Chapter Four: The Primordial Fireball

CHAPTER FOUR: The Primordial Fireball

4.1 The energy density of radiation in an expanding world-model

The observations of H_0, q_0 and ρ_0 discussed so far do not yet tell us whether there is a non-zero cosmological constant, or give the value of q_0 to within ~0.5. We must therefore seek other evidence for whether or not there was a 'Big Bang' with $R(t)=0$ in the past of our Universe. The strongest evidence for such a Big Bang comes from the microwave black-body background radiation, referred to earlier in Sec.1.1.4. To understand the importance of this background, consider first the behaviour of radiation in an expanding world-model.

If the Universe is filled with photons of energy density u_r joules/m^3 we can define an associated radiation density ρ_r through

$$u_r = \rho_r c^2$$

This energy density is associated with a radiation pressure p_r such that

$$p_r = u_r/3$$

so that if the photon flux is adiabatically expanded we can write

$$d(u_r V) = -p_r dV = -(1/3)u_r dV$$

$$u_r dV + V du_r = -(1/3)u_r dV$$

$-(4/3)(dV/V) = du_r/u_r$, from which

$$\rho_r(t) \sim u_r(t) \sim R^{-4}(t) \quad (4.1)$$

In a Universe where both mass and photons are conserved the radiation density ρ_r therefore falls off faster with t than does the matter density ρ_m, which varies as $R^{-3}(t)$ by equation (1.4). The early phases of any expanding model with a singularity $R(t) = 0$ must therefore be radiation-dominated, in the sense $\rho_r >> \rho_m$, unless the Universal photon flux in that model is strictly zero. This is a property of 'Big-Bang' models with radiation that has been known since the early work of Gamow and Lemaitre. Furthermore, any such model must spend some time with ρ_m sufficiently large that the opacity of the model will bring matter and radiation towards thermal equilibrium; if the time spent in this opaque state is sufficiently long, the radiation will acquire a black-body spectrum and both it and the matter will be characterisable by a temperature $T(t)$ where the number of photons $dN(t)$ in volume $V(t)$ with frequencies between ν and $\nu+d\nu$ is given by

$$dN(t) = 8\pi\nu^2 V(t)d\nu/c^3\{\exp(h\nu/kT(t)) - 1\}$$

$a = \frac{8\pi^5k^4}{15c^3h^3} = 7.564\times10^{-16}\ J.m^{-3}.K^{-4}$

and the energy density

$$u_r(t) = \int_{\nu=0}^{\nu=\infty} h\nu dN(t) = 8\pi^5k^4T^4(t)/15c^3h^3 = aT^4(t)$$

The constant a is known as the radiation density constant (= $4\sigma/c$ where σ is Stefan's radiation constant).

Models which achieve thermal equilibrium between matter and radiation at an early stage are known as 'hot' Big-Bang, or 'fireball', models. Evidently as $u_r(t) \sim R^{-4}(t)$ the temperature of these models varies as

$T(t) \sim R^{-1}(t)$, i.e.

(4.2) $\quad T(z) = T_0(1+z)$

Note that this is the same as the relation for the observed bolometric brightness temperature of a distant black-body source, derived in Sec. 2.4.1.

Gamow and his coworkers Alpher and Herman, while considering the problem of element synthesis by nuclear fusion in the 'hot Big Bang', realised that a radiative remnant of the dense past might still be observable as an isotropic black-body background with some low temperature T_0. From a number of assumptions about the mechanisms of element production and the conditions early in the expansion (i.e. at high z), they attempted to explain the origin of the elements and predicted a presently-observable black-body temperature T_0 of 5-28 K (Alpher and Herman, Phys. Rev., 73, 1089 (1949) and Rev. Mod. Phys., 22, 153 (1950), Phys. Rev., 84, 60 (1951)). A black-body spectrum at such a low temperature would peak in the far-infrared, and have most of its energy in the microwave region of the spectrum. There were no serious attempts to detect such a background in the 1950s - it would have been below the levels of sensitivity of microwave and infrared detectors that were then available. The popularity at that time of the Steady-State Theory (resulting at least in part from the early over-estimation of the Hubble parameter) may also have discouraged work towards detecting such a black-body background.

4.2 The discovery of the 2.7-K radiation

The first deliberate attempt to detect a black-body background at microwave frequencies was that of Roll and Wilkinson at Princeton, encouraged by the theorist Robert Dicke. While their apparatus was under construction in 1965 they learned that Penzias and Wilson, at the Bell Telephone Laboratories in Holmdel, New Jersey, had already detected an apparently isotropic 'excess' background temperature of (3.5 ± 1) K during observations of the nonthermal galactic (Milky-Way) background at 4080 MHz (λ7.4cm) with a 20-ft horn antenna originally built for satellite communications. The experimental results from the Holmdel antenna were then published by Penzias and Wilson in Ap.J., 142, 419 (1965) together with an interpretation of the excess radiation as the red-shifted 'primordial fireball' by Dicke, Peebles, Roll and Wilkinson in Ap.J., 142, 414 (1965). Roll and Wilkinson subsequently carried out their own measurements at λ3.2cm (Phys. Rev. Letters, 16, 405 (1966); Ann. Phys., 44, 289 (1967)) and obtained an 'excess' brightness temperature of (3.0 ± 0.5) K. The history of these experiments, and a description of the radio measurement techniques that they used, is given by Wilson in his 1978 Nobel Prize Lecture (Science, 205, 866 (1979)).

4.2.1 Properties of the 2.7 K radiation

The 'excess' temperature is the brightness temperature measured by a radiometer after all known discrete sources of radiation have been accounted for; the raw data must be corrected for radiation from the Earth, the atmosphere, the Milky Way, and discrete sources such as bright radio galaxies and quasars, the Sun, etc. The effects of losses in the receiving antenna must also be allowed for; an imperfection in the antenna can simulate a 'grey-body' radio source which has a fraction of the black-body emissivity at the temperature of the antenna. After allowance was made for all such factors, it was clear by 1972 (see Thaddeus, Ann.Rev.Astron.Astrophys., 10, 305) that there remained an isotropic, unpolarised invariable 'excess' of microwave radiation, whose brightness temperature averaged (2.72 ± 0.08) K over the wavelengths from λ73.5cm to λ3.3mm. The constancy of the excess brightness temperature with observing wavelength was consistent with, but did not prove, that the microwave excess radiation had a black-body spectrum.

A genuine black-body spectrum at 2.7 K should peak at a wavelength of 1.9 mm; an observation that the 'excess' radiation indeed falls off in intensity at wavelengths <1.9 mm is therefore crucial to establishing its black-body character. Unfortunately the atmosphere is opaque to most infrared wavelengths short of λ3mm, so attempts to verify the short-wavelength peak have to be made from detectors taken above the Earth's atmosphere. The most convincing evidence for the black-body form of the spectrum comes from measurements with a balloon-borne spectrophotometer in the range λ4-0.4mm by Woody and Richards (Phys. Rev. Letters, 42, 925 (1979)). This experiment confirmed the existence of the expected spectral peak, but at the same time showed significant departures from the theoretical Planckian spectrum, at a level of about 10-20% of the observed intensity. The total intensity in the background, integrated over the wavelength range of the experiment, also corresponded to a black-body temperature of (2.96 ± 0.05) K, rather than the (2.72 ± 0.08) K deduced from the radio data. The reason for these small but significant discrepancies remains unclear.

A small anisotropy has been found in the radiation by Cheng et al. (Ap.J., 232, L139 (1979)), who used balloon-borne microwave radiometers at 19.0, 24.8 and 31.4 GHz to measure the background at an altitude of 27 km, minimising corrections for atmospheric radiation. The anisotropic component of the radiation is well fitted by a dipole distribution of intensity with amplitude (2.99 ± 0.34) mK and direction R.A. = $(12^h3 \pm 0^h4)$, Decl. = $(-1^o \pm 6^o)$. This dipolar anisotropy can be interpreted as due to the Doppler Effect resulting from a motion of our Galaxy relative to the background radiation of 540 $km.s^{-1}$; if this interpretation is correct, then the anisotropy is due to a local peculiar motion of no cosmological significance. The same experiment placed a limit of 2 mK on nonpolar anisotropies. It is also known (Nanos, Ap.J., 232, 341 (1979)) that any linearly-polarized component of the background is < 1.6 mK in intensity.

There is good evidence that the 2.7 K radiation exists elsewhere than in the Solar System, an important test if it is to be interpreted as a genuinely Universal background. Surprisingly, some of the evidence was available even before Alpher and Herman made their predictions, but its significance was not realised. It had been known since 1940 that interstellar absorption lines due to the diatomic molecule CN were not confined to transitions involving the lowest rotational level in the electronic ground state. Adams (Ap.J., 93, 11 (1941)) had noted that both the transition of CN at λ3874.608 Å (which occurs from the J=0 rotational level in the electronic ground state) and the transition at λ3873.998 Å (which occurs from the J=1 rotational level in this state) were seen in the visible absorption spectrum of the star ζ Ophiuchi. McKellar (Publ. Dominion Obs. Canada, 7, 251 (1941)) deduced from Adams' observed absorption-line ratio that the effective excitation temperature of the J=0 and J=1 rotational level populations in the electronic ground state was ~ 2.3 K. As the transition between the J=0 and J=1 rotational levels in the ground state corresponds to a photon of wavelength 2.64mm, this excitation temperature could be ascribed to exposure of the interstellar CN molecules to a radiation field at about 2.3 K. Even though this interpretation was noted in the classic textbook 'Spectra of Diatomic Molecules' by G.Herzberg in 1950, its significance was overlooked until a paper by Field and Hitchcock in Ap.J., 146, 1 (1966) pointed out that this old observation was consistent with the existence of the 'primordial fireball' radiation on the line of sight to ζ Ophiuchi.

Since then various observers (see the review by Thaddeus) have shown that the level populations of molecular species in various directions through our galaxy correspond to an excitation temperature of 2.7 K. It is important that there is no convincing evidence that the excitation temperature is anywhere much less than this, so that 2.7 K appears to be a minimum temperature for interstellar molecules that are in thermal equilibrium. (Obviously the excitation temperature can be higher in proximity to galactic sources of radiation, such as stars, clusters, etc.) We therefore have a growing body of evidence that the '2.7 K radiation' is black-body in spectrum, isotropic and unpolarised, and remote in origin.

4.2.2 Is the 2.7 K radiation actually the 'Big Bang' fireball ?

Although the hot Big Bang models must leave a radiative relic, the existence of the 2.7 K background does not by itself prove that there was a hot Big Bang. A number of alternative explanations have been proposed: a) that the 2.7 K background arises from a hitherto unknown class of discrete (i.e. localised) microwave emitter (e.g. Wolfe and Burbidge, Ap.J., 156, 345 (1969)), b) that it is radiation from interstellar or intergalactic dust grains, or c) that it is distant radiation but from a Universe which began as a 'cold' Big Bang but was later heated throughout by some process of local (astrophysical) character rather than of global (cosmological) character.

The discrete-source model was originally proposed in an attempt to rescue the Steady-State model from the apparently damning evidence that there was an early dense phase of the Universe. This model has the difficulty that the 2.7 K background is too bright to be accountable for by known types of cosmic microwave emitters, such as stars, galaxies or quasars, and furthermore has a spectrum unlike any known examples of such emitters. If the hypothetical sources were actually distributed in redshift (as would be necessary in a Steady-State model), the observed black-body form of the spectrum would then have to be an accidental combination of a non-Planckian spectrum for each source conspiring with the effects of the redshift to add up to a simulated Planck Law for their integrated emission. The observed isotropy of the 2.7 K background also requires that a very large number of individual sources contribute the total radiation along any line of sight, so that a major new radio constituent of the Universe was being proposed. This long chain of assumptions is obviously not very economical in comparison with the simple argument of Sec.4.1, and the discrete-source model has not found many supporters.

The main difficulty of ascribing the background to radiation from dust is that grains are inefficient radiators at wavelengths much greater than the grain size and it is hard to conceive of a mechanism for producing a widespread medium comprised of grains some centimetres in size. If the dust were supposed to be in our Galaxy the isotropy of the 2.7 K background is again hard to understand, while if it were outside our Galaxy the energetic problem of keeping the dust warm enough to radiate a 2.7 K spectrum is nontrivial (a photon energy density comparable to that of the 'Big Bang relic' interpretation would anyway be required). It would also be surprising if the hypothetical medium could exist without having other detectable effects, such as extinction and reddening of visible light.

It is difficult to exclude possibility (c), but it does seem contrived by comparison with the naive 'hot Big Bang' argument.

4.3 Relativistic world-models near the singularity

4.3.1 The effect of pressure in the models

In our neo-Newtonian world-models we developed the relation (1.8)

$$\ddot{R} = (\Lambda R^3/3 - GM_0)/R^2, \text{ i.e.}$$

$$\ddot{R} = (\Lambda - 4\pi G\rho)R/3$$

The full analogue of this relation in a model based on General Relativity requires that the 'active density' ρ be comprised of the terms

$$(4.3)\quad \rho = \rho_m + \rho_r + 3p/c^2$$

where ρ_m and ρ_r are the mass and radiation densities and p is the total pressure (including pressures of the matter and of the radiation). The addition of ρ_r to ρ_m when computing the active density is intuitively reasonable, but the gravitational effect of pressure is a purely General-Relativistic effect whose Newtonian analogue seems obscure.

This needed modification to the theory developed in Chapter One does not invalidate everything that went before because ρ_r and p are insignificant in comparison with ρ_m for most epochs in any model which corresponds at all closely to the real Universe. Consider the radiative contributions in the present Universe. They are:

$$\rho_r = u_r/c^2 = aT^4/c^2 = 4.6 \times 10^{-31} \text{ kg/m}^3 \text{ for } T = 2.72 \text{ K}$$

while $p_r = u_r/3 = aT^4/3$, so $3p_r/c^2$ also $= aT^4/c^2$

and the total radiative contribution to the active density is 9.2×10^{-31} kg/m^3 for T = 2.72 K. This is less than 1% of the known mass density of the galaxies that was estimated in Chapter Three.

The pressure exerted by the galaxies could be estimated as

$$p_m = \rho_m\langle v^2\rangle/3$$

where $\langle v^2\rangle$ is their mean square random velocity around the Hubble flow. Sandage and Tammann claim that $\langle v\rangle$ is less than 50 km/s, and we would certainly be safe to put $\langle v\rangle < 300$ km/s, in which case

$$3p_m/c^2 = \rho_m\langle v^2\rangle/c^2 < 10^{-6}\rho_m$$

which shows that the pressure of the galaxies contributes negligibly to the active density in comparison with their mass density. This expression for the matter pressure also shows that we cannot neglect p_m if the random velocities in the matter are ever ~c.

4.3.2 The model at $R(t) \rightarrow 0$

The dynamical effects of ρ_r and p cannot be neglected in the limit $R(t) \rightarrow 0$, because as in (4.1) $\rho_r(t)$ varies as $R^{-4}(t)$ while $\rho_m(t)$ only varies as $R^{-3}(t)$. Also, as $R(t) \rightarrow 0$, kT_r will become $> m_0c^2$, where m_0 is the rest-mass of some class of material particle. When this occurs the matter must contain relativistic (v~c) particle-antiparticle pairs of this class, in equilibrium with the photons. If these pairs are in thermal equilibrium, their energy densities will each be $u_m = xaT^4 = xu_r$, the x's being numerical factors determined by the quantum statistics of that class of particle (see Section 4.4 below). If the matter is mainly comprised of relativistic particle-antiparticle pairs as $R(t)\rightarrow 0$, the matter pressure term $3p_m/c^2$ will be ~ ρ_m, so that the active gravitating density

$$(4.4) \quad \rho \rightarrow 2\rho_r + 2\rho_m = 2(1+\textstyle\sum x)\rho_r = 2X\rho_r$$

where the sum $\sum x$ is carried out over all types of particle pair that are present in the model. All exploding relativistic models in the neighbourhood of the singularity must therefore be governed by

$$\ddot{R} \rightarrow (\Lambda - 8\pi GX\rho_r)R/3$$

which for the 'hot Big Bang' models in thermal equilibrium can be written

$$\ddot{R} \rightarrow \Lambda R/3 - 8\pi GXaT^4(t_0)R^4(t_0)/3R^3c^2$$

As $R \rightarrow 0$ the second term dominates and the difference between models with different cosmological constants disappears, leaving all models obeying

$$\ddot{R} = -C/R^3, \quad C = 8\pi GXaT^4(t_0)R^4(t_0)/3c^2, \text{ i.e.}$$

$\dot{R}^2 = C/R^2 + \text{constant}$

As R --> 0 the constant becomes insignificant and all models satisfy

$\dot{R} = (\sqrt{C})/R$

so that the common R(t) form is

(4.5) $R(t) = (4C)^{1/4}\sqrt{t} = (32\pi GXa/3c^2)^{1/4}T(t_0)R(t_0)\sqrt{t}$

which can be combined with the relation $T(t)R(t) = T(t_0)R(t_0)$ from (4.2) to give

(4.6) $T(t) = (3c^2/32\pi GXa)^{1/4}t^{-1/2} = 1{\cdot}52\times 10^{10}.X^{-1/4}.t^{-1/2}$

This means that all exploding models have the same thermal history T(t), regardless of the subsequent evolution of their scale factor R(t) in the later matter-dominated era when $\rho_m > \rho_r$. They remain indistinguishable until either the Λ-term takes effect at some expanded size or the non-relativistic matter density ρ_m becomes comparable to the radiation and radiation-pressure densities ρ_r and $3p_r/c^2$, after which the individuality of the models asserts itself. This property of the 'fireball' models allows us to make a unique classification of the earliest phenomena in the Big Bang on the basis of temperature or elapsed time almost independent of initial conditions in the matter.

4.4 Classification of eras in the hot Big Bang

In the earliest stages of the hot Big Bang, the radiation temperature will be so high that many photons have sufficient energy to create particle-antiparticle pairs ($kT > m_0c^2$ is an approximate criterion for the temperature at which creation of particles of rest-mass m_0 will be significant). Recent theoretical work in particle physics has suggested that the temperature could not exceed about 2×10^{12} K. This notion of a limiting temperature stems from the observation that the number of different species of elementary particle appears to increase exponentially with increasing energy; the measured density of mass states n(E) increases with energy as e^{bE} with $b^{-1} = 160$ MeV. In thermodynamic equilibrium the mean energy of the mass states at a temperature T is then

$$\langle E\rangle = \int_0^\infty En(E)e^{-E/kT}dE \Big/ \int_0^\infty n(E)e^{-E/kT}dE$$

which is non-convergent if $kT > b^{-1}$. What would happen as T approached the maximum temperature would be that increases in the energy density would go into the creation of new massive particle-antiparticle pairs from the photons rather than into increasing the kinetic energy of pre-existing particles. The era $t < 10^{-5}$ sec is therefore difficult to describe in terms of our limited experience of particle physics, but at $t >$ few x 10^{-5} sec we can begin to apply familiar concepts. It is then convenient to label 'eras' of the development (cooling) of the hot Big Bang in terms of the states of matter-radiation equilibrium that will prevail during these eras.

4.4.1 The hadron era ($t <$ few x 10^{-5} sec)

As there are many species of hadrons (strongly interacting baryons and mesons) and only three known species of leptons (electrons, muons and heavy τ -leptons) and their associated neutrinos, the main constituents of the earliest era of the hot Big Bang will be hadrons, their antihadrons, and roughly equal numbers of photons. As the expansion proceeds and the temperature drops, the hadron-photon equilibrium

$$H + \tilde{H} \longleftrightarrow h\nu + h\nu$$

shifts towards the right; the heaviest hadrons (the massive baryons known as hyperons) will annihilate with their antiparticles to produce photon pairs as the baryon content cascades down to its lowest states (the nucleons). In so doing, the total number of baryons and antibaryons decreases, but the number of baryons minus the number of antibaryons is constant. This conserved quantity is known as the baryon number of the model.

A key feature of this era is that not every nucleon found an antinucleon with which to annihilate as kT fell below m_0c^2 for nucleons (protons and neutrons). In particular, a net excess of protons over antiprotons appears to have survived to the present day. There are two possible interpretations of this; either there was a net excess of baryons over anti-baryons at the outset (i.e. models with the baryon number non-zero initially) or processes occurred which managed to separate baryons from anti-baryons to some extent into 'pockets' which could no longer interact (the zero-baryon-number or 'matter-antimatter' models).

The contribution of relativistic hadrons to the pressure at high temperatures will not be negligible throughout the hadron era, and the effects on the energy-density scale factor X of annihilation of the most massive hyperon-antihyperon pairs to produce photons will certainly modify the T-t relation (4.6) during the first ~ 10^{-5} sec. We do not yet know enough about the strong interaction to specify the evolution of the model throughout the hadron era. Fortunately, the details of the hadron era are not preserved, and our conclusions about what eventually emerges from a hot Big Bang are not very sensitive to the unknown aspects of the hadron era.

4.4.2 The lepton era (few x $10^{-5} < t < \sim 10$ sec)

Once kT falls below $m_\pi c^2$, where m_π is the rest-mass of a pion, all remaining hadron-antihadron pairs have annihilated and the energy content of the Universe is shared among photons, leptons and their neutrinos, and the few remaining 'excess' nucleons. As the number of leptons will be about equal to the number of photons as long as $kT \gg m_0 c^2$ for the leptons, the surviving nucleons will be a minority of the particles present, so this is now the 'lepton era'. The leptons obey Fermi-Dirac statistics such that the probability of finding a lepton in a state with energy E and momentum p is

$$P(E) = (e^{E(p)/kT} + 1)^{-1}$$

and $E^2 = p^2c^2 + (m_0c^2)^2 \longrightarrow p^2c^2$ for $kT \gg m_0 c^2$

From normal 'particle-in-a-box' wave mechanics, we have that the number of available single-particle states with momenta in a range (dp_x, dp_y, dp_z) in a spatial volume V is $gV dp_x dp_y dp_z/h^3$, where g is the number of available spin states of given momentum (g = 2 for leptons). Thus the equilibrium number density of muon-antimuon or electron-positron pairs of a given type (for which g = 2) with momenta between p and p+dp is

$$n(p) = (8\pi/h^3)p^2 dp/(e^{E/kT} + 1)$$

and the total energy density in one of these lepton-antilepton pairs in equilibrium at temperature T is

$$\begin{aligned} u &= 2\int_0^\infty n(p)E(p)\,dp \\ &= (16\pi c/h^3)\int_0^\infty p^3 dp/(e^{pc/kT} + 1) \\ &= (7/4)aT^4 \end{aligned}$$

The neutrinos have only one spin state so for them g = 1 and each type of neutrino-antineutrino pair contributes $u = (7/8)aT^4$ in equilibrium.

These results justify the assertion made earlier (Section 4.3.1) that the relativistic particle-antiparticle pairs in statistical equilibrium will have energy densities $u = xaT^4$. In thermodynamic terms the only difference between relativistic gases of photons, leptons and neutrinos is their values of x=1, x=7/4 and x=7/8. For most of the lepton era, the temperature is below kT

$= m_\mu c^2$ where m_μ is the mass of a muon, so the muons annihilate early and the main contributions to the energy density are photons, electrons and electron and muon neutrinos. The neutrinos of the recently-discovered heavy τ-lepton (Perl et al., Phys.Letters, 63B, 466 (1976)) may also contribute, but it is not yet known if they are as light as electron or muon neutrinos, so we will exclude them here. In this case the total energy density after muon recombination is:

$$u_{tot} = u_r + u_e + u_n = (1 + 7/4 + 7/4)\,aT^4 = (9/2)\,aT^4$$

in which case through most of the lepton era X=9/2 and our earlier relation (4.6) becomes

$$(4.7) \qquad T(t) = (c^2/48\pi Ga)^{1/4} t^{-1/2}$$

which can be written $t = (T_{10})^{-2}$ sec, where T_{10} is the temperature in units of 10^{10}K. Equation (4.7) specifies the thermal history of the model for the lepton era after the muons have decayed.

The assumption of thermodynamic equilibrium between the radiation and the electron-positron pairs would not be valid if the model expanded faster than the rate at which interactions between the electrons and the photons could bring them to equilibrium. While the expansion proceeds as $R(t) \propto \sqrt{t}$, the scale-doubling time is ~ 4t, i.e. of order seconds throughout the lepton era. The characteristic time scale for bringing the pairs and photons to equilibrium will be ~ $1/\sigma_T nc$ where σ_T is the Thomson cross-section of an electron and n is the particle density. This time scale is ~10^{-21} sec at T ~ 10^{11}K, so equilibrium is achieved virtually instantaneously compared with the expansion rate.

The lepton era ends when the temperature falls below $kT = m_0c^2$ for an electron ($T = 5.9 \times 10^9$ K) at t ~ 7 sec; after this the electrons and positrons annihilate. The neutrino gas is effectively decoupled from the rest of the model at this time and expands to form a relic neutrino flux analogous to the 2.7 K relic photon flux.

To preserve macroscopic charge neutrality, it is necessary that the number of excess electrons over positrons at the end of the lepton era equals the number of excess protons over antiprotons, i.e. that the net proton number equal the net electron number. The mechanism by which this balance was obtained is an unknown facet of the 'original conditions' whose origin is obscure.

4.4.3 The radiation era (10 sec < t < $t(\rho_r=\rho_m)$)

Once the electron-positron pairs have annihilated, the energy density in photons greatly outweighs that in the particles, and the Universe enters a photon-dominated era that can be described fairly precisely by equation (4.6) with X=1; this is normally termed the 'radiation era'. To see that the photon background, rather than the neutrino background, dominates the dynamics, consider the entropy of the expanding model. The first and second laws of thermodynamics for fluids undergoing quasistatic processes can be expressed by

$$TdS = dU + pdV$$

where S is the entropy. For a relativistic gas satisfying $p = u/3 = U/3V$ this can be written

$$TV^{1/3}dS = V^{1/3}dU + (1/3)UV^{-2/3}dV = d(UV^{1/3}), \quad \text{from which}$$

$$dS = d(UV^{1/3})/TV^{1/3}$$

For a relativistic gas with $u = XaT^4$, $U = XaT^4V$, it follows that

$$dS = d(XaT^4V^{4/3})/TV^{1/3}, \quad \text{i.e.}$$

$$dS/dV = (4/3).XaT^3$$

The entropy of a volume V of relativistic gas at temperature T is therefore

$$S = (4/3).XaT^3.V$$

At the time of annihilation of the electrons and positrons, the neutrino gas is decoupled from the other components of the model, and is cooling with $T_n \propto 1/R(t)$. The annihilation of the electrons and positrons would reduce the entropy of the lepton-photon gas by decreasing its value of X from 11/4 to 1, and this must be compensated in an adiabatic (isentropic) expansion by an increase in the photon temperature, such that

$$(4/3)(1 + 7/4)aT_{ri}^3 = (4/3)aT_{rf}^3$$

where T_{ri} is the initial temperature of the radiation before electron-positron annihilation and T_{rf} is the final temperature of the radiation after the annihilation. The radiation temperature is therefore increased by a factor $(11/4)^{1/3}$ as a result of the electron-positron annihilation, while the neutrino temperature is unaffected. Throughout the later stages of the model $T_r = (11/4)^{1/3}.T_n$ and the photon energy density $u_r = aT^4$ is therefore 10.9 times the neutrino energy density $u_n = (7/4)aT^4$.

Once the annihilations are complete, the number of photons in the Universe is constant (absorption and re-emission processes involving excited states of the matter should cancel on average), so that the number density of photons decreases as R^{-3}. (This number density is proportional to the entropy of the photon gas per unit volume). The relative numbers of photons and baryons at the end of the lepton era must therefore have been essentially the same as their relative numbers now. The number density of photons in a black-body radiation field at temperature T is

$n_\gamma = 16\pi\zeta(3).(kT/ch)^3$, by integration over the black-body spectrum,

where ζ is the Riemann zeta function. $\zeta(3) = 1.202$, and this relation reduces to

$$n_\gamma = 2.03 \times 10^7 . T^3 \ m^{-3}$$

which for T = 2.72 K gives $n_\gamma \sim 4 \times 10^8 \ m^{-3}$. From Section 3.6.5, it is reasonable to take the mean matter density to be $\sim 4.6 \times 10^{-28}$ kg.m^{-3} now ($\rho_m(t_0) = 1000\rho_r(t_0)$ for convenience). In this case, n_b now is ~ 0.3 m^{-3} so that for the model we are living in

$$n_\gamma/n_b \sim 1.5 \times 10^9$$

As the photons were mainly produced from hadron-antihadron pairs the initial fractional excess of hadrons over antihadrons in models with non-zero baryon number must have been $<10^{-9}$. From this small excess we must trace the history of the structures of the matter content of our Universe.

The radiation era ends when ρ_r drops below ρ_m so that the dynamics of the model go over to those described in Chapter One. The red shift z_e and time t_e at which this occurs must satisfy

$$R(t_0)/R(t_e) = 1 + z_e = \rho_m(t_0)/\rho_r(t_0)$$

Note also that ρ_{rot} ~2ρ_r, so should ρ_m~2ρ_r as end of rad. era? →

With our estimates of $\rho_r(t_0)$ and $\rho_m(t_0)$, z_e must be ~1000, at which era the temperature $T(t_e)$ would be ~2720 K. The time after the singularity at which the radiation era ends can therefore be estimated by finding when the temperature should have dropped to 2720 K according to equation (4.6) with X=1. This is only an approximation because the validity of equation (4.6) cannot extend all the way to $\rho_r = \rho_m$, but the result shows that $t_E \sim 10^6$ yrs, so the radiation era is much longer than those which preceded it. Clearly, our data on discrete astronomical sources (galaxies, quasars), are confined to z<<1000 and thus apply entirely to the matter-dominated era.

Note that the estimates of red shift, temperature and time at the end of the radiation era are sensitive to the adopted value of $\rho_m(t_0)$; if this were increased by a factor f, z_e and T_e would be about a factor f larger, and t_e would be a factor f^2 shorter.

4.4.4 The plasma era (10 sec < t < t(T~3500))

Another important characterisation of the post-lepton era is based on the state of ionization of the matter; the 'plasma era' spans the times at which the temperature is high enough to maintain the matter essentially fully ionized. Chapter Five will demonstrate that the transition from fully-ionized to neutral matter signals the beginning of an era when fluctuations in the matter density can be gravitationally stabilised.

The ionization equilibrium of the matter at temperature T is governed by the Saha Equation, which is obtained from the following considerations. In thermal equilibrium, the ratio of populations of two atomic states, 1 and 2, with energies E_1 and E_2 is

$$N_2/N_1 = (g_2/g_1)e^{-(E_2-E_1)/kT}$$

where g_1 and g_2 are the degeneracies of the two states. In hydrogen, the degeneracy of the n-th orbital state is

$$g_n = (\text{electron spin degeneracy}) \times (\text{orbital degeneracy})$$

$$= 2n^2$$

so the ratio of populations of the n'th state and the ground state (n=1) is

$$N_n/N_1 = n^2e^{-(E_n + I)/kT}$$

where I is the ionization potential of the hydrogen atom (13.6 eV). At sufficiently low temperatures, most of the hydrogen will be in the ground state, so $N_1 \sim N_H$. At high temperatures, we should sum over all bound states, so that

$$N_H = \sum_n N_n = N_1 \sum_n n^2e^{-(E_n + I)/kT}$$

This sum diverges if taken to very high n. It is unrealistic to include in the sum those very high-n states that will be unstable to collisional ionization, so the expression for N_H should in fact be written

$$(4.8) \qquad N_H = N_1 \sum_n n^2p_ne^{-(E_n + I)/kT}$$

where p_n is the probability that the atom can remain in the n-th state during a collision at temperature T; p_n decreases as n and T increase. We must also consider the equilibrium neutral H <--> $p + e^-$

by computing the occupancy of the free-electron states with positive energies $E = p^2/2m$ (we can use this non-relativistic p-E expression throughout the plasma era). As the number of free-electron states with momenta between p and p+dp in spatial volume V is

$$dN(p) = V.4\pi p^2 dp/h^3,$$ we have

$$N_e(p)dp/N_1 = (g_p g_e/g_1).V.(4\pi p^2/h^3).e^{-((p^2/2m) + I)/kT}$$

As the volume available per electron is $V = 1/n_p$ where n_p is the proton number density, and the degeneracies are $g_p = 1$ for a free proton, $g_e = 2$ for a free electron and $g_1 = 2$, we can write

$$N_e n_p/N_1 = (4\pi/h^3) \int_0^\infty p^2 e^{-((p^2/2m) + I)/kT} dp$$

which on converting to number densities of electrons and atoms is

$$(4.9) \quad n_e n_p/n_1 = ((\sqrt{2\pi mk})/h)^3.T^{3/2}.e^{-I/kT}$$

This and equation (4.8) constitute the Saha Equations for ionization in thermal equilibrium.

The salient point is that the number of ionized hydrogen atoms goes up much faster than $e^{-I/kT}$ as the temperature increases, because of the $T^{3/2}$ factor in equation (4.9). This happens because there are many more free states than bound states available to the electrons as the temperature rises, so at higher temperatures the Boltzmann factor $e^{-I/kT}$ becomes offset by the state-density factor. As a result, hydrogen at most densities becomes fully ionized at temperatures well below that at which $kT \sim I$ ($T \sim 160{,}000$ K).

Taking a model Universe with $\rho_m(t_0) = 4.6 \times 10^{-28}$ kg/m^3 composed primarily of hydrogen (we will see below that this composition is a fair approximation), equations (4.8) and (4.9) can be used to show that the matter would be >99% ionized for temperatures >4400 K, but <1% ionized for temperatures <2800 K. We can therefore consider that the electrons and protons recombine to form neutral hydrogen atoms at a matter temperature $T_m \sim 3500$ K, which will be reached at a redshift $z_{rec} \sim 1285$ if T_m is close to T_r. We shall see that the matter and radiation are still strongly coupled by Thomson scattering up until recombination, so that this assumption is fair. If we use equation (4.6) with X=1 to describe the model at $t < t_{rec}$, we can estimate that $t_{rec} \sim 1.88 \times 10^{13}$ sec, or 5.9×10^5 years. Use of equation (4.6) is justified so long as the Universe does indeed have $t_{rec} < t_e$, i.e. so long as recombination occurs before the radiation ceases to be dynamically dominant. The validity of this assumption depends on the actual

Sunyaev-Zeldovich Equation

value of the factor f that was discussed at the end of Section 4.4.3.

4.5 Nucleosynthesis in the Fireball

Throughout the hadron era, kT was much greater than the photodisintegration energies of stable nuclei, so that no nuclei would emerge from the hadron era, only free protons and neutrons. During the lepton era, the surviving nucleons will come to statistical equilibrium with the leptons and neutrinos through the weak interactions:

$$p + \bar{\nu}_e \longleftrightarrow n + e^+$$

$$p + e^- \longleftrightarrow n + \nu_e$$

for which the time to reach equilibrium varies as ~ $(T_{10})^{-5}$; a factor ~ $(T_{10})^{-2}$ here derives from the energy-dependence of the cross-sections for these reactions and a factor ~ $(T_{10})^{-3}$ from the particle density. Given that the expansion time-scale of the model varies as $(T_{10})^{-2}$, the time to reach neutron-proton equilibrium remains shorter than the expansion time until the temperature reaches 10^{10} K, after which the neutron-proton ratio lags behind the temperature drop in the expanding model. After the electron-positron pairs annihilate the free neutron decay

(4.10) $\quad n \longrightarrow p + e^- + \bar{\nu}_e,$

whose half-life is 10.8 min, eventually depletes the neutrons. For $T > 10^{10}$K the relative numbers of neutrons and protons will therefore be determined by the Boltzmann factor

(4.11) $\quad n_n/n_p = e^{-(m_n - m_p)c^2/kT}$

where $(m_n - m_p)c^2 = 1.293$ MeV. It is a coincidence that the neutron-proton ratio is 'frozen in' at T ~ 10^{10} K just as this Boltzmann factor becomes an interesting function of temperature, i.e. the inverse beta decay rate is coincidentally of the same order of magnitude as the expansion time scale near kT ~ $(m_n - m_p)c^2$. The 'frozen-in' neutron-proton ratio can be obtained by putting T ~ 10^{10} K in equation (4.11), i.e. n_n/n_p ~ 0.22. When T falls below 10^{10} K not only does the expansion outstrip the equilibration of the neutrons and protons, but the equilibrating reactions are themselves shut off when the electrons and positrons annihilate at the end of the lepton era.

We can therefore see that almost any hot Big Bang model (except those which contrive to prevent neutron-proton equilibrium being reached by postulating large initial excesses of some form of particle) must predict the <u>same</u> neutron-proton ratio near the end of the lepton era. This fact in turn allows relatively straightforward predictions of the path of nucleosynthesis, i.e. of the formation of nuclei by fusion processes, during the early phases of the radiation era.

4.5.1 Synthesis of He^4 in the fireball

After the neutron-proton ratio is frozen in, the principal reactions involving neutrons are the free neutron decay and the formation and photodissociation of deuterons

$$n + p \longleftrightarrow D + h\nu$$

While $kT > 2.225$ MeV (the binding energy of the deuteron), photodissociation of the deuterons holds this reaction in an equilibrium which we can compute in the same manner as the Saha ionization-equilibrium equation (4.9):

$$n_n n_p / n_D = 4/3 . (\sqrt{2\pi k}/h)^3 . (m_n m_p / m_D)^{3/2} . T^{3/2} . e^{-B/kT} \qquad (4.12)$$

where B = 2.225 MeV. Given that the present baryon density $n_b(t_0) \sim 3 \times 10^{-1}\ m^{-3}$ (Section 4.4.3), the total baryon density at temperature T and earlier time t should be approximately

$$n_b(t) = n_b(t_0) . R^3(t_0)/R^3(t) = 3 \times 10^{-2} . (T/2.72)^3\ m^{-3}$$

using which we can calculate n_n, n_p and n_D as a function of T. Equation (4.12) then shows that the deuteron production-dissociation equilibrium shifts sharply in favour of deuteron production at temperatures below $T \sim 8 \times 10^8$ K, a condition which occurs during the radiation era at t ~ 200 sec. This leads to a second numerical coincidence : the fusion of the neutrons to form photo-stable deuterons becomes favoured just as the neutron abundance begins to be depleted noticeably by the free neutron decay.

Once deuterium begins to accumulate at $T < 8 \times 10^8$ K, it is fused to He^4 by reactions which amount to

$$D + D \longrightarrow He^4 + h\nu$$

These proceed so rapidly that essentially all of the deuterium produced is converted to He^4. If <u>every</u> neutron that was frozen in at $T = 10^{10}$ K ultimately ended up in a He^4 nucleus, we would initially have 18 neutrons and 82 protons among every 100 nucleons, and these 18 neutrons would have combined with 18 protons to leave 36% of the final mass of the model in the form of

He^4. In fact, if we allow for all of the known nuclear reactions that could occur in the radiation era and calculate the depletion of neutrons by the free neutron decay between the 'freezing-in' at $T=10^{10}$ K and the onset of deuterium stability at $T = 8x10^8 K$, we expect a somewhat lower He^4 mass fraction of 25%±3% (Wagoner, Ap.J., 179, 343 (1973); Yang et al., Ap.J., 227, 697 (1979)) in the 'standard' hot Big-Bang model.

Fusion of this helium to more massive nuclei is strongly inhibited by the instability of the putative nuclei of mass 5 and mass 8, so the 'standard' hot Big-Bang model predicts that the material emerging from the fireball would be about 25% helium by mass, unless either the neutron-proton equilibrium is somehow distorted by the initial conditions or our estimates of the (T, n_b, t) parameter combinations in the standard model are incorrect. If, for example, the temperature fell to $8x10^8$ K in a time that was significantly longer than 200 sec, then most of the neutrons would be removed by the free neutron decay before they could be incorporated into deuterons; this would 'shut down' the He^4 production. We will discuss some non-standard possibilities in Section 4.5.5 below, but first let us consider whether there is any evidence in the observed Universe for a primordial ~25% He^4 mass fraction that could be a possible 'fossil relic' of the proposed primordial fireball.

4.5.2 Observed He^4 abundances

The spectral lines of neutral helium atoms are mainly in the ultraviolet, so the spectroscopic evidence for the presence of He^4 comes mainly from observations of the visible spectra of systems that are likely to contain singly-ionised helium. The surface abundances of stars are not generally expected to be mixed with the products of nucleosynthesis in stellar cores, so these surface abundances in most cases should indicate the abundances in the interstellar gas at the time the stars were formed. In the massive OB stars of Population I in our galaxy, the observed He^4 mass fraction is about 0.28, while the fraction in solar cosmic rays is about 0.20.

Indirect evidence for the He^4 abundance in a variety of stellar sysems comes from calculations of stellar models. The luminosity of a homogeneous star with a mass near one solar mass depends on the mean molecular weight to the 15/2 power, and the mean molecular weight depends mainly on the assumed He^4 abundance. Recent solar models which fit the age and luminosity require an initial solar helium abundance of ~0.23 (Bahcall et al., Ap.J., 184, 1 (1973); Ulrich and Rood, Nature Physical Science, 241, 111 (1973)). Such models cannot however account for the reportedly low solar neutrino flux, and the inhomogenous models that have been proposed to match the neutrino flux lead to lower values of the initial solar He^4 abundance.

Studies of old star clusters give constraints on the helium abundance from attempts to fit the observed mass, luminosity and age parameters to stellar models. Typical derived aundances are ~0.25 (e.g. Rood, Ap.J., 184, 815 (1973); Bohm-Vitense and Szkody, Ap.J., 184, 211 (1973)), with uncertainties generally about ±0.06.

Optical and radio studies of the helium abundance in HII regions in our Galaxy and in neighbouring galaxies favour values between 0.25 and 0.30, while helium abundances inferred for Cepheid variables in the Magellanic Clouds and in M31 are ~0.29 (Iben and Tuggle, Ap.J., 197, 39 (1975)). In a study of the stellar content of dwarf spheroidal galaxies, Hirshfield (Yale thesis, 1978) found a helium abundance of 0.24±0.02.

We therefore have quite reasonable evidence for a 'global' He^4 mass fraction that is of order 0.23-0.29; depending on the nature of the object whose He^4 abundance was measured, the observed abundances should be decreased by from 5% to 2% to allow for the buildup of He^4 in the interstellar medium as a result of the processes of stellar evolution. The observations are therefore quite consistent with a 'global' primordial He^4 fraction that was 0.25±0.03. This agreement is widely interpreted as evidence in favour of the 'standard' hot Big-Bang models. The strength of this argument is the extreme difficulty we would have in explaining how the observed He^4 abundance could be so large and so uniform by any process other than Big-Bang synthesis.

The main cause for disquiet about this 'confirmation' of the predicted helium mass fraction is the existence of some old Population II stars with surface helium abundances that are about ten times lower than the mean abundances given above (e.g. Danziger, Ann. Rev. Astron. Astrophys., 8, 161 (1970)). Most of these helium-poor stars show other 'anomalous' element abundances which suggest that their surface compositions have been significantly altered by their 'internal' astrophysics, but this is not always the case. It is not yet clear whether these helium-poor stars can be dismissed as unusual systems in which local phenomena have succeeded in 'hiding' the primordial abundance, or alternatively whether they are the only places in the Galaxy where we see a truly primordial, but low, helium fraction that is in conflict with the standard Big-Bang model.