18. EXPERIMENTAL TESTS OF GRAVITATIONAL THEORY

Revised September 2003 by T. Damour (IHES, Bures-sur-Yvette, France).

Einstein's General Relativity, the current "standard" theory of gravitation, describes gravity as a universal deformation of the Minkowski metric:

$$g_{\mu\nu}(x^\lambda) = \eta_{\mu\nu} + h_{\mu\nu}(x^\lambda)\ , \text{ where } \eta_{\mu\nu} = \mathrm{diag}(-1,+1,+1,+1)\ . \quad (18.1)$$

Alternatively, it can be defined as the unique, consistent, local theory of a massless spin-2 field, $h_{\mu\nu}$, whose source must then be the total, conserved energy-momentum tensor [1]. General Relativity is classically defined by two postulates. One postulate states that the Lagrangian density describing the propagation and self-interaction of the gravitational field is

$$\mathcal{L}_{\rm Ein}[g_{\mu\nu}] = \frac{c^4}{16\pi G_N}\sqrt{g}g^{\mu\nu}R_{\mu\nu}(g)\ , \quad (18.2)$$

$$R_{\mu\nu}(g) = \partial_\alpha\Gamma^\alpha_{\mu\nu} - \partial_\nu\Gamma^\alpha_{\mu\alpha} + \Gamma^\beta_{\alpha\beta}\Gamma^\alpha_{\mu\nu} - \Gamma^\beta_{\alpha\nu}\Gamma^\alpha_{\mu\beta}\ , \quad (18.3)$$

$$\Gamma^\lambda_{\mu\nu} = \tfrac{1}{2}g^{\lambda\sigma}(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu})\ , \quad (18.4)$$

where G_N is Newton's constant, $g = -\det(g_{\mu\nu})$, and $g^{\mu\nu}$ is the matrix inverse of $g_{\mu\nu}$. A second postulate states that $g_{\mu\nu}$ couples universally, and minimally, to all the fields of the Standard Model by replacing everywhere the Minkowski metric $\eta_{\mu\nu}$. Schematically (suppressing matrix indices and labels for the various gauge fields and fermions, and for the Higgs doublet),

$$\begin{aligned}\mathcal{L}_{\rm SM}[\psi, A_\mu, H, g_{\mu\nu}] = &-\tfrac{1}{4}\sum\sqrt{g}g^{\mu\alpha}g^{\nu\beta}F^a_{\mu\nu}F^a_{\alpha\beta}\\ &-\sum\sqrt{g}\,\overline{\psi}\,\gamma^\mu D_\mu\psi\\ &-\tfrac{1}{2}\sqrt{g}g^{\mu\nu}\overline{D_\mu H}D_\nu H - \sqrt{g}\,V(H)\\ &-\sum\lambda\sqrt{g}\,\overline{\psi}\,H\psi\ ,\end{aligned} \quad (18.5)$$

where $\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}$, and where the covariant derivative D_μ contains, besides the usual gauge field terms, a (spin-dependent) gravitational contribution $\Gamma_\mu(x)$ [2]. From the total action $S_{\rm tot}[g_{\mu\nu},\psi,A_\mu,H] = c^{-1}\int d^4x(\mathcal{L}_{\rm Ein} + \mathcal{L}_{\rm SM})$ follows Einstein's field equations,

$$R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} = \frac{8\pi G_N}{c^4}T_{\mu\nu}\ . \quad (18.6)$$

Here, $R = g^{\mu\nu}R_{\mu\nu}$, $T_{\mu\nu} = g_{\mu\alpha}g_{\nu\beta}T^{\alpha\beta}$, and $T^{\mu\nu} = (2/\sqrt{g})\delta\mathcal{L}_{\rm SM}/\delta g_{\mu\nu}$ is the (symmetric) energy-momentum tensor of the Standard Model matter. The theory is invariant under arbitrary coordinate transformations: $x'^\mu = f^\mu(x^\nu)$. To solve the field equations Eq. (18.6), one needs to fix this coordinate gauge freedom. *E.g.*, the "harmonic gauge" (which is the analogue of the Lorentz gauge, $\partial_\mu A^\mu = 0$, in electromagnetism) corresponds to imposing the condition $\partial_\nu(\sqrt{g}g^{\mu\nu}) = 0$.

In this *Review*, we only consider the classical limit of gravitation (*i.e.*, classical matter and classical gravity). Considering quantum matter in a classical gravitational background already poses interesting challenges, notably the possibility that the zero-point fluctuations of the matter fields generate a nonvanishing vacuum energy density $\rho_{\rm vac}$, corresponding to a term $-\sqrt{g}\ \rho_{\rm vac}$ in $\mathcal{L}_{\rm SM}$ [3]. This is equivalent to adding a "cosmological constant" term $+\Lambda g_{\mu\nu}$ on the left-hand side of Einstein's equations Eq. (18.6), with $\Lambda = 8\pi G_N\rho_{\rm vac}/c^4$. Recent cosmological observations (see the following Reviews) suggest a positive value of Λ corresponding to $\rho_{\rm vac} \approx (2.3\times10^{-3}{\rm eV})^4$. Such a small value has a negligible effect on the tests discussed below. Quantizing the gravitational field itself poses a very difficult challenge because of the perturbative non-renormalizability of Einstein's Lagrangian. Superstring theory offers a promising avenue toward solving this challenge.

18.1. Experimental tests of the coupling between matter and gravity

The universality of the coupling between $g_{\mu\nu}$ and the Standard Model matter postulated in Eq. (18.5) ("Equivalence Principle") has many observable consequences. First, it predicts that the outcome of a local non-gravitational experiment, referred to local standards, does not depend on where, when, and in which locally inertial frame the experiment is performed. This means, for instance, that local experiments should neither feel the cosmological evolution of the universe (constancy of the "constants"), nor exhibit preferred directions in spacetime (isotropy of space, local Lorentz invariance). These predictions are consistent with many experiments and observations. The best limit on a possible time variation of the basic coupling constants concerns the fine-structure constant $\alpha_{\rm em}$, and has been obtained by analyzing a natural fission reactor phenomenon which took place at Oklo, Gabon, two billion years ago [4]. A conservative estimate of the (95% C.L.) Oklo limit on the variability of $\alpha_{\rm em}$ is (see second reference in [4])

$$-0.9\times10^{-7} < \frac{\alpha_{\rm em}^{\rm Oklo} - \alpha_{\rm em}^{\rm now}}{\alpha_{\rm em}} < 1.2\times10^{-7}\ , \quad (18.7)$$

which corresponds to the following limit on the average time derivative of $\alpha_{\rm em}$

$$-6.7\times10^{-17}{\rm yr}^{-1} < \dot{\alpha}_{\rm em}/\alpha_{\rm em} < 5.0\times10^{-17}{\rm yr}^{-1}\ . \quad (18.8)$$

Direct laboratory limits on the time variation of $\alpha_{\rm em}$ [5] are less stringent than Eq. (18.8). Recently, Ref. 6 claimed to have measured a cosmological change of $\alpha_{\rm em}$ of magnitude $\Delta\alpha_{\rm em}/\alpha_{\rm em} = (-0.72\pm0.18)\times10^{-5}$ around redshifts $z\simeq0.5-3.5$. When analyzed within various dilaton-like theoretical models of $\alpha_{\rm em}$ variability [7–9], such a large cosmological variation of $\alpha_{\rm em}$ appears incompatible with the combined set of other experimental limits, notably the Oklo bound quoted above, a comparable "rhenium" geological bound [10], and the limits on possible violations of the universality of free fall quoted below. See Refs. 11,12 for discussions of systematic errors in astronomical measurements of $\alpha_{\rm em}$, and Ref. 13 for a general review of the issue of "variable constants".

The highest precision tests of the isotropy of space have been performed by looking for possible quadrupolar shifts of nuclear energy levels [14]. The (null) results can be interpreted as testing the fact that the various pieces in the matter Lagrangian Eq. (18.5) are indeed coupled to one and the same external metric $g_{\mu\nu}$ to the 10^{-27} level. For astrophysical constraints on possible Planck-scale violations of Lorentz invariance, see Refs. 15,16.

The universal coupling to $g_{\mu\nu}$ postulated in Eq. (18.5) implies that two (electrically neutral) test bodies dropped at the same location and with the same velocity in an external gravitational field fall in the same way, independently of their masses and compositions. The universality of the acceleration of free fall has been verified at the 10^{-12} level both for laboratory bodies [17],

$$(\Delta a/a)_{\rm BeCu} = (-1.9\pm2.5)\times10^{-12}\ , \quad (18.9)$$

and for the gravitational accelerations of the Moon and the Earth toward the Sun [18],

$$(\Delta a/a)_{\rm MoonEarth} = (-3.2\pm4.6)\times10^{-13}\ . \quad (18.10)$$

See also Ref. 19 for a $\pm0.94\times10^{-12}$ limit on the fractional difference in free-fall acceleration toward the Sun of earth-core-like and moon-mantle-like laboratory bodies, and Ref. 20 for *short-range* tests of the universality of free-fall.

Finally, Eq. (18.5) also implies that two identically constructed clocks located at two different positions in a static external Newtonian potential $U(\boldsymbol{x}) = \sum G_N m/r$ exhibit, when intercompared by means of electromagnetic signals, the (apparent) difference in clock rate,

$$\frac{\tau_1}{\tau_2} = \frac{\nu_2}{\nu_1} = 1 + \frac{1}{c^2}[U(\boldsymbol{x}_1) - U(\boldsymbol{x}_2)] + O\left(\frac{1}{c^4}\right)\ , \quad (18.11)$$

independently of their nature and constitution. This universal gravitational redshift of clock rates has been verified at the 10^{-4} level by comparing a hydrogen-maser clock, flying on a rocket up to an altitude $\sim 10,000$ km, to a similar clock on the ground [21]. For more details and references on experimental gravity see, *e.g.*, Refs. 22 and 23.

18.2. Tests of the dynamics of the gravitational field in the weak field regime

The effect on matter of one-graviton exchange, *i.e.*, the interaction Lagrangian obtained when solving Einstein's field equations Eq. (18.6) written in, say, the harmonic gauge at first order in $h_{\mu\nu}$,

$$\Box h_{\mu\nu} = -\frac{16\pi G_N}{c^4}(T_{\mu\nu} - \tfrac{1}{2}T\eta_{\mu\nu}) + O(h^2) + O(hT)\ , \qquad (18.12)$$

reads $-(8\pi G_N/c^4)T^{\mu\nu}\Box^{-1}(T_{\mu\nu} - \frac{1}{2}T\eta_{\mu\nu})$. For a system of N moving point masses, with free Lagrangian $L^{(1)} = \sum_{A=1}^{N} - m_A c^2\sqrt{1 - \boldsymbol{v}_A^2/c^2}$, this interaction, expanded to order v^2/c^2, reads (with $r_{AB} \equiv |\boldsymbol{x}_A - \boldsymbol{x}_B|$, $\boldsymbol{n}_{AB} \equiv (\boldsymbol{x}_A - \boldsymbol{x}_B)/r_{AB}$)

$$L^{(2)} = \tfrac{1}{2}\sum_{A\neq B}\frac{G_N m_A m_B}{r_{AB}}\left[1 + \frac{3}{2c^2}(\boldsymbol{v}_A^2 + \boldsymbol{v}_B^2) - \frac{7}{2c^2}(\boldsymbol{v}_A\cdot\boldsymbol{v}_B) - \frac{1}{2c^2}(\boldsymbol{n}_{AB}\cdot\boldsymbol{v}_A)(\boldsymbol{n}_{AB}\cdot\boldsymbol{v}_B) + O\left(\frac{1}{c^4}\right)\right]\ . \qquad (18.13)$$

The two-body interactions Eq. (18.13) exhibit v^2/c^2 corrections to Newton's $1/r$ potential induced by spin-2 exchange. Consistency at the "post-Newtonian" level $v^2/c^2 \sim G_N m/rc^2$ requires that one also consider the three-body interactions induced by some of the three-graviton vertices, and other nonlinearities (terms $O(h^2)$ and $O(hT)$ in Eq. (18.12)),

$$L^{(3)} = -\frac{1}{2}\sum_{B\neq A\neq C}\frac{G_N^2 m_A m_B m_C}{r_{AB} r_{AC}\, c^2} + O\left(\frac{1}{c^4}\right)\ . \qquad (18.14)$$

All currently performed gravitational experiments in the solar system, including perihelion advances of planetary orbits, the bending and delay of electromagnetic signals passing near the Sun, and very accurate ranging data to the Moon obtained by laser echoes, are compatible with the post-Newtonian results Eqs. (18.12)–(18.14).

Similarly to what is done in discussions of precision electroweak experiments (see Section 10 in this *Review*), it is useful to quantify the significance of precision gravitational experiments by parameterizing plausible deviations from General Relativity. The addition of a mass-term in Einstein's field equations leads to a score of theoretical difficulties [24], which have not yet received any consensual solution. We shall, therefore, not consider here the ill-defined "mass of the graviton" as a possible deviation parameter from General Relativity (see, however, the phenomenological limits quoted in the Section "Gauge and Higgs Bosons" of this *Review*). Deviations from Einstein's pure spin-2 theory are then defined by adding new bosonic light or massless macroscopically coupled fields. The possibility of new gravitational-strength couplings leading (on small, and possibly large, scales) to deviations from Einsteinian (and Newtonian) gravity is suggested by String Theory [25], and by Brane World ideas [26]. For compilations of experimental constraints on Yukawa-type additional interactions, see Refs. [17,27,28] and the Section "Extra Dimensions" in this *Review*. Recent experiments have set limits on non-Newtonian forces below 0.1 mm [29].

Here, we shall focus on the parametrization of long-range deviations from relativistic gravity obtained by adding a massless scalar field φ, coupled to the trace of the energy-momentum tensor $T = g_{\mu\nu}T^{\mu\nu}$ [30]. The most general such theory contains an arbitrary function $a(\varphi)$ of the scalar field, and can be defined by the Lagrangian

$$\mathcal{L}_{\rm tot}[g_{\mu\nu},\varphi,\psi,A_\mu,H] = \frac{c^4}{16\pi G}\sqrt{g}(R(g) - 2g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi) + \mathcal{L}_{\rm SM}[\psi, A_\mu, H, \widetilde{g}_{\mu\nu}]\ , \qquad (18.15)$$

where G is a "bare" Newton constant, and where the Standard Model matter is coupled not to the "Einstein" (pure spin-2) metric $g_{\mu\nu}$, but to the conformally related ("Jordan-Fierz") metric $\widetilde{g}_{\mu\nu} = \exp(2a(\varphi))g_{\mu\nu}$. The scalar field equation $\Box_g\varphi = -(4\pi G/c^4)\alpha(\varphi)T$ displays $\alpha(\varphi) \equiv \partial a(\varphi)/\partial\varphi$ as the basic (field-dependent) coupling between φ and matter [31]. The one-parameter (ω) Jordan-Fierz-Brans-Dicke theory [30] is the special case $a(\varphi) = \alpha_0\varphi$, leading to a field-independent coupling $\alpha(\varphi) = \alpha_0$ (with ${\alpha_0}^2 = 1/(2\omega+3)$).

In the weak field, slow motion limit, appropriate to describing gravitational experiments in the solar system, the addition of φ modifies Einstein's predictions only through the appearance of two "post-Einstein" dimensionless parameters: $\overline{\gamma} = -2\alpha_0^2/(1+\alpha_0^2)$ and $\overline{\beta} = +\frac{1}{2}\beta_0\alpha_0^2/(1+\alpha_0^2)^2$, where $\alpha_0 \equiv \alpha(\varphi_0)$, $\beta_0 \equiv \partial\alpha(\varphi_0)/\partial\varphi_0$, φ_0 denoting the vacuum expectation value of φ. These parameters also show up naturally (in the form $\gamma_{\rm PPN} = 1 + \overline{\gamma}$, $\beta_{\rm PPN} = 1 + \overline{\beta}$) in phenomenological discussions of possible deviations from General Relativity [32,22]. The parameter $\overline{\gamma}$ measures the admixture of spin 0 to Einstein's graviton, and contributes an extra term $+\overline{\gamma}(\boldsymbol{v}_A - \boldsymbol{v}_B)^2/c^2$ in the square brackets of the two-body Lagrangian Eq. (18.13). The parameter $\overline{\beta}$ modifies the three-body interaction Eq. (18.14) by a factor $1 + 2\overline{\beta}$. Moreover, the combination $\eta \equiv 4\overline{\beta} - \overline{\gamma}$ parameterizes the lowest order effect of the self-gravity of orbiting masses by modifying the Newtonian interaction energy terms in Eq. (18.13) into $G_{AB}m_A m_B/r_{AB}$, with a body-dependent gravitational "constant" $G_{AB} = G_N[1 + \eta(E_A^{\rm grav}/m_A c^2 + E_B^{\rm grav}/m_B c^2) + O(1/c^4)]$, where $G_N = G\exp[2a(\varphi_0)](1+\alpha_0^2)$, and where $E_A^{\rm grav}$ denotes the gravitational binding energy of body A.

The best current limits on the post-Einstein parameters $\overline{\gamma}$ and $\overline{\beta}$ are (at the 68% confidence level): (i) $\overline{\gamma} = (2.1 \pm 2.3)\times 10^{-5}$, deduced from the additional Doppler shift experienced by radio-wave beams connecting the Earth to the Cassini spacecraft when they passed near the Sun [33], and (ii) $4\overline{\beta} - \overline{\gamma} = -0.0007 \pm 0.0010$ [18] from Lunar Laser Ranging measurements of a possible polarization of the Moon toward the Sun [34]. More stringent limits on $\overline{\gamma}$ are obtained in models (*e.g.*, string-inspired ones [35]) where scalar couplings violate the Equivalence Principle.

18.3. Tests of the dynamics of the gravitational field in the radiative and/or strong field regimes

The discovery of pulsars (*i.e.*, rotating neutron stars emitting a beam of radio noise) in gravitationally bound orbits [36,37] has opened up an entirely new testing ground for relativistic gravity, giving us an experimental handle on the regime of radiative and/or strong gravitational fields. In these systems, the finite velocity of propagation of the gravitational interaction between the pulsar and its companion generates damping-like terms at order $(v/c)^5$ in the equations of motion [38]. These damping forces are the local counterparts of the gravitational radiation emitted at infinity by the system ("gravitational radiation reaction"). They cause the binary orbit to shrink and its orbital period P_b to decrease. The remarkable stability of the pulsar clock has allowed Taylor and collaborators to measure the corresponding very small orbital period decay $\dot{P}_b \equiv dP_b/dt \sim -(v/c)^5 \sim -10^{-12}$ [37,39,40], thereby giving us a direct experimental confirmation of the propagation properties of the gravitational field, and, in particular, an experimental confirmation that the speed of propagation of gravity is equal to the velocity of light to better than a part in a thousand [41]. In addition, the surface gravitational potential of a neutron star $h_{00}(R) \simeq 2Gm/c^2R \simeq 0.4$ being a factor $\sim 10^8$ higher than the surface potential of the Earth, and a mere factor 2.5 below the black hole limit ($h_{00} = 1$), pulsar data are sensitive probes of the strong-gravitational-field regime.

Binary pulsar timing data record the times of arrival of successive electromagnetic pulses emitted by a pulsar orbiting around the center of mass of a binary system. After correcting for the Earth motion around the Sun, and for the dispersion due to propagation in the interstellar plasma, the time of arrival of the Nth pulse t_N can be described by a generic, parameterized "timing formula," [42] whose functional form is common to the whole class of tensor-scalar gravitation theories:

$$t_N - t_0 = F[T_N(\nu_p, \dot{\nu}_p, \ddot{\nu}_p)\,;\,\{p^K\}\,;\,\{p^{PK}\}]\ . \qquad (18.16)$$

Here, T_N is the pulsar proper time corresponding to the Nth turn given by $N/2\pi = \nu_p T_N + \frac{1}{2}\dot{\nu}_p T_N^2 + \frac{1}{6}\ddot{\nu}_p T_N^3$ (with $\nu_p \equiv 1/P_p$ the spin frequency of the pulsar, *etc.*), $\{p^K\} = \{P_b, T_0, e, \omega_0, x\}$ is the set of "Keplerian" parameters, (notably, orbital period P_b, eccentricity e and projected semi-major axis $x = a \sin i/c$), and $\{p^{PK}\} = \{k, \gamma_{\rm timing}, \dot{P}_b, r, s, \delta_\theta, \dot{e}, \dot{x}\}$ denotes the set of (separately measurable) "post-Keplerian" parameters. Most important among these are: the fractional periastron advance per orbit $k \equiv \dot{\omega} P_b/2\pi$, a dimensionful time-dilation parameter $\gamma_{\rm timing}$, the orbital period derivative $\dot{P}_b$, and the "range" and "shape" parameters of the gravitational time delay caused by the companion, r and s.

Without assuming any specific theory of gravity, one can phenomenologically analyze the data from any binary pulsar by least-squares fitting the observed sequence of pulse arrival times to the timing formula Eq. (18.16). This fit yields the "measured" values of the parameters $\{\nu_p, \dot{\nu}_p, \ddot{\nu}_p\}$, $\{p^K\}$, $\{p^{PK}\}$. Now, each specific relativistic theory of gravity predicts that, for instance, k, $\gamma_{\rm timing}$, $\dot{P}_b$, r, and s (to quote parameters that have been successfully measured from some binary pulsar data) are some theory-dependent functions of the Keplerian parameters and of the (unknown) masses m_1, m_2 of the pulsar and its companion. For instance, in General Relativity, one finds (with $M \equiv m_1 + m_2$, $n \equiv 2\pi/P_b$)

$$\begin{aligned} k^{\rm GR}(m_1,m_2) &= 3(1-e^2)^{-1}(G_N M n/c^3)^{2/3} \ , \\ \gamma^{\rm GR}_{\rm timing}(m_1,m_2) &= e n^{-1}(G_N M n/c^3)^{2/3} m_2(m_1+2m_2)/M^2 \ , \\ \dot{P}_b^{\rm GR}(m_1,m_2) &= -(192\pi/5)(1-e^2)^{-7/2}\left(1+\tfrac{73}{24}e^2+\tfrac{37}{96}e^4\right) \\ &\quad \times (G_N M n/c^3)^{5/3} m_1 m_2/M^2 \ , \\ r(m_1,m_2) &= G_N m_2/c^3 \ , \quad \text{and} \\ s(m_1,m_2) &= n x (G_N M n/c^3)^{-1/3} M/m_2 \ . \end{aligned} \tag{18.17}$$

In tensor-scalar theories, each of the functions $k^{\rm theory}(m_1,m_2)$, $\gamma^{\rm theory}_{\rm timing}(m_1,m_2)$, $\dot{P}_b^{\rm theory}(m_1,m_2)$, *etc.* is modified by quasi-static strong field effects (associated with the self-gravities of the pulsar and its companion), while the particular function $\dot{P}_b^{\rm theory}(m_1,m_2)$ is further modified by radiative effects (associated with the spin 0 propagator) [31,43].

Let us summarize the current experimental situation (see Ref. 44 for a more extensive review). In the first discovered binary pulsar PSR1913 + 16 [36,37], it has been possible to measure with accuracy the three post-Keplerian parameters k, $\gamma_{\rm timing}$ and $\dot{P}_b$. The three equations $k^{\rm measured} = k^{\rm theory}(m_1,m_2)$, $\gamma^{\rm measured}_{\rm timing} = \gamma^{\rm theory}_{\rm timing}(m_1,m_2)$, $\dot{P}_b^{\rm measured} = \dot{P}_b^{\rm theory}(m_1,m_2)$ determine, for each given theory, three curves in the two-dimensional mass plane. This yields *one* (combined radiative/strong-field) test of the specified theory, according to whether the three curves meet at one point, as they should. After subtracting a small ($\sim 10^{-14}$ level in $\dot{P}_b^{\rm obs} = (-2.4211 \pm 0.0014) \times 10^{-12}$), but significant Newtonian perturbing effect caused by the Galaxy [45], one finds that General Relativity passes this $(k - \gamma_{\rm timing} - \dot{P}_b)_{1913+16}$ test with complete success at the 10^{-3} level [37,39,40]

$$\begin{aligned} \left[\frac{\dot{P}_b^{\rm obs} - \dot{P}_b^{\rm galactic}}{\dot{P}_b^{\rm GR}[k^{\rm obs}, \gamma^{\rm obs}_{\rm timing}]}\right]_{1913+16} &= 1.0026 \pm 0.0006({\rm obs}) \pm 0.0021({\rm galactic}) \\ &= 1.0026 \pm 0.0022 \ . \end{aligned} \tag{18.18}$$

Here $\dot{P}_b^{\rm GR}[k^{\rm obs}, \gamma^{\rm obs}_{\rm timing}]$ is the result of inserting in $\dot{P}_b^{\rm GR}(m_1,m_2)$ the values of the masses predicted by the two equations $k^{\rm obs} = k^{\rm GR}(m_1,m_2)$, $\gamma^{\rm obs}_{\rm timing} = \gamma^{\rm GR}_{\rm timing}(m_1,m_2)$. This experimental evidence for the reality of gravitational radiation damping forces at the 0.3% level is illustrated in Fig. 18.1, which shows actual orbital phase data (after subtraction of a linear drift).

The discovery of the binary pulsar PSR1534 + 12 [46] has allowed one to measure the four post-Keplerian parameters k, $\gamma_{\rm timing}$, r and s, and thereby to obtain *two* (four observables minus two masses) tests of strong field gravity, without mixing of radiative effects [47]. General Relativity passes these tests within the measurement accuracy [37,47]. The most precise of these new, pure, strong-field tests is the one obtained by combining the measurements of k, γ, and s. Using the most recent data [48], one finds agreement at the 1% level:

$$\left[\frac{s^{\rm obs}}{s^{\rm GR}[k^{\rm obs}, \gamma^{\rm obs}_{\rm timing}]}\right]_{1534+12} = 1.000 \pm 0.007 \ . \tag{18.19}$$

It has also been possible to measure the orbital period change of PSR1534 + 12. General Relativity passes the corresponding $(k - \gamma_{\rm timing} - \dot{P}_b)_{1534+12}$ test with success at the 15% level [48].

The discovery of the binary pulsar PSR J1141 − 6545 [49] (whose companion is probably a white dwarf) has recently led to the measurement of the three post-Keplerian parameters k, $\gamma_{\rm timing}$ and $\dot{P}_b$ [50]. As in the PSR 1913 + 16 system, this yields *one* combined radiative/strong-field test of relativistic gravity. One finds that General Relativity passes this $(k - \gamma_{\rm timing} - \dot{P}_b)_{1141-6545}$ test with success at the 25% level [50]. Several other binary pulsar systems, of a nonsymmetric type (nearly circular systems made of a neutron star and a white dwarf), can also be used to test relativistic gravity [51–54]. The constraints on tensor-scalar theories provided by three binary-pulsar "experiments" have been analyzed in Ref. 43, and shown to exclude a large portion of the parameter space allowed by solar-system tests. Measurements of the pulse shape of PSR1913 + 16 [55] have detected a time variation of the pulse shape, compatible with the prediction [56] that the general relativistic spin-orbit coupling should cause a secular change in the orientation of the pulsar beam, with respect to the line of sight ("geodetic precession").

The tests considered above have examined the gravitational interaction on scales between a fraction of a millimeter and a few

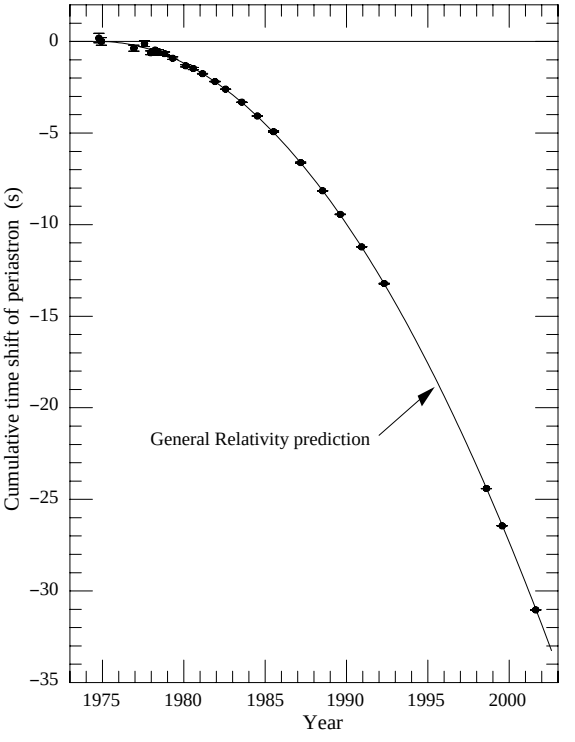

Figure 18.1: Accumulated shift of the times of periastron passage in the PSR 1913+16 system, relative to an assumed orbit with a constant period. The parabolic curve represents the general relativistic prediction, modified by Galactic effects, for orbital period decay from gravitational radiation damping forces. (Figure obtained with permission from Ref. 40.)

astronomical units. On the other hand, the general relativistic action on light and matter of an external gravitational field on a length scale ~ 100 kpc has been verified to $\sim 30\%$ in some gravitational lensing systems (see, *e.g.*, Ref. 57). Some tests on cosmological scales are also available. In particular, Big Bang Nucleosynthesis (see Section 20 of this *Review*) has been used to set significant constraints on the variability of the gravitational "constant" [58]. For other cosmological tests of the "constancy of constants," see the review [13].

18.4. Conclusions

All present experimental tests are compatible with the predictions of the current "standard" theory of gravitation: Einstein's General Relativity. The universality of the coupling between matter and gravity (Equivalence Principle) has been verified at the 10^{-12} level. Solar system experiments have tested all the weak-field predictions of Einstein's theory at better than the 10^{-3} level (and down to the 2×10^{-5} level for the post-Einstein parameter $\overline{\gamma}$). The propagation properties of relativistic gravity, as well as several of its strong-field aspects, have been verified at the 10^{-3} level in binary pulsar experiments. Recent laboratory experiments have set strong constraints on sub-millimeter modifications of Newtonian gravity.

Several important new developments in experimental gravitation are expected in the near future. The approved NASA Gravity Probe B mission [59] (a space gyroscope experiment; due for launch in December 2003) will directly measure the gravitational spin-orbit and spin-spin couplings, thereby measuring the parameter $\overline{\gamma}$ to better than the 10^{-5} level. The universality of free-fall acceleration should soon be tested to much better than the 10^{-12} level by some satellite experiments: the approved CNES MICROSCOPE [60] mission (10^{-15} level), and the planned (cryogenic) NASA-ESA STEP [61] mission (10^{-18} level). The recently constructed kilometer-size laser interferometers (notably LIGO [62] in the USA, and VIRGO [63] and GEO600 [64] in Europe), should soon directly detect gravitational waves arriving on Earth. As the sources of these waves are expected to be extremely relativistic objects with strong internal gravitational fields (*e.g.*, coalescing binary black holes), their detection will allow one to experimentally probe gravity in highly dynamical circumstances. Note finally that arrays of millisecond pulsars are sensitive detectors of (very low frequency) gravitational waves [65,66].

References:

1. S.N. Gupta, Phys. Rev. **96**, 1683 (1954);
   R.H. Kraichnan, Phys. Rev. **98**, 1118 (1955);
   R.P. Feynman, F.B. Morinigo, and W.G. Wagner, *Feynman Lectures on Gravitation*, edited by Brian Hatfield (Addison-Wesley, Reading, 1995);
   S. Weinberg, Phys. Rev. **138**, B988 (1965);
   V.I. Ogievetsky and I.V. Polubarinov, Ann. Phys. (NY) **35**, 167 (1965);
   W. Wyss, Helv. Phys. Acta **38**, 469 (1965);
   S. Deser, Gen. Rel. Grav. **1**, 9 (1970);
   D.G. Boulware and S. Deser, Ann. Phys. (NY) **89**, 193 (1975);
   J. Fang and C. Fronsdal, J. Math. Phys. **20**, 2264 (1979);
   R.M. Wald, Phys. Rev. **D33**, 3613 (1986);
   C. Cutler and R.M. Wald, Class. Quantum Grav. **4**, 1267 (1987);
   R.M. Wald, Class. Quantum Grav. **4**, 1279 (1987);
   N. Boulanger *et al.*, Nucl. Phys. **B597**, 127 (2001).
2. S. Weinberg, *Gravitation and Cosmology* (John Wiley, New York, 1972).
3. S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989).
4. A.I. Shlyakhter, Nature **264**, 340 (1976);
   T. Damour and F. Dyson, Nucl. Phys. **B480**, 37 (1996);
   Y. Fujii *et al.*, Nucl. Phys. **B573**, 377 (2000).
5. J.D. Prestage, R.L. Tjoelker, and L. Maleki, Phys. Rev. Lett. **74**, 3511 (1995);
   Y. Sortais *et al.*, Physica Scripta **95**, 50 (2001).
6. J.K. Webb *et al.*, Phys. Rev. Lett. **87**, 091301 (2001);
   M.T. Murphy *et al.*, Mon. Not. Roy. Astron. Soc. **327**, 1208 (2001).
7. K.A. Olive and M. Pospelov, Phys. Rev. **D65**, 085044 (2002).
8. H.B. Sandvik, J.D. Barrow, and J. Magueijo, Phys. Rev. Lett. **88**, 031302 (2002).
9. T. Damour, F. Piazza, and G. Veneziano, Phys. Rev. Lett. **89**, 081601 (2002); and Phys. Rev. **D66**, 046007 (2002).
10. K.A. Olive *et al.*, Phys. Rev. **D66**, 045022 (2002).
11. M.T. Murphy *et al.*, Astrophys. Space Sci. **283**, 577 (2003).
12. J.N. Bahcall, C.L. Steinhardt, and D. Schlegel, `astro-ph/0301507`.
13. J.P. Uzan, Rev. Mod. Phys. **75**, 403 (2003).
14. J.D. Prestage *et al.*, Phys. Rev. Lett. **54**, 2387 (1985);
    S.K. Lamoreaux *et al.*, Phys. Rev. Lett. **57**, 3125 (1986);
    T.E. Chupp *et al.*, Phys. Rev. Lett. **63**, 1541 (1989).
15. T. Jacobson, S. Liberati, and D. Mattingly, Nature **424**, 1019 (2003).
16. R.C. Myers and M. Pospelov, Phys. Rev. Lett. **90**, 211601 (2003).
17. Y. Su *et al.*, Phys. Rev. **D50**, 3614 (1994).
18. J.O. Dickey *et al.*, Science **265**, 482 (1994);
    J.G. Williams, X.X. Newhall, and J.O. Dickey, Phys. Rev. **D53**, 6730 (1996).
19. S. Baessler *et al.*, Phys. Rev. Lett. **83**, 3585 (1999).
20. G.L. Smith *et al.*, Phys. Rev. **D61**, 022001 (1999).
21. R.F.C. Vessot and M.W. Levine, Gen. Rel. Grav. **10**, 181 (1978);
    R.F.C. Vessot *et al.*, Phys. Rev. Lett. **45**, 2081 (1980).
22. C.M. Will, *Theory and Experiment in Gravitational Physics* (Cambridge University Press, Cambridge, 1993); and Living Rev. Rel. **4**, 4 (2001).
23. T. Damour, in *Gravitation and Quantizations*, ed. B. Julia and J. Zinn-Justin, Les Houches, Session LVII (Elsevier, Amsterdam, 1995), pp. 1–61.
24. H. van Dam and M.J. Veltman, Nucl. Phys. **B22**, 397 (1970);
    V.I. Zakharov, Sov. Phys. JETP Lett. **12**, 312 (1970);
    D.G. Boulware and S. Deser, Phys. Rev. **D6**, 3368 (1972);
    C. Aragone and J. Chela-Flores, Nuovo Cim. **10A**, 818 (1972);
    A.I. Vainshtein, Phys. Lett. **B39**, 393 (1972);
    C. Deffayet *et al.*, Phys. Rev. **D65**, 044026 (2002);
    M. Porrati, Phys. Lett. **B534**, 209 (2002);
    N. Arkani-Hamed, H. Georgi, and M.D. Schwartz, Annals Phys. **305**, 96 (2003);
    T. Damour, I.I. Kogan, and A. Papazoglou, Phys. Rev. **D67**, 064009 (2003);
    G. Dvali, A. Gruzinov, and M. Zaldarriaga, Phys. Rev. **D68**, 024012 (2003).
25. T.R. Taylor and G. Veneziano, Phys. Lett. **B213**, 450 (1988);
    S. Dimopoulos and G. Giudice, Phys. Lett. **B379**, 105 (1996);
    I. Antoniadis, S. Dimopoulos, and G. Dvali, Nucl. Phys. **B516**, 70 (1998).
26. V.A. Rubakov, Phys. Usp **44**, 871 (2001);
    I.I. Kogan, in *Proceedings of the XXXVIth Rencontres de Moriond, Electro-Weak Interactions and Unified Theories* (March 2001); `astro-ph/0108220`.
27. E. Fischbach and C.L. Talmadge, *The search for non-Newtonian gravity* (Springer-Verlag, New York, 1999).
28. J.C. Long, H.W. Chan, and J.C. Price, Nucl. Phys. **B539**, 23 (1999).
29. C.D. Hoyle *et al.*, Phys. Rev. Lett. **86**, 1418 (2001);
    J. Chiaverini *et al.*, Phys. Rev. Lett. **90**, 151101 (2003);
    J.C. Long *et al.*, Nature **421**, 922 (2003).

30. P. Jordan, *Schwerkraft und Weltall* (Vieweg, Braunschweig, 1955); M. Fierz, Helv. Phys. Acta **29**, 128 (1956); C. Brans and R.H. Dicke, Phys. Rev. **124**, 925 (1961).
31. T. Damour and G. Esposito-Farèse, Class. Quantum Grav. **9**, 2093 (1992).
32. A.S. Eddington, *The Mathematical Theory of Relativity* (Cambridge University Press, Cambridge, 1923); K. Nordtvedt, Phys. Rev. **169**, 1017 (1968); C.M. Will, Astrophys. J. **163**, 611 (1971).
33. B. Bertotti, L. Iess and P. Tortora, Nature, **425**, 374 (2003).
34. K. Nordtvedt, Phys. Rev. **170**, 1186 (1968).
35. T.R. Taylor and G. Veneziano, Phys. Lett. **B213**, 450 (1988); T. Damour and A.M. Polyakov, Nucl. Phys. **B423**, 532 (1994).
36. R.A. Hulse, Rev. Mod. Phys. **66**, 699 (1994).
37. J.H. Taylor, Rev. Mod. Phys. **66**, 711 (1994).
38. T. Damour and N. Deruelle, Phys. Lett. **A87**, 81 (1981); T. Damour, C.R. Acad. Sci. Paris **294**, 1335 (1982).
39. J.H. Taylor, Class. Quantum Grav. **10**, S167 (Supplement 1993).
40. J. Weisberg and J.H. Taylor, in *Radio Pulsars, ASP Conference Series* **302**, 93 (2003); `astro-ph/0211217`.
41. By contrast to the binary pulsar case, which does involve the gauge-invariant, helicity-two propagating degrees of freedom of the gravitational field, the recent measurement of light deflection from Jupiter (E.B. Fomalont and S.M. Kopeikin, `astro-ph/0302294`) does not depend (in spite of contrary claims: S.M. Kopeikin, Astrophys. J. **556**, L1 (2001)), at the considered precision level, on the propagation speed of gravity (see C.M. Will, Astrophys. J. **590**, 683 (2003) and S. Samuel, Phys. Rev. Lett. **90**, 231101 (2003)).
42. T. Damour and J.H. Taylor, Phys. Rev. **D45**, 1840 (1992).
43. T. Damour and G. Esposito-Farèse, Phys. Rev. **D54**, 1474 (1996); and Phys. Rev. **D58**, 042001 (1998).
44. I.H. Stairs, Living Rev. Rel. **6** 5 (2003).
45. T. Damour and J.H. Taylor, Astrophys. J. **366**, 501 (1991).
46. A. Wolszczan, Nature **350**, 688 (1991).
47. J.H. Taylor *et al.*, Nature **355**, 132 (1992).
48. I.H. Stairs *et al.*, Astrophys. J. **505**, 352 (1998); I.H. Stairs *et al.*, Astrophys. J. **581**, 501 (2002).
49. V.M. Kaspi *et al.*, Astrophys. J. **528**, 445 (2000).
50. M. Bailes *et al.*, Astrophys. J. **595**, L49 (2003).
51. C.M. Will and H.W. Zaglauer, Astrophys. J. **346**, 366 (1989).
52. T. Damour and G. Schäfer, Phys. Rev. Lett. **66**, 2549 (1991).
53. Ch. Lange *et al.*, Mon. Not. Roy. Astron. Soc. **326**, 274 (2001).
54. Z. Arzoumanian, in *Radio Pulsars, ASP Conference Series* **302**, 69 (2003); `astro-ph/0212180`.
55. M. Kramer, Astrophys. J. **509**, 856 (1998); J.M. Weisberg and J.H. Taylor, Astrophys. J. **576**, 942 (2002).
56. T. Damour and R. Ruffini, C. R. Acad. Sc. Paris **279**, série A, 971 (1974); B.M. Barker and R.F. O'Connell, Phys. Rev. D 12, 329 (1975).
57. A. Dar, Nucl. Phys. (Proc. Supp.) **B28**, 321 (1992).
58. J. Yang *et al.*, Astrophys. J. **227**, 697 (1979); T. Rothman and R. Matzner, Astrophys. J. **257**, 450 (1982); F.S. Accetta, L.M. Krauss, and P. Romanelli, Phys. Lett. **B248**, 146 (1990).
59. `http://einstein.stanford.edu`.
60. P. Touboul *et al.*, C.R.Acad. Sci. Paris **2** (série IV) 1271 (2001).
61. P.W. Worden, in *Proc. 7th Marcel Grossmann Meeting on General Relativity*, edited by R.J. Jantzen and G. MacKeiser, (World Scientific, Singapore, 1995), pp. 1569-1573.
62. `http://www.ligo.caltech.edu` .
63. `http://www.virgo.infn.it` .
64. `http://www.geo600.uni-hannover.de` .
65. V.M. Kaspi, J.H. Taylor and M.F. Ryba, Astrophys. J. **428**, 713 (1994).
66. A.N. Lommen and D.C. Backer, Bulletin of the American Astronomical Society **33**, 1347 (2001); and Astrophys. J. **562**, 297 (2001).

19. BIG-BANG COSMOLOGY

Written July 2001 by K.A. Olive (University of Minnesota) and J.A. Peacock (University of Edinburgh). Revised November 2003.

19.1. Introduction to Standard Big-Bang Model

The observed expansion of the Universe [1,2,3] is a natural (almost inevitable) result of any homogeneous and isotropic cosmological model based on general relativity. However, by itself, the Hubble expansion does not provide sufficient evidence for what we generally refer to as the Big-Bang model of cosmology. While general relativity is in principle capable of describing the cosmology of any given distribution of matter, it is extremely fortunate that our Universe appears to be homogeneous and isotropic on large scales. Together, homogeneity and isotropy allow us to extend the Copernican Principle to the Cosmological Principle, stating that all spatial positions in the Universe are essentially equivalent.

The formulation of the Big-Bang model began in the 1940s with the work of George Gamow and his collaborators, Alpher and Herman. In order to account for the possibility that the abundances of the elements had a cosmological origin, they proposed that the early Universe which was once very hot and dense (enough so as to allow for the nucleosynthetic processing of hydrogen), and has expanded and cooled to its present state [4,5]. In 1948, Alpher and Herman predicted that a direct consequence of this model is the presence of a relic background radiation with a temperature of order a few K [6,7]. Of course this radiation was observed 16 years later as the microwave background radiation [8]. Indeed, it was the observation of the 3 K background radiation that singled out the Big-Bang model as the prime candidate to describe our Universe. Subsequent work on Big-Bang nucleosynthesis further confirmed the necessity of our hot and dense past. (See the review on BBN—Sec. 20 of this *Review* for a detailed discussion of BBN.) These relativistic cosmological models face severe problems with their initial conditions, to which the best modern solution is inflationary cosmology, discussed in Sec. 19.3.5. If correct, these ideas would strictly render the term 'Big Bang' redundant, since it was first coined by Hoyle to represent a criticism of the lack of understanding of the initial conditions.

19.1.1. *The Robertson-Walker Universe*:

The observed homogeneity and isotropy enable us to describe the overall geometry and evolution of the Universe in terms of two cosmological parameters accounting for the spatial curvature and the overall expansion (or contraction) of the Universe. These two quantities appear in the most general expression for a space-time metric which has a (3D) maximally symmetric subspace of a 4D space-time, known as the Robertson-Walker metric:

$$ds^2 = dt^2 - R^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right] . \qquad (19.1)$$

Note that we adopt $c = 1$ throughout. By rescaling the radial coordinate, we can choose the curvature constant k to take only the discrete values $+1$, -1, or 0 corresponding to closed, open, or spatially flat geometries. In this case, it is often more convenient to re-express the metric as

$$ds^2 = dt^2 - R^2(t)\left[d\chi^2 + S_k^2(\chi)\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right] , \qquad (19.2)$$

where the function $S_k(\chi)$ is $(\sin\chi, \chi, \sinh\chi)$ for $k = (+1, 0, -1)$. The coordinate r (in Eq. (19.1)) and the 'angle' χ (in Eq. (19.2)) are both dimensionless; the dimensions are carried by $R(t)$, which is the cosmological scale factor which determines proper distances in terms of the comoving coordinates. A common alternative is to define a dimensionless scale factor, $a(t) = R(t)/R_0$, where $R_0 \equiv R(t_0)$ is R at the present epoch. It is also sometimes convenient to define a dimensionless or conformal time coordinate, η, by $d\eta = dt/R(t)$. Along constant spatial sections, the proper time is defined by the time coordinate, t. Similarly, for $dt = d\theta = d\phi = 0$, the proper distance is given by $R(t)\chi$. For standard texts on cosmological models see *e.g.*, Refs. [9–14].

19.1.2. *The redshift*:

The cosmological redshift is a direct consequence of the Hubble expansion, determined by $R(t)$. A local observer detecting light from a distant emitter sees a redshift in frequency. We can define the redshift as

$$z \equiv \frac{\nu_1 - \nu_2}{\nu_2} \simeq \frac{v_{12}}{c} , \qquad (19.3)$$

where ν_1 is the frequency of the emitted light, ν_2 is the observed frequency and v_{12} is the relative velocity between the emitter and the observer. While the definition, $z = (\nu_1 - \nu_2)/\nu_2$ is valid on all distance scales, relating the redshift to the relative velocity in this simple way is only true on small scales (*i.e.*, less than cosmological scales) such that the expansion velocity is non-relativistic. For light signals, we can use the metric given by Eq. (19.1) and $ds^2 = 0$ to write

$$\frac{v_{12}}{c} = \dot{R}\,\delta r = \frac{\dot{R}}{R}\,\delta t = \frac{\delta R}{R} = \frac{R_2 - R_1}{R_1} , \qquad (19.4)$$

where $\delta r(\delta t)$ is the radial coordinate (temporal) separation between the emitter and observer. Thus, we obtain the simple relation between the redshift and the scale factor

$$1 + z = \frac{\nu_1}{\nu_2} = \frac{R_2}{R_1} . \qquad (19.5)$$

This result does not depend on the non-relativistic approximation.

19.1.3. *The Friedmann-Lemaître equations of motion*:

The cosmological equations of motion are derived from Einstein's equations

$$\mathcal{R}_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu}\mathcal{R} = 8\pi G_{\rm N} T_{\mu\nu} + \Lambda g_{\mu\nu} . \qquad (19.6)$$

Gliner [15] and Zeldovich [16] seem to have pioneered the modern view, in which the Λ term is taken to the rhs and interpreted as particle-physics processes yielding an effective energy–momentum tensor $T_{\mu\nu}$ for the vacuum of $\Lambda g_{\mu\nu}/8\pi G_{\rm N}$. It is common to assume that the matter content of the Universe is a perfect fluid, for which

$$T_{\mu\nu} = -p g_{\mu\nu} + (p + \rho)\, u_\mu u_\nu , \qquad (19.7)$$

where $g_{\mu\nu}$ is the space-time metric described by Eq. (19.1), p is the isotropic pressure, ρ is the energy density and $u = (1, 0, 0, 0)$ is the velocity vector for the isotropic fluid in co-moving coordinates. With the perfect fluid source, Einstein's equations lead to the Friedmann-Lemaître equations

$$H^2 \equiv \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi\ G_{\rm N}\,\rho}{3} - \frac{k}{R^2} + \frac{\Lambda}{3} , \qquad (19.8)$$

and

$$\frac{\ddot{R}}{R} = \frac{\Lambda}{3} - \frac{4\pi G_{\rm N}}{3}\ (\rho + 3p) , \qquad (19.9)$$

where $H(t)$ is the Hubble parameter and Λ is the cosmological constant. The first of these is sometimes called the Friedmann equation. Energy conservation via $T^{\mu\nu}{}_{;\mu} = 0$, leads to a third useful equation [which can also be derived from Eq. (19.8) and Eq. (19.9)]

$$\dot{\rho} = -3H\,(\rho + p) . \qquad (19.10)$$

Eq. (19.10) can also be simply derived as a consequence of the first law of thermodynamics.

Eq. (19.8) has a simple classical mechanical analog if we neglect (for the moment) the cosmological term Λ. By interpreting $-k/R^2$ as a "total energy", then we see that the evolution of the Universe is governed by a competition between the potential energy, $8\pi G_{\rm N}\rho/3$ and the kinetic term $(\dot{R}/R)^2$. For $\Lambda = 0$, it is clear that the Universe must be expanding or contracting (except at the turning point prior to collapse in a closed Universe). The ultimate fate of the Universe is determined by the curvature constant k. For $k = +1$, the Universe will recollapse in a finite time, whereas for $k = 0, -1$, the Universe will expand indefinitely. These simple conclusions can be altered when $\Lambda \neq 0$ or more generally with some component with $(\rho + 3p) < 0$.

19.1.4. ***Definition of cosmological parameters***:

In addition to the Hubble parameter, it is useful to define several other measurable cosmological parameters. The Friedmann equation can be used to define a critical density such that $k = 0$ when $\Lambda = 0$,

$$\rho_c \equiv \frac{3H^2}{8\pi G_N} = 1.88 \times 10^{-26}\, h^2 \text{ kg m}^{-3} = 1.05 \times 10^{-5}\, h^2 \text{ GeV cm}^{-3} , \quad (19.11)$$

where the scaled Hubble parameter, h, is defined by

$$H \equiv 100\, h \text{ km s}^{-1} \text{ Mpc}^{-1} \Rightarrow H^{-1} = 9.78\, h^{-1} \text{ Gyr} = 2998\, h^{-1} \text{ Mpc} . \quad (19.12)$$

The cosmological density parameter $\Omega_{\rm tot}$ is defined as the energy density relative to the critical density,

$$\Omega_{\rm tot} = \rho/\rho_c . \quad (19.13)$$

Note that one can now rewrite the Friedmann equation as

$$k/R^2 = H^2(\Omega_{\rm tot} - 1) , \quad (19.14)$$

From Eq. (19.14), one can see that when $\Omega_{\rm tot} > 1$, $k = +1$ and the Universe is closed, when $\Omega_{\rm tot} < 1$, $k = -1$ and the Universe is open, and when $\Omega_{\rm tot} = 1$, $k = 0$, and the Universe is spatially flat.

It is often necessary to distinguish different contributions to the density. It is therefore convenient to define present-day density parameters for pressureless matter (Ω_m) and relativistic particles (Ω_r), plus the quantity $\Omega_\Lambda = \Lambda/3H^2$. In more general models, we may wish to drop the assumption that the vacuum energy density is constant, and we therefore denote the present-day density parameter of the vacuum by Ω_v. The Friedmann equation then becomes

$$k/R_0^2 = H_0^2(\Omega_m + \Omega_r + \Omega_v - 1) , \quad (19.15)$$

where the subscript 0 indicates present-day values. Thus, it is the sum of the densities in matter, relativistic particles and vacuum that determines the overall sign of the curvature. Note that the quantity $-k/R_0^2H_0^2$ is sometimes referred to as Ω_k. This usage is unfortunate: it encourages one to think of curvature as a contribution to the energy density of the Universe, which is not correct.

19.1.5. ***Standard Model solutions***:

Much of the history of the Universe in the standard Big-Bang model can be easily described by assuming that either matter or radiation dominates the total energy density. During inflation or perhaps even today if we are living in an accelerating Universe, domination by a cosmological constant or some other form of dark energy should be considered. In the following, we shall delineate the solutions to the Friedmann equation when a single component dominates the energy density. Each component is distinguished by an equation of state parameter $w = p/\rho$.

19.1.5.1. *Solutions for a general equation of state:*

Let us first assume a general equation of state parameter for a single component, w which is constant. In this case, Eq. (19.10) can be written as $\dot\rho = -3(1+w)\rho\dot R/R$ and is easily integrated to yield

$$\rho \propto R^{-3(1+w)} . \quad (19.16)$$

Note that at early times when R is small, the less singular curvature term k/R^2 in the Friedmann equation can be neglected so long as $w > -1/3$. Curvature domination occurs at rather late times (if a cosmological constant term does not dominate sooner). For $w \neq -1$, one can insert this result into the Friedmann equation Eq. (19.8) and if one neglects the curvature and cosmological constant terms, it is easy to integrate the equation to obtain,

$$R(t) \propto t^{2/[3(1+w)]} . \quad (19.17)$$

19.1.5.2. *A Radiation-dominated Universe:*

In the early hot and dense Universe, it is appropriate to assume an equation of state corresponding to a gas of radiation (or relativistic particles) for which $w = 1/3$. In this case, Eq. (19.16) becomes $\rho \propto R^{-4}$. The "extra" factor of $1/R$ is due to the cosmological redshift; not only is the number density of particles in the radiation background decreasing as R^{-3} since volumes scales as R^3, but in addition, each particle's energy is decreasing as $E \propto \nu \propto R^{-1}$. Similarly, one can substitute $w = 1/3$ into Eq. (19.17) to obtain

$$R(t) \propto t^{1/2} ; \qquad H = 1/2t . \quad (19.18)$$

19.1.5.3. *A Matter-dominated Universe:*

At relatively late times, non-relativistic matter eventually dominates the energy density over radiation (see Sec. 19.3.8). A pressureless gas ($w = 0$) leads to the expected dependence $\rho \propto R^{-3}$ from Eq. (19.16) and, if $k = 0$, we get

$$R(t) \propto t^{2/3} ; \qquad H = 2/3t . \quad (19.19)$$

19.1.5.4. *A Universe dominated by vacuum energy:*

If there is a dominant source of vacuum energy, V_0, it would act as a cosmological constant with $\Lambda = 8\pi G_N V_0$ and equation of state $w = -1$. In this case, the solution to the Friedmann equation is particularly simple and leads to an exponential expansion of the Universe

$$R(t) \propto e^{\sqrt{\Lambda/3}t} . \quad (19.20)$$

A key parameter is the equation of state of the vacuum, $w \equiv p/\rho$: this need not be the $w = -1$ of Λ, and may not even be constant [17,18,19]. It is now common to use w to stand for this vacuum equation of state, rather than of any other constituent of the Universe, and we use the symbol in this sense hereafter. We generally assume w to be independent of time, and where results relating to the vacuum are quoted without an explicit w dependence, we have adopted $w = -1$.

The presence of vacuum energy can dramatically alter the fate of the Universe. For example, if $\Lambda < 0$, the Universe will eventually recollapse independent of the sign of k. For large values of Λ (larger than the Einstein static value needed to halt any cosmological expansion or contraction), even a closed Universe will expand forever. One way to quantify this is the deceleration parameter, q_0, defined as

$$q_0 = -\left.\frac{R\ddot R}{\dot R^2}\right|_0 = \frac{1}{2}\Omega_m + \Omega_r + \frac{(1+3w)}{2}\Omega_v . \quad (19.21)$$

This equation shows us that $w < -1/3$ for the vacuum may lead to an accelerating expansion. Astonishingly, it appears that such an effect has been observed in the Supernova Hubble diagram [20–23] (see Fig. 19.1 below); current data indicate that vacuum energy is indeed the largest contributor to the cosmological density budget, with $\Omega_v = 0.72 \pm 0.05$ and $\Omega_m = 0.28 \pm 0.05$ if $k = 0$ is assumed [23].

The nature of this dominant term is presently uncertain, but much effort is being invested in dynamical models (*e.g.*, rolling scalar fields), under the catch-all heading of "quintessence."

19.2. Introduction to Observational Cosmology

19.2.1. ***Fluxes, luminosities, and distances***:

The key quantities for observational cosmology can be deduced quite directly from the metric.

(1) The *proper* transverse size of an object seen by us to subtend an angle $d\psi$ is its comoving size $d\psi\, S_k(\chi)$ times the scale factor at the time of emission:

$$d\ell = d\psi\ R_0 S_k(\chi)/(1+z) . \quad (19.22)$$

(2) The apparent flux density of an object is deduced by allowing its photons to flow through a sphere of current radius $R_0 S_k(\chi)$; but photon energies and arrival rates are redshifted, and the bandwidth $d\nu$ is reduced. The observed photons at frequency ν_0 were emitted at frequency $\nu_0(1+z)$, so the flux density is the luminosity at this frequency, divided by the total area, divided by $1+z$:

$$S_\nu(\nu_0) = \frac{L_\nu([1+z]\nu_0)}{4\pi R_0^2 S_k^2(\chi)(1+z)} \ . \tag{19.23}$$

These relations lead to the following common definitions:

$$\begin{aligned} \text{angular-diameter distance:} \quad & D_{\rm A} = (1+z)^{-1} R_0 S_k(\chi) \\ \text{luminosity distance:} \quad & D_{\rm L} = (1+z)\, R_0 S_k(\chi) \end{aligned} \tag{19.24}$$

These distance-redshift relations are expressed in terms of observables by using the equation of a null radial geodesic ($R(t)d\chi = dt$) plus the Friedmann equation:

$$\begin{aligned} R_0 d\chi = & \frac{1}{H(z)}\, dz = \frac{1}{H_0} \Big[(1 - \Omega_m - \Omega_v - \Omega_r)(1+z)^2 \\ & + \Omega_v (1+z)^{3+3w} \ + \ \Omega_m (1+z)^3 + \Omega_r (1+z)^4 \Big]^{-1/2} \, dz \ . \end{aligned} \tag{19.25}$$

The main scale for the distance here is the Hubble length, $1/H_0$.

The flux density is the product of the specific intensity I_ν and the solid angle $d\Omega$ subtended by the source: $S_\nu = I_\nu \, d\Omega$. Combining the angular size and flux-density relations thus gives the relativistic version of surface-brightness conservation:

$$I_\nu(\nu_0) = \frac{B_\nu([1+z]\nu_0)}{(1+z)^3} \ , \tag{19.26}$$

where B_ν is surface brightness (luminosity emitted into unit solid angle per unit area of source). We can integrate over ν_0 to obtain the corresponding total or bolometric formula:

$$I_{\rm tot} = \frac{B_{\rm tot}}{(1+z)^4} \ . \tag{19.27}$$

This cosmology-independent form expresses Liouville's Theorem: photon phase-space density is conserved along rays.

19.2.2. *Distance data and geometrical tests of cosmology*:

In order to confront these theoretical predictions with data, we have to bridge the divide between two extremes. Nearby objects may have their distances measured quite easily, but their radial velocities are dominated by deviations from the ideal Hubble flow, which typically have a magnitude of several hundred km s^{-1}. On the other hand, objects at redshifts $z \gtrsim 0.01$ will have observed recessional velocities that differ from their ideal values by $\lesssim 10\%$, but absolute distances are much harder to supply in this case. The traditional solution to this problem is the construction of the distance ladder: an interlocking set of methods for obtaining relative distances between various classes of object, which begins with absolute distances at the 10 to 100 pc level and terminates with galaxies at significant redshifts. This is reviewed in the review on Global cosmological parameters—Sec. 21 of this *Review*.

By far the most exciting development in this area has been the use of type Ia Supernovae (SNe), which now allow measurement of relative distances with 5% precision. In combination with Cepheid data from the HST key project on the distance scale, SNe results are the dominant contributor to the best modern value for H_0: 72 km s^{-1}Mpc^{-1} $\pm$ 10% [24]. Better still, the analysis of high-z SNe has allowed the first meaningful test of cosmological geometry to be carried out: as shown in Fig. 19.1 and Fig. 19.2, a combination of supernova data and measurements of microwave-background anisotropies strongly favors a $k = 0$ model dominated by vacuum energy. (See the review on Global cosmological parameters—Sec. 21 of this *Review* for a more comprehensive review of Hubble parameter determinations.)

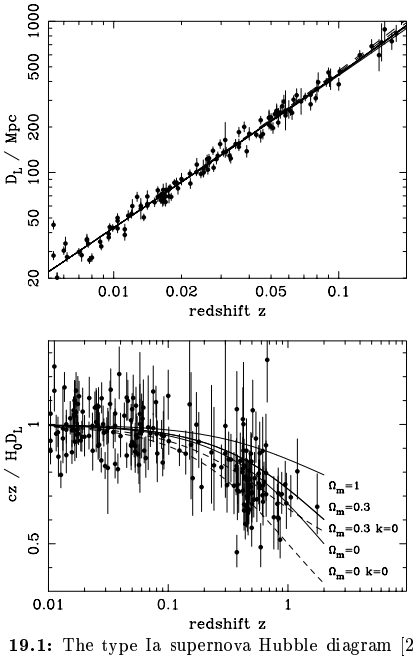

Figure 19.1: The type Ia supernova Hubble diagram [20–22]. The first panel shows that for $z \ll 1$ the large-scale Hubble flow is indeed linear and uniform; the second panel shows an expanded scale, with the linear trend divided out, and with the redshift range extended to show how the Hubble law becomes nonlinear. ($\Omega_r = 0$ is assumed.) Comparison with the prediction of Friedmann-Lemaître models appears to favor a vacuum-dominated Universe.

19.2.3. *Age of the Universe*:

The most striking conclusion of relativistic cosmology is that the Universe has not existed forever. The dynamical result for the age of the Universe may be written as

$$\begin{aligned} H_0 t_0 &= \int_0^\infty \frac{dz}{(1+z)H(z)} \\ &= \int_0^\infty \frac{dz}{(1+z)\left[(1+z)^2(1+\Omega_m z) - z(2+z)\Omega_v\right]^{1/2}} \ , \end{aligned} \tag{19.28}$$

where we have neglected Ω_r and chosen $w = -1$. Over the range of interest ($0.1 \lesssim \Omega_m \lesssim 1$, $|\Omega_v| \lesssim 1$), this exact answer may be approximated to a few % accuracy by

$$H_0 t_0 \simeq \tfrac{2}{3}\,(0.7\Omega_m + 0.3 - 0.3\Omega_v)^{-0.3} \ . \tag{19.29}$$

For the special case that $\Omega_m + \Omega_v = 1$, the integral in Eq. (19.28) can be expressed analytically as

$$H_0 t_0 = \frac{2}{3\sqrt{\Omega_v}} \ln \frac{1+\sqrt{\Omega_v}}{\sqrt{1-\Omega_v}} \quad (\Omega_m < 1) \ . \tag{19.30}$$

The most accurate means of obtaining ages for astronomical objects is based on the natural clocks provided by radioactive decay. The use of these clocks is complicated by a lack of knowledge of the initial conditions of the decay. In the Solar System, chemical fractionation of different elements helps pin down a precise age for the pre-Solar nebula of 4.6 Gyr, but for stars it is necessary to attempt an a priori calculation of the relative abundances of nuclei that result from supernova explosions. In this way, a lower limit for the age of stars in the local part of the Milky Way of about 11 Gyr is obtained [25].

The other major means of obtaining cosmological age estimates is based on the theory of stellar evolution. In principle, the main-sequence turnoff point in the color-magnitude diagram of a globular cluster should yield a reliable age. However, these have been controversial owing to theoretical uncertainties in the evolution model, as well as observational uncertainties in the distance, dust extinction and metallicity of clusters. The present consensus favors ages for the oldest clusters of about 12 Gyr [26,27].

These methods are all consistent with the age deduced from studies of structure formation, using the microwave background and large-scale structure: $t_0 = 13.7 \pm 0.2$ Gyr [28], where the extra accuracy comes at the price of assuming the Cold Dark Matter model to be true.

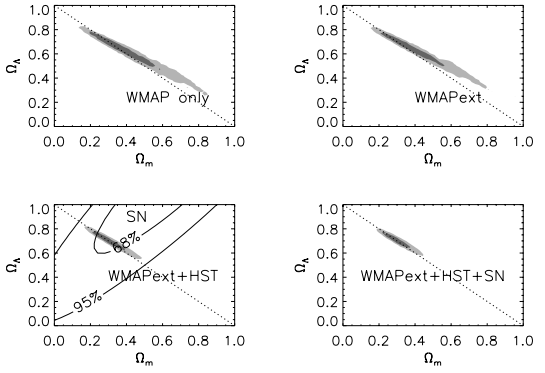

Figure 19.2: Likelihood-based confidence contours [28] over the plane Ω_Λ (*i.e.* Ω_v assuming $w = -1$) vs Ω_m. The SNe Ia results very nearly constrain $\Omega_v - \Omega_m$, whereas the results of CMB anisotropies (from the first-year WMAP data) favor a flat model with $\Omega_v + \Omega_m \simeq 1$. The intersection of these constraints is the most direct (but far from the only) piece of evidence favoring a flat model with $\Omega_m \simeq 0.3$.

19.2.4. *Horizon, isotropy, flatness problems*:

For photons, the radial equation of motion is just $c\,dt = R\,d\chi$. How far can a photon get in a given time? The answer is clearly

$$\Delta\chi = \int_{t_1}^{t_2} \frac{dt}{R(t)} \equiv \Delta\eta \ , \tag{19.31}$$

i.e., just the interval of conformal time. We can replace dt by $dR/\dot{R}$, which the Friedmann equation says is $\propto dR/\sqrt{\rho R^2}$ at early times. Thus, this integral converges if $\rho R^2 \to \infty$ as $t_1 \to 0$, otherwise it diverges. Provided the equation of state is such that ρ changes faster than R^{-2}, light signals can only propagate a finite distance between the Big Bang and the present; there is then said to be a particle horizon. Such a horizon therefore exists in conventional Big-Bang models, which are dominated by radiation ($\rho \propto R^{-4}$) at early times.

At late times, the integral for the horizon is largely determined by the matter-dominated phase, for which

$$D_{\rm H} = R_0\,\chi_{\rm H} \equiv R_0 \int_0^{t(z)} \frac{dt}{R(t)} \simeq \frac{6000}{\sqrt{\Omega z}}\,h^{-1}\,{\rm Mpc} \quad (z \gg 1) \ . \tag{19.32}$$

The horizon at the time of formation of the microwave background ('last scattering': $z \simeq 1100$) was thus of order 100 Mpc in size, subtending an angle of about 1°. Why then are the large number of causally disconnected regions we see on the microwave sky all at the same temperature? The Universe is very nearly isotropic and homogeneous, even though the initial conditions appear not to permit such a state to be constructed.

A related problem is that the $\Omega = 1$ Universe is unstable:

$$\Omega(a) - 1 = \frac{\Omega - 1}{1 - \Omega + \Omega_v a^2 + \Omega_m a^{-1} + \Omega_r a^{-2}} \ , \tag{19.33}$$

where Ω with no subscript is the total density parameter, and $a(t) = R(t)/R_0$. This requires $\Omega(t)$ to be unity to arbitrary precision as the initial time tends to zero; a universe of non-zero curvature today requires very finely tuned initial conditions.

19.3. The Hot Thermal Universe

19.3.1. *Thermodynamics of the early Universe*:

As alluded to above, we expect that much of the early Universe can be described by a radiation-dominated equation of state. In addition, through much of the radiation-dominated period, thermal equilibrium is established by the rapid rate of particle interactions relative to the expansion rate of the Universe (see Sec. 19.3.3 below). In equilibrium, it is straightforward to compute the thermodynamic quantities, ρ, p, and the entropy density, s. In general, the energy density for a given particle type i can be written as

$$\rho_i = \int E_i\,dn_{q_i} \ , \tag{19.34}$$

with the density of states given by

$$dn_{q_i} = \frac{g_i}{2\pi^2}\left(\exp[(E_{q_i} - \mu_i)/T_i] \pm 1\right)^{-1}\ q_i^2\,dq_i \ , \tag{19.35}$$

where g_i counts the number of degrees of freedom for particle type i, $E_{q_i}^2 = m_i^2 + q_i^2$, μ_i is the chemical potential, and the $\pm$ corresponds to either Fermi or Bose statistics. Similarly, we can define the pressure of a perfect gas as

$$p_i = \frac{1}{3}\int \frac{q_i^2}{E_i}\,dn_{q_i} \ . \tag{19.36}$$

The number density of species i is simply

$$n_i = \int dn_{q_i} \ , \tag{19.37}$$

and the entropy density is

$$s_i = \frac{\rho_i + p_i - \mu_i n_i}{T_i} \ , \tag{19.38}$$

In the Standard Model, a chemical potential is often associated with baryon number, and since the net baryon density relative to the photon density is known to be very small (of order 10^{-10}), we can neglect the chemical potential.

For photons, we can compute all of the thermodynamic quantities rather easily. Taking $g_i = 2$ for the 2 photon polarization states, we have

$$\rho_\gamma = \frac{\pi^2}{15}T^4 \ ; \quad p_\gamma = \frac{1}{3}\rho_\gamma \ ; \quad s_\gamma = \frac{4\rho_\gamma}{3T} \ ; \quad n_\gamma = \frac{2\zeta(3)}{\pi^2}T^3 \ , \tag{19.39}$$

with $2\zeta(3)/\pi^2 \simeq 0.2436$. Note that Eq. (19.10) can be converted into an equation for entropy conservation. Recognizing that $\dot{p} = s\dot{T}$, Eq. (19.10) becomes

$$d(sR^3)/dt = 0 \ . \tag{19.40}$$

For radiation, this corresponds to the relationship between expansion and cooling, $T \propto R^{-1}$ in an adiabatically expanding Universe. Note also that both s and n_γ scale as T^3.

19.3.2. *Radiation content of the Early Universe*:

At the very high temperatures associated with the early Universe, massive particles are pair produced, and are part of the thermal bath. If for a given particle species i we have $T \gg m_i$, then we can neglect the mass in Eq. (19.34) to Eq. (19.38), and the thermodynamic quantities are easily computed as in Eq. (19.39). In general, we can approximate the energy density (at high temperatures) by including only those particles with $m_i \ll T$. In this case, we have

$$\rho = \left(\sum_B g_B + \frac{7}{8}\sum_F g_F\right)\frac{\pi^2}{30}T^4 \equiv \frac{\pi^2}{30}\,N(T)\,T^4 \ , \tag{19.41}$$

Temperature	New Particles	$4N(T)$
$T < m_e$	γ's + ν's	29
$m_e < T < m_\mu$	$e^\pm$	43
$m_\mu < T < m_\pi$	$\mu^\pm$	57
$m_\pi < T < T_c$ †	π's	69
$T_c < T < m_{\rm strange}$	π's + u, $\bar{u}$, d, $\bar{d}$ + gluons	205
$m_s < T < m_{\rm charm}$	s, $\bar{s}$	247
$m_c < T < m_\tau$	c, $\bar{c}$	289
$m_\tau < T < m_{\rm bottom}$	$\tau^\pm$	303
$m_b < T < m_{\rm W,Z}$	b, $\bar{b}$	345
$m_{W,Z} < T < m_{\rm Higgs}$	$W^\pm$, Z	381
$m_H < T < m_{\rm top}$	H^0	385
$m_t < T$	t, $\bar{t}$	427

†T_c corresponds to the confinement-deconfinement transition between quarks and hadrons.

where $g_{B(F)}$ is the number of degrees of freedom of each boson (fermion) and the sum runs over all boson and fermion states with $m \ll T$. The factor of 7/8 is due to the difference between the Fermi and Bose integrals. Eq. (19.41) defines the effective number of degrees of freedom, $N(T)$, by taking into account new particle degrees of freedom as the temperature is raised. This quantity is plotted in Fig. 19.3 [29].

The value of $N(T)$ at any given temperature depends on the particle physics model. In the standard SU(3) × SU(2) × U(1) model, we can specify $N(T)$ up to temperatures of O(100) GeV. The change in N (ignoring mass effects) can be seen in the following table.

At higher temperatures, $N(T)$ will be model dependent. For example, in the minimal SU(5) model, one needs to add 24 states to $N(T)$ for the X and Y gauge bosons, another 24 from the adjoint Higgs, and another 6 (in addition to the 4 already counted in $W^\pm$, Z, and H) from the $\overline{5}$ of Higgs. Hence for $T > m_{\rm X}$ in minimal SU(5), $N(T) = 160.75$. In a supersymmetric model this would at least double, with some changes possibly necessary in the table if the lightest supersymmetric particle has a mass below m_t.

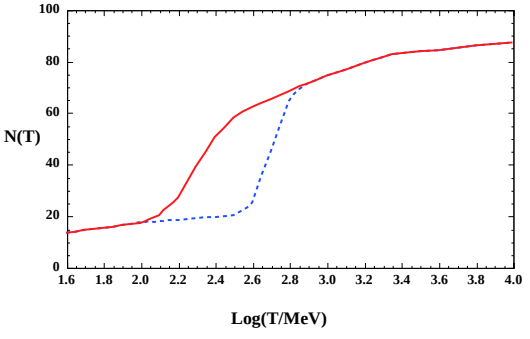

Figure 19.3: The effective numbers of relativistic degrees of freedom as a function of temperature. The sharp drop corresponds to the quark-hadron transition. The solid curve assume a QCD scale of 150 MeV, while the dashed curve assumes 450 MeV.

In the radiation-dominated epoch, Eq. (19.10) can be integrated (neglecting the T-dependence of N) giving us a relationship between the age of the Universe and its temperature

$$t = \left(\frac{90}{32\pi^3 G_{\rm N} N(T)}\right)^{1/2} T^{-2} \ , \tag{19.42}$$

Put into a more convenient form

$$t\, T_{\rm MeV}^2 = 2.4[N(T)]^{-1/2} \ , \tag{19.43}$$

where t is measured in seconds and $T_{\rm MeV}$ in units of MeV.

19.3.3. *Neutrinos and equilibrium*:

Due to the expansion of the Universe, certain rates may be too slow to either establish or maintain equilibrium. Quantitatively, for each particle i, as a minimal condition for equilibrium, we will require that some rate Γ_i involving that type be larger than the expansion rate of the Universe or

$$\Gamma_i > H \ . \tag{19.44}$$

Recalling that the age of the Universe is determined by H^{-1}, this condition is equivalent to requiring that on average, at least one interaction has occurred over the lifetime of the Universe.

A good example for a process which goes in and out of equilibrium is the weak interactions of neutrinos. On dimensional grounds, one can estimate the thermally averaged scattering cross section

$$\langle \sigma v \rangle \sim O(10^{-2}) T^2 / m_{\rm W}^4 \tag{19.45}$$

for $T \lesssim m_{\rm W}$. Recalling that the number density of leptons is $n \propto T^3$, we can compare the weak interaction rate, $\Gamma \sim n \langle \sigma v \rangle$, with the expansion rate,

$$\begin{aligned} H = \left(\frac{8\pi G_{\rm N}\rho}{3}\right)^{1/2} &= \left(\frac{8\pi^3}{90} N(T)\right)^{1/2} T^2/M_{\rm P} \\ &\sim 1.66 N(T)^{1/2} T^2/M_{\rm P}. \end{aligned} \tag{19.46}$$

The Planck mass $M_{\rm P} = G_{\rm N}^{-1/2} = 1.22 \times 10^{19}$ GeV.

Neutrinos will be in equilibrium when $\Gamma_{\rm wk} > H$ or

$$T > (500\, m_{\rm W}^4/M_{\rm P})^{1/3} \sim 1 \text{ MeV} \ . \tag{19.47}$$

The temperature at which these rates are equal is commonly referred to as the neutrino decoupling or freeze-out temperature and is defined by $\Gamma(T_d) = H(T_d)$.

At very high temperatures, the Universe is too young for equilibrium to have been established. For $T \gg m_{\rm W}$, we should write $\langle \sigma v \rangle \sim O(10^{-2})/T^2$, so that $\Gamma \sim 10^{-2} T$. Thus at temperatures $T \gtrsim 10^{-2} M_{\rm P}/\sqrt{N}$, equilibrium will not have been established.

For $T < T_d$, neutrinos drop out of equilibrium. The Universe becomes transparent to neutrinos and their momenta simply redshift with the cosmic expansion. The effective neutrino temperature will simply fall with $T \sim 1/R$.

Soon after decoupling, $e^\pm$ pairs in the thermal background begin to annihilate (when $T \lesssim m_e$). Because the neutrinos are decoupled, the energy released due to annihilation heats up the photon background relative to the neutrinos. The change in the photon temperature can be easily computed from entropy conservation. The neutrino entropy must be conserved separately from the entropy of interacting particles. A straightforward computation yields

$$T_\nu = (4/11)^{1/3} T_\gamma \simeq 1.9 \text{ K} \ . \tag{19.48}$$

Today, the total entropy density is therefore given by

$$s = \frac{4}{3}\frac{\pi^2}{30}\left(2 + \frac{21}{4}(T_\nu/T_\gamma)^3\right) T_\gamma^3 = \frac{4}{3}\frac{\pi^2}{30}\left(2 + \frac{21}{11}\right) T_\gamma^3 = 7.04\, n_\gamma \ . \tag{19.49}$$

Similarly, the total relativistic energy density today is given by

$$\rho_r = \frac{\pi^2}{30}\left[2 + \frac{21}{4}(T_\nu/T_\gamma)^4\right] T_\gamma^4 \simeq 1.68 \rho_\gamma \ . \tag{19.50}$$

In practice, a small correction is needed to this, since neutrinos are not totally decoupled at $e^\pm$ annihilation: the effective number of massless neutrino species is 3.04, rather than 3 [30].

This expression ignores neutrino rest masses, but current oscillation data require at least one neutrino eigenstate to have a mass exceeding 0.05 eV. In this minimal case, $\Omega_\nu h^2 = 5 \times 10^{-4}$, so the neutrino contribution to the matter budget would be negligibly small (which is our normal assumption). However, a nearly degenerate pattern of mass eigenstates could allow larger densities, since oscillation experiments only measure differences in m^2 values. Note that a 0.05-eV neutrino becomes non-relativistic at redshift $z = 2.2$, so the above expression for the total present relativistic density is really only an extrapolation. However, neutrinos are almost certainly relativistic at all epochs where the radiation content of the universe is dynamically significant.

19.3.4. ***Field Theory and Phase transitions***:

It is very likely that the Universe has undergone one or more phase transitions during the course of its evolution [31–34]. Our current vacuum state is described by $SU(3)_c \times U(1)_{em}$ which in the Standard Model is a remnant of an unbroken $SU(3)_c \times SU(2)_L \times U(1)_Y$ gauge symmetry. Symmetry breaking occurs when a non-singlet gauge field (the Higgs field in the Standard Model) picks up a non-vanishing vacuum expectation value, determined by a scalar potential. For example, a simple (non-gauged) potential describing symmetry breaking is $V(\phi) = \frac{1}{4}\lambda\phi^4 - \frac{1}{2}\mu^2\phi^2 + V(0)$. The resulting expectation value is simply $\langle\phi\rangle = \mu/\sqrt{\lambda}$.

In the early Universe, finite temperature radiative corrections typically add terms to the potential of the form $\phi^2 T^2$. Thus, at very high temperatures, the symmetry is restored and $\langle\phi\rangle = 0$. As the Universe cools, depending on the details of the potential, symmetry breaking will occur via a first order phase transition in which the field tunnels through a potential barrier, or via a second order transition in which the field evolves smoothly from one state to another (as would be the case for the above example potential).

The evolution of scalar fields can have a profound impact on the early Universe. The equation of motion for a scalar field ϕ can be derived from the energy-momentum tensor

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - \frac{1}{2}g_{\mu\nu}\partial_\rho\phi\partial^\rho\phi - g_{\mu\nu}V(\phi)\ . \tag{19.51}$$

By associating $\rho = T_{00}$ and $p = R^{-2}(t)T_{ii}$ we have

$$\begin{aligned} \rho &= \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}R^{-2}(t)(\nabla\phi)^2 + V(\phi) \\ p &= \frac{1}{2}\dot{\phi}^2 - \frac{1}{6}R^{-2}(t)(\nabla\phi)^2 - V(\phi) \end{aligned} \tag{19.52}$$

and from Eq. (19.10) we can write the equation of motion (by considering a homogeneous region, we can ignore the gradient terms)

$$\ddot{\phi} + 3H\dot{\phi} = -\partial V/\partial\phi\ . \tag{19.53}$$

19.3.5. ***Inflation***:

In Sec. 19.2.4, we discussed some of the problems associated with the standard Big-Bang model. However, during a phase transition, our assumptions of an adiabatically expanding universe are generally not valid. If, for example, a phase transition occurred in the early Universe such that the field evolved slowly from the symmetric state to the global minimum, the Universe may have been dominated by the vacuum energy density associated with the potential near $\phi \approx 0$. During this period of slow evolution, the energy density due to radiation will fall below the vacuum energy density, $\rho \ll V(0)$. When this happens, the expansion rate will be dominated by the constant V(0) and we obtain the exponentially expanding solution given in Eq. (19.20). When the field evolves towards the global minimum it will begin to oscillate about the minimum, energy will be released during its decay and a hot thermal universe will be restored. If released fast enough, it will produce radiation at a temperature $NT_R{}^4 \lesssim V(0)$. In this reheating process entropy has been created and the final value of RT is greater than the initial value of RT. Thus, we see that, during a phase transition, the relation $RT \sim$ constant need not hold true. This is the basis of the inflationary Universe scenario [35–37].

If during the phase transition the value of RT changed by a factor of $O(10^{29})$, the cosmological problems discussed above would be solved. The observed isotropy would be generated by the immense expansion; one small causal region could get blown up and hence our entire visible Universe would have been in thermal contact some time in the past. In addition, the density parameter Ω would have been driven to 1 (with exponential precision). Density perturbations will be stretched by the expansion, $\lambda \sim R(t)$. Thus it will appear that $\lambda \gg H^{-1}$ or that the perturbations have left the horizon, where in fact the size of the causally connected region is now no longer simply H^{-1}. However, not only does inflation offer an explanation for large scale perturbations, it also offers a source for the perturbations themselves through quantum fluctuations.

Early models of inflation were based on a first order phase transition of a Grand Unified theory [38]. Although these models led to sufficient exponential expansion, completion of the transition through bubble percolation did not occur. Later models of inflation [39,40], also based on Grand Unified symmetry breaking, through second order transitions were also doomed. While they successfully inflated and reheated, and in fact produced density perturbations due to quantum fluctuations during the evolution of the scalar field, they predicted density perturbations many orders of magnitude too large. Most models today are based on an unknown symmetry breaking involving a new scalar field, the inflaton, ϕ.

19.3.6. ***Baryogenesis***:

The Universe appears to be populated exclusively with matter rather than antimatter. Indeed antimatter is only detected in accelerators or in cosmic rays. However, the presence of antimatter in the latter is understood to be the result of collisions of primary particles in the interstellar medium. There is in fact strong evidence against primary forms of antimatter in the Universe. Furthermore, the density of baryons compared to the density of photons is extremely small, $\eta \sim 10^{-10}$.

The production of a net baryon asymmetry requires baryon number violating interactions, C and CP violation and a departure from thermal equilibrium [41]. The first two of these ingredients are expected to be contained in grand unified theories as well as in the non-perturbative sector of the standard model, the third can be realized in an expanding universe where as we have seen interactions come in and out of equilibrium.

There are several interesting and viable mechanisms for the production of the baryon asymmetry. While, we can not review any of them here in any detail, we mention some of the important scenarios. In all cases, all three ingredients listed above are incorporated. One of the first mechanisms was based on the out of equilibrium decay of a massive particle such as a superheavy GUT gauge of Higgs boson [42,43]. A novel mechanism involving the decay of flat directions in supersymmetric models is known as the Affleck-Dine scenario [44]. Recently, much attention has been focused on the possibility of generating the baryon asymmetry at the electro-weak scale using the non-perturbative interactions of sphalerons [45]. Because these interactions conserve the sum of baryon and lepton number, $B + L$, it is possible to first generate a lepton asymmetry (*e.g.*, by the out-of-equilibrium decay of a superheavy right-handed neutrino), which is converted to a baryon asymmetry at the electro-weak scale [46]. This mechanism is known as lepto-baryogenesis.

19.3.7. ***Nucleosynthesis***:

An essential element of the standard cosmological model is Big-Bang nucleosynthesis (BBN), the theory which predicts the abundances of the light element isotopes D, ^{3}He, ^{4}He, and ^{7}Li. Nucleosynthesis takes place at a temperature scale of order 1 MeV. The nuclear processes lead primarily to ^{4}He, with a primordial mass fraction of about 24%. Lesser amounts of the other light elements are produced: about 10^{-5} of D and ^{3}He and about 10^{-10} of ^{7}Li by number relative to H. The abundances of the light elements depend almost solely on one key parameter, the baryon-to-photon ratio, η. The nucleosynthesis predictions can be compared with observational determinations of the abundances of the light elements. Consistency between theory and observations leads to a conservative range of

$$3.4 \times 10^{-10} < \eta < 6.9 \times 10^{-10}\ . \tag{19.54}$$

η is related to the fraction of Ω contained in baryons, Ω_b

$$\Omega_b = 3.66 \times 10^7 \eta\, h^{-2}\ , \tag{19.55}$$

or $10^{10}\eta = 274\Omega_b h^2$. The WMAP result [28] for $\Omega_b h^2$ of 0.0224 ± 0.0009 translates into a value of $\eta = 6.15 \pm 0.25$. This result can be used to 'predict' the light element abundance which can in turn be compared with observation [47]. The resulting D/H abundance is in excellent agreement with that found in quasar absorption systems. It is in reasonable agreement with the helium abundance observed in extra-galactic HII regions (once systematic uncertainties are accounted

for) but is in poor agreement with the Li abundance observed in the atmospheres of halo dwarf stars. (See the review on BBN—Sec. 20 of this *Review* for a detailed discussion of BBN or references [48,49].)

19.3.8. ***The transition to a matter-dominated Universe***:

In the Standard Model, the temperature (or redshift) at which the Universe undergoes a transition from a radiation dominated to a matter dominated Universe is determined by the amount of dark matter. Assuming three nearly massless neutrinos, the energy density in radiation at temperatures $T \ll 1$ MeV, is given by

$$\rho_r = \frac{\pi^2}{30}\left[2 + \frac{21}{4}\left(\frac{4}{11}\right)^{4/3}\right] T^4 \ . \tag{19.56}$$

In the absence of non-baryonic dark matter, the matter density can be written as

$$\rho_m = m_{\rm N}\eta\, n_\gamma \ , \tag{19.57}$$

where $m_{\rm N}$ is the nucleon mass. Recalling that $n_\gamma \propto T^3$ [cf. Eq. (19.39)], we can solve for the temperature or redshift at the matter-radiation equality when $\rho_r = \rho_m$,

$$T_{eq} = 0.22\, m_{\rm N}\eta \qquad \text{or} \qquad (1+z_{\rm eq}) = 0.22\,\eta\frac{m_{\rm N}}{T_0} \ , \tag{19.58}$$

where T_0 is the present temperature of the microwave background. For $\eta = 5\times10^{-10}$, this corresponds to a temperature $T_{\rm eq} \simeq 0.1$ eV or $(1+z_{\rm eq}) \simeq 425$. A transition this late is very problematic for structure formation (see Sec. 19.4.5).

The redshift of matter domination can be pushed back significantly if non-baryonic dark matter is present. If instead of Eq. (19.57), we write

$$\rho_m = \Omega_m \rho_c \left(\frac{T}{T_0}\right)^3 \ , \tag{19.59}$$

we find that

$$T_{eq} = 0.9\frac{\Omega_m \rho_c}{T_0^3} \qquad \text{or} \qquad (1+z_{eq}) = 2.4\times10^4 \Omega_m h^2 \ . \tag{19.60}$$

19.4. The Universe at late times

19.4.1. ***The CMB***:

One form of the infamous Olbers' paradox says that, in Euclidean space, surface brightness is independent of distance. Every line of sight will terminate on matter that is hot enough to be ionized and so scatter photons: $T \gtrsim 10^3$ K; the sky should therefore shine as brightly as the surface of the Sun. The reason the night sky is dark is entirely due to the expansion, which cools the radiation temperature to 2.73 K. This gives a Planck function peaking at around 1 mm to produce the microwave background (CMB).

The CMB spectrum is a very accurate match to a Planck function [50]. (See the review on CBR–Sec. 23 of this *Review*.) The COBE estimate of the temperature is [51]

$$T = 2.725 \pm 0.002\,{\rm K} \ . \tag{19.61}$$

The lack of any distortion of the Planck spectrum is a strong physical constraint. It is very difficult to account for in any expanding universe other than one that passes through a hot stage. Alternative schemes for generating the radiation, such as thermalization of starlight by dust grains, inevitably generate a superposition of temperatures. What is required in addition to thermal equilibrium is that $T \propto 1/R$, so that radiation from different parts of space appears identical.

Although it is common to speak of the CMB as originating at "recombination," a more accurate terminology is the era of "last scattering," In practice, this takes place at $z \simeq 1100$, almost independently of the main cosmological parameters, at which time the fractional ionization is very small. This occurred when the age of the Universe was a few hundred thousand years. (See the review on CBR–Sec. 23 of this *Review* for a full discussion of the CMB.)

19.4.2. ***Matter in the Universe***:

One of the main tasks of cosmology is to measure the density of the Universe, and how this is divided between dark matter and baryons. The baryons consist partly of stars, with $0.002 \lesssim \Omega_* \lesssim 0.003$ [52] but mainly inhabit the IGM. One powerful way in which this can be studied is via the absorption of light from distant luminous objects such as quasars. Even very small amounts of neutral hydrogen can absorb rest-frame UV photons (the Gunn-Peterson effect), and should suppress the continuum by a factor $\exp(-\tau)$, where

$$\tau \simeq \left[\frac{n_{\rm HI}(z)}{(1+z)\sqrt{1+\Omega_m z}}\right] \Big/ \ 10^{-4.62}\ h\,{\rm m}^{-3}, \tag{19.62}$$

and this expression applies while the Universe is matter dominated ($z \gtrsim 1$ in the $\Omega_m = 0.3$ $\Omega_v = 0.7$ model). It is possible that this general absorption has now been seen at $z = 6.2$ [53]. In any case, the dominant effect on the spectrum is a 'forest' of narrow absorption lines, which produce a mean $\tau = 1$ in the Lyα forest at about $z = 3$, and so we have $\Omega_{\rm HI} \simeq 10^{-5.5}h^{-1}$. This is such a small number that clearly the IGM is very highly ionized at these redshifts.

The Lyα forest is of great importance in pinning down the abundance of deuterium. Because electrons in deuterium differ in reduced mass by about 1 part in 4000 compared to Hydrogen, each absorption system in the Lyα forest is accompanied by an offset deuterium line. By careful selection of systems with an optimal HI column density, a measurement of the D/H ratio can be made. This has now been done in 5 quasars, with relatively consistent results [49]. Combining these determinations with the theory of primordial nucleosynthesis yields a baryon density of $\Omega_b h^2 = 0.021 \pm 0.004$ (95% confidence). (See also the review on BBN—Sec. 20 of this *Review*.)

Ionized IGM can also be detected in emission when it is densely clumped, via bremsstrahlung radiation. This generates the spectacular X-ray emission from rich clusters of galaxies. Studies of this phenomenon allow us to achieve an accounting of the total baryonic material in clusters. Within the central $\simeq 1$ Mpc, the masses in stars, X-ray emitting gas and total dark matter can be determined with reasonable accuracy (perhaps 20% rms), and this allows a minimum baryon fraction to be determined [54,55]:

$$\frac{M_{\rm baryons}}{M_{\rm total}} \gtrsim 0.009 + (0.066 \pm 0.003)\, h^{-3/2} \ . \tag{19.63}$$

Because clusters are the largest collapsed structures, it is reasonable to take this as applying to the Universe as a whole. This equation implies a minimum baryon fraction of perhaps 12% (for reasonable h), which is too high for $\Omega_m = 1$ if we take $\Omega_b h^2 \simeq 0.02$ from nucleosynthesis. This is therefore one of the more robust arguments in favor of $\Omega_m \simeq 0.3$. (See the review on Global cosmological parameters—Sec. 21 of this *Review*.) This argument is also consistent with the inference on Ω_m that can be made from Fig. 19.2.

This method is much more robust than the older classical technique for weighing the Universe: '$L \times M/L$'. The overall light density of the Universe is reasonably well determined from redshift surveys of galaxies, so that a good determination of mass M and luminosity L for a single object suffices to determine Ω_m *if* the mass-to-light ratio is universal.

Galaxy redshift surveys allow us to deduce the galaxy luminosity function, ϕ, which is the comoving number density of galaxies; this may be described by the Schechter function, which is a power law with an exponential cutoff:

$$\phi = \phi^* \left(\frac{L}{L^*}\right)^{-\alpha} e^{-L/L^*}\,\frac{dL}{L^*} \tag{19.64}$$

The total luminosity density produced by integrating over the distribution is

$$\rho_L = \phi^* L^* \Gamma(2-\alpha) \ , \tag{19.65}$$

and this tells us the average mass-to-light ratio needed to close the Universe. Answers vary (principally owing to uncertainties in ϕ^*). In blue light, the total luminosity density is

$\rho_{\rm L} = 2 \pm 0.2 \times 10^8\ h L_\odot {\rm Mpc}^{-3}$ [56,57]. The critical density is $2.78 \times 10^{11}\ \Omega h^2 M_\odot {\rm Mpc}^{-3}$, so the critical M/L for closure is

$$(M/L)_{\rm crit,\,B} = 1390\,h \pm 10\% \ . \tag{19.66}$$

Dynamical determinations of mass on the largest accessible scales consistently yield blue M/L values of at least $300\,h$, but normally fall short of the closure value [58]. This was a long-standing argument against the $\Omega_m = 1$ model, but it was never conclusive because the stellar populations in objects such as rich clusters (where the masses can be determined) differ systematically from those in other regions.

19.4.3. *Gravitational lensing*:

A robust method for determining masses in cosmology is to use gravitational light deflection. Most systems can be treated as a geometrically thin gravitational lens, where the light bending is assumed to take place only at a single distance. Simple geometry then determines a mapping between the coordinates in the intrinsic source plane and the observed image plane:

$$\alpha(D_{\rm L}\theta_{\rm I}) = \frac{D_{\rm S}}{D_{\rm LS}}(\theta_{\rm I} - \theta_{\rm S}) \ , \tag{19.67}$$

where the angles $\theta_{\rm I}, \theta_{\rm S}$ and α are in general two-dimensional vectors on the sky. The distances $D_{\rm LS}$ etc. are given by an extension of the usual distance-redshift formula:

$$D_{\rm LS} = \frac{R_0 S_k(\chi_{\rm S} - \chi_{\rm L})}{1 + z_{\rm S}} \ . \tag{19.68}$$

This is the angular-diameter distance for objects on the source plane as perceived by an observer on the lens.

Solutions of this equation divide into weak lensing, where the mapping between source plane and image plane is one-to-one, and strong lensing, in which multiple imaging is possible. For circularly-symmetric lenses, an on-axis source is multiply imaged into a 'caustic' ring, whose radius is the Einstein radius:

$$\begin{aligned} \theta_{\rm E} &= \left(4GM\,\frac{D_{\rm LS}}{D_{\rm L} D_{\rm S}}\right)^{1/2} \\ &= \left(\frac{M}{10^{11.09} M_\odot}\right)^{1/2} \left(\frac{D_{\rm L} D_{\rm S}/D_{\rm LS}}{\rm Gpc}\right)^{-1/2} \quad {\rm arcsec} \end{aligned} \tag{19.69}$$

The observation of 'arcs' (segments of near-perfect Einstein rings) in rich clusters of galaxies has thus given very accurate masses for the central parts of clusters—generally in good agreement with other indicators, such as analysis of X-ray emission from the cluster IGM [59].

19.4.4. *Density Fluctuations*:

The overall properties of the Universe are very close to being homogeneous; and yet telescopes reveal a wealth of detail on scales varying from single galaxies to large-scale structures of size exceeding 100 Mpc. The existence of these structures must be telling us something important about the initial conditions of the Big Bang, and about the physical processes that have operated subsequently. This motivates the study of the density perturbation field, defined as

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \langle\rho\rangle}{\langle\rho\rangle} \ . \tag{19.70}$$

A critical feature of the δ field is that it inhabits a universe that is isotropic and homogeneous in its large-scale properties. This suggests that the statistical properties of δ should also be statistically homogeneous—*i.e.*, it is a stationary random process.

It is often convenient to describe δ as a Fourier superposition:

$$\delta(\mathbf{x}) = \sum \delta_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}} \ . \tag{19.71}$$

We avoid difficulties with an infinite universe by applying periodic boundary conditions in a cube of some large volume V. The cross-terms vanish when we compute the variance in the field, which is just a sum over modes of the power spectrum

$$\langle\delta^2\rangle = \sum |\delta_{\mathbf{k}}|^2 \equiv \sum P(k) \ . \tag{19.72}$$

Note that the statistical nature of the fluctuations must be isotropic, so we write $P(k)$ rather than $P(\mathbf{k})$. The $\langle\ldots\rangle$ average here is a volume average. Cosmological density fields are an example of an ergodic process, in which the average over a large volume tends to the same answer as the average over a statistical ensemble.

The statistical properties of discrete objects sampled from the density field are often described in terms of N-point correlation functions, which represent the excess probability over random for finding one particle in each of N boxes in a given configuration. For the 2-point case, the correlation function is readily shown to be identical to the autocorrelation function of the δ field: $\xi(r) = \langle\delta(x)\delta(x+r)\rangle$.

The power spectrum and correlation function are Fourier conjugates, and thus are equivalent descriptions of the density field (similarly, k-space equivalents exist for the higher-order correlations). It is convenient to take the limit $V \to \infty$ and use k-space integrals, defining a dimensionless power spectrum as $\Delta^2(k) = d\langle\delta^2\rangle/d\ln k = Vk^3P(k)/2\pi^2$:

$$\xi(r) = \int \Delta^2(k)\,\frac{\sin kr}{kr}\,d\ln k; \quad \Delta^2(k) = \frac{2}{\pi}k^3\int_0^\infty \xi(r)\,\frac{\sin kr}{kr}\,r^2\,dr \ . \tag{19.73}$$

For many years, an adequate approximation to observational data on galaxies was $\xi = (r/r_0)^{-\gamma}$, with $\gamma \simeq 1.8$ and $r_0 \simeq 5\,h^{-1}$ Mpc. Modern surveys are now able to probe into the large-scale linear regime where traces of the curved primordial spectrum can be detected [60].

19.4.5. *Formation of cosmological structure*:

The simplest model for the generation of cosmological structure is gravitational instability acting on some small initial fluctuations (for the origin of which a theory such as inflation is required). If the perturbations are adiabatic (*i.e.*, fractionally perturb number densities of photons and matter equally), the linear growth law for matter perturbations is simple:

$$\delta \propto \begin{cases} a(t)^2 & \text{(radiation domination; } \Omega_{\rm r} = 1) \\ a(t) & \text{(matter domination; } \Omega_{\rm m} = 1) \end{cases} \tag{19.74}$$

For low density universes, the present-day amplitude is suppressed by a factor $g(\Omega)$, where

$$g(\Omega) \simeq \frac{5}{2}\Omega_m \left[\Omega_m^{4/7} - \Omega_v + (1 + \Omega_m/2)(1 + \frac{1}{70}\Omega_v)\right]^{-1} \ , \tag{19.75}$$

is an accurate fit for models with matter plus cosmological constant. The alternative perturbation mode is isocurvature: only the equation of state changes, and the total density is initially unperturbed. These modes perturb the total entropy density, and thus induce additional large-scale CMB anisotropies [61]. Although the character of perturbations in the simplest inflationary theories are purely adiabatic, correlated adiabatic and isocurvature modes are predicted in many models; the simplest example is the curvaton, which is a scalar field that decays to yield a perturbed radiation density. If the matter content already exists at this time, the overall perturbation field will have a significant isocurvature component. Such a prediction is inconsistent with current CMB data [62], and most analyses of CMB and LSS data assume the adiabatic case to hold exactly.

Linear evolution preserves the shape of the power spectrum. However, a variety of processes mean that growth actually depends on the matter content:

(1) Pressure opposes gravity effectively for wavelengths below the horizon length while the Universe is radiation dominated. The *comoving* horizon size at $z_{\rm eq}$ is therefore an important scale:

$$D_{\rm H}(z_{\rm eq}) = \frac{2(\sqrt{2}-1)}{(\Omega_m z_{\rm eq})^{1/2} H_0} = \frac{16.0}{\Omega_m h^2}\,{\rm Mpc} \tag{19.76}$$

(2) At early times, dark matter particles will undergo free streaming at the speed of light, and so erase all scales up to the horizon—a process that only ceases when the particles go nonrelativistic. For light massive neutrinos, this happens at $z_{\rm eq}$; all structure up to the horizon-scale power-spectrum break is in fact erased. Hot(cold) dark matter models are thus sometimes dubbed large(small)-scale damping models.

(3) A further important scale arises where photon diffusion can erase perturbations in the matter–radiation fluid; this process is named Silk damping.

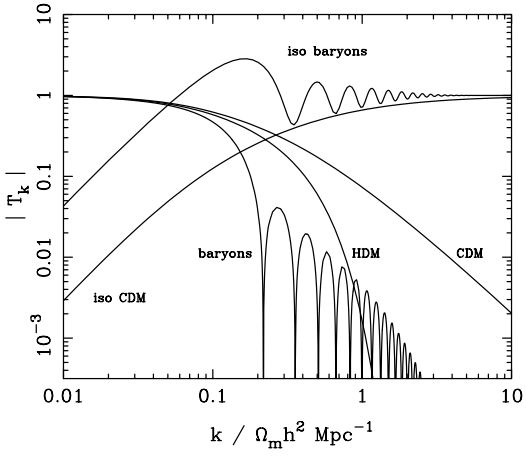

Figure 19.4: A plot of transfer functions for various models. For adiabatic models, $T_k \to 1$ at small k, whereas the opposite is true for isocurvature models. For dark-matter models, the characteristic wavenumber scales proportional to $\Omega_m h^2$. The scaling for baryonic models does not obey this exactly; the plotted cases correspond to $\Omega_m = 1$, $h = 0.5$.

The overall effect is encapsulated in the transfer function, which gives the ratio of the late-time amplitude of a mode to its initial value (see Fig. 19.4). The overall power spectrum is thus the primordial power-law, times the square of the transfer function:

$$P(k) \propto k^n T_k^2 \ . \tag{19.77}$$

The most generic power-law index is $n = 1$: the 'Zeldovich' or 'scale-invariant' spectrum. Inflationary models tend to predict a small 'tilt': $|n-1| \lesssim 0.03$ [13,14]. On the assumption that the dark matter is cold, the power spectrum then depends on 5 parameters: n, h, Ω_b, $\Omega_{\rm cdm}$ ($\equiv \Omega_m - \Omega_b$) and an overall amplitude. The latter is often specified as σ_8, the linear-theory fractional rms in density when a spherical filter of radius $8\,h^{-1}\,$Mpc is applied in linear theory. This scale can be probed directly via weak gravitational lensing, and also via its effect on the abundance of rich galaxy clusters. The favored value is approximately [63,64]

$$\sigma_8 = (0.7 \pm 15\%)(\Omega_m/0.3)^{-0.5} \ . \tag{19.78}$$

A direct measure of mass inhomogeneity is valuable, since the galaxies inevitably are biased with respect to the mass. This means that the fractional fluctuations in galaxy number, $\delta n/n$ may differ from the mass fluctuations, $\delta\rho/\rho$. It is commonly assumed that the two fields obey some proportionality on large scales where the fluctuations are small, $\delta n/n = b\delta\rho/\rho$, but even this is not guaranteed [65].

The main shape of the transfer function is a break around the horizon scale at $z_{\rm eq}$, which depends just on $\Omega_m h$ when wavenumbers are measured in observable units ($h\,$Mpc^{-1}). In principle, accurate data over a wide range of k could determine both Ωh and n, but in practice there is a strong degeneracy between these. For reasonable baryon content, weak oscillations in the transfer function may be visible, giving an alternative means of fixing the baryon content. Current data [60] favor $\Omega_m h \simeq 0.20$ and a baryon fraction of about 0.15 for $n = 1$ (see Fig. 19.5). In order to constrain n itself, it is necessary to examine data on anisotropies in the CMB.

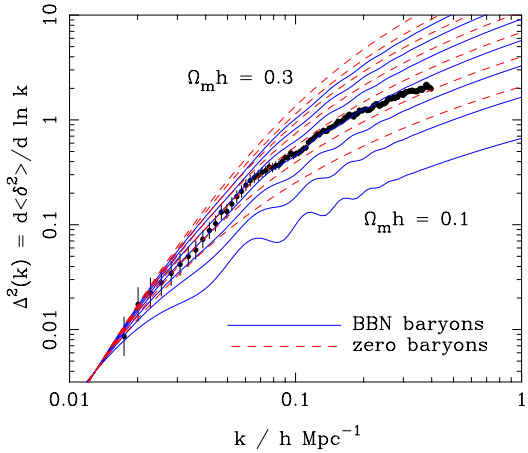

Figure 19.5: The galaxy power spectrum from the 2dFGRS, shown in dimensionless form, $\Delta^2(k) \propto k^3 P(k)$. The solid points with error bars show the power estimate. The window function correlates the results at different k values, and also distorts the large-scale shape of the power spectrum An approximate correction for the latter effect has been applied. The solid and dashed lines show various CDM models, all assuming $n = 1$. For the case with non-negligible baryon content, a big-bang nucleosynthesis value of $\Omega_b h^2 = 0.02$ is assumed, together with $h = 0.7$. A good fit is clearly obtained for $\Omega_m h \simeq 0.2$.

19.4.6. *CMB anisotropies*:

The CMB has a clear dipole anisotropy, of magnitude 1.23×10^{-3}. This is interpreted as being due to the Earth's motion, which is equivalent to a peculiar velocity for the Milky Way of

$$v_{\rm MW} \simeq 600\,{\rm km\,s^{-1}} \quad \text{towards} \quad (\ell, b) \simeq (270^\circ, 30^\circ) \ . \tag{19.79}$$

All higher-order multipole moments of the CMB are however much smaller (of order 10^{-5}), and interpreted as signatures of density fluctuations at last scattering ($\simeq 1100$). To analyze these, the sky is expanded in spherical harmonics as explained in the review on CBR–Sec. 23 of this *Review*. The dimensionless power per ln k or 'bandpower' for the CMB is defined as

$$\mathcal{T}^2(\ell) = \frac{\ell(\ell+1)}{2\pi} C_\ell \ . \tag{19.80}$$

This function encodes information from the three distinct mechanisms that cause CMB anisotropies:

(1) Gravitational (Sachs–Wolfe) perturbations. Photons from high-density regions at last scattering have to climb out of potential wells, and are thus redshifted.

(2) Intrinsic (adiabatic) perturbations. In high-density regions, the coupling of matter and radiation can compress the radiation also, giving a higher temperature.

(3) Velocity (Doppler) perturbations. The plasma has a non-zero velocity at recombination, which leads to Doppler shifts in frequency and hence shifts in brightness temperature.

Because the potential fluctuations obey Poisson's equation, $\nabla^2\Phi = 4\pi G\rho\delta$, and the velocity field satisfies the continuity equation $\nabla\cdot\mathbf{u} = -\dot{\delta}$, the resulting different powers of k ensure that the Sachs-Wolfe effect dominates on large scales and adiabatic effects on small scales.

The relation between angle and comoving distance on the last-scattering sphere requires the comoving angular-diameter distance to the last-scattering sphere; because of its high redshift, this is effectively identical to the horizon size at the present epoch, $D_{\rm H}$:

$$\begin{aligned} D_{\rm H} &= \frac{2}{\Omega_m H_0} \quad (\Omega_v = 0) \\ D_{\rm H} &\simeq \frac{2}{\Omega_m^{0.4} H_0} \quad (\text{flat} : \Omega_m + \Omega_v = 1) \end{aligned} \tag{19.81}$$

These relations show how the CMB is strongly sensitive to curvature: the horizon length at last scattering is $\propto 1/\sqrt{\Omega_m}$, so that this subtends an angle that is virtually independent of Ω_m for a flat model. Observations of a peak in the CMB power spectrum at relatively large scales ($\ell \simeq 225$) are thus strongly inconsistent with zero-Λ models with low density: current CMB data require $\Omega_m + \Omega_v \simeq 1 \pm 0.05$. (See *e.g.*, Fig. 19.2).

In addition to curvature, the CMB encodes information about several other key cosmological parameters. Within the compass of simple adiabatic CDM models, there are 9 of these:

$$\omega_c,\ \omega_b,\ \Omega_t,\ h,\ \tau,\ n_s,\ n_t,\ r,\ Q\ . \tag{19.82}$$

The symbol ω denotes the physical density, Ωh^2: the transfer function depends only on the densities of CDM (ω_c) and baryons (ω_b). Transcribing the power spectrum at last scattering into an angular power spectrum brings in the total density parameter ($\Omega_t \equiv \Omega_m + \Omega_v = \Omega_c + \Omega_b + \Omega_v$) and h: there is an exact geometrical degeneracy [66] between these that keeps the angular-diameter distance to last scattering invariant, so that models with substantial spatial curvature and large vacuum energy cannot be ruled out without prior knowledge of the Hubble parameter. Alternatively, the CMB alone cannot measure the Hubble parameter.

The other main parameter degeneracy involves the tensor contribution to the CMB anisotropies. These are important at large scales (up to the horizon scales); for smaller scales, only scalar fluctuations (density perturbations) are important. Each of these components is characterized by a spectral index, n, and a ratio between the power spectra of tensors and scalars (r). Finally, the overall amplitude of the spectrum must be specified (Q), together with the optical depth to Compton scattering owing to recent reionization (τ). The tensor degeneracy operates as follows: the main effect of adding a large tensor contribution is to reduce the contrast between low ℓ and the peak at $\ell \simeq 225$ (because the tensor spectrum has no acoustic component). The required height of the peak can be recovered by increasing n_s to increase the small-scale power in the scalar component; this in turn over-predicts the power at $\ell \sim 1000$, but this effect can be counteracted by raising the baryon density [67]. In order to break this degeneracy, additional data are required. For example, an excellent fit to the CMB data is obtained with a scalar-only model with zero curvature and $\omega_b = 0.023$, $\omega_c = 0.134$, $h = 0.73$, $n_s = 0.97$ [28], but this is indistinguishable from a model where tensors dominate at $\ell \lesssim 100$, but we raise ω_b to 0.03 and n_s to 1.2. This baryon density is too high for nucleosynthesis, which disfavors the high-tensor solution [68].

The reason the tensor component is introduced, and why it is so important, is that it is the only non-generic prediction of inflation. Slow-roll models of inflation involve two dimensionless parameters:

$$\begin{aligned} \epsilon &\equiv \frac{M_{\rm P}^2}{16\pi}\,(V'/V)^2 \\ \eta &\equiv \frac{M_{\rm P}^2}{8\pi}\,(V''/V) \end{aligned}\ , \tag{19.83}$$

where V is the inflaton potential, and dashes denote derivatives with respect to the inflation field. In terms of these, the tensor-to-scalar ratio is $r \simeq 12\epsilon$, and the spectral indices are $n_s = 1 - 6\epsilon + 2\eta$ and $n_t = -2\epsilon$. The natural expectation of inflation is that the quasi-exponential phase ends once the slow-roll parameters become significantly non-zero, so that both $n_s \neq 1$ and a significant tensor component are expected. These prediction can be avoided in some models, but it is undeniable that observation of such features would be a great triumph for inflation. Much future effort in cosmology will therefore be directed towards the question of whether the Universe contains anything other than scale-invariant scalar fluctuations.

References:

1. V.M. Slipher, Pop. Astr. **23**, 21 (1915).
2. K. Lundmark, MNRAS **84**, 747 (1924).
3. E. Hubble and M.L. Humason, Ap. J. **74**, 43 (1931).
4. G. Gamow, Phys. Rev. **70**, 572 (1946).
5. R.A. Alpher, H. Bethe, and G. Gamow, Phys. Rev. **73**, 803 (1948).
6. R.A. Alpher and R.C. Herman, Phys. Rev. **74**, 1737 (1948).
7. R.A. Alpher and R.C. Herman, Phys. Rev. **75**, 1089 (1949).
8. A.A. Penzias and R.W. Wilson, Ap. J. **142**, 419 (1965).
9. S. Weinberg, *Gravitation and Cosmology*, John Wiley and Sons (1972).
10. P.J.E. Peebles, *Principles of Physical Cosmology* Princeton University Press (1993).
11. G. Börner, *The Early Universe: Facts and Fiction*, Springer-Verlag (1988).
12. E.W. Kolb and M.S. Turner, *The Early Universe*, Addison-Wesley (1990).
13. J.A. Peacock, *Cosmological Physics*, Cambridge University Press (1999).
14. A.R. Liddle and D. Lyth, *Cosmological Inflation and Large-Scale Structure*, Cambridge University Press (2000).
15. E.B. Gliner, Sov. Phys. JETP 22, 378 (1966).
16. Y.B. Zeldovich, (1967), Sov. Phys. Uspekhi, **11**, 381 (1968).
17. P.M. Garnavich *et al.*, Ap. J. **507**, 74 (1998).
18. S. Perlmutter *et al.*, Phys. Rev. Lett. **83**, 670 (1999).
19. I. Maor *et al.*, Phys. Rev. **D65**, 123003 (2002).
20. A.G. Riess *et al.*, A. J. **116**, 1009 (1998).
21. S. Perlmutter *et al.*, Ap. J. **517**, 565 (1999).
22. A.G. Riess, PASP, **112**, 1284 (2000).
23. J.L. Tonry *et al.*, `astro-ph/0305008` (2003).
24. W.L. Freedman *et al.*, ApJ **553**, 47 (2001).
25. J.A. Johnson and M. Bolte, ApJ **554**, 888 (2001).
26. R. Jimenez and P. Padoan, ApJ **498**, 704 (1998).
27. E. Carretta *et al.*, ApJ **533**, 215 (2000).
28. D.N. Spergel *et al.*, `astro-ph/0302209` (2003).
29. M. Srednicki, R. Watkins and K. A. Olive, Nucl. Phys. B **310**, 693 (1988).
30. G. Mangano *et al.*, Phys. Lett. **B534**, 8 (2002).
31. A. Linde, Phys. Rev. **D14**, 3345 (1976).
32. A. Linde, Rep. Prog. Phys. **42**, 389 (1979).
33. C.E. Vayonakis, Surveys High Energ. Phys. **5**, 87 (1986).
34. S.A. Bonometto and A. Masiero, Riv. Nuovo Cim. **9N5**, 1 (1986).
35. A. Linde, A.D. Linde, *Particle Physics And Inflationary Cosmology*, Harwood (1990).
36. K.A. Olive, Phys. Rep. **190**, 3345 (1990).
37. D. Lyth and A. Riotto, Phys. Rep. **314**, 1 (1999).
38. A.H. Guth, Phys. Rev. **D23**, 347 (1981).
39. A.D. Linde, Phys. Lett. **108B**, 389 (1982).
40. A. Albrecht and P.J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
41. A.D. Sakharov, Sov. Phys. JETP Lett. **5**, 24 (1967).
42. S. Weinberg, Phys. Rev. Lett. **42**, 850 (1979).
43. D. Toussaint *et al.*, Phys. Rev. **D19**, 1036 (1979).
44. I. Affleck and M. Dine, Nucl. Phys. **B249**, 361 (1985).
45. V. Kuzmin, V. Rubakov, and M. Shaposhnikov, Phys. Lett. **B155**, 36 (1985).
46. M. Fukugita and T. Yanagida, Phys. Lett. **B174**, 45 (1986).
47. R.H. Cyburt, B.D. Fields, and K.A. Olive, Phys. Lett. **B567**, 227 (2003).
48. K.A. Olive, G. Steigman, and T.P. Walker, Phys. Rept. **333**, 389 (2000).
49. D. Kirkman *et al.*, `astro-ph/0302006` (2003).
50. D.J. Fixsen *et al.*, ApJ **473**, 576 (1996).
51. J.C. Mather *et al.*, ApJ **512**, 511 (1999).

52. S.M. Cole *et al.*, MNRAS **326**, 255 (2001).
53. R.H. Becker *et al.*, Astr. J. **122**, 2850 (2001).
54. S.D.M. White *et al.*, Nature **366**, 429 (1993).
55. S.W. Allen, R.W. Schmidt, A.C. Fabian, MNRAS, **334**, L11 (2002).
56. S. Folkes *et al.*, MNRAS **308**, 459 (1999).
57. M.L. Blanton *et al.*, Astr. J. **121**, 2358 (2001).
58. H. Hoekstra *et al.*, ApJ, **548**, L5 (2001).
59. S.W. Allen, MNRAS **296**, 392 (1998).
60. W.J. Percival *et al.*, MNRAS **327**, 1297 (2001).
61. G. Efstathiou and J.R. Bond, MNRAS **218**, 103 (1986).
62. C. Gordon and A. Lewis, Phys. Rev. **D67**, 123513 (2003).
63. M.L. Brown *et al.*, MNRAS, **341**, 1 (2003).
64. P.T.P. Viana, R.C. Nichol, A.R. Liddle, ApJ, **569**, L75 (2002).
65. A. Dekel and O. Lahav, ApJ **520**, 24 (1999).
66. G. Efstathiou and J.R. Bond, MNRAS, **304**, 75 (1999).
67. G.P. Efstathiou *et al.*, MNRAS **330**, L29 (2002).
68. G.P. Efstathiou, MNRAS **332**, 193 (2002).

20. BIG-BANG NUCLEOSYNTHESIS

Revised October 2003 by B.D. Fields (Univ. of Illinois) and S. Sarkar (Univ. of Oxford).

Big-bang nucleosynthesis (BBN) offers the deepest reliable probe of the early universe, being based on well-understood Standard Model physics [1]. Predictions of the abundances of the light elements, D, ^{3}He, ^{4}He, and ^{7}Li, synthesized at the end of the "first three minutes" are in good overall agreement with the primordial abundances inferred from observational data, thus validating the standard hot big-bang cosmology (see [5] for a recent review). This is particularly impressive given that these abundances span nine orders of magnitude — from $^4\mathrm{He/H} \sim 0.08$ down to $^7\mathrm{Li/H} \sim 10^{-10}$ (ratios by number). Thus BBN provides powerful constraints on possible deviations from the standard cosmology [2], and on new physics beyond the Standard Model [3].

20.1. Big-bang nucleosynthesis theory

The synthesis of the light elements is sensitive to physical conditions in the early radiation-dominated era at temperatures $T \lesssim 1$ MeV, corresponding to an age $t \gtrsim 1$ s. At higher temperatures, weak interactions were in thermal equilibrium, thus fixing the ratio of the neutron and proton number densities to be $n/p = \mathrm{e}^{-Q/T}$, where $Q = 1.293$ MeV is the neutron-proton mass difference. As the temperature dropped, the neutron-proton inter-conversion rate, $\Gamma_{n\leftrightarrow p} \sim G_\mathrm{F}^2 T^5$, fell faster than the Hubble expansion rate, $H \sim \sqrt{g_* G_\mathrm{N}}\, T^2$, where g_* counts the number of relativistic particle species determining the energy density in radiation. This resulted in departure from chemical equilibrium ("freeze-out") at $T_\mathrm{fr} \sim (g_* G_\mathrm{N}/G_F^4)^{1/6} \simeq 1$ MeV. The neutron fraction at this time, $n/p = \mathrm{e}^{-Q/T_\mathrm{fr}} \simeq 1/6$, is thus sensitive to every known physical interaction, since Q is determined by both strong and electromagnetic interactions while T_fr depends on the weak as well as gravitational interactions. Moreover the sensitivity to the Hubble expansion rate affords a probe of e.g. the number of relativistic neutrino species [6]. After freeze-out the neutrons were free to β-decay so the neutron fraction dropped to $\simeq 1/7$ by the time nuclear reactions began. A useful semi-analytic description of freeze-out has been given [7].

The rates of these reactions depend on the density of baryons (strictly speaking, nucleons), which is usually expressed normalized to the blackbody photon density as $\eta \equiv n_\mathrm{B}/n_\gamma$. As we shall see, all the light-element abundances can be explained with $\eta_{10} \equiv \eta \times 10^{10}$ in the range 3.4–6.9 (95% CL). Equivalently, this can be stated as the allowed range for the baryon mass density today, $\rho_\mathrm{B} = (2.3\text{–}4.7) \times 10^{-31}$ g cm^{-3}, or as the baryonic fraction of the critical density: $\Omega_\mathrm{B} = \rho_\mathrm{B}/\rho_\mathrm{crit} \simeq \eta_{10} h^{-2}/274 = (0.012\text{–}0.025) h^{-2}$, where $h \equiv H_0/100$ km s^{-1} Mpc$^{-1} = 0.72 \pm 0.08$ is the present Hubble parameter (see Cosmological Parameters review).

The nucleosynthesis chain begins with the formation of deuterium in the process $p(n,\gamma)$D. However, photo-dissociation by the high number density of photons delays production of deuterium (and other complex nuclei) well after T drops below the binding energy of deuterium, $\Delta_\mathrm{D} = 2.23$ MeV. The quantity $\eta^{-1}\mathrm{e}^{-\Delta_\mathrm{D}/T}$, i.e. the number of photons per baryon above the deuterium photo-dissociation threshold, falls below unity at $T \simeq 0.1$ MeV; nuclei can then begin to form without being immediately photo-dissociated again. Only 2-body reactions such as $\mathrm{D}(p,\gamma)^3\mathrm{He}$, $^3\mathrm{He}(\mathrm{D},p)^4\mathrm{He}$, are important because the density has become rather low by this time.

Nearly all the surviving neutrons when nucleosynthesis begins end up bound in the most stable light element ^{4}He. Heavier nuclei do not form in any significant quantity both because of the absence of stable nuclei with mass number 5 or 8 (which impedes nucleosynthesis via n^4He, p^4He or ^{4}He^{4}He reactions) and the large Coulomb barriers for reactions such as $\mathrm{T}(^4\mathrm{He},\gamma)^7$Li and $^3\mathrm{He}(^4\mathrm{He},\gamma)^7$Be. Hence the primordial mass fraction of ^{4}He, conventionally referred to as Y_p, can be estimated by the simple counting argument

$$Y_\mathrm{p} = \frac{2(n/p)}{1+n/p} \simeq 0.25\ . \qquad (20.1)$$

There is little sensitivity here to the actual nuclear reaction rates, which are however important in determining the other "left-over" abundances: D and ^{3}He at the level of a few times 10^{-5} by number relative to H, and ^{7}Li/H at the level of about 10^{-10} (when η_{10} is in the range 1–10). These values can be understood in terms of approximate analytic arguments [8]. The experimental parameter most important in determining Y_p is the neutron lifetime, τ_n, which normalizes (the inverse of) $\Gamma_{n\leftrightarrow p}$. (This is not fully determined by G_F alone since neutrons and protons also have strong interactions, the effects of which cannot be calculated very precisely.) The experimental uncertainty in τ_n used to be a source of concern but has recently been reduced substantially: $\tau_n = 885.7 \pm 0.8$ s.

The elemental abundances, calculated using the (publicly available [9]) Wagoner code [1,10], are shown in Fig. 20.1 as a function of η_{10}. The ^{4}He curve includes small corrections due to radiative processes at zero and finite temperature [11], non-equilibrium neutrino heating during $e^\pm$ annihilation [12], and finite nucleon mass effects [13]; the range reflects primarily the 1σ uncertainty in the neutron lifetime. The spread in the curves for D, ^{3}He and ^{7}Li corresponds to the 1σ uncertainties in nuclear cross sections estimated by Monte Carlo methods [14–15]. Recently the input nuclear data have been carefully reassessed [16–18], leading to improved precision in the abundance predictions. Polynomial fits to the predicted abundances and the error correlation matrix have been given [15,19]. The boxes in Fig. 20.1 show the observationally inferred primordial abundances with their associated uncertainties, as discussed below.

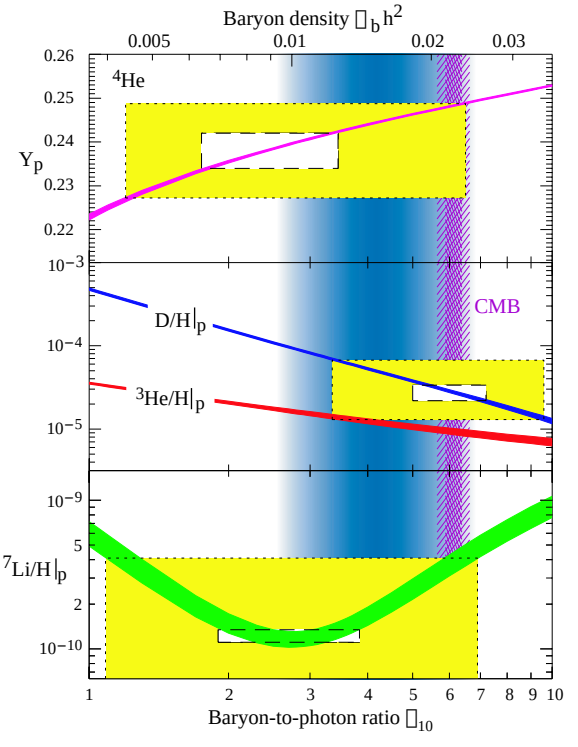

Figure 20.1: The abundances of ^{4}He, D, ^{3}He and ^{7}Li as predicted by the standard model of big-bang nucleosynthesis. Boxes indicate the observed light element abundances (smaller boxes: 2σ statistical errors; larger boxes: $\pm 2\sigma$ statistical and systematic errors added in quadrature). The narrow vertical band indicates the CMB measure of the cosmic baryon density. See full-color version on color pages at end of book.

20.2. Light Element Observations

BBN theory predicts the universal abundances of D, ^{3}He, ^{4}He, and ^{7}Li, which are essentially determined by $t \sim 180$ s. Abundances are however observed at much later epochs, after stellar nucleosynthesis has commenced. The ejected remains of this stellar processing can alter the light element abundances from their primordial values, but also produce heavy elements such as C, N, O, and Fe ("metals"). Thus one seeks astrophysical sites with low metal abundances, in order to measure light element abundances which are closer to primordial. For all of the light elements, systematic errors are an important and often dominant limitation to the precision with which primordial abundances can be inferred.

In recent years, high-resolution spectra have revealed the presence of D in high-redshift, low-metallicity quasar absorption systems (QAS), via its isotope-shifted Lyman-α absorption [20–24]. These are the first measurements of light element abundances at cosmological distances. It is believed that there are no astrophysical sources of deuterium [25], so any measurement of D/H provides a lower limit to primordial D/H and thus an upper limit on η; for example, the local interstellar value of $\mathrm{D/H} = (1.5 \pm 0.1) \times 10^{-5}$ [26] requires that $\eta_{10} \leq 9$. In fact, local interstellar D may have been depleted by a factor of 2 or more due to stellar processing; however, for the high-redshift systems, conventional models of galactic nucleosynthesis (chemical evolution) do not predict significant D/H depletion [27].

The 5 most precise observations of deuterium in QAS give $\mathrm{D/H} = (2.78 \pm 0.29) \times 10^{-5}$ [20–21], where the error is statistical only. However there remains concern over systematic errors, the dispersion between the values being much larger than is expected from the individual measurement errors ($\chi^2 = 12.4$ for 4 d.o.f.). Other lower values have been reported in different (damped Lyman-α) systems [22–23] and even the ISM value of D/H now shows unexpected scatter of a factor of 2 [28]. We thus conservatively bracket the observed values with an upper limit set by the non-detection of D in a high-redshift system, $\mathrm{D/H} < 6.7 \times 10^{-5}$ at 1σ [24], and a lower limit set by the local interstellar value [26]. These appear on Fig. 20.1, where it is clear that despite the observational uncertainties, the steep decrease of D/H with η makes it a sensitive probe of the baryon density. We are optimistic that a larger sample of D/H in high-redshift, low-redshift, and local systems will bring down systematic errors, and increase the precision with which η can be determined.

We observe ^{4}He in clouds of ionized hydrogen (H II regions), the most metal-poor of which are in dwarf galaxies. There is now a large body of data on ^{4}He and CNO in these systems [29]. These data confirm that the small stellar contribution to helium is positively correlated with metal production. Extrapolating to zero metallicity gives the primordial ^{4}He abundance [30] $Y_{\rm p} = 0.238 \pm 0.002 \pm 0.005$. Here the latter error is an estimate of the systematic uncertainty; this dominates, and is based on the scatter in different analyses of the physical properties of the H II regions [29,31]. Other extrapolations to zero metallicity give $Y_{\rm p} = 0.2443 \pm 0.0015$ [29], and $Y_{\rm p} = 0.2391 \pm 0.0020$ [32]. These are consistent (given the systematic errors) with the above estimate [30], which appears in Fig. 20.1.

The systems best suited for Li observations are metal-poor stars in the spheroid (Pop II) of our Galaxy, which have metallicities going down to at least 10^{-4} and perhaps 10^{-5} of the Solar value [33]. Observations have long shown [34–38] that Li does not vary significantly in Pop II stars with metallicities $\lesssim 1/30$ of Solar — the "Spite plateau" [34]. Recent precision data suggest a small but significant correlation between Li and Fe [35] which can be understood as the result of Li production from Galactic cosmic rays [36]. Extrapolating to zero metallicity one arrives at a primordial value [37] $\mathrm{Li/H}|_{\rm p} = (1.23 \pm 0.06) \times 10^{-10}$. One systematic error stems from the differences in techniques to determine the physical parameters (e.g., the temperature) of the stellar atmosphere in which the Li absorption line is formed. An alternative analysis [38] using a different set of stars (in a globular cluster) and a method that gives systematically higher temperatures yields $\mathrm{Li/H}|_{\rm p} = (2.19 \pm 0.28) \times 10^{-10}$; the difference with [37] indicates a systematic uncertainty of about a factor ~ 2. Another systematic error arises because it is possible that the Li in Pop II stars has been partially destroyed, due to mixing of the outer layers with the hotter interior [39]. Such processes can be constrained by the absence of significant scatter in Li-Fe [35], and by observations of the fragile isotope ^{6}Li [36]. Nevertheless, depletions by a factor as large as ~ 1.6 remain allowed [37,39]. Including these systematics, we estimate a primordial Li range of $\mathrm{Li/H}|_{\rm p} = (0.59 - 4.1) \times 10^{-10}$.

Finally, we turn to ^{3}He. Here, the only observations available are in the Solar system and (high-metallicity) H II regions in our Galaxy [40]. This makes inference of the primordial abundance difficult, a problem compounded by the fact that stellar nucleosynthesis models for ^{3}He are in conflict with observations [41]. Consequently, it is no longer appropriate to use ^{3}He as a cosmological probe; instead, one might hope to turn the problem around and constrain stellar astrophysics using the predicted primordial ^{3}He abundance [42].

20.3. Concordance, Dark Matter, and the CMB

We now use the observed light element abundances to assess the theory. We first consider standard BBN, which is based on Standard Model physics alone, so $N_\nu = 3$ and the only free parameter is the baryon-to-photon ratio η. (The implications of BBN for physics beyond the Standard Model will be considered below, §4). Thus, any abundance measurement determines η, while additional measurements overconstrain the theory and thereby provide a consistency check.

First we note that the overlap in the η ranges spanned by the larger boxes in Fig. 20.1 indicates overall concordance. More quantitatively, when we account for theoretical uncertainties as well as the statistical and systematic errors in observations, there is acceptable agreement among the abundances when

$$3.4 \leq \eta \leq 6.9 \ (95\%\ \mathrm{CL}). \tag{20.2}$$

However the agreement is much less satisfactory if we use only the quoted statistical errors in the observations. In particular, as seen in Fig. 20.1, ^{4}He and ^{7}Li are consistent with each other but favour a value of η which is lower by $\sim 2\sigma$ from that indicated by the D abundance. Additional studies are required to clarify if this discrepancy is real.

Even so the overall concordance is remarkable: using well-established microphysics we have extrapolated back to an age of ~ 1 s to correctly predict light element abundances spanning 9 orders of magnitude. This is a major success for the standard cosmology, and inspires confidence in extrapolation back to still earlier times.

This concordance provides a measure of the baryon content of the universe. With n_γ fixed by the present CMB temperature (see CMB Review), the baryon density is $\Omega_{\rm B} = 3.65 \times 10^{-3} h^{-2} \eta_{10}$, so that

$$0.012 \leq \Omega_{\rm B} h^2 \leq 0.025 \ (95\%\ \mathrm{CL}) \ , \tag{20.3}$$

a result that plays a key role in our understanding of the matter budget of the universe. First we note that $\Omega_{\rm B} \ll 1$, i.e., baryons cannot close the universe [43]. Furthermore, the cosmic density of (optically) luminous matter is $\Omega_{\rm lum} \simeq 0.0024 h^{-1}$ [44], so that $\Omega_{\rm B} \gg \Omega_{\rm lum}$: most baryons are optically dark, probably in the form of a $\sim 10^6$ K X-ray emitting intergalactic medium [45]. Finally, given that $\Omega_{\rm M} \sim 0.3$ (see Dark Matter, Cosmological Parameter Review), we infer that most matter in the universe is not only dark but also takes some non-baryonic (more precisely, non-nucleonic) form.

The BBN prediction for the cosmic baryon density can be tested through precision observations of CMB temperature fluctuations (see CMB Review). One can determine η from the amplitudes of the acoustic peaks in the CMB angular power spectrum, making it possible to compare two measures of η using very different physics, at two widely separated epochs [46]. In the standard cosmology, there is no change in η between BBN and CMB decoupling, thus, a comparison of $\eta_{\rm BBN}$ and $\eta_{\rm CMB}$ is a key test. Agreement would endorse the standard picture, and would open the way to sharper understanding of particle physics and astrophysics [54]. Disagreement could point to new physics during or between the BBN and CMB epochs.

The release of the first-year WMAP results are a landmark event in this test of BBN. As with other cosmological parameter determinations from CMB data, the derived $\eta_{\rm CMB}$ depends on the

adopted priors [47], in particular the form assumed for the power spectrum of primordial density fluctuations. If this is taken to be a scale-free power-law, the WMAP data implies $\Omega_{\rm B}h^2 = 0.024 \pm 0.001$ or $\eta_{10} = 6.58 \pm 0.27$, while allowing for a "running" spectral index lowers the value to $\Omega_{\rm B}h^2 = 0.0224 \pm 0.0009$ or $\eta_{10} = 6.14 \pm 0.25$ [48]; this latter range appears in Fig. 20.1. Other assumptions for the shape of the power spectrum can lead to baryon densities as low as $\Omega_{\rm B}h^2 = 0.019$ [49]. Thus outstanding uncertainties regarding priors are a source of systematic error which presently exceeds the statistical error in the prediction for η.

Even so, the CMB estimate of the baryon density is not inconsistent with the BBN range quoted in Eq. (20.3), and is in fact in good agreement with the value inferred from recent high-redshift D/H measurements [20]. However note that both ^{4}He and ^{7}Li are inconsistent with the CMB (as they are with D) given the error budgets we have quoted. The question then becomes more pressing as to whether this mismatch come from systematic errors in the observed abundances, and/or uncertainties in stellar astrophysics, or whether there might be new physics at work. Inhomogeneous nucleosynthesis can alter abundances for a given $\eta_{\rm BBN}$ but will overproduce ^{7}Li [50]. However a small excess of electron neutrinos over antineutrinos will lower the ^{4}He abundance below the standard BBN prediction without affecting the other elements [1]. Note that entropy generation by some non-standard process could have decreased η between the BBN era and CMB decoupling, however the lack of spectral distortions in the CMB rules out any significant energy injection upto a redshift $z \sim 10^7$ [51]. Interestingly, the CMB itself offers the promise of measuring ^{4}He [52] and possibly ^{7}Li [53] directly at $z \sim 300 - 1000$.

Bearing in mind the importance of priors, the promise of precision determinations of the baryon density using the CMB motivates using this value as an input to BBN calculations. Within the context of the Standard Model, BBN then becomes a zero-parameter theory, and the light element abundances are completely determined to within the uncertainties in $\eta_{\rm CMB}$ and the BBN theoretical errors. Comparison with the observed abundances then can be used to test the astrophysics of post-BBN light element evolution [54]. Alternatively, one can consider possible physics beyond the Standard Model (e.g., which might change the expansion rate during BBN) and then use all of the abundances to test such models; this is the subject of our final section.

20.4. Beyond the Standard Model

Given the simple physics underlying BBN, it is remarkable that it still provides the most effective test for the cosmological viability of ideas concerning physics beyond the Standard Model. Although baryogenesis and inflation must have occurred at higher temperatures in the early universe, we do not as yet have 'standard models' for these so BBN still marks the boundary between the established and the speculative in big bang cosmology. It might appear possible to push the boundary back to the quark-hadron transition at $T \sim \Lambda_{\rm QCD}$ or electroweak symmetry breaking at $T \sim 1/\sqrt{G_{\rm F}}$; however so far no observable relics of these epochs have been identified, either theoretically or observationally. Thus although the Standard Model provides a precise description of physics up to the Fermi scale, cosmology cannot be traced in detail before the BBN era.

Limits on particle physics beyond the Standard Model come mainly from the observational bounds on the ^{4}He abundance. This is proportional to the n/p ratio which is determined when the weak-interaction rates fall behind the Hubble expansion rate at $T_{\rm fr} \sim 1$ MeV. The presence of additional neutrino flavors (or of any other relativistic species) at this time increases g_*, hence the expansion rate, leading to a larger value of $T_{\rm fr}$, n/p, and therefore $Y_{\rm p}$ [6,55]. In the Standard Model, the number of relativistic particle species at 1 MeV is $g_* = 5.5 + \frac{7}{4}N_\nu$, where 5.5 accounts for photons and $e^\pm$, and N_ν is the number of (nearly massless) neutrino flavors (see Big Bang Cosmology Review). The helium curves in Fig. 20.1 were computed taking $N_\nu = 3$; the computed abundance scales as $\Delta\, Y_{\rm BBN} \simeq 0.013 \Delta N_\nu$ [7]. Clearly the central value for N_ν from BBN will depend on η, which is independently determined (with little sensitivity to N_ν) by the adopted D or ^{7}Li abundance. For example, if the best value for the observed primordial ^{4}He abundance is 0.238, then, for $\eta_{10} \sim 2$, the central value for N_ν is very close to 3. A maximum likelihood analysis on η and N_ν based on ^{4}He and ^{7}Li [56] finds the (correlated) 95% CL ranges to be $1.7 \leq \eta_{10} \leq 4.3$, and $1.4 \leq N_\nu \leq 4.9$. Similar results were obtained in another study [57] which presented a simpler method (FastBBN [9]) to extract such bounds based on χ^2 statistics, given a set of input abundances. Tighter bounds are obtained if less conservative assumptions are made concerning primordial abundances, e.g. adopting the 'low' D abundance [21] fixes $\eta_{10} = 5.6 \pm 0.6$ ($\Omega_{\rm B}h^2 = 0.02 \pm 0.002$) at 95% CL, and requires $N_\nu < 3.2$ [58] even if the 'high' ^{4}He abundance [29] is used. Using the CMB determination of η yields even tighter constraints, with $N_\nu = 3$ barely allowed at 2σ [59]! However if the discrepancy between the ^{4}He and D abundances is indeed due to a ν_e chemical potential, then N_ν can range up to 7.1 at 2σ [60].

It is clear that just as one can use the measured helium abundance to place limits on g_* [55], any changes in the strong, weak, electromagnetic, or gravitational coupling constants, arising e.g. from the dynamics of new dimensions, can be similarly constrained [61].

The limits on N_ν can be translated into limits on other types of particles or particle masses that would affect the expansion rate of the Universe during nucleosynthesis. For example consider 'sterile' neutrinos with only right-handed interactions of strength $G_{\rm R} < G_{\rm F}$. Such particles would decouple at higher temperature than (left-handed) neutrinos, so their number density ($\propto T^3$) relative to neutrinos would be reduced by any subsequent entropy release, e.g. due to annihilations of massive particles that become non-relativistic in between the two decoupling temperatures. Thus (relativistic) particles with less than full strength weak interactions contribute less to the energy density than particles that remain in equilibrium up to the time of nucleosynthesis [62]. If we impose $N_\nu < 4$ as an illustrative constraint, then the three right-handed neutrinos must have a temperature $3(T_{\nu_{\rm R}}/T_{\nu_{\rm L}})^4 < 1$. Since the temperature of the decoupled $\nu_{\rm R}$'s is determined by entropy conservation (see Big Bang Cosmology Review), $T_{\nu_{\rm R}}/T_{\nu_{\rm L}} = [(43/4)/g_*(T_{\rm d})]^{1/3} < 0.76$, where $T_{\rm d}$ is the decoupling temperature of the $\nu_{\rm R}$'s. This requires $g_*(T_{\rm d}) > 24$ so decoupling must have occurred at $T_{\rm d} > 140$ MeV. The decoupling temperature is related to $G_{\rm R}$ through $(G_{\rm R}/G_{\rm F})^2 \sim (T_{\rm d}/3\,{\rm MeV})^{-3}$, where 3 MeV is the decoupling temperature for $\nu_{\rm L}$s. This yields a limit $G_{\rm R} \lesssim 10^{-2} G_{\rm F}$. The above argument sets lower limits on the masses of new Z' gauge bosons in superstring models [63] or in extended technicolour models [64] to which such right-handed neutrinos would be coupled. Similarly a Dirac magnetic moment for neutrinos, which would allow the right-handed states to be produced through scattering and thus increase g_*, can be significantly constrained [65], as can any new interactions for neutrinos which have a similar effect [66]. Right-handed states can be populated directly by helicity-flip scattering if the neutrino mass is large enough and this can be used to used to infer e.g. a bound of $m_{\nu_\tau} \lesssim 1$ MeV taking $N_\nu < 4$ [67]. If there is mixing between active and sterile neutrinos then the effect on BBN is more complicated [68].

The limit on the expansion rate during BBN can also be translated into bounds on the mass/lifetime of particles which are non-relativistic during BBN resulting in an even faster speed-up rate; the subsequent decays of such particles will typically also change the entropy, leading to further constraints [69]. Even more stringent constraints come from requiring that the synthesized light element abundances are not excessively altered through photodissociations by the electromagnetic cascades triggered by the decays [70,71], or by the effects of hadrons in the cascades [80]. Such arguments have been applied to e.g. rule out a MeV mass ν_τ which decays during nucleosynthesis [73]; even if the decays are to non-interacting particles (and light neutrinos), bounds can be derived from considering their effects on BBN [74].

Such arguments have proved very effective in constraining supersymmetry. For example if the gravitino is light and contributes to g_*, the illustrative BBN limit $N_\nu < 4$ requires its mass to exceed ~ 1 eV [75]. In models where supersymmetry breaking is gravity mediated, the gravitino mass is usually much higher, of order the electroweak scale; such gravitinos would be unstable and decay after BBN. The constraints on unstable particles discussed above imply stringent bounds on the allowed abundance of such particles, which in turn impose powerful constraints on supersymmetric inflationary

cosmology [71,80]. These can be evaded only if the gravitino is massive enough to decay before BBN, i.e. $m_{3/2} \gtrsim 50$ TeV [76] which would be unnatural, or if it is in fact the LSP and thus stable [71,77]. Similar constraints apply to moduli — very weakly coupled fields in supergravity/string models which obtain an electroweak-scale mass from supersymmetry breaking [78].

Finally we mention that BBN places powerful constraints on the recently suggested possibility that there are new large dimensions in nature, perhaps enabling the scale of quantum gravity to be as low as the electroweak scale [79]. Thus Standard Model fields may be localized on a 'brane' while gravity alone propagates in the 'bulk'. It has been further noted that the new dimensions may be non-compact, even infinite [80] and the cosmology of such models has attracted considerable attention. The expansion rate in the early universe can be significantly modified so BBN is able to set interesting constraints on such possibilities [81].

References:

1. R.V. Wagoner *et al.*, Astrophys. J. **148**, 3 (1967).
2. R.A. Malaney and G.J. Mathews, Phys. Reports **229**, 145 (1993).
3. S. Sarkar, Rept. on Prog. in Phys. **59**, 1493 (1996).
4. D.N. Schramm and M.S. Turner, Rev. Mod. Phys. **70**, 303 (1998).
5. K.A. Olive *et al.*, Phys. Reports **333**, 389 (2000).
6. P.J.E. Peebles, Phys. Rev. Lett. **16**, 411 (1966).
7. J. Bernstein *et al.*, Rev. Mod. Phys. **61**, 25 (1989).
8. R. Esmailzadeh *et al.*, Astrophys. J. **378**, 504 (1991).
9. www-thphys.physics.ox.ac.uk/users/SubirSarkar/bbn.html.
10. L. Kawano, FERMILAB-PUB-92/04-A.
11. S. Esposito *et al.*, Nucl. Phys. **B568**, 421 (2000).
12. S. Dodelson and M.S. Turner, Phys. Rev. **D46**, 3372 (1992).
13. D. Seckel, hep-ph/9305311;
    R. Lopez and M.S. Turner, Phys. Rev. **D59**, 103502 (1999).
14. M.S. Smith *et al.*, Astrophys. J. Supp. **85**, 219 (1993).
15. G. Fiorentini *et al.*, Phys. Rev. **D58**, 063506 (1998).
16. K.M. Nollett and S. Burles, Phys. Rev. **D61**, 123505 (2000).
17. R.H. Cyburt *et al.*, New Astron. **6**, 215 (2001).
18. A. Coc *et al.*, astro-ph/0309480.
19. K.M. Nollett *et al.*, Astrophys. J. Lett. **552**, L1 (2001).
20. D. Kirkman *et al.*, astro-ph/0302006.
21. J.M. O'Meara *et al.*, Astrophys. J. **552**, 718 (2001).
22. S. D'Odorico *et al.*, Astron. & Astrophys. **368**, L21 (2001).
23. M. Pettini and D. Bowen, Astrophys. J. **560**, 41 (2001).
24. D. Kirkman *et al.*, Astrophys. J. **529**, 655 (2000).
25. R.I. Epstein *et al.*, Nature **263**, 198 (1976).
26. J. Linsky, Space Sci. Rev., **106**, 49 (2003).
27. B.D. Fields, Astrophys. J. **456**, 678 (1996).
28. G. Sonneborn *et al.*, Astrophys. J. **545**, 277 (2000).
29. Y.I. Izotov *et al.*, Astrophys. J. **527**, 757 (1999).
30. B.D. Fields and K.A. Olive, Astrophys. J. **506**, 177 (1998).
31. K.A. Olive and E. Skillman, New Astron. **6**, 119 (2001).
32. V. Luridiana *et al.*, Astrophys. J. **592**, 846 (2003).
33. N. Christlieb *et al.*, Nature **419**, 904 (2002).
34. M. Spite and F. Spite, Nature **297**, 483 (1982).
35. S.G. Ryan *et al.*, Astrophys. J. Lett. **530**, L57 (2000).
36. E. Vangioni-Flam *et al.*, New Astron. **4**, 245 (1999).
37. S.G. Ryan *et al.*, Astrophys. J. Lett. **530**, L57 (2000).
38. P. Bonifacio *et al.*, Astron. & Astrophys. **390**, 91 (2002).
39. M.H. Pinsonneault *et al.*, Astrophys. J. **574**, 389 (2002).
40. D.S. Balser *et al.*, Astrophys. J. **510**, 759 (1999).
41. K.A. Olive *et al.*, Astrophys. J. **479**, 752 (1997).
42. E. Vangioni-Flam *et al.*, Astrophys. J. **585**, 611 (2003).
43. H. Reeves *et al.*, Astrophys. J. **179**, 909 (1973).
44. M. Fukugita *et al.*, Astrophys. J. **503**, 518 (1998).
45. R. Cen and J.P. Ostriker, Astrophys. J. **514**, 1 (1999).
46. G. Jungman *et al.*, Phys. Rev. **D54**, 1332 (1996).
47. M. Tegmark *et al.*, Phys. Rev. **D63**, 043007 (2001).
48. D.N. Spergel *et al.*, Astrophys. J. Supp. **148**, 175 (2003).
49. A. Blanchard *et al.*, astro-ph/0304237;
    S. Bridle *et al.*, MNRAS **342**, L72 (2003).
50. K. Jedamzik and J.B. Rehm, Phys. Rev. **D64**, 023510 (2001).
51. D.J. Fixsen *et al.*, Astrophys. J. **473**, 576 (1996).
52. R. Trotta and S.H. Hansen, astro-ph/0306588.
53. P.C. Stancil *et al.*, Astrophys. J. **580**, 29 (2002).
54. R.H. Cyburt *et al.*, Phys. Lett. **B567**, 227 (2003).
55. G. Steigman *et al.*, Phys. Lett. **B66**, 202 (1977).
56. K.A. Olive and D. Thomas, Astropart. Phys. **11**, 403 (1999)..
57. E. Lisi *et al.*, Phys. Rev. **D59**, 123520 (1999).
58. S. Burles *et al.*, Phys. Rev. **D63**, 063512 (2001).
59. V. Barger *et al.*, Phys. Lett. **B566**, 8 (2003).
60. V. Barger *et al.*, Phys. Lett. **B569**, 123 (2003).
61. E.W. Kolb *et al.*, Phys. Rev. **D33**, 869 (1986);
    F.S. Accetta *et al.*, Phys. Lett. **B248**, 146 (1990);
    B.A. Campbell and K.A. Olive, Phys. Lett. **B345**, 429 (1995);
    K.M. Nollett and R. Lopez, Phys. Rev. **D66**, 063507 (2002).
62. K.A. Olive *et al.*, Nucl. Phys. **B180**, 497 (1981).
63. J. Ellis *et al.*, Phys. Lett. **B167**, 457 (1986).
64. L.M. Krauss *et al.*, Phys. Rev. Lett. **71**, 823 (1993).
65. J.A. Morgan, Phys. Lett. **B102**, 247 (1981).
66. E.W. Kolb *et al.*, Phys. Rev. **D34**, 2197 (1986);
    J.A. Grifols and E. Massó, Mod. Phys. Lett. **A2**, 205 (1987);
    K.S. Babu *et al.*, Phys. Rev. Lett. **67**, 545 (1991).
67. A.D. Dolgov *et al.*, Nucl. Phys. **B524**, 621 (1998).
68. K. Enqvist *et al.*, Nucl. Phys. **B373**, 498 (1992);
    A.D. Dolgov, Phys. Reports **370**, 333 (2002).
69. K. Sato and M. Kobayashi, Prog. Theor. Phys. **58**, 1775 (1977);
    D.A. Dicus *et al.*, Phys. Rev. **D17**, 1529 (1978);
    R.J. Scherrer and M.S. Turner, Astrophys. J. **331**, 19 (1988).
70. D. Lindley, MNRAS **188**, 15 (1979), Astrophys. J. **294**, 1 (1985).
71. J. Ellis *et al.*, Nucl. Phys. **B373**, 399 (1992);
    R.H. Cyburt *et al.*, Phys. Rev. **D67**, 103521 (2003).
72. M.H. Reno and D. Seckel, Phys. Rev. **D37**, 3441 (1988);
    S. Dimopoulos *et al.*, Nucl. Phys. **B311**, 699 (1989);
    K. Kohri, Phys. Rev. **D64**, 043515 (2001).
73. S. Sarkar and A.M. Cooper, Phys. Lett. **B148**, 347 (1984).
74. S. Dodelson *et al.*, Phys. Rev. **D49**, 5068 (1994).
75. J.A. Grifols *et al.*, Phys. Lett. **B400**, 124 (1997).
76. S. Weinberg, Phys. Rev. Lett. **48**, 1303 (1979).
77. M. Bolz *et al.*, Nucl. Phys. **B606**, 518 (2001).
78. G. Coughlan *et al.*, Phys. Lett. **B131**, 59 (1983).
79. N. Arkani-Hamed *et al.*, Phys. Rev. **D59**, 086004 (1999).
80. L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999);
    Phys. Rev. Lett. **83**, 4690 (1999).
81. J.M. Cline *et al.*, Phys. Rev. Lett. **83**, 4245 (1999);
    P. Binetruy *et al.*, Phys. Lett. **B477**, 285 (2000).

21. THE COSMOLOGICAL PARAMETERS

Written August 2003 by O. Lahav (University of Cambridge) and A.R. Liddle (University of Sussex).

21.1. Parametrizing the Universe

Rapid advances in observational cosmology are leading to the establishment of the first precision cosmological model, with many of the key cosmological parameters determined to one or two significant figure accuracy. Particularly prominent are measurements of cosmic microwave anisotropies, led by the first results from the Wilkinson Microwave Anisotropy Probe (WMAP) announced in February 2003 [1]. However the most accurate model of the Universe requires consideration of a wide range of different types of observation, with complementary probes providing consistency checks, lifting parameter degeneracies, and enabling the strongest constraints to be placed.

The term 'cosmological parameters' is forever increasing in its scope, and nowadays includes the parametrization of some functions, as well as simple numbers describing properties of the Universe. The original usage referred to the parameters describing the global dynamics of the Universe, such as its expansion rate and curvature. Also now of great interest is how the matter budget of the Universe is built up from its constituents: baryons, photons, neutrinos, dark matter, and dark energy. We are interested in describing the nature of perturbations in the Universe, through global statistical descriptions such as the matter and radiation power spectra. There may also be parameters describing the physical state of the Universe, most prominent being the ionization fraction as a function of time during the era since decoupling. Typical comparisons of cosmological models with observational data now feature about ten parameters.

21.1.1. *The global description of the Universe*:

Ordinarily, the Universe is taken to be a perturbed Robertson-Walker space-time with dynamics governed by Einstein's equations. This is described in detail by Olive and Peacock in this volume. Using the density parameters Ω_i for the various matter species and Ω_Λ for the cosmological constant, the Friedmann equation can be written

$$\sum_i \Omega_i + \Omega_\Lambda = \frac{k}{R^2 H^2}\,, \tag{21.1}$$

where the sum is over all the different species of matter in the Universe. This equation applies at any epoch, but later in this article we will use the symbols Ω_i and Ω_Λ to refer to the present values. A typical collection would be baryons, photons, neutrinos, and dark matter (given charge neutrality, the electron density is guaranteed to be too small to be worth considering separately).

The complete present state of the homogeneous Universe can be described by giving the present values of all the density parameters and the present Hubble parameter h, and indeed one of the density parameters can be eliminated using Eq. (21.1). These also allow us to track the history of the Universe back in time, at least until an epoch where interactions allow interchanges between the densities of the different species, which is believed to have last happened at neutrino decoupling shortly before nucleosynthesis. To probe further back into the Universe's history requires assumptions about particle interactions, and perhaps about the nature of physical laws themselves.

21.1.2. *Neutrinos*:

The standard neutrino sector has three flavors. For neutrinos of mass in the range 5×10^{-4} eV to 1 MeV, the density parameter in neutrinos is predicted to be

$$\Omega_\nu h^2 = \frac{\sum m_\nu}{94\,\mathrm{eV}}\,, \tag{21.2}$$

where the sum is over all families with mass in that range (higher masses need a more sophisticated calculation). We use units with $c = 1$ throughout. Recent results on atmospheric and solar neutrino oscillations [2] imply non-zero mass-squared differences between the three neutrino flavors. These oscillation experiments cannot tell us the absolute neutrino masses, but within the simple assumption of a mass hierarchy suggest a lower limit of $\Omega_\nu \approx 0.001$ on the neutrino mass density parameter.

For a total mass as small as 0.1 eV, this could have a potentially observable effect on the formation of structure, as neutrino free-streaming damps the growth of perturbations. Present cosmological observations have shown no convincing evidence of any effects from either neutrino masses or an otherwise non-standard neutrino sector, and impose quite stringent limits, which we summarize in Section 21.3.4. Consequently, the standard assumption at present is that the masses are too small to have a significant cosmological impact, but this may change in the near future.

The cosmological effect of neutrinos can also be modified if the neutrinos have decay channels, or if there is a large asymmetry in the lepton sector manifested as a different number density of neutrinos versus anti-neutrinos. This latter effect would need to be of order unity to be significant, rather than the 10^{-9} seen in the baryon sector, which may be in conflict with nucleosynthesis [3].

21.1.3. *Inflation and perturbations*:

A complete description of the Universe should include a description of deviations from homogeneity, at least in a statistical way. Indeed, some of the most powerful probes of the parameters described above come from studying the evolution of perturbations, so their study is naturally intertwined in the determination of cosmological parameters.

There are many different notations used to describe the perturbations, both in terms of the quantity used to describe the perturbations and the definition of the statistical measure. We use the dimensionless power spectrum Δ^2 as defined in Olive and Peacock (also denoted $\mathcal{P}$ in some of the literature). If the perturbations obey Gaussian statistics, the power spectrum provides a complete description of their properties.

From a theoretical perspective, a useful quantity to describe the perturbations is the curvature perturbation $\mathcal{R}$, which measures the spatial curvature of a comoving slicing of the space-time. A case of particular interest is the Harrison-Zel'dovich spectrum, which corresponds to a constant spectrum $\Delta^2_{\mathcal{R}}$. More generally, one can approximate the spectrum by a power-law, writing

$$\Delta^2_{\mathcal{R}}(k) = \Delta^2_{\mathcal{R}}(k_*) \left[\frac{k}{k_*}\right]^{n-1}\,, \tag{21.3}$$

where n is known as the spectral index, always defined so that $n = 1$ for the Harrison-Zel'dovich spectrum, and k_* is an arbitrarily chosen scale. The initial spectrum, defined at some early epoch of the Universe's history, is usually taken to have a simple form such as this power-law, and we will see that observations require n close to one, which corresponds to the perturbations in the curvature being independent of scale. Subsequent evolution will modify the spectrum from its initial form.

The simplest viable mechanism for generating the observed perturbations is the inflationary cosmology, which posits a period of accelerated expansion in the Universe's early stages [4]. It is a useful working hypothesis that this is the sole mechanism for generating perturbations. Commonly, it is further assumed to be the simplest class of inflationary model, where the dynamics are equivalent to that of a single scalar field ϕ slowly rolling on a potential $V(\phi)$. One aim of cosmology is to verify that this simple picture can match observations, and to determine the properties of $V(\phi)$ from the observational data.

Inflation generates perturbations through the amplification of quantum fluctuations, which are stretched to astrophysical scales by the rapid expansion. The simplest models generate two types, density perturbations which come from fluctuations in the scalar field and its corresponding scalar metric perturbation, and gravitational waves which are tensor metric fluctuations. The former experience gravitational instability and lead to structure formation, while the latter can influence the cosmic microwave background anisotropies. Defining slow-roll parameters, with primes indicating derivatives with respect to the scalar field, as

$$\epsilon = \frac{m_{\rm Pl}^2}{16\pi}\left(\frac{V'}{V}\right)^2 \quad ; \quad \eta = \frac{m_{\rm Pl}^2}{8\pi}\frac{V''}{V}\,, \tag{21.4}$$

which should satisfy $\epsilon, |\eta| \ll 1$, the spectra can be computed using the slow-roll approximation as

$$\Delta^2_{\mathcal{R}}(k) \simeq \frac{8}{3m_{\rm Pl}^4}\frac{V}{\epsilon}\bigg|_{k=aH} \quad ;$$
$$\Delta^2_{\rm grav}(k) \simeq \frac{128}{3m_{\rm Pl}^4} V\bigg|_{k=aH} . \qquad (21.5)$$

In each case, the expressions on the right-hand side are to be evaluated when the scale k is equal to the Hubble radius during inflation. The symbol '$\simeq$' indicates use of the slow-roll approximation, which is expected to be accurate to a few percent or better.

From these expressions, we can compute the spectral indices

$$n \simeq 1 - 6\epsilon + 2\eta \quad ; \quad n_{\rm grav} \simeq -2\epsilon . \qquad (21.6)$$

Another useful quantity is the ratio of the two spectra, defined by

$$r \equiv \frac{\Delta^2_{\rm grav}(k_*)}{\Delta^2_{\mathcal{R}}(k_*)} . \qquad (21.7)$$

The literature contains a number of definitions of r; this convention matches that of recent versions of CMBFAST [5] and of WMAP [6], while definitions based on the relative effect on the microwave background anisotropies typically differ by tens of percent. We have

$$r \simeq 16\epsilon \simeq -8 n_{\rm grav} , \qquad (21.8)$$

which is known as the consistency equation.

In general one could consider corrections to the power-law approximation, and indeed WMAP found some low-significance evidence that this might be needed, which we discuss later. However for now we make the working assumption that the spectra can be approximated by power laws. The consistency equation shows that r and $n_{\rm grav}$ are not independent parameters, and so the simplest inflation models give initial conditions described by three parameters, usually taken as $\Delta^2_{\mathcal{R}}$, n, and r, all to be evaluated at some scale k_*, usually the 'statistical centre' of the range explored by the data. Alternatively, one could use the parametrization V, ϵ, and η, all evaluated at a point on the putative inflationary potential.

After the perturbations are created in the early Universe, they undergo a complex evolution up until the time they are observed in the present Universe. While the perturbations are small, this can be accurately followed using a linear theory numerical code such as CMBFAST [5]. This works right up to the present for the cosmic microwave background, but for density perturbations on small scales non-linear evolution is important and can be addressed by a variety of semi-analytical and numerical techniques. However the analysis is made, the outcome of the evolution is in principle determined by the cosmological model, and by the parameters describing the initial perturbations, and hence can be used to determine them.

Of particular interest are cosmic microwave background anisotropies. Both the total intensity and two independent polarization modes are predicted to have anisotropies. These can be described by the radiation angular power spectra C_ℓ as defined in the article of Scott and Smoot in this volume, and again provide a complete description if the density perturbations are Gaussian.

21.1.4. *The standard cosmological model*:

We now have most of the ingredients in place to describe the cosmological model. Beyond those of the previous subsections, there is only one parameter which is essential, which is a measure of the ionization state of the Universe. The Universe is known to be highly ionized at low redshifts (otherwise radiation from distant quasars would be heavily absorbed in the ultra-violet), and the ionized electrons can scatter microwave photons altering the pattern of observed anisotropies. The most convenient parameter to describe this is the optical depth to scattering τ (i.e. the probability that a given photon scatters once); in the approximation of instantaneous and complete re-ionization, this could equivalently be described by the redshift of re-ionization $z_{\rm ion}$.

Table 21.1: The basic set of cosmological parameters. We give values as obtained using particular fit to a dataset known as WMAPext+2dF, described later. We cannot stress too much that the exact values and uncertainties depend on both the precise datasets used and the choice of parameters allowed to vary, and the effects of varying some assumptions will be shown later in Table 21.2. Limits on the cosmological constant depend on whether the Universe is assumed flat, while there is no established convention for specifying the density perturbation amplitude. Uncertainties are one-sigma/68% confidence unless otherwise stated.

Parameter	Symbol	Value
Hubble parameter	h	0.73 ± 0.03
Total matter density	$\Omega_{\rm m}$	$\Omega_{\rm m}h^2 = 0.134 \pm 0.006$
Baryon density	$\Omega_{\rm b}$	$\Omega_{\rm b}h^2 = 0.023 \pm 0.001$
Cosmological constant	Ω_Λ	See Ref. 7
Radiation density	$\Omega_{\rm r}$	$\Omega_{\rm r}h^2 = 2.47 \times 10^{-5}$
Neutrino density	Ω_ν	See Sec. 21.1.2
Density perturbation amplitude	$\Delta^2_{\mathcal{R}}(k_*)$	See Ref. 7
Density perturbation spectral index	n	$n = 0.97 \pm 0.03$
Tensor to scalar ratio	r	$r < 0.53$ (95% conf)
Ionization optical depth	τ	$\tau = 0.15 \pm 0.07$

The basic set of cosmological parameters is therefore as shown in Table 21.1. The spatial curvature does not appear in the list, because it can be determined from the other parameters using Eq. (21.1). The total present matter density $\Omega_{\rm m} = \Omega_{\rm dm} + \Omega_{\rm b}$ is usually used in place of the dark matter density.

As described in Sec. 21.4, models based on these ten parameters are able to give a good fit to the complete set of high-quality data available at present, and indeed some simplification is possible. Observations are consistent with spatial flatness, and indeed the inflation models so far described automatically generate spatial flatness, so we can set $k = 0$; the density parameters then must sum to one, and so one can be eliminated. The neutrino energy density is often not taken as an independent parameter. Provided the neutrino sector has the standard interactions the neutrino energy density while relativistic can be related to the photon density using thermal physics arguments, and it is currently difficult to see the effect of the neutrino mass although observations of large-scale structure have already placed interesting upper limits. This reduces the standard parameter set to eight. In addition, there is no observational evidence for the existence of tensor perturbations (though the upper limits are quite weak), and so r could be set to zero.* This leaves seven parameters, which is the smallest set that can usefully be compared to the present cosmological data set. This model is referred to by various names, including ΛCDM, the concordance cosmology, and the standard cosmological model.

Of these parameters, only $\Omega_{\rm r}$ is accurately measured directly. The radiation density is dominated by the energy in the cosmic microwave background, and the COBE FIRAS experiment has determined its temperature to be $T = 2.725 \pm 0.001$ Kelvin [8], corresponding to $\Omega_{\rm r} = 2.47 \times 10^{-5} h^{-2}$.

In addition to this minimal set, there is a range of other parameters which might prove important in future as the dataset further improves, but for which there is so far no direct evidence, allowing them to be set to a specific value. We discuss various speculative options in the next section. For completeness at this point, we mention one other interesting parameter, the helium fraction, which is a non-zero parameter that can affect the microwave anisotropies at a subtle level. Presently, big-bang nucleosynthesis provides the best measurement of this parameter, and it is usually fixed in microwave anisotropy studies, but the data are just reaching a level where allowing its variation may become mandatory.

* More controversially, one could argue that as no evidence against the Harrison-Zel'dovich spectrum has yet been seen, then n could be set to one. We will however allow it to vary.

21.1.5. ***Derived parameters*:**

The parameter list of the previous subsection is sufficient to give a complete description of cosmological models which agree with observational data. However, it is not a unique parametrization, and one could instead use parameters derived from that basic set. Parameters which can be derived from the set given above include the age of the Universe, the present horizon distance, the present microwave background and neutrino background temperatures, the epoch of matter-radiation equality, the epochs of recombination and decoupling, the epoch of transition to an accelerating Universe, the baryon-to-photon ratio, and the baryon to dark matter density ratio. The physical densities of the matter components, $\Omega_i h^2$, are often more useful than the density parameters. The density perturbation amplitude can be specified in many different ways other than the large-scale primordial amplitude, for instance, in terms of its effect on the cosmic microwave background, or by specifying a short-scale quantity, a common choice being the present linear-theory mass dispersion on a scale of $8\,h^{-1}$Mpc, known as σ_8.

Different types of observation are sensitive to different subsets of the full cosmological parameter set, and some are more naturally interpreted in terms of some of the derived parameters of this subsection than on the original base parameter set. In particular, most types of observation feature degeneracies whereby they are unable to separate the effects of simultaneously varying several of the base parameters, an example being the angular diameter/physical density degeneracy of cosmic microwave anisotropies.

21.2. Extensions to the standard model

This section discusses some ways in which the standard model could be extended. At present, there is no positive evidence in favor of any of these possibilities, which are becoming increasingly constrained by the data, though there always remains the possibility of trace effects at a level below present observational capability.

21.2.1. ***More general perturbations*:**

The standard cosmology assumes adiabatic, Gaussian perturbations. Adiabaticity means that all types of material in the Universe share a common perturbation, so that if the space-time is foliated by constant-density hypersurfaces, then all fluids and fields are homogeneous on those slices, with the perturbations completely described by the variation of the spatial curvature of the slices. Gaussianity means that the initial perturbations obey Gaussian statistics, with the amplitudes of waves of different wavenumbers being randomly drawn from a Gaussian distribution of width given by the power spectrum. Note that gravitational instability generates non-Gaussianity; in this context, Gaussianity refers to a property of the initial perturbations before they evolve significantly.

The simplest inflation models based on one dynamical field predict adiabatic fluctuations and a level of non-Gaussianity which is too small to be detected by any experiment so far conceived. For present data, the primordial spectra are usually assumed to be power laws.

21.2.1.1. *Non-power-law spectra:*

For typical inflation models, it is an approximation to take the spectra as power laws, albeit usually a good one. As data quality improves, one might expect this approximation to come under pressure, requiring a more accurate description of the initial spectra, particularly for the density perturbations. In general, one can write a Taylor expansion of $\ln \Delta^2_{\mathcal{R}}$ as

$$\ln \Delta^2_{\mathcal{R}}(k) = \ln \Delta^2_{\mathcal{R}}(k_*) + (n_* - 1) \ln \frac{k}{k_*} + \frac{1}{2} \left. \frac{dn}{d\ln k} \right|_* \ln^2 \frac{k}{k_*} + \cdots , \quad (21.9)$$

where the coefficients are all evaluated at some scale k_*. The term $dn/d\ln k|_*$ is often called the running of the spectral index [9], and has recently become topical due to a possible low-significance detection by WMAP. Once non-power-law spectra are allowed, it is necessary to specify the scale k_* at which quantities such as the spectral index are defined.

21.2.1.2. *Isocurvature perturbations:*

An isocurvature perturbation is one which leaves the total density unperturbed, while perturbing the relative amounts of different materials. If the Universe contains N fluids, there is one growing adiabatic mode and $N-1$ growing isocurvature modes. These can be excited, for example, in inflationary models where there are two or more fields which acquire dynamically important perturbations. If one field decays to form normal matter, while the second survives to become the dark matter, this will generate a cold dark matter isocurvature perturbation.

In general there are also correlations between the different modes, and so the full set of perturbations is described by a matrix giving the spectra and their correlations. Constraining such a general construct is challenging, though constraints on individual modes are beginning to become meaningful, with no evidence that any other than the adiabatic mode must be non-zero.

21.2.1.3. *Non-Gaussianity:*

Multi-field inflation models can also generate primordial non-Gaussianity. The extra fields can either be in the same sector of the underlying theory as the inflaton, or completely separate, an interesting example of the latter being the curvaton model [10]. Current upper limits on non-Gaussianity are becoming stringent, but there remains much scope to push down those limits and perhaps reveal trace non-Gaussianity in the data. If non-Gaussianity is observed, its nature may favor an inflationary origin, or a different one such as topological defects. A plausible possibility is non-Gaussianity caused by defects forming in a phase transition which ended inflation.

21.2.2. ***Dark matter properties*:**

Dark matter properties are discussed in the article by Drees and Gerbier in this volume. The simplest assumption concerning the dark matter is that it has no significant interactions with other matter, and that its particles have a negligible velocity. Such dark matter is described as 'cold,' and candidates include the lightest supersymmetric particle, the axion, and primordial black holes. As far as astrophysicists are concerned, a complete specification of the relevant cold dark matter properties is given by the density parameter $\Omega_{\rm cdm}$, though those seeking to directly detect it are as interested in its interaction properties.

Cold dark matter is the standard assumption and gives an excellent fit to observations, except possibly on the shortest scales where there remains some controversy concerning the structure of dwarf galaxies and possible substructure in galaxy halos. For all the dark matter to have a large velocity dispersion, so-called hot dark matter, has long been excluded as it does not permit galaxies to form; for thermal relics the mass must be above about 1 keV to satisfy this constraint, though relics produced non-thermally, such as the axion, need not obey this limit. However, there remains the possibility that further parameters might need to be introduced to describe dark matter properties relevant to astrophysical observations. Suggestions which have been made include a modest velocity dispersion (warm dark matter) and dark matter self-interactions. There remains the possibility that the dark matter comprises two separate components, *e.g.*, a cold one and a hot one, an example being if massive neutrinos have a non-negligible effect.

21.2.3. ***Dark energy*.**

While the standard cosmological model given above features a cosmological constant, in order to explain observations indicating that the Universe is presently accelerating, further possibilities exist under the general heading dark energy.† A particularly attractive possibility (usually called quintessence, though that word is used with various different meanings in the literature) is that a scalar field is responsible, with the mechanism mimicking that of early Universe inflation [11]. As described by Olive and Peacock, a fairly model-independent description of dark energy can be given just using the equation of state parameter w, with $w = -1$ corresponding to a cosmological constant.

† Unfortunately this is rather a misnomer, as it is the negative pressure of this material, rather than its energy, that is responsible for giving the acceleration.

In general, the function w could itself vary with redshift, though practical experiments devised so far would be sensitive primarily to some average value weighted over recent epochs. For high-precision predictions of microwave background anisotropies, it is better to use a scalar-field description in order to have a self-consistent evolution of the 'sound speed' associated with the dark energy perturbations.

Present observations are consistent with a cosmological constant, but it is quite common to see w kept as a free parameter to be added to the set described in the previous section. Most, but not all, researchers assume the weak energy condition $w \geq -1$. In the future it may be necessary to use a more sophisticated parametrization of the dark energy.

21.2.4. *Complex ionization history*:

The full ionization history of the Universe is given by specifying the ionization fraction as a function of redshift z. The simplest scenario takes the ionization to be zero from recombination up to some redshift $z_{\rm ion}$, at which point the Universe instantaneously re-ionizes completely. In that case, there is a one-to-one correspondence between τ and $z_{\rm ion}$ (that relation, however, also depending on other cosmological parameters).

While simple models of the re-ionization process suggest that rapid ionization is a good approximation, observational evidence is mixed, as it is difficult to reconcile the high optical depth inferred from the microwave background with absorption seen in some high-redshift quasar systems, and also perhaps with the temperature of the intergalactic medium at $z \simeq 3$. Accordingly, a more complex ionization history may need to be considered, and perhaps separate histories for hydrogen and helium, which will necessitate new parameters. Additionally, high-precision microwave anisotropy experiments may require consideration of the level of residual ionization left after recombination, which in principle is computable from the other cosmological parameters.

21.2.5. *Varying 'constants'*:

Variation of the fundamental constants of nature over cosmological times is another possible enhancement of the standard cosmology. There is a long history of study of variation of the gravitational constant G, and more recently attention has been drawn to the possibility of small fractional variations in the fine-structure constant. There is presently no observational evidence for the former, which is tightly constrained by a variety of measurements. Evidence for the latter has been claimed from studies of spectral line shifts in quasar spectra at redshifts of order two [12], but this is presently controversial and in need of further observational study.

21.2.6. *Cosmic topology*:

The usual hypothesis is that the Universe has the simplest topology consistent with its geometry, for example that a flat Universe extends forever. Observations cannot tell us whether that is true, but they can test the possibility of a non-trivial topology on scales up to roughly the present Hubble scale. Extra parameters would be needed to specify both the type and scale of the topology, for example, a cuboidal topology would need specification of the three principal axis lengths. At present, there is no direct evidence for cosmic topology, though the low values of the observed cosmic microwave quadrupole and octupole have been cited as a possible signature.

21.3. Probes

The goal of the observational cosmologist is to utilize astronomical objects to derive cosmological parameters. The transformation from the observables to the key parameters usually involves many assumptions about the nature of the objects, as well as about the nature of the dark matter. Below we outline the physical processes involved in each probe, and the main recent results. The first two subsections concern probes of the homogeneous Universe, while the remainder consider constraints from perturbations.

21.3.1. *Direct measures of the Hubble constant*:

In 1929 Edwin Hubble discovered the law of expansion of the Universe by measuring distances to nearby galaxies. The slope of the relation between the distance and recession velocity is defined to be the Hubble constant H_0. Astronomers argued for decades on the systematic uncertainties in various methods and derived values over the wide range, $40\,{\rm km\,s^{-1}\,Mpc^{-1}} \lesssim H_0 \lesssim 100\,{\rm km\,s^{-1}\,Mpc^{-1}}$.

One of the most reliable results on the Hubble constant comes from the Hubble Space Telescope Key Project [13]. The group has used the empirical period-luminosity relations for Cepheid variable stars to obtain distances to 31 galaxies, and calibrated a number of secondary distance indicators (Type Ia Supernovae, Tully-Fisher, surface brightness fluctuations and Type II Supernovae) measured over distances of 400 to 600 Mpc. They estimated $H_0 = 72 \pm 3$ (statistical) ± 7 (systematic) ${\rm km\,s^{-1}\,Mpc^{-1}}$.[‡] The major sources of uncertainty in this result are due to the metallicity of the Cepheids and the distance to the fiducial nearby galaxy (called the Large Magellanic Cloud) to which all Cepheid distances are measured relative to. Nevertheless, it is remarkable that this result is in such good agreement with the result derived from the WMAP CMB and large-scale structure measurements (see Table 21.2).

21.3.2. *Supernovae as cosmological probes*:

The relation between observed flux and the intrinsic luminosity of an object depends on the luminosity distance d_L, which in turn depends on cosmological parameters. More specifically

$$d_L = (1+z) r_e(z)\,, \tag{21.10}$$

where $r_e(z)$ is the coordinate distance. For example, in a flat Universe

$$r_e(z) = \int_0^z dz'/H(z')\,. \tag{21.11}$$

For a general dark energy equation of state $w(z) = p_Q(z)/\rho_Q(z)$, the Hubble parameter is, still considering only the flat case,

$$H^2(z)/H_0^2 = (1+z)^3\Omega_{\rm m} + \Omega_Q \exp[3X(z)]\,, \tag{21.12}$$

where

$$X(z) = \int_0^z [1+w(z')](1+z')^{-1} dz'\,, \tag{21.13}$$

and $\Omega_{\rm m}$ and Ω_Q are the present density parameters of matter and dark energy components. If a general equation of state is allowed, then one has to solve for $w(z)$ (parameterized, for example, as $w(z) = w =$ constant, or $w(z) = w_0 + w_1 z$) as well as for Ω_Q.

Empirically, the peak luminosity of supernova of Type Ia (SNe Ia) can be used as an efficient distance indicator (*e.g.*, Ref. 14). The favorite theoretical explanation for SNe Ia is the thermonuclear disruption of carbon-oxygen white dwarfs. Although not perfect 'standard candles', it has been demonstrated that by correcting for a relation between the light curve shape and the luminosity at maximum brightness, the dispersion of the measured luminosities can be greatly reduced. There are several possible systematic effects which may affect the accuracy of the SNe Ia as distance indicators, for example, evolution with redshift and interstellar extinction in the host galaxy and in the Milky Way, but there is no indication that any of these effects are significant for the cosmological constraints.

Two major studies, the 'Supernova Cosmology Project' and the 'High-z Supernova Search Team', found evidence for an accelerating Universe [15], interpreted as due to a cosmological constant, or to a more general 'dark energy' component. Recent results obtained by Tonry *et al.* [16] are shown in Fig. 21.1 (see also Ref. 17). The SNe Ia data alone can only constrain a combination of $\Omega_{\rm m}$ and Ω_Λ. When combined with the CMB data (which indicates flatness, *i.e.*, $\Omega_{\rm m} + \Omega_\Lambda \approx 1$), the best-fit values are $\Omega_{\rm m} \approx 0.3$ and $\Omega_\Lambda \approx 0.7$. Future experiments will aim to set constraints on the cosmic equation of state $w(z)$. However, given the integral relation between the luminosity distance and $w(z)$, it is not straightforward to recover $w(z)$ (*e.g.*, Ref. 18).

[‡] Unless stated otherwise, all quoted uncertainties in this article are one-sigma/68% confidence. It is common for cosmological parameters to have significantly non-Gaussian error distributions.

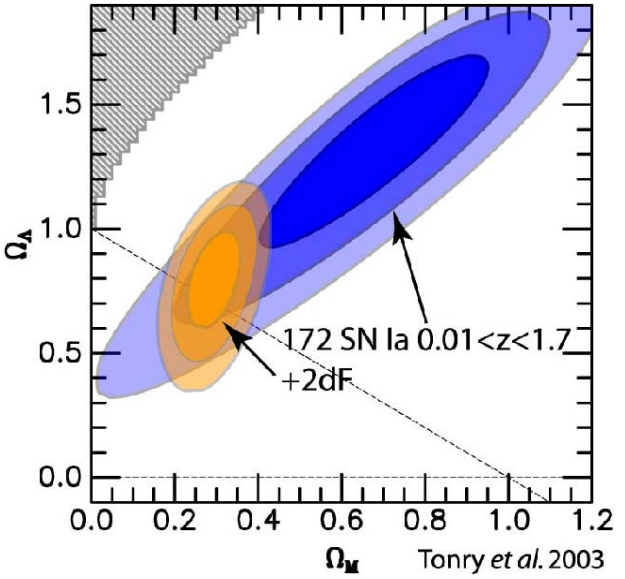

Figure 21.1: This shows the preferred region in the Ω_m–Ω_Λ plane from a study of 172 supernovae, and also how the constraints tighten when the 2dF galaxy redshift survey power spectrum is added as an additional constraint. [Reproduced with permission from Tonry *et al.* [16].] See full-color version on color pages at end of book.

21.3.3. *Cosmic microwave background*:

The physics of the cosmic microwave background (CMB) is described in detail by Scott and Smoot in this volume. Before recombination, the baryons and photons are tightly coupled, and the perturbations oscillate in the potential wells generated primarily by the dark matter perturbations. After decoupling, the baryons are free to collapse into those potential wells. The CMB carries a record of conditions at the time of decoupling, often called primary anisotropies. In addition, it is affected by various processes as it propagates towards us, including the effect of a time-varying gravitational potential (the integrated Sachs-Wolfe effect), gravitational lensing, and scattering from ionized gas at low redshift.

The primary anisotropies, the integrated Sachs-Wolfe effect, and scattering from a homogeneous distribution of ionized gas, can all be calculated using linear perturbation theory, a widely-used implementation being the CMBFAST code of Seljak and Zaldarriaga [5]. Gravitational lensing is also calculated in this code. Secondary effects such as inhomogeneities in the re-ionization process, and scattering from gravitationally-collapsed gas (the Sunyaev-Zel'dovich effect), require more complicated, and more uncertain, calculations.

The upshot is that the detailed pattern of anisotropies, quantified, for instance, by the angular power spectrum C_ℓ, depends on all of the cosmological parameters. In a typical cosmology, the anisotropy power spectrum [usually plotted as $\ell(\ell+1)C_\ell$] features a flat plateau at large angular scales (small ℓ), followed by a series of oscillatory features at higher angular scales, the first and most prominent being at around one degree ($\ell \simeq 200$). These features, known as acoustic peaks, represent the oscillations of the photon-baryon fluid around the time of decoupling. Some features can be closely related to specific parameters—for instance, the location of the first peak probes the spatial geometry, while the relative heights of the peaks probes the baryon density—but many other parameters combine to determine the overall shape.

The WMAP experiment [1] has provided the most accurate results to date on the spectrum of CMB fluctuations [19], with a precision determination of the temperature power spectrum up to $\ell \simeq 900$, shown in Fig. 21.2, and the first detailed measurement of the correlation spectrum between temperature and polarization [20] (the correlation having first been detected by DASI [21]). These are consistent with models based on the parameters we have described, and provide quite accurate determinations of many of them [7]. In this subsection, we will refer to results from WMAP alone, with the following section combining those with other observations. We note that as the parameter fitting is done in a multi-parameter space, one has to assume a 'prior' range for each of the parameters (*e.g.*, Hubble constant $0.5 < h < 1$), and there may be some dependence on these assumed priors.

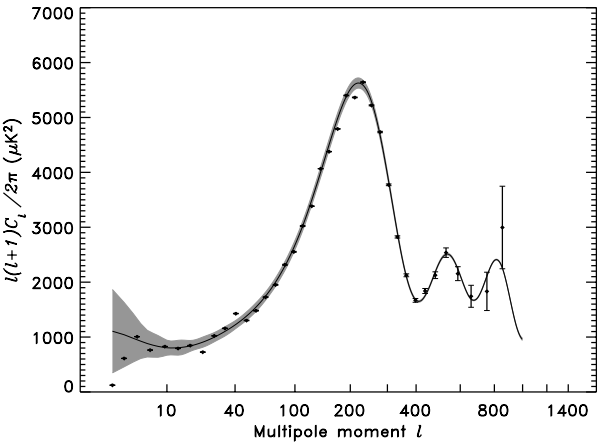

Figure 21.2: The angular power spectrum of the cosmic microwave background as measured by the WMAP satellite. The solid line shows the prediction from the best-fitting ΛCDM model [7]. The error bars on the data points (which are tiny for most of them) indicate the observational errors, while the shaded region indicates the statistical uncertainty from being able to observe only one microwave sky, known as cosmic variance, which is the dominant uncertainty on large angular scales. [Figure courtesy NASA/WMAP Science Team.]

WMAP provides an exquisite measurement of the location of the first acoustic peak, which directly probes the spatial geometry and yields a total density $\Omega_{\rm tot} \equiv \sum \Omega_i + \Omega_\Lambda$ of

$$\Omega_{\rm tot} = 1.02 \pm 0.02\,, \tag{21.14}$$

consistent with spatial flatness and completely excluding significantly curved Universes (this result does however assume a fairly strong prior on the Hubble parameter from other measurements; WMAP alone constrains it only weakly, and allows significantly closed Universes if h is small, e.g. $\Omega_{\rm tot} = 1.3$ for $h = 0.3$). It also gives a precision measurement of the age of the Universe. It gives a baryon density consistent with that coming from nucleosynthesis, and affirms the need for both dark matter and dark energy if the data are to be explained. For the spectral index of density perturbations, WMAP alone is consistent with a power-law spectrum, with spectral index $n = 0.99 \pm 0.04$, and in particular with a scale-invariant initial spectrum $n = 1$. It shows no evidence for dynamics of the dark energy, being consistent with a pure cosmological constant ($w = -1$).

One of the most interesting results, driven primarily by detection of large-angle polarization-temperature correlations, is the discovery of a high optical depth to re-ionization, $\tau \sim 0.17$, which roughly corresponds to a re-ionization redshift $z_{\rm ion} \sim 17$. This was higher than expected, though it appears it can be accommodated in models for development of the first structures which provide the ionizing flux.

In addition to WMAP, useful information comes from measurements of the CMB on small angular scales by, amongst others, the ACBAR and CBI experiments. Further, in 2002 the DASI experiment made the first measurement of the polarization anisotropies [21], again consistent with the standard cosmology, though not with sufficient accuracy to provide detailed constraints.

21.3.4. ***Galaxy clustering:***

The power spectrum of density perturbations depends on the nature of the dark matter. Within the Cold Dark Matter model, the shape of the power spectrum depends primarily on the primordial power spectrum and on the combination $\Omega_{\rm m}h$ which determines the horizon scale at matter-radiation equality, with a subdominant dependence on the baryon density. The matter distribution is most easily probed by observing the galaxy distribution, but this must be done with care as the galaxies do not perfectly trace the dark matter distribution. Rather, they are a 'biased' tracer of the dark matter. The need to allow for such bias is emphasized by the observation that different types of galaxies show bias with respect to each other. Further, the observed 3D galaxy distribution is in redshift space, *i.e.*, the observed redshift is the sum of the Hubble expansion and the line-of-sight peculiar velocity, leading to linear and non-linear dynamical effects which also depend on the cosmological parameters. On the largest length scales, the galaxies are expected to trace the location of the dark matter, except for a constant multiplier b to the power spectrum, known as the linear bias parameter. On scales smaller than 20 h^{-1} Mpc or so, the clustering pattern is 'squashed' in the radial direction due to coherent infall, which depends on the parameter $\beta \equiv \Omega_{\rm m}^{0.6}/b$ (on these shorter scales, more complicated forms of biasing are not excluded by the data). On scales of a few h^{-1} Mpc, there is an effect of elongation along the line of sight (colloquially known as the 'finger of God' effect) which depends on the galaxy velocity dispersion $\sigma_{\rm p}$.

21.3.4.1. *The galaxy power spectrum:*

The 2-degree Field (2dF) Galaxy Redshift Survey is now complete and publicly available, with nearly 230,000 redshifts.** Analyses of a subset of the full data (containing 160,000 redshifts) measured the power spectrum for $k > 0.02\,h\,{\rm Mpc}^{-1}$ with $\sim 10\%$ accuracy, shown in Fig. 21.3. The measured power spectrum is well fit by a CDM model with $\Omega_{\rm m}h = 0.18 \pm 0.02$, and a baryon fraction $\Omega_{\rm b}/\Omega_{\rm m} = 0.17 \pm 0.06$ [22]. The pattern of the galaxy clustering in redshift space is fitted by $\beta = 0.49 \pm 0.09$ and velocity dispersion $\sigma_{\rm p} = 506 \pm 52\,{\rm km\,s}^{-1}$ [23]; note that the two are strongly correlated. Combination of the 2dF data with the CMB indicates $b \sim 1$, in agreement with a 2dF-alone analysis of higher-order clustering statistics. Results for these parameters also depend on the length scale over which a fit is done, and the selection of the objects by luminosity, spectral type, or color. In particular, on scales smaller than 10 h^{-1}Mpc, different galaxy types are clustered differently. This 'biasing' introduces a systematic effect on the determination of cosmological parameters from redshift surveys. Prior knowledge from simulations of galaxy formation could help, but is model-dependent. We note that the present-epoch power spectrum is not sensitive to dark energy, so it is mainly a probe of the matter density.

The Sloan Digital Sky Survey (SDSS) is a project to image a quarter of the sky, and to obtain spectra of galaxies and quasars selected from the imaging data.†† A maximum likelihood analysis of early SDSS data by Szalay *et al.* [24] used the projected distribution of galaxies in a redshift bin around $z = 0.33$ to find $\Omega_{\rm m}h = 0.18 \pm 0.04$, assuming a flat ΛCDM model with $\Omega_{\rm m} = 1 - \Omega_\Lambda = 0.3$. The power spectrum of the latest version of SDSS redshift survey was published as this article was being finalized [25].

21.3.4.2. *Limits on neutrino mass from 2dFGRS:*

Large-scale structure data can put an upper limit on the ratio $\Omega_\nu/\Omega_{\rm m}$ due to the neutrino 'free streaming' effect [26]. By comparing the 2dF galaxy power spectrum with a four-component model (baryons, cold dark matter, a cosmological constant, and massive neutrinos), it was estimated that $\Omega_\nu/\Omega_{\rm m} < 0.13$ (95% confidence limit), giving $\Omega_\nu < 0.04$ if a concordance prior of $\Omega_{\rm m} = 0.3$ is imposed. The latter corresponds to an upper limit of about 2 eV on the total neutrino mass, assuming a prior of $h \approx 0.7$ [27]. The above analysis assumes that the primordial power spectrum is adiabatic, scale-invariant and Gaussian. Potential systematic effects include biasing of the galaxy distribution and non-linearities of the power spectrum. Additional cosmological data sets bring down this upper limit by a factor of two [28]. The analysis of WMAP+2dFGRS [7] derived $\Omega_\nu h^2 < 0.0067$ (95% CL).

Laboratory limits on absolute neutrino masses from tritium beta decay and especially from neutrinoless double-beta decay should, within the next decade, push down towards (or perhaps even beyond) the 0.1 eV level that has cosmological significance.

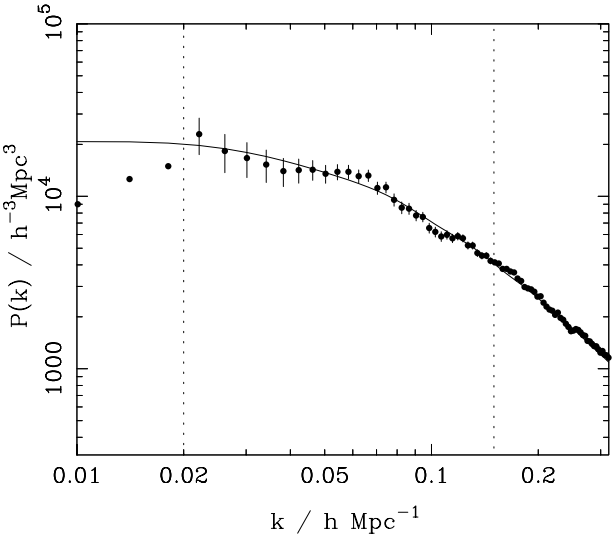

Figure 21.3: The galaxy power spectrum from the 2dF galaxy redshift survey as derived in Ref. 22. This plot shows $P(k) \propto \Delta^2(k)/k^3$, but with distances measured in redshift space and convolved with the survey geometry. The solid line shows a linear-theory ΛCDM fit (also convolved with the survey geometry) with $\Omega_{\rm m}h = 0.2$, $\Omega_{\rm b}/\Omega_{\rm m} = 0.15$, $h = 0.7$ and $n = 1$. Only the range $0.02\,h\,{\rm Mpc}^{-1} < k < 0.15\,h\,{\rm Mpc}^{-1}$, where perturbations are in the linear regime, was used to obtain that best fit. The error bars are correlated, but with known covariances. [Figure provided by Will Percival; see also Ref. 22.]

21.3.5. ***Clusters of galaxies:***

A cluster of galaxies is a large collection of galaxies held together by their mutual gravitational attraction. The largest ones are around 10^{15} solar masses, and are the largest gravitationally-bound structures in the Universe. Even at the present epoch they are relatively rare, with only a few percent of galaxies being in clusters. They provide various ways to study the cosmological parameters; here we discuss constraints from the measurements of the cluster number density and the baryon fraction in clusters.

21.3.5.1. *Cluster number density:* The first objects of a given kind form at the rare high peaks of the density distribution, and if the primordial density perturbations are Gaussian-distributed, their number density is exponentially sensitive to the size of the perturbations, and hence can strongly constrain it. Clusters are an ideal application in the present Universe. They are usually used to constrain the amplitude σ_8, as a box of side $8\,h^{-1}\,{\rm Mpc}$ contains about the right amount of material to form a cluster. The most useful observations at present are of X-ray emission from hot gas lying within the cluster, whose temperature is typically a few keV, and which can be used to estimate the mass of the cluster. A theoretical prediction for the mass function of clusters can come either from semi-analytic arguments or from numerical simulations. At present, the main uncertainty is the relation between the observed gas temperature and the cluster mass, despite extensive study using simulations. A recent analysis [29] gives

$$\sigma_8 = 0.78^{+0.30}_{-0.06} \qquad (95\%{\rm CL}) \tag{21.15}$$

for $\Omega_{\rm m} = 0.35$, with highly non-Gaussian error bars, but different authors still find a spread of values. Scaling to lower $\Omega_{\rm m}$ increases σ_8

** See `http://www.mso.anu.edu.au/2dFGRS`

†† See `http://www.sdss.org`

somewhat, and the result above is consistent with values predicted in cosmologies compatible with WMAP.

The same approach can be adopted at high redshift (which for clusters means redshifts approaching one) to attempt to measure σ_8 at an earlier epoch. The evolution of σ_8 is primarily driven by the value of the matter density $\Omega_{\rm m}$, with a sub-dominant dependence on the dark energy density. It is generally recognized that such analyses favor a low matter density, though there is not complete consensus on this, and at present this technique for constraining the density is not competitive with the CMB.

21.3.5.2. *Cluster baryon fraction:* If clusters are representative of the mass distribution in the Universe, the fraction of the mass in baryons to the overall mass distribution would be $f_{\rm b} = \Omega_{\rm b}/\Omega_{\rm m}$. If $\Omega_{\rm b}$, the baryon density parameter, can be inferred from the primordial nucleosynthesis abundance of the light elements, the cluster baryon fraction $f_{\rm b}$ can then be used to constrain $\Omega_{\rm m}$ and h (*e.g.*, Ref. 30). The baryons in clusters are primarily in the form of X-ray-emitting gas that falls into the cluster, and secondarily in the form of stellar baryonic mass. Hence, the baryon fraction in clusters is estimated to be

$$f_{\rm b} = \frac{\Omega_{\rm b}}{\Omega_{\rm m}} \simeq f_{\rm gas} + f_{\rm gal}\,, \qquad (21.16)$$

where $f_{\rm b} = M_{\rm b}/M_{\rm grav}$, $f_{\rm gas} = M_{\rm gas}/M_{\rm grav}$, $f_{\rm gal} = M_{\rm gal}/M_{\rm grav}$, and $M_{\rm grav}$ is the total gravitating mass.

This can be used to obtain an approximate relation between $\Omega_{\rm m}$ and h:

$$\Omega_{\rm m} = \frac{\Omega_{\rm b}}{f_{\rm gas} + f_{\rm gal}} \simeq \frac{\Omega_{\rm b}}{0.08h^{-1.5} + 0.01h^{-1}}\,. \qquad (21.17)$$

Big Bang Nucleosynthesis gives $\Omega_{\rm b}h^2 \approx 0.02$, allowing the above relation to be approximated as $\Omega_{\rm m}h^{0.5} \approx 0.25$ (*e.g.*, Ref. 31). For example, Allen *et al.* [32] derived a density parameter consistent with $\Omega_{\rm m} = 0.3$ from Chandra observations.

21.3.6. ***Clustering in the inter-galactic medium*:**

It is commonly assumed, based on hydrodynamic simulations, that the neutral hydrogen in the inter-galactic medium (IGM) can be related to the underlying mass distribution. It is then possible to estimate the matter power spectrum on scales of a few megaparsecs from the absorption observed in quasar spectra, the so-called Lyman-alpha forest. The usual procedure is to measure the power spectrum of the transmitted flux, and then to infer the mass power spectrum. Photo-ionization heating by the ultraviolet background radiation and adiabatic cooling by the expansion of the Universe combine to give a simple power-law relation between the gas temperature and the baryon density. It also follows that there is a power-law relation between the optical depth τ and $\rho_{\rm b}$. Therefore, the observed flux $F = \exp(-\tau)$ is strongly correlated with $\rho_{\rm b}$, which itself traces the mass density. The matter and flux power-spectra can be related by

$$P_{\rm m}(k) = b^2(k)\; P_F(k)\,, \qquad (21.18)$$

where $b(k)$ is a bias function which is calibrated from simulations. Croft *et al.* [33] derived cosmological parameters from Keck Telescope observations of the Lyman-alpha forest at redshifts $z = 2 - 4$. Their derived power spectrum corresponds to that of a CDM model, which is in good agreement with the 2dF galaxy power spectrum. A recent study using VLT spectra [34] agrees with the flux power spectrum of Ref. 33.

This method depends on various assumptions. Seljak *et al.* [35] pointed out that errors are sensitive to the range of cosmological parameters explored in the simulations, and the treatment of the mean transmitted flux. Combination of the Lyman-alpha data with WMAP suggested deviation from the scale-invariant $n = 1$ power spectrum [7,6], but Seljak *et al.* [35] have argued that the combined data set is still compatible with $n = 1$ model.

21.3.7. ***Gravitational lensing*:**

Images of background galaxies get distorted due to the gravitational effect of mass fluctuations along the line of sight. Deep gravitational potential wells such as galaxy clusters generate 'strong lensing', *i.e.*, arcs and arclets, while more moderate fluctuations give rise to 'weak lensing'. Weak lensing is now widely used to measure the mass power spectrum in random regions of the sky (see Ref. 36 for recent reviews). As the signal is weak, the CCD frame of deformed galaxy shapes ('shear map') is analyzed statistically to measure the power spectrum, higher moments, and cosmological parameters.

The shear measurements are mainly sensitive to the combination of $\Omega_{\rm m}$ and the amplitude σ_8. There are various systematic effects in the interpretation of weak lensing, *e.g.*, due to atmospheric distortions during observations, the redshift distribution of the background galaxies, intrinsic correlation of galaxy shapes, and non-linear modeling uncertainties. Hoekstra *et al.* [37] derived the result $\sigma_8\Omega_{\rm m}^{0.52} = 0.46^{+0.05}_{-0.07}$ (95% confidence level), assuming a ΛCDM model. Other recent results are summarized in Ref. 36. For a $\Omega_{\rm m} = 0.3, \Omega_\Lambda = 0.7$ cosmology, different groups derived normalizations σ_8 over a wide range, indicating that the systematic errors are still larger than some of the quoted error bars.

21.3.8. ***Peculiar velocities*:**

Deviations from the Hubble flow directly probe the mass fluctuations in the Universe, and hence provide a powerful probe of the dark matter. Peculiar velocities are deduced from the difference between the redshift and the distance of a galaxy. The observational difficulty is in accurately measuring distances to galaxies. Even the best distance indicators (*e.g.*, the Tully-Fisher relation) give an error of 15% per galaxy, hence limiting the application of the method at large distances. Peculiar velocities are mainly sensitive to $\Omega_{\rm m}$, not to Ω_Λ or quintessence. Extensive analyses in the early 1990s (*e.g.*, Ref. 38) suggested a value of $\Omega_{\rm m}$ close to unity. A more recent analysis [39], which takes into account non-linear corrections, gives $\sigma_8\Omega_{\rm m}^{0.6} = 0.49 \pm 0.06$ and $\sigma_8\Omega_{\rm m}^{0.6} = 0.63 \pm 0.08$ (90% errors) for two independent data sets. While at present cosmological parameters derived from peculiar velocities are strongly affected by random and systematic errors, a new generation of surveys may improve their accuracy. Two promising approaches are the 6dF near-infrared survey of 15,000 peculiar velocities[‡‡] and the kinematic Sunyaev-Zel'dovich effect.

21.4. Bringing observations together

Although it contains two ingredients—dark matter and dark energy—which have not yet been verified by laboratory experiments, the ΛCDM model is almost universally accepted by cosmologists as the best description of present data. The basic ingredients are given by the parameters listed in Sec. 21.1.4, with approximate values of some of the key parameters being $\Omega_{\rm b} \approx 0.04$, $\Omega_{\rm dm} \approx 0.26$, $\Omega_\Lambda \approx 0.70$, and a Hubble constant $h \approx 0.7$. The spatial geometry is very close to flat (and often assumed to be precisely flat), and the initial perturbations Gaussian, adiabatic, and nearly scale-invariant.

The most powerful single experiment is WMAP, which on its own supports all these main tenets. Values for some parameters, as given in Spergel *et al.* [7], are reproduced in Table 21.2. This model presumes a flat Universe, and so Ω_Λ is a derived quantity in this analysis, with best-fit value $\Omega_\Lambda = 0.73$.

However, to obtain the most powerful constraints, other data sets need to be considered in addition to WMAP. A standard data compilation unites WMAP with shorter-scale CMB measurements from CBI and ACBAR, and the galaxy power spectrum from the 2dF survey. In our opinion, this combination of datasets offers the most reliable set of constraints at present. In addition, it is possible to add the Lyman-alpha forest power spectrum data, but this has proven more controversial as the interpretation of such data has not reached a secure level.

Using the extended data set without the Lyman-alpha constraints produces no surprises; as compared to WMAP alone, the best-fit values move around a little within the uncertainties, and the error bars

[‡‡] See `http://www.mso.anu.edu.au/6dFGS/`

Table 21.2: Parameter constraints reproduced from Spergel *et al.* [7], both from WMAP alone and from the preferred data compilation of WMAP+CBI+ACBAR (known as WMAPext) plus 2dFGRS. The first two columns assume a power-law initial spectrum, while the third allows a running of the spectral index (in this case n is defined at a particular scale, and its value cannot be directly compared with the power-law case). Spatial flatness is assumed in the parameter fit. The parameter A is a measure of the perturbation amplitude; see Ref. 7 for details. Uncertainties are shown at one sigma, and caution is needed in extrapolating them to higher significance levels due to non-Gaussian likelihoods and assumed priors.

	WMAP alone power-law	WMAPext + 2dFGRS power-law	WMAPext + 2dFGRS running
$\Omega_{\rm m}h^2$	0.14 ± 0.02	0.134 ± 0.006	0.136 ± 0.009
$\Omega_{\rm b}h^2$	0.024 ± 0.001	0.023 ± 0.001	0.022 ± 0.001
h	0.72 ± 0.05	0.73 ± 0.03	0.71 ± 0.04
n	0.99 ± 0.04	0.97 ± 0.03	$0.93^{+0.04}_{-0.05}$
τ	$0.17^{+0.08}_{-0.07}$	0.15 ± 0.07	0.17 ± 0.06
A	0.9 ± 0.1	0.8 ± 0.1	0.84 ± 0.09
$dn/d\ln k$	-	-	$-0.031^{+0.023}_{-0.025}$

improve somewhat, as seen in Table 21.2. In this table we also show the effect of allowing the spectral index to vary with scale ('running'): the running is found to be consistent with zero and there are small drifts in the values and uncertainties of the other parameters.¶

However, inclusion of the Lyman-alpha data suggests a more radical development, with the running weakly detected at around 95% confidence, the spectral index making a transition from $n > 1$ on large scales to $n < 1$ on small scales [7,6]. The significance of this measurement is not high, and the result rather unexpected on theoretical grounds (it suggests that the power spectrum has a maximum which just happens to lie in the rather narrow range of scales that observations are able to probe, and the running is much larger than in typical inflation models giving a spectral index close to one). In our view it is premature to read much significance into this observation, though if true, it should rapidly be firmed up by new data.

The baryon density Ω_b is now measured with quite high accuracy from the CMB and large-scale structure, and shows reasonable agreement with the determination from big bang nucleosynthesis; Fields and Sarkar in this volume quote the range $0.009 \geq \Omega_{\rm b}h^2 \geq 0.023$. Given the sensitivity of the measurement, it is important to note that it has significant dependence on both the datasets and parameter sets chosen, as seen in Table 21.2.

While Ω_Λ is measured to be non-zero with very high confidence, there is no evidence of evolution of the dark energy density. The WMAP team find the limit $w < -0.78$ at 95% confidence from a compilation of data including SNe Ia data, where they impose a prior $w \geq -1$, with the cosmological constant case $w = -1$ giving an excellent fit to the data.

As far as inflation is concerned, the data provide good news and bad news. The good news is that WMAP supports all the main predictions of the simplest inflation models: spatial flatness and adiabatic, Gaussian, nearly scale-invariant density perturbations. But it is disappointing that there is no sign of primordial gravitational waves, with WMAP providing only a weak upper limit $r < 0.53$ at 95% confidence [6] (this assumes no running, and weakens significantly if running is allowed), and especially that no convincing deviations from scale-invariance have been seen. It is perfectly possible for inflation models to give $n \simeq 1$ and $r \simeq 0$, but in that limit, the observations give no clues as to the dynamical processes driving inflation. Tests have been made for various types of non-Gaussianity, a particular example being a parameter $f_{\rm NL}$ which measures a quadratic contribution to the perturbations and is constrained to $-58 < f_{\rm NL} < 134$ at 95% confidence [41] (this looks weak, but prominent non-Gaussianity requires the product $f_{\rm NL}\Delta_{\mathcal{R}}$ to be large, and $\Delta_{\mathcal{R}}$ is of order 10^{-5}).

¶ As we were finalizing this article, an analysis of WMAP combined with the SDSS galaxy power spectrum appeared [40], giving results in good agreement with those discussed here.

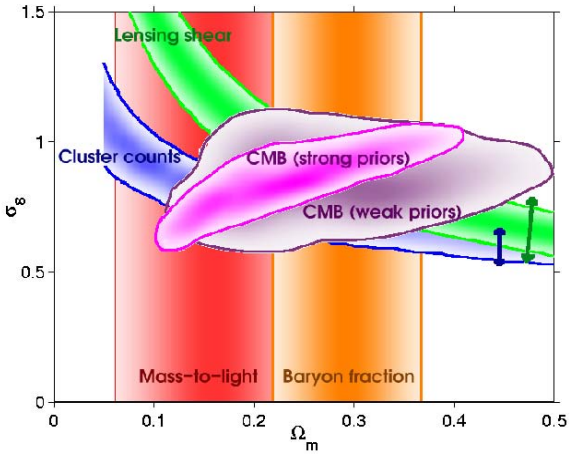

Figure 21.4: Various constraints shown in the $\Omega_{\rm m}$–σ_8 plane. [Figure provided by Sarah Bridle; see also Ref. 42.] See full-color version on color pages at end of book.

Two parameters which are still uncertain are $\Omega_{\rm m}$ and σ_8 (see Figure 21.4 and Ref. 42). The value of $\Omega_{\rm m}$ is beginning to be pinned down with some precision, with most observations indicating a value around 0.3, including the CMB anisotropies, the cluster number density, and gravitational lensing, though the latter two have a strong degeneracy with the amplitude of mass fluctuations σ_8. However, not all observations yet agree fully on this, for instance mass-to-light ratio measurements give $\Omega_{\rm m} \approx 0.15$ [43], and the fractional uncertainty remains significantly higher than one would like. Concerning σ_8, results from the cluster number density have varied quite a lot in recent years, spanning the range 0.6 to 1.0, primarily due to the uncertainties in the mass-temperature-luminosity relations used to connect the observables with theory. There is certainly scope for improving this calibration by comparison to mass measurements from strong gravitational lensing. The WMAP-alone measurements gives $\sigma_8 = 0.9 \pm 0.1$. However, this is not a direct constraint; WMAP only probes larger length scales, and the constraint comes from using WMAP to estimate all the parameters of the model needed to determine σ_8. As such, their constraint depends strongly on the assumed set of cosmological parameters being sufficient.

One parameter which is surprisingly robust is the age of the Universe. There is a useful coincidence that for a flat Universe the position of the first peak is strongly correlated with the age of the Universe. The WMAP-only result is 13.4 ± 0.3 Gyr (assuming a flat Universe). This is in good agreement with the ages of the oldest globular clusters [44] and radioactive dating [45].

21.5. Outlook for the future

The concordance model is now well established, and there seems little room left for any dramatic revision of this paradigm. A measure of the strength of that statement is how difficult it has proven to formulate convincing alternatives. For example, one corner of parameter space that has been explored is the possibility of abandoning the dark energy, and instead considering a mixed dark matter model with $\Omega_{\rm m} = 1$ and $\Omega_\nu = 0.2$. Such a model fits both the 2dF and WMAP data reasonably well, but only for a Hubble constant $h < 0.5$ [27,46]. However, this model is inconsistent with the HST key project value of h, the results from SNe Ia, cluster number density evolution, and baryon fraction in clusters.

Should there indeed be no major revision of the current paradigm, we can expect future developments to take one of two directions.

Either the existing parameter set will continue to prove sufficient to explain the data, with the parameters subject to ever-tightening constraints, or it will become necessary to deploy new parameters. The latter outcome would be very much the more interesting, offering a route towards understanding new physical processes relevant to the cosmological evolution. There are many possibilities on offer for striking discoveries, for example:

- The cosmological effects of a neutrino mass may be unambiguously detected, shedding light on fundamental neutrino properties;
- Detection of deviations from scale-invariance in the initial perturbations would indicate dynamical processes during perturbation generation, for instance, by inflation;
- Detection of primordial non-Gaussianities would indicate that non-linear processes influence the perturbation generation mechanism;
- Detection of variation in the dark energy density (*i.e.*, $w \neq -1$) would provide much-needed experimental input into the question of the properties of the dark energy.

These provide more than enough motivation for continued efforts to test the cosmological model and improve its precision.

Over the coming years, there are a wide range of new observations, which will bring further precision to cosmological studies. Indeed, there are far too many for us to be able to mention them all here, and so we will just highlight a few areas.

The cosmic microwave background observations will improve in several directions. The new frontier is the study of polarization, first detected in 2002. Data are imminent from balloon-based experiments including Maxipol and Boomerang, and with WMAP continuing to take data, they should be able to measure a polarization spectrum, as well as improve measures of the temperature-polarization cross-correlation (which is easier to measure as the temperature anisotropies are much larger). Dedicated ground-based polarization experiments, such as CBI and QUEST, promise powerful measures of the polarization spectrum in the next few years, and may be able to separately detect the two modes of polarization. Another area of development is pushing accurate power spectrum measurements to smaller angular scales, typically achieved by interferometry, which should allow measurements of secondary anisotropy effects, such as the Sunyaev-Zel'dovich effect, whose detection has already been tentatively claimed by CBI. Finally, we mention the Planck satellite, due to launch in 2007, which will make high-precision all-sky maps of temperature and polarization, utilizing a very wide frequency range for observations to improve understanding of foreground contaminants, and to compile a large sample of clusters via the Sunyaev-Zel'dovich effect.

Concerning galaxy clustering, the Sloan Digital Sky Survey is well underway, and currently expected to yield around 600,000 galaxy redshifts covering one quarter of the sky. Large samples of galaxy positions at high redshifts ($z \sim 1$) will begin to be obtained, for instance, by the DEEP2 survey using the Keck telescopes, and the VIRMOS survey on the VLT. The 6dF survey aims to take high-quality redshift and peculiar velocity data for a large sample of nearby galaxies, and has already taken around 40,000 of the planned 170,000 redshifts.

Still awaiting final approval is the SNAP satellite, which seeks to carry out a survey for Type Ia supernovae out to redshifts approaching two, which should in particular be a powerful probe of the dark energy. With large samples, it may be possible to detect evolution of the dark energy density, thus measuring its equation of state. SNAP is also able to carry out a large weak gravitational lensing survey, complementing those becoming possible with large-format CCDs on ground-based telescopes. Before SNAP, the ESSENCE project will significantly increase the size of the SNe Ia dataset.

The development of the first precision cosmological model is a major achievement. However, it is important not to lose sight of the motivation for developing such a model, which is to understand the underlying physical processes at work governing the Universe's evolution. On that side, progress has been much less dramatic. For instance, there are many proposals for the nature of the dark matter, but no consensus as to which is correct. The nature of the dark energy remains a mystery. Even the baryon density, now measured to an accuracy of a few percent, lacks an underlying theory able to predict it even within orders of magnitude. Precision cosmology may have arrived, but at present many key questions remain unanswered.

Acknowledgments:

ARL was supported in part by the Leverhulme Trust. We thank Sarah Bridle and Jochen Weller for useful comments on this article, and OL thanks members of the Cambridge Leverhulme Quantitative Cosmology and 2dFGRS Teams for helpful discussions.

References:

1. C.L. Bennett *et al.*, Astrophys. J. Supp. **148**, 1 (2003).
2. S. Fukuda *et al.*, Phys. Rev. Lett. **85**, 3999 (2000);
   Q. R. Ahmad *et al.*, Phys. Rev. Lett. **87**, 071301 (2001).
3. A.D. Dolgov *et al.*, Nucl. Phys. **B632**, 363 (2002).
4. For detailed accounts of inflation see E.W. Kolb and M.S. Turner, *The Early Universe*, Addison–Wesley (Redwood City, 1990);
   A.R. Liddle and D.H. Lyth, *Cosmological Inflation and Large-Scale Structure*, (Cambridge University Press, 2000).
5. U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 1 (1996).
6. H.V. Peiris *et al.*, Astrophys. J. Supp. **148**, 213 (2003).
7. D.N. Spergel *et al.*, Astrophys. J. Supp. **148**, 175 (2003).
8. J.C. Mather *et al.*, Astrophys. J. **512**, 511 (1999).
9. A. Kosowsky and M.S. Turner, Phys. Rev. **D52**, 1739 (1995).
10. D.H. Lyth and D. Wands, Phys. Lett. **B524**, 5 (2002);
    K. Enqvist and M.S. Sloth, Nucl. Phys. **B626**, 395 (2002);
    T. Moroi and T. Takahashi, Phys. Lett. **B522**, 215 (2001).
11. B. Ratra and P.J.E. Peebles, Phys. Rev. **D37**, 3406 (1988);
    C. Wetterich, Nucl. Phys. **B302**, 668 (1988);
    T. Padmanabhan, Phys. Rept. **380**, 235 (2003).
12. J.K. Webb *et al.*, Phys. Rev. Lett. **82**, 884 (1999);
    J.K. Webb *et al.*, Phys. Rev. Lett. **87**, 091301 (2001);
    J.K. Webb, M. Murphy, V. Flambaum, and S.J. Curran, Astrophys. Sp. Sci. **283**, 565 (2003).
13. W.L. Freedman *et al.*, Astrophys. J. **553**, 47 (2001).
14. A. Filippenko, `astro-ph/0307139`.
15. A.G. Riess *et al.*, Astron. J. **116**, 1009 (1998);
    P. Garnavich *et al.*, Astrophys. J. **509**, 74 (1998);
    S. Perlmutter *et al.*, Astrophys. J. **517**, 565 (1999).
16. J.L. Tonry *et al.*, Astrophys. J. **594**, 1 (2003).
17. R.A. Knop *et al.*, `astro-ph/0309368`.
18. I. Maor *et al.*, Phys. Rev. **D65**, 123003 (2002).
19. G. Hinshaw *et al.*, Astrophys. J. Supp. **148**, 135 (2003).
20. A. Kogut *et al.*, Astrophys. J. Supp. **148**, 161 (2003).
21. J. Kovac *et al.*, Nature **420**, 772 (2002).
22. W.J. Percival *et al.*, Mon. Not. Roy. Astr. Soc. **337**, 1068 (2002).
23. E. Hawkins *et al.*, in press, Mon. Not. Roy. Astr. Soc.
24. A.S. Szalay *et al.*, Astrophys. J. **591**, 1 (2003).
25. M. Tegmark *et al.*, `astro-ph/0310725`.
26. W. Hu, D. Eisenstein, and M. Tegmark, Phys. Rev. Lett. **80**, 5255 (1998).
27. O. Elgaroy and O. Lahav, JCAP **0304**, 004 (2003).
28. S. Hannestad, JCAP, **0305**, 004 (2003).
29. P.T.P. Viana *et al.*, Mon. Not. Roy. Astr. Soc., **346**, 319 (2003).
30. S.D.M. White *et al.*, Nature **366**, 429 (1993).
31. P. Erdogdu, S. Ettori, and O. Lahav, Mon. Not. Roy. Astr. Soc. **340**, 573 (2003).
32. S.W. Allen *et al.*, Mon. Not. Roy. Astr. Soc. **342**, 287 (2003).
33. R.A.C. Croft *et al.*, Astrophys. J. **581**, 20 (2002).
34. S. Kim *et al.*, `astro-ph/0308103`.

35. U. Seljak, P. McDonald, and A. Makarov, Mon. Not. Roy. Astr. Soc. **342**, L79 (2003).
36. P. Schneider, `astro-ph/0306465`;
A. Refregier, in press, Ann. Rev. Astron. Astrophys, `astro-ph/0307212`.
37. H. Hoekstra, H.K.C. Yee, and M. Gladders, Astrophys. J. **577**, 595 (2002).
38. A. Dekel, Ann. Rev. Astron. Astrophys. **32**, 371 (1994).
39. L. Silberman *et al.*, Astrophys. J. **557**, 102 (2001).
40. M. Tegmark *et al.*, `astro-ph/0310723`.
41. E. Komatsu *et al.*, Astrophys. J. Supp. **148**, 119 (2003).
42. S.L. Bridle *et al.*, Science **299**, 1532 (2003).
43. N. Bahcall *et al.*, Astrophys. J. **541**, 1 (2000).
44. B. Chaboyer and L.M. Krauss, Science **299**, 65 (2003).
45. R. Cayrel *et al.*, Nature **409**, 691 (2001).
46. A. Blanchard *et al.*, `astro-ph/0304237`.

22. DARK MATTER

Written September 2003 by M. Drees (Technical University, Munich) and G. Gerbier (Saclay, CEA).

22.1. Theory

22.1.1. *Evidence for Dark Matter*:

The existence of Dark (*i.e.*, non-luminous and non-absorbing) Matter (DM) is by now well established. The earliest [1], and perhaps still most convincing, evidence for DM came from the observation that various luminous objects (stars, gas clouds, globular clusters, or entire galaxies) move faster than one would expect if they only felt the gravitational attraction of other visible objects. An important example is the measurement of galactic rotation curves. The rotational velocity v of an object on a stable Keplerian orbit with radius r around a galaxy scales like $v(r) \propto \sqrt{M(r)/r}$, where $M(r)$ is the mass inside the orbit. If r lies outside the visible part of the galaxy and mass tracks light, one would expect $v(r) \propto 1/\sqrt{r}$. Instead, in most galaxies one finds that v becomes approximately constant out to the largest values of r where the rotation curve can be measured; in our own galaxy, $v \simeq 220$ km/s at the location of our solar system, with little change out to the largest observable radius. This implies the existence of a *dark halo*, with mass density $\rho(r) \propto 1/r^2$, *i.e.*, $M(r) \propto r$; at some point ρ will have to fall off faster (in order to keep the total mass of the galaxy finite), but we do not know at what radius this will happen. This leads to a lower bound on the DM mass density, $\Omega_{\rm DM} \gtrsim 0.1$, where $\Omega_X \equiv \rho_X/\rho_{\rm crit}$, $\rho_{\rm crit}$ being the critical mass density (*i.e.*, $\Omega_{\rm tot} = 1$ corresponds to a flat Universe).

The observation of clusters of galaxies tends to give somewhat larger values, $\Omega_{\rm DM} \simeq 0.2$ to 0.3. These observations include measurements of the peculiar velocities of galaxies in the cluster, which are a measure of their potential energy if the cluster is virialized; measurements of the *X-ray* temperature of hot gas in the cluster, which again correlates with the gravitational potential felt by the gas; and—most directly—studies of (weak) gravitational lensing of background galaxies on the cluster.

The currently most accurate, if somewhat indirect, determination of $\Omega_{\rm DM}$ comes from global fits of cosmological parameters to a variety of observations; see the Section on Cosmological Parameters for details. For example, using measurements of the anisotropy of the cosmic microwave background (CMB) and of the spatial distribution of galaxies, Ref. 2 finds a density of cold, non–baryonic matter

$$\Omega_{\rm nbm} h^2 = 0.111 \pm 0.006 \ , \tag{22.1}$$

where h is the Hubble constant in units of 100 km/(s·Mpc). Some part of the baryonic matter density [2],

$$\Omega_{\rm b} h^2 = 0.023 \pm 0.001 \ , \tag{22.2}$$

may well contribute to (baryonic) DM, *e.g.*, MACHOs [3] or cold molecular gas clouds [4].

The DM density in the "neighborhood" of our solar system is also of considerable interest. This was first estimated as early as 1922 by J.H. Jeans, who analyzed the motion of nearby stars transverse to the galactic plane [1]. He concluded that in our galactic neighborhood the average density of DM must be roughly equal to that of luminous matter (stars, gas, dust). Remarkably enough, the most recent estimates, based on a detailed model of our galaxy, find quite similar results [5]:

$$\rho_{\rm DM}^{\rm local} \simeq 0.3 \frac{\rm GeV}{\rm cm^3} \ ; \tag{22.3}$$

this value is known to within a factor of two or so.

22.1.2. *Candidates for Dark Matter*:

Analyses of structure formation in the Universe [6] indicate that most DM should be "cold", *i.e.*, should have been non-relativistic at the onset of galaxy formation (when there was a galactic mass inside the causal horizon). This agrees well with the upper bound [2] on the contribution of light neutrinos to Eq. (22.1),

$$\Omega_\nu h^2 \leq 0.0076 \quad 95\%\ {\rm CL} \tag{22.4}$$

Candidates for non-baryonic DM in Eq. (22.1) must satisfy several conditions: they must be stable on cosmological time scales (otherwise they would have decayed by now), they must interact very weakly with electromagnetic radiation (otherwise they wouldn't qualify as *dark* matter), and they must have the right relic density. Candidates include primordial black holes, axions, and weakly interacting massive particles (WIMPs).

Primordial black holes must have formed before the era of Big-Bang nucleosynthesis, since otherwise they would have been counted in Eq. (22.2) rather than Eq. (22.1). Such an early creation of a large number of black holes is possible only in certain somewhat contrived cosmological models [7].

The existence of axions [8] was first postulated to solve the strong *CP* problem of QCD; they also occur naturally in superstring theories. They are pseudo Nambu-Goldstone bosons associated with the (mostly) spontaneous breaking of a new global "Peccei-Quinn" (PQ) U(1) symmetry at scale f_a; see the Section on Axions in this *Review* for further details. Although very light, axions would constitute cold DM, since they were produced non-thermally. At temperatures well above the QCD phase transition, the axion is massless, and the axion field can take any value, parameterized by the "misalignment angle" θ_i. At $T \lesssim 1$ GeV the axion develops a mass m_a due to instanton effects. Unless the axion field happens to find itself at the minimum of its potential ($\theta_i = 0$), it will begin to oscillate once m_a becomes comparable to the Hubble parameter H. These coherent oscillations transform the energy originally stored in the axion field into physical axion quanta. The contribution of this mechanism to the present axion relic density is [8]

$$\Omega_a h^2 = \kappa_a \left(f_a/10^{12}\ {\rm GeV} \right)^{1.175} \theta_i^2 \ , \tag{22.5}$$

where the numerical factor κ_a lies roughly between 0.5 and a few. If $\theta_i \sim \mathcal{O}(1)$, Eq. (22.5) will saturate Eq. (22.1) for $f_a \sim 10^{11}$ GeV, comfortably above laboratory and astrophysical constraints [8]; this would correspond to an axion mass around 0.1 meV. However, if the post-inflationary reheat temperature $T_R > f_a$, cosmic strings will form during the PQ phase transition at $T \simeq f_a$. Their decay will give an additional contribution to Ω_a, which is often bigger than that in Eq. (22.5) [9], leading to a smaller preferred value of f_a, *i.e.*, larger m_a. On the other hand, values of f_a near the Planck scale become possible if θ_i is for some reason very small.

Weakly interacting massive particles (WIMPs) χ are particles with mass roughly between 10 GeV and a few TeV, and with cross sections of approximately weak strength. Their present relic density can be calculated reliably if the WIMPs were in thermal and chemical equilibrium with the hot "soup" of Standard Model (SM) particles after inflation. In this case their density would become exponentially (Boltzmann) suppressed at $T < m_\chi$. The WIMPs therefore drop out of thermal equilibrium ("freeze out") once the rate of reactions that change SM particles into WIMPs or vice versa, which is proportional to the product of the WIMP number density and the WIMP pair annihilation cross section into SM particles σ_A times velocity, becomes smaller than the Hubble expansion rate of the Universe. After freeze out, the co-moving WIMP density remains essentially constant. Their present relic density is then approximately given by (ignoring logarithmic corrections) [10]

$$\Omega_\chi h^2 \simeq const. \cdot \frac{T_0^3}{M_{\rm Pl}^3 \langle \sigma_A v \rangle} \simeq \frac{0.1\ {\rm pb} \cdot c}{\langle \sigma_A v \rangle} \ . \tag{22.6}$$

Here T_0 is the current CMB temperature, $M_{\rm Pl}$ is the Planck mass, c is the speed of light, σ_A is the total annihilation cross section of a pair of WIMPs into SM particles, v is the relative velocity between the two WIMPs in their cms system, and $\langle \ldots \rangle$ denotes thermal averaging. Freeze out happens at temperature $T_F \simeq m_\chi/20$ almost independently of the properties of the WIMP. This means that WIMPs are already non-relativistic when they decouple from the thermal plasma; it also implies that Eq. (22.6) is applicable if $T_R \gtrsim m_\chi/10$. Notice that the 0.1 pb in Eq. (22.6) contains factors of T_0 and $M_{\rm Pl}$; it is therefore quite intriguing that it "happens" to come out near the typical size of weak interaction cross sections.

The seemingly most obvious WIMP candidate is a heavy neutrino. However, an SU(2) doublet neutrino will have too small a relic density if its mass exceeds $M_Z/2$, as required by LEP data. One can suppress the annihilation cross section, and hence increase the relic density, by postulating mixing between a heavy SU(2) doublet and some "sterile" SU(2) $\times$ U(1)$_Y$ singlet neutrino. However, one also has to require the neutrino to be stable; it is not obvious why a massive neutrino should not be allowed to decay.

The currently best motivated WIMP candidate is therefore the lightest superparticle (LSP) in supersymmetric models [11] with exact R-parity (which guarantees the stability of the LSP). Searches for exotic isotopes [12] imply that a stable LSP has to be neutral. This leaves basically two candidates among the superpartners of ordinary particles, a sneutrino, and a neutralino. Sneutrinos again have quite large annihilation cross sections; their masses would have to exceed several hundred GeV for them to make good DM candidates. This is uncomfortably heavy for the lightest sparticle, in view of naturalness arguments. Moreover, the negative outcome of various WIMP searches (see below) rules out "ordinary" sneutrinos as primary component of the DM halo of our galaxy. (In models with gauge-mediated SUSY breaking the lightest "messenger sneutrino" could make a good WIMP [13].) The most widely studied WIMP is therefore the lightest neutralino. Detailed calculations [14] show that the lightest neutralino will have the desired thermal relic density Eq. (22.1) in at least four distinct regions of parameter space. χ could be (mostly) a bino or photino (the superpartner of the U(1)$_Y$ gauge boson and photon, respectively), if both χ and some sleptons have mass below $\sim$ 150 GeV, or if m_χ is close to the mass of some sfermion (so that its relic density is reduced through co-annihilation with this sfermion), or if $2m_\chi$ is close to the mass of the CP-odd Higgs boson present in supersymmetric models [15]. Finally, Eq. (22.1) can also be satisfied if χ has a large higgsino component.

Although WIMPs are attractive DM candidates because their thermal relic density naturally has at least the right order of magnitude, non-thermal production mechanisms have also been suggested, *e.g.*, LSP production from the decay of some moduli fields [16], from the decay of the inflaton [17], or from the decay of "Q–balls" (non-topological solitons) formed in the wake of Affleck-Dine baryogenesis [18]. Although LSPs from these sources are typically highly relativistic when produced, they quickly achieve kinetic (but not chemical) equilibrium if T_R exceeds a few MeV [19](but stays below $m_\chi/20$). They therefore also contribute to cold DM.

Primary black holes (as MACHOs), axions, and WIMPs are all (in principle) detectable with present or near-future technology (see below). There are also particle physics DM candidates which currently seem almost impossible to detect. These include the gravitino (the spin-3/2 superpartner of the graviton) [20], states from the "hidden sector" thought responsible for supersymmetry breaking [13], and the axino (the spin-1/2 superpartner of the axion) [21].

22.2. Experimental detection of Dark Matter

22.2.1. *The case of baryonic matter in our galaxy*:

The search for hidden galactic baryonic matter in the form of MAssive Compact Halo Objects (MACHOs) has been initiated following the suggestion that they may represent a large part of the galactic DM and could be detected through the microlensing effect [3]. The MACHO, EROS, and OGLE collaborations have performed a program of observation of such objects by monitoring the luminosity of millions of stars in the Large and Small Magellanic Clouds for several years. EROS concluded that MACHOs cannot contribute more than 20% to the mass of the galactic halo [22], while MACHO observed a signal at 0.4 solar mass and put an upper limit of 40%. Overall, this strengthens the need for non-baryonic DM, also supported by the arguments developed above.

22.2.2. *Axion searches*:

Axions can be detected by looking for $a \to \gamma$ conversion in a strong magnetic field [23]. Such a conversion proceeds through the loop-induced $a\gamma\gamma$ coupling, whose strength $g_{a\gamma\gamma}$ is an important parameter of axion models. Currently two experiments searching for axionic DM are taking data. They both employ high quality cavities. The cavity "Q factor" enhances the conversion rate on resonance, *i.e.*, for $m_a c^2 = \hbar\omega_{\rm res}$. One then needs to scan the resonance frequency in order to cover a significant range in m_a or, equivalently, f_a. The bigger of the two experiments, situated at the LLNL in California [24], started taking data in the first half of 1996. It uses very sophisticated "conventional" electronic amplifiers with very low noise temperature to enhance the conversion signal. Their first published results [25] exclude axions with mass between 2.9 and 3.3 μeV, corresponding to $f_a \simeq 4 \cdot 10^{13}$ GeV, as a major component of the dark halo of our galaxy, if $g_{a\gamma\gamma}$ is near the upper end of the theoretically expected range.

The smaller "CARRACK" experiment now under way in Kyoto, Japan [26] uses Rydberg atoms (atoms excited to a very high state, $n \simeq 230$) to detect the microwave photons that would result from axion conversion. This allows almost noise-free detection of single photons. Preliminary results of the CARRACK I experiment [27] exclude axions with mass in a narrow range around 10 μeV as major component of the galactic dark halo for some plausible range of $g_{a\gamma\gamma}$ values. This experiment is being upgraded to CARRACK II, which intends to probe the range between 2 and 50 μeV with sensitivity to all plausible axion models, if axions form most of DM [27].

22.2.3. *Basics of direct WIMP search*:

As stated above, WIMPs should be gravitationally trapped inside galaxies and should have the adequate density profile to account for the observed rotational curves. These two constraints determine the main features of experimental detection of WIMPs, which have been detailed in the reviews [28].

Their mean velocity inside our galaxy is expected to be similar to that of stars around the center of the galaxy, *i.e.*, a few hundred kilometers per second at the location of our solar system. For these velocities, WIMPs interact with ordinary matter through elastic scattering on nuclei. With expected WIMP masses in the range 10 GeV to 10 TeV, typical nuclear recoil energies are of order of 1 to 100 keV.

The shape of the nuclear recoil spectrum results from a convolution of the WIMP velocity distribution, usually taken as a shifted Maxwellian distribution, with the angular scattering distribution, which is isotropic to first approximation but forward-peaked for high nuclear mass (typically higher than Ge mass) due to the nuclear form factor. Overall, this results in a roughly exponential spectrum. The higher the WIMP mass, the higher the mean value of the exponential. This points to the need for low nuclear energy threshold detectors.

On the other hand, expected interaction rates depend on the product of the WIMP local flux and the interaction cross section. The first term is fixed by the local density of dark matter, taken as 0.3 GeV/cm^3 (see above), the mean WIMP velocity, typically 220 km/s, and the mass of the WIMP. The expected interaction rate then mainly depends on two unknowns, the mass and cross section of WIMP (with some uncertainty [5] due to the halo model). This is why the experimental observable, which is basically the scattering rate as a function of energy, is usually expressed as a contour in the WIMP mass—cross section plane.

The cross section depends on the nature of the couplings. For non-relativistic WIMPs one in general has to distinguish spin–independent and spin–dependent couplings. The former can involve scalar and vector WIMP and nucleon currents (vector currents are absent for Majorana WIMPs, e.g. the neutralino), while the latter involve axial vector currents (and obviously only exist if χ carries

spin). Due to coherence effects the spin–independent cross section scales approximately as the square of the mass of the nucleus, so higher mass nuclei, from Ge to Xe, are preferred for this search. For spin-dependent coupling, the cross section depends on the nuclear spin factor; the useful target nuclei are ^{19}F and ^{127}I.

Cross sections calculated in MSSM models induce rates of at most 1 evt day^{-1} kg^{-1} of detector, much lower than the usual radioactive backgrounds. This indicates the need for underground laboratories to protect against cosmic ray induced backgrounds, and for the selection of extremely radiopure materials.

The typical shape of exclusion contours can be anticipated from this discussion: at low WIMP mass, the sensitivity drops because of the detector energy threshold, whereas at high masses, the sensitivity also decreases because, for a fixed mass density, the WIMP flux decreases $\propto 1/m_\chi$. The sensitivity is best for WIMP masses near the mass of the recoiling nucleus.

22.2.4. *Status and prospects of direct WIMP searches*:

The first searches have been performed with ultra-pure semiconductors installed in pure lead and copper shields in underground environments [29]. Combining a priori excellent energy resolutions and very pure detector material, they produced the first limits on WIMP searches and until recently had the best performance (Heidelberg-Moscow, IGEX, COSME-II, HDMS) [29]. Without positive identification of nuclear recoil events, however, these experiments could only set limits, *e.g.*, excluding sneutrinos as major component of the galactic halo. Still, planned experiments using several tens of kgs to a ton of Germanium (many of which were designed for double-beta decay search)—GENIUS TF, GEDEON, MAJORANA—are based on only passive reduction of the external and internal electromagnetic, and neutron background by using segmented detectors, minimal detector housing, close electronics, and large liquid nitrogen shields.

To make further progress, active background rejection and signal identification questions have to be addressed. This has been the focus of many recent investigations and improvements. Active background rejection in detectors relies on the relatively small ionization in nuclear recoils due to their low velocity. This induces a reduction—quenching—of the ionization/scintillation signal for nuclear recoil signal events relative to e or γ induced backgrounds. Energies calibrated with gamma sources are then called "electron equivalent energies" (eee). This effects has been calculated and measured [29]. It is exploited in cryogenic detectors described later. In scintillation detectors, it induces in addition a difference in decay times of pulses induced by e/γ events vs nuclear recoils. Due to the limited resolution and discrimination power of this technique at low energies, this effect allows only a statistical background rejection. It has been used in NaI(Tl) (DAMA, NAIAD, Saclay NaI), in CsI(Tl)(KIMS), Xe (ZEPLIN) [29]. No observation of nuclear recoils has been reported by these experiments.

There are two experimental signatures of WIMP detection that would prove its astrophysical origin. One is the measurement of the strong daily forward/backward asymmetry of the nuclear recoil direction, due to the alternate sweeping of the WIMP cloud by the rotating Earth. Detection of this effect requires gaseous detectors or anisotropic response scintillators (stilbene). The second is the few percent annual modulation of the recoil rate due to the Earth speed adding to or subtracting from the speed of the Sun. This tiny effect can only be detected with large masses; nuclear recoil identification should also be performed, as the much larger background may also be subject to modulation.

The DAMA experiment operating 100 kg of NaI(Tl) in Gran Sasso has observed, with a statistical significance of 6.3 σ, an annually modulated signal with the expected phase, over a period of 7 years with a total exposure of around 100 000 kg·d , in the 2 to 6 keV (eee) energy interval [30]. This effect is attributed to a WIMP signal by the authors. If interpreted within the standard halo model described above, it would require a WIMP with $m_\chi \simeq 50$ GeV and $\sigma_{\chi p} \simeq 7 \cdot 10^{-6}$ pb (central values). This interpretation has however several unaddressed implications. In particular, the expected nuclear recoil rate from WIMPs should be of the order of 50% of the total measured rate in the 2–3 keV (eee) bin and 7% in the 4–6 keV (eee) bin. The rather large WIMP signal should be detectable by the pulse shape analysis. Moreover, the remaining, presumably e/γ induced, background would have to rise with energy; no explanation for this is given by the authors.

Annual modulation has also been searched for by NaI-32 (Zaragoza), DEMOS (Ge), ELEGANTS (NaI) [29]. No signal has been seen in these experiments, but their sensitivity is too low to contradict DAMA. The best current limit for a WIMP mass above 30 GeV has been produced in 2002 by Ge cryogenics detectors operated by EDELWEISS at 20 mK in the deep underground Fréjus lab [31]. They superseded the earlier CDMS results obtained at the much shallower Stanford site, with its large cosmic ray induced fast neutron flux [32]. The simultaneous measurement of the phonon signal and the ionization signal in such semiconductor detectors permits event by event discrimination between nuclear and electronic recoils down to 5 to 10 keV recoil energy. Assuming conventional WIMP halo parameters described above, EDELWEISS and CDMS exclude the DAMA signal at 99.8% CL. Varying the halo parameters, and/or including spin–dependent interactions compatible with the neutrino flux limit from the Sun, does not allow reconciliation of both results without finetuning [33]. The obtained sensitivity of $\sigma_{\chi p} \simeq 10^{-6}$ pb for the first time tests cross sections that can be accommodated in the MSSM [34].

Other cryogenic experiments like CRESST and ROSEBUD [35] use the scintillation of $CaWO_4$ as second variable for background discrimination, while CUORECINO will use TeO_2 in the purely thermal mode. In the coming years the cryogenic experimental programs of CDMS II, EDELWEISS II, CRESST II, CUORICINO, and ROSEBUD [35] intend to increase their sensitivity by a factor 100, by operating from few to 40 kg of detectors.

Liquid or two-phase Xenon detectors are rapidly coming on line. ZEPLIN has been operating 6 kg in the Boulby laboratory for about 1 year and announced a sensitivity close to that of EDELWEISS. With only 1.5 photoelectron/keV (eee), and a three-fold coincidence, searching for the WIMP signal in the 2–10 keV (eee) region is quite challenging. Neutron discrimination calibrations and trigger efficiency measurements will strengthen the reliability of this potentially powerful technique. With masses of 7 to 30 kg, ZEPLIN II and III aim at sensitivities down to 10^{-8} pb, while XMASS in Japan will soon operate 100 kg (ultimately 800 kg) at the SuperKamiokande site, using self-shielding and pulse shape analysis to lower the background [29]. XENON in US has approximately the same program.

On the other hand, the extended version of DAMA, LIBRA, starts operating 250 kg of NaI(Tl) in Gran Sasso, ANAIS will operate 107 kg of NaI(Tl) in Canfranc laboratory, KIMS, 80 kg of CsI(Tl)in Yang Yang lab in Korea, and ELEGANTS VI the large shield of 750 kg of NaI(Tl) [29].

There is also growing work in the development of low pressure Time Projection Chamber, the only convincing technique to measure the direction of nuclear recoils [36]. DRIFT, a 1m^3 volume detector is currently taking data. The small active mass of 200 g precludes competitive results. The sub-keV energy threshold gaseous detector MICROMEGAS is being investigated for WIMP search [36]. Other exotic techniques include the superheated droplet detectors SIMPLE, PICASSO, ORPHEUS, and ultra cold pure ^{3}He detector MacHe3 [29].

Sensitivities down to $\sigma_{\chi p}$ of 10^{-10} pb, as needed to probe large regions of MSSM parameter space [34], can be reached with detectors of typical masses of 1 ton [35], assuming nearly perfect background discrimination capabilities. Note that the expected WIMP rate is then 5 evts/ton/year for Ge. The ultimate neutron background will only be identified by its multiple interactions in a finely segmented or multiple interaction sensitive detector, and/or by operating detectors containing different target materials within the same set–up. Information on various neutron backgrounds calculations and measurements can be found in [37]. With intermediate mass of 10 to 30 kg and less efficient multiple interaction detection, a muon veto seems mandatory.

22.2.5. ***Status and prospects of indirect WIMP searches***:

WIMPs can annihilate and their annihilation products can be detected; these include neutrinos, gamma rays, positrons, antiprotons, and antinuclei [38]. These methods are complementary to direct detection and can explore higher masses and different coupling scenarios. "Smoking gun" signals for indirect detection are neutrinos coming from the center of the Sun or Earth, and monoenergetic photons from the halo.

WIMPs can be slowed down, captured, and trapped in celestial objects like the Earth or the Sun, thus enhancing their density and their probability of annihilation. This is a source of muon neutrinos which can interact in the Earth. Upward going muons can then be detected in large neutrino telescopes such as MACRO, BAKSAN, SuperKamiokande, Baikal, AMANDA, ANTARES, NESTOR, and the future large sensitive area IceCube [38]. The best upper limits, of $\simeq$ 3000 muons/km2/year, have been set by MACRO and BAKSAN [39]. However, at least in the framework of the MSSM, only the limits from the Sun are competitive with direct WIMP search limits. ANTARES (IceCube) will increase this sensitivity respectively by $\simeq$ one (two) order(s) of magnitude.

WIMP annihilation in the halo can give a continuous spectrum of gamma rays and (at one-loop level) also monoenergetic photon contributions from the $\gamma\gamma$ and γZ channels. The size of this signal depends very strongly on the halo model, but is expected to be most prominent towards the galactic center. Existing limits come from the EGRET satellite below 10 GeV, and from the WHIPPLE ground based telescope above 100 GeV [40]. However, only the planned space mission GLAST will be able to provide competitive SUSY sensitivities in both the continuous and γ line channels. Also, Atmospheric Cherenkov Telescopes like MAGIC, VERITAS, and HESS should be able to test some SUSY models, at large WIMP mass, for halo models showing a significant WIMP enhancement at the galactic center [38].

Diffuse continuum gammas could also give a signature due to their isotropic halo origin. The excess of GeV gamma-rays observed by EGRET [40] and attributed to a possible WIMP signal could however be due to classical sources.

The antiproton signal arises as another WIMP annihilation product in the halo. The signal is expected to be detectable above background only at very low energies. The BESS balloon-borne experiment indeed observed antiprotons below 1 GeV [41]. However, the uncertainties in the calculation of the expected signal and background energy spectra are too large to reach a firm conclusion. Precision measurements by the future experiments BESS, AMS2, and PAMELA may allow to disentangle signal and background [38].

A cosmic-ray positron flux excess at around 8 GeV measured by HEAT [42] has given rise to numerous calculations and conjectures concerning a possible SUSY interpretation. The need for an ad-hoc "boost" of expected flux to match the observed one and the failure to reproduce the energy shape by including a component from WIMP annihilation are illustrative of the difficulty to assign a Dark Matter origin to such measurements.

Last but not least, an antideuteron signal [43], as potentially observable by AMS2 or PAMELA, could constitute a signal for WIMP annihilation in the halo.

An interesting comparison of respective sensitivities to MSSM parameter space of future direct and various indirect searches has been performed with the DARKSUSY tool [44]. Searching for neutrinos from the Sun tests the spin-dependent WIMP couplings to matter, whereas direct searches are mostly sensitive to spin-independent couplings. Moreover, γ line searches are sensitive to higgsino-like neutralinos, whereas direct detection methods are more sensitive to gaugino-like neutralinos. However, it should be kept in mind that signals for WIMP annihilation in the halo strongly depend on various details of the halo model.

Numerous new experiments are in line to bring accurate measurements to constrain or discover Dark Matter.

References:

1. For a brief but delightful history of DM, see V. Trimble, in Proceedings of the *First International Symposium on Sources of Dark Matter in the Universe*, Bel Air, California, 1994; published by World Scientific, Singapore (ed. D.B. Cline).
2. See *Global cosmological parameters* in this *Review*.
3. B. Paczynski, Astrophys. J. **304**, 1 (1986);
   K. Griest, Astrophys. J. **366**, 412 (1991).
4. F. De Paolis *et al.*, Phys. Rev. Lett. **74**, 14 (1995).
5. M. Kamionkowski and A. Kinkhabwala, Phys. Rev. **D57**, 3256 (1998).
6. See *e.g.*, J.R. Primack, in the Proceedings of *Midrasha Mathematicae in Jerusalem: Winter School in Dynamical Systems*, Jerusalem, Israel, January 1997, `astro-ph/9707285` . There is currently some debate whether cold DM models correctly reproduce the DM density profile near the center of galactic haloes. See *e.g.*, R.A. Swaters *et al.*, Astrophys. J. **583**, 732 (2003).
7. B.J. Carr and S.W. Hawking, MNRAS **168**, 399 (1974).
8. See *Axions and other Very Light Bosons* in this *Review*.
9. R.A. Battye and E.P.S. Shellard, Phys. Rev. Lett. **73**, 2954 (1994);
   Erratum-ibid. **76**, 2203 (1996).
10. E.W. Kolb and M.E. Turner, *The Early Universe*, Addison-Wesley (1990).
11. For a review, see G. Jungman, M. Kamionkowski, and K. Griest, Phys. Reports **267**, 195 (1996).
12. See *Searches for WIMPs and other Particles* in this *Review*.
13. S. Dimopoulos, G.F. Giudice, and A. Pomarol, Phys. Lett. **B389**, 37 (1996).
14. See *e.g.*, J.R. Ellis *et al.*, Nucl. Phys. **B652**, 259 (2003);
    J.R. Ellis *et al.*, Phys. Lett. **B565**, 176 (2003);
    H. Baer *et al.*, JHEP **0306**, 054 (2003);
    A. Bottino *et al.*, `hep-ph/0304080` .
15. G. Griest and D. Seckel, Phys. Rev. **D43**, 3191 (1991).
16. T. Moroi and L. Randall, Nucl. Phys. **B570**, 455 (2000).
17. R. Allahverdi and M. Drees, Phys. Rev. Lett. **89**, 091302 (2002).
18. M. Fujii and T. Yanagida, Phys. Lett. **B542**, 80 (2002).
19. J. Hisano, K. Kohri, and M.M. Nojiri, Phys. Lett. **B505**, 169 (2001);
    X. Chen, M. Kamionkowski, and X. Zhang, Phys. Rev. **D64**, 021302 (2001).
20. M. Bolz, W. Buchmüller, and M. Plümacher, Phys. Lett. **B443**, 209 (1998).
21. L. Covi *et al.*, JHEP **0105**, 033 (2001).
22. MACHO Collab., C. Alcock *et al.*, Astrophys. J. **542**, 257 (2000);
    EROS Collab., AA 355, 39 (2000);
    OGLE Collab., AA 343, 10 (1999).
23. P. Sikivie, Phys. Rev. Lett. **51**, 1415 (1983), Erratum-ibid. **52**, 695 (1984).
24. H. Peng *et al.*, Nucl. Instrum. Methods **A444**, 569 (2000).
25. S. Asztalos *et al.*, Phys. Rev. **D64**, 092003 (2001).
26. M. Tada *et al.*, `physics/0101028` .
27. K. Yamamoto *et al.*, in *Heidelberg 2000, Dark matter in astro—and particle—physics*, `hep-ph/0101200` .
28. P.F. Smith and J.D. Lewin, Phys. Reports **187**, 203 (1990);
    J.R. Primack, D. Seckel, and B. Sadoulet, Ann. Rev. Nucl. Part. Sci. **38** 751 (1988).

29. For recent reviews on non cryogenic detectors, see *e.g.*, A. Morales in *Proceedings of IFMP2002*, Jaca (Spain) astro-ph/0211446;
Proceedings of Topics in Astroparticles and Underground Physics TAUP 2001, Nucl. Phys B (Proc. Suppl.) vol. 11 (2002);
TAUP 2003 Conference site:
http://int.phys.washington.edu/taup2003;
Proceedings of *Identification of Dark Matter*. IDM 2000, World Scientific, ed. N. Spooner and V. Kudryavtsev, (York, UK, 2000);
IDM 2002 Conference site:
http://www.shef.ac.uk/ phys/idm2002.html.
30. DAMA Collab., R. Bernabei *et al.*, Phys. Lett. **B480**, 23 (2000), and Riv. Nuovo Cimento **26**, 1 (2003).
31. EDELWEISS Collab., A. Benoit *et al.*, Phys. Lett. **B545**, 43 (2002).
32. CDMS Collab., D. Abrams *et al.*, Phys. Rev. **D66**, 122003 (2002);
CDMS Collab., D.S. Akerib *et al.*, Phys. Rev. **D68**, 082002 (2003).
33. C.J. Copi and L.M. Krauss, Phys. Rev. **67**, 103507 (2003);
A. Kurylov and M. Kamionkowski, hep-ph/0307185.
34. For a general introduction to SUSY, see the section devoted in this *Review of Particle Physics*. For a recent review on cross sections for direct detection, see J. Ellis *et al.*, Phys. Rev. **D67**, 123502 (2003).
35. For a recent review on cryogenic detectors, see *e.g.*, L. Mosca in Proceedings of IFMP2002. In addition to the TAUP and IDM Conference Proceedings, see also the Proceedings of *Int. Workshop on Low Temperature Detectors*, LTD9, AIP Conference Proceedings, Volume 605(2001).
36. Workshop on large TPC for low energy rare event detection, Paris, December 2002, http://www.unine.ch/phys/tpc.html.
37. These sites gather informations on neutrons from various underground labs:
http://www.physics.ucla.edu/wimps/nBG/nBG.html;
www.e15.physik.tu-muenchen.de/Tecnomusiq/Tecnomusiq.html.
38. L. Bergstrom, Rept. on Prog. in Phys. **63**, 793 (2000);
L. Bergstrom *et al.*, Phys. Rev. **D59**, 043506 (1999);
C. Tao, Phys. Scripta **T93**, 82 (2001).
39. MACRO Collab., M. Ambrosio *et al.*, Phys. Rev. **D60**, 082002 (1999);
BAKSAN Collab., M. M. Boliev *et al.*, Nucl. Phys. (Proc. Supp.) **B48**, 83 (1996).
40. EGRET Collab., D. Dixon *et al.*, New Astron. **3**, 539 (1998).
41. BESS Collab., S. Orito *et al.*, Phys. Rev. Lett. **84**, 1078 (2000).
42. HEAT Collab., S. W. Barwick *et al.*, Astrophys. J. **482**, L191 (2000).
43. F. Donato, N. Fornengo and P. Salati, Phys. Rev. **D62**, 043003 (2000).
44. DARKSUSY site: http://www.physto.se/ edsjo/darksusy/.

23. COSMIC MICROWAVE BACKGROUND

Revised September 2003 by D. Scott (University of British Columbia) and G.F. Smoot (UCB/LBNL).

23.1. Introduction

The energy content in radiation from beyond our Galaxy is dominated by the Cosmic Microwave Background (CMB), discovered in 1965 [1]. The spectrum of the CMB is well described by a blackbody function with $T = 2.725\,\mathrm{K}$. This spectral form is one of the main pillars of the hot Big Bang model for the early Universe. The lack of any observed deviations from a blackbody spectrum constrains physical processes over the history of the universe at redshifts $z \lesssim 10^7$ (see previous versions of this mini-review). However, at the moment, all viable cosmological models predict a very nearly Planckian spectrum, and so are not stringently limited.

Another observable quantity inherent in the CMB is the variation in temperature (or intensity) from one part of the microwave sky to another [2]. Since the first detection of these anisotropies by the *COBE* satellite [3], there has been intense activity to map the sky at increasing levels of sensitivity and angular resolution. A series of ground- and balloon-based measurements has recently been joined by the first results from NASA's Wilkinson Microwave Anisotropy Probe (*WMAP*) [4]. These observations have led to a stunning confirmation of the 'Standard Model of Cosmology.' In combination with other astrophysical data, the CMB anisotropy measurements place quite precise constraints on a number of cosmological parameters, and have launched us into an era of precision cosmology.

23.2. Description of CMB Anisotropies

Observations show that the CMB contains anisotropies at the 10^{-5} level, over a wide range of angular scales. These anisotropies are usually expressed by using a spherical harmonic expansion of the CMB sky:

$$T(\theta,\phi) = \sum_{\ell m} a_{\ell m}\, Y_{\ell m}(\theta,\phi).$$

The vast majority of the cosmological information is contained in the temperature 2 point function, *i.e.*, the variance as a function of separation θ. Equivalently, the power per unit $\ln\ell$ is $\ell\sum_m |a_{\ell m}|^2/4\pi$.

23.2.1. *The Monopole*:

The CMB has a mean temperature of $T_\gamma = 2.725 \pm 0.001\,\mathrm{K}\ (1\sigma)$ [5], which can be considered as the monopole component of CMB maps, a_{00}. Since all mapping experiments involve difference measurements, they are insensitive to this average level. Monopole measurements can only be made with absolute temperature devices, such as the FIRAS instrument on the *COBE* satellite [5]. Such measurements of the spectrum are consistent with a blackbody distribution over more than three decades in frequency. A blackbody of the measured temperature corresponds to $n_\gamma = (2\zeta(3)/\pi^2)\,T_\gamma^3 \simeq 411\,\mathrm{cm}^{-3}$ and $\rho_\gamma = (\pi^2/15)\,T_\gamma^4 \simeq 4.64\times10^{-34}\,\mathrm{g\,cm}^{-3} \simeq 0.260\,\mathrm{eV\,cm}^{-3}$.

23.2.2. *The Dipole*:

The largest anisotropy is in the $\ell = 1$ (dipole) first spherical harmonic, with amplitude $3.346 \pm 0.017\,\mathrm{mK}$ [4]. The dipole is interpreted to be the result of the Doppler shift caused by the solar system motion relative to the nearly isotropic blackbody field, as confirmed by measurements of the velocity field of local galaxies [6]. The motion of an observer with velocity $\beta = v/c$ relative to an isotropic Planckian radiation field of temperature T_0 produces a Doppler-shifted temperature pattern

$$\begin{aligned} T(\theta) &= T_0(1-\beta^2)^{1/2}/(1-\beta\cos\theta) \\ &\approx T_0\,\left(1+\beta\cos\theta+(\beta^2/2)\cos2\theta+O(\beta^3)\right). \end{aligned}$$

At every point in the sky one observes a blackbody spectrum, with temperature $T(\theta)$. The spectrum of the dipole is the differential of a blackbody spectrum, as confirmed by Ref. 7.

The implied velocity [8,4] for the solar system barycenter is $v = 368 \pm 2\,\mathrm{km\,s}^{-1}$, assuming a value $T_0 = T_\gamma$, towards $(\ell, b) = (263.85^\circ \pm 0.10^\circ, 48.25^\circ \pm 0.04^\circ)$. Such a solar system velocity implies a velocity for the Galaxy and the Local Group of galaxies relative to the CMB. The derived value is $v_{\rm LG} = 627 \pm 22\,\mathrm{km\,s}^{-1}$ toward $(\ell, b) = (276^\circ \pm 3^\circ, 30^\circ \pm 3^\circ)$, where most of the error comes from uncertainty in the velocity of the solar system relative to the Local Group.

The dipole is a frame dependent quantity, and one can thus determine the 'absolute rest frame' of the Universe as that in which the CMB dipole would be zero. Our velocity relative to the Local Group, as well as the velocity of the Earth around the Sun, and any velocity of the receiver relative to the Earth, is normally removed for the purposes of CMB anisotropy study.

23.2.3. *Higher-Order Multipoles*:

Excess variance in CMB maps at higher multipoles ($\ell \geq 2$) is interpreted as being the result of perturbations in the energy density of the early Universe, manifesting themselves at the epoch of the last scattering of the CMB photons. In the hot Big Bang picture, this happens at a redshift $z \simeq 1100$, with little dependence on the details of the model. The process by which the hydrogen and helium nuclei can hold onto their electrons is usually referred to as recombination [9]. Before this epoch, the CMB photons are tightly coupled to the baryons, while afterwards they can freely stream towards us.

Theoretical models generally predict that the $a_{\ell m}$ modes are Gaussian random fields, and all tests are consistent with this simplifying assumption [10]. With this assumption, and if there is no preferred axis, then it is the variance of the temperature field which carries the cosmological information, rather than the values of the individual $a_{\ell m}$s; in other words the power spectrum in ℓ fully characterizes the anisotropies. The power at each ℓ is $(2\ell+1)C_\ell/(4\pi)$, where $C_\ell \equiv \langle|a_{\ell m}|^2\rangle$, and a statistically isotropic sky means that all ms are equivalent. We use our estimators of the C_ℓs to constrain their expectation values, which are the quantities predicted by a theoretical model. For an idealized full-sky observation, the variance of each measured C_ℓ (the variance of the variance) is $[2/(2\ell+1)]C_\ell^2$. This sampling uncertainty (known as cosmic variance) comes about because each C_ℓ is χ^2 distributed with $(2\ell+1)$ degrees of freedom for our observable volume of the Universe. For partial sky coverage, $f_{\rm sky}$, this variance is increased by $1/f_{\rm sky}$ and the modes become partially correlated.

It is important to understand that theories predict the expectation value of the power spectrum, whereas our sky is a single realization. Hence the 'cosmic variance' is an unavoidable source of uncertainty when constraining models; it dominates the scatter at lower ℓs, while the effects of instrumental noise and resolution dominate at higher ℓs.

23.2.4. *Angular Resolution and Binning*:

There is no one-to-one conversion between the angle subtended by a particular wavevector projected on the sky and multipole ℓ. However, a single spherical harmonic $Y_{\ell m}$ corresponds to angular variations of $\theta \sim \pi/\ell$. CMB maps contain anisotropy information from the size of the map (or in practice some fraction of that size) down to the beam-size of the instrument, σ. One can think of the effect of a Gaussian beam as rolling off the power spectrum with the function $\mathrm{e}^{-\ell(\ell+1)\sigma^2}$.

For less than full sky coverage, the ℓ modes are correlated. Hence, experimental results are usually quoted as a series of 'band powers', defined as estimators of $\ell(\ell+1)C_\ell/2\pi$ over different ranges of ℓ. Because of the strong foreground signals in the Galactic Plane, even 'all-sky' surveys, such as *COBE* and *WMAP* involve a cut sky. The amount of binning required to obtain uncorrelated estimates of power also depends on the map size.

23.3. Cosmological Parameters

The current 'Standard Model' of cosmology contains around 10 free parameters (see the review on The Cosmological Parameters—Sec. 21 of this *Review*). The basic framework is the Friedmann-Robertson-Walker metric (*i.e.*, a universe that is approximately homogeneous and isotropic on large scales), with density perturbations laid down at early times and evolving into today's structures (see the review on Big-Bang Cosmology—Sec. 19 of this *Review*). These perturbations can be either 'adiabatic' (meaning that there is no change to the

entropy per particle for each species, *i.e.*, $\delta\rho/\rho$ for matter is $(3/4)\delta\rho/\rho$ for radiation) or 'isocurvature' (meaning that, for example, matter perturbations compensate radiation perturbations so that the total energy density remains unperturbed, *i.e.*, $\delta\rho$ for matter is $-\delta\rho$ for radiation). These different modes give rise to distinct phases during growth, and the adiabatic scenario is strongly preferred by the data. Models that generate mainly isocurvature type perturbations (such as most topological defect scenarios) are no longer considered to be viable.

Within the adiabatic family of models, there, is in principle, a free function describing how the comoving curvature perturbations, $\mathcal{R}$, vary with scale. In inflationary models, the Taylor series expansion of $\ln\mathcal{R}(\ln k)$ has terms of steadily decreasing size. For the simplest models, there are thus 2 parameters describing the initial conditions for density perturbations: the amplitude and slope of the power spectrum, $\left\langle|\mathcal{R}|^2\right\rangle \propto k^n$. This can be explicitly defined, for example, through:

$$\Delta^2_{\mathcal{R}} \equiv (k^3/2\pi^2)\left\langle|\mathcal{R}|^2\right\rangle,$$

and using $A^2 \equiv \Delta^2_{\mathcal{R}}(k_0)$ with $k_0 = 0.05\,\mathrm{Mpc}^{-1}$. There are many other equally valid definitions of the amplitude parameter (see also Sec. Olive & Peacock and Sec. Lahav & Liddle), and we caution that the relationships between some of them can be cosmology dependent. In 'slow roll' inflationary models this normalization is proportional to the combination $V^3/(V')^2$, for the inflationary potential $V(\phi)$. The slope n also involves V'', and so the combination of A and n can, in principle, constrain potentials.

Inflationary models can generate tensor (gravity wave) modes as well as scalar (density perturbation) modes. This fact introduces another parameter measuring the amplitude of a possible tensor component, or equivalently the ratio of the tensor to scalar contributions. The tensor amplitude $A_{\mathrm{T}} \propto V$, and thus one expects a larger gravity wave contribution in models where inflation happens at higher energies. The tensor power spectrum also has a slope, often denoted n_{T}, but since this seems likely to be extremely hard to measure, it is sufficient for now to focus only on the amplitude of the gravity wave component. It is most common to define the tensor contribution through r, the ratio of tensor to scalar perturbation spectra at large scales (say $k = 0.002\,\mathrm{Mpc}^{-1}$). There are other definitions in terms of the ratio of contributions to C_2, for example. Different inflationary potentials will lead to different predictions, e.g. for $\lambda\phi^4$ inflation, $r = 0.32$, while other models can have arbitrarily small values of r. In any case, whatever the specific definition, and whether they come from inflation or something else, the 'initial conditions' give rise to a minimum of 3 parameters: A, n, and r.

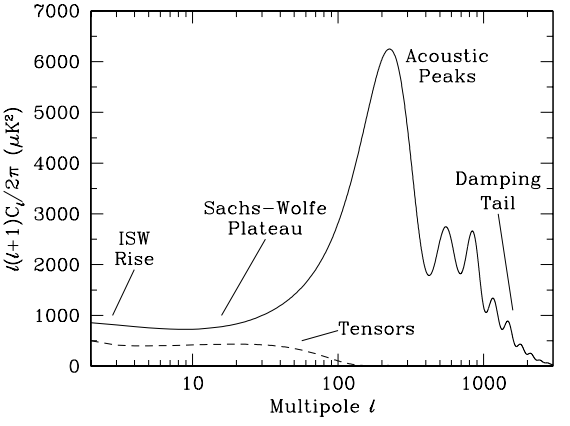

Figure 23.1: Plot of the theoretical CMB anisotropy power spectrum, using a standard ΛCDM model from `CMBFAST`. The x-axis is logarithmic here. The regions are labeled as in the text: the ISW Rise; Sachs-Wolfe Plateau; Acoustic Peaks; and Damping Tail. Also shown is the shape of the tensor (gravity wave) contribution, with an arbitrary normalization.

The background cosmology requires an expansion parameter (the Hubble Constant, H_0, often represented through $H_0 = 100\,h\,\mathrm{km\,s^{-1}Mpc^{-1}}$) and several parameters to describe the matter and energy content of the Universe. These are usually given in terms of the critical density, *i.e.*, for species 'x', $\Omega_{\mathrm{x}} = \rho_{\mathrm{x}}/\rho_{\mathrm{crit}}$, where $\rho_{\mathrm{crit}} = 3H_0^2/8\pi G$. Since physical densities $\rho_{\mathrm{x}} \propto \Omega_{\mathrm{x}}h^2 \equiv \omega_{\mathrm{x}}$ are what govern the physics of the CMB anisotropies, it is these ωs that are best constrained by CMB data. In particular CMB observations constrain $\Omega_{\mathrm{B}}h^2$ for baryons and $\Omega_{\mathrm{M}}h^2$ for baryons plus Cold Dark Matter.

The contribution of a cosmological constant Λ (or other form of Dark Energy) is usually included through a parameter which quantifies the curvature, $\Omega_{\mathrm{K}} \equiv 1 - \Omega_{\mathrm{tot}}$, where $\Omega_{\mathrm{tot}} = \Omega_{\mathrm{M}} + \Omega_\Lambda$. The radiation content, while in principle a free parameter, is precisely enough determined through the measurement of T_γ.

The main effect of astrophysical processes on the C_ℓs comes through reionization. The Universe became reionized at some redshift long after recombination, affecting the CMB through the integrated Thomson scattering optical depth:

$$\tau = \int_0^{z_{\mathrm{i}}} \sigma_{\mathrm{T}} n_{\mathrm{e}}(z) \frac{dt}{dz} dz,$$

where σ_{T} is the Thomson cross-section, $n_{\mathrm{e}}(z)$ is the number density of free electrons (which depends on astrophysics) and dt/dz is fixed by the background cosmology. In principle, τ can be determined from the small scale power spectrum together with the physics of structure formation and feedback processes. However, this is a sufficiently complicated calculation that τ needs to be considered as a free parameter.

Thus we have 8 basic cosmological parameters: A, n, r, h, $\Omega_{\mathrm{B}}h^2$, $\Omega_{\mathrm{M}}h^2$, Ω_{tot}, and τ. One can add additional parameters to this list, particularly when using the CMB in combination with other data sets. The next most relevant ones might be: $\Omega_\nu h^2$, the massive neutrino contribution; w ($\equiv p/\rho$), the equation of state parameter for the Dark Energy; and $dn/d\ln k$, measuring deviations from a constant spectral index. To these 11 one could of course add further parameters describing additional physics, such as details of the reionization process, features in the initial power spectrum, a sub-dominant contribution of isocurvature modes, *etc.*

As well as these underlying parameters, there are other quantities that can be derived from them. Such quantities include the actual Ωs of the various components (*e.g.*, Ω_{M}), the variance of density perturbations at particular scales (*e.g.*, σ_8), the age of the Universe today (t_0), the age of the Universe at recombination, reionization, *etc.*

23.4. Physics of Anisotropies

The cosmological parameters affect the anisotropies through the well understood physics of the evolution of linear perturbations within a background FRW cosmology. There are very effective, fast, and publicly-available software codes for computing the CMB anisotropy, polarization, and matter power spectra, *e.g.*, `CMBFAST` [12] and `CAMB` [13]. `CMBFAST` is the most extensively used code; it has been tested over a wide range of cosmological parameters and is considered to be accurate to better than the 1% level [14].

A description of the physics underlying the C_ℓs can be separated into 3 main regions, as shown in Fig. 23.1.

23.4.1. *The Sachs-Wolfe plateau:* $\ell \lesssim 100$:

The horizon scale (or more precisely, the angle subtended by the Hubble radius) at last scattering corresponds to $\ell \simeq 100$. Anisotropies at larger scales have not evolved significantly, and hence directly reflect the 'initial conditions.' The combination of gravitational redshift and intrinsic temperature fluctuations leads to $\delta T/T \simeq (1/3)\delta\phi/c^2$, where $\delta\phi$ is the perturbation to the gravitational potential. This is usually referred to as the 'Sachs-Wolfe' effect [15].

Assuming that a nearly scale-invariant spectrum of density perturbations was laid down at early times (*i.e.*, $n \simeq 1$, meaning equal power per decade in k), then $\ell(\ell+1)C_\ell \simeq$ constant at low ℓs. This effect is hard to see unless the multipole axis is plotted logarithmically (as in Fig. 23.1, but not Fig. 23.2).

Time variation in the potentials (*i.e.*, time-dependent metric perturbations) leads to an upturn in the C_ℓs in the lowest several

multipoles; any deviation from a total equation of state $w = 0$ has such an effect. So the dominance of the Dark Energy at low redshift makes the lowest ℓs rise above the plateau. This is sometimes called the 'integrated Sachs-Wolfe effect' (or ISW Rise), since it comes from the line integral of $\dot{\phi}$. It has been confirmed through correlations between the large-angle anisotropies and large-scale structure [16]. Specific models can also give additional contributions at low ℓ (*e.g.*, perturbations in the Dark Energy component itself [11]) but typically these are buried in the cosmic variance.

In principle, the mechanism that produces primordial perturbations would generate scalar, vector, and tensor modes. However, the vector (vorticity) modes decay with the expansion of the Universe. Tensors also decay when they enter the horizon, and so they contribute only to angular scales above about 1° (see Fig. 23.1). Hence some fraction of the low ℓ signal could be due to a gravity wave contribution, although small amounts of tensors are essentially impossible to discriminate from other effects that might raise the level of the plateau. However the tensors *can* be distinguished using polarization information (see Sec. 23.6).

23.4.2. *The acoustic peaks:* $100 \lesssim \ell \lesssim 1000$:

On sub-degree scales, the rich structure in the anisotropy spectrum is the consequence of gravity-driven acoustic oscillations occurring before the atoms in the universe became neutral. Perturbations inside the horizon at last scattering have been able to evolve causally and produce anisotropy at the last scattering epoch which reflects that evolution. The frozen-in phases of these sound waves imprint a dependence on the cosmological parameters, which gives CMB anisotropies their great constraining power.

The underlying physics can be understood as follows. When the proton-electron plasma was tightly coupled to the photons, these components behaved as a single 'photon-baryon fluid', with the photons providing most of the pressure and the baryons the inertia. Perturbations in the gravitational potential, dominated by the dark matter component, are steadily evolving. They drive oscillations in the photon-baryon fluid, with photon pressure providing the restoring force. The perturbations are quite small, $O(10^{-5})$, and so evolve linearly. That means each Fourier mode evolves independently and is described by a driven harmonic oscillator, with frequency determined by the sound speed in the fluid. Thus, there is an oscillation of the fluid density, with velocity $\pi/2$ out of phase and having amplitude reduced by the sound speed.

After the Universe recombined the baryons and radiation decoupled, and the radiation could travel freely towards us. At that point the phases of the oscillations were frozen-in, and projected on the sky as a harmonic series of peaks. The main peak is the mode that went through 1/4 of a period, reaching maximal compression. The even peaks are maximal *under*-densities, which are generally of smaller amplitude because the rebound has to fight against the baryon inertia. The troughs, which do not extend to zero power, are partially filled by the Doppler effect because they are at the velocity maxima.

An additional effect comes from geometrical projection. The scale associated with the peaks is the sound horizon at last scattering, which can be confidently calculated as a physical length scale. This scale is projected onto the sky, leading to an angular scale that depends on the background cosmology. Hence the angular position of the peaks is a sensitive probe of the spatial curvature of the Universe (*i.e.*, $\Omega_{\rm tot}$), with the peaks lying at higher ℓ in open universes and lower ℓ in closed geometry.

One last effect arises from reionization at redshift $z_{\rm i}$. A fraction of photons will be isotropically scattered at $z < z_{\rm i}$, partially erasing the anisotropies at angular scales smaller than those subtended by the Hubble radius at $z_{\rm i}$. This corresponds typically to ℓs above about a few 10s, depending on the specific reionization model. The acoustic peaks are therefore reduced by a factor $e^{-2\tau}$ relative to the plateau.

These acoustic peaks were a clear theoretical prediction going back to about 1970 [17]. Their empirical existence started to become clear around 1994 [18], and the emergence, over the following decade, of a coherent series of acoustic peaks and troughs is a triumph of modern cosmology. One can think of these peaks as a snapshot of stochastic standing waves. And, since the physics governing them is simple, then one can see how they encode information about the cosmological parameters.

23.4.3. *The damping tail:* $\ell \gtrsim 1000$:

The recombination process is not instantaneous, giving a thickness to the last scattering surface. This leads to a damping of the anisotropies at the highest ℓs, corresponding to scales smaller than that subtended by this thickness. One can also think of the photon-baryon fluid as having imperfect coupling, so that there is diffusion between the two components, and the oscillations have amplitudes that decrease with time. These effects lead to a damping of the C_ℓs, sometimes called Silk damping [19], which cuts off the anisotropies at multipoles above about 2000.

An extra effect at high ℓs comes from gravitational lensing, caused mainly by non-linear structures at low redshift. The C_ℓs are convolved with a smoothing function in a calculable way, partially flattening the peaks, generating a power-law tail at the highest multipoles, and complicating the polarization signal [20]. This is an example of a 'secondary effect', *i.e.*, the processing of anisotropies due to relatively nearby structures. Galaxies and clusters of galaxies give several such effects, but all are expected to be of low amplitude and are typically only important for the highest ℓs.

23.5. Current Anisotropy Data

There has been a steady improvement in the quality of CMB data that has led to the development of the present-day cosmological model. Probably the most robust constraints currently available come from the combination of the *WMAP* first year data [4] with smaller scale results from the CBI [21] and ACBAR [22] experiments. We plot these power spectrum estimates in Fig. 23.2. Other recent experiments, such as ARCHEOPS [23], BOOMERANG [24], DASI [25], MAXIMA [26], and VSA [27] also give powerful constraints, which are quite consistent with what we describe below. There have been some comparisons among data-sets [28], which indicate very good agreement, both in maps and in derived power spectra (up to systematic uncertainties in the overall calibration for some experiments). This makes it clear that systematic effects are largely under control. However, a fully self-consistent joint analysis of all the current data sets has not been attempted, one of the reasons being that it requires a careful treatment of the overlapping sky coverage.

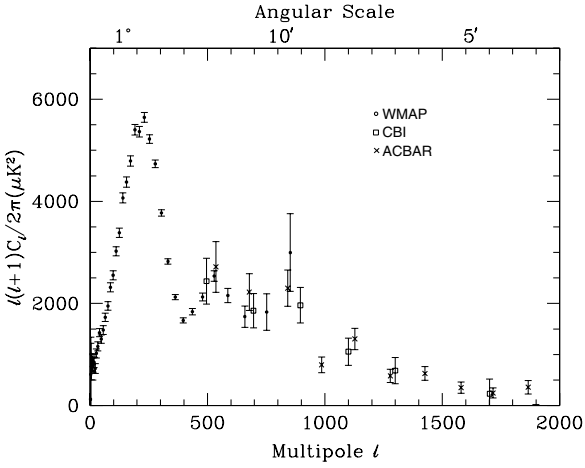

Figure 23.2: Band-power estimates from the *WMAP*, CBI, and ACBAR experiments. The *WMAP* data are the points, while squares are CBI and crosses ACBAR. We have shown only CBI and ACBAR data relevant for $\ell > 500$, and both experiments also probe to higher ℓ than shown. This plot represents only a fraction of experimental results, with several other data-sets being of similar quality. The multipole axis here is linear, so the Sachs-Wolfe plateau is hard to see. The acoustic peaks and damping region are very clearly observed, with no need for a theoretical curve to guide the eye.

Fig. 23.2 shows band-powers from the first year *WMAP* data [29], together with CBI and ACBAR data at higher ℓ. The points are in

very good agreement with a 'ΛCDM' type model, as described in the previous section, with several of the peaks and troughs quite apparent. For details of how these estimates were arrived at, the strength of any correlations between band-powers and other information required to properly interpret them, turn to the original papers [4,21,22].

23.6. CMB Polarization

Since Thomson scattering of an anisotropic radiation field also generates linear polarization, the CMB is predicted to be polarized at the roughly 5% level [30]. Polarization is a spin 2 field on the sky, and the algebra of the modes in ℓ-space is strongly analogous to spin-orbit coupling in quantum mechanics [31]. The linear polarization pattern can be decomposed in a number of ways, with two quantities required for each pixel in a map, often given as the Q and U Stokes parameters. However, the most intuitive and physical decomposition is a geometrical one, splitting the polarization pattern into a part that comes from a divergence (often referred to as the 'E-mode') and a part with a curl (called the 'B-mode') [32]. More explicitly, the modes are defined in terms of second derivatives of the polarization amplitude, with the Hessian for the E-modes having principle axes in the same sense as the polarization, while the B-mode pattern can be thought of simply as a 45° rotation of the E-mode pattern. Globally one sees that the E-modes have $(-1)^\ell$ parity (like the spherical harmonics), while the B-modes have $(-1)^{\ell+1}$ parity.

The existence of this linear polarization allows for 6 different cross power spectra to be determined from data that measure the full temperature and polarization anisotropy information. Parity considerations make 2 of these zero, and we are left with 4 potential observables: $C_\ell^{\rm TT}$, $C_\ell^{\rm TE}$, $C_\ell^{\rm EE}$, and $C_\ell^{\rm BB}$. Since scalar perturbations have no handedness, the B-mode power spectrum can only be generated by vectors or tensors. Hence, in the context of inflationary models, the determination of a non-zero B-mode signal is a way to measure the gravity wave contribution (and thus potentially derived the energy scale of inflation), even if it is rather weak. However, one must first eliminate the foreground contributions and other systematic effects down to very low levels.

The oscillating photon-baryon fluid also results in a series of acoustic peaks in the polarization power spectra. The main 'EE' power spectrum has peaks that are out of phase with those in the 'TT' spectrum, because the polarization anisotropies are sourced by the fluid velocity. The correlated component of the polarization and temperature patterns comes from correlations between density and velocity perturbations on the last scattering surface, which can be both positive and negative. There is no polarization 'Sachs-Wolfe' effect, and hence no large-angle plateau. However, scattering during a recent period of reionization can create a polarization 'bump' at large angular scales.

The strongest upper limits on polarization are at the roughly 10 μK level from the POLAR [33] experiment at large angular scales and the PIQUE [34] and COMPASS [35] experiments at smaller scales. The first measurement of a polarization signal came in 2002 from the DASI experiment [36], which provided a convincing detection, confirming the general paradigm, but of low enough significance that it lends little constraint to models. As well as the E-mode signal, DASI also made a statistical detection of the TE correlation.

More recently the *WMAP* experiment was able to measure the TE cross-correlation power spectrum with high precision [37]. The results are shown in Fig. 23.3, along with some estimates from the DASI experiment. The detected shape of the cross-correlation power spectrum provides supporting evidence of the adiabatic nature of the perturbations, as well as directly constraining the thickness of the last scattering surface. Since the polarization anisotropies are generated in this scattering surface, the existence of correlations at angles above about a degree demonstrate that there were super-Hubble fluctuations at the recombination epoch.

Perhaps the most intriguing result from the polarization measurements is at the largest angular scales ($\ell < 10$), where there is an excess signal compared to that expected from the temperature power spectrum alone. This is precisely the signal expected from an early period of reionization, arising from Doppler shifts during the partial scattering at $z < z_{\rm i}$. It seems to indicate that the first stars (presumably the source of the ionizing radiation) formed around $z = 20$.

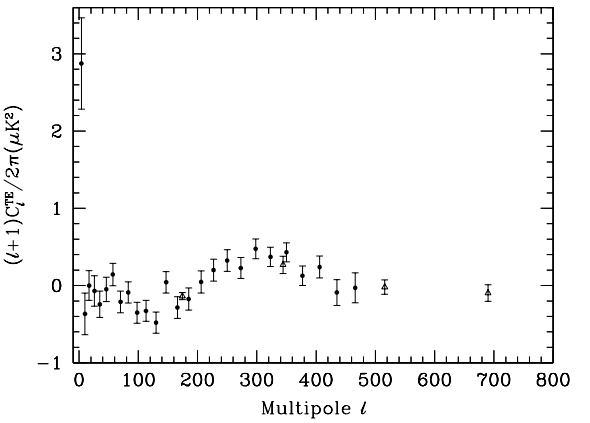

Figure 23.3: Cross power spectrum of the temperature anisotropies and E-mode polarization signal from *WMAP* (points), together with some estimates from DASI (triangles) which extend to higher ℓ. Note that the DASI bands are much wider in ℓ than those of *WMAP*. Also note that the y-axis is not multiplied by the additional ℓ, which helps to show both the large and small angular scale features.

23.7. Complications

There are a number of issues which complicate the interpretation of CMB anisotropy data, some of which we sketch out below.

23.7.1. *Foregrounds*:

The microwave sky contains significant emission from our Galaxy and from extra-galactic sources. Fortunately, the frequency dependence of these various sources are in general substantially different than the CMB anisotropy signals. The combination of Galactic synchrotron, bremsstrahlung and dust emission reaches a minimum at a wavelength of roughly 3 mm (or about 100 GHz). As one moves to greater angular resolution, the minimum moves to slightly higher frequencies, but becomes more sensitive to unresolved (point-like) sources.

At frequencies around 100 GHz and for portions of the sky away from the Galactic Plane the foregrounds are typically 1 to 10% of the CMB anisotropies. By making observations at multiple frequencies, it is relatively straightforward to separate the various components and determine the CMB signal to the few per cent level. For greater sensitivity it is necessary to improve the separation techniques by adding spatial information and statistical properties of the foregrounds compared to the CMB.

The foregrounds for CMB polarization are expected to follow a similar pattern, but are less well studied, and are intrinsically more complicated. Whether it is possible to achieve sufficient separation to detect B-mode CMB polarization is still an open question. However, for the time being, foreground contamination is not a major issue for CMB experiments.

23.7.2. *Secondary Anisotropies*:

With increasingly precise measurements of the primary anisotropies, there is growing theoretical and experimental interest in 'secondary anisotropies.' Effects which happen at $z \ll 1000$ become more important as experiments push to higher angular resolution and sensitivity.

These secondary effects include gravitational lensing, patchy reionization and the Sunyaev-Zel'dovich (SZ) effect [38]. This is Compton scattering ($\gamma e \to \gamma' e'$) of the CMB photons by a hot electron gas, which creates spectral distortions by transferring energy from the electrons to the photons. The effect is particularly important for clusters of galaxies, through which one observes a partially Comptonized spectrum, resulting in a decrement at radio wavelengths and an increment in the submillimeter. This can be used to find and study individual clusters and to obtain estimates of the Hubble

constant. There is also the potential to constrain the equation of state of the Dark Energy through counts of clusters as a function of redshift [39].

23.7.3. *Higher-order Statistics*:

Although most of the CMB anisotropy information is contained in the power spectra, there will also be weak signals present in higher-order statistics. These statistics will measure primordial non-Gaussianity in the perturbations, as well as non-linear growth of the fluctuations on small scales and other secondary effects (plus residual foreground contamination). Although there are an infinite variety of ways in which the CMB could be non-Gaussian, there is a generic form to consider for the initial conditions, where a quadratic contribution to the curvature perturbations is parameterized through a dimensionless number $f_{\rm NL}$. This weakly non-linear component can be constrained through measurements of the bispectrum or Minkowski functionals for example, and the result from *WMAP* is $-58 < f_{\rm NL} < 134$ (95% confidence region) [10].

23.8. Constraints on Cosmologies

The most important outcome of the newer experimental results is that the standard cosmological paradigm is in good shape. A large amount of high precision data on the power spectrum is adequately fit with fewer than 10 free parameters. The framework is that of Friedmann-Robertson-Walker models, which have nearly flat geometry, containing Dark Matter and Dark Energy, and with adiabatic perturbations having close to scale invariant initial conditions.

Within this framework, bounds can be placed on the values of the cosmological parameters. Of course, much more stringent constraints can be placed on models which cover a restricted number of parameters, e.g. assuming that $\Omega_{\rm tot} = 1$, $n = 1$ or $r = 0$. More generally, the constraints depend upon the adopted priors, even if they are implicit, for example by restricting the parameter freedom or the ranges of parameters (particularly where likelihoods peak near the boundaries), or by using different choices of other data in combination with the CMB. When the data become even more precise, these considerations will become less important, but for now we caution that restrictions on model space and choice of priors need to be kept in mind when adopting specific parameter values and uncertainties.

There are some combinations of parameters that fit the CMB anisotropies almost equivalently. For example, there is a nearly exact geometric degeneracy, where any combination of $\Omega_{\rm M}$ and Ω_Λ that gives the same angular diameter distance to last scattering will give nearly identical C_ℓs. There are also other near degeneracies among the parameters. Such degeneracies can be broken when using the CMB data in combination with other cosmological data sets. Particularly useful are complementary constraints from galaxy clustering, the abundance of galaxy clusters, weak gravitational lensing measurements, Type Ia supernova distances and the distribution of Lyman α forest clouds. For an overview of some of these other cosmological constraints, see the review on The Cosmological Parameters—Sec. 21 of this *Review*.

The combination of *WMAP*, CBI and ACBAR, together with weak priors (on h and $\Omega_{\rm B}h^2$ for example), and within the context of a 6 parameter family of models (which fixes $\Omega_{\rm tot} = 1$), yields the following results [40]: $A = 2.7(\pm0.3) \times 10^{-9}$, $n = 0.97 \pm 0.03$, $h = 0.73 \pm 0.05$, $\Omega_{\rm B}h^2 = 0.023 \pm 0.001$, $\Omega_{\rm M}h^2 = 0.13 \pm 0.01$ and $\tau = 0.17 \pm 0.07$. Note that for h, the CMB data alone provide only a very weak constraint, unless spatial flatness or some other cosmological data are used. For $\Omega_{\rm B}h^2$ the precise value depends sensitively on how much freedom is allowed in the shape of the primordial power spectrum (see the review on Big-Bang nucleosynthesis—Sec. 20 of this *Review*). For the optical depth τ, the error bar is large enough that apparently quite different results can come from other combinations of data.

The best constraint on $\Omega_{\rm tot}$ is 1.02 ± 0.02. This comes from including priors from h and supernova data. Slightly different, but consistent results come from using different data combinations.

The 95% confidence upper limit on r is 0.53 (including some extra constraint from galaxy clustering). This limit is stronger if we restrict ourselves to $n < 1$ and weaker if we allow $dn/d\ln k \neq 0$.

There are also constraints on parameters over and above the basic 8 that we have described. But for such constraints it is necessary to include additional data in order to break the degeneracies. For example the addition of the Dark Energy equation of state, w adds the partial degeneracy of being able to fit a ridge in (w, h) space, extending to low values of both parameters. This degeneracy is broken when the CMB is used in combination with independent H_0 limits, for example [13], giving $w < -0.5$ at 95% confidence. Tighter limits can be placed using restrictive model-spaces and/or additional data.

For the optical depth τ, the error bar is large enough that apparently quite different results can come from other combinations of data. The constraint from the combined *WMAP* $C_\ell^{\rm TT}$ and $C_\ell^{\rm TE}$ data is $\tau = 0.17 \pm 0.04$, which corresponds (within reasonable models) to a reionization redshift $9 < z_{\rm i} < 30$ (95% CL) [37]. This is a little higher than some theoretical predictions and some suggestions from studies of absorption in high-z quasar spectra [42]. The excitement here is that we have direct information from CMB polarization which can be combined with other astrophysical measurements to understand when the first stars formed and brought about the end of the cosmic dark ages.

23.9. Particle Physics Constraints

CMB data are beginning to put limits on parameters which are directly relevant for particle physics models. For example there is a limit on the neutrino contribution $\Omega_\nu h^2 < 0.0076$ (95% confidence) from a combination of *WMAP* and galaxy clustering data from the 2dFGRS project [43]. This directly implies a limit on neutrino mass, assuming the usual number density of fermions which decoupled when they were relativistic.

A combination of the *WMAP* data with other data-sets gives some hint of a running spectral index, *i.e.*, $dn/d\ln k \neq 0$ [40]. Although this is still far from resolved [44], things will certainly improve as new data come in. A convincing measurement of a non-zero running of the index would be quite constraining for inflationary models [45].

One other hint of new physics lies in the fact that the quadrupole and some of the other low ℓ modes seem anomalously low compared with the best-fit ΛCDM model [29]. This is what might be expected in a universe which has a large scale cut-off to the power spectrum, or is topologically non-trivial. However, because of cosmic variance, possible foregrounds *etc.*, the significance of this feature is still a matter of debate [46].

In addition it is also possible to put limits on other pieces of physics [47], for example the neutrino chemical potentials, time variation of the fine-structure constant, or physics beyond general relativity. Further particle physics constraints will follow as the anisotropy measurements increase in precision.

Careful measurement of the CMB power spectra and non-Gaussianity can in principle put constraints on high energy physics, including ideas of string theory, extra dimensions, colliding branes, *etc.* At the moment any calculation of predictions appears to be far from definitive. However, there is a great deal of activity on implications of string theory for the early Universe, and hence a very real chance that there might be observational implications for specific scenarios.

23.10. Fundamental Lessons

More important than the precise values of parameters is what we have learned about the general features which describe our observable Universe. Beyond the basic hot Big Bang picture, the CMB has taught us that:

- The Universe recombined at $z \simeq 1100$ and started to become ionized again at $z \simeq 10$–30.
- The geometry of the Universe is close to flat.
- Both Dark Matter and Dark Energy are required.
- Gravitational instability is sufficient to grow all of the observed large structures in the Universe.
- Topological defects were not important for structure formation.
- There are 'synchronized' super-Hubble modes generated in the early Universe.
- The initial perturbations were adiabatic in nature.
- The perturbations had close to Gaussian (*i.e.*, maximally random) initial conditions.

It is very tempting to make an analogy between the status of the cosmological 'Standard Model' and that of particle physics. In cosmology there are about 10 free parameters, each of which is becoming well determined, and with a great deal of consistency between different measurements. However, none of these parameters can be calculated from a fundamental theory, and so hints of the bigger picture, 'physics beyond the Standard Model' are being searched for with ever more challenging experiments.

Despite this analogy, there are some basic differences. For one thing, many of the cosmological parameters change with cosmic epoch, and so the measured values are simply the ones determined today, and hence they are not 'constants', like particle masses for example (although they *are* deterministic, so that if one knows their values at one epoch, they can be calculated at another). Moreover, the number of parameters is not as fixed as it is in the particle physics Standard Model; different researchers will not necessarily agree on what the free parameters are, and new ones can be added as the quality of the data improves. In addition parameters like τ, which come from astrophysics, are in principle calculable from known physical processes, although this is currently impractical. On top of all this, other parameters might be 'stochastic' in that they may be fixed only in our observable patch of the Universe.

In a more general sense the cosmological 'Standard Model' is much further from the underlying 'fundamental theory' which will provide the values of the parameters from first principles. On the other hand, any genuinely complete 'theory of everything' must include an explanation for the values of these cosmological parameters as well as the parameters of the Standard Model.

23.11. Future Directions

With all the observational progress in the CMB and the tying down of cosmological parameters, what can we anticipate for the future? Of course there will be a steady improvement in the precision and confidence with which we can determine the appropriate cosmological model and its parameters. We can anticipate that the evolution from one year to four years of *WMAP* data will bring improvements from the increased statistical accuracy and from the more detailed treatment of calibration and systematic effects. Ground-based experiments operating at the smaller angular scales will also improve over the next few years, providing significantly tighter constraints on the damping tail. In addition, the next CMB satellite mission, *Planck*, is scheduled for launch in 2007, and there are even more ambitious projects currently being discussed.

Despite the increasing improvement in the results, it is also true that the addition of the latest experiments has not significantly changed the cosmological model (apart from a suggestion of higher reionization redshift perhaps). It is therefore appropriate to ask: what should we expect to come from *Planck* and from other more grandiose future experiments, including the proposed *Inflation Probe* or *CMBPol*? *Planck* certainly has the the advantage of high sensitivity and a full sky survey. A detailed measurement of the third acoustic peak provides a good determination of the matter density; this can only be done by measurements which are accurate relative to the first two peaks (which themselves constrained the curvature and the baryon density). A detailed measurement of the damping tail region will also significantly improve the determination of n and any running of the slope. *Planck* should also be capable of measuring $C_\ell^{\rm EE}$ quite well, providing both a strong check on the Standard Model and extra constraints that will improve parameter estimation.

A set of cosmological parameters are now known to roughly 10% accuracy, and that may seem sufficient for many people. However, we should certainly demand more of measurements which describe *the entire observable Universe*! Hence a lot of activity in the coming years will continue to focus on determining those parameters with increasing precision. This necessarily includes testing for consistency among different predictions of the Standard Model, and searching for signals which might require additional physics.

A second area of focus will be the smaller scale anisotropies and 'secondary effects.' There is a great deal of information about structure formation at $z \ll 1000$ encoded in the CMB sky. This may involve higher-order statistics as well as spectral signatures. Such investigations can also provide constraints on the Dark Energy equation of state, for example. *Planck*, as well as experiments aimed at the highest ℓs, should be able to make a lot of progress in this arena.

A third direction is increasingly sensitive searches for specific signatures of physics at the highest energies. The most promising of these may be the primordial gravitational wave signals in $C_\ell^{\rm BB}$, which could be a probe of the $\sim 10^{16}$ GeV energy range. Whether the amplitude of the effect coming from inflation will be detectable is unclear, but the prize makes the effort worthwhile.

Anisotropies in the CMB have proven to be the premier probe of cosmology and the early Universe. Theoretically the CMB involves well-understood physics in the linear regime, and is under very good calculational control. A substantial and improving set of observational data now exists. Systematics appear to be well understood and not a limiting factor. And so for the next few years we can expect an increasing amount of cosmological information to be gleaned from CMB anisotropies, with the prospect also of some genuine surprises.

References:

1. A.A. Penzias and R. Wilson, Astrophys. J. **142**, 419 (1965); R.H. Dicke *et al.*, Astrophys. J. **142**, 414 (1965).
2. M. White, D. Scott, and J. Silk, Ann. Rev. Astron. & Astrophys. **32**, 329 (1994); W. Hu and S. Dodelson, Ann. Rev. Astron. & Astrophys. **40**, 171 (2002).
3. G.F. Smoot *et al.*, Astrophys. J. **396**, L1 (1992).
4. C.L. Bennett *et al.*, Astrophys. J. Supp. **148**, 1 (2003).
5. J.C. Mather *et al.*, Astrophys. J. **512**, 511 (1999).
6. S. Courteau *et al.*, Astrophys. J. **544**, 636 (2000).
7. D.J. Fixsen *et al.*, Astrophys. J. **420**, 445 (1994).
8. D.J. Fixsen *et al.*, Astrophys. J. **473**, 576 (1996); A. Kogut *et al.*, Astrophys. J. **419**, 1 (1993).
9. S. Seager, D.D. Sasselov, and D. Scott, Astrophys. J. Supp. **128**, 407 (2000).
10. E. Komatsu *et al.*, Astrophys. J. Supp. **148**, 119 (2003).
11. W. Hu and D.J. Eisenstein, Phys. Rev. **D59**, 083509 (1999); W. Hu *et al.*, Phys. Rev. **D59**, 023512 (1999).
12. U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996).
13. A. Lewis, A. Challinor, A. Lasenby, Astrophys. J. **538**, 473 (2000).
14. U. Seljak *et al.*, Physical Review D, in press, `astro-ph/0306052`.
15. R.K. Sachs and A.M. Wolfe, Astrophys. J. **147**, 73 (1967).
16. M.R. Nolta *et al.*, Astrophysical J., in press, `astro-ph/0305097`.
17. P.J.E. Peebles and J.T. Yu, Astrophys. J. **162**, 815 (1970); R.A. Sunyaev and Ya.B. Zel'dovich, Astrophysics & Space Science **7**, 3 (1970).
18. D. Scott, J. Silk, and M. White, Science **268**, 829 (1995).
19. J. Silk, Astrophys. J. **151**, 459 (1968).
20. M. Zaldarriaga and U. Seljak, Phys. Rev. **D58**, 023003 (1998).
21. T.J. Pearson *et al.*, Astrophys. J. **591**, 556 (2003).
22. C.L. Kuo *et al.*, Astrophysical J., in press, `astro-ph/0212289`.
23. A. Benoit *et al.*, Astronomy & Astrophysics **399**, L19 (2003).
24. J.E. Ruhl *et al.*, Astrophysical J., in press, `astro-ph/0212229`.
25. N.W. Halverson *et al.*, Astrophys. J. **568**, 38 (2002).
26. A.T. Lee *et al.*, Astrophys. J. **561**, L1 (2001).
27. P.F. Scott *et al.*, Monthly Not. Royal Astron. Soc. **341**, 1076 (2003).
28. M.E. Abroe *et al.*, Astrophysical J., in press, `astro-ph/0308355`.
29. G. Hinshaw *et al.*, Astrophys. J. Supp. **148**, 135 (2003).
30. W. Hu, M. White, New Astron. **2**, 323 (1997).
31. W. Hu, M. White, Phys. Rev. **D56**, 596 (1997).
32. M. Zaldarriaga and U. Seljak, Phys. Rev. **D55**, 1830 (1997); M. Kamionkowski, A. Kosowsky, and A. Stebbins, Phys. Rev. **D55**, 7368 (1997).

33. B.G. Keating *et al.*, Astrophysical J., in press, `astro-ph/0107013`.
34. M.M. Hedman *et al.*, Astrophys. J. **548**, L111 (2001).
35. P.C. Farese *et al.*, Astrophysical J., in press, `astro-ph/0308309`.
36. J. Kovac *et al.*, Nature, 420, 772 (2002).
37. A. Kogut *et al.*, Astrophys. J. Supp. **148**, 161 (2003).
38. R.A. Sunyaev and Ya.B. Zel'dovich, Ann. Rev. Astron. Astrophys. **18**, 537 (1980);
    M. Birkinshaw, Phys. Rep. **310**, 98 (1999).
39. J.E. Carlstrom, G.P. Holder, and E.D. Reese, Ann. Rev. Astron. & Astrophys. **40**, 643 (2002).
40. D.N. Spergel *et al.*, Astrophys. J. Supp. **148**, 175 (2003).
41. W.L. Freedman *et al.*, Astrophys. J. **553**, 47 (2001).
42. X. Fan *et al.*, Astrophys. J. **123**, 1247 (2002).
43. M. Colless *et al.*, Monthly Not. Royal Astron. Soc. **328**, 1039 (2001).
44. U. Seljak, P. McDonald, and A. Makarov, Monthly Not. Royal Astron. Soc. **342**, L79 (2003).
45. H.V. Peiris *et al.*, Astrophys. J. Supp. **148**, 213 (2003).
46. A. de Oliveira-Costa *et al.*, Physical Review D, in press, `astro-ph/0307282`;
    G. Efstathiou, Monthly Notices of the Royal Astronomical Society, in press, `astro-ph/0306431`.
47. M. Kamionkowski and A. Kosowsky, Ann. Rev. Nucl. Part. Sci. **49**, 77 (1999).

24. COSMIC RAYS

Revised March 2002 by T.K. Gaisser and T. Stanev (Bartol Research Inst., University of Delaware).

24.1. Primary spectra

The cosmic radiation incident at the top of the terrestrial atmosphere includes all stable charged particles and nuclei with lifetimes of order 10^6 years or longer. Technically, "primary" cosmic rays are those particles accelerated at astrophysical sources and "secondaries" are those particles produced in interaction of the primaries with interstellar gas. Thus electrons, protons and helium, as well as carbon, oxygen, iron, and other nuclei synthesized in stars, are primaries. Nuclei such as lithium, beryllium, and boron (which are not abundant end-products of stellar nucleosynthesis) are secondaries. Antiprotons and positrons are also in large part secondary. Whether a small fraction of these particles may be primary is a question of current interest.

Apart from particles associated with solar flares, the cosmic radiation comes from outside the solar system. The incoming charged particles are "modulated" by the solar wind, the expanding magnetized plasma generated by the Sun, which decelerates and partially excludes the lower energy galactic cosmic rays from the inner solar system. There is a significant anticorrelation between solar activity (which has an alternating eleven-year cycle) and the intensity of the cosmic rays with energies below about 10 GeV. In addition, the lower-energy cosmic rays are affected by the geomagnetic field, which they must penetrate to reach the top of the atmosphere. Thus the intensity of any component of the cosmic radiation in the GeV range depends both on the location and time.

There are four different ways to describe the spectra of the components of the cosmic radiation: (1) By particles per unit rigidity. Propagation (and probably also acceleration) through cosmic magnetic fields depends on gyroradius or *magnetic rigidity, R*, which is gyroradius multiplied by the magnetic field strength:

$$R = \frac{p\,c}{Z\,e} = r_L\,B\,. \qquad (24.1)$$

(2) By particles per energy-per-nucleon. Fragmentation of nuclei propagating through the interstellar gas depends on energy per nucleon, since that quantity is approximately conserved when a nucleus breaks up on interaction with the gas. (3) By nucleons per energy-per-nucleon. Production of secondary cosmic rays in the atmosphere depends on the intensity of nucleons per energy-per-nucleon, approximately independently of whether the incident nucleons are free protons or bound in nuclei. (4) By particles per energy-per-nucleus. Air shower experiments that use the atmosphere as a calorimeter generally measure a quantity that is related to total energy per particle.

The units of differential intensity I are $[\mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}\mathcal{E}^{-1}]$, where $\mathcal{E}$ represents the units of one of the four variables listed above.

The intensity of primary nucleons in the energy range from several GeV to somewhat beyond 100 TeV is given approximately by

$$I_N(E) \;\approx\; 1.8\;E^{-\alpha}\;\frac{\text{nucleons}}{\mathrm{cm}^2\ \mathrm{s}\ \mathrm{sr}\ \mathrm{GeV}}\,, \qquad (24.2)$$

where E is the energy-per-nucleon (including rest mass energy) and $\alpha\ (\equiv \gamma + 1) = 2.7$ is the differential spectral index of the cosmic ray flux and γ is the integral spectral index. About 79% of the primary nucleons are free protons and about 70% of the rest are nucleons bound in helium nuclei. The fractions of the primary nuclei are nearly constant over this energy range (possibly with small but interesting variations). Fractions of both primary and secondary incident nuclei are listed in Table 24.1. Figure 24.1 [1] shows the major components as a function of energy at a particular epoch of the solar cycle. There has been a series of more precise measurements of the primary spectrum of protons and helium in the past decade [2–6].

The spectrum of electrons and positrons incident at the top of the atmosphere is steeper than the spectra of protons and nuclei,

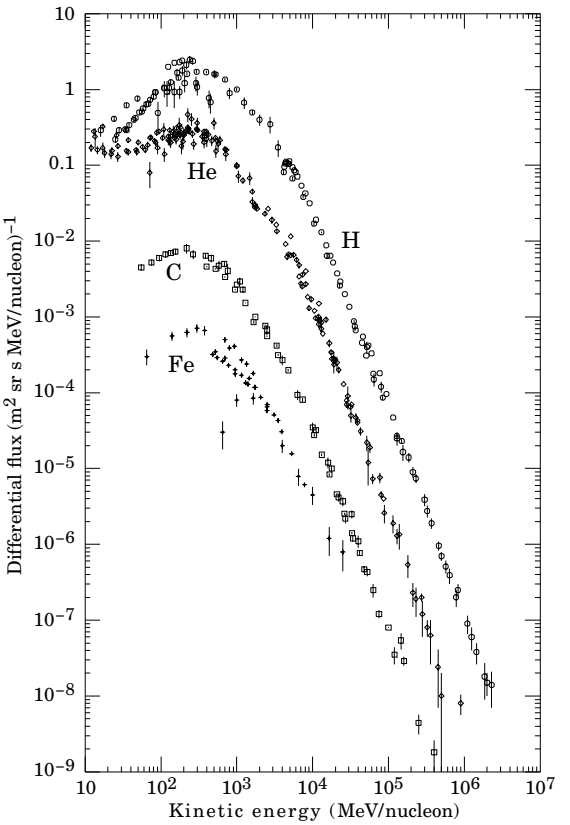

Figure 24.1: Major components of the primary cosmic radiation (from Ref. 1).

Table 24.1: Relative abundances F of cosmic-ray nuclei at 10.6 GeV/nucleon normalized to oxygen ($\equiv$ 1) [7]. The oxygen flux at kinetic energy of 10.6 GeV/nucleon is 3.26×10^{-6} cm^{-2} s^{-1} sr^{-1} (GeV/nucleon)$^{-1}$. Abundances of hydrogen and helium are from Ref. [5,6].

Z	Element	F	Z	Element	F
1	H	540	13–14	Al-Si	0.19
2	He	26	15–16	P-S	0.03
3–5	Li-B	0.40	17–18	Cl-Ar	0.01
6–8	C-O	2.20	19–20	K-Ca	0.02
9–10	F-Ne	0.30	21–25	Sc-Mn	0.05
11–12	Na-Mg	0.22	26–28	Fe-Ni	0.12

as shown in Fig. 24.2. The positron fraction decreases from ~ 0.2 below 1 GeV [9–11] to ~ 0.1 around 2 GeV and to ~ 0.05 in at the highest energies for which it is measured ($5 - 20$ GeV) [12]. This behavior refers to measurements made during solar cycles of positive magnetic polarity and at high geomagnetic latitude. Ref. 11 discusses the dependence of the positron fraction on solar cycle and Ref. 5 studies the geomagnetic effects.

The ratio of antiprotons to protons is $\sim 2 \times 10^{-4}$ [13,14] at around 10–20 GeV, and there is clear evidence [15–17] for the kinematic suppression at lower energy that is the signature of secondary antiprotons. The $\overline{p}/p$ ratio also shows a strong dependence on the phase and polarity of the solar cycle [18] in the opposite sense to

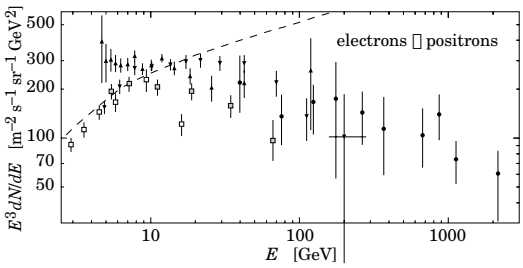

Figure 24.2: Differential spectrum of electrons plus positrons multiplied by E^3 (data summary from Ref. 8). The dashed line shows the proton spectrum multiplied by 0.01.

that of the positron fraction. There is at this time no evidence for a significant primary component either of positrons or of antiprotons.

24.2. Cosmic rays in the atmosphere

Figure 24.3 shows the vertical fluxes of the major cosmic ray components in the atmosphere in the energy region where the particles are most numerous (except for electrons, which are most numerous near their critical energy, which is about 81 MeV in air). Except for protons and electrons near the top of the atmosphere, all particles are produced in interactions of the primary cosmic rays in the air. Muons and neutrinos are products of the decay of charged mesons, while electrons and photons originate in decays of neutral mesons.

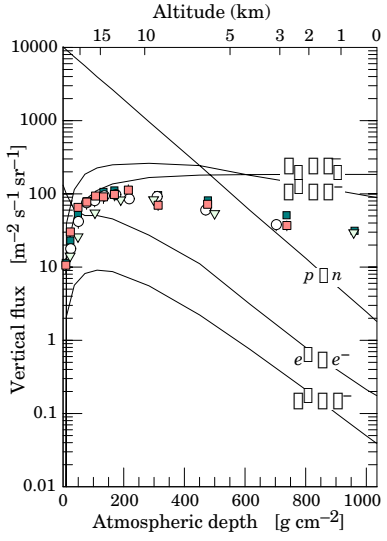

Figure 24.3: Vertical fluxes of cosmic rays in the atmosphere with $E > 1$ GeV estimated from the nucleon flux of Eq. (24.2). The points show measurements of negative muons with $E_\mu > 1$ GeV [4,19,20,21].

Most measurements are made at ground level or near the top of the atmosphere, but there are also measurements of muons and electrons from airplanes and balloons. Fig. 24.3 includes recent measurements of negative muons [4,19,20,21]. Since $\mu^+(\mu^-)$ are produced in association with $\nu_\mu(\overline{\nu}_\mu)$, the measurement of muons near the maximum of the intensity curve for the parent pions serves to calibrate the atmospheric ν_μ beam [22]. Because muons typically lose almost two GeV in passing through the atmosphere, the comparison near the production altitude is important for the sub-GeV range of $\nu_\mu(\overline{\nu}_\mu)$ energies.

The flux of cosmic rays through the atmosphere is described by a set of coupled cascade equations with boundary conditions at the top of the atmosphere to match the primary spectrum. Numerical or Monte Carlo calculations are needed to account accurately for decay and energy-loss processes, and for the energy-dependences of the cross sections and of the primary spectral index γ. Approximate analytic solutions are, however, useful in limited regions of energy [23]. For example, the vertical intensity of nucleons at depth X (g cm^{-2}) in the atmosphere is given by

$$I_N(E, X) \approx I_N(E, 0)\, e^{-X/\Lambda} , \qquad (24.3)$$

where Λ is the attenuation length of nucleons in air.

The corresponding expression for the vertical intensity of charged pions with energy $E_\pi \ll \epsilon_\pi = 115$ GeV is

$$I_\pi(E_\pi, X) \approx \frac{Z_{N\pi}}{\lambda_N} I_N(E_\pi, 0)\, e^{-X/\Lambda} \frac{X\, E_\pi}{\epsilon_\pi} . \qquad (24.4)$$

This expression has a maximum at $t = \Lambda \approx 120$ g cm^{-2}, which corresponds to an altitude of 15 kilometers. The quantity $Z_{N\pi}$ is the spectrum-weighted moment of the inclusive distribution of charged pions in interactions of nucleons with nuclei of the atmosphere. The intensity of low-energy pions is much less than that of nucleons because $Z_{N\pi} \approx 0.079$ is small and because most pions with energy much less than the critical energy ϵ_π decay rather than interact.

24.3. Cosmic rays at the surface

24.3.1. ***Muons***: Muons are the most numerous charged particles at sea level (see Fig. 24.3). Most muons are produced high in the atmosphere (typically 15 km) and lose about 2 GeV to ionization before reaching the ground. Their energy and angular distribution reflect a convolution of production spectrum, energy loss in the atmosphere, and decay. For example, 2.4 GeV muons have a decay length of 15 km, which is reduced to 8.7 km by energy loss. The mean energy of muons at the ground is ≈ 4 GeV. The energy spectrum is almost flat below 1 GeV, steepens gradually to reflect the primary spectrum in the 10–100 GeV range, and steepens further at higher energies because pions with $E_\pi > \epsilon_\pi \approx 115$ GeV tend to interact in the atmosphere before they decay. Asymptotically ($E_\mu \gg 1$ TeV), the energy spectrum of atmospheric muons is one power steeper than the primary spectrum. The integral intensity of vertical muons above 1 GeV/c at sea level is ≈ 70 m^{-2}s^{-1}sr^{-1} [24,25], with recent measurements [26–28] tending to give lower normalization by 10-15%. Experimentalists are familiar with this number in the form $I \approx 1$ cm^{-2} min^{-1} for horizontal detectors.

The overall angular distribution of muons at the ground is $\propto \cos^2\theta$, which is characteristic of muons with $E_\mu \sim 3$ GeV. At lower energy the angular distribution becomes increasingly steep, while at higher energy it flattens, approaching a $\sec\theta$ distribution for $E_\mu \gg \epsilon_\pi$ and $\theta < 70°$.

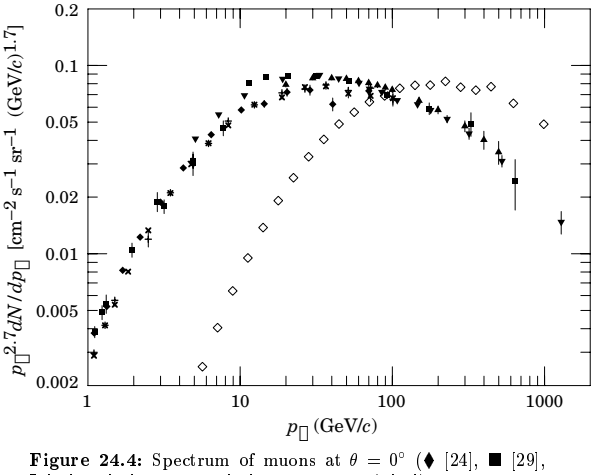

Figure 24.4: Spectrum of muons at $\theta = 0°$ (⧫ [24], ■ [29], ▼ [30], ▲ [31], × and + [26], and $\theta = 75°$ ◊ [32]).

Figure 24.4 shows the muon energy spectrum at sea level for two angles. At large angles low energy muons decay before reaching the surface and high energy pions decay before they interact, thus the average muon energy increases. An approximate extrapolation formula valid when muon decay is negligible ($E_\mu > 100/\cos\theta$ GeV) and the curvature of the Earth can be neglected ($\theta < 70^\circ$) is

$$\frac{dN_\mu}{dE_\mu} \approx \frac{0.14\, E_\mu^{-2.7}}{\mathrm{cm}^2\ \mathrm{s}\ \mathrm{sr}\ \mathrm{GeV}} \times \left\{ \frac{1}{1+\dfrac{1.1 E_\mu \cos\theta}{115\,\mathrm{GeV}}} + \frac{0.054}{1+\dfrac{1.1 E_\mu \cos\theta}{850\,\mathrm{GeV}}} \right\}, \qquad (24.5)$$

where the two terms give the contribution of pions and charged kaons. Eq. (24.5) neglects a small contribution from charm and heavier flavors which is negligible except at very high energy [33].

The muon charge ratio reflects the excess of π^+ over π^- in the forward fragmentation region of proton initiated interactions together with the fact that there are more protons than neutrons in the primary spectrum. The charge ratio is between 1.1 and 1.4 from 1 GeV to 100 GeV [24,26,27]. Below 1 GeV there is a systematic dependence on location due to geomagnetic effects [26,27].

24.3.2. ***Electromagnetic component***: At the ground, this component consists of electrons, positrons, and photons primarily from electromagnetic cascades initiated by decay of neutral and charged mesons. Muon decay is the dominant source of low-energy electrons at sea level. Decay of neutral pions is more important at high altitude or when the energy threshold is high. Knock-on electrons also make a small contribution at low energy [34]. The integral vertical intensity of electrons plus positrons is very approximately 30, 6, and 0.2 $\mathrm{m}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}$ above 10, 100, and 1000 MeV respectively [25,35], but the exact numbers depend sensitively on altitude, and the angular dependence is complex because of the different altitude dependence of the different sources of electrons [34,35,36]. The ratio of photons to electrons plus positrons is approximately 1.3 above a GeV and 1.7 below the critical energy [36].

24.3.3. ***Protons***: Nucleons above 1 GeV/c at ground level are degraded remnants of the primary cosmic radiation. The intensity is approximately represented by Eq. (24.3) with the replacement $t \to t/\cos\theta$ for $\theta < 70^\circ$ and an attenuation length $\Lambda = 123$ g cm^{-2}. At sea level, about 1/3 of the nucleons in the vertical direction are neutrons (up from $\approx 10\%$ at the top of the atmosphere as the n/p ratio approaches equilibrium). The integral intensity of vertical protons above 1 GeV/c at sea level is $\approx 0.9\ \mathrm{m}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}$ [25,37].

24.4. Cosmic rays underground

Only muons and neutrinos penetrate to significant depths underground. The muons produce tertiary fluxes of photons, electrons, and hadrons.

24.4.1. ***Muons***: As discussed in Section 27.6 of this *Review*, muons lose energy by ionization and by radiative processes: bremsstrahlung, direct production of e^+e^- pairs, and photonuclear interactions. The total muon energy loss may be expressed as a function of the amount of matter traversed as

$$-\frac{dE_\mu}{dX} = a + b\,E_\mu\ , \qquad (24.6)$$

where a is the ionization loss and b is the fractional energy loss by the three radiation processes. Both are slowly varying functions of energy. The quantity $\epsilon \equiv a/b$ (≈ 500 GeV in standard rock) defines a critical energy below which continuous ionization loss is more important than radiative losses. Table 24.2 shows a and b values for standard rock as a function of muon energy. The second column of Table 24.2 shows the muon range in standard rock ($A = 22$, $Z = 11$, $\rho = 2.65$ g cm^{-3}). These parameters are quite sensitive to the chemical composition of the rock, which must be evaluated for each experimental location.

Table 24.2: Average muon range R and energy loss parameters calculated for standard rock [38]. Range is given in km-water-equivalent, or 10^5 g cm^{-2}.

E_μ	R	a	b_{brems}	b_{pair}	b_{nucl}	$\sum b_i$
GeV	km.w.e.	MeV g^{-1} cm^2	———	10^{-6} g^{-1} cm^2		———
10	0.05	2.17	0.70	0.70	0.50	1.90
100	0.41	2.44	1.10	1.53	0.41	3.04
1000	2.45	2.68	1.44	2.07	0.41	3.92
10000	6.09	2.93	1.62	2.27	0.46	4.35

The intensity of muons underground can be estimated from the muon intensity in the atmosphere and their rate of energy loss. To the extent that the mild energy dependence of a and b can be neglected, Eq. (24.6) can be integrated to provide the following relation between the energy $E_{\mu,0}$ of a muon at production in the atmosphere and its average energy E_μ after traversing a thickness X of rock (or ice or water):

$$E_\mu = (E_{\mu,0} + \epsilon)\, e^{-bX} - \epsilon\ . \qquad (24.7)$$

Especially at high energy, however, fluctuations are important and an accurate calculation requires a simulation that accounts for stochastic energy-loss processes [39].

There are two depth regimes for Eq. (24.7). For $X \ll b^{-1} \approx 2.5$ km water equivalent, $E_{\mu,0} \approx E_\mu(X) + aX$, while for $X \gg b^{-1}$ $E_{\mu,0} \approx (\epsilon + E_\mu(X))\exp(bX)$. Thus at shallow depths the differential muon energy spectrum is approximately constant for $E_\mu < aX$ and steepens to reflect the surface muon spectrum for $E_\mu > aX$, whereas for $X > 2.5$ km.w.e. the differential spectrum underground is again constant for small muon energies but steepens to reflect the surface muon spectrum for $E_\mu > \epsilon \approx 0.5$ TeV. In the deep regime the shape is independent of depth although the intensity decreases exponentially with depth. In general the muon spectrum at slant depth X is

$$\frac{dN_\mu(X)}{dE_\mu} = \frac{dN_\mu}{dE_{\mu,0}}\frac{dE_{\mu,0}}{dE_\mu} = \frac{dN_\mu}{dE_{\mu,0}}\, e^{bX}\ , \qquad (24.8)$$

where $E_{\mu,0}$ is the solution of Eq. (24.7) in the approximation neglecting fluctuations.

Fig. 24.5 shows the vertical muon intensity versus depth. In constructing this "depth-intensity curve," each group has taken account of the angular distribution of the muons in the atmosphere, the map of the overburden at each detector, and the properties of the local medium in connecting measurements at various slant depths and zenith angles to the vertical intensity. Use of data from a range of angles allows a fixed detector to cover a wide range of depths. The flat portion of the curve is due to muons produced locally by charged-current interactions of ν_μ.

24.4.2. ***Neutrinos***: Because neutrinos have small interaction cross sections, measurements of atmospheric neutrinos require a deep detector to avoid backgrounds. There are two types of measurements: contained (or semi-contained) events, in which the vertex is determined to originate inside the detector, and neutrino-induced muons. The latter are muons that enter the detector from zenith angles so large (*e.g.*, nearly horizontal or upward) that they cannot be muons produced in the atmosphere. In neither case is the neutrino flux measured directly. What is measured is a convolution of the neutrino flux and cross section with the properties of the detector (which includes the surrounding medium in the case of entering muons).

Contained and semi-contained events reflect neutrinos in the sub-GeV to multi-GeV region where the product of increasing cross section and decreasing flux is maximum. In the GeV region the neutrino flux and its angular distribution depend on the geomagnetic location of the detector and, to a lesser extent, on the phase of the solar cycle. Naively, we expect $\nu_\mu/\nu_e = 2$ from counting neutrinos of the two flavors coming from the chain of pion and muon decay. This ratio is only slightly modified by the details of the decay kinematics, but the fraction of electron neutrinos gradually decreases above a

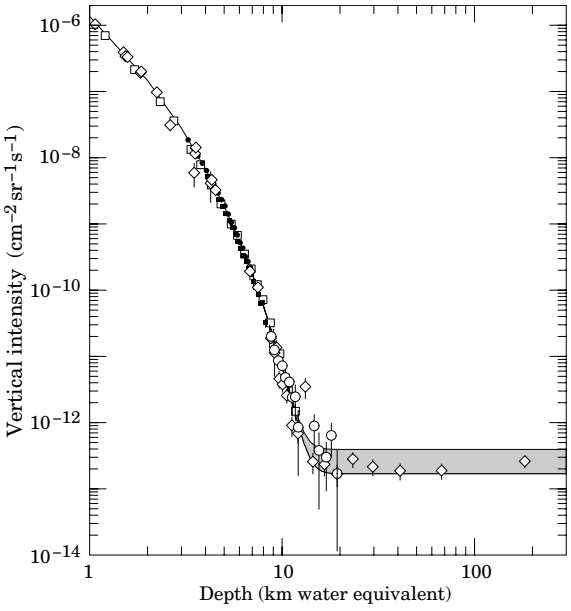

Figure 24.5: Vertical muon intensity vs depth (1 km.w.e. = 10^5 g cm^{-2} of standard rock). The experimental data are from: ◇: the compilations of Crouch [40], □: Baksan [41], ○: LVD [42], •: MACRO [43], ■: Frejus [44]. The shaded area at large depths represents neutrino-induced muons of energy above 2 GeV. The upper line is for horizontal neutrino-induced muons, the lower one for vertically upward muons.

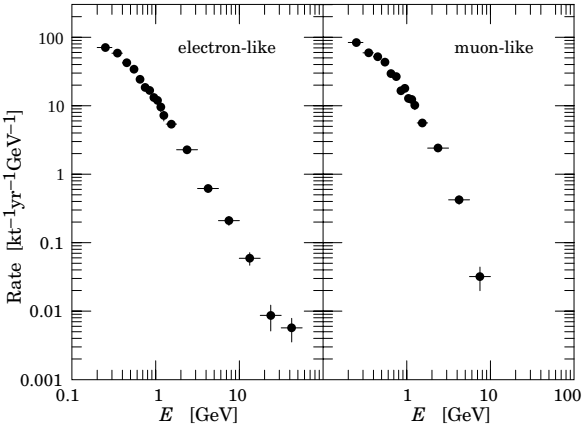

Figure 24.6: Sub-GeV and multi-GeV neutrino interactions from SuperKamiokande [45]. The plot shows the spectra of visible energy in the detector.

GeV as parent muons begin to reach the ground before decaying. Experimental measurements have to account for the ratio of $\overline{\nu}/\nu$, which have cross sections different by a factor of 3 in this energy range. In addition, detectors generally have different efficiencies for detecting muon neutrinos and electron neutrinos which need to be accounted for in comparing measurements with expectation. Fig. 24.6 shows the distributions of the visible energy in the Super-Kamiokande detector [45] for electron-like and muon-like charged current neutrino interactions. Contrary to expectation, the numbers of the two classes of events are similar rather than different by a factor of two. The exposure for the data sample shown here is 50 kiloton-years. The falloff of the muon-like events at high energy is a consequence of the poor containment for high energy muons. Corrections for detection efficiencies and backgrounds are, however, insufficient to account for the large difference from the expectation.

Two well-understood properties of atmospheric cosmic rays provide a standard for comparison of the measurements of atmospheric neutrinos. These are the "secθ effect" and the "east-west effect". The former refers originally to the enhancement of the flux of > 10 GeV muons (and neutrinos) at large zenith angles because the parent pions propagate more in the low density upper atmosphere where decay is enhanced relative to interaction. For neutrinos from muon decay, the enhancement near the horizontal becomes important for $E_\nu > 1$ GeV and arises mainly from the increased pathlength through the atmosphere for muon decay in flight. Fig. 24.7 from Ref. 46 shows a comparison between measurement and expectation for the zenith angle dependence of multi-GeV electron-like (mostly ν_e) and muon-like (mostly ν_μ) events separately. The ν_e show an enhancement near the horizontal and approximate equality for nearly upward ($\cos\theta \approx -1$) and nearly downward ($\cos\theta \approx 1$) events. There is, however, a very significant deficit of upward ($\cos\theta < 0$) ν_μ events, which have long pathlengths comparable to the radius of the Earth. This pattern has been interpreted as evidence for oscillations involving muon neutrinos [45]. (See the article on neutrino properties in this *Review.*) Including three dimensional effects in the calculation of atmospheric neutrinos may change somewhat the expected angular distributions of neutrinos at low energy [47], but it does not change the fundamental expectation of up-down symmetry, which is the basis of the evidence for oscillations.

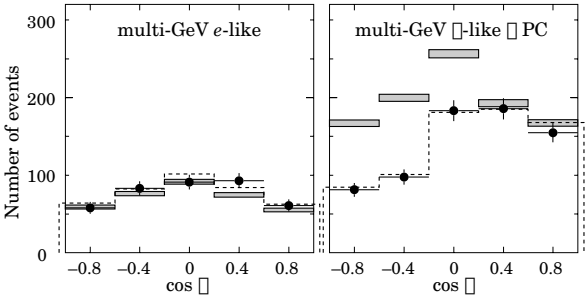

Figure 24.7: Zenith-angle dependence of multi-GeV neutrino interactions from SuperKamiokande [46]. The shaded boxes show the expectation in the absence of any oscillations. The lines show fits with some assumed oscillation parameters, as described in Ref. 46.

The east-west effect [48,49] is the enhancement, especially at low geomagnetic latitudes, of cosmic rays incident on the atmosphere from the west as compared to those from the east. This is a consequence of the fact that the cosmic rays are postively charged nuclei which are bent systematically in one sense in the geomagnetic field. Not all trajectories can reach the atmosphere from outside the geomagnetic field. The standard procedure to see which trajectories are allowed is to inject antiprotons outward from near the top of the atmosphere in various directions and see if they escape from the geomagnetic field without becoming trapped indefinitely or intersecting the surface of the Earth. Any direction in which an antiproton of a given momentum can escape is an allowed direction from which a proton of the opposite momentum can arrive. Since the geomagnetic field is oriented from south to north in the equatorial region, antiprotons injected toward the east are bent back towards the Earth. Thus there is a range of momenta and zenith angles for which positive particles cannot arrive from the east but can arrive from the west. This east-west asymmetry of the incident cosmic rays induces a similar asymmetry on the secondaries, including neutrinos. Since this is an azimuthal effect, the resulting asymmetry is independent of possible oscillations, which depend on pathlength (equivalently zenith angle), but not on azimuth. Fig. 24.8 (from Ref. 50) is a comparison of data and expectation for this effect, which serves as a consistency check of the measurement and analysis.

Muons that enter the detector from outside after production in charged-current interactions of neutrinos naturally reflect a higher

Table 24.3: Measured fluxes (10^{-13} cm^{-2} s^{-1} sr^{-1}) of neutrino-induced muons as a function of the effective minimum muon energy E_μ.

$E_\mu >$	1 GeV	1 GeV	1 GeV	2 GeV	3 GeV	3 GeV
Ref.	CWI [51]	Baksan [52]	MACRO [53]	IMB [54]	Kam [55]	SuperK [56]
F_μ	2.17±0.21	2.77±0.17	2.29 ± 0.15	2.26±0.11	1.94±0.12	1.74±0.07

energy portion of the neutrino spectrum than contained events because the muon range increases with energy as well as the cross section. The relevant energy range is $\sim 10 < E_\nu < 1000$ GeV, depending somewhat on angle. Neutrinos in this energy range show a $\sec\theta$ effect similar to muons (see Eq. (24.5)). This causes the flux of horizontal neutrino-induced muons to be approximately a factor two higher than the vertically upward flux. The upper and lower edges of the horizontal shaded region in Fig. 24.5 correspond to horizontal and vertical intensities of neutrino-induced muons. Table 24.3 gives the measured fluxes of upward-moving neutrino-induced muons averaged over the lower hemisphere. Generally the definition of minimum muon energy depends on where it passes through the detector. The tabulated effective minimum energy estimates the average over various accepted trajectories.

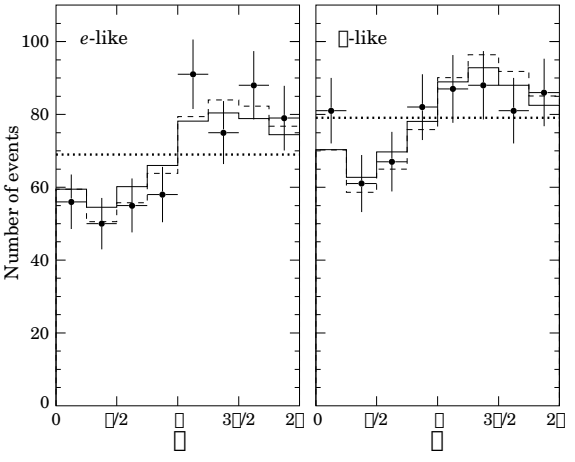

Figure 24.8: Azimuthal dependence of ∼GeV neutrino interactions from SuperKamiokande [50]. The cardinal points of the compass are S, E, N, W starting at 0. These are the direction from which the particles arrive. The lines show the expectation based on two different calculations, as described in Ref. 50.

24.5. Air showers

So far we have discussed inclusive or uncorrelated fluxes of various components of the cosmic radiation. An air shower is caused by a single cosmic ray with energy high enough for its cascade to be detectable at the ground. The shower has a hadronic core, which acts as a collimated source of electromagnetic subshowers, generated mostly from $\pi^0 \to \gamma\gamma$. The resulting electrons and positrons are the most numerous particles in the shower. The number of muons, produced by decays of charged mesons, is an order of magnitude lower.

Air showers spread over a large area on the ground, and arrays of detectors operated for long times are useful for studying cosmic rays with primary energy $E_0 > 100$ TeV, where the low flux makes measurements with small detectors in balloons and satellites difficult.

Greisen [57] gives the following approximate expressions for the numbers and lateral distributions of particles in showers at ground level. The total number of muons N_μ with energies above 1 GeV is

$$N_\mu(>1\ \mathrm{GeV}) \;\approx\; 0.95\times10^5\left(N_e/10^6\right)^{3/4}\ , \qquad (24.9)$$

where N_e is the total number of charged particles in the shower (not just $e^\pm$). The number of muons per square meter, ρ_μ, as a function of the lateral distance r (in meters) from the center of the shower is

$$\rho_\mu = \frac{1.25\,N_\mu}{2\pi\ \Gamma(1.25)}\left(\frac{1}{320}\right)^{1.25} r^{-0.75}\left(1+\frac{r}{320}\right)^{-2.5}\ , \qquad (24.10)$$

where Γ is the gamma function. The number density of charged particles is

$$\rho_e = C_1(s,d,C_2)\,x^{(s-2)}(1+x)^{(s-4.5)}(1+C_2x^d)\ . \qquad (24.11)$$

Here s, d, and C_2 are parameters in terms of which the overall normalization constant $C_1(s,d,C_2)$ is given by

$$C_1(s,d,C_2) = \frac{N_e}{2\pi r_1^2}\,[\,B(s,4.5-2s) + C_2\,B(s+d,\,4.5-d-2s)]^{-1}\ , \qquad (24.12)$$

where $B(m,n)$ is the beta function. The values of the parameters depend on shower size (N_e), depth in the atmosphere, identity of the primary nucleus, etc. For showers with $N_e \approx 10^6$ at sea level, Greisen uses $s = 1.25$, $d = 1$, and $C_2 = 0.088$. Finally, x is r/r_1, where r_1 is the Molière radius, which depends on the density of the atmosphere and hence on the altitude at which showers are detected. At sea level $r_1 \approx 78$ m. It increases with altitude.

The lateral spread of a shower is determined largely by Coulomb scattering of the many low-energy electrons and is characterized by the Molière radius. The lateral spread of the muons (ρ_μ) is larger and depends on the transverse momenta of the muons at production as well as multiple scattering.

There are large fluctuations in development from shower to shower, even for showers of the same energy and primary mass—especially for small showers, which are usually well past maximum development when observed at the ground. Thus the shower size N_e and primary energy E_0 are only related in an average sense, and even this relation depends on depth in the atmosphere. One estimate of the relation is [58]

$$E_0 \;\sim\; 3.9\times10^6\ \mathrm{GeV}\ (N_e/10^6)^{0.9} \qquad (24.13)$$

for vertical showers with $10^{14} < E < 10^{17}$ eV at 920 g cm^{-2} (965 m above sea level). Because of fluctuations, N_e as a function of E_0 is not the inverse of Eq. (24.13). As E_0 increases the shower maximum (on average) moves down into the atmosphere and the relation between N_e and E_0 changes. At the maximum of shower development, there are approximately 2/3 particles per GeV of primary energy.

Detailed simulations and cross-calibrations between different types of detectors are necessary to establish the primary energy spectrum from air-shower experiments [58,59]. Figure 24.9 shows the "all-particle" spectrum. The differential energy spectrum has been multiplied by $E^{2.7}$ in order to display the features of the steep spectrum that are otherwise difficult to discern. The steepening that occurs between 10^{15} and 10^{16} eV is known as the *knee* of the spectrum. The feature around 10^{19} eV is called the *ankle* of the spectrum.

Measurements with small air shower experiments in the knee region differ by as much as a factor of two, indicative of systematic uncertainties in interpretation of the data. (For a recent review see Ref. 60.) In establishing the spectrum shown in Fig. 24.9, efforts have been made to minimize the dependence of the analysis on the primary composition. Ref. 61 uses an unfolding procedure to obtain the spectra of the individual components, giving a result for the all-particle spectrum between 10^{15} and 10^{17} eV that lies toward the

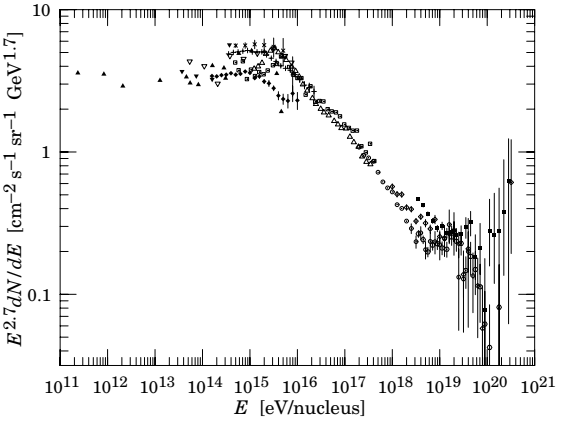

Figure 24.9: The all-particle spectrum: □ [58], ▲ [62], ▼ [63], ▽ [64], △ [65], + [66], × [67], ♦ [68]. References for the high energy portion of the spectrum are given in Fig. 24.10.

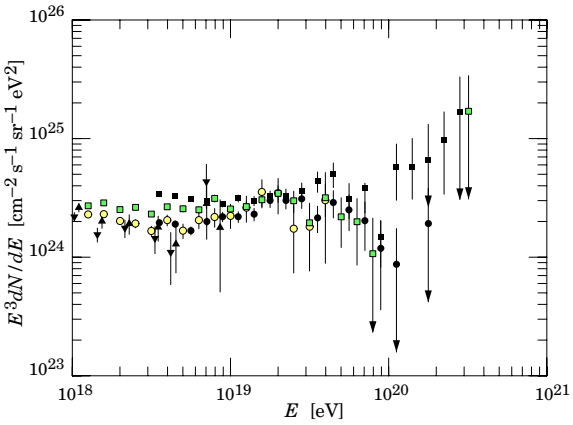

Figure 24.10: Expanded view of the highest energy portion of the cosmic-ray spectrum: o [69] (stereo), □ [69] (monocular) ■ [78], • [79](Hi-Res monocular, preliminary), ▲ [80](if protons) ▼ [80](if iron).

upper range of the data shown in Fig. 24.9. In the energy range above 10^{17} eV, the Fly's Eye technique [69] is particularly useful because it can establish the primary energy in a model-independent way by observing most of the longitudinal development of each shower, from which E_0 is obtained by integrating the energy deposition in the atmosphere.

If the cosmic ray spectrum below 10^{18} eV is of galactic origin, the *knee* could reflect the fact that some (but not all) cosmic accelerators have reached their maximum energy. Some types of expanding supernova remnants, for example, are estimated not to be able to accelerate particles above energies in the range of 10^{15} eV total energy per particle. Effects of propagation and confinement in the galaxy [70] also need to be considered.

The *ankle* has the classical characteristic shape [71] of a higher energy population of particles overtaking a lower energy population. A possible interpretation is that the higher energy population represents cosmic rays of extragalactic origin. If this is the case and if the cosmic rays are cosmological in origin, then there should be a cutoff around 5×10^{19} eV, resulting from interactions with the microwave background [72,73]. It is therefore of special interest that several events have been assigned energies above 10^{20} eV [74–77].

Figure 24.10 gives an expanded view of the high energy end of the spectrum. The differential flux is multiplied by E^3, a procedure that amplifies small systematic differences in energy assignments into sizeable differences in rate. Current discussion focuses on systematic effects with small data samples from a steep spectrum. At issue is whether or not the spectrum of the highest energy cosmic rays indeed continues well past the cutoff expected for a cosmological distribution of sources. If it does, the implication would be that some sources of the highest energy particles must be relatively nearby. For example, the attenuation length for protons at 2×10^{20} eV is 30 Mpc [81]. Both cosmic accelerators ("bottom-up") and massive cosmological relics ("top-down") have been suggested [82].

References:

1. J.A. Simpson, Ann. Rev. Nucl. and Part. Sci. **33**, 323 (1983).
2. M. Boezio *et al.*, Astrophys. J. **518**, 457 (1999).
3. W. Menn *et al.*, Astrophys. J. **533**, 281 (2000).
4. R. Bellotti *et al.*, Phys. Rev. **D60**, 052002 (1999).
5. AMS Collaboration, Phys. Lett. **B490**, 27 (2000).
6. T. Sanuki *et al.*, Astrophys. J. **545**, 1135 (2000).
7. J.J. Engelmann *et al.*, Astron. & Astrophys. **233**, 96 (1990); See also *Cosmic Abundances of Matter* (ed. C. Jake Waddington) A.I.P. Conf. Proceedings No. 183 (1988) p. 111.
8. S.W. Barwick *et al.*, Astrophys. J. **498**, 779 (1998).
9. M. Boezio *et al.*, Astrophys. J. **552**, 635 (2000).
10. J. Alcaraz *et al.*, Phys. Lett. **B484**, 10 (2000).
11. J.M. Clem & P.A. Evenson, Astrophys. J. **568**, 216 (2002).
12. M.A. DuVernois *et al.*, Astrophys. J. **559**, 296 (2000).
13. M. Hof *et al.*, Astrophys. J. **467**, L33 (1996) See also G. Basini *et al.*, *Proc. 26th Int. Cosmic Ray Conf.*, Salt Lake City, **3**, 101 (1999).
14. A.S. Beach *et al.*, Phys. Rev. Lett. **87**, 271101 (2001).
15. J.W. Mitchell *et al.*, Phys. Rev. Lett. **76**, 3057 (1996).
16. M. Boezio *et al.*, Astrophys. J. **487**, 415 (1997).
17. S. Orito *et al.*, Phys. Rev. Lett. **84**, 1078 (2000).
18. Y. Asaoka *et al.*, Phys. Rev. Lett. **88**, 51101 (2002).
19. R. Bellotti *et al.*, Phys. Rev. **D53**, 35 (1996).
20. M. Boezio *et al.*, Phys. Rev. **D62**, 032007 (2000).
21. S. Coutu *et al.*, Phys. Rev. **D62**, 032001 (2000).
22. D.H. Perkins, Astropart. Phys. **2**, 249 (1994).
23. T.K. Gaisser, *Cosmic Rays and Particle Physics*, Cambridge University Press (1990).
24. M.P. De Pascale *et al.*, J. Geophys. Res. **98**, 3501 (1993).
25. P.K.F. Grieder, *Cosmic Rays at Earth*, Elsevier Science (2001).
26. J. Kremer *et al.*, Phys. Rev. Lett. **83**, 4241 (1999).
27. M. Motoki *et al.*, Proc. 27th Int. Cosmic Ray Conf. (Hamburg) **3**, 927 (2001).
28. P. Le Coultre on behalf of the L3 Collaboration, Proc. 27th Int. Cosmic Ray Conf. (Hamburg) **3**, 974 (2001)..
29. O.C. Allkofer, K. Carstensen, and W.D. Dau, Phys. Lett. **B36**, 425 (1971).
30. B.C. Rastin, J. Phys. **G10**, 1609 (1984).
31. C.A. Ayre *et al.*, J. Phys. **G1**, 584 (1975).
32. H. Jokisch *et al.*, Phys. Rev. **D19**, 1368 (1979).
33. C.G.S. Costa, Astropart. Phys. **16**, 193 (2001).
34. S. Hayakawa, *Cosmic Ray Physics*, Wiley, Interscience, New York (1969).
35. R.R. Daniel and S.A. Stephens, Revs. Geophysics & Space Sci. **12**, 233 (1974).
36. K.P. Beuermann and G. Wibberenz, Can. J. Phys. **46**, S1034 (1968).
37. I.S. Diggory *et al.*, J. Phys. **A7**, 741 (1974).
38. D.E. Groom, N.V. Mokhov, and S.I. Striganov, "Muon stopping-power and range tables," Atomic Data and Nuclear Data Tables, 78, 183 (2001).
39. P. Lipari and T. Stanev, Phys. Rev. **D44**, 3543 (1991).

40. M. Crouch, in *Proc. 20th Int. Cosmic Ray Conf.*, Moscow, **6**, 165 (1987).
41. Yu.M. Andreev, V.I. Gurentzov, and I.M. Kogai, in *Proc. 20th Int. Cosmic Ray Conf.*, Moscow, **6**, 200 (1987).
42. M. Aglietta *et al.*, (LVD Collaboration), Astropart. Phys. **3**, 311 (1995).
43. M. Ambrosio *et al.*, (MACRO Collaboration), Phys. Rev. **D52**, 3793 (1995).
44. Ch. Berger *et al.*, (Frejus Collaboration), Phys. Rev. **D40**, 2163 (1989).
45. Y. Fukuda *et al.*, Phys. Rev. Lett. **81**, 1562 (1998).
46. K. Scholberg *et al.*, in *Proc. Eighth Int. Workshop on Neutrino Telescopes* (ed. Milla Baldo Ceolin, edizioni papergraf, Venezia, 1999) p. 183.
47. G. Battistoni *et al.*, Astropart. Phys. **12**, 315 (2000).
48. T.H. Johnson, Phys. Rev. **43**, 834 (1933). See also T.H. Johnson & E.C. Street, Phys. Rev. **44**, 125 (1933).
49. L.W. Alvarez and A.H. Compton, Phys. Rev. **43**, 835 (1933).
50. T. Futagami *et al.*, Phys. Rev. Lett. **82**, 5194 (1999).
51. F. Reines *et al.*, Phys. Rev. Lett. **15**, 429 (1965).
52. M.M. Boliev *et al.*, in *Proceedings 3rd Int. Workshop on Neutrino Telescopes*(ed. Milla Baldo Ceolin), 235 (1991).
53. M. Ambrosio *et al.*, (MACRO) Phys. Lett. **B434**, 451 (1998). The number quoted for MACRO is the average over 90% of the lower hemisphere, $\cos\theta < -0.1$; see F. Ronga *et al.*, `hep-ex/9905025`.
54. R. Becker-Szendy *et al.*, Phys. Rev. Lett. **69**, 1010 (1992); *Proc. 25th Int. Conf. High-Energy Physics*, Singapore (ed. K.K. Phua and Y. Yamaguchi, World Scientific, 1991) p. 662.
55. S. Hatakeyama *et al.*, Phys. Rev. Lett. **81**, 2016 (1998).
56. Y. Fukuda *et al.*, Phys. Rev. Lett. **82**, 2644 (1999).
57. K. Greisen, Ann. Rev. Nucl. Sci. **10**, 63 (1960).
58. M. Nagano *et al.*, J. Phys. **G10**, 1295 (1984).
59. M. Teshima *et al.*, J. Phys. **G12**, 1097 (1986).
60. S.P. Swordy *et al.*, astro-ph/0202159v1, Astropart. Phys. (to be published).
61. H. Ulrich *et al.*, *Proc. 27th Int. Cosmic Ray Conf.*, Hamburg, **1**, 97 (2001).
62. N.L. Grigorov *et al.*, Yad. Fiz. **11**, 1058 (1970) and *Proc. 12th Int. Cosmic Ray Conf.*, Hobart, **2**, 206 (1971).
63. K. Asakimori *et al.*, *Proc. 23rd Int. Cosmic Ray Conf.*, Calgary, **2**, 25 (1993); *Proc. 22nd Int. Cosmic Ray Conf.*, Dublin, **2**, 57 and 97 (1991).
64. T.V. Danilova *et al.*, *Proc. 15th Int. Cosmic Ray Conf.*, Plovdiv, **8**, 129 (1977).
65. Yu. A. Fomin *et al.*, *Proc. 22nd Int. Cosmic Ray Conf.*, Dublin, **2**, 85 (1991).
66. M. Amenomori *et al.*, Astrophys. J. **461**, 408 (1996).
67. F. Arqueros *et al.*, Astron. & Astrophys. **359**, 682 (2000).
68. M.A.K. Glasmacher *et al.*, Astropart. Phys. **10**, 291 (1999).
69. D.J. Bird *et al.*, Astrophys. J. **424**, 491 (1994).
70. V.S. Ptuskin *et al.*, Astron. & Astrophys. **268**, 726 (1993).
71. B. Peters, Nuovo Cimento **22**, 800 (1961).
72. K. Greisen, Phys. Rev. Lett. **16**, 748 (1966).
73. G.T. Zatsepin and V.A. Kuz'min, Sov. Phys. JETP Lett. **4**, 78 (1966).
74. J. Linsley, Phys. Rev. Lett. **10**, 146 (1963).
75. D.J. Bird *et al.*, Astrophys. J. **441**, 144 (1995).
76. N. Hayashima *et al.*, Phys. Rev. Lett. **73**, 3941 (1994).
77. M. Nagano & A.A. Watson, Rev. Mod. Phys. **72**, 689 (2000).
78. M. Takeda *et al.*, Phys. Rev. Lett. **81**, 1163 (1998) (as updated by M. Teshima, private communication).
79. P. Sokolsky, private communication.
80. M. Ave *et al.*, `astro-ph/0112253`.
81. V.S. Berezinskii *et al.*, *Astrophysics of Cosmic Rays*, North-Holland (1990).
82. P. Bhattacharjee & G. Sigl, Phys. Reports **327**, 109 (2000).