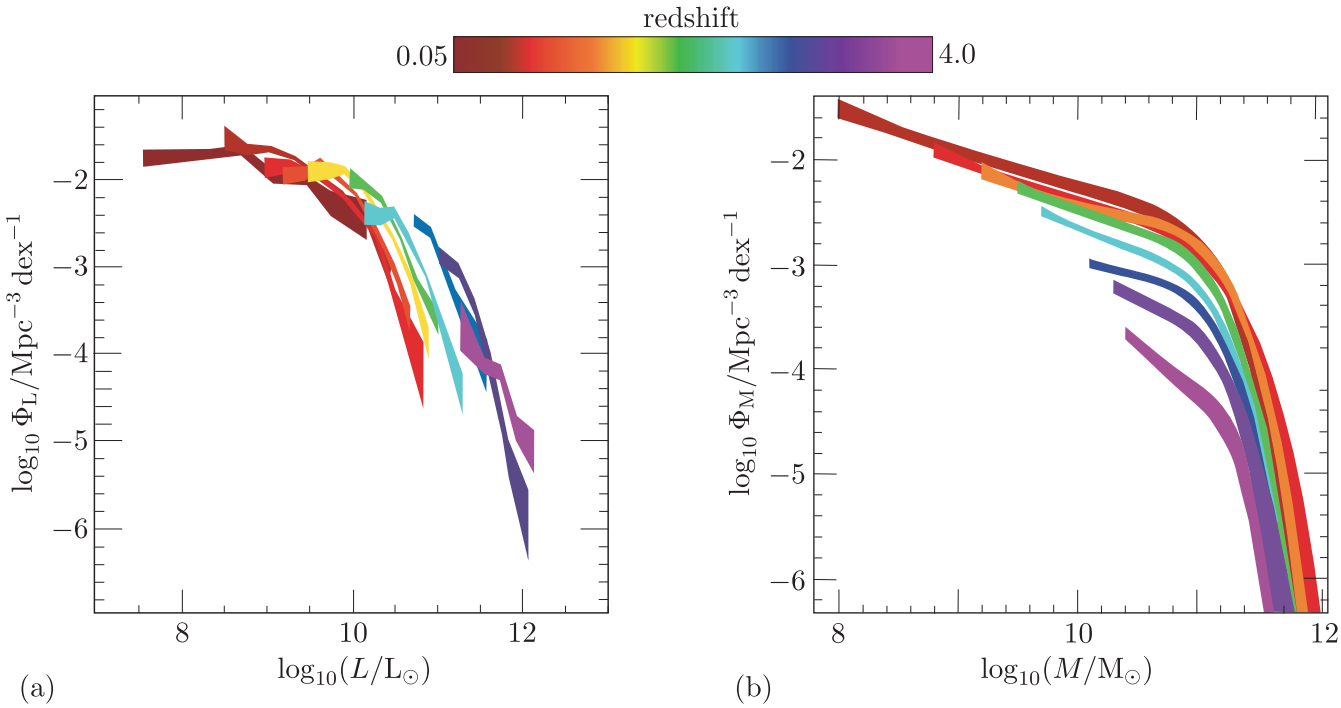


Figure 11.12 (a) The galaxy luminosity function from far-UV observations and (b) the corresponding mass function. In both cases, the vertical axis shows the logarithm of the number of galaxies per cubic megaparsec, per factor of ten (dex) in the horizontal quantity. The coloured bands show the luminosity and mass functions for the indicated redshifts.

- The highest redshift curves drop off at higher luminosity in Figure 11.12a than those at low redshift. Suggest an explanation for this decrease in the number of UV-bright galaxies with time.
- UV-bright stars are hot, blue stars. The far-UV is therefore a star-formation tracer. As the available gas in galaxies has declined, the amount of star formation has dropped from its peak at $z \approx 2$. This causes the point at which the far-UV luminosity function steepens to move left in the plot as we progress towards $z = 0$.
- Similarly, explain the increase in the number of massive galaxies at lower redshifts.
- Galaxy mergers and the continuing inflow of gas increase the mass of the most massive galaxies, meaning the mass function moves upwards with decreasing redshift. (The faint end of the mass function will also move down in time as these galaxies are removed, but faint galaxies are hard to find at very high redshifts.)

In Section 10.3.3 you saw that the observed and theoretical halo mass functions for galaxy clusters are in good agreement. It is also very informative to compare the mass functions for dark-matter halos with those for observed galaxies. Figure 11.13 shows a comparison between the theoretical halo mass function and observations of the stellar mass function, similar to those shown in Figure 11.12b.

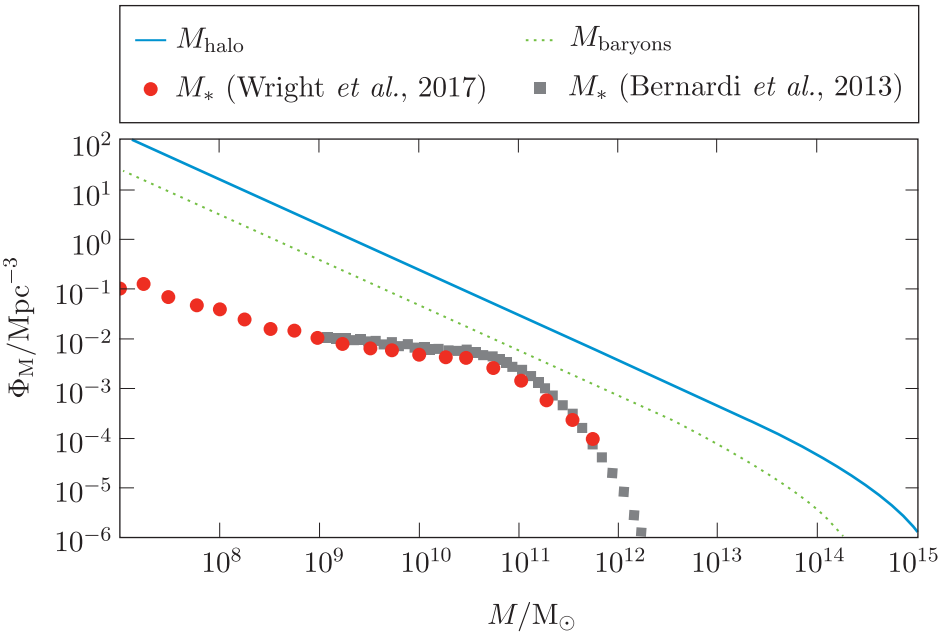


Figure 11.13 The mass function of dark-matter halos (solid blue line). The dotted line is offset to the left by a factor of about six, and shows the baryonic component of those halos. The points show the observed mass of stars in individual galaxies, according to two different surveys.

The shape of the stellar mass function is quite different from that of the dark-matter halos. The halo mass function has a uniform, fairly steep slope across a large mass range (starting to drop off at the very highest masses). The halo mass function in the range $M < 5 \times 10^{13}\,{\rm M}_\odot$ can be written as:

$$\frac{{\rm d}n}{{\rm d}M_{\rm halo}} = (1.72 \times 10^8\,{\rm M}_\odot^{0.9}\,{\rm Mpc}^{-3}) M_{\rm halo}^{-1.9} \tag{11.11}$$

where we have used the derivative with respect to mass rather than logarithmic mass to simplify the analysis.

In contrast, the stellar mass function has a shallow slope at low mass (sometimes known as the **faint-end slope**) and a steep slope at high mass. The turnover point between the two regimes occurs as a characteristic mass M_* corresponding to galaxies of mass fairly similar to the Milky Way.

The next example and exercise explore what can be learned from the different shapes of the halo and stellar mass functions.

Example 11.1

Using Equation 11.11, determine the number density of dark-matter halos in the mass range $10^{10} < M_{\rm halo} < 10^{11}\,{\rm M}_\odot$.

Solution

The differential number of dark-matter halos per cubic megaparsec is given by rearranging Equation 11.11 to give:

$$\mathrm{d}n = (1.72 \times 10^8\,\mathrm{M}_\odot^{0.9}\,\mathrm{Mpc}^{-3}) M_{\mathrm{halo}}^{-1.9}\,\mathrm{d}M_{\mathrm{halo}}$$

This expression can be integrated to determine the number density of dark-matter halos within a particular mass range, M_1 to M_2, as follows:

$$\begin{aligned} n &= \int_{M_1}^{M_2} (1.72 \times 10^8\,\mathrm{M}_\odot^{0.9}\,\mathrm{Mpc}^{-3}) M_{\mathrm{halo}}^{-1.9}\,\mathrm{d}M_{\mathrm{halo}} \\ &= (1.72 \times 10^8\,\mathrm{M}_\odot^{0.9}\,\mathrm{Mpc}^{-3}) \int_{M_1}^{M_2} M_{\mathrm{halo}}^{-1.9}\,\mathrm{d}M_{\mathrm{halo}} \end{aligned}$$

Evaluating the integral gives

$$n = (1.72 \times 10^8\,\mathrm{M}_\odot^{0.9}\,\mathrm{Mpc}^{-3}) \left[\frac{M^{-0.9}}{-0.9}\right]_{M_1}^{M_2} \tag{11.12}$$

$$= (1.91 \times 10^8\,\mathrm{M}_\odot^{0.9}\,\mathrm{Mpc}^{-3}) \left[M_1^{-0.9} - M_2^{-0.9}\right] \tag{11.13}$$

We can now substitute in $M_1 = 10^{10}\,\mathrm{M}_\odot$ and $M_2 = 10^{11}\,\mathrm{M}_\odot$ to give $n = 0.167$ dark-matter halos Mpc^{-3} within the specified mass range.

The next exercise involves some direct comparisons between the halo and stellar mass functions. A useful parameter to introduce at this point is the **star formation efficiency**, ϵ_*. The relationship between a galaxy's stellar mass and the mass of the halo in which it resides depends on ϵ_* as well as the baryon fraction within that halo, f_B (i.e. the fraction of the total halo mass consisting of baryons), according to:

$$M_* = \epsilon_* f_\mathrm{B} M_{\mathrm{halo}} \tag{11.14}$$

An efficiency of $\epsilon_* = 1$ would correspond to all the baryons in a dark-matter halo turning into stars.

Exercise 11.6

(a) Assuming that all dark-matter halos have a universal baryon fraction $f_\mathrm{B} = 0.15$ and a star-formation efficiency of $\epsilon_* = 1$, use Equations 11.11 and 11.14 to predict the number density of galaxies in the following stellar mass ranges:

(i) $10^8 < M_* < 10^9\,\mathrm{M}_\odot$
(ii) $10^{10} < M_* < 10^{11}\,\mathrm{M}_\odot$
(iii) $10^{11.5} < M_* < 10^{12.5}\,\mathrm{M}_\odot$.

(b) Compare your predictions from part (a) with the *observed* galaxy number densities in the middle of each mass range (Figure 11.13), and comment on what this suggests about the assumptions made.

Exercise 11.6 shows that stars must form much less efficiently in low- and high-mass halos compared to halos with intermediate mass. In fact, both the baryon fraction and star formation efficiency are thought to depend on the halo mass, and this is the explanation for why the shapes of the stellar mass and luminosity function are different from that of the halo mass function.

The processes of stellar and AGN feedback are thought to strongly influence star formation in the least massive and most massive halos. Modern simulations attempt to incorporate these effects to reproduce the observed galaxy luminosity function and other galaxy observations.

11.3.3 Clustering and the two-point correlation function

As well as cataloguing the properties of individual galaxies, the clustering together of galaxies on the sky also provides an important tool for cosmology. The clustering of galaxies can be measured via the angular **two-point correlation function**, $\omega(\theta)$. This is a measure of the distribution of angular separation distances between pairs of galaxies on the sky, here denoted by θ. To obtain this measure, all possible pairs of galaxies are considered, and so it captures whether or not it is more probable for galaxies to be clustered close to another galaxy or not.

To compute $\omega(\theta)$ we can take a galaxy and define a thin annulus (ring) around its location on the sky, with inner and outer angular radii of θ and $\theta + \mathrm{d}\theta$, respectively. We use $R(\theta)$ to denote the number of galaxies expected to be found within the annulus if they are randomly distributed – this can be calculated if you know the total number of galaxies in a survey catalogue that is being considered, and the sky area covered. $P(\theta)$ is then defined as the number of galaxies actually observed within that annulus.

The angular correlation function $\omega(\theta)$ is defined as

$$\omega(\theta) = \frac{\langle P(\theta) \rangle}{R(\theta)} - 1 \tag{11.15}$$

where $\langle P(\theta) \rangle$ is obtained by measuring $P(\theta)$ for each galaxy in a survey, and then taking an average.

If the distance to the galaxies is known, then the angular separation, θ, can be converted into a physical separation, r, and a corresponding *physical* two-point correlation function, $\xi(r)$, can be determined. Figure 11.14 shows observations of the galaxy two-point correlation function $\xi(r)$ from a low-redshift sky survey. Here $\xi(r)$ is calculated in a slightly more sophisticated way that accounts for biases that arise from working with different types of galaxy catalogues, but still represents the excess frequency of finding two galaxies at a particular distance from one another compared to a random distribution. Note the steep slope on the left-hand side and the 'bump' at around $100h^{-1}\,\mathrm{Mpc}$, caused by baryon acoustic oscillations, discussed below. (Recall that h represents the Hubble constant, H_0, in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.)

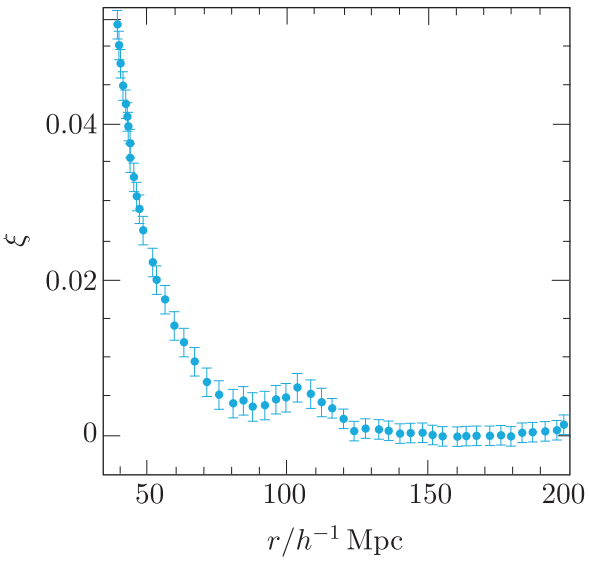


Figure 11.14 Observed galaxy two-point correlation function on large scales.

- What does the steep slope of the two-point correlation function at small r indicate about how galaxies are distributed?
- It is more probable for galaxies to be found at small distances from their neighbours than at large distances. In other words, galaxies cluster together, as predicted from cosmological simulations.

The 'bump' feature in Figure 11.14 indicates that there is an excess of galaxy pairs having a separation of $\sim 100h^{-1}$ Mpc (around 150 Mpc for $h = 0.677$). This is caused by a phenomenon known as baryon acoustic oscillations.

11.3.4 Baryon acoustic oscillations

Baryon acoustic oscillations (BAOs) originated in the oscillating photon–baryon fluid prior to recombination. A snapshot of these oscillations in the early Universe was frozen into the distribution of baryonic gas at decoupling. Hence, when the baryons cooled to form galaxies, a signature of these oscillations was imprinted into the distribution of galaxies. Today, the signals associated with BAOs are the largest recognisable structural feature in the Universe.

BAOs had a fixed physical size at the time of decoupling, because of the characteristic speed of sound in the photon–baryon fluid. All acoustic waves in this fluid travelled at a (highly relativistic) speed of $c_s \approx c/\sqrt{3}$. When the radiation pressure supported waves moving outward with respect to the centre of the overdensity dropped suddenly at recombination (see Section 11.1.1), these outward-moving perturbations had all reached the same distance $d_s = c_s t_{rec}$, known as the sound horizon (see further discussion in Chapter 7). Since the sound horizon is a fixed physical size (determined by the time of photon–baryon decoupling, t_{rec}), observational measurements of the scale on which the BAO signal is found at different

redshifts can be used as a standard ruler to measure the expansion of the Universe. Measurements of the BAO signal have contributed to pinning down the present value of the Hubble parameter, and it is hoped that measuring BAO features at multiple redshifts with future surveys can help to determine the equation of state and nature of dark energy.

Evolution of BAOs

Figure 11.15 illustrates how an initial overdensity of dark matter, baryons and photons evolves in a situation where the photon–baryon fluid is highly overpressured. The horizontal axis shows the *co-moving* size of the perturbation (remember that the true *physical* distance at a given redshift scales with $1 + z$). The vertical axis indicates the mass profile of the perturbation (i.e. the amount of mass for each type of material at a given radius). We can think of the mass profile as the amplitude of the perturbation as it propagates out. Note that the values on the vertical axis increase from one panel to the next as time increases. This is because the perturbation grows and the dark-matter halo collapses.

We can work through Figure 11.15 to explore how the different particle types interact and hence how the BAO signal arose. First, consider a dark-matter overdensity (i.e. density perturbation) into which more dark matter falls under gravity. The photon–baryon fluid also falls into this overdensity, but radiation pressure causes the fluid to rebound, which drives an acoustic wave outwards at the speed of sound.

- Panel (a) of Figure 11.15 shows the evolution of the density perturbation considered above shortly after its formation. Interpret the lines in this panel.
- The dark matter is starting to accumulate at the centre of the perturbation. Most of the mass is concentrated within about 10 Mpc co-moving distance (corresponding to 1.5 kpc physical distance at redshift $z = 6824$). Meanwhile, the photons and baryons, which are tightly coupled together at this redshift, are already expanding well beyond the concentration of dark matter, reaching a co-moving distance of 30 Mpc.

As we move to (b), dark matter is continuing to fall into the perturbation from larger radii. However, particle energy stops the dark matter from collapsing further, so it remains as a diffuse, overdense blob with a co-moving size of $\sim$10 Mpc. The perturbation in the photon–baryon fluid continues to propagate out at the speed of sound.

Between panels (b) and (c), last scattering occurs. We can see the photons keep expanding out in panels (c) and (d). However, the baryons remain almost fixed in place, concentrated in a spherical shell at a co-moving distance of $\sim$150 Mpc from the initial perturbation, which corresponds to a physical scale of $\approx$140 kpc at recombination, i.e. $z \approx 1100$.

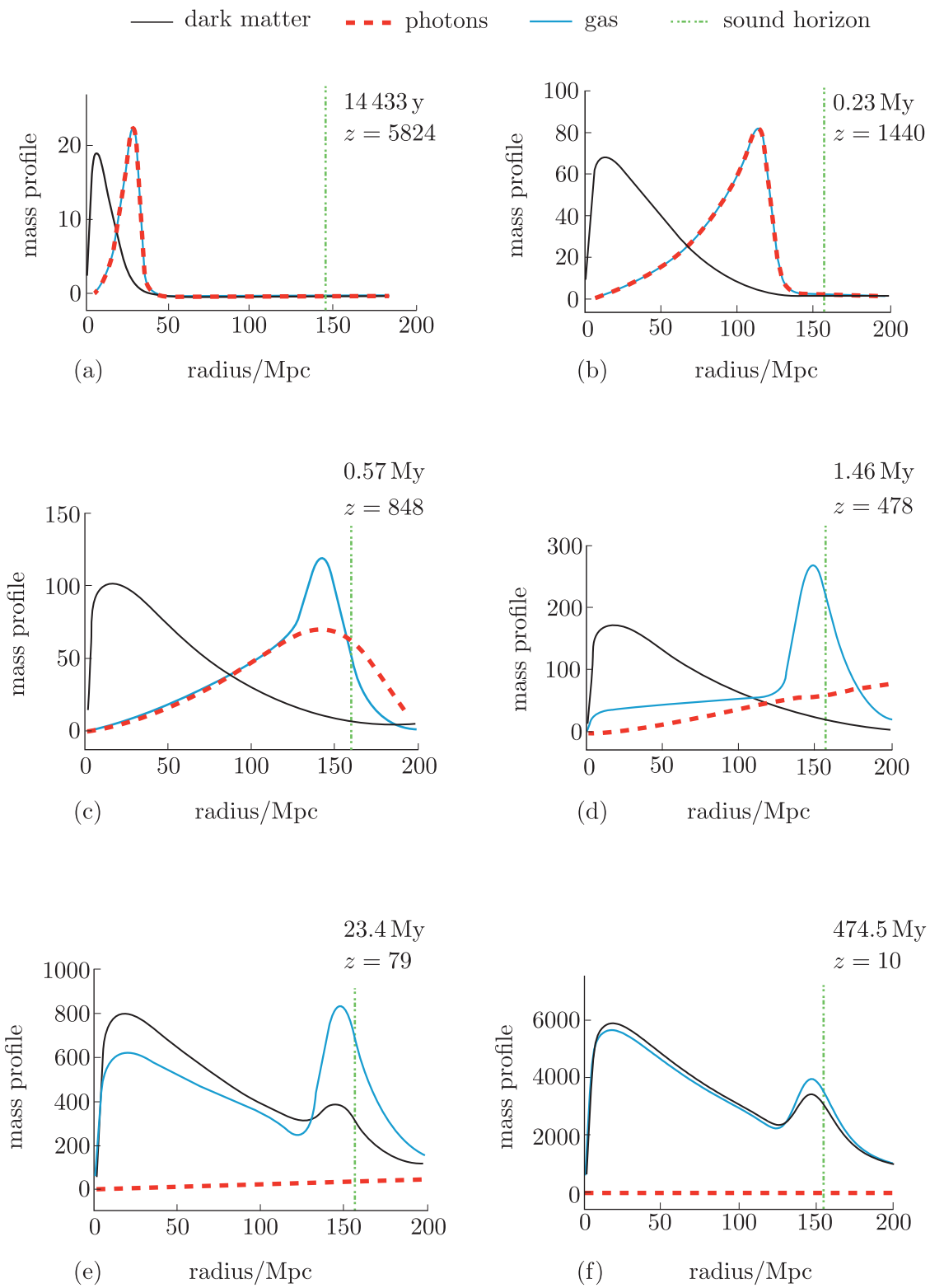


Figure 11.15 The evolution of a single, initially highly overpressured perturbation in the photon–baryon fluid, as well as the corresponding dark-matter perturbation, at high redshift. The horizontal axis shows *comoving* distance, and so reflects the *present-day* size of the regions considered.

Between panels (b) and (c), the dominant force on baryons stops being the electromagnetic interaction with photons, and starts becoming the gravitational attraction to the dark matter. Therefore, baryons begin falling into the nearest dark-matter halo. However, because baryons make up a significant fraction ($\approx$15%) of matter, this also has a gravitational effect on the dark matter, and so the dark-matter perturbation starts to expand a little towards the baryonic perturbation as well, as becomes more evident in panel (d).

By panel (d), about a million years after recombination and last scattering, we can see these collapse processes at work: the density of baryonic matter is now increasing in the inner regions. This continues through panels (e) and (f), until the mass profile amplitudes of the inner dark-matter and baryonic overdensities are almost equal.

We can also see from panels (d)–(f) that the dark matter is beginning to cluster on larger scales around the baryonic matter peak at 150 Mpc. Hence, by panel (f), at $z = 10$, dark matter and baryonic matter have become highly coupled. This process gives us two peaks in the matter profile: one peak around $\sim$10 Mpc seeded the most massive halos we see today; and a smaller peak at 150 Mpc, which is the BAO signature and represents a characteristic scale on which galaxies tend to be separated. The observable consequence of this smaller, outer peak is the bump feature in the two-point correlation function seen in Figure 11.14.

11.4 Summary of Chapter 11

- Before recombination, baryons were physically coupled to radiation, and the radiation pressure in the photon–baryon fluid was high. After decoupling, the baryonic gas collapsed into nearby dark-matter overdensities. These overdensities continued to grow during a period known as the dark ages, from $z \approx 1100$, when the Universe was $\sim$380 000 years old and lasting for around 200 million years, until $z \approx 20$.
- Self-interaction among baryons allows them to cool in a way dark matter cannot, so they can collapse into smaller clumps, including individual stars. The **cooling timescale** $t_{\rm cool}$ depends on gas density, temperature and the available cooling mechanisms, characterised by a **cooling rate** $\Lambda_{\rm cool}$ according to

$$t_{\rm cool} = \frac{\gamma n k_{\rm B} T}{2\Lambda_{\rm cool}} \qquad \text{(Eqn 11.1)}$$

  where γ is the number of degrees of freedom of the dominant particle type.
- The first (**Population III**) stars probably formed around $z \approx 17$ and were likely very massive. They ionised **Strömgren spheres** in the interstellar medium (ISM) around them.

- The size of a Strömgren sphere is limited by recombination to a radius of

$$R_{\max} \approx \left[\frac{3Q}{4\pi n_{\rm i} n_{\rm e} \alpha(T)}\right]^{1/3} \qquad \text{(Eqn 11.5)}$$

- The sphere's expansion slows exponentially as recombination becomes more significant, with expansion characterised approximately by

$$R(t) \approx R_{\max}\left[1 - \exp(-t/t_{\rm r})\right]^{1/3} \qquad \text{(Eqn 11.6)}$$

- The dense ISM within galaxies can recombine, allowing star formation to continue. However, recombination was extremely rare in intergalactic space at that time. The merged Strömgren spheres from entire galaxies rapidly ionised the diffuse **intergalactic medium** (IGM), probably by $z \approx 6$. It remains ionised to this day.
- Matter cycles in and out of stars over time. Stellar winds and supernovae release material back to the ISM, but some remains locked in compact **stellar remnants**.
- The nuclear-processed matter ejected from stars causes chemical enrichment of the ISM, spreading out astronomical metals within it and increasing the metallicity of the host galaxy.
- Even when stars die, explosions and mergers of compact objects can be sources of additional chemistry and feedback via supernovae and other outflows.
- **Supermassive black holes** (SMBHs) exist in the centre of galaxies. These may form by mergers of smaller compact objects through the process of **mass segregation**. SMBHs can form **active galactic nuclei** (AGN) and their associated **jets** and **radio lobes**.
- **Stellar feedback** and **AGN feedback** can both increase or decrease the star-formation rate of a galaxy, potentially driving outflows from the galaxy's gas reservoir.
- Galaxy **number counts** can be used to investigate the geometry of the Universe, while **luminosity functions** and **stellar mass functions** provide sophisticated tests of theoretical models and computer simulations that predict the evolution of structure in the Universe.
- The mass function for dark-matter halos can be modelled as a power law below the scale of galaxy clusters, while the stellar mass function has a more complicated form due to a varying efficiency of star formation with halo mass.
- The galaxy **two-point correlation function** measures the clustering of galaxies as a function of separation distance, and is a useful cosmological probe.
- **Baryon acoustic oscillations** (BAOs) are acoustic waves in the early Universe that imprint a 'bump' in the two-point correlation function at a characteristic size scale determined by the sound horizon at recombination. This scale is used as a standard ruler in cosmology.