# 1
# An Introduction to Cosmology

## 1.1
## Unity of Physical Laws

Central to our thinking about the structure of the universe is the belief that the laws of nature ought to be the same throughout. This assumption is supported, for instance, by the observation that the spectra of elements found on Earth and those of distant galaxies contain the same atomic spectral lines. Another firmly held belief is that physical laws are but manifestations of a few underlying fundamental principles. To discover those principles is, in Newton's words, the "business of experimental philosophy" (*Opticks*, 1706).

The formulation of *classical mechanics* by Isaac Newton represents a giant milestone in this quest. In his masterpiece *Philosophiae Naturalis Principia Mathematica*, published in 1687, Newton showed that physical phenomena can be explained by a simple set of laws expressed in mathematical form. In particular, he demonstrated that the planetary orbits could be accounted for by the same law of gravitation as the motion of terrestrial objects in free fall. Newton's law of gravitation correctly describes the motion of a wide variety of celestial objects, from binary stars to galaxies.

One of the most profound unifications of physical laws was initiated by Michael Faraday and accomplished by James Clerk Maxwell in the 1860s. Maxwell's unified *theory of electricity and magnetism* – the ultimate achievement of 19th century physics – predicted the existence of the electromagnetic field that propagates through space with the speed of light. This prediction forms the basis for another remarkable insight of his:

> *"We can scarcely avoid the inference that light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena."*

The first experimental confirmation of the existence of electromagnetic waves was made in 1887 by Heinrich Hertz.

*Relativity, Astrophysics and Cosmology* Radoje Belušević

ISBN: 978-3-527-40764-4

The laws of electromagnetism are invariant under the Lorentz transformation of space and time coordinates (unlike Newtonian mechanics, which is invariant under the conceptually simpler Galileo transformation), and are thus the origin of Albert Einstein's *special theory of relativity* (1905). Einstein's theory brought about a unification of the concepts of space and time into a four-dimensional *spacetime continuum*, and also of energy and mass through the relation $E = mc^2$, where $c$ is the speed of light *in vacuo*.

To address problems involving many particles (many degrees of freedom), Rudolf Clausius, James Clerk Maxwell, Ludwig Boltzmann, Josiah Willard Gibbs and others developed *statistical physics*. The concept of probability was introduced by Boltzmann to explain the apparent irreversibility of the macroscopic world. For instance, it was found that the second law of thermodynamics does not have an absolute validity, but rather an extremely high probability:

> *In any isolated macroscopic system the only allowed processes are those evolving from a less probable to a more probable macrostate, i.e., those involving no entropy decrease.*

This law results from the fact that there are always many more disordered states than there are ordered ones.

The *kinetic theory of gases* was another great accomplishment of Maxwell. It explained the concept of temperature in terms of a chaotic motion of molecules, thus bridging the gap between mechanics and thermodynamics.

The existence of *discrete energy levels* was conjectured by Boltzmann in 1872. Unlike Max Planck, who believed that light is emitted discontinously but travels through space as a classical electromagnetic wave, Einstein assumed that the energy in a light beam propagates as *field quanta*, called *photons*. The photon energy, E, and momentum, $p$, are related to the wave frequency of light, $\nu$, through $E = h\nu$ and $p = h\nu/c$ ($h$ is Planck's constant).

A momentous step – although not fully appreciated for almost a century – towards a unified theory of particles and waves was made by a mathematician, William Rowan Hamilton. In 1834, he realized that there was a similarity between the Hamilton–Jacobi equation in mechanics and the Fermat principle in optics:

> *The propagation of a particle in a variable potential is formally equivalent to the propagation of light in a medium with a changing index of refraction.*

Hamilton's insight bore fruit in 1923 when Louis de Broglie suggested that the wave-particle duality of radiation should have its counterpart in a *particle–wave duality* of matter. According to de Broglie, a wave with wavelength $\lambda$ that propagates in an infinite medium has associated with it a particle, or *quantum*, of momentum $p = h/\lambda$. This concept is at the heart of Erwin Schrödinger's wave mechanics and its probabilistic view of physical phenomena, which was introduced into *quantum theory* in 1926 by Max Born. A striking empirical verification of electron waves came in 1927, when Clinton Davisson and Lester Germer observed the phenomenon of electron diffraction.

The advent of quantum mechanics in the 1920s (Max Born, Paul Dirac, Werner Heisenberg, Pascual Jordan, Wolfgang Pauli and Erwin Schrödinger) culminated in the conceptual foundation of *quantum field theory* (Dirac, 1927) and the *relativistic quantum theory* of the electron (Dirac, 1928), two cornerstones of contemporary physics. Dirac's relativistic quantum theory proved to be exceptionally fruitful, enabling him to predict the existence of spin in 1928 and antimatter in 1931. Two decades later, these developments led to the creation of a relativistic theory of photons and electrons, called *quantum electrodynamics* (QED), by Richard Feynman, Julian Schwinger and Shinichiro Tomonaga. Their work, in turn, laid foundations for the unification of weak and electromagnetic interactions, which was accomplished in the 1960s through the *electroweak theory* of Sheldon Glashow, Abdus Salam and Steven Weinberg. A similar theory describing the strong interactions between quarks (the building blocks of matter), called *quantum chromodynamics* (QCD), was proposed in the 1970s by David Gross, David Politzer and Frank Wilczek. Both theories are based on the principle of local gauge invariance, which states that

> *A physical theory described by a Lagrangian that is invariant under certain symmetry transformations should not be affected when these transformations are performed at an arbitrary spacetime point.*

All current theories of the fundamental interactions between the basic constituents of matter are *gauge theories*.

An intimate union between physics and geometry was proposed in 1915 by Albert Einstein. In his *general theory of relativity* (GRT), the gravitational field is a manifestation of the curvature of spacetime. The gravitational and electromagnetic fields are similar in the sense that they propagate in space and time with a speed not exceeding that of light.

A century and a half after Charles Darwin put forward his *theory of evolution* which explains the origin of species, we believe that the stage is set for us to recount the creation and evolution of the universe itself. Everything we presently know indicates that the universe began in an extremely hot and dense state about 14 billion years ago. The conditions that existed in the universe when it was only one ten-billionth of a second old are routinely recreated in terrestrial laboratories using particle accelerators. The fact that our theories accurately describe particle interactions under these conditions means that we can retrace the evolution of the universe almost to its inception, and probably even predict its ultimate fate. However, the evidence of accelerated cosmic expansion presented in Section 3.9 suggests that either a fundamentally new theory of gravity is required, or that we do not know what constitutes most of the energy density of the universe (see Fig. 1.5).

## 1.2
## The Cosmological Hypothesis

The *Friedmann–Lemaître model* of cosmic evolution is based on the *cosmological hypothesis*, which states that the universe at large looks the same from any position within it. The differential equations describing the cosmic evolution were originally derived from Einstein's field equations by Russian mathematician and physicist Aleksandr Friedmann in 1922 [1]. Friedmann's cosmological model received crucial support seven years later when Edwin Hubble published his discovery that nearly all galaxies appear to be moving away from us, and that the farther a galaxy is, the faster it is receding [2]. The universe is expanding! This observation can be expressed in terms of the *theoretical Hubble law*, according to which the speed difference between two "nearby" points in space is directly proportional to their separation:

$$\text{Recession velocity} = \mathrm{H} \times \text{distance} \tag{1.1}$$

The coefficient of proportionality is a function of time: $\mathrm{H} = \mathrm{H}(t)$. For the correct general-relativistic interpretation of Hubble's law see Section 3.1.

The founder of modern physical cosmology was Georges Lemaître, who proposed his model of the evolving universe in 1927 [3] (at the time he was not aware of Friedmann's work). Lemaître was the first to consider the possibility that space expanded from a state of near-infinite density:

> *"The world has proceeded from the condense to the diffuse… We can conceive of space beginning with the primeval atom and the beginning of space being marked by the beginning of time…"* [4].

He also suggested that there ought to be some evidence for this "explosive" origin of the universe,[1] which Fred Hoyle later termed the Big Bang. An apparent relic from the early universe is the *cosmic microwave background radiation* (CMB) of temperature $\mathrm{T} \approx 3\,\mathrm{K}$, first detected by Arno Penzias and Robert Wilson in 1964 [5]. The CMB spectrum is very close to a *thermal* Planck distribution, which implies that the radiation has almost completely relaxed to

**1** Since the universe was initially much more isotropic and homogeneous than it is now (inhomogeneities grow due to gravity), the expansion could not have originated from a single center. That is, the birth of the universe was not "explosive" because, by definition, explosion is driven by a pressure gradient.

thermodynamic equilibrium (see Section 1.3). Clearly, this could not have happened in a universe that is transparent to radiation. As the detection of very remote radio sources indicates, the universe has been optically thin to radio waves through much of its history. These observations support the view that the CMB is a remnant from early epochs when the universe was sufficiently dense and hot for thermodynamic equilibrium to be established.

The universe appears to be remarkably *isotropic* and *homogeneous* at scales larger than $10^9$ light-years. By isotropic we mean that it looks the same in all directions. This notion relates to a quality of space rather than its matter content. If the universe is isotropic around every point, it is also homogeneous (has constant density). To see this, imagine a pair of intersecting spheres about two observers. The density on each sphere is constant by isotropy, and it must be the same constant since they intersect. In a universe that is both isotropic and homogeneous, any part of it is a representative of the whole. As Richard Feynman put it:

> *"It would be embarassing to find, after stating that we live on an ordinary planet about an ordinary star in an ordinary galaxy, that our place in the universe is extraordinary . . ."* [6].

If the assumption of large-scale *isotropy* and *homogeneity* of the universe (the cosmological hypothesis) is valid, then *every* observer will have the impression that all astronomical objects are receding from him. A homogeneous and isotropic universe does not have a center! At present there are no compelling empirical reasons to abandon this hypothesis. Indeed, (a) the galaxies seem to be distributed more or less isotropically around us (see Fig. 1.1), and to recede from us equally in all directions; (b) the temperature of the cosmic background radiation is found to be the same, to roughly one part in $10^5$, in every direction in the sky; and (c) the observed abundances of the light elements synthesized in the early universe – which range over nine orders of magnitude – limit the large-scale anisotropy to less than one part in $10^8$.

In our "immediate" vicinity (over regions of about $10^{10}$ light-years in diameter), the universe appears to be both isotropic and homogeneous. However, since we see remote regions of the universe as they were billions of years ago, we do not really know if the universe is homogeneous and isotropic when observed from other places or at other times. The cosmological hypothesis, therefore, is no more than a reasonable supposition, which has the great merit of keeping our calculations simple.

The cosmological hypothesis allows us to "smear out" any existing structures in the universe into an idealized fluid. We can imagine placing clocks at rest with respect to the expanding cosmic fluid and setting them to read the same reference time $t$ when the fluid density and temperature reach certain

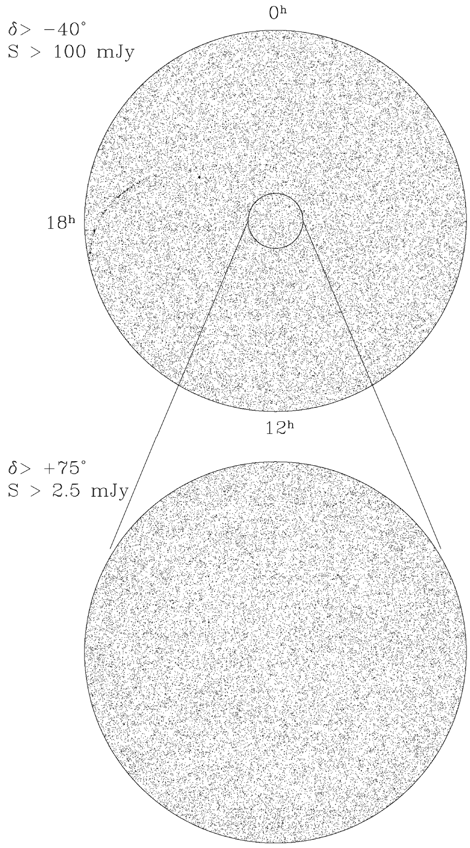


**Fig. 1.1** Angular distribution of some two million radio sources stronger than $S \simeq 2.5$ mJ ($1\,\text{J} = 10^{-26}\,\text{Wm}^{-2}\text{Hz}^{-1}$), including radio galaxies, quasars, ultraluminous starburst galaxies even at cosmological distances and low-luminosity active galactic nuclei [7]. The sources have been projected onto the plane of the figure in such a way that equal solid angles on the sky project to equal areas on the plane. The radial coordinate used in the map is defined as $d = (1 - \sin\delta)q^{1/2}$, where $\delta$ is the declination. The outer boundary corresponds to declination $\delta = 0$. Reproduced by permission, courtesy of J. J. Condon.

values. Using these synchronized clocks, the physical state of the universe will depend on the *cosmic time t* in the same way everywhere.

## 1.3 Thermal Radiation in an Expanding Universe

The idea of the thermodynamic history of an expanding universe was introduced in 1934 by Richard Tolman [8], who showed that the temperature of *thermal* (*blackbody*) *radiation* decreases during cosmic expansion.

The expansion of the universe, or any region within it, is assumed to be adiabatic, which means that no heat exchange is involved in the process. This assumption is valid provided that the expansion is reversible (see Section 1.5) and the universe is homogeneous throughout. Since irreversible processes may occur during the cosmic evolution (e.g., phase transitions that release latent heat and thus increase entropy), the adiabaticity condition is only an approximation.[2]

The *entropy* $\mathcal{S}$ of an adiabatically expanding system is conserved: $\mathrm{d}\mathcal{S} = 0$. We can therefore use the thermodynamic identity (C.60) to write

$$\mathrm{dE} = -\mathrm{P}\,\mathrm{dV} \tag{1.2}$$

where E is the total energy, P the pressure and $\mathrm{V} = 4\pi\mathrm{R}^3/3$ the volume of a typical spherical region in the universe. If we denote the total energy and mass densities by $\varepsilon$ and $\varrho$, respectively, then

$$\mathrm{E} = \varepsilon\,\mathrm{V} = \varrho c^2\,\mathrm{V} \tag{1.3}$$

(see (C.31)). Expression (1.2) now reads

$$\mathrm{d}\left(\varepsilon\,\mathrm{R}^3\right) = -\mathrm{P}\,\mathrm{d}\left(\mathrm{R}^3\right) \tag{1.4}$$

or equivalently

$$\mathrm{d}\left[\mathrm{R}^3\left(\varepsilon + \mathrm{P}\right)\right] = \mathrm{R}^3\mathrm{dP} \tag{1.5}$$

Setting $\mathrm{P} = \varepsilon/3$ for radiation and relativistic matter (see (C.50)), we obtain

$$\frac{\mathrm{d}\varepsilon}{\varepsilon} = -\frac{4}{3}\frac{\mathrm{d}\left(\mathrm{R}^3\right)}{\mathrm{R}^3} \tag{1.6}$$

Hence,

$$\varepsilon \propto \mathrm{R}^{-4} \qquad \text{relativistic matter, radiation} \tag{1.7}$$

where $\mathrm{R}(t)$ is the global *scale parameter of the expansion*.

**2** Our intuitive ideas about the thermodynamic properties of conventional systems may not apply to the universe as a whole. From the appearance of structures (order) in the universe we infer that, for a gravitational system, stable equilibrium states are not homogeneous. This does not contradict the laws of statistical physics, since only a system in stable external conditions can be expected, over a sufficiently long time, to reach a state of thermodynamic equilibrium. In GRT, the metric properties of spacetime may be regarded as variable "external conditions" to which large regions of the universe are subject [9]. Consequently, the application of the second law of thermodynamics does not necessarily imply that the universe ought to reach a state of thermodynamic equilibrium [10].

The relation between $\varepsilon_\gamma$ and the temperature of the radiation, $T_\gamma$, is obtained by integrating *Planck's formula* for the spectral energy density of thermal (blackbody) radiation:

$$\varepsilon_\gamma(T_\gamma) = \int_0^\infty \frac{8\pi h}{c^3} \frac{\nu^3 d\nu}{e^{h\nu/kT_\gamma} - 1} \tag{1.8}$$

(see Appendix C), which yields the *Stefan–Boltzmann law*

$$\varepsilon_\gamma(T_\gamma) = \frac{\pi^2}{15} \frac{k^4}{(\hbar c)^3} T_\gamma^4 \equiv a T_\gamma^4 \tag{1.9}$$

With $T_\gamma$ expressed in degrees kelvin, *Stefan's constant a* has the value

$$a \approx 7.6 \times 10^{-15} \frac{\text{erg}}{\text{cm}^3\,\text{K}^4} \approx 4.7 \times 10^{-3} \frac{\text{eV}}{\text{cm}^3\,\text{K}^4} \tag{1.10}$$

The numerical value of the *Boltzmann constant* k is given in (C.1).

Planck's formula depends on a single parameter, $T_\gamma$. This reflects the fact that the energy spectrum of photons in thermodynamic equilibrium is completely determined by its temperature. Thermodynamic equilibrium is the state of maximum uniformity and highest entropy, characterized by a unique temperature T throughout.

The number of photons per cubic centimeter is given by (see Eqs (C.10) and (C.68))

$$n_\gamma(T_\gamma) = \int_0^\infty \frac{d\varepsilon_\gamma}{h\nu} \approx 0.243 \left(\frac{kT_\gamma}{\hbar c}\right)^3 \approx 20\, T_\gamma^3 \tag{1.11}$$

The present temperature of the cosmic microwave background radiation (CMB) is $T_0 \approx 2.73\,\text{K}$ (see Section 1.4). Hence,

$$\varepsilon_{0\gamma} \approx 0.26 \frac{\text{eV}}{\text{cm}^3} \qquad n_{0\gamma} \approx 410 \frac{\text{photons}}{\text{cm}^3} \tag{1.12}$$

(see Eq. (C.69)). The average photon energy

$$E_\gamma \equiv \frac{\varepsilon_\gamma}{n_\gamma} \approx 2.70\, kT_\gamma \Rightarrow E_{0\gamma} \approx 6.3 \times 10^{-4}\,\text{eV} \tag{1.13}$$

is comparable to the kinetic energy of rotation for a small molecule, such as CN or $H_2O$. The fossil CMB makes up about one percent of the static noise detected by a home TV antenna. The peak of the radiation spectrum is at a frequency $\nu_{\text{peak}} = 160.4\,\text{GHz}$ ($\lambda_{\text{peak}} = 1.9\,\text{mm}$).

From (1.3), (1.7) and (1.9) it follows that

$$E_\gamma, T_\gamma \propto R^{-1} \Rightarrow \lambda_\gamma \propto R \tag{1.14}$$

As the universe expands, the energy and temperature of the radiation fall in inverse proportion to R. Since $\lambda_\gamma \propto \mathrm{R}$, the expansion of the universe produces a *redshift*, that is, the wavelength of light stretches along with the expansion. Consider a light signal emitted from a source in a distant galaxy. *Redshift is a measure of the scale parameter of the universe when the signal was emitted by the source* (see (1.55)). The more distant a galaxy, the higher its redshift. The redshift of an object is directly related to the *cosmic time* in which that object exists. For the correct general-relativistic interpretation of (1.14) see Section 3.1.

The law $\nu \propto \mathrm{R}^{-1}$ applies to the momentum of any particle, relativistic or not. To see this, consider the de Broglie wavelength $\lambda = h/p$. This quantity grows with the global scale parameter of the expansion, *as if* the de Broglie waves were standing waves trapped in a box that expands with the universe.

It is easy to show that freely expanding blackbody radiation retains the Planck spectrum, but with a temperature that drops in inverse proportion to the scale of the expansion. Suppose the size of the universe changes by a factor $\eta$. With $\tilde{\nu} = \nu/\eta$ and $\widetilde{\mathrm{T}}_\gamma = \mathrm{T}_\gamma/\eta$, it follows from (1.7) and (1.8) that the radiation energy density

$$\mathrm{d}\tilde{\mathcal{E}}_\gamma = \frac{\mathrm{d}\mathcal{E}_\gamma}{\eta^4} \propto \left(\frac{\nu}{\eta}\right)^3 \frac{\mathrm{d}(\nu/\eta)}{\mathrm{e}^{h\nu/\mathrm{kT}_\gamma} - 1} = \frac{\tilde{\nu}^3 \mathrm{d}\tilde{\nu}}{\mathrm{e}^{h\tilde{\nu}/\mathrm{k}\widetilde{\mathrm{T}}_\gamma} - 1} \tag{1.15}$$

The existence of a low-temperature blackbody radiation produced by the redshifted light left over from the early universe was predicted by Ralph Alpher and Robert Herman in 1948 [11]. Their prediction is an outgrowth of George Gamow's seminal work on primordial nucleosynthesis (see Section 1.11). Gamow recognized that the early universe must have been very hot,[3] for otherwise the formation of elements would not have proceeded fast enough to offset the rapid decrease in density caused by cosmic expansion. If, on the other hand, the nucleosynthesis had proceeded too fast, the lightest nuclei would have fused into heavier nuclei, in contradiction with the observed abundance of elements. This could have been prevented only if, at the epoch of nucleosynthesis, the universe was filled with intense, hot radiation that would photodisintegrate nuclei as fast as they could be formed. As the universe expanded and cooled, the element formation eventually stopped, leaving behind matter composed of light nuclei. These nuclei later combined with electrons to form atoms. At that point the radiation temperature was so low that the photons had too little energy to be absorbed by atoms, and were at last free to follow their own thermal history. This happened some four hundred thousand years after the Big Bang, when the universe was about 1/1000th its

**3** When the average density of matter in the universe was comparable to that of air at sea level (about 1/800th the density of water), its temperature was 2.73 billion degrees! The average density today is about one proton per cubic meter.

present size. As we have seen, freely expanding thermal radiation retains the Planck spectrum, but with a temperature that drops in inverse proportion to $R(t)$. Alpher and Herman found that, in order to account for the observed abundances of light elements, they had to assume a ratio of photon to nuclear densities of about $10^9$. Based on estimates of the present cosmic density of nuclear particles, they predicted that the temperature of this fossil radiation would have dropped by now to a few degrees kelvin (see Section 1.10).

To see why the CMB might be expected to have a thermal spectrum, consider the adiabatic expansion of a system comprising particles and heat quanta. If the mean number density of a particle species is n, then the ratio of its heat capacity to that of the CMB is given by (C.98):

$$\frac{C_V^{\text{mat.}}}{C_V^{\text{rad.}}} = \frac{1.5nk}{4aT_\gamma^3} \tag{1.16}$$

Since $T_\gamma \propto R^{-1}$ and $n_m \propto R^{-3}$ (see (1.21)), the ratio $C_V^{\text{mat.}}/C_V^{\text{rad.}}$ remains constant in the process of adiabatic expansion or contraction as long as the number of particles is conserved.

At temperatures $T > 10^{10}$ K ($E_\gamma > 3$ MeV), photons were in equilibrium with electrons and positrons through the processes of pair production and particle–antiparticle annihilation (see Section 1.6). Before the recombination of the $e^+e^-$ pairs, when their heat capacity was comparable to that of the CMB, the electron-positron plasma, the nucleons and the CMB were tightly thermally coupled, to one temperature $T(t)$. As the expanding universe cooled, the coupling between matter and radiation became weaker. Consequently, the temperature of radiation dropped in inverse proportion to R, and that of nonrelativistic matter fell as $R^{-2}$ (see (1.20)). As matter consists mainly of atomic hydrogen, the ratio of its heat capacity to that of the CMB is about $10^{-10}$ (see (1.83)). Because of its enormous heat capacity, the radiation temperature closely follows the $R^{-1}$law. This explains why the CMB spectrum is nearly thermal [12, 13].

## 1.4
## The Discovery and Properties of CMB

In 1964, Arno Penzias and Robert Wilson detected a constant, low-level radio noise with a wavelength $\lambda = 7.35$ cm in an antenna built for satellite communication [5]. The detected signal did not seem to be of galactic origin, as its intensity did not change appreciably in the direction of the nearby Andromeda galaxy nor with time, and so it could not have come from the center of our galaxy. Indeed, if the signal detected by Penzias and Wilson came from the center of the Milky Way, its intensity would peak once per *sidereal day*, when the galactic center was high in the sky (see Appendix A.25). The signal also did not vary with the altitude above the horizon (i.e., with the thickness of the atmosphere), which meant that it could not have originated in the Earth's atmosphere. From these observations they inferred that the universe was uniformly filled with a cosmic background radiation of temperature $T = 3.5\,K$, which was isotropic and unpolarized within their measurement precision.[4] Subsequent measurements in the wavelength range 0.03 cm to 75 cm have shown that the CMB possesses a spectrum of $2.725 \pm 0.002\,K$ blackbody radiation [14] (see Fig. 1.2). The CMB temperature has been determined to a precision of 0.1%, which makes it the best known cosmological parameter.[5]

Recall that CMB photons interacted with matter for the last time some 400 000 years after the Big Bang, when electrons and protons combined to form hydrogen atoms. The universe was an opaque "fog" of free electrons before this *recombination epoch*, and became transparent to the cosmic microwave background radiation afterwards. Thus, when we look at the sky in any direction, we can expect to see photons that originate from a *last-scattering surface* (or *cosmic photosphere*) at a redshift corresponding to $t \approx 4 \times 10^5$ years after the Big Bang. Similarly, when we look at the surface of the Sun, we observe photons last scattered by the hot plasma of its photosphere, the temperature of which is roughly equal to that of the universe at the epoch of recombination. Since there are about $10^9$ photons per nucleon in the universe, the transition from the ionized primordial plasma to neutral atoms did not significantly alter the CMB spectrum. The large photon-to-nucleon ratio also implies that it is very unlikely for the CMB to be produced in astrophysical processes such

**4** Radio astronomers customarily describe the intensity of radio noise detected at a given wavelength in terms of the temperature of the walls of an opaque box within which the radio noise would have the observed intensity. Any body at a temperature above absolute zero emits radio noise, produced by the thermal motions of electrons within the body. Inside a box with opaque walls, the intensity of radio noise at a given wavelength depends only on the temperature of the walls: the higher the temperature, the more intense the noise. In a sense, the antenna used by Penzias and Wilson was in a box, if we think of the universe as a huge cavity filled with thermal radiation.

**5** It is difficult to measure a radiation temperature of a few degrees kelvin because the signal is much weaker than the electrical noise in the amplifier circuits of the receiver. Suppose a receiver is sensitive to radiation with wavelength $\lambda = 2 - 5\,\text{mm}$, corresponding to a bandwidth $f \approx 10\,\text{GHz} = 10^{10}\,\text{Hz}$. The power $\mathcal{P}$ into the amplifier from the antenna is $\mathcal{P} = kTf \approx 1.4 \times 10^{-23}(\text{J/K}) \times 3\,\text{K} \times 10^{10}\,\text{Hz} \approx 0.4\,\text{pW}$. Penzias and Wilson, who intended to measure the intensity of the radio waves emitted out of the plane of our galaxy, compared the power coming from the antenna with the power produced by an artificial source cooled with liquid helium to $T \approx 4\,\text{K}$. The electrical noise in the amplifier circuits is the same in both cases, and therefore cancels out in the comparison.

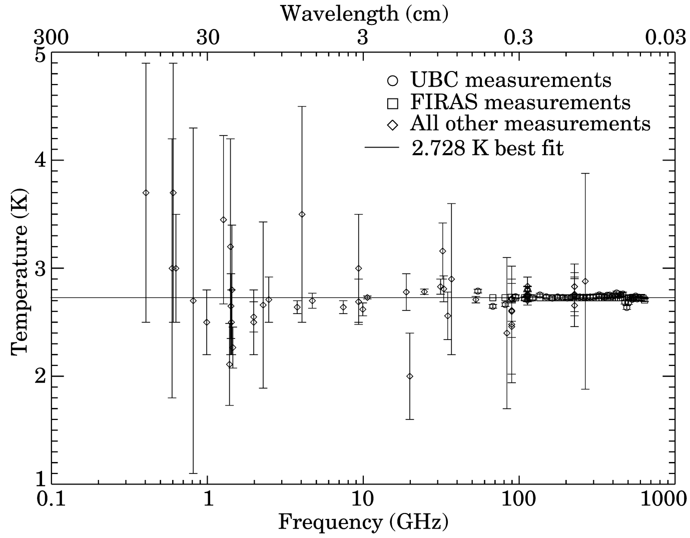


**Fig. 1.2** Observed CMB temperature as a function of frequency [15].

as the absorption and re-emission of starlight by cold dust, or the absorption or emission by plasmas.

Before the recombination epoch, Compton scattering tightly coupled CMB photons to electrons, which in turn coupled to protons via electomagnetic interactions. As a consequence, CMB photons and nucleons in the early universe behaved as a single "photon–nucleon fluid" in a gravitational potential well created by primeval variations in the density of matter. Outward pressure from CMB photons, acting against the inward force of gravity, set up *acoustic oscillations* that propagated through the photon–nucleon fluid, exactly like sound waves in air. The frequencies of these oscillations are now seen imprinted on the CMB temperature fluctuations. Gravity, an attractive force, caused the primordial density perturbations across the universe to grow with time. The temperature anisotropies in the CMB are interpreted as a snapshot of the early stages of this growth, which eventually resulted in the formation of galaxies.

CMB photons can propagate for billions of years through the tenuous intergalactic medium before reaching the Earth. Microwaves with wavelengths shorter than $\lambda \simeq 3\,\text{cm}$ are strongly absorbed by water molecules in the Earth's atmosphere. At $\lambda \gtrsim 30\,\text{cm}$, the CMB is swamped by radiation from interstellar gas within our own galaxy, the Milky Way. Thus, from sea level on Earth, the CMB can be detected only within the limited wavelength range $3\,\text{cm} \lesssim \lambda \lesssim 30\,\text{cm}$. The spectrum of the CMB, which peaks at $\lambda \approx 2\,\text{mm}$, has been measured at short wavelengths by detectors carried on satellites, sound-

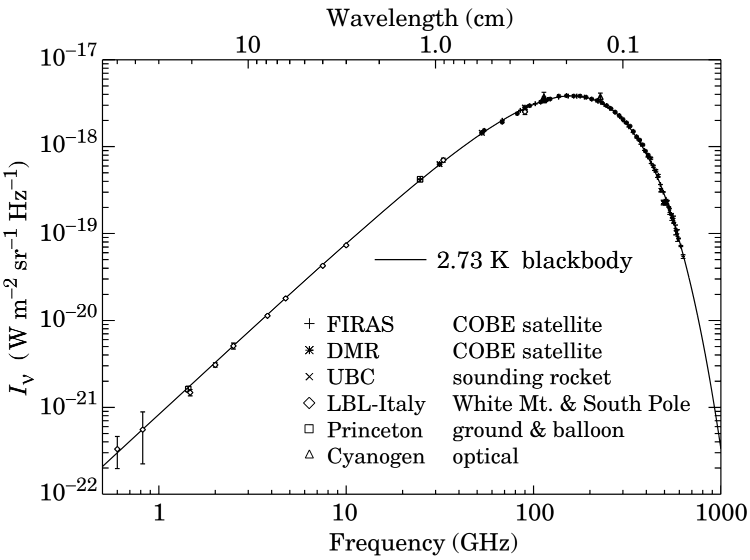


**Fig. 1.3** Precise measurements of the CMB spectrum [15]. Radio waves with wavelengths of up to one meter are known as "microwave radiation". The observed CMB spectrum is much closer to that of a perfect blackbody than the spectrum of any man-made source.

ing rockets and stratospheric balloons, and by larger instruments deployed at mountain altitudes and on the South Pole.

Some of the most accurate determinations of the frequency spectrum and temperature fluctuations of the CMB come from: the *Cosmic Background Explorer* (COBE) [16, 17], a satellite experiment that took data for four years after its launch in 1989; the balloon-borne microwave telescopes *BOOMERANG* (which completed an 8000 km flight over Antarctica in 1999) [18] and *MAXIMA* (which flew twice, in 1998 and 1999) [19]; and the *Wilkinson Microwave Anisotropy Probe* (WMAP), a spacecraft carrying a pair of back-to-back telescopes (launched in 2001). The CMB spectrum agrees with that of a blackbody to within 50 parts per million (see Fig. 1.3). This agreement indicates that the early universe was once in thermodynamic equilibrium.

The COBE *FIRAS* instrument was a cryogenically cooled Michelson interferometer that used very sensitive thermal detectors called *bolometers*, and was calibrated to within 0.001 K by an external blackbody (XCAL). The instrument was designed to measure precisely the spectrum of the CMB over a wavelength range from 0.1 to 10 mm. As the COBE orbited the Earth at an altitude of 900 km once every 103 minutes, it viewed a circle on the sky about 90 degrees away from the Sun, and gradually scanned the entire sky over the course of the year. The FIRAS instrument, which was aligned with the satellite's spin axis, measured the *difference* betwen the CMB and the XCAL blackbody spectrum.

The *DMR* detector on board COBE had a pair of separate antennas that pointed at two different regions of the sky. The signals received by the antennas differed slightly if the two regions were not equally bright. DMR was the first instrument to detect tiny (tens of $\mu$K) variations in the temperature of the CMB across the sky. These fluctuations arise from primeval lumpiness in the distribution of matter, as explained earlier.

BOOMERANG and MAXIMA were focal-plane arrays of cryogenic bolometers on small, offaxis telescopes suspended from stratospheric balloons. Millimeter-wave radiation was absorbed and measured as a minute temperature rise in each bolometer (a micromachined mesh of silicon nitride) by a tiny germanium *thermistor* (temperature-dependent semiconductor resistor). The bolometers were kept at a fraction of a degree above absolute zero in order to achieve high sensitivity.

To eliminate spurious signals, WMAP measures temperature differences between two points in the sky using microwave receivers coupled to back-to-back telescopes. By measuring the CMB at five different frequencies (from 23 to 94 GHz), WMAP can also subtract the foreground radiation that could be confused with CMB anisotropy: synchrotron radiation from electrons orbiting in magnetic fields, radiation from hot ionized gas, and thermal radiation from interstellar dust. A projection over the full sky of the CMB radiation measured by WMAP is shown in Fig. 1.4.

If the intensity of the radiation incident on an electron is anisotropic, the *Thomson scattering* process converts some of that anisotropy into *polarization* of the scattered photon: an electron irradiated by unpolarized light with anisotropic intensity emits linearly polarized light. The degree of linear polarization is directly related to the quadrupolar anisotropy in the photons when they last scatter [21] (in a quadrupolar anisotropy, intensity maxima and minima are separated by 90 degrees). Thus, the CMB spectrum today bears a unique imprint of the state of the universe at the recombination epoch, when matter and radiation essentially decoupled from each other. Based on the observed temperature fluctuations, one can predict the level of polarization of the CMB with essentially no free parameters. Variations in the level of CMB polarization were first measured by the DASI interferometric array at the Amundsen–Scott South Pole Station, and were found to be in excellent agreement with the predictions of the standard cosmological model [22, 23]. DASI, which operates at centimeter wavelengths, consists of thirteen 20 cm-diameter telescopes with separations spanning 25–121 cm. The effective response for each of the $13(13-1)/2 = 78$ independent pairs of telescopes is similar to a two-slit interference pattern. The detected polarization signal was about ten times fainter than that due to the temperature fluctuations of the

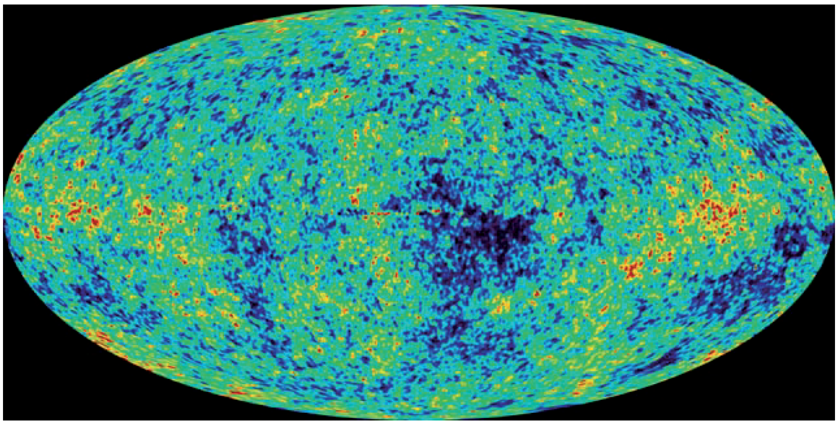

**Fig. 1.4** The imprint of primordial seeds (matter density fluctuations) on the cosmic microwave background as seen by WMAP [20]. The oval shape of this "map" is a projection to display the whole sky, similar to the way in which the globe of the Earth can be projected as an oval. The colors represent temperature variations, with red (blue) indicating regions that are warmer (cooler) than the average temperature of 2.73 K.

CMB. This is to be expected, since only for those photons that last scattered in an optically thin region (i.e, at the very last instants of the decoupling process) was the polarization not washed out by subsequent rescattering [21].

Astronomers describe the fluctuations in the CMB by its *angular power spectrum*, a plot of the strength of the fluctuations versus their angular size. The precise shape of the angular power spectrum of the CMB has been measured, for instance, by the CBI microwave telescope array [24, 25]. The instrument is located at an altitude of 5 km in northern Chile, and operates at frequencies from 26 to 36 GHZ. The angular resolution of the CBI telescope is sufficiently high to discern the primeval clumps of matter, which eventually evolved to become the clusters of galaxies we see today. The CBI is also capable of observing the *Sunyaev–Zeldovich scattering* of CMB photons by the hot electrons in clusters of galaxies. Measurements of this effect can be used to study the evolution of clusters and to determine the Hubble parameter (see Chapter 3).

At the epoch of recombination, the universe was filled with a red, uniformly bright glow of blackbody radiation. As the universe continued to expand, the spectrum of cosmic background radiation shifted to the infrared. To the human eye, the universe would then have appeared completely dark. A few hundred million years after the Big Bang, the ultraviolet radiation from the first stars and quasi-stellar objects (*quasars*) re-ionized the intergalactic medium, heating it well above the temperature of the CMB. We know that most of the intergalactic medium is ionized because astronomers have observed only small "islands" of neutral hydrogen. These islands show up as absorption lines in the spectra of very distant quasars. Observations of several such spectra indicate that the *re-ionization epoch* ended at a redshift $z \approx 6$ [26, 27].

The sum of two blackbody spectra of different temperatures does not result in a blackbody spectrum. Thus, *any* deviation from a perfectly homogeneous and isotropic universe causes a spectral distortion. The remarkable precision with which the CMB spectrum is fitted by a Planckian distribution, and the fact that the cosmic microwave background accounts for more than 99% of the radiant energy in the universe (see below), set a limit on possible spectral distortions at roughly the fractional level of $10^{-4}$ of the CMB energy for $t \gtrsim 1$ year after the Big Bang [28].

Recent observations using Earth-based telescopes have provided the first direct evidence that the CMB radiation has decreased over cosmic time. The cosmic background radiation will excite the fine-structure levels of the ground state of certain atoms and molecules when the energy separation of the levels is similar to the CMB peak frequency. The relative populations of excited levels are determined from the absorption lines seen in the spectra of distant quasars. Based on this information, one can calculate the CMB temperature in a gas cloud at high redshift. By detecting absorption from the first fine-structure level of neutral carbon atoms in an intergalactic cloud along the line-of-sight to a quasar, astronomers have found the excitation temperature of this atomic level to be $7.4 \pm 0.8$ K [29]. The measured temperature agrees nicely with 7.58 K, the value expected for the CMB radiation at the observed redshift ($z \simeq 1.78$). For a similar measurement of the cosmic background temperature at $z = 3.025$ see [30].[6]

The first measurement of the CMB temperature was in fact made using this method, although it was not recognized as such until after the detection of the CMB by Penzias and Wilson. In 1941, Walter Adams observed narrow absorption lines in the spectrum of a star in the constellation Ophiuchus ("the serpent bearer"), which originated in an interstellar gas cloud between the Earth and the star [32]. The absorption lines were identified in the same year by Andrew McKellar as being due to the diatomic molecules CH and CN [33]. *Cyanogen* (CN) has a visible absorption line at 3874 Å, corresponding to transitions from the ground state to a vibrationally excited state. Both states are split into rotational energy levels, distinguished by the angular momentum J. McKellar noticed that the cyanogen molecules in the interstellar cloud were

**6** Measurements of this kind are affected by substantial systematic uncertainties due to the unknown physical conditions in the absorbing clouds and the presence of additional sources of excitation (e.g., collisional excitation and fluorescence induced by the local ultraviolet radiation field). The CMB temperature at different redshifts has also been inferred from measurements of the Sunyaev–Zeldovich effect at radio and microwave frequencies [31].

absorbing light not only in the dipole ($|\Delta J| = 0$) transition from the $J = 0$ ground state, but also in the transition from the first excited state with $J = 1$, which is at an excitation energy corresponding to a wavelength of 2.64 mm. From the relative strength of the two absorption lines, he could infer the relative number of CN molecules in the two rotational states. Then, by assuming that the molecules in the rotational ground state were in statistical equilibrium with those in the first rotationally excited state, he computed the temperature of the system to be 2.3 K. McKellar did not speculate as to what the source of the rotational excitation might be, but merely stated: *"It can be calculated that the 'rotational' temperature of interstellar space is 2 K."*

Cyanogen excitation has been used to measure the *present* CMB temperature very accurately: $T_{\text{CMB}} = 2.729^{+0.023}_{-0.031}$ [34]. This provides an important calibration point, since it determines the present temperature of the cosmic microwave background radiation far from the Solar System [35].

The evidence for the cosmic background radiation represents one of the most important discoveries in cosmology. The CMB carries valuable information about the properties of our universe that cannot be obtained in any other way. Indeed, the CMB radiation has traveled through the entire observable universe, and so its appearance reflects the expansion history and overall geometry of the universe. We discuss this subject in more detail in Section 1.18.

**The Dark Sky Paradox**

Why is the sky dark at night if the universe is uniformly filled with stars? Naïvely, one would expect the night sky to be as bright per unit area as the Sun's surface. To see this, consider any large spherical shell enclosing the Earth at its center. The amount of light produced by stars within this shell can be readily calculated. Now consider a shell of twice the radius. It contains four times as many stars, but they are on the average only one-quarter as bright as the stars within the first shell; hence, their contribution to the brightness of the night sky is about the same. As we consider larger and larger shells, the amount of starlight continues to increase without limit. However, we must allow for the fact that the intervening stars may intercept radiation from more distant stars. This effect reduces the inferred sky brightness to that at the surface of an average star. If we take our Sun to be such an average star, then the night sky should be as bright per unit area as the Sun's surface. This translates into a total brightness for the night sky of nearly 100 000 suns [36]. Yet the night sky is very dark (apart from the light emitted by the Milky Way), which leads to an apparent contradiction. This paradox, clearly stated by the astronomer Heinrich Wilhelm Olbers in 1823, was already known to Johannes Kepler (1571–1630) (see [37] and references therein).

The resolution of the paradox is that *stars do not shine for long enough to light up the entire sky* [37]. It turns out that the expansion of the universe, which increases the volume of space and redshifts the light, reduces the intensity of intergalactic radiation only by a relatively modest amount [37, 38]. The darkness of the night sky is *not* due to: (a) absorption of starlight; (b) clustering of stars into galaxies; or (c) the finite size of the universe, just to mention a few of the many explanations offered over the past four centuries. Of course, the night sky is not completely dark: it is filled with light from all the stars and galaxies, and there is also the cosmic microwave radiation. This extragalctic light is the brightest in the microwave range. The peak intensity of CMB is about 385 $\mathrm{MJy\,sr^{-1}}$ ($1\,\mathrm{Jy} = 10^{-26}\,\mathrm{Wm^{-2}\,Hz^{-1}}$), or approximately $4\times10^{-15}\,\mathrm{erg\,cm^{-2}s^{-1}sr^{-1}\,Hz^{-1}}$.

The energy density of the CMB, $\varepsilon_{0\gamma} \approx 4\times10^{-13}\mathrm{erg\,cm^{-3}} \approx 0.26\,\mathrm{eVcm^{-3}}$, is much larger than the energy density of all the photons emitted by all the stars throughout the history of the universe. The present luminosity density of galaxies in our immediate neighborhood (within $10^{21}$ km or so) is about $3\times10^{-32}\mathrm{erg\,s^{-1}cm^{-3}}$. As a very rough estimate, let us assume that galaxies have been radiating photons at this rate for the entire age of the universe, $t_0 \approx 14\times10^9\,\mathrm{yr} \approx 4.2\times10^{17}\mathrm{s}$. This gives $1.1\times10^{-14}\,\mathrm{erg\,cm^{-3}} \approx 0.007\,\mathrm{eVcm^{-3}}$ for the present energy density of starlight, or about 3% of the energy density in the CMB. We can, therefore, approximate the energy density of photons in the universe with that of the CMB, and assume that photons are neither created nor destroyed.

## 1.5
## Thermodynamic Equations for Matter

If the number density of particles with mass $m$ (protons, neutrons, electrons, etc.) is $\mathrm{n}_m$, then their pressure is (see (C.48) and (C.50))

$$\mathrm{P}_m = \mathrm{n}_m \mathrm{kT}_m \qquad \text{ideal gas} \qquad (1.17)$$

$$\mathrm{P}_r = \varrho_r c^2/3 \qquad \text{relativistic matter, radiation} \qquad (1.18)$$

The energy density of nonrelativistic matter is given by (see C.49)

$$\varepsilon_m = \underbrace{\mathrm{n}_m mc^2}_{\text{rest energy density}} + \underbrace{\mathrm{n}_m 3\,\mathrm{kT}_m/2}_{\text{kinetic energy density}} \approx \varrho_m c^2 \qquad (1.19)$$

where $\varrho = \mathrm{n}m$ is the mass density (see (C.72)). According to (1.17) and (1.18), nonrelativistic matter exerts a negligible pressure ($mc^2 \gg \mathrm{kT} \Rightarrow \mathrm{P} = n\mathrm{kT} \ll \varrho c^2$), whereas ultrarelativistic particles (including photons and neutrinos) exert a pressure that is proportional to their energy density ($\mathrm{P} = \varrho c^2/3$). Note that $T_m$ is *not* the temperature of matter in thermodynamic equilibrium with radiation.

Setting $N \equiv n_m V$ and substituting (1.17) and (1.19) in (1.2), we obtain

$$\left(\frac{3}{2}\,kT_m + mc^2\right) dN + \frac{3}{2}Nk\, dT_m = -\frac{N}{V}\,kT_m dV$$

If we assume that the number of particles is conserved, this becomes

$$\frac{3}{2}\frac{dT_m}{T_m} = -\frac{dV}{V}$$

from which we infer that

$$T_m \propto R^{-2} \qquad \text{nonrelativistic matter} \tag{1.20}$$

Since $N = \text{const.}$,

$$d\left(\frac{4\pi}{3}\,R^3 n_m\right) = 0 \Rightarrow n_m \propto R^{-3}$$

For nonrelativistic particles, $\varrho_m = n_m m$. Thus,

$$\varrho_m \propto R^{-3} \qquad \text{nonrelativistic matter} \tag{1.21}$$

We note that (1.20) and (1.21) were derived under the following assumptions: (a) thermodynamic equations for radiation and nonrelativistic matter can be decoupled; and (b) the number of particles is constant over a sufficiently long period of time. Therefore, if the radiation and matter are decoupled, they will cool at different rates due to their different adiabatic exponents:

$$T \propto R^{-3(\gamma-1)} \qquad \gamma = \begin{cases} 4/3 & \text{radiation} \\ 5/3 & \text{nonrelativistic particles} \end{cases} \tag{1.22}$$

where $\gamma$ is the ratio of heat capacities at constant pressure and volume (see (C.103) and (C.104)).

We infer from (1.14) and (1.20) that

$$T_r \geq T_m \tag{1.23}$$

which means that, in an expanding universe, nonrelativistic matter and radiation cannot be in thermodynamic equilibrium at a cosmic timescale. The Hubble expansion ensures that the universe as a whole is not in a state of complete thermodynamic equilibrium, only a partial one. This prevents its otherwise imminent *thermal death*.

One of the basic assumptions about the early universe is that it was so dense and hot that interactions, which are now far from equilibrium, were in equilibrium at that time. They were driven out of equilibrium, at one stage or another, by cosmic expansion. As the expanding universe cooled and became less dense, the interaction rates eventually fell below the expansion rate. For much of its early history, however, the universe was very nearly in thermodynamic equilibrium. As long as particle interaction rates were higher than the rate of expansion, the universe was evolving through a series of nearly thermal states. Successive departures from equilibrium were essential for the evolution of the universe into its present state; without them the past history of the universe would be irrelevant.

Suppose the universe is in thermodynamic equilibrium at time $t_0$. If the expansion rate at a later time is faster than the characteristic thermalization rate between matter and radiation, the two entities will have different temperatures: $\mathrm{T}_r \neq \mathrm{T}_m$. Using (C.85), we can express the change of *entropy* per unit time as

$$\frac{\mathrm{d}\mathcal{S}}{\mathrm{d}t} = \left(\frac{1}{\mathrm{T}_m} - \frac{1}{\mathrm{T}_r}\right)\frac{đQ}{\mathrm{d}t} \tag{1.24}$$

where $\mathcal{S} = \mathcal{S}_m + \mathcal{S}_r$ and $đQ$ is the amount of heat exchanged between matter and radiation: $đQ_m = -đQ_r = đQ$. If the expansion rate is very fast compared with the thermalization rate, then $đQ/\mathrm{d}t \approx 0$; if it is very slow, $\mathrm{T}_r \approx \mathrm{T}_m$.

## 1.6 Universe in Transition

*"The changing of bodies into light, and light into bodies, is very conformable to the course of nature, which seems delighted with transmutations."*

Isaac Newton

When the temperature of the dense cosmic plasma was so high that the radiation energy exceeded $2mc^2$, where $m$ is the rest mass of a charged particle, photons could readily convert into pairs of these particles and their antiparticles. Below that temperature, particle–antiparticle annihilations into photons were not compensated by the pair production. Thus, depending on the temperature, the early universe was populated by different kinds of elementary particles at different times.

Photons were in equilibrium with electrons and positrons through the processes of pair production and particle–antiparticle annihilation,

$$e^+ + e^- \leftrightarrow \gamma + \gamma$$

as long as the photon energy was larger than the rest mass of an electron–positron pair (the electron rest mass is $m_e \approx 0.5 \times 10^6\,\mathrm{eV} \approx 5.8 \times 10^9\,\mathrm{K}$;