232

12

Structure Formation: Baryons and Photons

If the only matter in the universe were cold dark matter, then structure formation would be driven solely by gravity. Adding baryonic matter, capable of absorbing, emitting, and scattering light, complicates the process of structure formation. As we saw in Section 11.2, before the time of decoupling at $z_{\text{dec}} = 1090$, the baryonic matter was coupled to the photons, thanks to the ability of photons and electrons to scatter from each other. At $z > 1090$, therefore, the interaction of baryonic matter with photons *prevented* dense baryonic lumps from forming. At lower redshifts, however, the interaction of baryonic matter with photons *encouraged* dense baryonic lumps to form, since the ability to radiate away excess thermal energy is necessary to make objects with high mass density.

The average density of baryonic matter today, at $t_0 \approx 13.7\,\text{Gyr}$, is

$\Omega_{b,0} = \frac{\rho_b}{\rho_{cr}} = 0.048 \Rightarrow$

$$\rho_{\text{bary},0} = 4.2 \times 10^{-28}\,\text{kg}\,\text{m}^{-3} = 6.2 \times 10^{9}\,\text{M}_\odot\,\text{Mpc}^{-3}. \tag{12.1}$$

(432)
$\rho_{cr,0} = \frac{\varepsilon_{c,0}}{c^2} \simeq 8.7 \times 10^{-27}\,\frac{\text{kg}}{\text{m}^3} \simeq 1.28 \times 10^{11}\,\frac{M_\odot}{\text{Mpc}^3}$

However, some parts of the universe are far above average when it comes to baryonic density. Let's look at a typical suburban location in a luminous galaxy: the region within a few hundred parsecs of the Sun. In the solar neighborhood, we find that the density of stars and interstellar gas is $\rho_{\text{sn}} \approx 0.095\,\text{M}_\odot\,\text{pc}^{-3} \approx 6.4 \times 10^{-21}\,\text{kg}\,\text{m}^{-3}$. This represents an overdensity

$$\delta_{\text{sn}} = \frac{\rho_{\text{sn}} - \rho_{\text{bary},0}}{\rho_{\text{bary},0}} \sim 2 \times 10^{7} \tag{12.2}$$

relative to the average baryonic density of the universe today. Now let's look at an individual main sequence star: the Sun itself. The Sun's average internal density is $\rho_\odot \approx 1400\,\text{kg}\,\text{m}^{-3}$, representing an overdensity $\delta_\odot \sim 3 \times 10^{30}$ relative to the average baryonic density today.[1] However, although baryons are capable of forming very dense objects, the majority of baryonic matter today is still in the form of low density intergalactic gas. To understand why some of the

[1] You are slightly less dense than the Sun, so $\delta_{\text{you}} \sim 2 \times 10^{30}$.

baryonic matter in the universe forms condensed knots such as stars, while most remains low in density, we start by making a census of the baryonic matter in the universe today.

12.1 Baryonic Matter Today

Although we know the average baryon density $\rho_{\mathrm{bary},0}$ quite well, the task of making a more detailed census of neutrons and protons (and their electron sidekicks) is frustratingly difficult. For example, in Section 7.1, we attempted to find the mass density of stars today, $\rho_{\star,0}$. Since stars glow at wavelengths that astronomers are highly experienced at detecting, stars should be the easiest baryonic component to detect. Nevertheless, we made only the rough estimate (Equation 7.4)

$$\rho_{\star,0} \approx 4 \times 10^8\,\mathrm{M}_\odot\,\mathrm{Mpc}^{-3} \approx 3 \times 10^{-29}\,\mathrm{kg\,m}^{-3}, \tag{12.3}$$

representing about 7 percent of the baryonic mass. Estimates of the other contributions to the baryonic matter are equally rough, if not more so. With that caveat in mind, Figure 12.1 shows how the baryonic component of the universe is divided up. The wedge labeled "Stars, etc." represents the 7% of the baryonic matter that is in the form of stars, stellar remnants, brown dwarfs, and planets. Proceeding clockwise around the pie chart of Figure 12.1, about 1% of the baryonic mass is in the *interstellar gas* that fills the space between stars within the luminous portion of a galaxy. About 3% is in the *circumgalactic gas* that is gravitationally bound within the dark halo of a galaxy, but lying outside the main distribution of the galaxy's stars. (Since there is no clean dividing line between the interstellar gas and the circumgalactic gas, you can think of them together as gas associated with individual galaxies, providing about 4% of the baryonic matter in the universe.) Approximately 4% of the baryonic matter is in the *intracluster gas* that is gravitationally bound within a cluster of galaxies, but is not bound to any particular galaxy within the cluster; we have already encountered intracluster gas (Figure 7.3) in the form of the hot, X-ray emitting gas of the Coma cluster. Thus, we conclude that baryonic matter associated with gravitationally bound systems (galaxies and clusters of galaxies) contributes just 15 percent of the total baryonic matter in the universe.

GAS INFLOW ⇓ ISM

GAS OUTFLOW ⇓ CSM

Where are the rest of the baryons? In the 1990s, astronomers began talking of a "missing baryon problem," when they began to realize that the baryonic matter in galaxies and clusters falls far short of $\rho_{\mathrm{bary},0}$. The missing baryons, however, aren't truly missing; they are simply in a low-density and inconspicuous intergalactic medium. As shown in Figure 12.1, about 40% of the baryonic matter is in the *diffuse intergalactic gas*, which consists of gas widely distributed outside galaxies and clusters, at a temperature $T < 10^5$ K. The remaining $\sim$ 45% of the

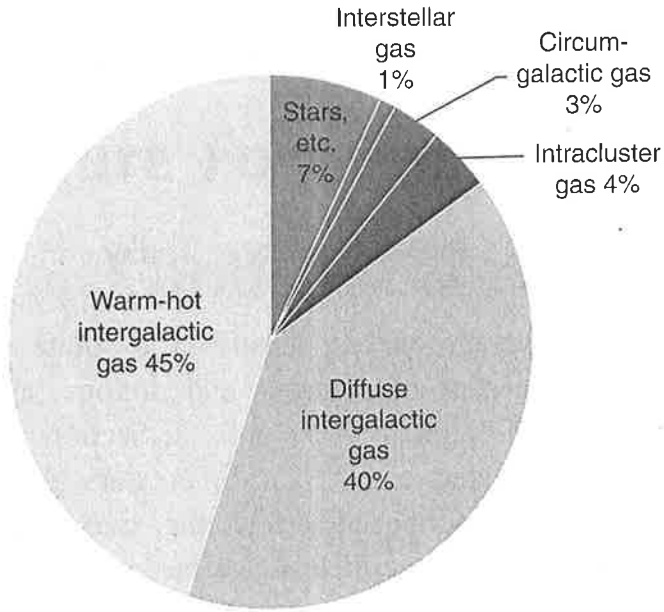


Figure 12.1 An approximate division of the current baryonic mass of the universe into its various components.

baryonic matter is in the *warm-hot intergalactic gas*, which is hotter than the diffuse intergalactic gas, typically having $10^5\,\mathrm{K} < T < 10^7\,\mathrm{K}$.[2] The warm-hot intergalactic gas is found in long filaments between clusters, as compared to the more smoothly distributed diffuse intergalactic gas.

DIG
WHIG

One striking characteristic of intergalactic gas is its low density. Most of the volume in the universe today is filled with diffuse intergalactic gas with $\delta \leq 0$. Even the warm-hot intergalactic gas, which tends to be higher in density, typically lies in the overdensity range $3 < \delta < 300$. Another striking characteristic of intergalactic gas today is its high degree of ionization. In most of the intergalactic gas, the fractional ionization of hydrogen, X, is very close to one. However, when we looked at the physics of recombination in Section 8.3, we noted that in the early universe, between $z = 1480$ and $z = 1260$, the fractional ionization of hydrogen plummeted from $X = 0.9$ to $X = 0.1$. Obviously, something happened after the epoch of recombination that *reionized* the hydrogen in the universe.

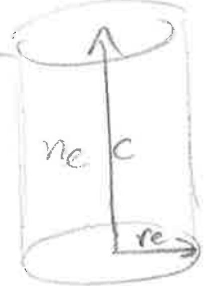
How was hydrogen reionized? To answer that question, it helps to ask the preliminary question, "*When* was hydrogen reionized?" Hydrogen was mostly neutral just after recombination ($z \approx 1380$); it is mostly ionized today ($z \approx 0$). The time that elapses between $z \approx 1380$ and $z \approx 0$ is $t \approx 0.99998 t_0 \approx 13.7\,\mathrm{Gyr}$. If we could pin down the time of reionization just a little more closely than that, we would have a useful clue about what physical mechanisms might be ionizing the hydrogen. (I am focusing on the reionization of hydrogen, and ignoring

[2] The fact that gas at 10^7 K is called "warm-hot" rather than "hot-hot" may seem like laughable understatement. However, it is being contrasted with the X-ray-emitting intracluster gas, which can reach $T \sim 10^8$ K or more.

helium, for the same reason I focused on hydrogen while discussing recombination in Section 8.2. Adding helium, and other heavier elements, makes the analysis more complicated mathematically, without changing the basic physical conclusions.)

12.2 Reionization of Hydrogen

Although we mainly think of the cosmic microwave background as telling us about the epoch of last scattering ($z_{\mathrm{ls}} \approx 1090$), it also contains useful information about the universe at lower redshifts. The reionized intergalactic gas provides an obstacle course of free electrons that the photons of the CMB must pass through to reach our microwave antennas. Each free electron has a cross-section $\sigma_e =$ $\pi r_e^2 \simeq$ $6.65 \times 10^{-29}\,\mathrm{m}^2$ for scattering with a photon. ($r_e \sim 5\times 10^{-15}$ m) The rate at which a CMB photon scatters from free electrons in the reionized gas is (compare to Equation 8.14)



$$\Gamma = n_e \sigma_e c, \tag{12.4}$$

where n_e is the number density of free electrons. If the baryonic matter is reionized starting at a time t_*, then the optical depth for scattering from the reionized gas is (compare to Equation 8.43) $\sigma_e = \pi r_e^2$

$$\tau_* = \int_{t_*}^{t_0} \Gamma(t)dt = c\sigma_e \int_{t_*}^{t_0} n_e(t)dt. \tag{12.5}$$

If the optical depth of the reionized gas were $\tau_* \gg 1$, then each CMB photon would be scattered many times in passing through the reionized gas, losing all information about its original direction of motion. Our view of the cosmic microwave background would then be completely smeared out, as if we were looking through a translucent screen. The fact that we can see temperature fluctuations down to small angular scales in the CMB, as shown in Figure 8.3, tells us that the optical depth of the reionized gas must be $\tau_* < 1$. ($\tau_* \ll 1$ as a matter of fact)

Looking out at the distant last scattering surface through the nearby reionized gas is like looking out through a slightly frosted window rather than one made of perfectly transparent glass. As a consequence, the CMB shows a small amount of smearing on small angular scales, corresponding to large values of the multipole l. The temperature fluctuations measured by the *Planck* satellite show slight suppression at high l compared to what you would expect in the absence of reionization. The amount of suppression is consistent with an optical depth $\tau_* = 0.066 \pm 0.016$ for the reionized gas. That is, about one CMB photon in 15 scatters from a free electron at low redshift.

Planck 2016

We can use the optical depth τ_* to estimate the time t_* of reionization, if we make some assumptions. First, let's assume the baryonic portion of the universe

$\tau \simeq 0.055 \pm .013 \Rightarrow z_{\mathrm{reion}} \simeq 8. \pm 1.3$

is pure hydrogen, either in the form of neutral atoms, with number density n_{H}, or in the form of free protons, with number density n_p. With this assumption,

$$n_{\text{H}} + n_p = n_{\text{bary}} = \frac{n_{\text{bary},0}}{a^3}. \tag{12.6}$$

Next, let's assume that the hydrogen undergoes complete, instantaneous reionization at the time t_*. In that case, the number density of free electrons before reionization is $n_e = 0$, and the number density after reionization ($t > t_*$) is

$$n_e = n_p = \frac{n_{\text{bary},0}}{a^3}. \tag{12.7}$$

Plugging this estimate for the number density of free electrons into the relation for optical depth (Equation 12.5), we find

$$\tau_* = \Gamma_0 \int_{t_*}^{t_0} \frac{dt}{a(t)^3}, \tag{12.8}$$

where

$$\Gamma_0 = c\sigma_e n_{\text{bary},0} = 1.58 \times 10^{-4}\,\text{Gyr}^{-1} \approx 0.0023 H_0 \tag{12.9}$$

is the rate at which photons would scatter from free electrons today, if the baryonic matter were a perfectly uniform distribution of fully ionized hydrogen.

Changing the variable of integration from t to a, equation 12.8 becomes

$$\tau_* = \Gamma_0 \int_{a(t_*)}^{1} \frac{da}{\dot{a}a^3} = \Gamma_0 \int_{a(t_*)}^{1} \frac{da}{H(a)a^4}, \tag{12.10}$$

using the fact that $H = \dot{a}/a$. Alternatively, we can find the redshift of reionization, z_*, by making the substitution $1 + z = 1/a$:

$$\tau_* = \Gamma_0 \int_0^{z_*} \frac{(1+z)^2 dz}{H(z)}. \tag{12.11}$$

During recent times, when the matter-dominated universe has been giving way to the lambda-dominated universe, the Hubble parameter has been, from Equation 5.96 $\frac{H^2}{H_0^2} = \frac{\Omega_{m,0}}{a^3} + (1 - \Omega_{m,0})$

$$H(z) = H_0 \left[\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}\right]^{1/2}. \tag{12.12}$$

Inserting this functional form for $H(z)$ into Equation 12.11 gives an integral with the analytic solution

$$\tau_* = \frac{2}{3\Omega_{m,0}} \frac{\Gamma_0}{H_0} \left([\Omega_{m,0}(1+z_*)^3 + \Omega_{\Lambda,0}]^{1/2} - 1 \right). \tag{12.13}$$

Using the relevant parameters from the Benchmark Model, the optical depth for scattering in the reionized universe is

$$\tau_* = 0.00485 \left([0.31(1+z_*)^3 + 0.69]^{1/2} - 1 \right). \tag{12.14}$$

0.055 ± 0.013

Given the observed optical depth, $\tau_* = 0.066 \pm 0.016$, we find that the redshift of reionization was $z_* = 7.8 \pm 1.3$ (6.8), corresponding to an age for the universe $t_* \sim$ 650 Myr (990). Thus, the "era of neutrality," when the baryonic matter consisted mainly of neutral atoms, was a relatively brief interlude in the history of the universe, with $t_* - t_{\rm rec} \sim 0.05 t_0$.

One way to ionize hydrogen is to bombard it with photons of energy $hf > Q = 13.6\,\text{eV}$. The obvious place to look for sources of ionizing photons at $z \geq 8$ is in galaxies. The highest redshift galaxies that have been discovered (so far) are at $z \sim 10$ ($z \simeq 10$–11?), so we know that galaxies were present at the time of reionization. One source of ionizing photons in galaxies is the hot, luminous O stars that contribute much of the luminosity in star-forming galaxies. Only the most massive O stars, those with $M \geq 30\,\text{M}_\odot$, are hot enough to contribute significantly to the background of ionizing ultraviolet radiation. As an example, an O star with $M \approx 30\,\text{M}_\odot$ produces ionizing photons at a rate $\dot{N}_* \approx 5 \times 10^{48}\,\text{s}^{-1}$. Thus, during its entire lifetime $t \approx 6\,\text{Myr} \approx 2 \times 10^{14}\,\text{s}$, the star will produce $N_* \approx \dot{N}_* t \approx 10^{63}$ ionizing photons.

see §3 of Windhorst et al (2018) ApJS, 234, 41

Another source of ionizing photons in a galaxy is an *active galactic nucleus*, or AGN. An AGN is a compact central region of a galaxy that is luminous over a broad range of the spectrum, including the range $hf > 13.6\,\text{eV}$ required to ionize hydrogen. AGNs are compact because they are fueled by accretion of matter onto a black hole. Luminous galaxies usually have a supermassive black hole at their center. The central black hole of our own galaxy, for instance, has a mass $M_{\rm bh} \approx 4 \times 10^6\,\text{M}_\odot$, and thus a Schwarzschild radius $2GM_{\rm bh}/c^2 \approx 0.08\,\text{AU}$. The galaxy NGC 4889, one of the two brightest galaxies in the Coma cluster (Figure 7.2), has a central black hole with mass $M_{\rm bh} \approx 2 \times 10^{10}\,\text{M}_\odot$ and Schwarzschild radius $2GM_{\rm bh}/c^2 \approx 400\,\text{AU}$. If a supermassive black hole is accreting gas, the heated gas can emit light before it slips through the event horizon; the energy of the light can be as much as $0.1mc^2$, where m is the mass of the accreted gas. A significant fraction of this emitted light takes the form of ionizing ultraviolet photons with $hf > 13.6\,\text{eV}$, coming from a region not far outside the Schwarzschild radius of the black hole. For a luminous AGN, the rate of production of ionizing photons is approximately

see W18 and also B. Smith et al 2018 ApJ, 853, 191

$r_s(2\times10^{10} M_\odot) = 0.002$ pc

$$\dot{N}_* \approx 3 \times 10^{56}\,\text{s}^{-1} \left(\frac{L_{\rm AGN}}{10^{13}\,\text{L}_\odot} \right). \tag{12.15}$$

The most luminous active galactic nuclei, with $L > 10^{13}\text{L}_\odot$, are often referred to as *quasars*.[3] A quasar with $L \sim 10^{13}\,\text{L}_\odot$ can emit as many ionizing photons in a month as our example O star does in its entire 6 million year lifetime.

[3] The term "quasar" is short for "quasi-stellar object," referring to the fact that these distant compact light sources are, like stars, unresolved by our telescopes. In other respects, quasars are dissimilar to stars; they are not powered by nuclear fusion, and they are tremendously more luminous than any single star.

Let this sink in: A QSO produces $10^{13} L_\odot$ within $r \lesssim 10\, d_{\rm Pluto} \simeq 400$ AU
[By the latest count, inside Pluto's orbit our SS contains $1.00\, L_\odot$

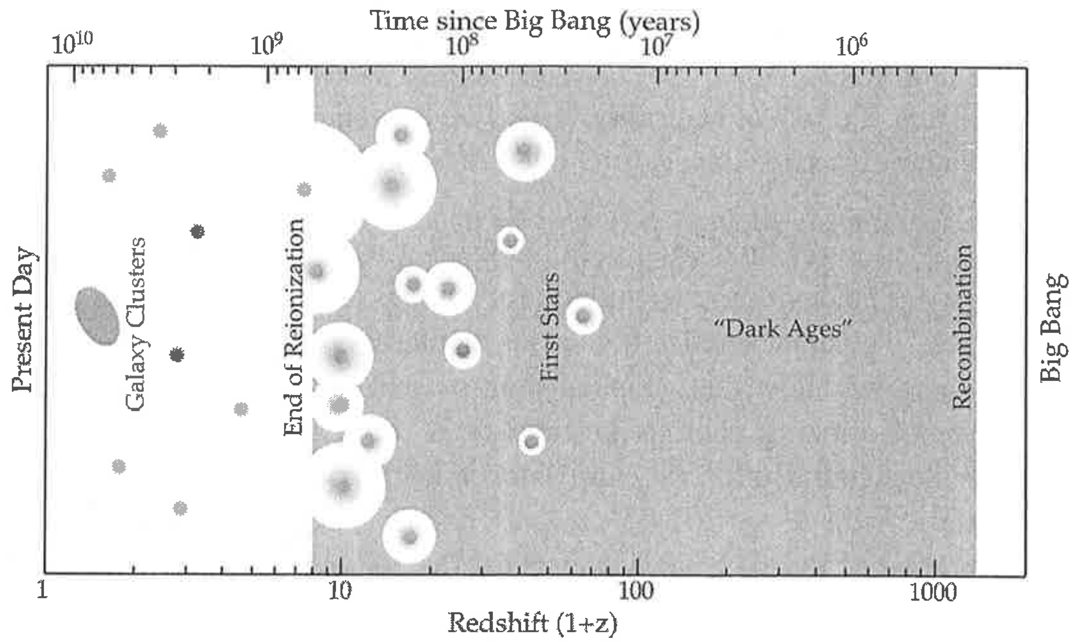


Figure 12.2 Schematic overview of the epoch of reionization; redshift decreases, and thus time increases, from right to left. The gray of the "Dark Ages" represents a gas of neutral atoms, while the white intergalactic spaces of the present day represent ionized gas.

12.3 The First Stars and Quasars

The time between recombination and the formation of the first stars and AGN is known to astronomers as the "Dark Ages" of the universe. The adjective "Dark," however, refers merely to the absence of starlight. The photons of the CMB were present during the Dark Ages. Indeed, immediately after the time of last scattering, at $z_{\rm ls} \approx 1090$, the temperature of the cosmic background radiation was $T \approx 2970\,{\rm K}$, about the temperature of an M star. If Heinrich Olbers had lived at the time of last scattering (admittedly an extremely large "if"), he would not have formulated Olbers' paradox. At $t \approx t_{\rm ls} \approx 0.37\,{\rm Myr}$, the entire sky *was* as bright as the surface of a star.

2.72(1091)
$T_0(1+z) =$

By the time the cosmic background radiation had cooled to a temperature $T \approx 140\,{\rm K}$, at $z \approx 50$, the universe was filled with a cosmic far infrared background, utterly ineffective at ionizing hydrogen. However, as shown in Figure 12.2, a redshift $z \approx 50$, corresponding to a cosmic time $t \approx 50\,{\rm Myr}$, is about the time when the very first stars began to emit light.[4] The Dark Ages ended as increasingly large numbers of stars and AGN poured out photons, some of them with $hf > 13.6\,{\rm eV}$. Figure 12.2 illustrates, in a schematic way, how regions of reionized gas began to grow around isolated galaxies, until the regions merged to form a single ionized intergalactic medium at $z \sim 8$.

[4] Exactly when the first stars formed is understandably conjectural, given our lack of direct observational evidence; most estimates fall in the range $z = 50 \rightarrow 20$, corresponding to $t = 50 \rightarrow 180\,{\rm Myr}$.

One complicating factor with using light from O stars and AGN to reionize intergalactic gas is that much of the ionizing radiation never escapes into intergalactic space. Instead, it is absorbed by the gas and dust within the galaxy in which the O star or AGN exists. The escape fraction, $f_{\rm esc}$, represents the fraction of ionizing photons that actually leak out into the intergalactic gas. The escape fraction is poorly determined; as an optimistic first guess, we can adopt $f_{\rm esc} = 0.2$, both from hot stars and from AGN. Now, let's consider a comoving cubic megaparsec of space, expanding along with the universal expansion, and compute how many ionizing photons we need to reionize its hydrogen. The comoving number density of baryons in intergalactic space is

$$n_{\rm bary} = 0.25\,{\rm m}^{-3} = 7.3 \times 10^{66}\,{\rm Mpc}^{-3}. \tag{12.16}$$

In our pure hydrogen approximation, ionizing this many neutral hydrogen atoms requires a comoving number density of ionizing photons equal to

see B. Smith et al 2018 (ApJ 853, 191 for a discussion $f_{esc} \simeq 0.1-0.3$!

$$n_* = \frac{n_{\rm bary}}{f_{\rm esc}} = 3.7 \times 10^{67}\,{\rm Mpc}^{-3}\left(\frac{0.2}{f_{\rm esc}}\right). \tag{12.17}$$

Thus, assuming an escape fraction $f_{\rm esc} \approx 0.2$, ionizing a comoving cubic megaparsec of hydrogen requires $N_* \sim 4 \times 10^{67}$ ionizing photons; this is the number created by $\sim 40\,000$ O stars, or by a $10^{13} L_\odot$ quasar shining for 4000 yr.

Given the immense ultraviolet luminosity of a quasar compared to even the hottest, brightest stars, it is tempting to assume that quasars perform most of the reionization. However, quasars, in addition to being breathtakingly luminous, are breathtakingly rare. Surveys like the 2dF survey and the Sloan Digital Sky Survey permit us to make quantitative statements about the scarcity of quasars as a function of time (or equivalently, of redshift). Figure 12.3 summarizes how the number of luminous quasars per comoving cubic megaparsec has changed over time. The comoving number density of quasars was greatest at a redshift $z \approx 2.5$, corresponding to a cosmic time $t \approx 2.6$ Gyr, long after reionization was complete. The comoving number density of quasars at $z_* \approx 8$ is more uncertain. Even the optimistic extrapolation shown in Figure 12.3 suggests that the number density was one luminous quasar for every $\sim 10^{10}$ comoving cubic megaparsecs. If reionizing one comoving cubic megaparsec of hydrogen requires 4000 years' worth of a quasar's luminosity, then reionizing $\sim 10^{10}$ comoving cubic megaparsecs would require $\sim 40\,000$ Gyr. Since the age of the universe at the time of reionization was only $t \sim 0.65$ Gyr, luminous quasars could have reionized only a small fraction of the intergalactic hydrogen.

Solution: Weak AGN or very faint Quasars $\Leftrightarrow$ Accretion disks around 30–300 $M_\odot$ BHs (Ligo!) See Windhorst et al 2018 ApJS 234, 41

Adding the ionizing radiation from lower-luminosity AGN to that from luminous quasars will help to boost the number of ionizing photons created. We know that in the present universe, there are more AGN with low luminosity than there are quasars with high luminosity. However, if the luminosity function of AGN at $z_* \sim 8$ is at all similar to that at $z \sim 0$, then even adding the efforts of the

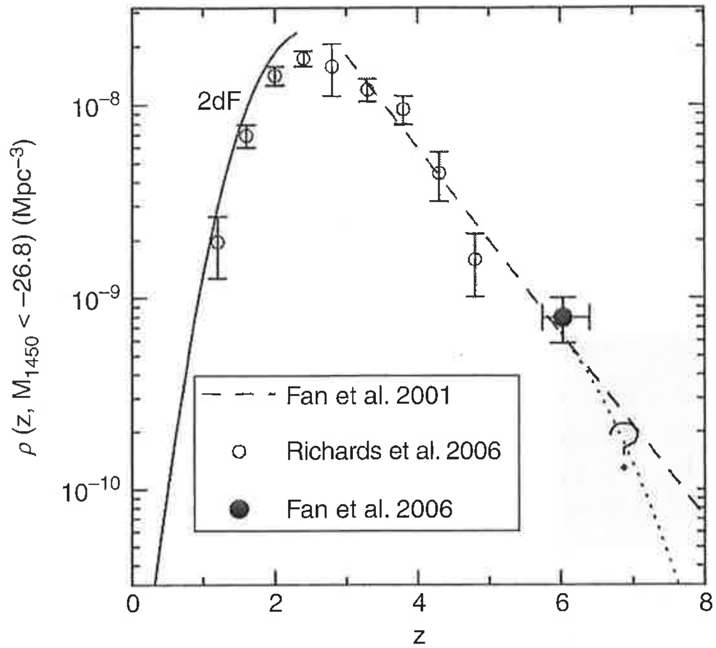


Figure 12.3 Evolution of the comoving number density of luminous quasars, based on data from the 2dF survey and the Sloan Digital Sky Survey. Redshift decreases, and thus time increases, from right to left. [Fan 2012, *RAA*, **12**, (*8*), 865]

lower-luminosity AGN will be inadequate to reionize the universe using the light from active galactic nuclei alone. It seems that most of the photons that reionized the universe came from hot, luminous stars. or from LIGO BHs in binaries with $M_1 \gtrsim 30 M_\odot$ $M_2 \gtrsim 20 M_\odot$.

At the present time, some galaxies are actively forming stars, while others are quiescent. However, if we average over a large volume containing a representative sample of galaxies, we find that the star formation rate today is $\dot{\rho}_{\star,0} \approx 20\,000\,\mathrm{M_\odot\,Myr^{-1}\,Mpc^{-3}}$. Since the baryon density today is $\rho_{\mathrm{bary},0} \approx 6 \times 10^9\,\mathrm{M_\odot\,Mpc^{-3}}$, this means that 3 parts per million of the universe's baryons are being converted into stars every million years. However, studies of star formation at higher redshifts reveal that the star formation rate has varied with time. Figure 12.4, for instance, shows an observationally based estimate of the star formation rate per comoving cubic megaparsec as a function of redshift. The rate at which stars form today (within a comoving volume) is down by a factor of 10 from its maximum in the redshift range $z = 4 \to 1$, corresponding to cosmic times $t = 1.5 \to 6\,\mathrm{Gyr}$.[5]

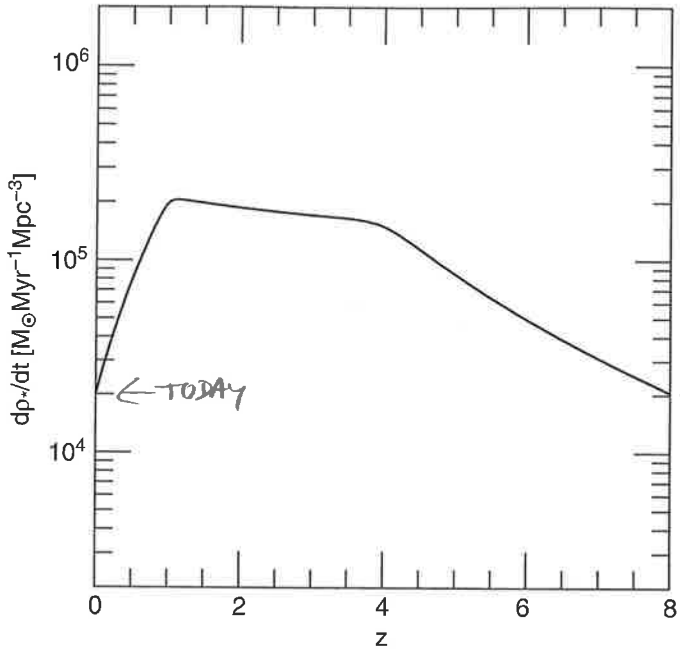
At the time of reionization, $z_* \approx 8$, the star formation rate per comoving volume was about the same as it is today:

$$\dot{\rho}_\star(z = 8) \approx 20\,000\,\mathrm{M_\odot\,Myr^{-1}\,Mpc^{-3}}. \tag{12.18}$$

However, not all stars are equally useful for reionizing the intergalactic gas. In a comoving cubic megaparsec where stars are created at a rate of $20\,000\,\mathrm{M_\odot\,Myr^{-1}}$,

[5] The intense star-forming era from $z = 4$ to $z = 1$ is when all the action was, baryonically speaking. We live today in a boring era, with fewer new stars and only feeble AGN.

241



← END OF REIONIZATION AT $z \simeq 7$-8.

Figure 12.4 Evolution of the comoving density of star formation in the universe. Redshift decreases, and thus time increases, from right to left. [adapted from Yüksel *et al.* 2008, *ApJ*, **683**, 5]

only $\sim 2000\,M_\odot\,Myr^{-1}$, or about 10 percent of the total, will take the form of O stars with $M \geq 30\,M_\odot$, hot enough to emit a significant number of ionizing photons. If we make the simplifying assumption that all the O stars have $M \approx 30\,M_\odot$, this implies the production of $\sim$ 67 new O stars per comoving cubic megaparsec every million years. Since the lifetime of a $30\,M_\odot$ star is $\sim 6\,Myr$, this means there will be about 400 O stars present per comoving cubic megaparsec at any time, as long as the star formation rate remains at the level given by Equation 12.18. With 400 O stars each emitting ionizing photons at a rate $\dot{N}_* \approx$ Pg 237! → $5 \times 10^{48}\,s^{-1}$, this means that the total rate of ionizing photon production per comoving cubic megaparsec is

$$\dot{n}_* \approx (5 \times 10^{48}\,s^{-1})(400\,Mpc^{-3}) \approx 2 \times 10^{51}\,s^{-1}\,Mpc^{-3} \approx 6 \times 10^{64}\,Myr^{-1}\,Mpc^{-3}. \quad (12.19)$$

If we compare this rate to the total number n_* of ionizing photons required to reionize the hydrogen in the same comoving volume (Equation 12.17), we find that the star formation has to continue for a time

That is t_H at $z \simeq 8.3$

$$t = \frac{n_*}{\dot{n}_*} \approx \frac{3.7 \times 10^{67}\,Mpc^{-3}}{6 \times 10^{64}\,Myr^{-1}\,Mpc^{-3}}\left(\frac{0.2}{f_{esc}}\right) \approx 600\,Myr\left(\frac{0.2}{f_{esc}}\right) \quad (12.20)$$

in order to reionize the intergalactic gas. As long as $f_{esc} \geq 0.2$, this time is not impossibly long compared to the age of the universe at the time it was reionized, $t_* \approx 650\,Myr$. Although many of the details of reionization remain to be worked but see WB: Since most SF at z≥7 occurred at z≃7-8!, you likely need the help from stellar mass BH's in binaries!

out, this back-of-envelope accounting indicates that massive stars are capable of emitting most of the photons required to reionize the intergalactic gas.

Reionizing the baryonic universe with starlight requires forming stars at a comoving rate $\dot{\rho}_\star \approx 20\,000\,\mathrm{M}_\odot\,\mathrm{Myr}^{-1}\,\mathrm{Mpc}^{-3}$ for a time $t \approx 600\,\mathrm{Myr}$. By the time reionization is complete, the comoving number density of stars that have been formed is $\rho_\star = \dot{\rho}_\star t \approx 1.2 \times 10^7\,\mathrm{M}_\odot\,\mathrm{Mpc}^{-3}$, or about 0.2% of the total baryon density. Thus, converting one part in 500 of the baryonic mass into stars has the side effect of reionizing the remaining baryonic gas. At times $t > t_*$, the ongoing star formation sketched out in Figure 12.4 drives the stellar mass density up to $\rho_{\star,0} \approx 4 \times 10^8\,\mathrm{M}_\odot\,\mathrm{Mpc}^{-3}$ today. The accompanying production of ionizing photons from O stars (with help from AGN) enables the intergalactic medium to remain ionized despite the existence of radiative recombination.

and stellar mass BH's accretion disks at z≥6-8

12.4 Making Galaxies

The current mass density of stars, $\rho_\star \approx 4 \times 10^8\,\mathrm{M}_\odot\,\mathrm{Mpc}^{-3}$, produces a luminosity density $\Psi_V = 1.1 \times 10^8\,\mathrm{L}_{\odot,V}\,\mathrm{Mpc}^{-3}$. In addition, this mass density implies a stellar number density $n_\star \sim 10^9\,\mathrm{Mpc}^{-3}$ (the number we used back in Section 2.1, discussing Olbers' paradox). These stars are not uniformly distributed in space; instead, they tend to be contained within galaxies, which consist of a relatively small concentration of stars and interstellar gas in the midst of a larger halo, consisting mainly of dark matter with only a tenuous circumgalactic gas of baryons.

The observed luminosity function for galaxies, $\Phi(L)$, is defined so that $\Phi(L)dL$ is the number density of galaxies in the luminosity range $L \rightarrow L + dL$. It is found that the luminosity function is well fitted by the function[6]

POWER LAW FAINT END ($L \ll L^*$) EXPONENTIAL BRIGHT END ($L \gg L^*$)

$$\Phi(L)dL = \Phi^* \left(\frac{L}{L^*}\right)^\alpha \exp\left(-\frac{L}{L^*}\right) \frac{dL}{L^*}. \tag{12.21}$$

$\alpha \simeq -2$ at z≥6!

Surveys in the V band find a power-law slope $\alpha \approx -1$, a normalization $\Phi^* \approx 0.005\,\mathrm{Mpc}^{-3}$, and a characteristic luminosity $L_V^* \approx 2 \times 10^{10}\,\mathrm{L}_{\odot,V}$, comparable to the luminosity of our own galaxy. Figure 12.5 shows a characteristic luminosity function for galaxies.

Thanks to the exponential cutoff in the Schechter function, galaxies with $L > L^*$ are exponentially rare. Although a few galaxies with $L \approx 10L^*$ exist, such as NGC 4889 and NGC 4874 in the Coma cluster, they are very uncommon. They exist only in rich clusters of galaxies, where they have grown to vast size by cannibalizing other galaxies.[7] Our own galaxy, with $L_V \approx L_V^* \approx 2 \times 10^{10}\,\mathrm{L}_{\odot,V}$, is about "average"

[6] This function is called the Schechter luminosity function, after the astronomer Paul Schechter, who pioneered its use.

[7] Although it sounds rather gruesome, "cannibalism" is the usual technical term for a merger between a small galaxy and a large one.

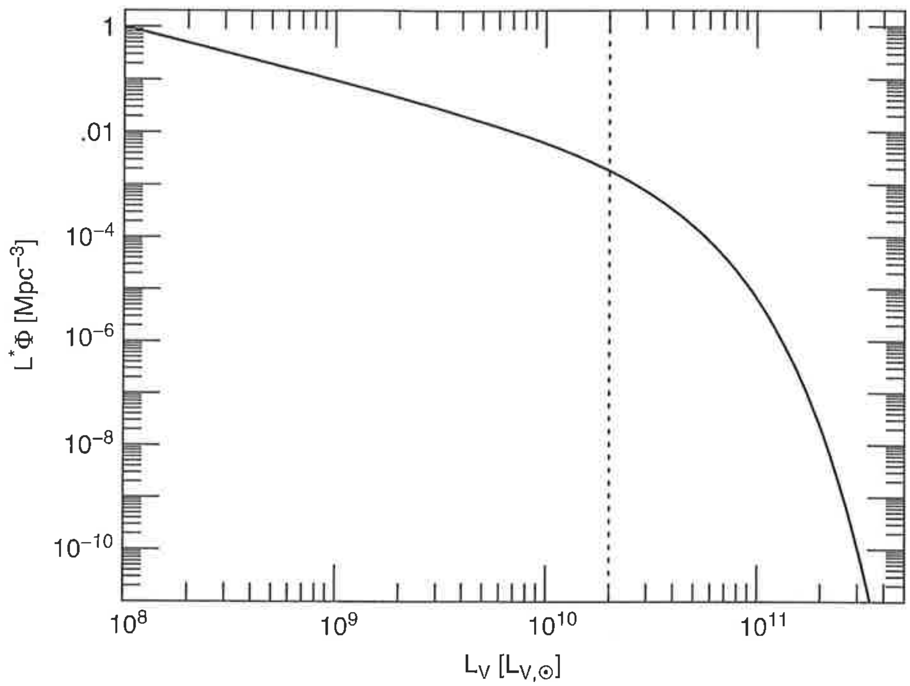


Figure 12.5 The luminosity function of galaxies today, observed in the V band. A Schechter function is assumed, with $\alpha = -1$ and $\Phi^* = 0.005\,\text{Mpc}^{-3}$. The vertical dotted line indicates the value of $L_V^* = 2 \times 10^{10}\,L_{\odot,V}$.

has a baryonic mass $M_{\text{bary}} \approx 1.2 \times 10^{11}\,\text{M}_\odot$, including all its interstellar and circumgalactic gas. Its total mass, provided mostly by the dark halo, is $M_{\text{tot}} \approx (1 \to 2) \times 10^{12}\,\text{M}_\odot$, depending on how far the halo extends. The exponential cutoff in the luminosity function at $L_V > L_V^*$ indicates that it is very difficult to make galaxies with a baryonic mass $M_{\text{bary}} > 10^{11}\,\text{M}_\odot$ and a total mass $M_{\text{tot}} > 10^{12}\,\text{M}_\odot$. It is definitely possible to make dark halos with a mass greater than $10^{12}\,\text{M}_\odot$; the dark halo of the Coma cluster, for instance, has $M \approx 2 \times 10^{15}\,\text{M}_\odot$. However, highly massive dark halos embrace multiple smaller galaxies rather than a single gargantuan galaxy. This hints that the difficulty in making jumbo galaxies, with $L_V > 2 \times 10^{10}\,\text{L}_{\odot,V}$ and $M_{\text{bary}} > 10^{11}\,\text{M}_\odot$, must have something to do with the properties of baryons rather than the properties of dark matter.

To see why there is an upper limit on the size of galaxies, we need a simplified model of how a galaxy forms. Consider a spherical overdense region at the time of radiation–matter equality ($t_{rm} \approx 0.050\,\text{Myr}$); this sphere will eventually become a luminous, star-filled galaxy and its surrounding dark halo. Initially, the relation among the mass M of the sphere, its radius R, and its overdensity δ is given by Equation 11.5:

(11.5) $\Longrightarrow$

$$M = \frac{4\pi}{3}\rho_m(t)[1 + \delta(t)]R(t)^3, \tag{12.22}$$

where $\rho_m(t) = \rho_{m,0}(1 + z)^3$ is the average mass density of the universe at time t and redshift $z = 1/a(t) - 1$.

At first, the overdensity is small: $\delta(t_{rm}) = \delta_{rm} \ll 1$. As long as $\delta(t) \ll 1$, the sphere's expansion is nearly indistinguishable from the Hubble expansion of the universe. However, the sphere reaches a maximum radius, and begins to collapse under its self-gravity, at the time $t_{\rm coll}$ when $\delta(t_{\rm coll}) \approx 1$. Given that $\delta \propto a \propto t^{2/3}$ ← (11.58) during the matter-dominated epoch, we can make the approximation that $t_{\rm coll} \approx \delta_{rm}^{-3/2} t_{rm}$, or

$$1 + z_{\rm coll} \approx \delta_{rm}(1 + z_{rm}). \tag{12.23}$$

Thus, regions with a higher initial overdensity δ_{rm} will collapse at an earlier time, corresponding to a higher redshift.

At the moment when the sphere starts to collapse, its density $\bar{\rho}$ is

$$\bar{\rho}(t_{\rm coll}) \approx 2\rho_m(t_{\rm coll}) \approx 2\rho_{m,0}(1 + z_{\rm coll})^3. \tag{12.24}$$

After the sphere collapses, it oscillates in and out a few times before coming into an equilibrium state; it is now a gravitationally bound halo with a radius $R_{\rm halo} \approx R(t_{\rm coll})/2$. The process by which the collapsing structure comes into equilibrium is called "virialization," since the resulting halo obeys the virial theorem. Since the radius of the virialized halo is half the radius it had at $t_{\rm coll}$, the average density of the halo is now

$$\bar{\rho}_{\rm halo} \approx 8\bar{\rho}(t_{\rm coll}) \approx 16\rho_{m,0}(1 + z_{\rm coll})^3. \tag{12.25}$$

From the average density of a galaxy's virialized halo, you can deduce when it started its collapse. For instance, the mass of our own galaxy, contained mainly in its dark halo, is (from Equation 7.12)

(7.12)

$$M_h = 1.9 \times 10^{12}\,{\rm M}_\odot \left(\frac{R_{\rm halo}}{0.15\,{\rm Mpc}}\right), \tag{12.26}$$

going from 8.2 → 150 kpc =150 kpc!

giving an average density

$$\bar{\rho}_h = \frac{3M}{4\pi R_{\rm halo}^3} = 1.4 \times 10^{14}\,{\rm M}_\odot\,{\rm Mpc}^{-3} \left(\frac{R_{\rm halo}}{0.15\,{\rm Mpc}}\right)^{-2}$$

using $\rho_{m,0} \simeq 0.32\rho_{crit} \simeq 0.32 \times 1.28 \times 10^{11}$ (4.32) (+12.1)

$$= 3400\rho_{m,0}\left(\frac{R_{\rm halo}}{0.15\,{\rm Mpc}}\right)^{-2}. \tag{12.27}$$

Combined with Equation 12.25, this implies that our own galaxy started its collapse at a redshift given by the relation

I.e. for $R_h(GXY) \simeq 150$ kpc

$$1 + z_{\rm coll} \approx 6\left(\frac{R_{\rm halo}}{0.15\,{\rm Mpc}}\right)^{-2/3}. \tag{12.28}$$

≃20 Y_{HL}

⇒ $z_{coll} \simeq 5$! (LATE!)

The process of virialization is not gentle; it involves subclumps of baryonic gas slamming into each other, creating shocks that heat the gas. A possible end state for the hot gas is a spherical distribution in hydrostatic equilibrium.

245

(As an example of such an end state, consider the hot intracluster gas in the Coma cluster, shown in Figure 7.3.) We can compute how hot the gas must be to remain in hydrostatic equilibrium. From Equation 7.41, a sphere of gas in hydrostatic equilibrium obeys the relation

$$(7.41) \Rightarrow \quad M(r) = \frac{kT_{\rm gas}(r)r}{G\mu}\left[-\frac{d\ln\rho_{\rm gas}}{d\ln r} - \frac{d\ln T_{\rm gas}}{d\ln r}\right], \tag{12.29}$$

where $M(r)$ is the total mass within a radius r and μ is the mean mass per gas particle. For an ionized primordial mix of hydrogen and helium, with $Y = Y_p = 0.24$, the mean mass is $\mu = 0.59m_p$. For simplicity, assume a single temperature $T_{\rm gas}$ for all the gas in the halo, and a power law $\rho_{\rm gas} \propto r^{-\beta}$ for the gas profile. Evaluated at $r = R_{\rm halo}$, where $M(r) = M_{\rm tot}$, the equation of hydrostatic equilibrium yields

$$kT_{\rm gas} = \frac{GM_{\rm tot}\mu}{\beta R_{\rm halo}}. \tag{12.30}$$

This temperature is known as the *virial temperature* for gas in a virialized halo. Since the halo radius is

$$(12.25) \Rightarrow \quad R_{\rm halo} = \left(\frac{3M_{\rm tot}}{4\pi\bar{\rho}_{\rm halo}}\right)^{1/3} = \left(\frac{3M_{\rm tot}}{64\pi\rho_{m,0}(1+z_{\rm coll})^3}\right)^{1/3}, \tag{12.31}$$

we can rewrite the virial temperature as a function of a halo's mass $M_{\rm tot}$ and the redshift $z_{\rm coll}$ at which its collapse began:

$$kT_{\rm gas} = \frac{4}{\beta}\left(\frac{\pi}{3}\right)^{1/3} G\mu\rho_{m,0}^{1/3}M_{\rm tot}^{2/3}(1+z_{\rm coll}). \tag{12.32}$$

Massive halos that collapsed early (and thus have high density) have the highest virial temperature.

If we assume the gas in the halo is hot enough to be mostly ionized, then $\mu \approx 0.6m_p$. Dark halos today have $\rho \propto r^{-2}$; if the hot gas started with a similar profile to the dark matter, we can take $\beta \approx 2$, and compute a numerical value for the temperature of the hot gas:

$$T_{\rm gas} \approx 1.0 \times 10^6\,{\rm K}\left(\frac{M_{\rm tot}}{10^{12}\,{\rm M}_\odot}\right)^{2/3}\left(\frac{1+z_{\rm coll}}{5}\right). \tag{12.33}$$

Here, I have scaled the value of $1 + z_{\rm coll}$ to a collapse starting at redshift $z_{\rm coll} = 4$, at the start of the intense star-forming era, and consistent with the value for our own galaxy.

A universe with cold dark matter is a "bottom-up" universe, in which low-mass halos tend to collapse at earlier times than high-mass halos. However, the right panel of Figure 11.5 reminds us that the root mean square density fluctuation, $\delta M/M$, is not strongly dependent on mass; at $M \sim 10^{12}\,{\rm M}_\odot$, the dependence is $\delta M/M \propto M^{-0.14}$. Thus, the redshift of collapse should be only weakly dependent on halo mass. Consider, for example, a halo of mass $M_{\rm tot}$ that started as a modestly unusual 2σ density fluctuation; that is, at the time of radiation-matter equality, it had an overdensity δ_{rm} equal to $2 \times (\delta M/M)_{rm}$. For a halo mass $M_{\rm tot} = 10^{12}\,{\rm M}_\odot$, a 2σ fluctuation started its collapse at a redshift $z_{\rm coll} \approx 3.4$, and thus had a virial gas temperature $T_{\rm gas} \approx 9 \times 10^5\,{\rm K}$. For a much smaller halo mass, $M_{\rm tot} = 10^{10}\,{\rm M}_\odot$, a 2σ fluctuation started its collapse earlier, at $z_{\rm coll} \approx 6.7$, and had a virial temperature $T_{\rm gas} \approx 70\,000\,{\rm K}$. For a much larger halo mass, $M_{\rm tot} = 10^{14}\,{\rm M}_\odot$, a 2σ fluctuation started its collapse later, at $z_{\rm coll} \approx 1.0$, and had a virial temperature $T_{\rm gas} \approx 9 \times 10^6\,{\rm K}$.

This dependence of virial temperature on halo mass is the key to understanding why a low-mass halo can form a luminous, dense, star-filled galaxy and a high-mass halo cannot. To form a dense galaxy at the center of the dark halo, the baryonic gas must be able to cool by emitting light that escapes into intergalactic space. As it cools, the gas is no longer supported by pressure in a state of hydrostatic equilibrium, and falls to the halo's center. The scarcity of galaxies with mass $M_{\rm tot} > 10^{12}\,{\rm M}_\odot$ results from the fact that the hotter baryonic gas in higher mass halos is less efficient at radiating away its thermal energy.

Consider a $10^{10}\,{\rm M}_\odot$ halo, with a virial temperature $T_{\rm gas} \approx 70\,000\,{\rm K}$. At this temperature, although the hydrogen is ionized by collisions with other gas particles, the helium atoms are still able to retain one of their electrons. A He^+ ion is able to radiate energy efficiently by line emission, as the remaining bound electron is excited to higher levels by collisions, then emits light as it falls back to the ground state. In general, halos with $T_{\rm gas} < 10^6\,{\rm K}$ don't have their hydrogen and helium completely ionized, and can cool quickly by line emission from He^+ or, at lower temperatures, from neutral He and H.

By contrast, halos with $T_{\rm gas} > 10^6\,{\rm K}$ have hydrogen and helium that is almost completely ionized. In these halos, the ionized gas cools primarily by *bremsstrahlung*, also called free-free emission. Bremsstrahlung radiation is produced when a free electron is accelerated as it passes near a free proton or positively charged ion.[8] For a fully ionized primordial mix of hydrogen and helium, the luminosity density for bremsstrahlung emission is

$$\Psi = 5.3 \times 10^{-32}\,{\rm watts\,m^{-3}} \left(\frac{\rho_{\rm gas}}{10^{-24}\,{\rm kg\,m^{-3}}}\right)^2 \left(\frac{T}{10^6\,{\rm K}}\right)^{1/2}. \tag{12.34}$$

Since the low-density gas is highly transparent, all this luminosity is able to escape from the halo. The energy that is being radiated away is the thermal energy of the gas, which has energy density

[8] "Bremsstrahlung," in German, literally means "braking radiation." As the electron moves past the proton or ion, it emits a photon and loses kinetic energy. The alternative name of "free-free radiation" refers to the fact that the electron starts out free and ends free, without being captured by the proton or ion.

$$\varepsilon = \frac{3}{2}nkT = 2.1 \times 10^{-14}\,\mathrm{J\,m^{-3}} \left(\frac{\rho_{\rm gas}}{10^{-24}\,\mathrm{kg\,m^{-3}}}\right)\left(\frac{T}{10^6\,\mathrm{K}}\right). \qquad (12.35)$$

The time it takes the ionized gas to cool by bremsstrahlung emission is then

$$t_{\rm cool} = \frac{\varepsilon}{\Psi} = 13\,\mathrm{Gyr}\left(\frac{\rho_{\rm gas}}{10^{-24}\,\mathrm{kg\,m^{-3}}}\right)^{-1}\left(\frac{T}{10^6\,\mathrm{K}}\right)^{1/2}. \qquad (12.36)$$

Thus, if gas at $T > 10^6$ K is to cool in times less than the age of the universe, it must have a density $\rho_{\rm gas} > 10^{-24}\,\mathrm{kg\,m^{-3}}$. Is this a plausible density for gas in a virialized halo?

Suppose that the baryonic gas makes up a fraction f of the total mass of the virialized halo; if the baryon fraction in the halo is the same as that of the universe as a whole, we expect $f = 0.048/0.31 = 0.15$. The average mass density of the baryonic gas is then, making use of Equation 12.25,

(12.25) $\Rightarrow$

$$\begin{aligned}\bar\rho_{\rm bary} &= f\bar\rho_{\rm halo} \approx 16 f \rho_{m,0}(1+z_{\rm coll})^3 \\ &\approx 0.8 \times 10^{-24}\,\mathrm{kg\,m^{-3}}\left(\frac{f}{0.15}\right)\left(\frac{1+z_{\rm coll}}{5}\right)^3.\end{aligned} \qquad (12.37)$$

We conclude that a virialized halo with $M_{\rm tot} \sim 10^{12}\,\mathrm{M}_\odot$ that starts its collapse at $z_{\rm coll} > 4$ will be hot enough to cool by bremsstrahlung (from Equation 12.33), and will be dense enough (from Equation 12.37) to cool in a time shorter than the age of the universe. However, we can also show that it is statistically unlikely for a halo with mass much larger than $10^{12}\,\mathrm{M}_\odot$ to collapse at a high enough redshift to be able to cool.

To begin our statistical analysis, consider the very first $10^{14}\,\mathrm{M}_\odot$ halo to have collapsed in the entire directly observable universe. (If the baryons in this halo could cool to form a single luminous galaxy, it would have ten times the baryonic mass of the huge cannibal galaxy NGC 4889.) Since the last scattering surface lies at a proper distance $d_p(t_0) \approx 14\,000$ Mpc, the total amount of mass inside the last scattering surface (and thus visible to our telescopes) is

$$M = \rho_{m,0}\frac{4\pi}{3}d_p(t_0)^3 \approx 4.3 \times 10^{23}\,\mathrm{M}_\odot. \qquad (12.38)$$

We can divide this mass into 4.3×10^9 different regions, each of mass $10^{14}\,\mathrm{M}_\odot$. The very first $10^{14}\,\mathrm{M}_\odot$ halo to collapse is the one region out of 4.3 billion that had the highest overdensity at the time of radiation–matter equality. You can think of it as having won a "density lottery" with a probability $P = 1/4.3 \times 10^9 \approx 2.3 \times 10^{-10}$ of drawing the winning ticket. In a Gaussian distribution, this probability is equivalent to a 6.2σ deviation. From the cold dark matter $\delta M/M$ distribution (Figure 11.5), we can compute that a region of mass $M = 10^{14}\,\mathrm{M}_\odot$ with a 6.2σ overdensity begins its collapse at a redshift $z_{\rm coll} \approx 5.2$. After virialization, its gas has a virial temperature $T_{\rm gas} \approx 2.3 \times 10^7$ K (from Equation 12.33) and an average density $\bar{\rho}_{gas} \approx 1.5 \times 10^{-24}\,\mathrm{kg\,m^{-3}}$ (from Equation 12.37). The cooling time for the gas, from Equation 12.36, is therefore $t_{cool} \approx 42\,\mathrm{Gyr}$, longer than the present age of the universe. Moreover, the $10^{14}\,\mathrm{M_\odot}$ halos that collapse later will have slightly lower virial temperatures, but much lower densities. Therefore, the later-collapsing halos have even longer cooling times than the pioneering $10^{14}\,\mathrm{M_\odot}$ halo.

To adapt the "spherical cow" joke common among physicists, I've been using a "spherical, isothermal, virialized cow" model for the formation of galaxies. The true picture, unsurprisingly, is more complex. In particular, computer simulations of galaxy formation indicate that in a more realistic non-spherical collapse, not all the baryonic gas is shock-heated to the virial temperature T_{gas}. The relatively cold gas that escapes being heated is able to flow to the center of the halo on time scales shorter than the cooling time of Equation 12.36. These "cold flows" of gas, as they are called, permit the formation of the first galaxies at higher redshifts, signaling the end of the Dark Ages and the beginning of reionization. The portion of the baryonic gas that *is* shock-heated, however, must obey the cooling time argument. Thus, the spherical, isothermal, virialized model explains why the hot intracluster gas in a massive halo (like that of the Coma cluster) fails to form a single gargantuan galaxy with baryonic mass $M_{bary} \gg 10^{11}\,\mathrm{M_\odot}$.

12.5 Making Stars

Suppose that a collapsed, virialized halo contains gas that is below the virial temperature T_{gas} of Equation 12.30. This can be either because the gas is part of a cold flow that was never shock-heated, or because it cooled rapidly by line emission or bremsstrahlung. The gas is out of hydrostatic equilibrium, and falls toward the center of the dark halo. What happens to the cool, infalling gas then? We know it isn't all swallowed by the central supermassive black hole. In the local universe, the mass of a galaxy's central black hole is less than 1% of the total baryonic mass of the galaxy. (Even the extraordinarily massive black hole in the galaxy NGC 4889, with $M_{bh} \sim 2 \times 10^{10}\,\mathrm{M_\odot}$, is small compared to the total baryonic mass of that bloated cannibal galaxy, $M_{bary} \sim 2 \times 10^{12}\,\mathrm{M_\odot}$.) If $10^5\,\mathrm{M_\odot}$ of gas cools and falls inward, we know it doesn't form into a single "megastar." We also know that it doesn't form a trillion "microstars," each of mass $\sim 10^{-7}\,\mathrm{M_\odot}$. Instead, it forms about a million stars and brown dwarfs, with a typical mass $M_\star \sim 0.1\,\mathrm{M_\odot}$ and a power-law tail to higher masses, as illustrated in the initial mass function of Figure 7.1.

In our own galaxy, stars are observed to form in the dense central cores of *molecular clouds*, regions of interstellar gas that are relatively cold and dense; in these regions hydrogen takes the form of molecules (H_2) rather than individual atoms. In the dense cores of molecular clouds, the mass density is as high as

$\rho_{\rm core} \sim 10^{-15}\,{\rm kg\,m^{-3}}$. This is more than 10^{12} times $\rho_{\rm bary,0}$, the mean density of baryons today. However, it is still less than 10^{-18} times $\rho_\odot$, the mean density of the Sun. In the interstellar gas of our galaxy, the helium mass fraction has been raised, by pollution from early generations of stars, from its primordial value $Y_p = 0.24$ to a current value $Y = 0.27$.[9] The helium mass fraction is $X = 0.72$, leaving a mass fraction $Z = 0.01$ in other elements, primarily oxygen and carbon. Part of the carbon and oxygen is in the form of molecules and radicals such as CO, CH, and OH; however, some condenses into tiny dust grains made of silicates (minerals containing oxygen and silicon) or graphite (pure carbon). A molecular cloud core is dusty enough to be opaque at visible wavelengths; cores seen against a background of stars are called "dark nebulae."[10] A typical temperature for a molecular cloud core is $T_{\rm core} \approx 20\,{\rm K}$. This temperature results from a balance between heating by cosmic rays (high-energy charged particles that can penetrate the opaque core) and cooling by far infrared radiation from dust grains.

The dynamical time in a molecular cloud core is (Equation 11.13)

$$t_{\rm dyn} = \frac{1}{(4\pi G\rho_{\rm core})^{1/2}} \approx 1.1 \times 10^{12}\,{\rm s} \left(\frac{\rho_{\rm core}}{10^{-15}\,{\rm kg\,m^{-3}}}\right)^{-1/2} . \qquad (12.39)$$

The mean molecular mass in a molecular cloud, given its mixture of H_2 and He, is $\mu = 2.3m_p$. This results in an isothermal sound speed

$$c_s = \left(\frac{kT_{\rm core}}{\mu}\right)^{1/2} \approx 270\,{\rm m\,s^{-1}} \left(\frac{T_{\rm core}}{20\,{\rm K}}\right)^{1/2} , \qquad (12.40)$$

only slightly slower than the sound speed in air at room temperature. The Jeans length in the molecular cloud core is then (Equation 11.20)

$$\lambda_J = 2\pi c_s t_{\rm dyn} \approx 1.9 \times 10^{15}\,{\rm m} \left(\frac{\rho_{\rm core}}{10^{-15}\,{\rm kg\,m^{-3}}}\right)^{-1/2} \left(\frac{T_{\rm core}}{20\,{\rm K}}\right)^{1/2} . \qquad (12.41)$$

This means that the baryonic Jeans mass within a dense molecular cloud core is

$$\begin{aligned} M_J = \frac{4\pi}{3}\rho_{\rm core}\lambda_J^3 &\approx 3 \times 10^{31}\,{\rm kg} \left(\frac{\rho_{\rm core}}{10^{-15}\,{\rm kg\,m^{-3}}}\right)^{-1/2} \left(\frac{T_{\rm core}}{20\,{\rm K}}\right)^{3/2} \\ &\approx 15\,{\rm M_\odot} \left(\frac{\rho_{\rm core}}{10^{-15}\,{\rm kg\,m^{-3}}}\right)^{-1/2} \left(\frac{T_{\rm core}}{20\,{\rm K}}\right)^{3/2} . \end{aligned} \qquad (12.42)$$

Since objects smaller than M_J are pressure-supported, this seems to indicate that regions of a molecular cloud that are less massive than $\sim 15\,{\rm M_\odot}$ cannot collapse to form stars. A look at the initial mass function of stars, plotted in Figure 7.1,

[9] O, the futility of stars! Their mass provides $< 10\%$ of the baryonic density today, and even after 13 billion years on the job, the helium they produce has increased the helium mass fraction Y by only $\sim 10\%$, just as the starlight they emit has increased the photon energy density ε_γ by only $\sim 10\%$.

[10] A dark nebula isn't called "dark" because it fails to interact with light (the way dark matter is dark). It's called "dark" because it absorbs light (the way dark chocolate is dark).

reveals that this conclusion is nonsense. Stars with $M_\star > 15\,\mathrm{M}_\odot$ are O stars, the rarest type of star; the preferred mass for stars is actually $M_\star \sim 0.1\,\mathrm{M}_\odot$.

The reason why molecular clouds can make stars smaller than the Jeans mass is that collapsing cores can cool. Consider a molecular cloud core of mass $M_{\rm core} \approx 15\,\mathrm{M}_\odot$, just above the Jeans mass. It contains a total of $N = M_{\rm core}/\mu = 7.8 \times 10^{57}$ gas particles and has a thermal energy

$$E_{\rm core} = N\left(\frac{3}{2}kT_{\rm core}\right) = 3.2 \times 10^{36}\,\mathrm{J}\left(\frac{M_{\rm core}}{15\,\mathrm{M}_\odot}\right)\left(\frac{T_{\rm core}}{20\,\mathrm{K}}\right). \tag{12.43}$$

The core starts its collapse. If energy didn't flow out of (or into) the core, the first law of thermodynamics tells us that the thermal energy would increase at the rate

$$\frac{dE_{\rm core}}{dt} = -P_{\rm core}\frac{dV_{\rm core}}{dt} = -\frac{NkT_{\rm core}}{V_{\rm core}}\frac{dV_{\rm core}}{dt}. \tag{12.44}$$

Since the core's volume is $V_{\rm core} \propto R_{\rm core}^3$, this implies

$$\frac{dE_{\rm core}}{dt} = -3NkT_{\rm core}\left(\frac{1}{R_{\rm core}}\frac{dR_{\rm core}}{dt}\right) = -2E_{\rm core}\left(\frac{1}{R_{\rm core}}\frac{dR_{\rm core}}{dt}\right). \tag{12.45}$$

If the core is to remain at a constant temperature $T_{\rm core}$ during its collapse, it must radiate away the increased thermal energy at the same rate it is generated. The luminosity required to keep the core at a constant temperature (let's call it the "isothermal luminosity") is

$$L_{\rm core}^{\rm iso} = -\frac{dE_{\rm core}}{dt} = \frac{2E_{\rm core}}{R_{\rm core}}\frac{dR_{\rm core}}{dt}. \tag{12.46}$$

For a freely collapsing core, $dR_{\rm core}/dt \approx R_{\rm core}/t_{\rm dyn}$, and the isothermal luminosity at $T_{\rm core} = 20\,\mathrm{K}$ is

$$L_{\rm core}^{\rm iso} \approx \frac{2E_{\rm core}}{t_{\rm dyn}} \approx 0.015\,\mathrm{L}_\odot\left(\frac{M_{\rm core}}{15\,\mathrm{M}_\odot}\right)\left(\frac{\rho_{\rm core}}{10^{-15}\,\mathrm{kg\,m^{-3}}}\right)^{1/2}. \tag{12.47}$$

Thus, as long as the core radiates away energy at a minimum rate $L^{\rm iso} \sim 0.015\,\mathrm{L}_\odot$, it can maintain a constant temperature of $T_{\rm core} = 20\,\mathrm{K}$ as it collapses on a dynamical time. But what will be the actual luminosity L of a molecular cloud core?

The main source of emission from a dusty molecular cloud core is the far infrared light from its dust grains. Looking at a dusty dark nebula, as in the left panel of Figure 12.6, our first guess might be that since it's an opaque object with a well-defined temperature $T_{\rm core}$, its luminosity is that of a blackbody. For a spherical core, this would be $L_{\rm bb} = 4\pi R_{\rm core}^2 \sigma_{\rm sb} T_{\rm core}^4$, where $\sigma_{\rm sb} = 5.67 \times 10^{-8}\,\mathrm{watts\,m^{-2}\,K^{-4}}$ is the Stefan–Boltzmann constant for blackbody radiation. However, a molecular cloud core that is opaque at visible wavelengths is not necessarily opaque at infrared wavelengths. Interstellar dust grains are tiny, with a radius of 100 nanometers or less; for such small grains, the cross-section for absorbing visible and infrared light is a decreasing function of the wavelength λ.

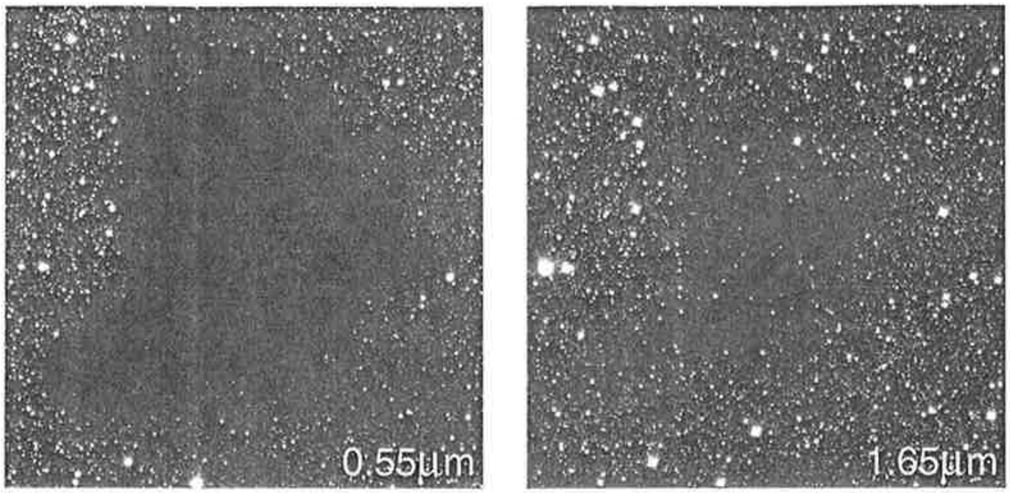


Figure 12.6 The dark nebula Barnard 68, which is at a distance $d \approx 125\,\mathrm{pc}$ from the Earth and has a radius $R_{\mathrm{core}} \approx 15\,000\,\mathrm{AU} \approx 3 \times 10^6\,\mathrm{R}_\odot$. Left: V-band image, at a wavelength $\lambda_{\mathrm{vis}} \approx 550\,\mathrm{nm}$. Right: Near infrared image, at a wavelength $\lambda_{\mathrm{nir}} \approx 1650\,\mathrm{nm}$. [European Southern Observatory]

Figure 12.6 shows a nearby dark nebula that is highly opaque at visible wavelengths, with $\tau \approx 23$ through its center at $\lambda_{\mathrm{vis}} = 550\,\mathrm{nm}$. However, at near infrared wavelengths it is much less opaque: $\tau \approx 4$ at $\lambda_{\mathrm{nir}} = 1650\,\mathrm{nm}$, so bright background stars can be glimpsed through the dusty core at this wavelength. At far infrared wavelengths, the core is largely transparent, with $\tau \approx 0.03$ at $\lambda_{\mathrm{fir}} \approx 10^5\,\mathrm{nm} \approx 0.1\,\mathrm{mm}$. The optical depth of the core will increase as it is compressed to smaller radii; if n_{dust} is the average number density of dust grains in the core, then $n_{\mathrm{dust}} \propto R_{\mathrm{core}}^{-3}$ and therefore $\tau \propto n_{\mathrm{dust}} R_{\mathrm{core}} \propto R_{\mathrm{core}}^{-2}$. However, different cores have different dust properties, so we can't say in the general case at what degree of compression a core will become opaque at far infrared wavelengths.

Mathematically, we can take into account the possible non-opacity of the core by writing its luminosity as

$$L_{\mathrm{core}} = 4\pi R_{\mathrm{core}}^2 \cdot f_e \sigma_{\mathrm{sb}} T_{\mathrm{core}}^4, \tag{12.48}$$

where the *efficiency factor* is $f_e \leq 1$. If the core is highly opaque at the wavelength of emission, with $\tau \gg 1$, then we expect $f_e \approx 1$. However, if it is largely transparent, with $\tau < 1$, then we expect $f_e < 1$. Scaled to the properties of our standard molecular cloud core, the core's radius is

$$R_{\mathrm{core}} = \left(\frac{3M_{\mathrm{core}}}{4\pi\rho_{\mathrm{core}}}\right)^{1/3} \approx 10^4\,\mathrm{AU}\left(\frac{M_{\mathrm{core}}}{15\,\mathrm{M}_\odot}\right)^{1/3}\left(\frac{\rho_{\mathrm{core}}}{10^{-15}\,\mathrm{kg\,m^{-3}}}\right)^{-1/3}, \tag{12.49}$$

and the luminosity actually emitted from the core at a temperature $T_{\mathrm{core}} = 20\,\mathrm{K}$ is

$$L_{\mathrm{core}} \approx 1100\,\mathrm{L}_\odot f_e \left(\frac{M_{\mathrm{core}}}{15\,\mathrm{M}_\odot}\right)^{2/3}\left(\frac{\rho_{\mathrm{core}}}{10^{-15}\,\mathrm{kg\,m^{-3}}}\right)^{-2/3}. \tag{12.50}$$

Any efficiency factor greater than the extremely modest value $f_e \sim 10^{-5}$ will enable the actual luminosity (Equation 12.50) to be larger than the isothermal luminosity required to maintain a constant temperature (Equation 12.47).

Let's see what happens as our $15\,\mathrm{M}_\odot$ core collapses at a constant temperature $T_{\rm core} = 20\,\mathrm{K}$. Since the Jeans mass, from Equation 12.42, has the dependence $M_J \propto \rho^{-1/2}T^{3/2}$, then if the temperature is constant, the Jeans mass *decreases* as the core becomes smaller and its density thus becomes larger. Consider what happens when the radius of the core decreases from $R_{\rm core}$ to $R_{\rm core}/4^{1/3} \approx 0.63R_{\rm core}$, and the density thus increases from $\rho_{\rm core}$ to $4\rho_{\rm core}$. The Jeans mass has now fallen by a factor of $1/2$, and the $15\,\mathrm{M}_\odot$ core is now unstable, splitting to form a pair of $7.5\,\mathrm{M}_\odot$ fragments. These fragments continue to collapse until their density increases by another factor of 4, then split into a total of four $3.75\,\mathrm{M}_\odot$ fragments; and so on, and so forth.

Hierarchical fragmentation, as this process of repeated subdivision is called, naturally produces a power-law distribution of stellar masses, as long as the fragmentation process is slightly inefficient. Suppose, for example, that fragmentation has a failure rate of 1%. That is, if you start with 100 cores, each of mass $M_{\rm core}$, 99 of them will split, forming 198 fragments each of mass $M_{\rm core}/2$, but one will collapse directly to form a star of mass $M_\star = M_{\rm core}$. If the failure rate is the same at the next step, then of the 198 fragments of mass $M_{\rm core}/2$, we expect that 196 will split, forming 392 subfragments, each of mass $M_{\rm core}/4$, but two will collapse directly to form stars of mass $M_\star = M_{\rm core}/2$. As this process proceeds, you will expect one star of mass $M_{\rm core}$, two of mass $M_{\rm core}/2$, four of mass $M_{\rm core}/4$, and so forth. In general, after the nth round of fragmentation, there will be 2^n stars that each have mass $2^{-n}M_{\rm core}$. Thus, the number of stars produced per logarithmic interval of mass will be

$$\frac{dN}{d\log M} \propto \frac{1}{M}, \tag{12.51}$$

which is equivalent to an initial mass function $\chi(M) \propto M^{-2}$. (The actual initial mass function has the steeper dependence $\chi \propto M^{-2.3}$. This is a sign that the simple hierarchical fragmentation model doesn't capture all the physics involved.)

Starting with a core mass $M_{\rm core}$, after n rounds of fragmentation, the mass of each fragment will be $M_{\rm frag} = 2^{-n}M_{\rm core}$ and the density of each fragment will be $\rho_{\rm frag} = 2^{2n}\rho_{\rm core}$. What halts the fragmentation process after a finite number of rounds, and prevents it from proceeding *ad infinitum*, producing an infinite number of infinitesimal fragments? Notice that the isothermal luminosity required to keep the fragments at a constant temperature $T_{\rm core} = 20\,\mathrm{K}$ has the value (Equation 12.47)

$$L_{\rm frag}^{\rm iso} \approx 0.015\,\mathrm{L}_\odot \left(\frac{M_{\rm frag}}{15\,\mathrm{M}_\odot}\right)\left(\frac{\rho_{\rm frag}}{10^{-15}\ \mathrm{kg\,m^{-3}}}\right)^{1/2}. \tag{12.52}$$

However, since $M_{\text{frag}} \propto 2^{-n}$ and $\rho_{\text{frag}}^{1/2} \propto 2^{n}$, the isothermal luminosity is the same for every value of n. That is, every fragment, regardless of size, must radiate at the same luminosity to maintain a temperature of 20 K. However, since the actual luminosity is proportional to the square of the fragment's radius, smaller fragments will have lower luminosity. After some number of rounds of fragmentation, the small fragments will find it impossible to radiate away enough energy to keep cool.

The ratio of the actual luminosity of a fragment (Equation 12.50) to its required isothermal luminosity (Equation 12.47) is

$$\frac{L_{\text{frag}}}{L_{\text{frag}}^{\text{iso}}} \approx 73\,000 f_e \left(\frac{M_{\text{frag}}}{15\text{M}_\odot}\right)^{-1/3} \left(\frac{\rho_{\text{frag}}}{10^{-15}\,\text{kg}\,\text{m}^{-3}}\right)^{-7/6}. \tag{12.53}$$

After n rounds of fragmentation, the characteristic fragment mass will be $M_{\text{frag}} = 2^{-n} M_{\text{core}} \approx 2^{-n}(15\,\text{M}_\odot)$, and the characteristic fragment density will be $\rho_{\text{frag}} = 2^{2n}\rho_{\text{core}} \approx 2^{2n}(10^{-15}\,\text{kg}\,\text{m}^{-3})$. The four-fold increase of density with each round of fragmentation means that the ratio of the actual luminosity to the required isothermal luminosity will *decrease* with each round of fragmentation:

$$\frac{L_{\text{frag}}}{L_{\text{frag}}^{\text{iso}}} \approx 73\,000 f_e 2^{n/3} 2^{-7n/3} \approx 73\,000 f_e 2^{-2n}. \tag{12.54}$$

Fragmentation stops when this ratio drops to one, and the temperature of the collapsing fragments starts to rise; this causes the Jeans mass to increase instead of decrease as the fragments become denser. From Equation 12.54, we find that cooling the fragments that result from $n = 9$ rounds of fragmentation would require an unphysically high efficiency $f_e \approx 3$. Thus, the maximum possible number of fragmentation rounds is $n = 9$, producing a minimum fragment mass $M_{\text{frag}} \sim 2^{-9} M_{\text{core}} \sim 0.03\,\text{M}_\odot$, within the broad maximum of the initial mass function (Figure 7.1). Once fragments can no longer keep cool, they become *protostars*, objects that are almost, but not quite, in hydrostatic equilibrium. Protostars gradually contract inward as they radiate away energy, but only on time scales much longer than the dynamical time. A protostar with $M > 0.08\,\text{M}_\odot$ becomes a star when the density and temperature at its center become large enough for nuclear fusion to begin.[11]

Hierarchical fragmentation requires that collapsing molecular cloud cores are able to cool efficiently through multiple rounds of fragmentation. If molecular cloud cores are very low in dust (which happens if they are deficient in heavy elements such as carbon, oxygen, and silicon), this drives down f_e, and halts fragmentation at an earlier stage. Thus, star formation in regions with low abundances of heavy elements has an initial mass function that peaks at a higher mass. When the very first stars formed at the end of the Dark Ages, there was no carbon,

[11] A protostar with $M < 0.08\,\text{M}_\odot$ should really be called a "proto-brown-dwarf."

oxygen, and silicon; thus, there were no dust grains at all. The first stars could cool only through much less efficient processes involving molecular hydrogen. Thus, it is likely that the very first stars that formed in the universe were not the result of hierarchical fragmentation, but instead were all extremely massive stars, comparable in size to the Jeans mass of the gas from which they formed. These ultramassive stars would have lived fast and died young, ejecting the carbon, oxygen, and other heavy elements that they formed into the surrounding gas.

The hierarchical fragmentation scenario obviously represents a highly simplified "spherical isothermal cow that reproduces by fission." Realistic models of star formation must take into account, among other physical processes, the effects of magnetic fields and of turbulence within molecular clouds. In addition, the effects of angular momentum must be acknowledged. An initially slowly spinning fragment, thanks to conservation of angular momentum, will rotate more and more rapidly as it contracts adiabatically. The net result will be a protostar, made from the material with lower angular momentum, surrounded by a rotationally supported protoplanetary disk. Protoplanetary disks, it is found observationally, are unstable to the formation of the objects that we call "planets." Planets are a common, but minor, side effect of star formation. In the Solar System, for instance, the total mass of all the planets is $M_{\rm pl} = 2.67 \times 10^{27}$ kg, with more than two-thirds of this mass provided by Jupiter. The total planet mass thus represents a small fraction of the Sun's mass: $M_{\rm pl} = 0.0013\,{\rm M}_\odot$. If the Sun is not unusual in having a planetary system with a mass $\sim 0.13\%$ of its own mass, then we can estimate the density parameter of planets:

$$\Omega_{\rm pl,0} \sim 0.0013\Omega_{\star,0} \sim (0.0013)(0.003) \sim 4 \times 10^{-6}. \tag{12.55}$$

From one point of view, planets are unimportant, because they provide just a few parts per million of the mass-energy density of the universe. From another point of view, however, planets are vitally important; planets of the right size, at the right distance from their parent star, provide a hospitable environment for the evolution of beings who ask the questions "Where do we come from?", "What are we?", and "Where are we going?"

Exercises

12.1 For the Schechter luminosity function of galaxies (Equation 12.21), find the number density of galaxies more luminous than L, as a function of L^*, Φ^*, and α. In the limit $L \to 0$, show why $\alpha = -1$ leads to problems, mathematically speaking. What is a plausible physical solution to this mathematical problem? [Hint: an acquaintance with incomplete gamma functions will be useful.]

12.2 For the Schechter luminosity function of galaxies, find the total luminosity density Ψ as a function of L^*, Φ^*, and α. What is the numerical value of

Epilogue

A book dealing with an active field like cosmology can't really have a neat, tidy ending. Our understanding of the universe is still growing and evolving. During the twentieth century, the growing weight of evidence pointed toward the Hot Big Bang model, in which the universe started in a hot, dense state, but gradually cooled as it expanded. At the end of the twentieth century and the beginning of the twenty-first, cosmological evidence was gathered at an increasing rate, refining our knowledge of the universe. As I write this epilogue, on a clear spring day in the year 2016, the available evidence is explained by a Benchmark Model that is spatially flat and that has an expansion, which is currently accelerating. It seems that 69% of the energy density of the universe is contributed by a cosmological constant (or other form of "dark energy" with negative pressure). Only 31% of the energy density is contributed by matter (and only 4.8% is contributed by the familiar baryonic matter of which you and I are made).

However, many questions about the cosmos remain unanswered. Here are a few of the questions that currently nag at cosmologists:

- *What are the precise values of cosmological parameters such as H_0, $\Omega_{m,0}$, and $\Omega_{\Lambda,0}$?* Much effort has been invested in determining these parameters, but still they are not pinned down absolutely.
- *What is the dark matter?* It can't be made entirely of baryons. It can't be made entirely of neutrinos. Most of the dark matter must be in the form of some exotic stuff that has not yet been detected in laboratories.
- *What is the dark energy?* Is it vacuum energy that plays the role of a cosmological constant, or is it some other component of the universe with $-1 < w < -1/3$? If it is vacuum energy, is it provided by a false vacuum, driving a temporary inflationary stage, or are we finally seeing the true vacuum energy?
- *What drove inflation during the early universe?* Our knowledge of the particle physics behind inflation is still sadly incomplete. Indeed, some

cosmologists pose the questions, "Did inflation take place at all during the early universe? Is there another way to resolve the flatness, horizon, and monopole problems?"

- *Why is the universe expanding?* At one level, this question is easily answered. The universe is expanding today because it was expanding yesterday. It was expanding yesterday because it was expanding the day before yesterday... However, when we extrapolate back to the Planck time, we find that the universe was expanding then with a Hubble parameter $H \sim 1/t_P$. What determined this set of initial conditions? In other words, "What put the Bang in the Big Bang?"

The most interesting questions, however, are those that we are still too ignorant to pose correctly. For instance, in ancient Egypt, a list of unanswered questions in cosmology might have included "How high is the dome that makes up the sky?" and "What's the dome made of?" Severely erroneous models of the universe obviously give rise to irrelevant questions. The exciting, unsettling possibility exists that future observations will render the now-promising Benchmark Model obsolete. I hope, patient reader, that learning about cosmology from this book has encouraged you to become a cosmologist yourself, and to join the scientists who are laboring to make my book a quaint, out-of-date relic from a time when the universe was poorly understood.