Chapter 15
Origin, Evolution, and Destiny of the Observable Universe

15.1 Hotter Than Anything Else

Regardless of the direction in which we look out into space, almost all of the distant galaxies are flying apart, dispersing and moving away at speeds that increase with their distance, as if they had been ejected by a cosmic bomb. Astronomers call it the "big bang."

We can envision this early state by putting the observed expansion of the galaxies in reverse and pushing the galaxies back closer together until a time about 13.7 billion years ago, when the universe was incredibly small and all of its mass was compressed to a very high density. It marks the beginning of the observable universe, but no one knows what happened before the big bang propelled the expanding universe into existence (Focus 15.1).

Focus 15.1 Before the Big Bang

How did it all begin? The existing theory falls apart at the first crucial instant, at the beginning of the big bang, and can't be extended to anything that occurred before that. Moreover, there is no observational evidence for prior events.

Einstein's *General Theory of Relativity* cannot be used back then because the equations contain a singularity at the beginning of the big bang, when a non-zero parameter is divided by zero and conditions cannot be defined. So we've just pushed the mystery of the ultimate origin of the universe back about 13.7 billion years, to a point that science cannot penetrate.

An inflation theory does describe what could have happened in the first fraction of a second of the big bang, when the universe was just 10^{-35} s old. Guth (1981) and Linde (1982) describe inflation; Narlikar and Padmanabhan (1991) have described inflation for astronomers.

During inflation the universe was driven by a repulsive gravity, unlike the attracting kind we are used to, and operated on a very small scale in both

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_15, © Springer-Verlag Berlin Heidelberg 2013

space and time, blowing the universe up, enlarging it by an enormous factor. Owing to its inherent instability, the burst of inflation soon decayed away and came to an end, in a time far less than one second, releasing its remaining energy into material particles and creating the heat of the big bang, the primeval fireball.

This accelerated expansion in the first miniscule moments of the big bang, this inflation, supposedly obliterated evidence of previous space, time, energy and matter, erasing previous history. That cosmic forgetfulness closes the door to the very beginning, conveniently avoiding the question of ultimate origins, the original genesis, and removing it from any observational consequences.

In other words, according to this theory the big bang or its immediate consequences destroyed all evidence of what came before. Or the big bang might have initiated time, on a day without a yesterday. So there is no before. Or perhaps the explanation lies outside space and time.

In any event, the existing equations and theories fail to explain how the observed universe began. So we still don't know how the universe came into being, and it remains a captivating mystery.

Because gases become hotter when they are compressed and cool when they expand, the observed universe must have been incredibly hot in its earliest, most compact state. As we look back in time, at the most distant regions, the universe becomes increasingly hot, eventually becoming so exceptionally hot that radiation was the most powerful force, dominating the expansion of the universe.

In the earliest moments of the big bang, there were no stars or galaxies, only intense radiation and subatomic particles from which the material universe subsequently grew. During this hot beginning, matter was then being created by radiation and vice versa. Some of the incredibly energetic radiation was being transformed into electrons and their anti-matter counterparts, the positrons or positive electrons, and just as often an electron would collide with a positron to make radiation again.

If we let γ denote a photon of the energetic radiation, or a gamma ray, then the equilibrium condition can be written in short hand notation as:

$$\gamma + \gamma \rightleftarrows e^- + e^+, \tag{15.1}$$

where e^- denotes an electron and e^+ denotes a positron, the anti-matter particle of the electron. The double arrow means that the reaction goes in both directions at the same rate. The forward process, from left to right, is known as *electron–positron pair creation*, and the reverse one is called *pair annihilation*.

Neutrons and protons were also around, and these subatomic particles would also turn back and forth into each other, through reactions that included electron neutrinos and electron antineutrinos, as well as electrons and positrons. The equilibrium between radiation and these subatomic particles continued as long as it was hot enough to create positrons; however, the radiation quickly cooled as the result of the expansion of the universe into a greater volume. Electron–positron pair creation continued only as long as the thermal energy of the radiation, $kT_r(t)$, with radiation temperature $T_r(t)$ at time t, was greater than the rest mass energy of the electron, or for $kT_r(t) \geq m_e c^2$. Here the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, the mass of the electron is $m_e = 9.1094 \times 10^{-31}$ kg, and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. Using these constants, the equilibrium stopped when the radiation cooled to a temperature of $T_r(t) = 5.93 \times 10^9$ K. After that, no more positrons were made and the leftover positrons were then consumed by interactions with electrons. The neutrons, protons, and electrons that remained eventually gathered together to create the material universe we have today.

In the early stages, the radiation controlled the expansion, with an effective energy-mass density, $\rho_r(t)$, at time, t, given by:

$$\rho_r(t) = \frac{aT_r^4(t)}{c^2} \approx 0.842 \times 10^{-32}\, T_r^4(t) \text{ kg m}^{-3}, \tag{15.2}$$

where the radiation constant $a = 7.5657 \times 10^{-16}$ J m^{-3} K^{-4}, and we have divided by the square of the speed of light, $c = 2.9979 \times 10^8$ m s^{-1}, to convert the energy density into an equivalent mass density.

If we denote the effective mass of the radiation as $M_r(t)$, assume the universe is expanding at a constant velocity $V = R(t)/t$ for radius $R(t)$ at time t, and equate the kinetic energy of expansion to the gravitational potential energy of the radiation, then

$$\frac{1}{2} M_r(t) \left[\frac{R(t)}{t}\right]^2 = \frac{GM_r^2(t)}{R(t)}, \tag{15.3}$$

or

$$M_r(t) = \frac{R^3(t)}{2Gt^2}, \tag{15.4}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2} and

$$M_r(t) = \frac{4\pi}{3} R^3(t) \rho_r(t) = \frac{4\pi}{3} R^3(t) \frac{aT_r^4(t)}{c^2}. \tag{15.5}$$

Collecting terms, we obtain an approximate equation for the radiation temperature $T_r(t)$ at time, t, as long as the radiation is controlling the expansion.

$$T_r(t) = \left[\frac{3c^2 M_r(t)}{4\pi a R^3(t)}\right]^{1/4} = \left[\frac{3c^2}{8\pi G a}\right]^{1/4} \frac{1}{t^{1/2}} \approx 2.15 \times 10^{10} \frac{1}{t^{1/2}} \text{ K}, \tag{15.6}$$

where the time t is in seconds for the numerical approximation, or equivalently

$$t \approx 4.6 \times 10^{20} \frac{1}{[T_r(t)]^2} \text{ s}. \tag{15.7}$$

The radiation temperature has, for example, dropped to 5.93×10^9 K in just about 13 s after the big bang.

In the early stages, the radiation controlled the expansion of the universe because it was incredibly hot. Just 1 s after the big bang, the radiation had a temperature of about 20 billion, or 2×10^{10} K. As the universe continued to expand, the radiation steadily cooled, and the matter eventually took over the expansion. However, the big bang was so intense and so hot that we are still immersed within the radiation.

15.2 Three Degrees Above Absolute Zero

15.2.1 An Unexpected Source of Noise

The discovery of the faint afterglow of the big bang was a serendipitous event, involving a horn-reflector antenna that had been used at the Bell Telephone Laboratories in the first tests of a communication satellite. Arno Penzias (1933–) and Robert Wilson (1936–) were measuring the temperatures of noises in the horn-antenna system so they could make accurate measurements of the intensity of several extragalactic radio sources. A persistent, ubiquitous, and unvarying noise source was detected at a signal frequency of 4,080 MHz = 4.08×10^9 Hz, or a wavelength of 7.35 cm, contributing an antenna temperature of only 3 degrees above zero, or about 3 K. It was equally strong in all directions, wherever the antenna was pointed, independent of the time of day and year and with no dependence on the location of any known cosmic radio source.

Penzias and Wilson did not know what they had found and avoided any mention of the cosmological implications in their publication (Penzias and Wilson 1965), which had the modest title, *A Measurement of Excess Antenna Temperature at* 4,090 *MHz*. However, a group at Princeton University, which was attempting to make a similar measurement at the time, drew attention to the implications in a companion paper (Dicke et al. 1965). The unexpected source of noise was the faint, cooled relic of the hot big bang, now known as the *three-degree cosmic microwave background radiation*, because it has a temperature of about 3 K and it originated before the stars and galaxies were formed, lying behind them.

This particular discovery was not entirely unanticipated. In the late 1940s and early 1950s, George Gamow (1904–1968), Ralph A. Alpher (1921–2007), James W. Follin (1919–2007), and Robert C. Herman (1914–1997) had speculated that the 1-billion-degree, or 10^9 K, radiation of the early universe would have cooled to about 5 K during the past billions of years of expansion, but nobody had attempted to observe the relic radiation (Gamow 1948, 1956; Alpher and Herman 1948). Penzias and Wilson were also unaware of this previous calculation until after their discovery. They received the 1976 Nobel Prize in Physics for their discovery of the cosmic microwave background radiation.

15.2.2 Blackbody Spectrum

At the high temperatures during the early history of the expanding universe, the radiation and subatomic particles frequently interacted, achieving thermal equilibrium characterized by single temperature. Later, when the universe thinned out and cooled by expanding into a greater volume, the matter and radiation quit interacting, going their separate ways. However, the radiation would have retained its thermal nature as it cooled and the temperature slowly decreased.

A perfect thermal radiator is known as a *blackbody*, which absorbs all thermal radiation falling on it and reflects none – hence, the term *black*. The distribution of the radiation emitted by the blackbody, its spectrum, peaks at a wavelength that is inversely proportional to the temperature, dropping precipitously at shorter wavelengths and falling off gradually at longer ones (Sect. 2.4).

The expansion of the universe preserves the blackbody spectrum of the radiation for all time. No process can destroy its shape, but the location of maximum intensity will stretch to increasingly longer wavelengths as time goes on and the radiation gets colder. The wavelength of peak intensity, λ_{max}, is inversely proportional to the radiation temperature, T, and the Wien displacement law (Sect. 2.4) specifies that wavelength as $\lambda_{max} = 0.0029/T$ m. In the present epoch, with a temperature of only about 3° above absolute zero, or at 3 K, the blackbody radiation intensity peaks at a wavelength of about 0.001 m or 0.1 cm. Unfortunately, the Earth's atmosphere absorbs cosmic radiation at this short wavelength where the most intense radiation occurs.

The definitive spectral measurements therefore had to be made from above the atmosphere using NASA's *COsmic Background Explorer* (*COBE*), launched on 18 November 1989. Less than two months after *COBE* went into orbit, but a quarter-century after the discovery of the cosmic radiation, John C. Mather (1946–) reported the combined results of millions of *COBE* spectral measurements at an American Astronomical Society meeting near Washington, DC. The spectrum fit the Planck blackbody curve with a precision of 1 part in 10,000 (Fig. 15.1), establishing a temperature of precisely 2.725 K, with an uncertainty of 0.002 K (Mather et al. 1994).

Such a thermal spectrum could not have happened in the universe as it is now. Matter currently has a very different temperature than the background radiation. In other words, the observed spectrum is proof that the observable universe must have expanded from a very hot, dense state in the past, when matter and radiation were in thermal equilibrium and at the same temperature.

Every part of space now is filled with background radiation. From its temperature, of $T = 2.725$ K and we can specify the frequency $\nu_{\max}$ from the Wien displacement law in frequency (Sect. 2.4) as $\nu_{\max} \approx 2.8\, kT/h \approx 5.88 \times 10^{10}\ T \approx 1.60 \times 10^{11}$ Hz, where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1} and the Planck constant

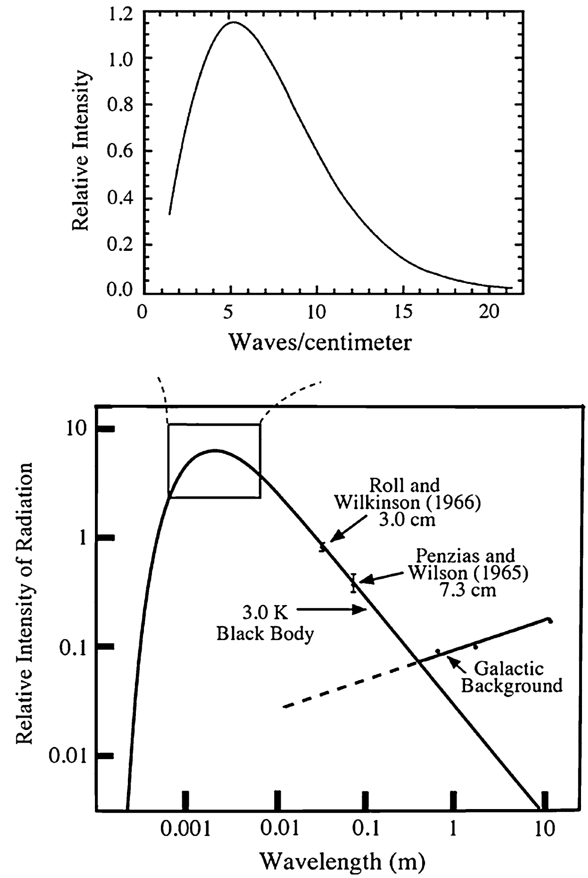

Fig. 15.1 Spectrum of the cosmic microwave background radiation The intensity of the cosmic microwave background radiation plotted as a function of wavelength. This thermal radiation was formed about 390,000 years after the big bang, which occurred about 14 billion years ago. The observed radiation has a nearly perfect blackbody spectrum. Pioneering measurements by Arno A. Penzias (1933–) and Robert W. Wilson (1936–) in 1965 and Peter G. Roll (1935–) and David T. Wilkinson (1935–2002) in 1966, at 7.35 and 3.0 cm wavelength, respectively, are compared to the expected spectrum of a three-degree blackbody and radiation from our Galaxy (*bottom*). The full spectrum at millimeter wavelengths (*top*) was obtained from instruments aboard the *COsmic Background Explorer* (*COBE*) in late 1989. These data are so accurate that the error bars of the individual points all lie within the width of the plot curve. This solid line, which matches the shape and peak location of the observed data, corresponds to a thermal radiator, or blackbody, with a temperature of 2.725 K

$h = 6.6261 \times 10^{-34}$ J s. The number density N_{CMB} of photons in the background radiation can be determined from:

$$N_{CMB} = \frac{aT^4}{h\nu_{max}} \approx 4 \times 10^8 \text{ m}^{-3}, \tag{15.8}$$

where the radiation energy density is aT^4, the radiation constant $a = 7.5656 \times 10^{-16}$ J m^{-3} K^{-4}, and $h\nu_{max}$ is the photon energy. Thus, every cubic meter of space in the observable universe contains about a half billion photons of the cosmic microwave radiation. These are tiny bundles of radiation energy that originated about 14 billion years ago.

15.2.3 As Smooth as Silk

What alerted astronomers to the importance of the background radiation was its equal brightness wherever one looked, indicating that it uniformly fills all of space (Wilson and Penzias 1967). This spatial isotropy satisfied one of the basic tenets of modern cosmology, *the cosmological principle*, which asserts that except for local irregularities, the universe presents the same aspect from every point.

But the radiation seemed too uniform. The *COBE* instruments could detect no regions brighter than others to 0.0003 K, or 1 part in 10,000, on angular scales from minutes of arc to 180°. Yet, the background radiation ought to have concentrations in it, which acted as seeds for the subsequent formation of the material universe. They must have acted as a template or blueprint, encoding the information required to explain the subsequent formation of stars and galaxies.

15.2.4 Cosmic Ripples

In 1992, George Smoot (1945–) and his colleagues announced measurements of the temperature fluctuations in the cosmic microwave background radiation using four years of data gathered by *COBE* (Smoot et al. 1992; Bennett et al. 1993, 1996). Aftersubtracting the known microwave emission of the Milky Way and using mathematical averaging techniques on about 100 million observations, the *COBE* team found that the temperature varies ever so slightly over large angular sizes. The sensitive instrument detected minute temperature differences no larger than a hundred-thousandth, or 10^{-5} K. Mather and Smoot were awarded the 2006 Nobel Prize in Physics for their discovery of the blackbody form and anisotropy of the cosmic microwave background radiation.

COBE was pushed to the limits of its sensitivity, with evidence that wasn't quite definitive, mainly because it mapped the cosmic radiation with coarse angular resolution greater than 7°. In the subsequent decade, more than 20 experiments were therefore carried out from the ground and balloon platforms, bringing the temperature fluctuations into shaper focus with angular resolutions as fine as a few minutes of arc. However, the ground-based and balloon experiments only

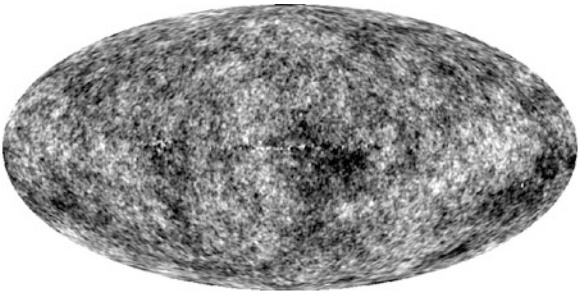

Fig. 15.2 Maps of the infant universe An all-sky view of the three-degree cosmic microwave background radiation emitted from the universe in its infancy, just 390,000 years after the big bang that occurred 13.7 billion years ago. The data, taken in 2003 from the *Wilkinson Microwave Anisotropy Probe* (WMAP) are shown here after seven years of data analysis. The temperature fluctuations range up to 0.0002K above and below the average value. Darker regions are cooler and lighter regions are hotter. These temperature fluctuations provided the seeds from which galaxies subsequently grew. (Courtesy of the NASA/*COBE* and NASA/*WMAP* science teams.)

glimpsed small portions of the sky for a limited time, so there was a possibility that they might not be truly representative of all the background radiation.

It was therefore time for another satellite experiment that would scan the entire sky without the confusion of microwave radiation from the atmosphere and ground. This time, the spacecraft would not only detect the cosmic ripples; it instead would determine their distribution and characteristic sizes, filling in the gaps between the large features seen with *COBE* and the smaller features detected by other instruments.

David T. Wilkinson (1935–2002), of Princeton University, joined Charles L. "Chuck" Bennett (1956–) of the Goddard Space Flight Center to create a small team of experts and design a spacecraft that could accomplish the goal within the modest, for NASA, budget cap of $70 million in 1994 dollars. The resultant *Microwave Anisotropy Probe* (*MAP*) was approved in mid-1996 and launched on 30 June 2001. The name was changed in early 2003 to *Wilkinson Microwave Anisotropy Probe* (*WMAP*), to honor Wilkinson after his death. Instruments aboard *WMAP* provided definitive measurements of the rippling departures from uniformity (Fig. 15.2), with temperature fluctuations of 1 part in 100,000, or at about 0.00003 K (Bennett et al. 2003). This anisotropy, at the level of $\Delta T/T = (1.1 \pm 0.1) \times 10^{-5}$, is given with other physical properties of the background radiation in Table 15.1. Hu and Dodelson (2002) have reviewed cosmic microwave background anisotropies.

When combined with previous measurements, the *WMAP* instruments showed that temperature variations are concentrated within certain angular sizes that are displayed in an angular power spectrum – a plot of the relative strength of the hot and cold spots against their angular sizes (Fig. 15.3). This spectrum is not flat – it is rippled. Gravity explains the ripples, the relative amplitudes of which can be used to infer the gravitational pull that caused them.

Table 15.1 Physical properties of the cosmic microwave background radiation

Parameter	Name	Value
$T_0 = T_{CMB}$	Temperature	2.725 ± 0.002 K
N_{CMB}	Photon density	$(410.4 \pm 0.9) \times 10^6$ m^{-3}
ρ_{CMB}	Mass-energy density of photons	4.648×10^{-31} kg m^{-3}
T_1	Dipole anisotropy	0.003346 ± 0.000017 K
T_1/T_0	Dipole anisotropy/temperature	0.001228
$\Delta T/T_0$	Anisotropy, dipole removed	$(1.1 \pm 0.1) \times 10^{-5}$
$T_2 = Q_{rms}$	Quadrupole moment	$(8 \pm 2) \times 10^{-6}$ K

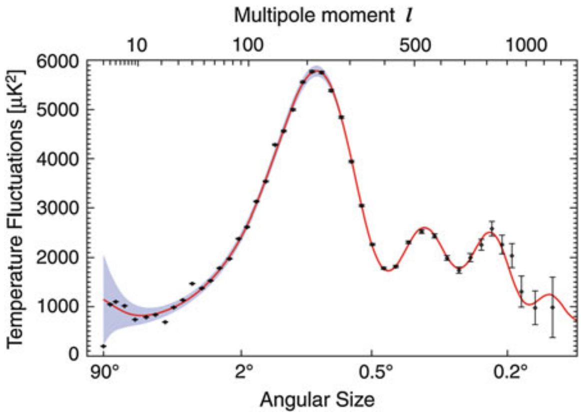

Fig. 15.3 **Ripple data** The angular fluctuation strength, or power, of the cosmic microwave background radiation in which temperature fluctuations, in units of square micro kelvin (10^{-6} K and designated μK), are displayed as a function of their angular extent in degrees, denoted 0. This plot shows the relative brightness for the all-sky map observed from the *Wilkinson Microwave Anisotropy Probe* (*WMAP*) (see Fig. 15.2) at various sizes. The solid line is the model that best fits the observed data (*solid dots*); the gray band represents uncertainties in the model. Anisotropy data obtained by previous experiments are denoted by dots with error bars. The observed power spectrum has been compared to other astronomical observations and different theoretical models, providing estimates for the amount of dark matter and dark energy in the universe (see Fig. 15.10). (Courtesy of the NASA/*WMAP* Science Team.)

The ratio of the heights of the first and second peak of the angular power spectrum was used to determine the amount of "ordinary" matter with which we are familiar, the baryonic type that comprises atoms. The neutrons and protons found in the nuclei of all atoms are *baryons*.

When the height of the third peak was compared to the other two, scientists estimated the amount of dark, nonbaryonic matter. The comparison indicated that dark matter is five times more abundant than ordinary baryonic matter, and that the combined gravitational pull of both kinds of types of matter is not enough to stop the future expansion of the universe (Table 15.2).

The *COBE* and *WMAP* results have carried cosmology beyond the esoteric realms of theoretical speculation and into precise scientific tests. Definitive new

Table 15.2 Cosmological parameters inferred from *WMAP* observations[a]

Parameter	Name	Value
H_0	Hubble constant[a]	71.0 ± 2.5 km s^{-1} Mpc^{-1}
t_0	Age of expanding universe	$(13.75 \pm 0.13) \times 10^9$ years (after the big bang)
t_{eq}	Equality of matter and radiation (redshift $z_{eq} = 3196 \pm 133$)	76,000 ± 5,000 years
t_{dec}	Decoupling (recombination) (redshift $z_{dec} = 1090.89 \pm 0.69$)	$(3.79 \pm 0.05) \times 10^5$ years (after the big bang)
t_{reion}	Reionization time (redshift $z_{reion} = 10.5 \pm 1.2$)	$(3.5 \pm 1.5) \times 10^8$ years (after the big bang)
$\Omega_b h^2$	Baryonic matter ($\Omega_b = 0.0449 \pm 0.0028$)	0.02258 ± 0.00057
$\Omega_c h^2$	Dark matter density ($\Omega_c = 0.222 \pm 0.026$)	0.1109 ± 0.0056
$\Omega_m h^2$	Total matter density ($\Omega_m = 0.267 \pm 0.026$)	0.1335 ± 0.0056
Ω_Λ	Dark energy density	0.734 ± 0.029
Ω_{tot}	Total density: matter + Energy	1.08 ± 0.09

[a] Parameter values are from *WMAP* only, adapted from Jarosik et al. (2011). The Hubble constant $H_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, or $h \approx 0.71$ for these estimates. A more accurate value indicates that $H_0 = 75$ km s^{-1} Mpc^{-3}. The parameter Ω is the ratio of the specified quantity to the critical amount required to keep the expansion of the universe on the brink of closure. The matter density parameter, for example, is $\Omega_m = \rho_m(t_0)/\rho_C$. Where $\rho_m(t_0)$ is the total mass density, in visible and invisible form, at the present time t_0, and $\rho_C = 3H_0^2/(8\pi G) \approx 1.0 \times 10^{-26}\,\mathrm{kg\,m^{-3}}$ for $H_0 = 75\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ is the critical mass density required to stop the expansion of the universe in the future. The total density parameter Ω_{tot} is the sum of the contributions from visible matter, dark matter, and dark energy, and $\Omega_{tot} = 1.00$ is consistent with inflation and a universe that is described by Euclidean geometry without space curvature.

observational descriptions of the background radiation, with refined cosmological consequences, are expected from the *Planck* mission, launched in May 2009. The initial *Planck* results, announced on March 21, 2013, indicate a Hubble constant of $H_0 = 67.15 \pm 1.2$ km s^{-1} Mpc^{-1}, and a universe with a baryonic (normal) matter content of 4.9 percent ($\Omega_b = 0.049$), a dark matter content of 26.8 percent ($\Omega_c = 0.268$), a total matter density of 31.7 percent ($\Omega_m = 0.317$) and a dark energy content of 68.3 percent ($\Omega_\Lambda = 0.683$). The values of these parameters inferred from the *WMAP* results are given in Table 15.2.

15.3 The Beginning of the Material Universe

15.3.1 The First Three Minutes

George Gamow (1904–1968) and his colleagues proposed that the first elements were formed during the big bang that propelled the universe into expansion (Alpher et al.1948, Sect. 10.5). As they supposed, the lightest atomic nuclei were synthesized before any atoms were created, during the first three minutes following the big bang explosion; the nuclei of less abundant, heavier atoms were manufactured at a later time, within the cores of stars.

The Japanese astrophysicist Chushiro Hayashi (1920–2010) improved their scenario by noting that the temperature was hot enough in the immediate aftermath of the big bang to create neutrinos and positrons, the antimatter particles of the electron, and equilibrium was established between electrons, positrons, neutrons, protons and radiation (Hayashi 1950).

The positrons, denoted e^+, interacted with the neutrons, denoted n, to form protons, p, by the reaction:

$$e^+ + n \rightleftarrows p + \bar{\nu}_e, \tag{15.9}$$

where $\bar{\nu}_e$ denotes an antielectron neutrino. The double arrow in these equations indicates that the reverse reactions happen just as often as the forward reactions, so positrons and neutrons were being produced just as often as they were being consumed. The neutrons also interacted with the electron neutrinos, ν_e to form protons by:

$$\nu_e + n \rightleftarrows p + e^-, \tag{15.10}$$

where e^- denotes an electron.

And in the meantime, the radiation was producing electron–positron pairs, which turned back into radiation just as rapidly, through the reactions:

$$\gamma + \gamma \rightleftarrows e^- + e^+, \tag{15.11}$$

where γ denotes a photon of gamma ray radiation.

During the first few seconds of the expanding universe, when the radiation temperature was greater than 10^9 K, there were no atoms, just radiation and subatomic particles, like neutrons, protons, electrons, positrons, and neutrinos. The equilibrium number density ratio of neutrons, N_n, and protons, N_P, just depended on the mass difference between neutrons and protons, denoted $m_n - m_p$, and the temperature T, according to the relation (Hayashi 1950):

$$\frac{N_n}{N_p} = \exp\left[\frac{-\left(m_n - m_p\right)c^2}{kT}\right], \tag{15.12}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, the neutron mass $m_n = 1.6749274 \times 10^{-27}$ kg, the proton mass $m_p = 1.6726218 \times 10^{-27}$ kg, and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. The mass difference $\Delta m = m_n - m_p = 2.3056 \times 10^{-30}$ kg, and the energy $\Delta E = (m_n - m_p)\, c^2 = \Delta mc^2 \approx 2.07 \times 10^{-13}$ J, which corresponds to a temperature $T = \Delta E/k \approx 1.5 \times 10^{10}$ K.

As the temperature decreased, the number density ratio N_n/N_p also decreased and the protons outnumbered the heavier neutrons. Eventually, the reactions were no longer in thermodynamic equilibrium, neutrons could no longer be created, and the neutron-to-proton ratio became "frozen-in" at an amount that determined the abundance of helium synthesized. Modern computations by David Schramm (1945–1977) and others have conclusively demonstrated that all of the hydrogen and most of the helium nuclei, which now are found in the universe, were indeed synthesized in the immediate aftermath of the big bang (Sect. 10.5, Peebles et al. 1991).

Example: Big-Bang nucleosynthesis of helium
In the first moments of the big bang, the neutrons and protons were in thermodynamic equilibrium with a number density ratio $N_n/N_p = \exp[-\Delta mc^2/(kT)]$, where Δm is the difference between the mass of the neutron and the mass of the proton, c is the speed of light, $\Delta mc^2 \approx 2.07 \times 10^{-13}$ J, the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1} and T is the temperature. Just a few seconds after the big bang, the temperature had cooled to just above 10^9 K, and the production of both positrons and neutrons stopped. The leftover positrons then were consumed by interactions with electrons, and an equilibrium was established in which the relative amounts of neutrons and protons were governed by their mass difference and the temperature. Hayashi (1950) estimated that this frozen-in abundance ratio was $N_n/N_p \approx 0.25$. Alpher et al. (1953) obtained lower amounts of between 0.17 and 0.22 using detailed calculations that depended on the time it takes a free neutron to decay into a proton; outside a nucleus free neutrons are unstable and have a mean lifetime of 881.5 ± 1.3 s or about 14 min 42 s. A lower limit for the "frozen-in" neutron-proton ratio is obtained when the thermal energy kT equals $m_e c^2$, the rest-mass energy of the electron or $N_n/N_p = \exp[-(m_n - m_p)/m_e] \approx \exp(-2.53) \approx 0.08$, where the electron mass $m_e = 9.10938 \times 10^{-31}$ kg.

If N_n and N_p respectively denote the number densities of neutrons and protons before helium nuclei were synthesized, then $N_n/2$ helium nuclei will be formed, since each helium nucleus contains two neutrons and two protons, and the number of protons left over is $N_p - N_n$. The relative number densities of helium, $N(^4He)$, nuclei to hydrogen, $N(H)$, nuclei is:

$$\frac{N(^4He)}{N(H)} = \frac{N_n/2}{N_p - N_n} \approx \frac{1}{12} \approx 0.08. \tag{15.13}$$

The helium mass fraction, Y, is:

$$Y = \frac{4N(^4He)}{N(H) + 4(^4He)} \approx 0.25. \tag{15.14}$$

About one quarter of the mass of the material universe, in baryons, was synthesized into helium in the first few minutes of the expanding universe, and this is consistent with the amount of helium that is now observed in the universe.

Once neutron production stopped, the neutrons and protons could begin combining to form the nuclei of deuterium and helium atoms in amounts governed by the frozen-in abundance ratio of neutrons and protons. All of the protons that did not participate in forming these deuterium and helium nuclei eventually became the nuclei of hydrogen atoms. The production of light atomic nuclei was over, with vastly more hydrogen left behind than anything else.

15.3.2 Formation of the First Atoms, and the Amount of Invisible Dark Matter

Whole atoms were not formed until the expanding universe cooled enough for electrons to combine with protons and helium nuclei to form long-lived hydrogen and helium atoms. This recombination occurred about 400,000 years after the big bang, when the temperature had fallen to about 3,000 K. The rate of recombination was then higher than the rate of ionization by the intense radiation. By the end of recombination, all of the nuclei and electrons had been bound up in atoms, and the universe became transparent to the radiation that then could travel through space without scattering off free, unattached electrons. The cosmic microwave background radiation that we observe in the present was released back then, almost 14 billion years ago.

Because there is no stable nucleus of atomic mass 5 or 8, elements heavier than helium (of mass 4) could not be synthesized by successive collisions with protons (of mass 1). Big-bang nucleosynthesis therefore stopped at helium 4. Heavier elements needed to be synthesized inside stars where the densities are high enough for triple collisions of helium to form carbon, rather than the big bang in which the density had become too low by the time helium nuclei were formed for triple collisions to become significant at the prevailing temperature.

The nuclei of the hydrogen and deuterium atoms and most of the nuclei of helium atoms that now are present in the universe were synthesized in the first 3 min of the expansion, in the immediate aftermath of the big bang and about 14 billion years ago. All of the hydrogen found in stars and interstellar space, or in the Earth's water, and in our body, was produced by this big-bang. And every time you buy a floating party balloon, which has been inflated by helium, you are getting atoms made about 14 billion years ago. Deuterium is destroyed inside stars and, although helium is synthesized in main-sequence stars, the amount of helium formed inside stars over the lifetime of the expanding universe is no more than 10 % of what is now observed in cosmic objects (Hoyle and Tayler 1964).

The cosmological implications of big-bang nucleosynthesis are profound! The agreement of light-element abundances and predictions from the primordial nuclear reactions works only if the density of ordinary matter in the universe – in both visible and invisible forms – is less than 10 % of the critical mass density, ρ_C, required to eventually stop the expansion of the universe.

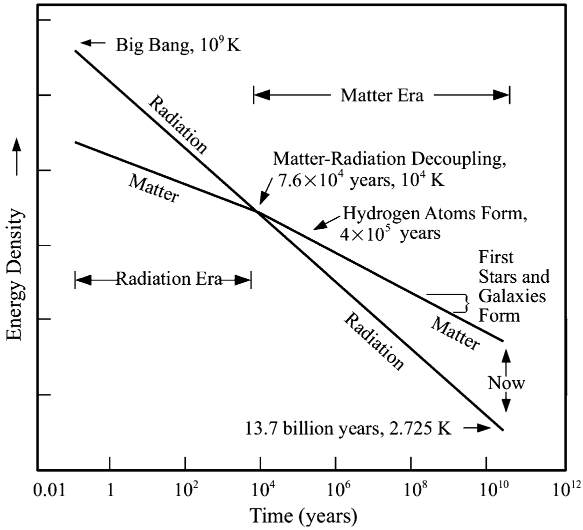

Fig. 15.4 Key events in the expanding universe Shortly after the big bang (*left*) the temperature was about 10^9 K for both radiation and matter, and with its greater energy density radiation dominated the expansion of the universe. With increasing time, the radiation energy density dropped faster than the matter energy density, until the two became equal about 76,000 years after the big bang when the temperature had dropped to about 10^4 K. Thereafter the radiation went its separate way, cooling to the 2.725 K we now detect in the cosmic microwave background radiation 13.7×10^9 years after the big bang (*right*). The matter became cool enough for protons and electrons to recombine and make hydrogen atoms about 400,000 years after the big bang, at a temperature of about 3,000 K, and the first stars and galaxies formed between 10^8 and 10^9 years after the big bang

Observations of cosmic objects with low heavy-element (metals) content indicate that the primordial abundance of helium, or the amount created before the first stars formed and synthesized any helium or heavy elements, is roughly $Y = 0.25$, or 25 % by mass. *WMAP* observations of patterns in the cosmic microwave radiation provide evidence for the presence of helium long before the first stars formed with $Y = 0.326 \pm 0.075$ (Jarosik et al. 2011). The abundance of deuterium, D, relative to hydrogen, H, which is $D/H \geq 10^{-5}$, sets an upper limit to the baryon density parameter of $\Omega_B = \rho_{B0}/\rho_C \leq 0.025\, h^{-2} \approx 0.05$, where the Hubble constant $H_0 = 100\, h \approx 75$ km s^{-1} Mpc^{-1} and the critical mass density required to close the expanding universe in the future is $\rho_C = 3H_0^2/(8\pi G) \approx 10^{-26}\,\text{kg m}^{-3}$ for $H_0 = 75\,\text{km s}^{-1}\,\text{Mpc}^{-1}$. The number of baryons is equal to the number of neutrons and protons in matter in either seen or unseen forms.

When combined, the big-bang nucleosynthesis constraints of all four of the light nuclei, ^4He, D, ^3He, and ^7Li, provide limits to the baryon to photon number density ratio, $\eta = N_B/N_\gamma \approx 3 \times 10^{-10}$, which has remained unchanged since the epoch of electron–positron annihilation a few seconds after the big bang. Here N_B

is the number density of baryons and N_γ is the number density of photons, so the universe contains less than one baryon (neutrons and protons) per billion photons. During the time of *big-bang nucleosynthesis*, the universe was a dilute gas of radiation photons, contaminated by only trace amounts of baryons. At present the number density of photons in the cosmic microwave background radiation is $N_{\gamma 0} \approx 4 \times 10^8\,\mathrm{m}^{-3}$. The baryon density now, $\rho_{B0} = m_p N_{\gamma 0} \eta \approx 2 \times 10^{-28}\,\mathrm{kg\,m}^{-3}$, where the proton mass $m_p = 1.673 \times 10^{-27}$ kg, and the baryon density parameter $\Omega_B = \rho_{B0}/\rho_C \approx 0.05$ (Copi et al. 1995).

Thus, when measurements of the Hubble constant are considered, the results of big-bang nucleosynthesis and the observations of the abundance of light elements indicate that the baryon density is now 0.05 or 5 %, of the critical mass density required to ever halt the current expansion of the universe in the future.

A completely independent estimate of the baryon density is provided by the power spectrum of the anisotropies in the cosmic microwave background radiation. The relative amplitudes of the peaks in the spectrum constrain it to $\Omega_B = 0.044 \pm 0.004$, in excellent agreement with the conclusions based on big-bang nucleosynthesis.

15.3.3 History of the Expanding Universe

As the universe grew larger, the radiation energy density, $\rho_r(t)$, decreased more rapidly than the mass density, $\rho_m(t)$. Eventually, at the time $t_{eq} \approx$ 76,000 years after the big bang, the mass-energy density of the radiation had become equal to that of the matter. Thereafter it was mass that dominated the expansion of the universe. This and other critical times in the expansion are illustrated in Fig. 15.4. Loeb and Barkana (2001) have reviewed the re-ionization of the universe by the first stars and quasars.

If $R(t)$ denotes the radius, or scale factor, of the universe at cosmic time, t, the radiation temperature, $T_r(t)$, falls off as $1/R(t)$. This means that the radiation energy density (Tolman 1934):

$$\rho_r(t) = \frac{aT_r^4(t)}{c^2} \propto \frac{1}{R^4(t)}, \tag{15.15}$$

where the radiation constant $a = 7.5657 \times 10^{-16}\,\mathrm{J\,m^{-3}\,K^{-4}}$, the speed of light $c = 2.9979 \times 10^8\,\mathrm{m\,s^{-1}}$, and the symbol $\propto$ denotes proportional to. The mass density, $\rho_m(t)$, doesn't decrease as rapidly with increasing time and radius, since it thins out with increasing volume, or with

$$\rho_m(t) \propto \frac{1}{R^3(t)}. \tag{15.16}$$

Since

$$T_r(t) \propto \frac{1}{R(t)}, \quad (15.17)$$

the ratio

$$\frac{\rho_m(t)T_r(t)}{\rho_r(t)} = T_r(t_{eq}) \quad (15.18)$$

is a constant for all times t. Here t_{eq} denotes the time when the radiation energy density is equal to the matter density, or when $\rho_r(t_{eq}) = \rho_m(t_{eq})$

Example: Radiation energy density and mass energy density
As the universe expands, the radiation energy density $\rho_r(t)$ at time t will fall off as the inverse fourth power of the radius of the universe, $R(t)$, since the radiation temperature $T_r(t)$ falls off as $1/R(t)$. The mass energy density $\rho_m(t)$ falls off as the inverse cube of the radius, since the volume increases as the radius cubed. This means that $\rho_m(t)T_r(t)/\rho_r(t) =$ constant for all time t. At the present time t_0 we have $T_r(t_0) = 2.725$ K, the $\rho_r(t_0) = a\,T_r^4(t_0)/c^2 \approx 0.842 \times 10^{-32}\,T_r^4(t_0) \approx 4.64 \times 10^{-31}$ kg m^{-3}, where the radiation constant $a = 7.5657 \times 10^{-16}$ J m^{-3} K^{-4} and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. *WMAP* observations indicate that the density parameter Ω_b of normal baryonic matter, like protons and neutrons that are found in atoms, is $\Omega_b = \rho_m(t_0)/\rho_c \approx 0.04$ (Table 15.2), where $\rho_m(t_0)$ is the present mass density in both visible and invisible baryons and the critical mass density required to stop the expansion of the universe in the future is $\rho_c = 3H_0^2/(8\pi G) \approx 1.0 \times 10^{-26}$ kg m^{-3} (Sect. 14.6), the Hubble constant $H_0 = 100\,h$ km s^{-1} Mpc^{-1} $= 3.24 \times 10^{-18}\,h$ s^{-1}, with $h \approx 0.75$, 1 Mpc $= 3.0857 \times 10^{19}$ km, and the Newtonian gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. This means that $\rho_m(t_0) \approx 0.04\,\rho_c \approx 4 \times 10^{-28}$ kg m^{-3}, about a thousand times greater than the present radiation energy density, and that $\rho_m(t_0)T_r(t_0)/\rho_r(t_0) \approx 2 \times 10^3$.

The radiation energy density and mass energy density were equal at a time $t_{eq} \approx 76{,}000$ years $\approx 2.40 \times 10^{12}$ s, where one year $= 3.1557 \times 10^7$ s, after the big bang (Table 15.2). At this time, the radiation temperature has dropped to $T_r(t_{eq}) = 2.15 \times 10^{10}\ t_{eq}^{-1/2} \approx 10^4$ K, and $\rho_m(t_{eq})T_r(t_{eq})/\rho_r(t_{eq}) = T_r(t_{eq}) \approx 10^4$, close enough to the value of this ratio now considering the uncertainties.

The radiation dominated era, when the radiation energy density is greater than the matter density, has a radius that scales as the square root of time, t, or

$$R(t) \propto t^{1/2} \quad \text{for}\, \rho_r(t) > \rho_m(t)\, \text{or for}\; t < t_{eq}, \tag{15.19}$$

and when the matter dominates the expansion of the universe, its radius is:

$$R(t) \propto t^{2/3}\; \text{for}\, \rho_m(t) > \rho_r(t)\, \text{or for}\; t > t_{eq}. \tag{15.20}$$

If radiation was emitted at time t, then its redshift, z, is given by

$$1 + z = \frac{R(t_0)}{R(t)}, \tag{15.21}$$

and the radiation temperature, $T_r(t)$ at that time is given by

$$T_r(t) = (1+z)\, T_r(t_0) = 2.725\,(1+z), \tag{15.22}$$

where the temperature of the background radiation $T_r(t_0) = 2.725 \pm 0.002$ K right now, at time t_0.

WMAP observations indicate that the radiation energy density became equal to the matter density at a redshift of $z_{eq} = 3196 \pm 133$ (Jarosik et al. 2011). Since that time, denoted as t_{eq} for the equality of radiation and matter, the matter dominated the expansion and $R(t)$ grew as $t^{2/3}$, so the transition from the radiation dominated era to the matter dominated era occurred at time t_{eq} where:

$$\frac{t_{eq}}{t_0} = \left[\frac{R(t_0)}{R\left(t_{eq}\right)}\right]^{3/2} = (1+z)^{3/2}, \tag{15.23}$$

and that

$$t_{eq} = \frac{t_0}{\left(1 + z_{eq}\right)^{3/2}} \approx 76,000\ \text{years} \tag{15.24}$$

after the big bang. This indicates that matter has dominated the expansion of the universe for all but a relatively small fraction of the age $t_0 \approx 13.7 \times 10^9$ years of the expanding universe.

Another crucial time is the *decoupling time*, or *recombination time*, t_{dec}, when the temperature had fallen to about 3,000 K, which was cool enough for electrons to begin recombining with protons and helium nuclei to form long-lived hydrogen and helium atoms. That is, the rate of recombination to form atoms was then higher than the rate of atomic ionization by the intense radiation. By the end of recombination, all the nuclei had been bound up in atoms, and the universe became transparent to the radiation; so cosmic microwave background radiation that we observe at the present time was released back then, about 13.3 billion years ago.

The decoupling redshift corresponding to this temperature is $z_{dec} \approx 3{,}000/2.725 \approx 1100$, and $t_{dec} = t_0/(1100)^{3/2} \approx 377{,}000$ years. The *WMAP* estimates give $z_{dec} = 1090.89 \pm 0.69$. This also marks the beginning of the dark ages of the expanding universe, for there were no sources of radiation other than the gradually cooling and darkening cosmic background radiation until stars and galaxies

Table 15.3 Crucial times during the expansion of the universe

Time (after the big bang)	Redshift, z	Temperature (K)	Key events
10^{-14} s	10^{27}	10^{27}	Inflation ends, $\Omega_m + \Omega_\Lambda = 1$.
10 s	4×10^9	10^{10}	Neutron and positron production stops.
3 min	4×10^8	10^9	Big-bang nuclosynthesis ends, light elements H, D, He formed
$t_{eq} \approx$ 76,000 year	3,196	10^4	Radiation domination equals matter domination, transfer from $R(t) \propto t^{1/2}$ to $R(t) \propto t^{2/3}$
$t_{dec} \approx$ 377,000 year	1,100	3,000	Decoupling (recombination) time. radiation decouples from matter, hydrogen atoms recombine, universe becomes transparent to background fluctuations, dark ages begin
$t_{reion} \approx 3.5 \times 10^8$ year	10.5 ± 1.2	31	Reionization time, first stars and galaxies form, universe re-ionized by their radiation, and dark ages end
6×10^9 year	1	5	Dark energy begins acceleration of universe expansion
$t_0 = 13.7 \times 10^9$ year	0	2.725	Age of expanding universe, today, present epoch

formed, about 350 million years after the big bang, they provided beacons of bright light that could ionize surrounding matter. However, by then the universe had thinned out enough that the low-density ionized hydrogen remained transparent, ending the dark ages.

A summary of these milestones in the history of the expanding universe is given in Table 15.3.

Example: When did the first stars and galaxies form?
When atoms coalesced to form stars and galaxies, these shining beacons of intense radiation illuminated the former darkness and re-ionized nearby matter. The *WMAP* observations suggest that this began at a redshift $z_{reion} = 10.5 \pm 1.2$. This would happen at a background radiation temperature $T = (1 + z)\, 2.725 = 31$ K. The corresponding time, t_{reion}, is $t_{reion} = t_0/(1 + z_{reion})^{3/2} \approx 3.5 \times 10^8$ years or about 350 million years after the big bang. Since matter decoupled from radiation, and the dark ages began, about 76,000 years after the big bang, the dark ages lasted about 350 million years and stars and galaxies have been around for about $t_0 - t_{reion} \approx 13.4$ billion years, where the big bang occurred about $t_0 = 13.75$ billion years ago. These stars and galaxies have redshifts of 10.5 or less.

15.4 The First Stars and Galaxies

15.4.1 Pulling Primordial Material Together

Immediately after the big bang, the radiation and matter were distributed smoothly, with almost no structure at all; because the subatomic neutrons and protons were then in thermal equilibrium with the radiation, the material universe must have had a smooth beginning. Departures from complete uniformity in the cosmic microwave background radiation were about one part in one hundred thousand, or with fluctuations ΔT in the temperature T of $\Delta T/T \approx 10^{-5}$. Yet, the material universe we see today has gathered together into stars, galaxies, and clusters of galaxies, so something must have given them shape and form. They are composed of baryonic matter, which includes neutrons, protons, and atoms of any sort, and thus consist of ordinary matter that we encounter or experience in everyday life.

When left alone, a spread out distribution of matter will coalesce about small, initial concentrations, as a result of their gravitational pull on surrounding material, so any slight perturbations in otherwise uniform matter will grow and eventually contract. Given enough time, gravity might magnify the extremely slight irregularities in the initial mass distribution, eventually providing the concentrated structures, the observed stars and galaxies, by gravitational collapse.

The problem with this scenario, first recognized by Georges Lemaître (1894–1966) and rigorously derived by Evgeny Lifshitz (1915–1985), is that the initial irregularities could not have grown fast enough to account for the observed stars and galaxies (Lemaître 1934; Lifshitz 1946). Chance density fluctuations in this matter would have grown too slowly to overcome and resist the overall expansion of the universe, which pushes and tears the material apart as soon as it starts to gather together. Even a billion years is not enough time for the fluctuations in ordinary matter density to gravitationally pull the primeval matter together into luminous stars and galaxies.

We can follow the basic argument by supposing that an initial density perturbation $\Delta\rho_m$ in the mass density ρ_m was of order $\Delta\rho_m/\rho_m = 10^{-5}$, since the subatomic matter was in thermal equilibrium with the radiation with temperature fluctuations of $\Delta T/T = 10^{-5}$. In the matter-dominated era the perturbations will slowly grow with increasing time, t, and the scale factor $R(t)$ as $\Delta\rho_m/\rho_m \propto t^{2/3} \propto R(t)$, but even over the past 13 billion years $R(t)$ has only increased by a factor of a thousand, or 10^3, so the initial perturbations would have grown to $\Delta\rho_m/\rho_m = 10^{-2}$, much smaller than the amount of $\Delta\rho_m/\rho_m = 1$ now observed in galaxies.

However, we have considered only ordinary matter, and the paradox could be resolved if the extra gravitational pull of much greater amounts of invisible dark matter helped clump and shape the expanding universe, pulling together primordial fluctuations in perfectly ordinary matter and forming the seeds from which the first stars and galaxies grew.

According to the *cold dark matter* scenario, galaxies formed first, than gravitationally merged and consolidated into clusters and super-clusters of galaxies as the universe expanded and evolved; for pioneering papers on the cold dark matter hypothesis see Lang (1999).

Understanding how gravity causes the perturbations in ordinary matter to grow in an expanding universe and eventually become galaxies requires studying the interaction between ordinary matter and dark matter. That interaction causes a region of space with more ordinary matter than average to oscillate, sending out waves known as baryonic acoustic oscillations. These sound waves were first predicted in 1970, suggested by *WMAP* fluctuations of the cosmic microwave background in 1999, and measured as rippling imprints in the distribution of galaxies using Sloan Digital Sky Survey data in 2005 (Peebles and Yu 1970; Eisenstein et al. 2005).

15.4.2 When Stars Began to Shine

With the help of cold dark matter, the first stars and galaxies appeared more than 10 billion years ago. We can observe these embryonic galaxies when they were cosmic infants; the light now reaching us began its journey long before the Sun came into existence.

Bromm and Larson (2004) have discussed the first stars; Bromm and Yoshida (2011) provided a review of the first galaxies; Brodie and Stader (2006) discussed extragalactic globular clusters and galaxy formation, and Kravtsov and Borgani (2012) has reviewed the formation of galaxy clusters.

Kennicutt and Evans (2012) have reviewed star formation near and far. Shapley (2011) has reviewed the physical properties of galaxies from redshifts $z = 2$–4, and Giavalisco (2002) has discussed Lyman-break galaxies. Sanders and Mirabel (1996) have reviewed luminous infrared galaxies.

Each galaxy may have formed through the gravitational collapse of a larger, protogalactic cloud, which would become a rotating disk like the Milky Way (Eggen et al. 1962). These flattened, spinning galaxies often show spiral structure, with arms of gas and dust in which new stars are forming.

Not all galaxies have a disk or spiral shape, and the most massive are the giant, rounded, featureless elliptical galaxies. They may result from the collision and subsequent merger of two spiral galaxies. During the encounter, the ordered rotational motions of the stars in the spiral galaxies are transformed by tidal forces, which tear their disks and arms apart and randomize the orbits of their stars. When the merger is complete, a single elliptical galaxy remains, composed of old stars with little or no gas and dust left to form new stars. Many of the giant elliptical galaxies are found in the cores of dense clusters of galaxies where collisions should be frequent on a cosmic time scale.

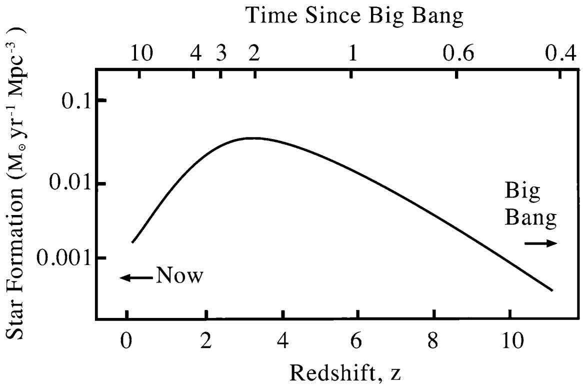

Fig. 15.5 Star–formation rates The star-formation rate, in solar masses per year per cubic Megaparsec, or $M_\odot$ year^{-1} Mpc^{-3}, plotted as a function of redshift, z (*bottom axis*) and time since the beginning of the expanding universe (*top axis*), in units of 10^9 years, or 1 billion years and a G year. The rate of star formation peaked at a redshift of about 3, or roughly 2 billion years after the expansion began, and this rate subsequently has decreased as gravitation pulls more material into stars. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

When astronomers use infrared telescopes, aboard the *Herschel* and *Spitzer* spacecraft, they can peer behind veils of local interstellar dust to see infant stars in distant galaxies. Some *starburst galaxies* are very powerful infrared emitters, with an infrared luminosity greater than a million million, or 10^{12}, times that of the Sun, and an infrared output that is 100 times their visible-light emission. An exceptional amount of interstellar dust in these galaxies absorbs the intense ultraviolet radiation produced by enhanced star formation, and the dust reradiates in the infrared part of the electromagnetic spectrum. Their intense infrared emission and implied dust suggests that these galaxies are forming stars more vigorously than our present-day Milky Way (Fig. 15.5).

Starburst galaxies at large redshift of about $z = 3$ are found by a clever comparison of the ultraviolet and visible radiation of galaxies. The technique is related to the main ultraviolet spectral transitions of hydrogen, the Lyman alpha line at 121.7 nm, which occurs between electron orbits with quantum numbers $m = 2$ and $n = 1$, and all the other Lyman transitions at larger m and the same n, culminating at the Lyman limit at 91.2 nm at very large m. The observed Lyman limit at large redshifts is Doppler shifted into visible wavelengths at $\lambda = 91.2\ (1 + z)$, which is at 360 nm for $z = 3$. Radiation at wavelengths lower than the Lyman limit is almost completely absorbed by the alpha transition of neutral hydrogen in the star-forming regions. At large redshifts the sharp decrease, or break, of the emitted spectrum has been Doppler shifted into the visible region. The starburst galaxies are therefore also known as *Lyman-break galaxies*.

About 2 billion years after the big bang and roughly 12 billion years ago, some starburst galaxies had an exceptionally high rate of star formation, which exceeded

100 stars per year for hundreds of millions of years – much greater than the rate in most galaxies and currently in the Milky Way, which is about one new star every year. Scarcely any stars were being formed during the first half billion years of the expanding universe, perhaps because the gravitational forces were still pulling the galaxies together, and after the bursts of star formation there may have been less material available for forming new stars, since some of it already had been used up in creating other stars.

The fast pace of star formation in galaxies in the early universe could not be continued for long times. It would use up the interstellar gas from which stars are formed in much less time than the age of the universe. So the bursts of star formation could be associated with rare circumstances, perhaps feeding off gas stirred up as a result of collisions or close encounters between galaxies.

The mergers of galaxies may not be the dominant method of high star growth. It could be associated with a voracious consumption of hydrogen gas, which has been observed in greater abundance back then when compared to more recent times. A steady supply of gas may have streamed in from filaments of dark matter.

When the first stars formed out of collapsing clouds of gas and ignited the nuclear reactions that make them shine, the early universe consisted of nothing more than the light elements, hydrogen and helium. These young stars must have been uncontaminated by heavier elements. Some of these "infant" stars most likely were very massive, perhaps with about 100 times as much mass as the Sun; therefore, they would have a relatively short lifetime on the cosmic time-scale. The first massive stars would have exploded as supernovae, spewing out ashes of dust made of heavy elements synthesized within them and spawning the next generation of stars.

The interstellar medium would have become steadily enriched in heavy elements as subsequent generations of massive stars were formed, lived, and expired explosively. They would have seeded their surroundings with elements such as carbon, oxygen, and iron, which were needed for the formation of Earth-like planets and life.

Whether a galaxy is young or old, there will always be many more stars of low mass than there are massive stars (see Sect. 10.1). The stellar mass distribution will also depend on the evolution of stars, which varies with mass. The initial mass distribution can be inferred from the observed stellar luminosity function, or the number of stars of different absolute luminosities, by using the stellar mass-luminosity relation together with a model of how the star formation rate varies with time.

The initial mass distribution for stars more massive than the Sun was quantified by the Cornell astronomer Edwin E. Salpeter (1924–2008), who showed that the number of stars with masses in the range M to $M + dM$ within a specified volume of space, is proportional to $M^{-2.35}$ (Salpeter 1955). In other words, the number of stars in each mass range decreases rapidly with increasing mass.

The oldest stars in our Milky Way Galaxy, which were formed when the universe was only about 1 billion years old, are deficient in heavy elements when compared to stars that are now forming in the Milky Way. No one has yet found a completely pure star that formed out of uncontaminated hydrogen. Perhaps such stars now are inaccessible to direct observation, awaiting the next space telescope that can peer deeper into the remote past.

15.5 The Evolution of Galaxies

15.5.1 Active Galactic Nuclei

By peering out at galaxies that are located at vastly different distances or redshifts, astronomers have shown that the entire observable universe evolves and has a history, and that the properties of galaxies change over vast time scales. The light we detect from a remote galaxy has traveled for a very long time, and was emitted in the galaxy's infancy many billions of years ago. And when our telescopes observe a nearby galaxy, its light may have been generated a few million years ago, after the galaxy has aged for billions of years.

Because no significant change in the equilibrium of galaxies can be produced without a substantial change in the distribution of mass and angular momentum, it was long believed that no significant departures from a stable equilibrium in their shape, form, mass, or luminosity would be produced during most of their lifetimes. Nevertheless, it now is known that the centers of galaxies are locations of pronounced activity that disrupts the expected equilibrium and that galaxies tend to be more active in their youth.

The American astronomer Carl K. Seyfert (1911–1960) provided early observational evidence that the central regions of some galaxies are not in equilibrium when he examined the intense blue centers of certain spiral nebulae – a type subsequently named *Seyfert galaxies* (Seyfert 1943). Although most spirals exhibit spectral lines in absorption, similar to the absorption spectra of stars, the central regions of Seyfert galaxies exhibit intense emission lines of the type produced by ionized emission nebulae (also see Sect. 11.1). They are the emission lines of oxygen [O II], [O III], nitrogen [N II], neon [Ne III] and sulphur [S II], as well as the permitted emission transitions of unionized hydrogen, H, and ionized helium, He II (Table 15.4).

The emission lines of Seyfert galaxies are unexpectedly wide. High-speed motions of the ions and hydrogen atoms have widened the emission lines, and the velocities implied from their widths, when interpreted by the Doppler effect, are up to 8,500 km s^{-1}. Because the central masses derived from the rotation curves of spiral nebulae are no more than 10^{11} solar masses, the escape velocities of the central regions are only a few hundred km s^{-1}. The observed motions at the centers of Seyfert galaxies are therefore far in excess of the expected escape velocities, and they provide the first evidence for violent explosive events in the nuclei of galaxies. Their matter could be flowing out into intergalactic space; some of the Seyfert galaxies exhibit bright filaments that suggest the ejection of gas.

Table 15.4 Intense emission lines found in Seyfert galaxies[a]

Element	Wavelength (nm)	Element	Wavelength (nm)
[O II]	372.62	He I	447.25
[O II]	372.89	He II	468.57
He II	376.89	Hβ	486.1332
Hθ	379.86	[O III]	495.891
He II + Hη	383.56	[O III]	500.684
[Ne III]	386.875	He I	587.56
Hς + He I	388.89	He II	597.7
[Ne III]	396.746	[O I]	630.0304
Hε	397.01	[O I]	636.3776
[S II]	406.85	[N II]	644.803
[S II]	407.65	Hα	656.281
Hδ	410.17	[N II]	658.341
He II + Hγ	433.86	[S II]	671.647
	+ 434.047	[S II]	673.085
[O III]	436.32		

[a] Adapted from Seyfert (1943)

Weedman (1977) has summarized our then current knowledge of Seyfert galaxies. Early considerations of violent activity in the nuclei of galaxies can be found in Ambartsumian (1958), Burbidge et al. (1963) and Lynden-Bell (1969).

Powerful cosmic radio sources provide additional evidence for intense activity in the central regions of young galaxies. But since there was no radio technique for establishing distances, the optical wavelength counterparts of the radio sources had to be used to determine how far away they were. Research groups led by Joseph L. Pawsey (1908–1962) in Australia and by Martin Ryle (1918–1984) in Cambridge, England built interferometers that were used to obtain accurate positions and identify the optical counterparts of the brightest radio sources, named by the constellation they appeared in. The Australian group identified the source named Virgo A – also numbered 3C 274 in the third Cambridge catalogue of bright radio sources – with giant elliptical galaxy, M 87 (Bolton et al. 1949), and an accurate position established with the Cambridge radio interferometer was used to identify the radio source Cygnus A, numbered 3C 405, with another elliptical galaxy (Baade and Minkowski 1954). Like the nuclei of Seyfert galaxies, the optical counterpart of Cygnus A emits strong "forbidden" emission lines of [O II], [O III], [N II] and [Ne III] with widths corresponding to velocities of a least 1,000 km s^{-1}.

Moreover, the Doppler shift of the central wavelengths of these spectral lines exhibited a redshift of 0.0561, and a recession velocity of 16,820 km s^{-1}. Using the Hubble law with this redshift, the radio galaxy Cygnus A lies at a distance of about 224 Mpc, or 731 million light-years, for a Hubble constant $H_0 = 75$ km s^{-1} Mpc^{-1}, and the apparent radio luminosity and distance can be combined to infer an enormous absolute radio luminosity of about 10^{38} J s^{-1}. It is emitting as much power at radio wavelengths as the visible luminosity of a million million stars like the Sun, of absolute luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, whose

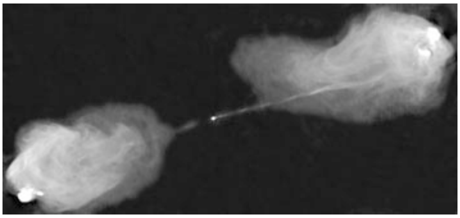

Fig. 15.6 Radio galaxy Cygnus A The radio galaxy Cygnus A, listed as 3C 405 in the third Cambridge catalogue of bright radio sources, which has a radio output 1 million times more powerful than the radio emission of a normal galaxy like the Milky Way. This radio image, taken with the Very Large Array at a wavelength of 6 cm with a field of view of 0.038 × 0.022 degrees, shows two narrow, straight radio-emitting jets of particles that protrude in opposite directions from a giant elliptical galaxy at the center. The redshift of the optically visible elliptical is $z = 0.056075$, indicating a distance of about 224 Mpc or 780 million light-years, and a linear extent for the radio galaxy of about 1 million light-years from end to end. The radio jets probably were ejected along the rotation axis of a super-massive black hole located within a central elliptical galaxy. It had to be active for tens of millions of years to produce the two radio lobes. (Courtesy of NRAO/AUI/NSF.)

most intense radiation is also at the visible, optical wavelengths with relatively dim radio emission.

When the optical image of the elliptical galaxy associated with Cygnus A is combined with maps of the radio signals, it is found that the radio emission is not confined to its visible counterpart; instead it is concentrated in two radio lobes that are separated from the central visible galaxy by hundreds of thousands of light-years. It is as if the radio-emitting clouds were expelled from the central elliptical galaxy, which is detectable only at optically visible wavelengths. The astonishing radio power is attributed to the nonthermal synchrotron radiation of high-speed electrons supplied from the visible center along two oppositely directed jets that feed the radio lobes (Fig. 15.6). These dual jets remain extraordinarily straight and surprisingly stable, energizing the radio lobes and pushing them farther and farther apart.

Harris and Krawcynski (2006) discussed x-ray emission from extragalactic jets; Bridle and Perley (1984) have reviewed extragalactic radio jets; Kellermann and Pauliny-Toth (1981) have reviewed our knowledge of compact radio sources, and Miley (1980) has discussed the structure of extended radio sources.

If the *radio galaxy* has been sending out radio power at the present rate at an estimated million-year lifetime, then it has emitted radio energy equivalent to the complete annihilation of about a 100,000 stars.

Example: Feeding the radio lobes of Cygnus A

The radio galaxy Cygnus A has a redshift of $z = 0.056$, which from the Hubble law $cz = H_0 D$ provides a distance, D, of 224 Mpc $= 6.91 \times 10^{24}$ m and about 730 million light-years for a Hubble constant of $H_0 = 75$ km s^{-1} Mpc^{-1} where the speed of light is $c = 2.9979 \times 10^5$ km s^{-1} and

1 Mpc = 3.0857×10^{22} m. Its two lobes are separated by an angle of $\theta \approx 0.038$ degrees $\approx 137''$ and 6.63×10^{-4} radians, using the conversion factor of 1 radian = 2.06265×10^5 s of arc. The minimum amount of time, t, that the central elliptical galaxy has been feeding the radio lobes is $t = \theta D/(2c)$, where the distance between the visible galaxy and one radio lobe is $\theta D/2$. That time is $t = 7.6 \times 10^{12}$ s = 0.24×10^6 years, where 1 year = 3.1557×10^7 s.

The absolute radio luminosity of Cygnus A is about $L_R \approx 1.3 \times 10^{38}$ J s^{-1} $\approx 3.4 \times 10^{11}$ $L_\odot$, where $L_\odot = 3.828 \times 10^{26}$ J s^{-1} is the visible absolute luminosity of the Sun. Thus, the radio power emitted by Cygnus A is comparable to the visible-light power of 340 billion stars like the Sun, exceeding that of our Milky Way Galaxy of about 100 billion stars. Over its lifetime of a quarter of a million years, Cygnus A emits about 10^{51} J of radio energy, which is comparable to the rest mass energy of 5,594 solar masses, or 5,594 $M_\odot c^2$ for a solar mass $M_\odot = 1.989 \times 10^{30}$ kg and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. Finding a radiation mechanism comparable to the complete annihilation of 5,594 stars was initially problematic, until super-massive black holes provided the answer.

Even more dramatic sources of energy were found still deeper in space and generated a longer time ago. As with radio galaxies, the discovery of the first quasar resulted from the accurate location of a bright radio source, which was determined when the Moon happened to pass in front of it. As the radio astronomer Cyril Hazard (1928–), then at the University of Sydney in Australia, realized, a careful timing of the disappearance and reappearance of the occulted source would establish a precise position, since the location of the Moon's edge is known accurately for any time.

In 1962 Hazard and his colleagues used the occultation method to show that 3C 273 is a double radio source, one component of which apparently coincided with a blue stellar object. This coincidence prompted Maarten Schmidt (1929–) to obtain an optical spectrum of the blue object using the 5 m (200 inch) Palomar telescope, which indicated an exceptionally high recession velocity of 0.16 % of the speed of light. When he told his colleague Jesse Greenstein (1909–2002) about the discovery, Greenstein produced a list of emission line wavelengths for the optically visible counterpart of another radio source 3C 48, and within minutes they had found that it is rushing away with an even faster motion at 37 % of the speed of light.

When these velocities are used to infer distances using the Hubble law, it is found that 3C 48 and 3C 273 are located at distances of billions of light-years. And when their observed luminosities are combined with these distances, it was found that they are shining with the visible blue light of 10 million million, or 10^{13}, Sun-like stars.

The sequence leading to the discovery of these hitherto unknown objects of tremendous velocity, distance, and luminosity happened so quickly that the article reporting the major discoveries appeared together in a six-page sequence in the journal *Nature* (Hazard et al. 1963; Schmidt 1963; Greenstein and Matthews 1963).

Because the bright objects appeared star-like in visible light, they became known as *quasistellar radio sources*, a term that soon was shortened to *quasars*. The quasars had, in fact, been ignored as stars on optical photographs for years. Once quasars were known, astronomers located others by obtaining optical spectra of bright, blue-colored, star-like objects that are located well outside the plane of the Milky Way, where stars are not supposed to be, and measuring the large redshifts characteristic of remote quasars. Thousands of quasars have now been discovered in this way, some of them emitting intense radio signals and many more silent ones with their radios turned off.

Astronomers gradually came to realize that quasars are brilliant, tiny cores, sometimes smaller than the solar system, embedded in much larger, extremely active galaxies, whose outer parts are difficult to detect in the intense quasar glare. From its vantage point in space, the *Hubble Space Telescope* resolved the core quasar light and removed it from the computerized images to detect the faint, fuzzy halo of a host galaxy that is as large as the elliptical galaxies found at the centers of many intense radio sources.

Quasars are believed to be very luminous versions of the same blue nuclei that Seyfert observed in the center of nearby spiral galaxies. The visible-light emission of quasars exhibits the same emission lines as both the Seyfert galaxies and the central elliptical galaxies of radio galaxies (Lynden-Bell 1969). Lang et al. (1975) present a composite Hubble diagram that includes normal galaxies, radio galaxies and quasi-stellar objects in the context of the evolution of the universe.

Seyfert galaxies, radio galaxies, and quasars all belong to a common class, known collectively as *active galactic nuclei*. Modern astronomers are now investigating active galactic nuclei using the *Hubble Space Telescope*, the *Spitzer Space Telescope* and large ground based telescopes operating at visible, infrared, millimeter and radio wavelengths. Fabian (2012) has discussed observational evidence for active galactic nuclei feedback. Ho (2008) reviewed nuclear activity in nearby galaxies; Crenshaw et al. (2003) have reviewed evidence for mass loss from the nuclei of active galaxies; Osterbrock (1991) provided a review of active galactic nuclei; Sulentic et al. (2000) have described broad emission lines in active galactic nuclei; Osterbrock and Mathews (1986) reviewed emission-line regions of active galaxies and QSOs; Weymann et al. (1981) have reviewed absorption lines in the spectra of quasi-stellar objects; and Ulrich et al. (1997) have discussed the variability of active galactic nuclei.

The active galactic nuclei radiate so powerfully over the entire range of the electromagnetic spectrum that they cannot possibly consist of ordinary stars, which emit most of their luminous output in a narrow band of wavelengths grouped around visible light. However, super-massive black holes can account for the prodigious energy output, violent activity, and rapid variations of active galactic nuclei, as well as jets of material that moves out of them at extremely high relativistic speeds that approach the speed of light.

15.5.2 Super-Massive Black Holes

As independently proposed by the astrophysicists Edwin E. Salpeter (1924–), at Cornell University, and Yakov B. Zeldovich (1914–1987) in Moscow, the tremendous luminosity of every radio galaxy and quasar most likely is supplied by a *super-massive black hole*, which emits luminous radiation as its powerful gravity pulls in surrounding stars and gas (Salpeter 1964; Zeldovich 1964). The gravitational pull of a mass equivalent to 100 million Suns is needed to balance the visible quasar luminosity; otherwise its radiation pressure would blow away the quasar. Such a super-massive black hole would be sufficiently small and powerful enough to explain the tiny sizes and the colossal brightness of quasars. The super-massive black hole's rotational energy is used to accelerate charged particles and spew them out in diametrically opposite directions along its rotation axis at about the speed of light, continuously feeding the two radio lobes commonly found symmetrically placed from the center of radio galaxies and quasars.

Rees (1984) has reviewed black hole models for active galactic nuclei, and Begelman et al. (1984) and Lang (1999) have reviewed the theory of extragalactic radio sources. Longair (2011) provides a detailed treatment of the high-energy astrophysics used to describe cosmic radio and x-ray emission.

Kormendy and Richstone (1995) have reviewed the search for super-massive black holes in galactic nuclei. The classic example is M 87, a giant elliptical galaxy whose central spinning disk of hot gas indicates that a super-massive black hole resides at its center. M 87 is close enough to measure the motions of stars, and their increasing velocities toward the center indicate that billions of solar masses must be crammed within a very small, unseen volume to keep the high-velocity stars from flying into space (Sargent et al. 1978; Macchetto et al. 1997; Gebhardt and Thomas 2009).

A one-sided jet of gas emerges from the center of M 87 and stretches out into one of the two lobes of the radio galaxy Virgo A – numbered 3C 274 in the Cambridge survey (Fig. 15.7). The motions of bright concentrations in the jet indicate that they are traveling outward at about half the speed of light. And Very Long Baseline Interferometry observations with widely separated radio telescopes reveal that the M 87 jet emerges from a region at most 6 light-years across, most likely harboring the super-massive black hole that produces the jet.

Monstrous, super-massive black holes seem to inhabit the centers of all galaxies. They are massive, scaled-up versions of stellar black holes, with millions if not billions of times the mass of the Sun packed into a region only a few light-years across. Like their stellar counterparts, the super-massive black holes cannot be observed directly. Their presence is inferred from the orbital motion of nearby

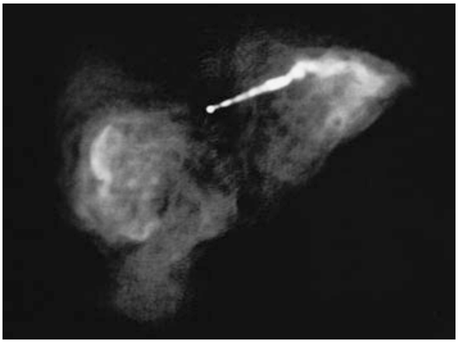

Fig. 15.7 Radio jet from M 87 The bright radio source Virgo A, also designated 3C 274, coincides with M 87, a giant elliptical galaxy with a redshift of $z = 0.00436$ and a distance of about 17.4 Mpc or 56.8 million light-years. M 87 is the largest and brightest galaxy within the Virgo cluster of galaxies. The core of M 87 contains a super-massive black hole of about 6.3 billion solar masses, or $6.3 \times 10^9 M_\odot$, which is 1,500 times more massive than the black hole at the center of our Milky Way Galaxy. This radio map, made with the Very Large Array, shows two elongated lobes, one on either side of the center of M 87, apparently fed by the super-massive black hole. The most intense radio emission comes from a jet that emerges from the core of the galaxy and extends about 5,000 light-years into one of the two lobes. The observed high-speed motion of bright knots in the jet implies that its radio-emitting electrons are traveling at nearly the speed of light. (Courtesy of NRAO/AUI/NSF.)

visible stars the trajectories of which are guided by the otherwise invisible black holes.

The faster the stars are moving, the more gravity – and therefore mass – is needed to hold the stars in their orbits. By measuring the sharp rise in orbital velocity at close distances from galaxy centers, astronomers have weighed unseen super-massive black holes in nearby giant elliptical galaxies, which are the brightest galaxies in clusters of galaxies. The central black-hole powerhouse in relatively nearby galaxies, designated M 87, NGC 3842, and NGC 4889, tips the scales at 6.3, 9.7 and 21 billion solar masses, respectively. Without a gravitational pull equivalent to about 10 billion Sun-like stars, the close, fast-moving stars would fly away from the galaxies (McConnell et al. 2011). Such central, super-massive black holes most likely reside in more distant galaxies that are too far away to resolve central stars and measure their motions.

Quasars and active galactic nuclei become increasingly numerous as we look deeper into space, at larger redshifts. The number density of unobscured quasars peaks at a redshift between $z = 2$ and $z = 3$, which indicates that the hot, luminous spurt of activity happened in the distant past, about 10 billion years ago and shortly after the first galaxies were born. At smaller redshifts and closer distances, corresponding to an old age, there are relatively few quasars. The spurt of activity apparently became worn out and used up as the galaxies grew older.

To power the youthful activity of a quasar, there has to be about 1 solar mass per year of gas flowing into the super-massive black hole. Therefore, billions of stars or the equivalent amount of gas must be consumed as its active nucleus evolves during the course of billions of years. The supply dwindles away over time and the activity dies down, but the black hole does not disappear.

Most galaxies probably contain super-massive black holes at their center. Those in the older, nearby galaxies are the starving remains of former quasars, with a dwindling supply of material that once fed a higher rate of activity. They are found in ordinary nearby galaxies, such as Andromeda, whose cores are surviving fossils of former quasars. Our Galaxy, the Milky Way, is almost as old as the observable universe, and it contains a central super-massive black hole. However, its mass is equivalent to only about 1 million stars like the Sun rather than the billions in some super-massive black holes (see Sect. 14.1).

15.5.3 Gamma-Ray Bursts

The brightest sources found in the universe, at least so far, are the gamma-ray bursts whose duration is measured in seconds or less and which never reappear in exactly the same part of the sky. They emit energy at a gamma-ray wavelength shorter than 10^{-11} m, so each photon of a gamma ray is about 100,000 times more energetic than a photon of visible light. When it was found that they originate in remote galaxies, it was realized that the observed gamma ray bursts might radiate, for a few seconds, gamma-ray energy far in excess of the visible-light energy emitted by galaxies. These gamma-ray bursts can briefly become the brightest electromagnetic events in the universe.

The initial burst is usually followed by a longer-lived "afterglow" emitted at longer wavelengths, from x-rays to optical and radio of galaxies. Gamma-ray bursts are attributed to intense radiation emitted during a supernova explosion when a rapidly rotating, high mass star collapses to form a neutron star or black hole.

The discovery of the gamma ray bursts was the unexpected result of defense satellite observations designed to detect clandestine nuclear bomb explosions in the Earth's atmosphere, on the Moon, or in outer space, but brief, intense gamma-ray bursts were instead found to be coming from the distant Cosmos.

The secret *Vela* satellites were launched and operated in identical pairs on opposite sides of a circular orbit around the Sun, eventually with sufficient time resolution to determine the direction of the source from the difference in arrival time of its radiation at two or more satellites. Although a nuclear bomb was never detected, unexpected flashes of gamma rays were discovered coming from different places in deep space, with an existence measured in seconds. Moreover, the lag between the burst arrival times at the two defense satellites indicated that they originated far beyond the solar system, and were therefore of cosmic origin.

The gamma-ray bursts were kept secret for about five years, until the Los Alamos scientists, Ray W. Klebesadel (1932–), Ian B. Strong (1930–) and Roy A. Olson (1924–), described the discovery in the *Astrophysical Journal*

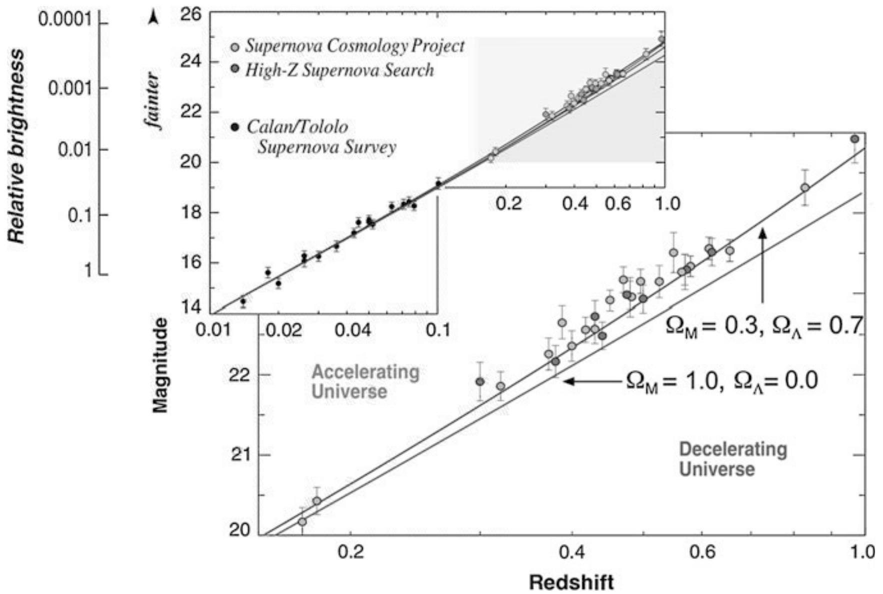

Fig. 15.8 The accelerating expansion of the universe The Hubble diagram plot of the apparent magnitude of Type Ia supernovae plotted as a function of their redshift. At a redshift below about $z = 0.1$, there is a linear fit to the data, but at larger redshifts, the observations begin to diverge from a straight line. The curved departures for distant supernovae at high redshift indicate an acceleration in which the speed of expansion is increasing. The observed data can be compared to cosmological models with different values of the omega parameter, Ω. It is the ratio of the inferred density to the critical mass density needed to stop the expansion of the universe in the future. The subscript Λ denotes the cosmological constant, a possible form of dark energy, and the subscript m denotes matter. (Adapted from Saul Perlmutter, *Physics Today* April 2003. From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

(Klebesadel et al. 1973). Civilian astronomers might blame the military for keeping such an important discovery hidden for years, but it is also likely that the gamma-ray bursts would have never been discovered if it wasn't for the defense satellites. No other government agency was likely to fund a speculative program designed to look for such totally unknown, unsuspected, and unprecedented events in outer space.

The long lasting afterglow of the cosmic gamma-ray bursts was eventually detected at visible wavelengths, enabling the distances to be determined by spectroscopy and the Hubble law, and the enormous energies realized (Van Paradijs et al. 1997; Frail et al. 1997; Metzger et al. 1997; Kulkarni et al. 1998). If these bursts emit radiation in all directions, then the total gamma ray luminosity of the burst can briefly exceed the combined visible-light luminosity of millions of galaxies. A lower gamma ray luminosity is inferred if the bursts are beamed, like a pulsar's radio emission. At least some of the gamma ray bursts are attributed to powerful extragalactic supernovae explosions of very massive stars whose centers collapse into black holes.

Gehrels et al. (2009) have reviewed gamma-ray bursts in the *Swift* era; Fishman and Meegan (1995) discussed earlier observations of gamma ray bursts. Woosley and Bloom (2006) have presented the supernova-gamma-ray burst connection, and Mészáros (2002) has provided a review of theories of gamma-ray bursts. Paradijs et al. (2000) have reviewed gamma-ray burst afterglows; and Weiler, Panagia, and Sramek (2002) have summarized our knowledge of radio emission from supernovae and gamma-ray bursts. Fender and Belloni (2004) have discussed GRS 1915 + 105 and the disc-jet coupling in accreting black hole systems.

15.6 Dark Energy, the Cosmological Constant, and How it All Ends

15.6.1 Discovery of Dark Energy

The redshift, z, of a galaxy increases with its distance, D, according to the Hubble law $cz = H_0 D$, where c is the speed of light and H_0 is the Hubble constant. Since distances can only be independently measured for nearby galaxies, the law is expressed as a redshift – magnitude relation applicable at larger redshifts. This relation is given by

$$m = 5\,\log(cz/H_0) + M + 25, \tag{15.25}$$

where the apparent magnitude is denoted by m and the absolute magnitude by M. If all galaxies have the same absolute magnitude, then a plot of m against log z will describe a straight line whose slope is related to the Hubble constant.

At redshifts greater than one, observable departures from linearity in the apparent magnitude – log redshift diagram could occur due to changes in the expansion velocity. If the expansion of the universe is slowing down and decelerating, for example, then the apparent brightness would be greater than expected from uniform expansion, and the apparent magnitude less than expected since greater brightness means a smaller magnitude.

Supernova explosions of Type Ia, denoted by SNe Ia, provide a useful method for measuring such possible effects (Reiss et al. 1995, 1996; Branch, 1988). Since they are the result of a thermonuclear explosion of a white dwarf star that has grown above the Chandrasekhar limit (Sect. 13.4), there is little variation in their absolute magnitude of $M_B = 19.6 \pm 0.2$ in blue light,and they are so luminous that they can be detected at relatively large redshifts of $z = 0.5$–1.5.

In the late 1980s, a group led by Saul Perlmutter (1959–) at the Lawrence Berkeley National Laboratory began a dedicated search for Type Ia supernovae to measure how fast the expansion of the universe was slowing down, due to the gravitational pull of its combined matter. By 1998, his *Supernova Cosmology*

Project, and a rival group, dubbed the *High-z Supernova Search*, announced that the light from the distant supernovae was fainter than predicted, which meant that the galaxies are speeding up, expanding at a quickening pace, and accelerating instead of slowing down by gravity (Perlmutter et al. 1998, 1999; Riess et al. 1988; Schmidt et al. 1988). In other words, the distant galaxies were not where they were supposed to be, and the space they are in seems to be expanding at a faster rate as time goes on.

The redshift – magnitude diagram of SNe Ia has a straight-line, linear shape at low redshifts, as expected from the Hubble law, but unexpected nonlinear effects appear at large redshifts (Fig. 15.8). The line describing the data indicates that the galaxies are expanding at a quickening pace.

Saul Perlmutter, Brian P. Schmidt (1967–), and Adam G. Riess (1969–) were awarded the 2011 Nobel Prize in Physics for the discovery of the accelerating expansion of the universe through observations of distant supernovae.

Distant galaxies are being accelerated by the anti-gravity push of a mysterious *dark energy*. So the fate of the universe is no longer supposed to depend on its mass, but rather on its energy. If dark energy retains its vigor, the universe will not stop expanding.

Dark energy, which pushes matter apart, is not the same as dark matter, which encourages attraction. But the discovery of dark energy did do away with the need for overwhelming amounts of dark matter to keep the universe poised on the edge of future collapse, but never quite pulling it there. No more than one quarter of the critical mass density is now imagined to reside in mass of any kind, and astronomer's observations of such a low-density universe are now widely accepted. Dark energy has taken over; perhaps keeping the universe at the brink of closure within "flat" space, which is described by Euclidean geometry.

Frieman et al. (2008) have reviewed dark energy and the acceleration of the expansion of the universe. Ratra and Vogeley (2008) have reviewed the current standard model for evolution of the universe, including big-bang cosmology, inflation, dark matter and dark energy, as well as the formation and observations of galaxies and stars.

15.6.2 Using the Cosmological Constant to Describe Dark Energy

The trouble is, nobody understands this mysterious something, this dark energy that permeates space and eventually overwhelms the gravitational self-attraction of the entire material universe. But an old idea, termed the cosmological constant, has been revived to give dark energy another name and couch it in mathematical terms. That's the anti-gravity fudge factor that Einstein introduced to stabilize a non-moving universe against collapse.

When the expanding universe had not yet been discovered, it looked as if the unrelenting, universal attraction of gravity would cause the eventual collapse of an unmoving universe, so Einstein (1917a, b) inserted a mathematical fix, the cosmological constant, into his relativity equations. The extra term represented the repulsive force of an unknown and undetected form of energy that permeated space and exerted a sort of outward pressure that opposed gravity and kept the universe from collapsing.

In a little more than a decade, it was discovered that the galaxies are moving away from us in a cosmic expansion, so the universe wasn't static, or non-moving, after all. Einstein therefore abandoned the cosmological constant, and stated that the *ad hoc* term was greatly detrimental to the formal beauty of his theory.

But the artifice stubbornly refused to die, and has been repeatedly invoked whenever cosmologists have had trouble reconciling their theories with observations. All they had to do was revive the term, stick it in the relevant equations, and adjust its value. As an example, Georges Lemaître (1894–1966) constructed a model universe with two periods of accelerated expansion, one at the beginning and one later, and a more gentle coasting period in between (Fig. 15.9), proclaiming that the expansion thus took place in three phases: a first period of rapid expansion a second period of slowing-up, followed by a third period of accelerated expansion (Lemaître 1931a, b). It is doubtless in the third period we find ourselves today. This interpretation is somewhat similar to some modern explanations of dark energy, beginning with rapid inflation and with a currently accelerated expansion that might invoke the cosmological constant.

Peebles and Ratra (2003) have provided an extensive review of the cosmological constant and dark energy, whereas Carroll et al. (1992) provided a review of the cosmological constant in another context.

The equations that describe a homogeneous, isotropic expanding universe, given in Chap. 14, Sect. 14.6 for zero cosmological constant $\Lambda = 0$, have to be rewritten to allow for a non-zero value of this constant. The Hubble expansion parameter $H(t)$ is then given by:

$$H^2(\mathrm{t}) = \left(\frac{\dot{R}(t)}{R(t)}\right)^2 = \left(\frac{dR(t)/dt}{R(t)}\right)^2 = \frac{8\pi}{3}G\rho(t) - \frac{kc^2}{R^2(t)} + \frac{\Lambda}{3}, \tag{15.26}$$

where $R(t)$ is the scale factor of the universe, $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2} is the gravitational constant, $\rho(t)$ is the mass-energy density, which is equal to the radiation energy density $\rho_r(t)$ in the early radiation-dominated era and the matter density $\rho_m(t)$ in the current matter-dominated era, the space curvature constant is $k = -1$, 0 or +1, and the cosmological constant is denoted by Λ.

Contemporary observations are consistent with $k = 0$, and in this situation, we have:

$$\frac{H^2(t)}{H^2(t_0)} = \frac{8\pi G\rho(t)}{3H_0^2} + \frac{\Lambda}{3H_0^2} = \frac{\rho(t)}{\rho_c} + \frac{\Lambda}{3H_0^2}, \tag{15.27}$$

With

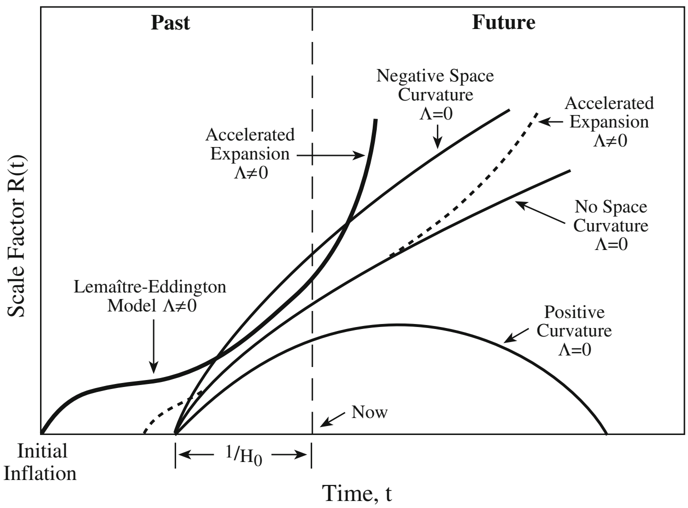

Fig. 15.9 Models of the expanding universe Schematic representation of various cosmological models showing the size, or scale factor R(t), of the expanding universe as a function of time, t. The approximate age since the expansion began is given by $1/H_0$ where H_0 is the Hubble constant. Since the expansion age was once thought to be smaller than the age of the oldest rocks on Earth, Georges Lemaître (1894–1966) and Arthur Eddington (1822–1944) independently used a cosmological repulsion term with Einstein's *General Theory of Relativity* to permit an adjustable age for the universe. The adjustable term is symbolized by a non-zero value for the cosmological constant Λ. In this interpretation (*thick solid line*), the universe began with expansion against gravity, followed by an essentially non-moving stagnation in which gravitation and cosmological repulsion were nearly in balance. This coasting period was then followed by an accelerated expansion driven by the cosmological repulsion. Three models with no cosmological constant, or with $\Lambda = 0$, describe three future possibilities for the universe with no dark energy (*thin solid lines*). It can become closed with positive space curvature, forever open with negative space curvature, and always open and no curvature of space. In recent times, this last option, of never ending expansion in space without curvature, has been modified by a non-zero cosmological constant to give a boost to some age estimates and permit an accelerated expansion by a mysterious dark energy (*dashed lines*)

$$\rho_C = \frac{3H_0^2}{8\pi G} = 1.879 \times 10^{-26}\, h^2\, \mathrm{kg\,m^{-3}} \approx 1.0 \times 10^{-26}\, \mathrm{kg\,m^{-3}}, \tag{15.28}$$

with the Hubble constant $H_0 = 100\, h$ km s^{-1} Mpc^{-1}, $h \approx 0.75$, and 1 Mpc $= 3.0857 \times 10^{19}$ km.

Once the cosmological term has grown large enough, it dominates the right side of the expansion-parameter equation and we have

$$\dot{R}(t) \approx \left(\frac{\Lambda}{3}\right)^{\frac{1}{2}} R(t), \tag{15.29}$$

which has the solution

$$R(t) \propto \exp\left[\left(\frac{\Lambda}{3}\right)^{\frac{1}{2}} t\right] = \exp(H_C t), \tag{15.30}$$

where the Hubble expansion parameter $H(t)$ has become a constant H_C, and the universe enters an exponential expansion.

The other Friedmann equation that involves the pressure P becomes:

$$\frac{\ddot{R}(t)}{R(t)} = \frac{d^2R(t)/dt}{R(t)} = -\frac{4\pi G}{3c^2}\left(\rho\,(t)c^2 + 3P\right) + \frac{\Lambda}{3}. \tag{15.31}$$

This equation implies that with large positive cosmological constant Λ the term $d^2R(t)/dt$ becomes positive and the expansion of the universe accelerates, as opposed to deceleration that could occur without such a term.

The present value of the deceleration parameter q_0 for any value of Λ is given by

$$q_0 = \frac{\ddot{R}(t_0)}{R(t_0)H^2_{0''/>}} = \frac{\Omega_m(t_0)}{2} - \frac{\Lambda c^2}{3H_0^2}, \tag{15.32}$$

where the pressure is now negligible and $H_0 = H(t_0) = 100\ h$ km s^{-1} Mpc^{-1} is the Hubble constant and $h \approx 0.75$.

The power of the cosmological constant is measured from its density parameter Ω_Λ given by

$$\Omega_\Lambda = \Omega_\Lambda(t_0) = \frac{\Lambda c^2}{3H_0^2}, \tag{15.33}$$

while the matter density parameter is given in its usual way

$$\Omega_m = \Omega_m(t_0) = \frac{\rho_m(t_0)}{\rho_c}. \tag{15.34}$$

More than a decade of observations of type SNe Ia has confirmed that their redshift – magnitude relation goes non-linear at large redshifts, permitting astronomers to measure parameters that describe the unexpected effect. The initial results, obtained in 1998, were interpreted by a total matter density parameter $\Omega_m = 0.28 \pm 0.085$, for invisible and visible matter assuming a cosmological constant of $\Lambda = 0$ (Perlmutter et al. 1998), but it was soon realized that the universe was expanding, propelled by a dark energy that could be interpreted in terms of a non-zero cosmological constant (Riess et al. 1998). Ten years later, data for hundreds of SNe Ia were used to suggest a cosmological-constant density factor of $\Omega_\Lambda = 0.713 \pm 0.027$ for a flat universe with a space curvature constant of $k = 0$ (Kowalski et al. 2008). When combined with other data, tight limits of $\Omega_m = 0.237 \pm 0.010$ were also realized (Fig. 15.10) and the previous Table 15.2 of *WMAP* results).

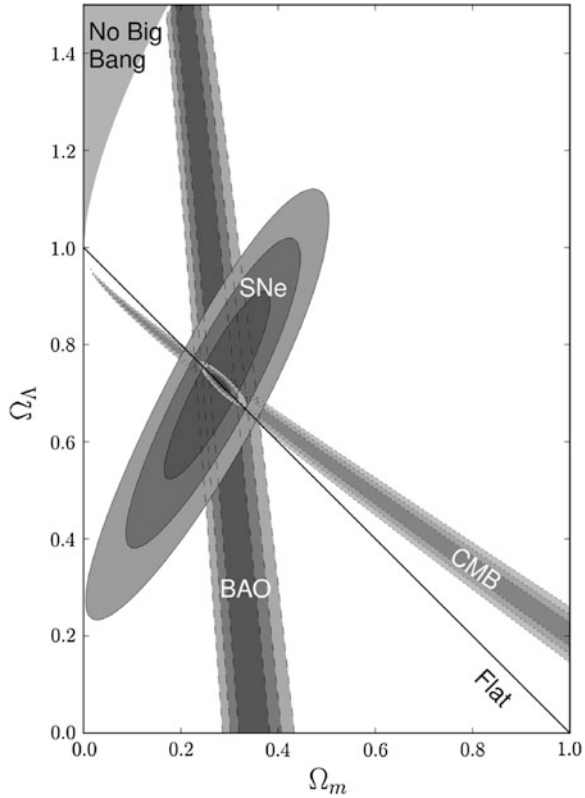

Fig. 15.10 Dark mass and dark energy constraints Three independent sets of observations provide constraints to the mass and energy content of the universe. Studies of high-redshift Type Ia supernovae, designated SNe, constrain the difference between the density of matter and the density of dark energy in the universe. This diagram illustrates the results from the Union 2.1 compilation of 833 SNe drawn from 19 datasets from 2008 to 2011. Anisotropies in the cosmic microwave background radiation, denoted CMB, constrain their sum. Investigations of Baryon Acoustic Oscillations, abbreviated BAO, detected in data from the Sloan Digital Sky Survey provide other constraints. Here the densities are given in terms of the omega parameter, denoted by Ω. It is the ratio of the inferred density to the critical mass density needed to stop the expansion in the future. The subscript Λ denotes the cosmological constant, a possible form of dark energy, while the subscript *m* denotes matter. Galaxy observations indicate a mass density of at most $\Omega_m = 0.3$. The theoretical expectation of an inflationary universe without spatial curvature requires $\Omega_m + \Omega_\Lambda = 1.0$, and significant dark energy with $\Omega_\Lambda = 0.7$. The observations are consistent with such a flat universe, described by ordinary Euclidean geometry. [Courtesy of the Supernova Cosmology Project, whose compilation of 580 SNe is available at http://supernova/lbl.gov/ - also see N. Suzuki et al, "The *Hubble Space Telescope* cluster supernova survey: V. Improving the dark energy constraints above z > 1 and building an early-type-hosted supernova sample", *Astrophysical Journal* **746**, 85 (2012).]

As also previously mentioned, recent *Planck* mission results indicate $\Omega_m = 0.317$ and $\Omega_\Lambda = 0.683$ also with a flat universe of $\Omega_m + \Omega_\Lambda = 1.000$.

The SNe Ia redshift–magnitude diagram, the *WMAP* fluctuations in the cosmic microwave background radiation, and other astronomical observations, have been combined to show that $\Omega_\Lambda = 0.728 \pm 0.015$, and $\Omega_m h^2 = 0.1334 \pm 0.0156$ with $h \approx 0.75$, so $\Omega_m + \Omega_\Lambda \approx 1$ with a space curvature constant $k = 0$ and non-curved Euclidean space. In this condition, at any time, t,

$$\Omega_m + \Omega_\Lambda = \Omega_m(t) + \Omega_\Lambda(t) = 1. \tag{15.35}$$

Allen et al. (2011) have reviewed the determination of cosmological parameters from observations of galaxy clusters. Leibundgut (2001) has reviewed the cosmological implications of observations of Type Ia supernovae; and Branch (1998) has reviewed Type Ia supernovae and the Hubble constant .

15.6.3 When Stars Cease to Shine

Ever since the discovery of the expansion of the universe, we have known that the universe is slowly and inexorably approaching an end. There has never been any known force that can prevent the observable universe from steadily moving into darkness. As Georges Lemaître (1894–1966) so eloquently stated "The evolution of the world can be compared to a display of fireworks that has just ended: some few red wisps, ashes, and smoke. Standing on a well-chilled cinder, we see the slow fading of the suns, and we try to recall the vanished brilliance of the origin of the worlds (Lemaître 1931a, 1931, 1950)."

Dark energy has been found, but that discovery may not help matters. It gives an extra outward push to the expansion and further reduces the power of mass to stop the expanding universe in the future. If the acceleration caused by dark energy continues unabated at the current rate, all the galaxies will be moving apart so quickly that they cannot communicate with one another in about 150 billion years, disappearing over the cosmic horizon.

Eventually, in about 100 trillion years, all of the interstellar gas and dust from which new stars condense finally will be used up, and new stars will cease to form in any galaxy. But if dark energy weakens as time goes on, expending its strength, then gravity and mass may take over and eventually pull back on the outward moving galaxies, ultimately reversing the expansion and dragging the universe back, melting it down and remaking the big bang. So no one knows for sure just how it will all end.