General Relativity and Cosmology : JAP 2000

Lecture notes for cosmology: Biman B. Nath

Raman Research Institute, biman@rri.ernet.in

1 Cosmological observations

Like all other fields of science, the history of cosmology is full of zigzag steps, with intricate feedbacks between theory and observations. In this course, I would not like to go through the historical development and would try to present things as we know them now. Those who are interested in the history, could look up Peebles's 'Principle of physical cosmology'.

There are a few important observations which lead us to make some assumptions for our universe and which then enable us to study it in the framework of general relativity (GR). These are mainly the observations of (1) recession velocities, (2) number counts and (3) the cosmic microwave background radiation (CMBR).

1.1 Recession velocities

Soon after the discovery that our Milky Way is a galaxy of stars, and that there are many other galaxies in the universe, far away from us, astronomers began looking at the spectra of galaxies. Edwin Hubble then announced in a paper in 1929 that he found a linear relation between the radial velocity (as derived from the redshift) and distances of 24 galaxies. He had of course subtracted the contribution of the solar motion to the radial velocity (see Fig.1). His value of the constant K in $v = Kr$, of 500 km/sec/Mpc was too large though, as his distances were small. We call this constant K the Hubble constant and denote by H_0. This observation is borne out by modern results too.

What one plots in actuality (since one doesn't really know the distance all the time) is the velocity v and the magnitude m. Now, since $m - M = 5(\log_{10} r_{Mpc}) + 25$, one expects from the Hubble relation (where B is a constant),

$$\log v = 0.2(m - M) + B. \tag{1}$$

The predicted slope of 0.2 has been now verified for sources at very large distances. There are two difficulties in testing this accurately. Firstly, if galaxies have some extra velocities in addition to the Hubble radial velocity, then one cannot accurately test the relation. We now know that galaxies do have what is called 'peculiar velocities' (velocities with respect to Hubble radial velocity) of order ± 500 km/sec. So, we would have to look at sources at large distances so that Hubble velocity is much larger than 500 km/sec. Secondly, one needs to know the absolute magnitude M, that is one needs sources with standard luminosities, and which are bright enough to be seen from a large distance. A favourite standard candle is Type I supernovae, which are very bright, and the scatter in their standard luminosity is very small. One study by Reiss, Press & Kirshner (1996) find a slope of 0.02010 ± 0.0035.

The value of the Hubble constant is more difficult to determine, because one needs to know the actual distances accurately. This has been (one of) the Holy Grails of observational cosmology and a matter of furious debate for many decades. It looks like now that the measured values by different methods are converging to a value of $H_0 = 70 \pm 7$ km/sec/Mpc. By the way, in metric units, this is around $\sim 4.4 \times 10^{-17}$ sec$^{-1} \sim \frac{1}{14\,\mathrm{Gyr}}$.

Let us now look at the Hubble relation abit more carefully. The velocity field $\vec{v} = H\vec{r}$ has a very interesting property, that it is invariant under translation, and of course under rotation. Consider a galaxy 1 at position r_1. It would have a recession velocity of $\vec{v}_1 = H\vec{r}_1$ with respect to us. Similarly, the galaxy 2 at $\vec{r}_2$ would have a velocity $\vec{v}_2 = H\vec{r}_2$. For observers in galaxy 1, the galaxy 2 would be at a distance $\vec{r}_2 - \vec{r}_1$, and its velocity would be $\vec{v}_2 - \vec{v}_1$, and it is easy to see that the Hubble relation is valid for observers in galaxy 1 as well ($\vec{v}_2 - \vec{v}_1 = H(\vec{r}_2 - \vec{r}_1)$. This means that galaxies are not receding only from us, but from one another.

1.2 Number counts

One way to find the distribution of galaxies in space is to find the number of sources brighter than a given flux per unit solid angle, $N(> f)$, where f is the limiting flux. We know that $f = L/(4\pi r^2)$ where L is the luminosity. Suppose all galaxies have the same luminosity L, then all galaxies brighter than $f = L/(4\pi r^2)$ would be closer than r. The volume of space in one steradian of the sky out to distance r is $V = r^3/3$. So, if galaxies were distributed homogeneously, with mean number density n, then the number of galaxies per steradian brighter than f would be (equal to number of galaxies within r),

$$N(> f) = nV = n\frac{r^3}{3} = \frac{n}{3}\left(\frac{L}{4\pi f}\right)^{3/2} . \tag{2}$$

Actually, the luminosities have a distribution, so this equation applies to a class of galaxies with luminosity L_i with mean number density n_i. Summing over all types of galaxies, we have,

$$N(> f) = \frac{1}{3}\sum_i n_i \left(\frac{L_i}{4\pi f}\right)^{3/2} \propto f^{-3/2} . \tag{3}$$

Now, since we have the apparent magnitude $m = -2.5 \log f +$ constant, or, equivalently, $f \propto 10^{-0.4m}$, we have,

$$N(< m) \propto 10^{0.6m} . \tag{4}$$

Hubble had announced as early as 1926 that the galaxy distribution followed this law, down to a blue magnitude of 17, corresponding to a distance about ten times that of Virgo cluster. In the next decade, he pushed it down to $m = 21$. One modern example of this is the study by Shanks (1991) one can see that the above relation is valid. The bump around $b_J \sim 13$ (in the Durham system, for magnitude measured in the band $\lambda \sim 4000$–5500 Å) is due to the local concentration in and around Virgo cluster. One also sees that at the faint end there is a slight deviation. The problem is that, as we will soon see, for very faint galaxies which are on average at large distances, the light is shifted to red, and one needs to know this correction to be able to determine the magnitude, and for which one needs accurate determination of redshift, which is difficult. Also, at that distance (or so far back in time) it is possible that galaxies were young and did not look like galaxies today; moreover one has to take into account the possible evolution of the spacetime. All these problems make cosmology a difficult and a challenging subject.

Galaxy surveys like the Las Campanas survey has been used to find the clumping of matter as a function of the lengthscale in the universe, which shows the deviations from homogeneity. These studies show that at small lengthscales the universe is certainly inhomogeneous, but the deviation, characterised by $\delta\rho/\rho$ where ρ is the mean density, decreases rapidly with increasing length scales. In other words if you average the density over a large region you are going to be close to the mean density. At length scales more than 100 Mpc, $\delta\rho/\rho \leq 10^{-2}$. There has been claims that the distribution of galaxies in the universe is like a fractal. But careful studies show that (e.g., the review by Lahav et al.1998) the 'dimension' of the distribution of galaxies approaches a value of 3, equivalent to a homogeneous distribution, when one looks at a large volume of space.

So, we see that there is overwhelming evidence for a homogeneous distribution of galaxies. Also, it is isotropic, as we do not know of any difference of number counts in different lines of sight. By the way, if the universe is isotropic at every point, it is bound to be homogeneous.

1.3 CMBR

One of the most important observations in cosmology is that of the cosmic microwave background radiation (CMBR). Penzias and Wilson discovered in 1964 that there was an excess of radio noise in the sky, which they could not attribute to any known phenomenon, and which was equivalent to a temperature 3.7 ± 1 K (which means that a blackbody with the same temperature would cause the noise). Incidentally, there was a hint of such a radiation as early as 1939 from the excitation of the interstellar cyanogen radical CN into its first excited state, and the radiation was supposed to have a temperature of 2.3 ± 1 K. But it was forgotten soon somehow (this is how science progresses!). Well, after the discovery by Penzias and Wilson, the measurement has been refined, especially by the COBE satellite launched in 1990, and the temperature is now known to be $T = 2.728 \pm 0.002$ K, in the wavelength region of 0.5 to 5 mm.

This radiation is also known to be incredibly isotropic. There is a dipole due to the motion of the Sun. When one takes that out, the resulting isotropy is better than 10 parts in a million, i.e., the temperature is constant to that accuracy across the sky.

2 Cosmological principle

These observations lead us to enunciate the so-called cosmological principle, which states that our universe is homogeneous and isotropic. We also know that the universe is expanding in time. So, the homogeneity we talk about means that the universe is homogeneous only on slices through space-time with constant cosmic time. We would have to be careful here about the definition of homogeneity. From SR we know that the universe would look different for observers moving with different velocities. So, we define a set of observers, called the comoving observers, who will see the Hubble law for galaxy redshifts in its simple form. These observers can determine the local density ρ of matter at time t. Then the cosmological principle would mean that ρ is a function only of t and does not depend on the location of the observer.

2.1 A metric for the universe

We can now use the FRW metric we derived sometime ago on the basis of the cosmological principle,

$$ds^2 = -dt^2 + R^2(t)\left[\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\right]. \tag{5}$$

2.2 Geometry

Consider first the case of $k = 0$. At any moment t_0, the line element with $dt = 0$, is,

$$\begin{aligned} dl^2 &= R_0^2[dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)] \\ &= dx^2 + dy^2 + dz^2 \end{aligned} \tag{6}$$

where,

$$\begin{aligned} x &= R_0 r \sin\theta\cos\phi, \\ y &= R_0 r \sin\theta\sin\phi, \\ z &= R_0 r \cos\theta. \end{aligned} \tag{7}$$

Obviously this is the metric for flat 3-dimensional Euclidean space. We would refer to as the flat metric and call the corresponding universe the **flat** universe.

Consider next, $k = +1$. We define a new coordinate $\chi(r)$ such that,

$$d\chi^2 = \frac{dr^2}{1-r^2}, \tag{8}$$

Then we have $r = \sin\chi$, and for the line element at $t = t_0$,

$$\begin{aligned} dl^2 &= R_0^2[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2)] \\ &= dw^2 + dz^2 + dy^2 + dz^2, \end{aligned} \tag{9}$$

where,

$$\begin{aligned} w &= R_0 \cos\chi, \\ x &= R_0 r \sin\chi\sin\theta\cos\phi, \\ y &= R_0 r \sin\chi\sin\theta\sin\phi, \\ z &= R_0 r \sin\chi\cos\theta. \end{aligned} \tag{10}$$

These equations imply that

$$w^2 + x^2 + y^2 + z^2 = R_0^2, \tag{11}$$

so that the 3-surface is a 3-dimensional sphere in 4-dimensional Euclidean space (Fig. ? shows the geometry with $\phi = 0 = y$). The surface is defined by the coordinate range, $0 \le \chi \le \pi$, $0 \le \theta \le \pi$, and $0 \le \phi < 2\pi$. The 2-surfaces with $\chi =$ constant, which appear as circles in the figure, are 2-spheres of surface area,

$$A_\chi = \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}(R_0\sin\chi d\theta)(R_0\sin\chi\sin\theta d\phi) = 4\pi R_0^2\sin^2\chi. \tag{12}$$

(Recall that for the case when the line element is $dl^2 = a^2(d\theta^2 + \sin^2\theta d\phi^2)$, the surface area is $A = \int\int a^2 \sin\theta d\theta d\phi$.) Here (θ, ϕ) are the usual polar coordinates of these 2-spheres. This area is zero at the North Pole, increases to a maximum at the equator and then decreases to zero again at the South Pole. The volume of this surface is given by,

$$V = \int_{\chi=0}^{\pi}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}(R_0 d\chi)(R_0 \sin\chi d\theta)(R_0 \sin\chi \sin\theta d\phi) = 2\pi^2 R_0^3 = 2\pi^2 R^3(t_0)\,. \quad (13)$$

This 3-space is a generalization of a 2-sphere to a 3-dimensional entity, and is called the 3-sphere. It is clearly bounded and the corresponding universe is called a **closed** universe.

For $k = -1$, one can show that the line element can be written as, (with $r = \sinh\chi$)

$$\begin{aligned} dl^2 &= R_0^2[d\chi^2 + \sinh^2\chi(d\theta^2 + \sin^2\theta d\phi^2)] \\ &= -dw^2 + dz^2 + dy^2 + dz^2\,, \end{aligned} \quad (14)$$

where,

$$\begin{aligned} w &= R_0 \cosh\chi, \\ x &= R_0 r \sinh\chi \sin\theta \cos\phi, \\ y &= R_0 r \sinh\chi \sin\theta \sin\phi, \\ z &= R_0 r \sinh\chi \cos\theta\,. \end{aligned} \quad (15)$$

These equations imply that

$$w^2 - x^2 - y^2 - z^2 = R_0^2\,, \quad (16)$$

so that the 3-surface is a 3-dimensional hyperboloid in 4-dimensional Euclidean space (Fig. ? shows the geometry with $\phi = 0 = y$). The surface is defined by the coordinate range, $0 \le \chi < \infty$, $0 \le \theta \le \pi$, and $0 \le \phi < 2\pi$. The surface area of the 2-surfaces with $\chi =$ constant is equal to $A_\chi = 4\pi R_0^2 \sinh^2\chi$, so that the surface area increase to infinity with χ. This is an **open** universe.

3 Dynamics of the universe

We can now use Einstein's equations to determine how the universes with different geometries behave. We idealize the universe as filled with a perfect fluid (that is the stress-energy tensor is diagonal), with

$$T_{\mu\nu} = (\rho + p)U_\mu U_\nu + p g_{\mu\nu}\,. \tag{17}$$

Since $U^\mu = (1,0,0,0)$ and $U_\mu = (-1,0,0,0)$, we have, $T_{00} = \rho$, $T_{rr} = pR^2/(1-kr^2)$, $T_{\theta\theta} = pR^2 r^2$, $T_{\phi\phi} = pR^2 r^2 \sin^2\theta$, where we have used the fact that, $g_{00} = -1$, $g_{rr} = R^2/(1-kr^2)$, $g_{\theta\theta} = R^2 r^2$, $g_{\phi\phi} = R^2 r^2 \sin^2\theta$.

The nonzero Christoffel symbols are,

$$\begin{array}{lll} \Gamma^0_{rr} = \frac{R\dot{R}}{1-kr^2} & \Gamma^0_{\theta\theta} = R\dot{R}r^2 & \Gamma^0_{\phi\phi} = R\dot{R}r^2\sin^2\theta \\ \Gamma^r_{rr} = \frac{kr}{1-kr^2} & \Gamma^r_{\theta\theta} = -r(1-kr^2) & \Gamma^r_{\phi\phi} = -r\sin^2\theta(1-kr^2) \\ \Gamma^r_{0r} = \frac{\dot{R}}{R} & \Gamma^\theta_{0\theta} = \frac{\dot{R}}{R} & \Gamma^\phi_{0\phi} = \frac{\dot{R}}{R} \\ \Gamma^\theta_{r\theta} = \Gamma^\phi_{r\phi} = \frac{1}{r} & \Gamma^\phi_{\theta\phi} = \cot\theta & \Gamma^\theta_{\phi\phi} = -\sin\theta\cos\theta \end{array} \tag{18}$$

So that the Ricci tensors are,

$$R^0_0 = \frac{3\ddot{R}}{R}, \qquad R^r_r = R^\theta_\theta = R^\phi_\phi = \frac{2\dot{R}^2}{R^2} + \frac{2k}{R^2} + \frac{\ddot{R}}{R}, \tag{19}$$

and the Ricci scalar is $R = 6\left(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2} + \frac{\ddot{R}}{R}\right)$. The Einstein tensor $G_{00} = \frac{3\dot{R}^2}{R} + \frac{3k}{R^2}$, and $G_{rr} = \frac{-R}{1-kr^2}\left(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2} + \frac{2\ddot{R}}{R}\right)$. So, the time component of the Einstein's equation gives,

$$\frac{3\dot{R}^2}{R^2} + \frac{3k}{R^2} - \Lambda = 8\pi\rho\,, \tag{20}$$

and the space components give,

$$\frac{\dot{R}^2}{R^2} + \frac{k}{R^2} + \frac{2\ddot{R}}{R} - \Lambda = -8\pi p\,. \tag{21}$$

Differentiating the first equation with respect to t, multiplying through by $(1/8\pi)$ and adding the result to the second equation (after multiplying it by $-3\dot{R}/8\pi R$, we get,

$$\dot{\rho} + 3p\frac{\dot{R}}{R} = -\frac{3}{8\pi}\frac{\dot{R}}{R}\left(\frac{3\dot{R}^2}{R^2} + \frac{3k}{R^2} - \Lambda\right) = -3\rho\frac{\dot{R}}{R}, \tag{22}$$

where we have used the first equation again. Multiplying this by R^3 we can write,

$$\frac{d}{dt}(\rho R^3) = -p\frac{d}{dt}(R^3)\,. \tag{23}$$

This is basically the statement of conservation of energy (and can also be derived from $T^{\mu\nu}_{;\nu} = 0$).

In this equation, p includes all types of pressure, e.g., pressure due to random motion of stars and galaxies, radiation pressure, and so forth. But observations tell us that at the present epoch p is much smaller than the energy density ρc^2 due to matter (about a millionth or so). (This ratio is $p/\rho c^2 \sim (v/c)^2 \sim 10^{-6}$, where v is the average random velocity of galaxies, which is of order a few hundred km/sec.) Let us then take $p=0$. Then the second Einstein's equation gives after integration,

$$R(\dot{R}^2+k)-\frac{1}{3}\Lambda R^3 = C\,, \tag{24}$$

where C is a constant of integration. Using the first Einstein's equation we then get $C = \frac{8}{3}\pi R^3\rho$, which is a constant if $p=0$, from the above equation for energy conservation. So, now the first Einstein's equation can be written as,

$$\dot{R}^2 = \frac{C}{R}+\frac{1}{3}\Lambda R^2 - k\,, \tag{25}$$

which is known as the Friedmann's equation. We can use the concept of effective potentials here. We define,

$$\dot{R}^2 = (-k) - V_M(R) \qquad V_M = -C/R\,, \tag{26}$$

where we have put $\Lambda = 0$ for the time being. (See Fig 12.3 in Schutz.) The universe exists only in the regions where $-k$ exceeds $V_M(R)$, so that $\dot{R}^2 > 0$. Since we know that at present $\dot{R} > 0$, there are three possible futures. If $k=-1$, the universe expands to infinity with finite terminal velocity. If $k=0$ then the universe expands to infinity with ever decreasing speed, and if $k=+1$ then it reaches a maximum radius $R=C$, at which it has a turning point and then it recollapses. Notice that all models originate at $R=0$; there is no turning point for small R – this is the Big Bang.

The effect of including Λ is shown in the next plot. For a $k=0$ universe, it can accelerate after a point of time (and recent observations indicate that this might be the case for our universe). $\Lambda < 0$ can make a $k=-1$ universe contract again.

The time derivative of the Friedmann's equation gives us,

$$\dot{R}\ddot{R} = -\frac{C}{2}\frac{\dot{R}}{R^2}+\frac{1}{3}\Lambda R\dot{R} = -\frac{4}{3}\pi\rho\dot{R}R+\frac{1}{3}\Lambda R\dot{R}\,. \tag{27}$$

We define the dimensionless *deceleration parameter*,

$$q \equiv -\frac{R\ddot{R}}{\dot{R}^2}\,, \tag{28}$$

so that one can write,

$$\rho = \frac{3}{4\pi}q(\dot{R}/R)^2+\frac{1}{4\pi}\Lambda\,. \tag{29}$$

If we define the *critical density* ρ_c as,

$$\rho_c = \frac{3}{8\pi}(\dot{R}/R)^2 (\equiv \frac{3}{8\pi G}(\dot{R}/R)^2)\,, \tag{30}$$

and write $\Omega_m = \rho/\rho_c$, and $\Omega_\Lambda = \frac{\Lambda}{3}(R/\dot{R})^2$, then,

$$q = \frac{\Omega_m}{2} - \Omega_\Lambda\,. \tag{31}$$

Also, the Friedmann's equation(25) can be written as,

$$1 = \Omega_k + \Omega_m + \Omega_\Lambda\,, \tag{32}$$

where we wrote $\Omega_k = -k/(\dot{R}^2)$.

One interesting model of the universe is that of de Sitter. This universe is empty ($p = 0 = \rho$) and is flat. So, we have,

$$3\dot{R}^2/R^2 - \Lambda = 0\,, \tag{33}$$

or,

$$\dot{R}/R = \sqrt{\frac{1}{3}\Lambda}\,, \tag{34}$$

which on integration becomes,,

$$R = A\exp[\sqrt{\frac{1}{3}\Lambda t}]\,, \tag{35}$$

where A is a constant. This model is mainly of historical interest, although recently there has been the concept of inflation, in which for a certain interval of time the expansion is exponential as the above equation.

3.1 Propagation of light

Consider the way an observer O receives light from a receding galaxy. Since we assume that the time slices are homogeneous, we can, without any loss of generality, take O to be the origin of coordinates $r = 0$. For a radial null geodesic for a photon ($ds^2 = 0 = d\theta = d\phi$), then, the FRW metric gives,

$$\frac{dt}{R(t)} = \pm\frac{dr}{\sqrt{1-kr^2}}\,, \tag{36}$$

where the $-$ sign corresponds to an approaching light ray. Consider a light ray emanating from a galaxy with world-line at $r = r_1$, at coordinate time t_1, and received by O at coordinate time t_o (at $r = 0$). Then we get,

$$\int_{t_1}^{t_o}\frac{dt}{R(t)} = -\int_{r_1}^{0}\frac{dr}{\sqrt{1-kr^2}} = f(r_1)\,, \tag{37}$$

where $f(r_1) = \arcsin r_1, r_1, \sinh^{-1} r_1$ for $k = +1, 0, -1$ respectively.

Next consider two successive light rays emanating from this galaxy at times $t = t_1$ and $t_1 + dt_1$, and received by O at t_o and $t_o + dt_o$ respectively. Then we have,

$$\int_{t_1+dt_1}^{t_o+dt_o}\frac{dt}{R(t)} = \int_{t_1}^{t_o}\frac{dt}{R(t)}\,, \tag{38}$$

since each side is equal to the same function $f(r_1)$. So, we have,

$$\int_{t_1+dt_1}^{t_o+dt_o}\frac{dt}{R(t)} - \int_{t_1}^{t_o}\frac{dt}{R(t)} = \int_{t_o}^{t_o+dt_o}\frac{dt}{R(t)} - \int_{t_1}^{t_1+dt_1}\frac{dt}{R(t)} = 0\,, \tag{39}$$

and assuming that $R(t)$ does not vary greatly in the intervals dt_o and dt_1, we can take it outside the integrals and write,

$$\frac{dt_o}{R(t_o)} = \frac{dt_1}{R(t_1)}\,. \tag{40}$$

Hence the interval as measured by O is $R(t_o)/R(t_1)$ times the interval measured by the observer in the frame of the receding galaxy. Therefore, the observer O will see a redshift z which is defined by,

$$1+z=\nu_1/\nu_o=R(t_o)/R(t_1)\,, \tag{41}$$

where ν_1 and ν_o are the frequencies measured by the emitter and receiver, respectively. This redshift is called the **Cosmological redshift**. This is not to be confused with Doppler shift though.

If the galaxy in consideration is nearby, then the cosmic time of emission and reception differ by a small amount, dt, say, that is, $t_o=t_1+dt$, and we we have,

$$1+z=\frac{R(t_o)}{R(t_o-dt)}\approx\frac{R(t_o)}{R(t_o)-\dot{R}(t_o)dt}\sim 1+\frac{\dot{R}(t_o)}{R(t_o)}dt\,, \tag{42}$$

to first order in dt. We also have,

$$\int_{t_1}^{t_o}\frac{dt}{R(t)}=\int_{t_1}^{t_1+dt}\frac{dt}{R(t)}\approx\frac{dt}{R(t_1)}=\frac{dt}{R(t_o-dt)}\approx\frac{dt}{R(t_o)}\,. \tag{43}$$

But for small r,

$$\int_{t_1}^{t_o}\frac{dt}{R(t)}=f(r_1)\approx r_1\,, \tag{44}$$

and so, $\frac{dt}{R(t_o)}\sim r_1$. Which means that, for small distances,

$$z\approx\dot{R}(t_o)r_1\sim\frac{\dot{R}(t_o)}{R(t_o)}(R(t_o)r_1)\,. \tag{45}$$

This is basically the Hubble law.

Consider a particle now, which is not necessarily massless. The geodesic equation of motion is then,

$$\frac{dU^\mu}{d\lambda} + \Gamma^\mu_{\nu\alpha} U^\nu \frac{dx^\alpha}{d\lambda} = 0\,, \tag{46}$$

where λ is some parameter on the curve. The $\mu = 0$ component of this equation is (choosing the parameter λ to be the proper length ds),

$$\frac{dU^0}{ds} + \Gamma^0_{\nu\alpha} U^\nu \frac{dx^\alpha}{ds} = 0\,, \tag{47}$$

For the FRW metric, the only non-zero Christoffel symbol of $\Gamma^0_{\nu\alpha}$ is $\Gamma^0_{ij} = (\dot{R}/R)g_{ij}$, and remembering that $g_{ij}U^iU^j = |\vec{u}|^2$, one can write,

$$\frac{dU^0}{ds} + \frac{\dot{R}}{R}|\vec{u}|^2 = 0\,. \tag{48}$$

Since $-(U^0)^2 + |\vec{u}|^2 = -1$, it follows that $U^0 dU^0 = |\vec{u}|d|\vec{u}|$, and the geodesic equation can be written as,

$$\frac{1}{U^0}\frac{d|\vec{u}|}{ds} + \frac{\dot{R}}{R}|\vec{u}| = 0\,. \tag{49}$$

Finally, since $U^0 \equiv dt/ds$, this equation reduces to $|\dot{\vec{u}}|/|\vec{u}| = -\dot{R}/R$, which implies that $|\vec{u}| \propto R^{-1}$. So, the momentum of freely propagating particles 'redshifts' as R^{-1}.

3.2 Cosmological definition of distance

We must have a better definition of distance though. What do we mean by distance? It is not coordinate distance, because one does not know how to measure it. Proper distance is also difficult, because this distance between the emitting galaxy and receiving galaxy is zero, as light travels in null lines. The distance between the emitting galaxy *now* and the receiving galaxy is also difficult – the emitting galaxy may not even exist. Let us then find out how do we actually measure distances in the local universe.

One way is to measure the apparent luminosity f of a source of known luminosity L. The distance is then $\sqrt{L/4\pi f}$. With these considerations in mind, we define the **luminosity distance** as,

$$d_L^2 = \frac{L}{4\pi f}\,. \tag{50}$$

In an expanding universe, however, the the interval of time during which a certain amount of energy is received is longer than the interval of emission by virtue of redshift, and hence the number of photons received per unit time is reduced by a factor of $1+z$. In addition to this, the energy of each photon decreases by a factor $1+z$ due to redshift again.

Consider light emanating from a galaxy P at time t_1, and observed by us 'now' at O at a time $t = t_0\,(t_1 < t_0)$. The light will have spread out over the surface of a sphere with centre at the event P_0 $(t = t_0, r = r_1)$ and passing through the event $O_0(t = t_0, r = 0)$. The surface area of the sphere is the same as that of the sphere centred on O_0 passing through P_0, due to homogeneity. The line element for this sphere $(t = t_0, r = r_1)$ is ,

$$dl^2 = R^2(t_0)r_1^2(d\theta^2 + \sin\theta d\phi^2)\,. \tag{51}$$

Since this is the usual line element for a sphere of radius $R(t_0)r_1$, so the sphere has surface area $4\pi R^2(t_0)r_1^2$. So the observed intensity is,

$$f = \frac{L}{4\pi r_1^2 R^2(t_0)(1+z)^2} \,. \tag{52}$$

So, we can write the luminosity distance as,

$$d_L = r_1 R(t_0)(1+z) \,. \tag{53}$$

We will evaluate it in terms of the redshift z of the galaxy P once we have derived a relationship between z and the age of the universe.

One related topic is that of the angular size of objects. Consider two neighbouring null lines from two end points A and B of an object. We choose the coordinates of A and B as (r_1, θ_1, ϕ_1) and $(r_1, \theta_1 + d\theta_1, \phi_1)$ while we are at the origin. The proper distance between A and B can be obtained by putting $t = t_1$ constant, $r = r_1 =$ constant, $\phi = \phi_1 =$ constant and $d\theta = d\theta_1$, and we get,

$$dl^2 = r_1^2 R^2(t_1) d\theta_1^2 = l^2 \,, \tag{54}$$

where l is the spacelike separation between A and B in the rest frame of the object. So,

$$d\theta_1 = \frac{l}{r_1 R(t_1)} = \frac{l(1+z)}{r_1 R(t_0)} \,. \tag{55}$$

If we define an angular distance as $d\theta = l/d_A$, then one has $d_L = d_A(1+z)^2$.

3.3 K correction

Note that the formula $f = L/(4\pi d_L^2)$ is valid to the bolometric fluxes and luminosities, integrating over all frequencies. If one wants flux or luminosity within a band, one needs to (1) transform the frequency so that one gets the flux F_ν from the luminosity $L_{\nu(1+z)}$, and (2) transform the bandwidth of the observation into the bandwidth of emission. Since the fractional bandwidth does not change with redshift, one has,

$$\nu F_\nu = \frac{\nu(1+z) L_{\nu(1+z)}}{4\pi d_L^2} \,. \tag{56}$$

So,

$$F_\nu = \frac{(1+z) L_{\nu(1+z)}}{4\pi d_L^2} \qquad , F_\lambda = \frac{L_{\nu(1+z)}}{(1+z) 4\pi d_L^2} \,. \tag{57}$$

The difference between $\nu(1+z)L_{\nu(1+z)}$ and νL_ν leads to the so-called K-correction, which is given in terms of magnitudes as (to be added to the apparent magnitude),

$$K(\nu, z) = 2.5 \log\left(\frac{\nu(1+z) L_{\nu(1+z)}}{\nu L_\nu}\right) . \tag{58}$$

For example, in a given band, like V, one writes,

$$m_V = M_V + 5 \log\left(\frac{d_L(z)}{10\mathrm{pc}}\right) + K(\nu_V, z) \,. \tag{59}$$

Earlier the observations were made only in photographic blue, and the correction used be large because the flux drops by a large amount at the 4000 Å edge due to the

Balmer edge in hydrogen and the H and K lines of ionized calcium. With modern data, though, one can use bands in infrared (R or I) to observe galaxies at $z \sim 0.5$ and then compare the fluxes with B and V band data from nearby galaxies, reducing the uncertainty and magnitude of the K-correction.

This assumes though that there is no evolution of galaxies with time. This has been a big hurdle in using high redshift galaxies for determining the cosmological parameters. This is why the recent works using distant SN Ia is important, since the absolute magnitude of the SN Ia depends on the Chandrasekhar mass limit which does not evolve with time, and so evolution does not matter in these studies.

3.4 Number counts revisited

With the expressions for the luminosity and angular distances, we can discuss the number counts observations to high redshift range. Consider a shell between redshifts z and $z+dz$. The physical volume is given by the surface area $4\pi d_A(z)^2$ and the thickness of the shell is $(cdt/dz)dz$. Also assume that the numer density varies as $n(z) = n_0(1+z)^3$, that is the objects are conserved. So, the number in this redshift range is,

$$\frac{dN}{dz} = n_0(1+z)^3 d_A(z)^2 \frac{cdt}{dz}\,, \tag{60}$$

where N is the number of sources with redshift less than z per steradian. One, however, does not have surveys till a certain redshift. In practice, one has data which is complete to a given flux or magnitude. Consider a single class of objects with the same luminosity L and let S be the flux. The general case of a range of luminosities can be easily calculated by summing over. The luminosity distance is $d_L = \sqrt{L/(4\pi S)}$ and the number count is given by,

$$\frac{dN}{ds} = n_0(1+z)^3 \frac{d_L^2}{(1+z)^4} \frac{d(d_L)}{dS} \frac{dz}{d(d_L)} \frac{cdt}{dz}\,. \tag{61}$$

Here, $d(d_L)/dS = -\frac{1}{2}S^{-1.5}\sqrt{L/4\pi}$ and $d_L^2 = L/(4\pi S)$, so that,

$$\frac{dN}{dS} = \frac{n_0(L/4\pi)^{1.5}}{2S^{2.5}} \big((1+z)^{-1} \frac{dz}{d(d_L)} \frac{cdt}{dz}\big)\,. \tag{62}$$

The factor in the brackets involve the cosmological corrections to the Euclidean expression of dN/dS. The total intensity from all sources is,

$$\int SdN = \int S \frac{n_0(L/4\pi)^{1.5}}{2S^{2.5}} \big((1+z)^{-1} \frac{dz}{d(d_L)} \frac{cdt}{dz}\big) dS\,, \tag{63}$$

where without the cosmological corrections this would have been $\int S^{-1.5} dS$ which diverges as $S \to 0$ (a statement of the Olber's paradox).

From the next section (on the age of the universe) we will see that for a universe which is matter dominated and where $\Lambda = 0$, (where $x = 1/1+z$)

$$\frac{dt}{dz} = \frac{-H_0^{-1}}{(1+z)^2\sqrt{(1+\Omega_m z)}} \sim \frac{-H_0^{-1}}{(1+z)^2(1+q_0 z+\ldots)}\,. \tag{64}$$

For the luminosity distance, we derive an approximate result here. We had $d_A = r_1 R(t_0)/(1+z)$ where we have to find r_1 by following the past light cone (since $R(t) =$

$R(t_0)/(1+z)$),

$$R_0 \int \frac{dr}{\sqrt{1-kr^2}} = \int_{t_{em}}^{t_o} (1+z) c dt = \frac{c}{H_0} \int_0^z \frac{dz}{(1+z)(1+q_0 z)} = \frac{cz}{H_0} (1 + \frac{z}{2}(-1-q_0) + \cdots) . \tag{65}$$

So we have (since $r_1 \sim r$ + third order terms from $\sin r$ etc, the LHS is close to $R_0 r$).

$$d_A \approx R_0 r/(1+z) = \frac{cz}{H_0}(1+\frac{z}{2}(-3-q_0)+\cdots) , \tag{66}$$

and

$$d_L = d_A(1+z)^2 = \frac{cz}{H_0}(1+\frac{z}{2}(1-q_0)+\cdots) . \tag{67}$$

Using this we have,

$$\frac{d(d_L)}{dz} = \frac{c}{H_0}(1+z(1-q_0)+\ldots) . \tag{68}$$

one gets,

$$\frac{dN}{dS} = \frac{n_0 (L/4\pi)^{1.5}}{2S^{2.5}} (\frac{1}{(1+z)^3(1+q_0 z+\ldots)(1+z(1-q_0)+\ldots)}) . \tag{69}$$

The q_0 dependence cancels out in this approximation and one has,

$$\frac{dN}{dS} = \frac{n_0 (L/4\pi)^{1.5}}{2S^{2.5}} (\frac{1+O(z^2)}{(1+z)^4}) . \tag{70}$$

So the correction term decreases the number counts below the usual expectation and so the curve flattens. Since the correction term is $\sim (1+z)^4$ so by $z \sim 0.25$ the correction is a factor of ~ 0.5 and and so counts till higher redshifts the flattening is substantial.

This source counts were used to rule out the Steady State model of the universe even before the discovery of the CMBR. It was found that the source counts of radio sources and quasars were steeper than the Euclidean and flattened at only very small fluxes, and the fit to the curve meant that $n_0 \propto (1+z)^9$ which meant that either the density or the luminosity of these sources evolved in the universe, which is opposite to the assumption of $n(z)$ being constant in the Steady State universe.

4 Standard cosmology

4.1 Age of the universe

The Friedmann equation can be integrated to give the age of the universe. First, we write the energy conservation equation as,

$$d[R^3(p+\rho)] = R^3 dp . \tag{71}$$

This means that $wR^3 d\rho = (1+w)R^3 d\rho + (1+w)\rho dR^3$, or $(1+w)dR^3/R^3 = -d\rho/\rho$. For an equation of state of type $p = w\rho$, this shows that the energy density evolves as $\rho R^{3(1+w)} =$ constant, or $\rho \propto R^{-3(1+w)}$. When radiation dominates, $p = \rho/3$, and

$\rho \propto R^{-4}$. And when matter dominates $p = 0$, $\rho \propto R^{-3}$. The Friedmann equation then becomes, (putting $\Lambda = 0$, multiplying by R^2 and dividing by R_0^2)

$$\begin{aligned}
\left(\frac{\dot{R}}{R_0}\right)^2 + \frac{k}{R_0^2} &= \frac{8\pi G}{3}\rho_0 \frac{R_0}{R} \qquad (MD) \\
\left(\frac{\dot{R}}{R_0}\right)^2 + \frac{k}{R_0^2} &= \frac{8\pi G}{3}\rho_0 \left(\frac{R_0}{R}\right)^2 \qquad (RD)\,.
\end{aligned} \tag{72}$$

Here $R_0 = R(t_0)$. Now, we also have $k/R_0^2 = (\Omega_0 - 1)(\dot{R_0}^2/R_0^2) = (\Omega_0 - 1)H_0^2$. Here, Ω_0 is the present value of Ω_m. So, we can write the age of the universe in terms of $1 + z = R_0/R$,as, (where $x = R/R_0$)

$$\begin{aligned}
t &\equiv \int_0^{R(t)} \frac{dR'}{\dot{R}'} \\
&= \frac{1}{H_0}\int_0^{(1+z)^{-1}} \frac{dx}{\sqrt{1 - \Omega_0 + \Omega_0 x^{-1}}} \qquad (MD) \\
&= \frac{1}{H_0}\int_0^{(1+z)^{-1}} \frac{dx}{\sqrt{1 - \Omega_0 + \Omega_0 x^{-2}}} \qquad (RD)\,.
\end{aligned} \tag{73}$$

We have put the zero of time as when R is extrapolates back to zero. These integrals give the results,

$$\begin{aligned}
t_{MD} &= H_0^{-1}\frac{\Omega_0}{2(\Omega_0 - 1)^{3/2}} \\
&\quad \times \left[\cos^{-1}\left(\frac{\Omega_0 z - \Omega_0 + 2}{\Omega_0 z + \Omega_0}\right) - \frac{2(\Omega_0 - 1)^{1/2}(\Omega_0 z + 1)^{1/2}}{\Omega_0(1+z)}\right],
\end{aligned} \tag{74}$$

for $\Omega_0 > 1$, and

$$\begin{aligned}
t_{MD} &= H_0^{-1}\frac{\Omega_0}{2(1 - \Omega_0)^{3/2}} \\
&\quad \times \left[-\cosh^{-1}\left(\frac{\Omega_0 z - \Omega_0 + 2}{\Omega_0 z + \Omega_0}\right) - \frac{2(1 - \Omega_0)^{1/2}(\Omega_0 z + 1)^{1/2}}{\Omega_0(1+z)}\right],
\end{aligned} \tag{75}$$

for $\Omega_0 < 1$. For $\Omega_0 = 1$, $t = (2/3)H_0^{-1}(1+z)^{-3/2}$. One can find the present age of a matter dominated universe by putting $z = 0$ in these equations.

Notice that the age of the universe is a decreasing function of Ω_m. Larger Ω_m means faster deceleration, which corresponds to a more rapidly expanding universe early on. In the limit $\Omega_m \to 0$, $t \to H_0^{-1}(1+z)^{-1}$.

The present age of a matter dominated universe ($\Omega_0 = 1$) is $13.4 \times 10^9 h_{50}^{-1}$ yr. This is consistent with the observations .

For a radiation dominated universe, one has,

$$t = H_0^{-1}\frac{\sqrt{\Omega_0(1+z)^2} - \sqrt{\Omega_0(1+z)^2 - \Omega_0 + 1}}{(\Omega_0 - 1)(1+z)}, \tag{76}$$

which gives for the present age,

$$t_0 = H_0^{-1}\frac{\sqrt{\Omega_0} - 1}{\Omega_0 - 1}\,. \tag{77}$$

4.2 Distances in cosmology: Formulae

Now we can relate the radial coordinate r_1 of a galaxy to its redshift z and time t_1 it emitted a photon that we observe now (t_0). For $k=0$, we have, $t \propto (1+z)^{-3/2} \propto R(t)^{3/2}$ for a matter dominated universe.

$$r_1 = \int_{t_1}^{t_0} \frac{dt}{R(t)} = \frac{1}{R_0} t_o^{2/3} t_1^{-2/3} dt = \frac{3t_0}{R_0}(1-(t_1/t_0)^{1/3}) = \frac{2c}{R_0 H_0}(1-(1+z)^{-1/2}) \,. \quad (78)$$

The luminosity distance is therefore given by,

$$d_L = r_1 R_0 (1+z) = \frac{2c}{H_0}[(1+z) - \sqrt{(1+z)}] \,. \quad (79)$$

For, say, $k=-1$, we have instead,

$$\int_0^{r_1} \frac{dr}{(1-r^2)^{1/2}} = \int_{t_1}^{t_0} \frac{dt}{R(t)} \,, \quad (80)$$

which is abit more complicated to solve. One gets,

$$\begin{aligned} r_1 &= \frac{\sqrt{1-2q_0}}{q_0^2(1+z)}[q_0 z + (1-q_0)(1-(1+2zq_0)^{1/2})] \\ d_l &= \frac{c}{H_0}\frac{1}{q_0^2}(q_0 z + (q_0-1)[(1+2zq_0)^{1/2}-1]) \end{aligned} \quad (81)$$

Notice that for larger q_0 the luminosity distance is smaller for the same redshift. This means that the apparent magnitude m for a source of standard luminosity will be smaller for larger values of q_0.

Also the angular size of an object of proper size l at redshfit z would be,

$$\Delta\theta = \frac{l(1+z)^2}{d_L} = \frac{lH_0}{c}\frac{(1+z)^{3/2}}{\sqrt{(1+z)}-1} \,, \quad (82)$$

for a flat universe. Differentiating this gives the result that the angular size of an object of constant size is minimum at a redshift $z = z_m$ given by,

$$z_m = 1.25, \qquad \theta_{min} = 6.75 \frac{lH_0}{c} \,. \quad (83)$$

4.3 Thermodynamics in the early universe

As we go back in time, we will find that since the momentum of particles were larger, the temperature of the universe was higher. It was also dense and due to the large number density the plasma was possibly in thermal equilibrium. We need to look at this more carefully, to find out which particles were in thermal equilibrium when and how it affected the history of the universe.

We need to review some of the basics of thermodynamics before we can discuss the thermal history of the universe. For a species of particles in kinetic equilibrium the phase space occupancy f is given by the Bose-Einstein or Fermi-Dirac statistics,

$$f(\vec{p}) = \frac{1}{\exp((E-\mu)/k_B T) \pm 1}\,, \tag{84}$$

where μ is the chemical potential of the species. The number density n, energy density ρ and pressure p of a dilute, weakly-interacting gas of particles with g degrees of freedom is given as,

$$\begin{aligned}
n &= \frac{g}{\hbar^3}\int f(\vec{p})d^3p \propto \int_m^\infty \frac{\sqrt{E^2-m^2}}{\exp[(E-\mu)/k_BT]\pm 1} E\,dE \\
\rho &= \frac{g}{\hbar^3}\int E(\vec{p}) f(\vec{p})d^3p \propto \int_m^\infty \frac{\sqrt{E^2-m^2}}{\exp[(E-\mu)/k_BT]\pm 1} E^2 dE \\
p &= \frac{g}{\hbar^3}\int \frac{|\vec{p}|^2}{3E} f(\vec{p})d^3p \propto \int_m^\infty \frac{(E^2-m^2)^{3/2}}{\exp[(E-\mu)/k_BT]\pm 1} E\,dE\,.
\end{aligned} \tag{85}$$

In the relativistic limit ($T \gg m$) and for $T \gg \mu$,

$$\begin{aligned}
\rho &= \begin{cases} (\pi^2/30) g \frac{k_B^4}{c^3\hbar^3} T^4 & \text{(BE)} \\ (7/8)(\pi^2/30) g \frac{k_B^4}{c^3\hbar^3} T^4 & \text{(FD)} \end{cases} \\
n &= \begin{cases} (\zeta(3)/\pi^2) g \frac{k_B^3}{c^3\hbar^3} T^3 & \text{(BE)} \\ (3/4)(\zeta(3)/\pi^2) g \frac{k_B^3}{c^3\hbar^3} T^3 & \text{(FD)} \end{cases} \\
p &= \rho/3\,,
\end{aligned} \tag{86}$$

where $\zeta(3) = 1.20206..$ is the Riemann zeta function of 3. In the non-relativistic limit the expressions are same for bosons and fermions,

$$\begin{aligned}
n &= g\left(\frac{mk_BT}{2\pi}\right)^{3/2} \frac{1}{\hbar^3} \exp[-(mc^2-\mu)/k_BT] \\
\rho &= mnc^2 \\
p &= nk_BT \ll \rho\,.
\end{aligned} \tag{87}$$

Since the energy density of a non-relativistic species ($m \gg T$) is exponentially smaller than that of a relativistic species, it is a good approximation to include only the relativistic particles in the sums for the energy density and pressure, and we can write,

$$\rho = \frac{\pi^2}{30} g_* T^4\,, \tag{88}$$

where g_* is the effective degrees of freedom for species with $m \ll T$,

$$g_* = \sum_{i=bosons} g_i (\frac{T_i}{T})^4 + \frac{7}{8} \sum_{i=fermions} g_i (\frac{T_i}{T})^4 . \tag{89}$$

During the early radiation dominated era ($t \lesssim 4 \times 10^{10}$ s) (in the units $\hbar = k_B = 1$ and where $m_{PL} = \sqrt{\hbar c^5/G} = 1.22 \times 10^{19}$ GeV)

$$\begin{aligned} H &= 1.66 g_*^{1/2} \frac{T^2}{m_{PL}} \\ t &= 0.3 g_*^{-1/2} \frac{m_{PL}}{T^2} \end{aligned} \tag{90}$$

The entropy of a comoving volume (the physical volume being $V = R^3$) can be written as,

$$T dS = d(\rho V) + p dV = V d\rho + \rho dV + p dV , \tag{91}$$

which can be rewritten as,

$$\begin{aligned} dS &= \frac{V}{T} \frac{d\rho}{dT} dT + \frac{(p+\rho)}{T} dV \\ &= (\frac{\partial S}{\partial T})_V dT + (\frac{\partial S}{\partial V})_T dV . \end{aligned} \tag{92}$$

Therefore,

$$(\frac{\partial S}{\partial T})_V = \frac{V}{T} \frac{d\rho}{dT}, \qquad (\frac{\partial S}{\partial V})_T = \frac{(p+\rho)}{T} . \tag{93}$$

So that,

$$\frac{\partial^2 S}{\partial T \partial V} = \frac{1}{T} \frac{du}{dT} = -\frac{(p+\rho)}{T^2} + \frac{1}{T} \frac{d(p+\rho)}{dT} , \tag{94}$$

and finally,

$$T \frac{dp}{dT} = (p+\rho) . \tag{95}$$

This means that,

$$dS = \frac{1}{T} d[(p+\rho)V] - (\rho+p)V \frac{dT}{T^2} = d\left[\frac{(\rho+p)V}{T} + \text{const}\right] . \tag{96}$$

So, the entropy per comoving volume is $S = R^3(\rho + p)/T$, upto a constant. Remember that the first law of of thermodynamics can be written as $d[(p+\rho)V] = V dp$, and so (since $dp = (\rho + p) dT/T$),

$$d\left[\frac{(\rho+p)V}{T}\right] = 0 , \tag{97}$$

which means that in thermal equilibrium, the entropy per comoving volume is conserved.

So, we define an entropy density, s as

$$s \equiv \frac{S}{V} = \frac{p+\rho}{T}\,. \tag{98}$$

The entropy density is dominated by relativistic particles and can be written as,

$$s = \frac{2\pi^2}{45} g_{*s} T^3\,, \tag{99}$$

where,

$$g_{*s} = \sum_{i=bosons} g_i (\frac{T_i}{T})^3 + \frac{7}{8} \sum_{i=fermions} g_i (\frac{T_i}{T})^3\,. \tag{100}$$

In reality for most of the history of the universe all the species had a common temperature and $g_{*s} \sim g_*$, although now they have different values. At any rate, for the discussion of the early universe, we will replace g_{*s} by g_*.

Conservation of S implies that $s \propto R^{-3}$ and so that $g_{*s}T^3R^3$ remains a constant as the universe expands. That is, the temperature evolves as $T \propto g_{*s}^{-1/3} R^{-1}$. So, whenever g_{*s} is a constant we get the result $T \propto R^{-1}$. the factor of $g_{*s}^{-1/3}$ enters here because whenever a particle species becomes non-relativistic and disappears, its entropy is transferred to the other relativistic species still present in the thermal plasma, causing the temperature to decrease slightly less slowly.

One important example of this is when the temperature drops below the mass of electrons and the they become non-relativistic, and their entropy is transferred to the photons. Now it turns out that the neutrinos had decoupled before this happened so that they were not in the thermal plasma any longer. In the early universe the neutrinos were kept in equilibrium via reactions of the sort $\bar{\nu}\nu \to e^+e^-$ and $\nu e \to \nu e$ etc. Now the cross-sections for these weak interaction processes is of order $\sigma \sim G_F^2 T^2$ where G_F is the Fermi coupling constant. The number density of particles is $n \sim T^3 \sim R^{-3}$, so that the interaction rate $\Gamma_{int} = n\sigma|v| \sim G_F^2 T^5$. Comparing this with the Hubble expansion rate for the radiation dominated universe, one finds that,

$$\frac{\Gamma_{int}}{H} \sim \frac{G_F^2 T^5}{T^2/m_{PL}} \sim \left(\frac{T}{1\,MeV}\right)^3\,. \tag{101}$$

This means that at temperatures above 1 MeV, the interaction rate was greater than the expansion rate of the universe and so the neutrinos were in good thermal contact with the plasma. When T drops below 1 MeV, the neutrino interactions are too slow to keep them in equilibrium, so the neutrinos **decouple** from the plasma. After that the temperature of the neutrinos evolve as $T \propto R^{-1}$.

But shortly after that, the temperature also drops below the mass of electron and the entropy in e^+e^- pairs is transferred to the photons (not to the decoupled neutrinos). For $T \gtrsim m_e$, the particle species in thermal equilibrium with photons include the photons ($g = 2$) and e^+e^- pairs ($g = 4$), so that $g_* = 11/2$. For $T \ll m_e$, only the photons are in equilibrium and so $g_* = 2$. For particles in thermal equilibrium with the photons $g_*(RT)^3$ remains constant, which means that after the decoupling of e^+e^-, the value of RT must be larger than that before the decoupling, by a factor $[(11/2)/2]^{1/3}$. So that RT_γ is larger than RT_ν by a factor $(11/4)^{1/3}$, while RT_ν remains constant. So today the ratio of T and T_ν is

$$\frac{T}{T_\nu} = \left(\frac{11}{4}\right)^{1/3} = 1.40\,, \tag{102}$$

which gives $T_\nu \sim 1.96$ K today. Also, the present value of g_* is $2 + \frac{7}{8} \times 2 \times 3 \times (4/11)^{4/3} = 3.36$, taking into account three massless neutrino species. So that the total energy density in relativistic particles is $\rho_R = (\pi^2/30) g_* T^4 = 8.09 \times 10^{-34}$ g cm^{-3}. Also the number density of photons $n_\gamma = \frac{2\zeta(3)}{\pi^2} T^3 = 422$ cm^{-3} assuming $T = 2.75$ K.

4.4 A brief thermal history of the universe

This above example already shows some of the important points that one has to bear in mind while discussing the early history of the universe. Firstly, whether or not a species is in thermal equilibrium depends on whether the interaction rates (which keep it in equilibrium) is slower or faster than the expansion rate. Secondly, the effects like entropy transfers should be taken into account.

We will not discuss the very early universe as things get more and more speculative as one goes back in time. At the earliest epochs, going back to the Planck epoch ($t \sim 10^{-43}$ s and $T \sim 10^{19}$ GeV), the universe was a plasma of relativistic particles, including quarks, leptons etc. Possibly a number of spontaneous symmetry breaking phase transitions took place during the very early universe – a GUT phase transition around $T \sim 10^{14-19}$ GeV and the electroweak phase transition around 300 GeV.

At around $T \sim 100$ to 300 MeV ($t \sim 10^{-5}$ s), another phase transition is supposed to have occurred corresponding to the quark-hadron transition.

The next milestone is the time of nucleosynthesis, which we will discuss soon in detail, around $T \sim 10$ to 0.1 MeV ($t \sim 10^{-2}$ to 100 s). Then one has the decoupling of neutrinos as we have already discussed.

At some epoch, the universe passes from being radiation dominated to matter dominated. Since the density is of order $\rho_m = 1.89 \times 10^{-29} \Omega_0 h^2$ g cm^{-3}. Comparing this with the value of ρ_R obtained earlier, and using the fact that $\rho_R/\rho_m \propto (1+z)$, it follows that the redshift of the epoch of matter and radiation equality is given by,

$$1 + z_{eq} = 2.32 \times 10^4 \Omega_0 h^2 , \tag{103}$$

which corresponds to $t_{eq} = (2/3) H_0^{-1} \Omega_0^{-1/2} (1+z_{eq})^{-3/2} \sim 1.4 \times 10^3 (\Omega_0 h^2)^{-2}$ yr.

Then at some epoch matter and radiation got decoupled. This happened when the interaction rate of photons got slower than the expansion rate. The interaction rate of photons is given by $\Gamma_\gamma = n_e \sigma_T$ where σ_T is the Thomson cross-section, $\sigma_T = 6.65 \times 10^{-25}$ cm^2. To determine this we need to calculate n_e. One should be careful here to take into account the fact that as the universe cools, electrons and protons can recombine to form atoms, and this would decrease the number density of free electrons. This is called the recombination era and we can determine when it happened from Saha's equation. In thermal equilibrium, at temperatures much less than m_i, the number density of a species i is given by,

$$n_i = g_i \left(\frac{m_i k_B T}{2\pi} \right)^{3/2} \frac{1}{\hbar^3} \exp\left(\frac{\mu_i - m_i c^2}{T} \right) . \tag{104}$$

In chemical equilibrium, the process $p + e \to H + \gamma$ guarantees that $\mu_p + \mu_e = \mu_H$. So that we can write,

$$\frac{n_e n_p}{n_H n_{tot}} = \frac{n_e^2}{(n_{tot} - n_e) n_{tot}} = \frac{x^2}{1-x} = \frac{1}{n_{tot}} \left(\frac{m_e k_B T}{2\pi \hbar^2} \right)^{3/2} \exp(-B/k_B T) , \tag{105}$$

where B is the binding energy of hydrogen, $B = m_p + m_e - m_H = 13.6$ eV. We have defined the fractional ionization as,

$$x \equiv \frac{n_p}{n_{tot}}\,, \tag{106}$$

where n_{tot} is the total baryon density, and the fact that, $g_p = g_e = 2$ and $g_H = 4$. This is the Saha's equation for equilibrium ionization fraction.

It is found that for values of $\Omega_b h^2 = 1, 0.1, 0.01$, the universe recombines around the redshift of 1200-1400, if one defines a recombination epoch when 90% of the electrons have combined with protons. One also finds that the duration of the recombination era was short. Taking $1+z = 1300$ as the recombination redshift, the temperature at recombination is then 3575 K, and the age of the universe (using the matter dominated universe expression) is,

$$t_{rec} = (2/3)H_0^{-1}\Omega_0^{-1/2}(1+z_{rec})^{-3/2} = 4.4\times 10^{12}(\Omega_0 h^2)^{-1/2}\,\mathrm{s}\,. \tag{107}$$

4.5 Big bang nucleosynthesis

One of the supports for the big bang hypothesis comes from its prediction of formation of the light elements H, D, He and Li. Originally it was proposed by Gamow, Herman and Alpher, although they though it was possible to form all the elements in the primeval plasma. But the absence of any stable nuclei with atomic mass $A = 5$ and $A = 8$ makes it impossible to go beyond Li in the brief time and relatively low baryon density available in the early universe. We now know that carbon and other heavy elements are synthesised in the cores of red giant stars with the triple-α reaction, and this happens at temperatures around 10^8 K and density of order 10^7 gm/cc, and one has a time scale of 10^{14} sec. In the early universe, when He is formed, the temperature is around 10^9 K but the density is around $2\times10^{-29}\Omega_B h^2 (T/T_0)^3$ gm/cc $= 2\times10^{-5}$ gm/cc, and the time scale is only 3 minutes. So the probability of a three-body collision is negligible. Because of the low density, we will concentrate mainly on reactions involving abundant particles, that is, photons, neutrinos and the electron positron plasma.

Neutrinos interact only via the weak interaction and the typical cross-section is,

$$\sigma_{weak} = 2\times10^{-32}\mathrm{cm}^2\left(\frac{E}{1\,\mathrm{erg}}\right)^2 = 5\times10^{-44}\mathrm{cm}^2\left(\frac{E}{1\,\mathrm{MeV}}\right)^2. \qquad (108)$$

The interaction rate is given by,

$$\begin{aligned}\langle n\sigma v\rangle &= \int\left[\frac{4\pi g_* p^2 dp/h^3}{\exp(pc/kT)+1}\times 2\times10^{-32}\mathrm{cm}^2\left(\frac{pc}{1\,\mathrm{erg}}\right)^2 c\right]\\ &= 4\pi g_*\frac{31}{32}\Gamma(6)\zeta(6)\frac{(kT)^5}{h^3c^2}\times 2\times10^{-32}\mathrm{sec}^{-1}\\ &= 1.1\left(\frac{T}{10^{10}K}\right)^5\mathrm{sec}^{-1}, \end{aligned} \qquad (109)$$

for $g_* = 2$. We need to compare this with the Hubble expansion rate. If we take the density to be equal to the critical density, then $H = \sqrt{8\pi G\rho/3}$. The density is determined by the thermal equilibrium density of photons, neutrinos and electron-positron pairs. We have,

$$H = \sqrt{\frac{8\pi G}{3}\frac{aT^4 + 3\times(7/8)aT^4 + 2\times(7/8)aT^4}{c^2}} = 0.5\left(\frac{T}{10^{10}K}\right)^2\mathrm{sec}^{-1}. \qquad (110)$$

So the weak interactions will freeze out when the temperature is of order $T_f/10^{10} = (0.5/1.1)^{1/3} = 0.77$. One of the weak interactions that freezes out is the transformation of protons into neutrons and back: $p + e^- \leftrightarrow n + \nu$ and $p + \bar{\nu} \leftrightarrow n + e^+$. The energy difference is $E = (m_n - m_p)c^2 = 1.3$ MeV, the neutron to proton ratio freezes out, with a value of $n/p = \exp(-E/kT_f) = 0.14$. These neutrons then decay as free neutrons with a mean lifetime of 887 ± 2 sec, until temperatures fall enough to allow deuterium to form.

The binding energy of deuterium is 2.2 MeV, and the temperature at the freeze out of the weak interaction is only about 0.7 MeV. So one would expect deuterium to form quite easily. But the reaction $p + n \leftrightarrow d + \gamma$ has two rare particles on the left side and one rare particles on the right side. Since the photon to baryon ratio is about 3×10^9, the formation of deuterium will not be favoured until $\exp(-\Delta E/kT) =$

$10^{-9.5}$ which occurs around $T = 10^{9.1}$ K. The age of the universe then is $t = 10^{2.1}$ sec. Now, by this time, about 14 % of the neutrons decay into protons, so the net neutron fraction available for deuterium formation is $0.14(1-0.14) = 0.12$. All these deuterons get incorporated into He^4, so that the final helium abundance by weight is $Y_{prim} = 0.24$. Now, this number depends only weakly on η and is determined mainly by the strength of the weak interactions, the mass difference between proton and neutron and the number of particle types with masses less than 1 MeV that contribute to the expansion rate of the universe during the freeze out. A change of, say, 20 % in the weak interaction rate or the expansion rate during the first 3 minutes would change the predicted value of Y. The close agreement between the predicted value and the observations is a very important confirmation of the big bang model.

Now the deuterium abundance can be used to find out the baryon density of the universe. Let us simplify the reaction network that makes helium to $d + d \to \mathrm{He} + \gamma$. The binding energy here is 24 MeV, so the reverse reaction will not occur since deuterium does not form until $kT < 1$ MeV. The deuteron fraction for this reaction is,

$$\frac{dX_d}{dt} = -2\alpha(T) n_B X_d^2 = -2\alpha(T_1(t/t_1)^{-1/2}) n_B(t_1)(t/t_1)^{-3/2} X_d^2\,, \tag{111}$$

where $\alpha(T)$ is the 'recombination coefficient' for deuterons which will be small at high T, peak at some intermediate temperature and then exponentially suppressed at low temperature by the Coulomb barrier. The solution of this equation is,

$$X_d^{-1} = 2n_B(t_1)\int_{t_1}^{\infty} \alpha(T_1(t/t_1)^{-1/2})(t/t_1)^{-3/2} dt + X_d(t_1)^{-1}\,. \tag{112}$$

The relation between time and temperature is,

$$\frac{1.68 a T^4}{c^2} = \frac{3}{32\pi G t^2} \tag{113}$$

since the deuterium forms after the annihilation of the thermal electron-positron plasma. So that,

$$t = \frac{1.78 \times 10^{20} K^2}{T^2}\,. \tag{114}$$

Almost all the deuterium will be incorporated into helium so that the final deuterium abundance is only slightly dependent on $X_d(t_1)$. Changing variables to $T = T_1\sqrt{t_1/t}$ we get,

$$X_d = \frac{1}{n_B(t_1)} \frac{T_1}{4t_1 \int_0^{T_1} \alpha(T) dT} = \frac{T_1^3}{n_B(t_1)} \frac{1}{7.1 \times 10^{20} K^2 \int_0^{T_i} \alpha(T) dT}\,. \tag{115}$$

Since $n_\gamma(t_1) \propto T_1^3$, $X_d \propto \eta^{-1}$.

Comparison with the observed abundances, one gets an allowed range of the baryon abundance of $2.5 < \eta_{10} < 6$. Now, since the present photon density is $n_\gamma = 412$ /cc for $T_0 = 2.728$ K, we have the present baryon density $n_B(t_0) = 0.412\eta_{10} \times 10^{-7}$ /cc. The critical density for $h = 1$ corresponds to $n_B = 1.12 \times 10^{-5}$/cc at the present epoch. So,

$$\Omega_B h^2 = \frac{0.412\eta_{10} \times 10^{-7}}{1.12 \times 10^{-5}} = 0.00368\eta_{10}\,. \tag{116}$$

So the above range corresponds to $\Omega_B h^2 = 0.01425 \pm 44$ %. For $h = 0.65$, this gives $\Omega_B = 0.034$, which is much less than the measured Ω, which means that most of the matter in the universe is *not* baryonic.

5 Inflation

Although the standard big bang model has been very successful in explaining the observations, there are a few serious problems. One of them is the 'homogeneity problem'. The CMBR is observed to be isotropic to an accuracy of order 10^{-5}, which means that the last scattering surface is homogeneous within the above accuracy over a length scale at least as long as the past light cone l_p at the time t_{rec} of recombination (last scattering). In comoving coordinates, (for $t > t_{eq}$),

$$l_p^c(t) = \int_t^{t_0} dt' R(t')^{-1} = 3t_0(1-(t/t_0)^{1/3})\,. \tag{117}$$

The comoving radius of the forward light cone (the horizon) $l_f^c(t)$ is given by,

$$l_f^c(t) = \int_0^t dt' R(t')^{-1} \approx 3t_0(t/t_0)^{1/3}\,. \tag{118}$$

Regions separated by distances larger than the horizon size cannot be causally connected. But for $t = t_{rec}$ one finds that $l_p^c > l_f^c$. Therefore, no causal processes can explain the homogeneity.

The other problem is the 'flatness problem'. From observations it is now known that $0.2 < \Omega < 2$, that is it is close to being critical. However, in standard cosmology, $\Omega = 1$ is an unstable fixed point during the expansion of the universe. For $\Omega = 1$, $H^2 = \frac{8\pi G}{3}\rho_c$, whereas in general, one has from Friedmann equation (25),

$$H^2R^2 - \frac{8\pi G R^2}{3}\rho = \text{const}\,. \tag{119}$$

So, we have,

$$\Omega^{-1} - 1 = \frac{\text{const}'}{\rho R^2} \tag{120}$$

Suppose we want to have $\Omega = 0.1$ to 2 now; what would be its value at $z = 10^4$? Firstly, const'=$(-0.5...9)\rho_0 R_0^2$, and we know that $\rho R^2 \propto (1+z)^3(1+z)^{-2} \propto (1+z)^{-1}$, so that at $z = 10^4$, the value of Ω ranges from 0.9991 to 1.00005. This means that to get a value of Ω close to unity now, there must be some very effective mechanism for setting the initial value of Ω. This problem becomes more acute when one goes to higher redshifts. In radiation dominated phase, $\rho R^2 \propto T^2$, so that $|1-\Omega^{-1}| \propto T^{-2}$. Let us write the above equation in the following manner,

$$H^2 + \varepsilon T^2 = \frac{8\pi G}{3}\rho\,, \tag{121}$$

where $\varepsilon = k/(RT)^2$ is a constant, k being the curvature constant. Notice that $\varepsilon \propto (RT)^{-2} \propto S^{-2/3}$, where $S \propto (RT)^3$ is the entropy. So, one has,

$$\frac{\rho-\rho_c}{\rho} = \frac{3}{8\pi G}\frac{\varepsilon T^2}{\rho} \propto T^{-2}\,, \tag{122}$$

in the radiation dominated phase. At $T = 10^{14}$ GeV, for example, (corresponding to the scale of grand unification), ratio was $(\rho-\rho_c)/\rho \sim 10^{-50}$. It therefore requires some fine tuning of initial conditions.

The other problem has to do with topological defects, especially monopoles. If the spontaneous symmetry breaking occurred at $T \sim 10^{15}$ GeV ($t \sim 10^{-36}$ sec), then the

Higgs field could have been uniform on length scales of order $ct \sim 3\times10^{-26}$cm and the density of monopoles could have been of order 10^{78} per cc. After the expansion of the universe since then (by a factor 10^{81}), and after possible annihilation with oppositely charged monopoles, the expected current density of monopoles is still of order one per cubic meter. With their large mass, this means that $\Omega \sim 10^{15}$, and so there is a problem.

In 1981, Guth proposed (although there were independent suggestions by Starobinsky and Kazanas) that having a sufficiently long phase in the very early universe during which the scale factor expands exponentially, $R(t) \propto e^{Ht}$, can solve all these problems. This is the inflationary model.

Let t_i denote the onset of inflation and t_R the end of it, so that $\Delta t = t_R - t_i$ is the period of inflation. During the inflation, the forward light cone increases exponentially compared to a model without inflation, whereas the past light cone is not affected for $t \geq t_R$. Therefore, if Δt is sufficiently large, $l_f^c(t_R)$ will be larger than $l_p^c(t_R)$. One essentially needs,

$$e^{\Delta t H} \geq \frac{l_p^c(t_R)}{l_f^c(t_R)} \sim (t_0/t_R)^{1/2}, \tag{123}$$

which is in turn proportional to $(T_R/T_0) \sim 10^{27}$, for $T_R \sim 10^{14}$ GeV and $T_0 \sim 10^{-13}$ GeV (the present CMBR temperature). So, in order to solve the homogeneity problem one needs a period of inflation with $\Delta t \gg 50H^{-1}$.

This can also solve the flatness problem. The important point is that for reasons that we cannot go into here the temperature at t_i and t_R (which are essentially the epochs of beginning of inflation and that of reheating, respectively) are comparable. So, the entropy increases during inflation by a factor $\exp(3H\Delta t)$ and so ε decreases by a factor of $\exp(-2H\Delta t)$. So for the numbers used above, one has $\varepsilon_{after} \sim 10^{-54}\varepsilon_{before}$. Hence $(\rho-\rho_c)/\rho$ need not be closer to unity then than it is now. As a matter of fact, inflation predicts that at present time Ω would be unity to a high accuracy.

Inflation also solves another problem, by providing a causal way of generating density fluctuation to form large structures. Since the Hubble radius does not change much during inflation, structures with fixed comoving scales can form inside the Hubble radius during inflation, then leave the horizon, becoming bigger than it, and then enter again.

5.1 Realization of inflation

Although in the inflationary phase, R increases exponentially, it is adequate to have an accelerated expansion, so that $R \propto t^q$ where $q > 1$. In that case the Hubble radius, in comoving coordinates, is $H^{-1} \propto t^{1-q}$. One only needs the Hubble radius in comoving coordinates to decrease in time (so that structures forming inside with fixed comoving scale can leave later). This needs $q > 1$; in other words one wants $\ddot{R}(t) > 0$.

Now, the difference of equation (20) and equation (21) gives,

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}(\rho+3p), \tag{124}$$

so that an accelerated growth requires,

$$p < -\frac{1}{3}\rho. \tag{125}$$

So one needs negative pressure, and to discuss this one needs to use quantum field theory. One can show in the context of field theory that if the dynamics of the universe is dictated by a (scalar) field ϕ and the Lagrangian is dominated by the potential term $V(\phi)$, then one can have negative pressure if some conditions are met. The details depend on aspects of field theory that we cannot go into.

5.2 Fluctuations from inflation

It is thought that quantum fluctuations on microscopic scales could have grown to larger than the universe during the inflationary phase. These perturbations will then enter the horizon again later and act as perturbations needed for structure formation. For perturbations larger than $c_s t(\sim c_s/H)$ one can ignore pressure gradients, since sounds waves cannot cross the perturbation in one Hubble time. In the absence of pressure gradients, the perturbations will grow like a homogeneous universe,

$$\Omega^{-1} - 1 = \frac{\text{const}}{\rho R^2}\,. \tag{126}$$

Assuming that $\Omega \sim 1$ at early epochs and that the fluctuations are small ($\Delta\rho \ll \rho$), one has,

$$-\rho R^2(\Omega^{-1} - 1) \sim \rho_{crit} R^2 \Delta\Omega \sim \Delta\rho R^2 \sim \text{const}\,, \tag{127}$$

and so,

$$\Delta\phi = \frac{G\Delta M}{R} = \frac{4\pi}{3}\frac{G\Delta\rho_0(RL)^3}{RL} \propto L^{-2} \tag{128}$$

where L is the comoving size of the perturbation. Note that this is independent of the scale factor and so it does not change during the expansion of the universe. During inflation, the magnitude of $\Delta\phi$ for perturbations with physical scale c/H does not change, but since this physical scale is RL and R changes by several orders of magnitude during inflation, it means that the magnitude of $\Delta\phi$ remains same for many orders of magnitude of L. This is called the scale-invariant spectrum of perturbation—with equal power on all scales, originally predicted by Harrison (1970, PRD, 1, 2726), Zel'dovich (1972, MNRAS, 160, 1p) and Peebles and Yu (1970, ApJ, 162, 815), and is often called the Harrison-Zel'dovich spectrum of perturbation.

Notice that this also means that $\Delta\rho/\rho \propto L^{-2}$ and so the the universe becomes more homogeneous as one goes to larger scales.

5.3 Topological defects

Topological defects arise naturally in the context of spontaneous symmetry breaking. Modern particle physics theories invoke such symmetry breaking to try to explain the existence of various interactions starting from some sort of unified interaction. For example, the electroweak theory incorporates the long range electromagnetic force mediated by massless photons and the short range weak nuclear force mediated by massive W and Z bosons. In the context of spontaneous symmetry breaking, one has a symmetric models but whose lowest energy states are not symmetric.

Under spontaneous symmetry breaking, it is possible to have defects, just like in phase transitions. The defects can be like points (monopoles), one dimensional (cosmic strings), two dimensional (domain walls) and so on. We can view these defects as being regions of trapped energy density which is frozen from the time of phase transitions. For example, cosmic strings can be characterized by the mass per unit length μ, which is about 10^{22} gm/cm, if the strings form at 10^{16} GeV, at the breaking of the symmetries in the grand unified theories.

Cosmic strings has been invoked to help in structure formation models, and topological defects in general have been suggested as being the cause for ultra-high energy cosmic rays. They ought to, however, leave their tell-tale signatures in the CMBR anisotropy measurements, and it is not yet clear whether one has seen these signs.

6 Radiation in the universe

6.1 Diffuse backgrounds

If a radiation field is homogeneous and isotropic like the universe, one only needs to consider a small region of space. Consider a small cubicle region bounded by co-moving mirrors. As in a barber shop with mirrors on two opposing walls, one sees an infinite number of images upto infinity, and this is just like a homogeneous and isotropic universe. The equation of radiation transfer is,

$$\frac{dI_\nu}{ds} = j_\nu - \alpha_\nu I_\nu\,, \tag{129}$$

where I_ν is the specific intensity in erg/cm^2/sec/sr/Hz, s is the path length along the ray, j_ν is the emissivity in erg/cm^3/sec/sr/Hz, and α_ν is the absorption coefficient in cm^{-1}. The optical depth $\tau_n u = \int \alpha_\nu ds$; so we can write,

$$\frac{dI_\nu}{d\tau_\nu} = \frac{j_\nu}{\alpha_\nu} - I_\nu = S_\nu - I_\nu\,, \tag{130}$$

where the source function is S_ν.

6.2 Olber's Paradox

Consider j_ν to be due to blackbody stars with number density n, radius R and temperature T_*. Then the luminosity per unit frequency of a single star is $L_\nu = 4\pi^2 R^2 B_\nu(T_*)$. The emissivity is $j_\nu = nL_\nu/4\pi = n\pi R^2 B_\nu(T_*)$. The absorption coefficient can be written as the reciprocal of the mean free path, and is $\alpha_\nu = n\pi R^2$. Solving the radiation transfer equation, one gets,

$$I_\nu = \exp(-\tau_\nu) I_\nu(0) + [1-\exp(-\tau_\nu)] B_\nu(T_*)\,. \tag{131}$$

If the path length $s \to \infty$, then the intensity approaches $B_\nu(T_*)$, the surface brightness of a star. This is the Olber's paradox : why is then the night sky dark? Obviously, something is wrong with our assumptions.

One needs to take the fact that the universe is expanding and the effect of the finite age of the universe, as we will see below.

6.3 Cosmological equation of transfer

We will neglect scattering here, because for diffuse background radiation, the effect of scattering cancels out when one averages over solid angle. For individual sources, however, scattering might be important. One can then find out the total optical depth by using $ds = c(dt/dz)dz$.

For diffuse backgrounds, we will replace I_ν by a quantity that does not change during the expansion. The number of photons per mode, $I_\nu/(2h\nu[\nu/c]^2)$, evaluated at $\nu = \nu_0(1+z)$ is one such quantity. Using this, we have,

$$\frac{\partial}{\partial z}\left(\frac{I_{\nu_0(1+z)}}{(1+z)^3}\right) = \frac{cdt/dz}{(1+z)^3}\left(j_{\nu 0(1+z)} - \alpha_{\nu 0(1+z)} I_{\nu 0(1+z)}\right). \tag{132}$$

Another way of looking at this equation is to think in terms of the photon distribution function $n(\nu, t)$ and how it changes in time. Let the number of photons per unit volume in a bandwidth $\nu, \nu+\delta\nu$, is $\delta n = n(\nu,t)\delta\nu$. First consider the case when there is no destruction or creation of photons. In that case, the distribution function changes because of changes in (1) volume due to expansion, (2) redshift of photon and in (3) change in the bandwidth. Since $nR^3 =$ constant and $\nu R =$ constant, the first term for $\frac{\partial n}{\partial t}$ is $-3n(\dot{R}/R)$. The second term is $-\frac{\partial n}{\partial \nu}\frac{\partial \nu}{\partial t} = \nu(\dot{R}/R)\frac{\partial n}{\partial \nu}$. (We put a negative sign in front since as time progresses, photons of lower energy are come into the bandwidth and if $\frac{\partial n}{\partial \nu}$ is positive, this decrease can described by adding a negative sign.) Since $n \propto \delta\nu^{-1}$, the third term is $n(\dot{R}/R)$. So, finally one has,

$$\frac{\partial n}{\partial t} = \nu\frac{\dot{R}}{R}\frac{\partial n}{\partial \nu} - 2\frac{\dot{R}}{R}n. \tag{133}$$

Now, using the energy density $u(\nu,t) = n(\nu,t)\times h\nu$, the factor 2 changes to 3, and taking into the possibility of a volume emissivity j, one can write,

$$\frac{\partial u}{\partial t} = \nu\frac{\dot{R}}{R}\frac{\partial u}{\partial \nu} - 3\frac{\dot{R}}{R}u + 4\pi j(\nu,t). \tag{134}$$

Then since $I = cu/4\pi$, one has,

$$\frac{d}{dt}I_\nu = \frac{\partial I}{\partial t} - \nu\frac{\dot{R}}{R}\frac{\partial I}{\nu} = -3\frac{\dot{R}}{R}I_\nu + cj_\nu\,, \tag{135}$$

and so,

$$\frac{d}{dt}(I_\nu R^3) = cj_\nu R^3\,, \tag{136}$$

which is equivalent to the equation we derived earlier, as $R = 1/1+z$.

Consider Olber's paradox again. Suppose we have a luminosity density $L = 4\pi\int j_\nu d\nu$. The comoving luminosity density is constant, since $j(\nu,z) = j(\nu,z=0)(1+z)^3$. We will ignore α since it is very small so that we have,

$$I_{\nu_0} = \int j_{\nu_0(1+z)}cdt\,. \tag{137}$$

For the total intensity we integrate over the frequency,

$$I = \int I_{\nu_0}d\nu_0 = \int\int j_{\nu_0(1+z)}cdtd\nu_0 = \frac{L}{4\pi}\int\frac{cdt}{(1+z)}\,, \tag{138}$$

which shows that the total (bolometric) intensity is reduced by a factor of $(1+z)^{-1}$, and is also limited by the age of the universe.

For an example of diffuse backgrounds, consider the X-ray background produced by a hot intergalactic medium (IGM) with $n_e = n_e(0)(1+z)^3$. The emissivity of a hot plasma is $j_\nu = An_e^2\exp(-h\nu/kT)/\sqrt{T}$ and the absorption coefficient is negligible. So we have,

$$\left.\frac{I_{\nu_0(1+z)}}{(1+z)^3}\right|_{z=0} = An_e^2(0)\frac{c}{H_0}\int\frac{(1+z)^6\exp(-h\nu_0(1+z)/kT(z))/\sqrt{T}dz}{(1+z)^3(1+z)^2\sqrt{1+\Omega_{m0}z}}\,, \tag{139}$$

so that,

$$I_{\nu 0} = An_e^2(0)\frac{c}{H_0}\int\frac{(1+z)\exp(-h\nu_0(1+z)/kT(z))dz}{\sqrt{1+\Omega_{m0}z}}\,. \tag{140}$$

This turns out to be exceedingly small for any realistic value of $n_e(0)$.

6.4 CMBR distortion from electron scattering

It is possible for scattering by electrons in ionized gas (either in the early epochs or after reionization (see later)) to transfer energy from gas to the radiation field and thus change the radiation field. This is important in view of the possible distortion of the CMBR. If the gas is non-relativistic (temperatures less than 10^8 K) then the effect of electron scattering on the spectrum of radiation can be determined by using the Kompaneets equation (Kompaneets 1957, Sov. Phys JETP, 4 , 730) which we will not derive here (but see either Rybicki & Lightman or 'High energy astrophysics' by Katz),

$$\frac{\partial n}{\partial y} = x^{-2}\frac{\partial}{\partial x}\big(x^4(n+n^2+\frac{\partial n}{\partial x})\big)\,, \tag{141}$$

where n is the number of photons per mode, $x = h\nu/kT_e$ and the Kompaneets y parameter is defined by,

$$dy = \frac{kT_e}{m_e c^2} n_e \sigma_T c dt\,. \tag{142}$$

Note that for a blackbody spectrum, $n = 1/(exp(h\nu/kT_\gamma)-1)$. The y parameter is the electron scattering optical depth multiplied by the electron temperature in units of the electron rest mass. One should note that in deriving this one has used a Maxwellian distribution of the electrons. Also note that scattering conserves the photons and so the general outcome is a Bose-Einstein distribution of photons.

Sunyaev and Zel'dovich (1969, Ap. & Sp. Sci, 4, 301) found a simple solution to the Kompaneets equation for the case when the blackbody radiation field has a temperature $T_\gamma < T_e$. Let us write $f = T_e/T_\gamma$. Since the distortion is small, one takes the photon field to be a blackbody, $n = 1/(\exp(fx)-1)$. Then,

$$n^2+n+\frac{\partial n}{\partial x} = \frac{(1-f)\exp(fx)}{(\exp(fx)-1)^2} = (1-f^{-1})\frac{\partial n}{\partial x}\,. \tag{143}$$

One can then find that, (after redifining $x = h\nu/kT$)

$$\frac{\Delta n}{n} = y\big(\frac{x^2e^x(e^x+1)}{(e^x-1)^2} - \frac{4xe^x}{e^x-1}\big)\,, \tag{144}$$

which in the limit $x \ll 1$ approaches $-2y$ and in the limit $x \gg 1$ approaches x^2y. So in the Rayleigh-Jeans part, where $n \propto T_\gamma$, the effective temperature is lowered by an amount,

$$\frac{\Delta T}{T} = -2y\,. \tag{145}$$

The hot electrons boost the energy of the photons and the effective temperature is lowered at long wavelengths. This is what is expected when the CMBR photons traverse through a cluster which has the hot intra-cluster gas. Sunyaev and Zel'dovich argued that the plasma in a rich cluster is hot enough to cause a substantial distortion and this is called the SZ effect. The effect has been observed and provides important clues to the structure of clusters. Also it is now known that if clusters and intra-cluster gas are present back to redshift $z \sim 1$, the integrated effect can be important and will add to the anisotropy of the CMBR at arc minute scales.

With typical values of $n_e \sim 10^{-3}$ cm^{-3}, $T_e \sim 5$ keV and a line of sight length of ~ 200 kpc, one gets an optical depth of $\sim 10^{-3}$ and the distortion at long wavelengths,

$$\frac{\Delta T}{T} \sim -2\tau\frac{kT_e}{m_e c^2} \sim 10^{-5}\,, \tag{146}$$

Although detecting this signal through the noise of the radio sources in and behind the cluster is a big challenge, this has been detected (Birkinshaw 1999 Phys Rep, 310, 97).

COBE measurements have shown that $y \leq 1.5\times10^{-5}$ which provides constraints to any hot diffuse intergalactic medium. The energy density transferred from the hot electrons can be calculated by considering the energy density $U \propto x^3 n dx$, which means that,

$$\frac{\partial U}{\partial y} = \int x \frac{\partial}{\partial x}\left(x^4 \frac{\partial n}{\partial x}\right) dx, \tag{147}$$

which gives, after integrating by parts twice,

$$\begin{aligned} \frac{\partial U}{\partial y} &= -\int \left(x^4 \frac{\partial n}{\partial x}\right) dx \\ &= 4\int x^3 n dx = 4U. \end{aligned} \tag{148}$$

So the limit of $y \leq 1.5\times10^{-5}$ means that the fractional energy deposition to the CMBR has been $\Delta U/U < 6\times10^{-6}$. Note that at $z > 7$ the inverse Compton cooling time scale is less than a Hubble time and the transfer of energy of electrons to the photons before this epoch is going to distort the CMBR. The Compton cooling time scale is,

$$t_C \sim \frac{m_e c^2}{4\sigma_T c U_{rad}} = \frac{5\times10^{19}\text{sec}}{(1+z)^4}, \tag{149}$$

which should be compared to the Hubble time $3\times10^{17} h^{-1}(1+z)^{-1.5}$ sec.