Chapter 3

Cosmology

3.1 Foundations

The pivotal moment in the history of modern cosmology was the "Great Debate" about the physical size of the Milky Way galaxy between Harlow Shapley and Heber D. Curtis, which took place on April 26th 1920. The general consensus was that Shapley's position was stronger, concluding that the Milky was was quite large, and included such objects like the Large and Small Magellenic Clouds and the Andromeda galaxy (not known at that time to be an actually galaxy in itself). A few years later in 1923, Edwin Hubble showed that Andromeda was actually very far away, and thus much bigger than previously thought. While contradicting Shapley's argument and making the Milky Way itself smaller, Hubble ended up making the universe *much* larger.

In 1926, Hubble proposed his eponymous distance-redshift relation, with recession velocity linearly proportional to distance,

$$\vec{\mathbf{v}} = H\vec{\mathbf{r}}, \tag{3.1}$$

where H is the Hubble parameter (in units of inverse time) and $\vec{\mathbf{r}}$ and $\vec{\mathbf{v}}$ are the position and velocity in space as measured by an observer. The Hubble relation is the first of four "pillars" of modern cosmoology. An immediate consequence of is the homogeneity of space: observers at all points in space measure the same recession. If an observer in the Milky Way (see Fig. 3.1) measures the velocities of galaxies at relative positions $\vec{\mathbf{r}}_1$ and $\vec{\mathbf{r}}_2$:

$$\begin{aligned} \vec{\mathbf{v}}_1 &= H\vec{\mathbf{r}}_1 \\ \vec{\mathbf{v}}_2 &= H\vec{\mathbf{r}}_2, \end{aligned} \tag{3.2}$$

then an observer in galaxy 1 will measure the velocity of galaxy 2 to be

$$\vec{\mathbf{v}}_{21} = H(\vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1). \tag{3.3}$$

This linear scaling relation is a direct result of the homogeneous expansion of the universe.

In 1926, with only crude techniques for measuring extragalactic distances, Hubble determined a value for H of $\sim$ 500 km/s/Mpc, much larger than the currently accepted value of $H = 72 \pm 8$

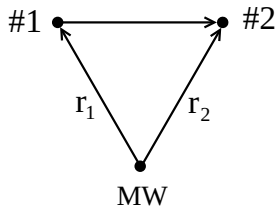

Figure 3.1: Two galaxies as seen from the Milky Way. An observer in galaxy #1 sees galaxy #2 receding with velocity $H(\vec{r_2} - \vec{r_1})$.

km/s/Mpc. This measurement proceeds through a chain of primary, secondary and tertiary "distance indicators," most of which involve an empirically established relation between some distance-independent observable quantity and the intrinsic luminosity of the object considered. The most important link in this chain is the period-luminosity relation for Cepheid variables. This is then used to calibrate indicators that can be seen at greater distances: type Ia supernovae, the Faber-Jackson and Tully-Fisher relations, surface brightness fluctuations in elliptical galaxies and the luminosity function of planetary nebulae. The Cepheids are themselves calibrated by measuring the distance to the Large Magellanic cloud by a suite of methods that includes the light echo from supernova 1987a, Cepheids measured in the Milky Way and RR Lyrae stars. A second route to Cepheid calibration is the exquisitely accurate geometric distance to NGC 4258 as determined from the observed acceleration of the water masers in its active galactic nucleus.

The second pillar of modern cosmology is its apparent *isotropy*. All manner of cosmological phenomena – counts of distant galaxies, radio sources, X-ray sources and gamma ray bursts – appear to be independent of angular position in the sky. The sky looks roughly the same in all directions and at all wavelengths (after accounting for the additional absorption and emission from the nearby galactic plane). Along with a general form of Copernicus' principle, arguing against the earth being the center of the universe, i. e. there is no "special place" in the universe, isotropy implies homogeneity, thus suggesting that on the largest spatial scales, density perturbations all become smoothed out.

The third pillar of modern cosmology is the measured abundances of the light elements, particularly hydrogen and helium. From basic nuclear reaction calculations, the amount of helium produced in stars cannot be more than about three times that of the heavier elements:

$$\frac{\Delta Y}{\Delta Z} \leq 3. \tag{3.4}$$

With measured cosmic abundances closer to 70% H, 28% He, and 2% ($Z >$He), it is clear that the major portion of helium in the universe must *not* have been produced in stars. In 1948, Gamow, Alpher, and Herman ran the expansion of the universe "backwards," heating up the photon background to the point that would have been necessary to produce a cosmic mass fraction of 25% in helium from primordial nuclear reactions. Given the present-day density of baryons, they predicted a red-shifted background photon temperature of 6 °K. Then, in 1965, Penzias and Wilson were the first to experimentally measure a static radio signal corresponding to a thermal temperature of ~ 3 °K, what is now known as the Cosmic Microwave Background (CMB) radiation, spatially isotropic to about one part in 10^5.

To develop a rudimentary model for the dynamics of an expanding universe, we need to introduce one fundamental result from general relativity. A direct analogy to Newton's first and second theorems, *Birkhoff's theorem* states that for any system with spherical symmetry, concentric shells are only affected by the mass and energy interior to their radius, which collectively behaves as a point particle at the origin. In an infinite universe with uniform density ρ_0, the specific Newtonian energy

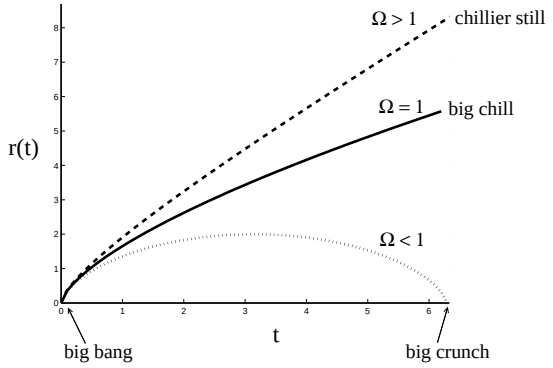

Figure 3.2: Three futures.

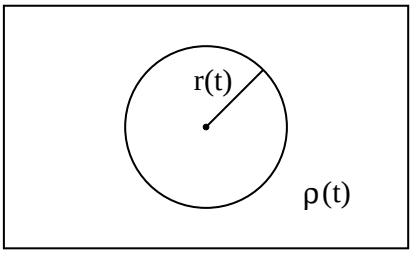

Figure 3.3: The rectangular region represents the universe with uniform density $\rho(t)$. The circle represents a spherical shell whose growth we follow with Newtonian dynamics.

of a spherical shell with radius r is

$$E = \frac{1}{2}\dot{r}^2 - \frac{GM_r}{r} = \frac{1}{2}H_0^2 r^2 - \frac{4\pi}{3}G\rho_0 r^2. \tag{3.5}$$

It is convenient to define the cosmological density parameter Ω_0, the fraction of "critical" density needed to just close the expansion ($E = 0$):

$$\Omega_0 \equiv \frac{8\pi G\rho_0}{3H_0^2}. \tag{3.6}$$

Here and throughout the discussion of cosmology, the subscript "0" will generally refer to the value of a given parameter at the present epoch. The three interesting possibilities for Ω_0 are

$$\begin{array}{lcll} \Omega_0 > 1 & : & E < 0 & \text{closed, bound} \\ \Omega_0 = 1 & : & E = 0 & \text{critical, "Einstein-de Sitter"} \\ \Omega_0 < 1 & : & E > 0 & \text{open, unbound,} \end{array} \tag{3.7}$$

each corresponding to a qualitatively different history and future for the universe (see Fig. 3.2). One of the major goals of modern cosmology is to conclusively measure Ω_0, and thus determine the global geometry of the universe.

3.2 Newtonian spherical model

In this Section we will develop the simplest model for cosmic expansion, following the dynamics of an expanding spherical shell of matter with radius $r(t)$ in Newtonian gravity (see Fig. 3.3). For

any bound system ($\Omega_0 > 1$), there is a maximum radius of expansion $r_{\max}$ for any shell. At this turn-around point, the shell velocity is zero, so the energy determines the radius

$$E = -\frac{GM_r}{r_{\max}}. \tag{3.8}$$

Plugging into equation (3.5), we can solve for the expansion velocity

$$\frac{dr}{dt} = v = \sqrt{2GM_r}\left(\frac{1}{r} - \frac{1}{r_{\max}}\right)^{1/2}. \tag{3.9}$$

Following standard techniques for solving central-force problems, make a change of variables:

$$\frac{r}{r_{\max}} = \sin^2 x, \qquad \frac{dr}{r_{\max}} = 2\sin x \cos x dx, \tag{3.10}$$

giving for equation (3.9)

$$\frac{2\sin x\cos x}{\cot x}dx = \sqrt{\frac{2GM}{r_{\max}^3}}dt. \tag{3.11}$$

Converting to the Keplerian *eccentric anomaly* η with an equivalent eccentricity of $e = 1$, let $x = \eta/2$ and then

$$\left(\frac{2GM}{r_{\max}^3}\right)^{1/2}\int dt = \frac{1}{2}\int(1-\cos\eta)d\eta. \tag{3.12}$$

This gives parametric solutions of $t(\eta)$ and $r(\eta)$:

$$t(\eta) = \frac{1}{2}\left(\frac{3}{8\pi G\rho_{\max}}\right)^{1/2}(\eta - \sin\eta) \tag{3.13}$$

and

$$r(\eta) = \frac{r_{\max}}{2}(1-\cos\eta), \tag{3.14}$$

where we have introduced the density

$$\rho_{\max} \equiv \frac{3M}{4\pi r_{\max}^3}$$

that corresponds to the density of the matter within the spherical shell at the point of maximum radius $r_{\max}$.

From equations (3.13) and (3.14), we can derive expressions for the Hubble and density parameters $H(\eta)$ and $\Omega(\eta)$ for this particular spherical perturbation

$$H(\eta) \equiv \frac{dr/dt}{r} = \frac{1}{r}\frac{dr/d\eta}{dt/d\eta} = 2\left(\frac{8\pi G\rho_{\max}}{3}\right)^{1/2}\frac{\sin\eta}{(1-\cos\eta)^2} \tag{3.15}$$

and

$$\Omega(\eta) = \frac{8\pi G\rho}{3H^2} = \frac{2}{1+\cos\eta}. \tag{3.16}$$

For under-dense regions with $\Omega < 1$ the shell is unbound and there is no maximum radius. Fortunately, all of the above results can nonetheless carry forward, with the difference that hyperbolic trigonometric functions are substituted for ordinary trigometric functions:

$$\begin{aligned} 1 - \cos\eta &\rightarrow \cosh\eta - 1 \\ \eta - \sin\eta &\rightarrow \sinh\eta - \eta \\ \cos\eta &\rightarrow \cosh\eta \\ \sin\eta &\rightarrow \sinh\eta \end{aligned} \tag{3.17}$$

Since under-dense shells have positive energy, the expansion velocity at infinity approaches a finite, positive value. The radius $r_{\max}$ marks transition from decelerated to free expansion.

$$E = \frac{2GM}{r_{\max}} = \frac{1}{2}\left(\frac{dr}{dt}\right)_{\infty}^{2}. \tag{3.18}$$

Though we have used time as the abscissa in Fig. 3.2, it is not an observable quantity. Much more useful is the redshift of photons emitted at some early time.

For small values of η, both the over-dense and the under-dense spheres behave similarly:

$$t(\eta) \approx \frac{1}{12}\left(\frac{3}{8\pi G\rho_{\max}}\right)^{1/2}\eta^3 \tag{3.19}$$

$$r(\eta) \approx \frac{r_{\max}}{4}\eta^2, \tag{3.20}$$

giving the important result for early times

$$r \propto t^{2/3} \tag{3.21}$$

which is independent of the actual value for Ω.

While neither the trigonometric or hyperbolic solutions are valid for $\Omega = 1$, there are still some well-defined quantities that can be derived by taking the appropriate limits from Eqn. 3.16 ($\eta \rightarrow 0, \Omega \rightarrow 1$):

$$t = \left(\frac{3}{4\pi G\rho}\right)^{1/2}\frac{\eta^3/6}{(\eta^2/2)^{3/2}} = \left(\frac{1}{6\pi G\rho}\right)^{1/2}. \tag{3.22}$$

The dimensionless parameter Ht is a good way to describe the global expansion of the universe. For a critical density,

$$Ht = \frac{(\eta - \sin\eta)\sin\eta}{(1-\cos\eta)^2} \rightarrow \frac{2}{3}. \tag{3.23}$$

This is a good point for a quick digression to define one of the most important terms in cosmology, the *redshift*. Due to Hubble's law of expansion, we can equate the distance to objects with their observed recessional velocity. This velocity is most commonly determined by measuring the change in wavelength of light emitted by that object due to its velocity relative to the distant observer. Since most objects in the universe are moving away from us, we call this Doppler shift "redshift",

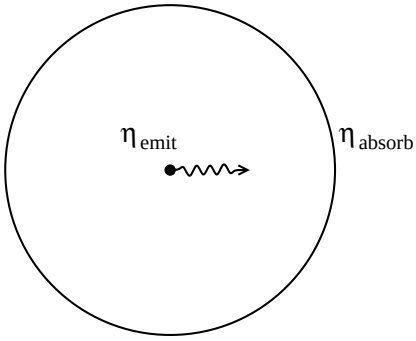

Figure 3.4: A photon is emitted at eccentric anomaly η_{emit} and absorbed at η_{absorb}.

Table 3.1: Struture in the universe

	$r_{1/2}$	$M(V_{1/2})/(\rho_u V_{1/2})$	$t_{collapse}/t_0$
ellipticals	2 h^{-1} kpc	10^{11}	10^{-3}
clusters of galaxies	1 h^{-1} Mpc	10^3	$10^{-0.5}$
superclusters	20 h^{-1} Mpc	3	not yet
voids	50 h^{-1} Mpc	1/3	never will

since the red part of the visible spectrum has longer wavelength than the blue. For a photon that has wavelength $\lambda_{\rm emit}$ in the rest frame of the emitter and wavelength $\lambda_{\rm obs}$ in the rest frame of the observer, the redshift z is defined as

$$1 + z \equiv \frac{\lambda_{\rm obs}}{\lambda_{\rm emit}}. \tag{3.24}$$

For nonrelativistic velocities $v \ll c$, the redshift can be approximated as

$$z = \frac{\Delta\lambda}{\lambda_{\rm emit}} = \frac{v}{c} = H_0 \Delta t = H_0 \frac{dt}{d\eta}\Delta\eta = \frac{dr/d\eta}{r}\Delta\eta = \frac{\Delta r}{r} \quad , \tag{3.25}$$

where Δt is the difference in cosmological time between the emission and detection events. Similarly, $\Delta\eta$ is the difference as measured in the eccentric anomaly parameter (see Fig. 3.4). Thus we see how the redshift can act as a more easily measurable proxy for the time and η variables. More remarkably, redshift exactly tracts the size of any shell and the of the universe as a whole. The wavelength of photon and the universe by the same amount.

3.3 Perturbations: the spherical model

The perfectly uniform universe considered above would be a boring place to live. Fortunately our universe isn't. Small scale density perturbations present at some early epoch grow by gravitational condensation into larger perturbations and eventually form the non-linear structures that are the bread and butter of astronomers. A caricature of the kinds of structures observed in the universe is given in Table 3.1

To apply the spherical shell model to structure formation in an expanding background universe, it is helpful to think about the perturbation as having an initial density enhancement of

$$\frac{\Delta\rho}{\rho} = \frac{\rho_p - \rho_u}{\rho_u}, \tag{3.26}$$

where ρ_p is the density inside the spherical perturbation and ρ_u is the unperturbed background density. Consider two spherical shell models evolving in parallel (see Fig. 3.5), each measuring $t(\eta)$ and $r(\eta)$ according to its own parameterization (recall that the dependence of r and t on η are functions of the density parameter). Give one shell an overdensity of ρ_p and the companion shell the same density as the background ρ_u. Then by setting $t_p(\eta_p) = t_u(\eta_u)$ in equations (3.13) and (3.23), we can measure the time to collapse. Assuming a critical background ($\Omega_u = 1$), we have

$$\left(\frac{3}{4\pi G\rho_p}\right)^{1/2} \frac{\eta_p - \sin\eta_p}{(1-\cos\eta_p)^{3/2}} = \left(\frac{1}{6\pi G\rho_u}\right)^{1/2}. \tag{3.27}$$

Solving for the density of the perturbation,

$$\frac{\rho_p}{\rho_u} = \frac{9}{2}\frac{(\eta_p - \sin\eta_p)^2}{(1-\cos\eta_p)^3}, \tag{3.28}$$

so at early times (small η_p),

$$\frac{\rho_p}{\rho_u} \approx 1 + \frac{3}{10}\eta_p^2. \tag{3.29}$$

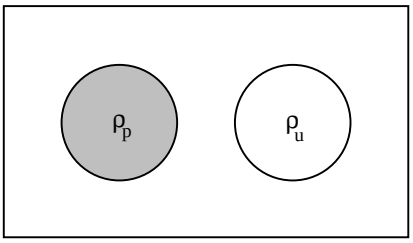

Figure 3.5: The growth of a perturbation with density ρ_p compared to a region with density ρ_u.

Combining with equation (3.19), we find that at early times, like the radius, the density grows algebraically as

$$\frac{\Delta\rho_p}{\rho_u} \sim t^{2/3}. \tag{3.30}$$

The local Hubble parameter evolves as

$$\frac{H_p}{H_u} = \frac{3}{2}\frac{\eta_p(\eta_p - \sin\eta_p)}{(1-\cos\eta_p)^2} \approx 1 - \frac{1}{20}\eta_p^2. \tag{3.31}$$

Note that the smaller value for the Hubble parameter corresponds to the perturbation expanding more slowly than the background. From equations (3.29) and (3.31), we find the fractional difference in expansion rates

$$\frac{\Delta H}{H} = -\frac{1}{3}\frac{\Delta\rho}{\rho_u} \quad . \tag{3.32}$$

Dropping the assumption that $\Omega_u = 1$ and carrying the calculation to higher order in η one finds

$$\frac{\Delta H}{H} = -\frac{1}{3}\frac{\Delta\rho}{\rho_u}\Omega_u^{4/7} \tag{3.33}$$

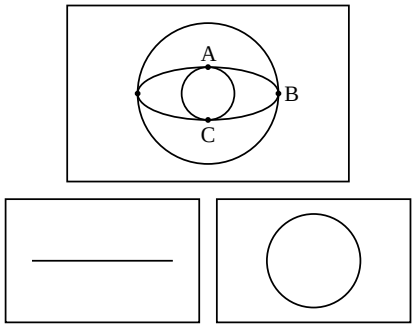

Figure 3.6: Overdense regions collapse to sheets and filaments (bottom left) while underdense regions expand to form voids that are nearly spherical (bottom right). Take the ellipsoidal region in the top figure to be overdense. Point A will decelerate more rapidly than it would if only the inscribed sphere were comparably overdense. Point B will decelerate *less* rapidly than it would it the entire circumscribed region were comparably overdense. But since the inscribed and circumscribed spherical regions would collapse at the same rate, the ellipsoid becomes flatter with time. Now take the ellipsoidal region in the top figure to be underdense. Point A will decelerate less rapidly than it would if only the inscribed sphere were comparably underdense. Point B will decelerate *more* rapidly than it would if the entire circumscribed region were comparably overdense. But since the inscribed and circumscribed spherical regions would collapse at the same rate, the ellipsoid becomes rounder with time.

(although in the literature one often sees 4/7 approximated as 0.6).

The nonlinear growth regime is said to begin at the point where $\eta_p \sim \pi$, which is also the turn-around point for an over-dense perturbation that stops expanding and begins to collapse. Again by setting $t_u = t_p$, we find the over-density of the perturbation at $r_{\rm max}$:

$$\left(\frac{1}{6\pi G\rho_u}\right)^{1/2} = \frac{\pi}{2}\left(\frac{3}{8\pi G\rho_{p,\rm max}}\right)^{1/2} \Rightarrow \frac{\rho_{p,\rm max}}{\rho_u} = \frac{9\pi^2}{16}. \tag{3.34}$$

The collapse itself is complete when $\eta_p = 2\pi$, which gives a singularity in the density $[r_p(2\pi) = 0]$. However, for real astrophysical objects, the random particle motions acquired from the collapse contribute a positive kinetic energy $K_{\rm eq}$ to the object at equilibrium. Recalling the virial theorem from Chapter 1, we set the total energy at equilibrium to be the total potential energy at the turn-around point $W_{\rm max}$, where the kinetic energy is zero.

$$W_{\rm eq} + K_{\rm eq} = \frac{1}{2}W_{\rm eq} = W_{\rm max} \tag{3.35}$$

For a constant mass M, the gravitational potential energy of a system goes as $\sim M^2/r$, so equation (3.35) implies that $r_{\rm eq} = r_{\rm max}/2$, and the collapsed density is $\rho_{p,\rm collapse} = 8\rho_{p,\rm max}$. During this time of collapse from $\eta_p = \pi \rightarrow 2\pi$, the background universe continued to expand. Combining equations (3.27) and (3.34), we find the background density at time of collapse

$$\frac{\rho_{u,\rm collapse}}{\rho_{u,\rm max}} = \frac{1}{4}, \tag{3.36}$$

giving the density contrast of the final collapsed object to be

$$\left(\frac{\rho_p}{\rho_u}\right)_{\rm collapse} = 18\pi^2. \tag{3.37}$$

A qualitative explanation for the shapes of these collapsed objects (or alternatively expanded underdense regions) is presented in Figure 3.6.

3.4 Recombination

While the total density parameter Ω can possibly be determined by global measurements of the universe's expansion, the individual components in matter (Ω_m), radiation (Ω_r), and "normal" baryonic matter (Ω_b) can be determined more directly by local measurements. The critical density, where $\Omega = 1$, is easily calculated in familiar units:

$$\rho_{\rm crit} = \frac{3H_0^2}{8\pi G} = 1.88 \times 10^{-29} h^2 \ {\rm g/cm}^3, \tag{3.38}$$

where h is the current value of the Hubble parameter in units of 100 km/s/Mpc (from recent observations $h = 0.72$). If the density is entirely in baryons, then $\rho_{\rm crit}$ corresponds to a number density of $n_b = 1.13 \times 10^{-5}$ cm^{-3}, of course averaged over all of intergalactic space. This is a tricky measurement to make since we cannot observe mass directly, but only infer it from the observable luminosity coming from stars and galaxies. From the galaxy luminosity function introduced in Section 1.3, the total density in galaxies is

$$\rho_{\rm gal} = \left\langle \frac{M}{L} \right\rangle \int \Phi(L) L dL = \left\langle \frac{M}{L} \right\rangle \Phi^\star L^\star \Gamma(\alpha + 2) = \left\langle \frac{M}{L} \right\rangle 1.05 \times 10^8 h L_\odot \ {\rm Mpc}^{-3}. \tag{3.39}$$

Within gaaxies measurements of baryonic mass in gas and dust and stars, the average mass-to-light ratio is

$$\left\langle \frac{M}{L} \right\rangle_{\rm gal} \sim 10 \frac{M_\odot}{L_\odot}, \tag{3.40}$$

(or perhaps even smaller) giving a density parameter for galactic baryons of $\rho_{b,{\rm gal}} = 7 \times 10^{-32} h$ g/cm^3 or $\Omega_{b,{\rm gal}} = 0.004$, well below the critical value needed to bound the universe gravitationally. Even if we include all the dark matter indirectly observed in clusters of galaxies(see Section 2.4), with

$$\left\langle \frac{M}{L} \right\rangle_{\rm clus} \approx 350 h^{-1} \frac{M_\odot}{L_\odot}, \tag{3.41}$$

dynamic measurements still only give $\Omega_m = 0.20 \pm 0.05$.

Well, what about the radiation? As Einstein famously showed in 1905, mass and energy are interchangeable, so the enormous number of cosmic background photons should also slow the expansion of the universe. This is a quantity more easily measured directly, since the energy density of a blackbody radiation field is

$$\rho_r = \frac{\sigma T_r^4}{c^3}, \tag{3.42}$$

where σ is the Stefan-Boltzmann constant. For the observed CMB temperature of $T_r = 2.728$ °K, this gives $\rho_r = 4.7 \times 10^{-34}$ g/cm^3; $\Omega_r = 3.5 \times 10^{-5}$, which we can safely neglect when calculating present-day expansion dynamics. But this was not always the case. Looking back in time, when the universe was physically smaller (or at least everything was closer together), the radiation density

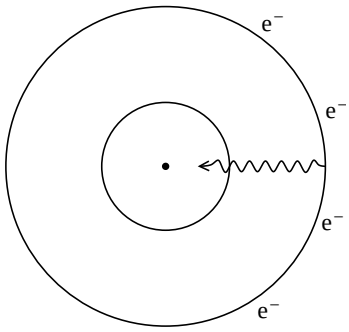

Figure 3.7: A photon heads toward the Milky Way from the "surface of last scattering" at redshift $(1+z) \approx 1500$. If reionization occurs too soon, the photon will be re-scattered

was higher, even more so than the matter density. Thinking of the CMB as a collection of particles (photons) in a box (the universe) with scale length a, it is clear that as the box expands, the density of particles decreases as $n \sim a^{-3}$, just like the massive particles. However, unlike the massive particles, the individual photons are redshifted with the expansion, losing energy and thus mass like $E \sim a^{-1}$, so the photon energy density decreases with an expanding universe as $\rho_r \sim a^{-4}$.

Since the expansion is directly related to the expansion factor $(1+z) = a^{-1}$, we can solve for the radiation density as a function of redshift (our favorite measurement of time in cosmology):

$$\frac{\rho_r}{\rho_m} = \frac{\rho_{r0}}{\rho_{m0}} \frac{(1+z)^4}{(1+z)^3}. \tag{3.43}$$

Simply extrapolating back in time tells us that while today the universe appears to be matter-dominated, at a redshift of $(1+z) \sim 4000$ and greater, the universe was radiation-dominated. As the universe expanded and cooled, the background photons eventually were redshifted below the point where they had enough energy to ionize a hydrogen atom (the well-known Rydberg: 13.6 eV). So at some point in the expansion, the hot plasma of ionized hydrogen that made up the early universe should have cooled sufficiently for the protons and electrons to combine into neutral hydrogen without getting immediately reionized by the energetic background photons.

To calculate the redshift of this recombination (technically it should be called *combination*, but the widely-accepted term is *re*combination), we need to introduce the *Saha equation*. For any reversible physical or chemical reaction in thermal equilibrium at a temperature T such as

$$A + B \rightleftharpoons C + \chi(\text{energy}), \tag{3.44}$$

where χ is the amount of energy released in the reaction, the relative concentrations of the three species (n_A, n_B, n_C) can be determined by the Saha equation:

$$\frac{n_A n_B}{n_C} = \frac{g_A g_B}{g_C} \left(\frac{2\pi k T}{h^2} \frac{m_A m_B}{m_C} \right)^{3/2} \exp\left(\frac{-\chi}{kT} \right). \tag{3.45}$$

Here g_i is the quantum degeneracy of the i^{th} species and h is Plank's constant (*not* the scaled Hubble parameter). For ionized hydrogen nuclei (protons) and electrons, both of which are fermions with half-integer spin, $g_{H^+} = 2$ and $g_{e^-} = 2$ and when combined, the total number of spin combinations is $g_{H^0} = 4$. At the time of hydrogen recombination, it is safe to say that practically all of the primordial helium (which has a higher ionization potential $\chi_{\rm He}$) has already recombined with free electrons and is thus ignorable in the Saha equation.

While the process of recombination actually takes place over a finite period of time, for definiteness we will define the instantaneous event of recombination as the point where $n_{H^+} = n_{H^0}$: the number densities of neutral and ionized hydrogen are equal. The total (free and bound) number density of electrons today is related to the total number of baryons by

$$n_{e,\mathrm{tot}}(z=0) = \frac{3}{4} n_{b,\mathrm{tot}}(z=0), \tag{3.46}$$

approximating mass fractions as $X = 0.75$ and $Y = 0.25$. Assuming conservation of electron number, the total number density of electrons at a redshift $(1+z)$ was

$$n_{e,\mathrm{tot}}(z) = (1+z)^3 n_{e,\mathrm{tot}}(0). \tag{3.47}$$

The temperature of the cosmic background radiation changes linearly with redshift so

$$T(z) = (1+z)T(0). \tag{3.48}$$

From the recombination criterion of $n_{H^+} = n_{H^0}$, we have

$$n_{e,\mathrm{free}}(z_{\mathrm{rec}}) = \frac{1}{2} n_{e,\mathrm{tot}}(z_{\mathrm{rec}}). \tag{3.49}$$

Plugging into equation (3.45), we can solve for the redshift of recombination:

$$\frac{n_{e,\mathrm{tot}}(0)}{2}(1+z_{\mathrm{rec}})^3 = \frac{2\cdot 2}{4}\left(\frac{2\pi k T(0)(1+z_{\mathrm{rec}})m_e}{h^2}\right)^{3/2} \exp\left(\frac{-\chi}{kT(0)(1+z_{\mathrm{rec}})}\right). \tag{3.50}$$

The only remaining unknown is $n_{e,tot}(0)$, the total electron density in the universe today. For simplicity, consider $\Omega_b = 1$, which gives $n_{e,tot}(0) = 8.5 \times 10^{-6}$ cm^{-3} and then a recombination redshift $(1 + z_{\mathrm{rec}}) \approx 1510$. Taking the recent results of the WMAP satellite, we have $\Omega_b = 0.047$, and $(1 + z_{rec}) = 1090$. Note that the temperature at recombination was 3000 K, giving a temperature smaller by almost a factor of 100 than the temperature corresponding to the ionization energy of hydrogen. The phase space factors strongly favor ionization.

Recall from Section 1.11 the derivation of the *Jeans mass*, the minimum mass necessary for an overdense perturbation to collapse under its own gravity:

$$M_J = \frac{4}{3}\pi\rho\left(\frac{\pi v_s^2}{G\rho}\right)^{3/2}.$$

Before recombination, the ionized hydrogen gas is opaque to the background photons, so the dominant source of pressure in the gas comes from the photons which are relativistic particles, giving a very large sound speed $v_s \approx c/\sqrt{3}$ and thus large Jeans mass. After recombination, the sound speed drops to that of an ideal gas of massive, non-relativistic particles with $v_s \approx 1$ km/s. So the corresponding Jeans mass also changes dramatically during recombination.

$$\begin{aligned} (1+z) > (1+z_{\mathrm{rec}}) \quad &: \quad v_s \sim \frac{c}{\sqrt{3}} \\ &\quad\;\; M_J \sim 10^{16} M_\odot \\ (1+z) < (1+z_{\mathrm{rec}}) \quad &: \quad v_s \sim 1 \text{ km/s} \\ &\quad\;\; M_J \sim 10^{5} M_\odot \end{aligned} \tag{3.51}$$

It would appear that galaxies with $M_{\rm gal} \sim 10^9 - 10^{12} M_\odot$ and even clusters of galaxies with masses up to $M_{\rm clus} \sim 10^{15} M_\odot$, are too small to have collapsed before recombination. Objects with masses smaller than the Jeans mass have imaginary growth exponents – they oscillate. This creates a major problem in models without dark matter, with only photons and baryons. The photons and baryons oscillate (see Fig. 3.14 below) but in the process the photons leak out of the perturbation. Letting $\lambda_{\rm rec}$) be the mean free path of a photon at recombination, photons will leak out of masses smaller than the *Silk* mass,

$$M_S \sim \left[\lambda_{\rm rec} \begin{pmatrix} \text{number of scattering} \\ \text{events in } t \sim 1/H_0 \end{pmatrix}^{1/2}\right]^3 \cdot [\text{density}] \sim 10^{12} M_\odot \quad . \tag{3.52}$$

The Silk mass is yet larger for low Ω. This would seem to imply that perturbations on the scale of galaxies could not have survived Silk damping prior to recombination. Zeldovich and his school therefore favored a scenario in which clusters formed first, collapsing to "pancakes" which then fragment into galaxies. But dark matter couples only gravitationall to the photons and baryons, and dark matter perturbations continue to grow, albeit slowly, prior to recombination. Dark matter obviates the need for the Zeldovich "top down" scenario.

3.5 From Fields to Objects: The Press-Schechter Recipe

Linear perturbation theory is quite useful for determining growth rates at early times, but what do we do when a whole field of fluctuations transitions to non-linear growth, precipitating objects of various sizes? Also, for a given wave-number, some large fluctuations will have had time to grow into objects (e.g. galaxy, cluster, etc.), while others would still be in the early linear growth phases even today. We want to understand how to make the transition from the linear problem of a superposition of density wave perturbations in an otherwise uniform fluid to the fully non-linear description of discrete objects of varying masses distributed throughout empty space. In the analysis that follows, we will consider primarily the perturbations in the dark matter fields that grow through self-gravitation, largely independent of the ionization of the baryons, i. e. we will ignore the effects of recombination.

Consider a density perturbation in the early universe with magnitude

$$\delta \equiv \frac{\rho_p - \rho_u}{\rho_u}.$$

This perturbation will grow according to the linear theory presented above, but at an amplitude δ_c non-linear effects become important. By convention δ_c is taken to be 1.69, which is the amplitude a perturbation would have at the time of maximum expansion in and $\Omega = 1$ universe if it had grow with only the linear growth rate. Including the non-linear effects gives $\delta = 9\pi^2/16 - 1$.

We assume the initial density perturbations in the early universe were like superpositions of plane waves with wave-number k, random phase, and amplitude δ. The probability distribution of amplitudes δ is assumed to be Gaussian with a variance σ^2, and the density perturbations can have $\delta < 0$ (underdense) or $\delta > 0$ (overdense). We will primarily be interested in the positive density perturbations, since those are the ones that can form objects. Each value of k corresponds to a typical volume and thus mass M determined by the initial background density $\rho_{u,i}$. For each mass

M there is a whole range of superposed random-phase modes with different values of δ determined by the normal distribution. Thus we can define the variance as a function of mass:

$$\sigma^2(M) = \delta^2 = \left\langle \left(\frac{\delta\rho}{\rho} \right)^2 \right\rangle_M . \tag{3.53}$$

In the limit of small σ^2, the distribution of δ looks like the Dirac delta-function centered on $\delta = 0$, giving no perturbations at that mass scale. The larger the value of $\sigma^2(M)$, the more structure will evolve with mass M. There is a straightforward connection between the variance on scale M and the power spectrum $P(k)$ at wavenumber k,

$$P(k) \propto k^n \quad . \tag{3.54}$$

From these initial conditions, we want to estimate the fraction of masses $F(M)$ which have $\delta > \delta_c$ and have thus collapsed to form bound objects by the present epoch. This fraction simply is the integral of the upper tail of the normalized distribution function:

$$F(M) = \frac{1}{\sqrt{2\pi}\sigma(M)} \int_{\delta_c}^{\infty} \exp\left(\frac{-\delta^2}{2\sigma^2(M)} \right) d\delta = \frac{1}{\sqrt{\pi}} \int_{q_c}^{\infty} e^{-q^2} dq = \frac{1}{2}\mathrm{erfc}(q_c), \tag{3.55}$$

where we have made the substitution $q \equiv \frac{\delta}{\sqrt{2}\sigma(M)}$ and $\mathrm{erfc}(x)$ is the complimentary error function.

The characteristic mass M^* is that which corresponds to

$$\sigma^2(M^*) = \frac{1}{2}\delta_c^2. \tag{3.56}$$

The power-law relation of equation (3.54) gives a variance of

$$\sigma^2(M) = AM^{-(3+n)/3} = \frac{\delta_c^2}{2} \left(\frac{M}{M^*} \right)^{-(3+n)/3} , \tag{3.57}$$

giving further

$$q_c = \left(\frac{M}{M^*} \right)^{-(3+n)/n} . \tag{3.58}$$

At this point Press and Schechter equate two quantities which are clearly not the same thing,

$$\begin{pmatrix} \text{fraction of} \\ \text{masses } M \\ \text{with } \delta > \delta_c \end{pmatrix} \approx \begin{pmatrix} \text{fraction of universe} \\ \text{collapsed in objects} \\ \text{with masses } > M \end{pmatrix} \quad . \tag{3.59}$$

This *ansatz* (althought swindle might be more appropriate – see Fig. 3.8) allows one to compute the fractional number density of collapsed objects with masses in the range $(N, N + dM)$,

$$N(M)dM = \frac{-\bar{\rho}}{M} \frac{\partial F}{\partial M} dM, \tag{3.60}$$

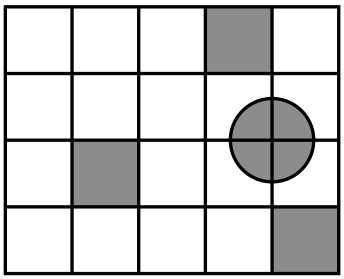

Figure 3.8: The Press-Schechter recipe starts with the probability that a cell on a predetermined grid is sufficiently dense to collapse. But what about objects that span the cell boundaries?

where $\bar{\rho}/M$ is the number of possible objects with mass M per unit volume. Note the additional negative sign in equation (3.60) that is needed because number densities are always non-negative and $F(M)$ is always non-increasing, so $\frac{\partial F}{\partial M}$ is non-positive. More precisely,

$$-\frac{\partial F}{\partial M} = -\frac{1}{\sqrt{\pi}}\frac{dq}{dq_c}\frac{\partial q_c}{\partial M} = \frac{1}{\sqrt{\pi}}\exp\left[-\left(\frac{M}{M^*}\right)^{\frac{n+3}{6}}\right]\left(\frac{n+3}{6}\right)\frac{1}{M}\left(\frac{M}{M^*}\right)^{\frac{n+3}{6}}. \tag{3.61}$$

Combining equations (3.60) and (3.61), we get

$$N(M)dM = \frac{1}{2\sqrt{\pi}}\left(1+\frac{n}{3}\right)\frac{\bar{\rho}}{M^2}\left(\frac{M}{M^*}\right)^{\frac{n+3}{6}}\exp\left[-\left(\frac{M}{M^*}\right)^{\frac{n+3}{6}}\right], \tag{3.62}$$

which actually normalizes to

$$\int_0^\infty N(M)dM = \frac{1}{2}, \tag{3.63}$$

and *not* to unity, as should be expected since all collapsed objects have less than infinite mass. This factor of 1/2 is can be explained by invoking an argument based on the "excursion sets," a topic outside the scope of this discussion.

The P-S recipe assumes the instantaneous destruction of substructure when a new structure forms. This is obviously not the case – galaxies continue to exist and flourish within clusters of galaxies. Best estimates are that roughly 10% of the dark matter substructure survives the collapse of the next generation in the hierarchy, but this is very uncertain. A much larger fraction of the baryonic structure survives because the baryons form stars at the bottoms of the dark matter potential wells.

The shape of the $N(M)$ distribution for the currently favored perturbaton spectrum differs significantly from the shape of luminosity function for galaxies. The galaxy distribution cuts of more sharply at the bright end and is shallower at the faint end. One might suppose that this is the result of the crude approximations involved, but N-body simulations give results remarkably similar to the P-S recipe. The distribution of baryon masses does indeed appear to be different. Much has been made lately of the large number of low mass dark matter subsystems expected to persist in the dark matter halos of galaxies. These far exceed the observed numbers of dwarf galaxies. Star formation may have failed to occur in these sub-halos.

How does the critical mass M^* change with time? From simple scaling comparisons, we see that

for an overdense region in an expanding universe,

$$\begin{aligned} t_c &\sim \delta_c^{-3/2} \sim \eta^{-3} \\ t &\sim t_c \eta^3 \\ \delta &\sim \eta^2 \\ M^* &\sim t_c^{6/n+3} \sim (1+z)^{-2}. \end{aligned} \tag{3.64}$$

This seems to suggest that for high redshift, the characteristic collapsed mass should be relatively small, but a number of very massive clusters of galaxies (large collapsed objects) have been found at $z \sim 0.8$!

Voit has presented a simple approach to calculating M^* His reasoning goes as follows:

- Use the currently observed values of $N(M)$ to derive (M/M^*).
- Extrapolate M^* to some observed redshift z where we can evaluate $\exp(-M/M^*)$.
- Use the temperature T as the independent variable instead of the mass M. The densities of the most recently collapsed objects vary as $(1+z)^3$, so the radii of of these same objects (at fixed mass) vary as $(1+z)^{-1}$. The observed temperatures vary as M/R so one finds that $T \sim (1+z)$ at fixed mass.

The somewhat surprising result that clusters of galaxies are no bigger today than they were at $z \sim 0.8$ is resolved if one drops the assumption (common until the discovery of the acceleration of the universe) that $\Omega = 1$. In low *Omega* universes, growth rapidly ceases once Ω deviates significantly from unity.

3.6 Adding General Relativity to Cosmology

Using only the Newtonian law of gravity and the concepts of absolute time and space, we have been able to construct a remarkably good model for the expansion of the universe and the formation of structure from the collapse of initial density perturbations. However, Einstein showed that these basic Newtonian concepts can break down in the regime of high velocities, high energy and mass densities, and large distances where the finite speed of light becomes more important. Replacing a world where space and time are absolute and objects travel in straight lines at constant velocity, Einstein worked in a world where space and time were mixed together in a curved manifold. There are no longer global inertial reference frames, but all objects still follow straight trajectories when viewed in free-fall, under the influence of only gravity.

The mathematics behind Einstein's general theory of relativity is elegant and sophisticated, and would take up an entire course in itself. Fortunately, we can make a good deal of progress towards a general relativistic cosmological model with the addition of just a couple key ideas:

- In the absence of non-gravitational forces, particles (including photons), follow curved paths through space-time called geodesics, curves whose tangent vectors are "parallel transported."

- Mass density, energy density, and pressure, all contribute contribute the curvature of that space time. gravity.

To accurately measure distances and motion in curved space-time, we need to use the *metric tensor* $g_{\mu\nu}$. The geodesic equations of motion can be determined entirely from the metric and its first derivatives. The infinitesimal distance between two space-time events is given by

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu. \tag{3.65}$$

Here we have used the implicit sum notation over repeated upper and lower indices. For the more familiar special relativistic form of equation (3.65), the distance ds^2 is known as the Lorentz invariant distance and is the same in all reference frames (hence "invariant"). The metric in these flat Lorentz frames takes a special symbol $\eta_{\mu\nu}$. In cartesian coordinates,

$$\eta_{\mu\nu}(t,x,y,z) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.66}$$

and in spherical coordinates,

$$\eta_{\mu\nu}(t,r,\theta,\phi) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta \end{pmatrix}. \tag{3.67}$$

Here and for the rest of our relativistic cosmology discussion, we adopt units where the speed of light $c = 1$. Then the proper time $d\tau$ (the time measured by a clock moving with the particle) for a particle moving in flat space-time is given by

$$d\tau^2 = -ds^2 = dt^2 - dx^2 - dy^2 - dz^2 \tag{3.68}$$

so for a photon with velocity $dl/dt = 1$, the proper time elapsed along its trajectory is zero in *any* reference frame.

The metric for non-flat spacetimes is in general quite complicated. There are, however, several simple metrics, the simplest perhaps being Schwarzschild metric for the vacuum outside a spherically symmetric mass. It was derived by Karl Schwarzschild in 1916, just a year after Einstein's original formulation of General Relativity. And as we saw above with the introduction of Birkhoff's theorem, there are a great many astrophysical systems that are made up of spherically symmetric mass distributions. For an interior total mass M, the Schwarzschild metric is

$$g_{\mu\nu} = \begin{pmatrix} -(1-\frac{2GM}{r}) & 0 & 0 & 0 \\ 0 & (1-\frac{2GM}{r})^{-1} & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2\sin^2\theta \end{pmatrix}. \tag{3.69}$$

With this metric, the radial coordinate r should not be thought of as a conventional distance, but it does have the property of giving the area of a section of space with constant r by

$$A = r^2 \sin\theta d\theta d\phi. \tag{3.70}$$

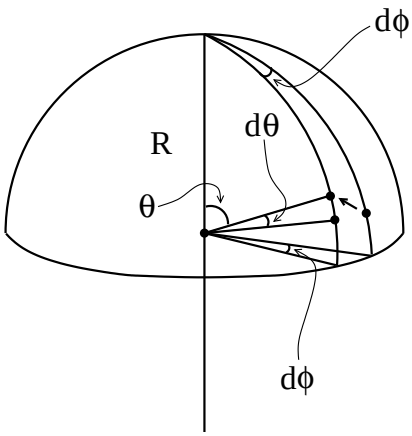

Figure 3.9: Calculating the infinitesimal distance *ds* between two points on the surface of a sphere.

As mentioned above, the curvature described by the metric $g_{\mu\nu}$ is created by mass, energy, and pressure, as incorporated in the *stress-energy tensor*, which can be described by a scalar density, vector momentum, and 3×3 tensor of momentum flux:

$$T_{\mu\nu} = \begin{bmatrix} \rho & \rho\vec{\mathbf{v}} \\ \rho\vec{\mathbf{v}} & \vec{\nabla}(\rho\vec{\mathbf{v}}) \end{bmatrix}. \tag{3.71}$$

The upper left term is just the mass-energy density. The lower right term is the momentum density flux. The off diagonal terms are the momentum density.

Specifying the stress-energy tensor is relatively easy. Computing the associated curvature is more difficult. An *Einstein tensor* $G_{\mu\nu}$, is constructed from the metric tensor and its derivatives, It is the general relativistic analog of the gravitational potential. The complications are hidden in the construction of the Einstein tensor, which is related to the stress-energy tensor by the deceptively simple *field equations*,

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \quad , \tag{3.72}$$

While Newtonian gravity, as described by the global gravitational potential has no real analogy in general relativity, the field equations do reduce directly to Poisson's equation $\nabla^2\Phi = 4\pi G\rho$ in the non-relativistic limit. Since the mass density ρ is the leading non-relativistic term in the tensor component T_{00}, the Newtonian potential Φ can be then compared to the leading order curvature terms R_{00} and g_{00}.

As suggested by the curvature singularity in the Schwarzschild metric at $r = 0$, it does *not* represent a spatially homogeneous space, but rather has a physically distinguishable origin. Since one of our central cosmological assumptions is the homogeneity of space, we would like a metric in which all points in the space appear physically identical. The above forms of the Minkowski metric [eqns. (3.66) and (3.67)] are homogeneous, but also geometrically flat, so are of limited practical interest. Better, consider the metric on the surface of a smooth sphere of radius R. Adopting the standard spherical coordinates (θ, ϕ), we can see from Figure 3.9 that the infinitesimal spatial distance dl is given by

$$dl^2 = R^2(d\theta^2 + \sin^2\theta d\phi^2) \quad . \tag{3.73}$$

We define the "longitudinal distance" coordinate χ as the distance along a great circle from the

pole so $\chi \equiv R\theta$ and the spatial metric becomes

$$dl^2 = d\chi^2 + R^2 \sin^2\left(\frac{\chi}{R}\right) d\phi^2 \quad . \tag{3.74}$$

For a surface of positive curvature, like a sphere, the "angular diameter distance," x, is defined such that the angle subtended by an object, when multiplied by the angular diameter distance, gives the physical size of the object. Thus

$$x \equiv R \sin\left(\frac{\chi}{R}\right) \tag{3.75}$$

and

$$dx^2 = \cos^2\left(\frac{\chi}{R}\right) d\chi^2. \tag{3.76}$$

The spatial metric is then given by

$$dl^2 = \frac{dx^2}{1 - x^2/R^2} + x^2 d\phi^2. \tag{3.77}$$

Homogeneous, two-dimensional spaces can also have negative curvature, although they are somewhat more difficult to visualize than the surface of a sphere. Often described as saddle geometries, they have the same form of the spatial metric only with the following substitutions:

$$\begin{aligned} \sin^2 &\rightarrow \sinh^2 \\ 1 - \frac{x^2}{R^2} &\rightarrow 1 + \frac{x^2}{R^2}. \end{aligned} \tag{3.78}$$

The 3-D analogue of the 2-sphere can be described with a combination of either one length and two angles (longitudinal and transverse) or three angles and one constant radius of curvature. In terms of three angles (χ, θ, ϕ) and a radius R, the metric looks like

$$dl^2 = d\chi^2 + R^2 \sin^2\left(\frac{\chi}{R}\right) [d\theta^2 + \sin^2\theta d\phi^2]. \tag{3.79}$$

Combining with a time-like coordinate t and defining a scale factor $a(t) \equiv R(t)/R_0$ [with present-day scale $a(t_0) = 1$], we get the space-time metric of a homogeneous closed universe in terms of the comoving angular diameter distance x:

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dx^2}{1 - x^2/R_0^2} + x^2(d\theta^2 + \sin^2\theta d\phi^2)\right]. \tag{3.80}$$

3.7 Cosmography

To begin our more in-depth discussion of cosmography (the study of distances in cosmology), let us review the types of distance measurements already introduced. Different text books and papers unfortunately use different conventions, but at this point there is nothing to do but make sure we have our own terms well-defined and internally consistent.

- χ : comoving longitudinal distance

- x : comoving transverse distance
- $a = (1+z)^{-1}$: scale factor
- $D_A \equiv ax$: angular diameter distance
- D_L : luminosity distance
- R_0 : present-day radius of curvature

For a spatially homogeneous universe, the Einstein field equations reduce to the *Friedmann equations*, which govern the dynamics of the expansion. Formulated in terms of the scale factor $a(t)$, the Friedmann equations are

$$\ddot{a} = -\frac{4\pi}{3}G\left(\rho + \frac{3p}{c^2}\right) + \frac{1}{3}\Lambda a, \tag{3.81}$$

and

$$\dot{a} = -\frac{8\pi G\rho}{3}a^2 \mp \frac{c^2}{R_0^2} + \frac{1}{3}\Lambda a^2, \tag{3.82}$$

where ρ is the density, p the pressure. Λ the so-called "cosmological constant." The pressure term represents a fundamental departure from Newtonian gravity, Einstein added the cosmological constant to the curvature side of zero out the acceleration of the universe implicit implicit in the Friedmann equations, since the universe was, at the time, thought to be static. Today something like the cosmological constant may appear on the stress-energy side of the equation, in the form of a vacuum energy density

Recall the first law of thermodynamics, with the internal energy U in a volume V given by $\rho c^2 V$:

$$dU + pdV = c^2 V d\rho + c^2\rho dV + pdV = 0 \Rightarrow d\rho + \left(\rho + \frac{p}{c^2}\right)dV = 0. \tag{3.83}$$

Since the volume grows as $V \sim a^3$,

$$\frac{dV}{V} = 3\frac{da}{a}, \tag{3.84}$$

so combining with equation (3.83),

$$a\frac{d\rho}{da} + 3\left(\rho + \frac{p}{c^2}\right) = 0. \tag{3.85}$$

Multiplying equation (3.81) by $2\dot{a}$ and using (3.85) to eliminate the pressure term, we have

$$2\dot{a}\ddot{a} = \frac{8\pi}{3}G\rho\left(2\rho a\dot{a} + \frac{d\rho}{da}a^2\dot{a}\right) + \frac{2}{3}\Lambda a\dot{a}. \tag{3.86}$$

Integrating equation (3.86) gives

$$\dot{a}^2 = \frac{8\pi}{3}G\rho a^2 + \frac{1}{3}\Lambda a^2 - k \quad . \tag{3.87}$$

which looks much like equation (3.82) except that the latter involves the radius of curvature of the universe, R_0, while the former gives us no information about the scale on which the universe is curved.

Both the matter density and radiation density decrease with the expansion ($\rho_{\rm matter} \sim a^{-3}$; $\rho_{\rm rad} \sim a^{-4}$), but the vacuum energy density remains constant ($\rho_\Lambda \sim a^0$), hence the term cosmological *constant*. In the limit where the expansion is dominated by Λ, equation (3.87) becomes

$$\dot{a} \approx \sqrt{\frac{\Lambda}{3}} a, \tag{3.88}$$

giving infinite, exponential expansion

$$a(t) \sim \exp\left(\sqrt{\frac{\Lambda}{3}} t\right). \tag{3.89}$$

This is thought to be the behavior of the very early universe during the inflationary period and may also be the ultimate fate of the expansion if there is in fact a non-zero Λ today.

At early times, yet after the completion of the inflationary period, the universe can be described as a gas of photons and ultra-relativistic massive particles. For such a gas, the equation of state is

$$p = \frac{c^2}{3}\rho, \tag{3.90}$$

so equation (3.85) gives

$$\frac{d\rho_{\rm rad}}{\rho_{\rm rad}} = -4\frac{da}{a}, \tag{3.91}$$

consistent with the relation presented above ($\rho_{\rm rad} \sim a^{-4}$). For a universe with zero curvature and negligible Λ, equations (3.87) and (3.91) give

$$\left(\frac{\dot{a}}{a}\right)^2 = -\frac{1}{4}\left(\frac{\dot{\rho}_{\rm rad}}{\rho_{\rm rad}}\right)^2 = \frac{8\pi}{3} G \rho_{\rm rad}, \tag{3.92}$$

which can be integrated to give

$$\rho_{\rm rad}(t) = \frac{3}{32\pi G} t^{-2}. \tag{3.93}$$

Note the similarity to the expression for matter-dominated expansion in equation (3.22).

Recall the cosmological density parameter Ω_0 introduced in Section 3.1:

$$\Omega_0 \equiv \frac{8\pi G \rho_0}{3H_0^2}.$$

Similarly, we can define multiple density parameters for each of the "ingredients" of the universe: matter, radiation, and a cosmological constant. Again using the subscript 0 to indicate present-day values, we have

$$\Omega_{m0} = \frac{8\pi G \rho_{m0}}{3H_0^2}, \tag{3.94a}$$

$$\Omega_{r0} = \frac{8\pi G \rho_{r0}}{3H_0^2}, \tag{3.94b}$$

$$\Omega_{\Lambda 0} = \frac{\Lambda}{3H_0^2}. \tag{3.94c}$$

With a time-varying value of the Hubble parameter $H(t) = \dot{a}/a$, equation (3.87) becomes

$$\dot{a}^2 = \frac{H_0^2\Omega_{m0}}{a} + \frac{H_0^2\Omega_{r0}}{a^2} + H_0^2\Omega_{\Lambda 0}a^2 \mp \frac{c^2}{R_0^2}. \tag{3.95}$$

From here, it is easy to derive an expression for the present-day ($a_0 = 1$) geometry of the universe, in terms of the curvature radius R_0:

$$\mp R_0^2 = \frac{(c/H_0)^2}{1 - \Omega_{m0} - \Omega_{r0} - \Omega_{\Lambda 0}}. \tag{3.96}$$

If the Ω's add up to unity, we get a flat universe with infinite radius of curvature. The departure from flatness is sometimes described by Ω_k, the "curvature density parameter":

$$\Omega_k \equiv 1 - \sum_i \Omega_i. \tag{3.97}$$

We can now apply all the above results from the Friedmann equation to the problem of measuring distances in an expanding metric. It is often easiest to think of infinitesimal distance changes over length scales that are small compared to the curvature and time scales that are short compared to the age of the universe. The *comoving longitudinal distance* is a measure of how long a distant, ancient meter stick would appear today. This is *longer* than the distance the photon itself would have measured along its path with an odometer because the space that the photon moved through a billion years ago has since expanded and moved further away from the observer. Because of this, it is sometimes called the "tachyon distance," the distance that would be with an odometer by a particle moving with infinite velocity. From equations (3.80) and (3.76), the comoving longitudinal meter stick is given by

$$\begin{aligned} d\chi &= \frac{cdt}{a(t)} = c(1+z)dt = \frac{-cdz}{(1+z)\dot{a}} \\ &= \frac{-cdz}{H_0}[\Omega_{m0}(1+z)^3 + \Omega_{r0}(1+z)^4 + \Omega_{\Lambda 0} + \Omega_{k0}(1+z)^2]^{-1/2} \end{aligned} \tag{3.98}$$

and then the total comoving longitudinal distance between two events at redshifts z_1 and z_2 is

$$\chi_{12} = \int_{z_1}^{z_2} d\chi. \tag{3.99}$$

Another distance measurement that is very important in observational cosmology is the *luminosity distance* D_L, which is the apparent distance to an object of a given bolometric luminosity $L_{\rm bol}$ and observed fluc $f_{\rm bol}$, assuming a Newtonian inverse-square law for flux traveling through a flat, static space:

$$f_{\rm bol} = \frac{L_{\rm bol}}{4\pi D_L^2}. \tag{3.100}$$

The *angular diameter distance* is also defined with respect to the appearance of an object in flat space. An object of proper size l at an angular diameter distance D_A will appear to subtend an angle in the sky of $d\phi = l/D_A$. As is shown in Figure 3.10, the curvature of light rays due to the lensing by

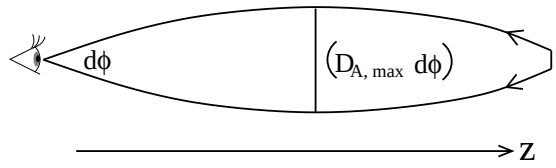

Figure 3.10: An object subtending an angle $d\phi$ will have a different physical size at different redshifts. The angular diameter distance D_A transforms angle into length.

intervening mass can cause D_A to actually decrease with increasing redshift. The luminosity distance is related to the angular diameter distance D_A and the comoving angular distance x through

$$D_L = \frac{D_A}{1+z} = \frac{x}{(1+z)^2}. \tag{3.101}$$

For the case of a closed, matter-dominated universe ($\Omega_r = \Omega_\Lambda = 0, \Omega_m > 1$), analytic expressions exist for the comoving distances χ and x and the volume element dV:

$$\chi = \frac{2c}{H_0(\Omega_{m0}-1)^{1/2}} \left[\tan^{-1}\left(\frac{1+\Omega_{m0}z}{\Omega_{m0}-1}\right)^{1/2} - \tan^{-1}\left(\frac{1}{\Omega_{m0}-1}\right)^{1/2} \right], \tag{3.102}$$

$$x = \frac{2c}{H_0\Omega_{m0}^2(1+z)} \left[\Omega_{m0}z + (\Omega_{m0}-z)((\Omega_{m0}z+1)^{1/2}-1)\right], \tag{3.103}$$

sometimes called the Mattig formula. The comoving volume element dV for a section of solid angle $d\Omega$ is

$$dV = d\Omega R_0^2 \sin^2\left(\frac{\chi}{R_0}\right) d\chi = d\Omega R_0^2 \left(\frac{1}{2} - \frac{1}{2}\cos\frac{2\chi}{R_0}\right) d\chi \tag{3.104}$$

which integrates to give a comoving volume for a cone extending away from the observer a distance χ:

$$V = d\Omega \frac{R_0^3}{2}\left(\frac{\chi}{R_0} - \frac{1}{2}\sin\frac{2\chi}{R_0}\right) = d\Omega\frac{R_0^3}{2}\left(\sin^{-1}\frac{x}{R_0} - \frac{x}{R_0}\sqrt{1-(x/R_0^2)}\right). \tag{3.105}$$

3.8 Cosmological Parameters

In many of the formulae above, we have repeatedly referred to various cosmological parameters like the Hubble "constant" H_0 and the density parameters Ω_{m0}, Ω_{r0}, and $\Omega_{\Lambda 0}$. Since we understand how the different densities evolve with redshift, the present-day values of the Ω's is enough to extrapolate backwards in time to give the density and thus dynamics of the universe at an earlier time (higher redshift). Without knowing the values of these cosmological parameters today, it is impossible to quantify the distance to objects or events in the sky. However, precisely because of this relation between distance and density and the Hubble constant, we have at our disposal a number of interesting observational tools to measure the cosmological parameters.

For example, universe models with a non-zero cosmological constant have recently become fashionable and eventhough nobody understands the underlying physics, we can nonetheless measure its value and how it affects the dynamics of the universe. Equation (3.94a) suggests that we can define a vacuum energy density ρ_Λ as

$$\rho_\Lambda \equiv \frac{\Lambda}{8\pi G}. \tag{3.106}$$

Then from the first law of thermodynamics [eqn. (3.83)],

$$\rho_\Lambda c^2 dV + p_\Lambda dV = 0, \tag{3.107}$$

giving the equation of state for the pure vacuum:

$$p_\Lambda = -\frac{\rho_\Lambda}{c^2}. \tag{3.108}$$

This unusual equation of state suggests peculiar, "anti-gravitational" forces that push the universe apart. The Friedmann equation for a matter plus vacuum universe can be written

$$\ddot{a} = -\frac{4\pi}{3} G a (\rho_m - 3\rho_\Lambda). \tag{3.109}$$

Since the matter density falls off as a^{-3} with the expansion, any non-zero vacuum energy will eventually dominate, giving an *accelerating* universe with $\ddot{a} > 0$. Since, until recently, the universe has been thought to be decelerating, the conventional parameter used to measure such an effect is called the *deceleration* parameter q_0:

$$q_0 \equiv -\left(\frac{\ddot{a}a}{\dot{a}^2}\right) = \frac{\Omega_{m0}}{2} - \Omega_{\Lambda 0}. \tag{3.110}$$

For certain values of Ω_m and Ω_Λ, there are actually static solutions where $\dot{a} = \ddot{a} = 0$. In fact, the desire to find such static solutions to the Friedmann equation is what originally motivated Einstein to introduce the cosmological constant in the first place.

In 1970, Sandage (Physics Today, 23[2], 34) famously described cosmology as "a search for two parameters:" H_0 and q_0. Today, along with H_0 and q_0, observational cosmologists are trying (and in many cases, quite successfully) to measure the different critical density factors Ω_{m0}, Ω_{r0}, $\Omega_{\Lambda 0}$, and the matter density in "normal," baryonic matter Ω_{b0}. In addition to the density factors (the "ingredients" of the universe), we would also like to know the amplitude of the initial density perturbations and the spectral index for the perturbation distribution (the "recipe" building the universe into galaxies and large-scale structure).

But then again, Sandage may have had it right after all. His two parameters are (more or less) the time scale of the universe and its radius of curvature. One might argue that all of the other "cosmological" parameters are mere physics, in the same way that the internal composition of a star will affect its lifetime through nuclear reaction rates and opacities and the like. While this physics is certainly interesting, it is not fundamentally cosmological.

Some of the classical tests for determining these parameters are described below:

Angular diameter-redshift test

Since the angular diameter D_A is a function of $(H_0, \Omega_{m0}, \Omega_{\Lambda 0})$, by measuring the angular size $d\phi$ on the sky of a collection of objects of known physical size at different redshifts, we can fit $D_A(z)$ to the best cosmological model parameters. For a "standard meter stick" of length L at redshift $(z+1)$, the angular size in the sky is

$$d\phi(z+1) = \frac{L}{D_A(H_0, \Omega_{m0}, \Omega_{\Lambda 0}, z)}. \tag{3.111}$$

Now all we need are good cosmological meter sticks. Good meter sticks can be of all different lengths, as long as we can determine their lengths using other observational features. Some possibilities are:

- Double-lobe radio galaxies: the physical size of the radio jets is related to their velocity and emission.
- The size of elliptical galaxy and galaxy cluster cores is often related to the velocity dispersion of the system or the surface brightness of the object.
- The image separation of a gravitational lens. In this case, we would measure the Einstein radius:

$$\theta_E = 4\pi \frac{\sigma^2}{c^2} \frac{D_{LS}}{D_S},$$

as introduced in Section 2.2. While we were not careful there to specify what sort of distance was meant by D_S and D_{LS}, with hindsight these can be seen to be angular diameter distances from the source to the observer and from the lens to the source.

Number-redshift test

By counting up the total number of objects of a given brightness as a function of redshift, we can measure the comoving volume $V(\Omega_m, \Omega_\Lambda)$. The major problem with this method is that, while the total mass within a comoving volume element dV will remain roughly constant with the expansion of the universe, the actual objects that fill this region will evolve and also merge. This means that neither the total number of objects nor their intrinsic brightness is constant with redshift. The Hubble Deep Field program (since superceded by the Ultradeep Field) detected about 100-1000 times more high-redshift galaxies than predicted by a no-evolution model. So while not being an ideal method for independently measuring Ω_{m0} and $\Omega_{\Lambda 0}$, the number-redshift test, along with a decent cosmological model, can give a great deal of information about galaxy evolution and hierarchical structure formation.

Luminosity-redshift test

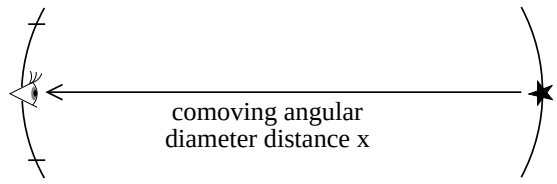

Figure 3.11: If the pupil of the observer at redshift zero subtends a solid angle $d\Omega$ as seen from the source, the fraction of the total luminosity received by the observer is $d\Omega/4\pi$. A square one unit length on a side subtends a solid angle 1 unit length$/4\pi x^2$, where x is the co-moving angular diameter distance.

As defined in equation (3.100), the observed bolometric flux from a distant object is proportional to the inverse square of the luminosity distance, or alternatively

$$f_{\rm bol}(obs) = \frac{L_{\rm bol}(emit)}{4\pi x^2} \left(\frac{1}{1+z}\right)^2. \tag{3.112}$$

Here the comoving transverse distance x is used to give the surface area of a distant sphere surrounding the source (see Fig. 3.11), thus the solid angle factor of the observer. Since flux is measured in units of [energy/area/time], two factors of $1/(1+z)$ are required to account for the photon redshift reducing the energy of any given photon and the time dilation of a receding source emitting fewer

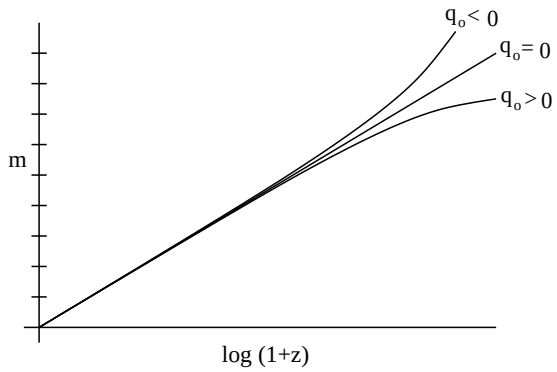

Figure 3.12: Apparent magnitude versus redshift for the different universes: decelerating, coasting, and accelerating.

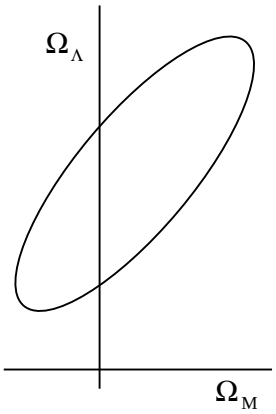

Figure 3.13: Constraints on the matter density and the vacuum energy density as determined from supernovae. The most probable values of Ω_M and Ω_Λ lie inside the ellipse. The location of the first accoustic peak in the CMB gives a second ellipse, elongated perpendicular to the one shown.

photons per second as measured by the observer. The central problem for the luminosity-redshift test is finding a "standard candle," an object of known (or at least scalable) intrinsic luminosity $L_{\rm bol}$. Then by plotting the observed magnitude m as a function of redshift, we can fit the data to different values of the deceleration parameter q_0 (see Fig. 3.12). In the 60s and 70s the most promising standard candle was the population of brightest cluster elliptical galaxies, which were grouped around a single absolute magnitude with a variance of $\sigma \sim 0.3 - 0.45$ mag. Just like the number-redshift test, the problem with the elliptical galaxies was that of evolution, introducing systematic errors at higher redshifts.

Today, by far the most successful form of the luminosity-redshift test has been the use of type-Ia supernovae, which have a strong correlation between the time for the peak emission to fade and the intrinsic luminosity. Here the trick is just to find these supernovae at large redshifts shortly after they explode, and then follow them over a period of weeks or months to measure their waxing and waning. The results of many such observations suggest that the universe is in fact expanding at an *accelerating* rate with $q_0 < 0$, consistent with a non-zero cosmological constant. Since the deceleration parameter is a function of Ω_Λ and Ω_m, the luminosity-redshift test is actually most sensitive to a linear combination of the two density parameters, as can be seen in the "error ellipse" in Figure 3.13.

Recent observations by Riess and collaborators have pushed type Ia supernova observations out to $z \sim 1.5$ at which epoch the universe was still decelerating. This allows measurement of the third derivative of the expansion parameter a which when rendered dimensionless with factors of $\dot{a}$ and a gives a "cosmic jerk" parameter. The measured value of the cosmic jerk is consistent a flat universe presently dominated by Λ.

3.9 Density perturbations in an expanding universe: Jeans's approach revisited

As we showed in Section 1.11 for a static space-time, small density perturbations in an otherwise uniform fluid will either grow through gravitational instability or oscillate like acoustic waves held up from collapsing by thermal and radiation pressure. However, unlike in Section 1.11, now we want to investigate the growth of perturbations in an *expanding* fluid. Recall that the Jeans' length λ_J dependence on the sound velocity v_s in the gas:

$$\lambda_J^2 = \frac{\pi v_s^2}{G\rho_0}. \tag{3.113}$$

As the universe expands adiabatically, the fluid is dominated by radiation pressure at first, giving the sound speed for a relativistic gas: $v_s^2 = c^2/3$. During this period, the corresponding Jeans length is very large, causing most modes to undergo accoustic oscillation. Only when the gas cools enough for the hydrogen ions to combine with the background electrons and form neutral atomic hydrogen (transparent to the background radiation) does the pressure become gas-dominated, dramatically lowering the sound speed and thus allowing the modes to grow under gravitational collapse.

Following the same general approach of Section 1.11, we begin by writing the fluid equations for a gas moving with a velocity $\vec{\mathbf{v}}$, pressure p, density ρ, and gravitational potential Φ. For Eulerian space with fixed coordinates, the mass continuity equation is

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{\mathbf{v}}) = 0, \tag{3.114}$$

the Euler equation

$$\frac{\partial \vec{\mathbf{v}}}{\partial t} + (\vec{\mathbf{v}} \cdot \vec{\nabla})\vec{\mathbf{v}} = -\vec{\nabla}\Phi - \frac{1}{\rho}\vec{\nabla} p, \tag{3.115}$$

and the Poisson equation is

$$\nabla^2 \Phi_1 = 4\pi G \rho_1. \tag{3.116}$$

Unlike our earlier approach, where the average velocity was taken to be zero, for the expanding universe we must take into consideration the bulk velocity of the fluid according to Hubble's law $\vec{\mathbf{v}} = H\vec{\mathbf{x}}$. This formulation of the fluid problem strongly suggests the use of comoving Lagrangian coordinates. In comoving coordinates, the time derivative can be re-written as

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \vec{\mathbf{v}} \cdot \vec{\nabla}. \tag{3.117}$$

Then the continuity equation for a Hubble flow in three dimensions becomes

$$\frac{d\rho}{dt} = -\rho \vec{\nabla} \cdot \vec{\mathbf{v}} = -3H\rho \tag{3.118}$$

and Euler's equation can be written

$$\frac{d\vec{\mathbf{v}}}{dt} = -\frac{1}{\rho}\vec{\nabla} p - \vec{\nabla}\Phi. \tag{3.119}$$

Just as with the Eulerian formalism, we will consider small deviations from the steady-state equilibrium solutions to the fluid equations 3.118 and 3.119. For small perturbation δ, we have the following expressions for the density, velocity, pressure, and potential of the fluid:

$$\vec{\mathbf{v}}(\vec{\mathbf{x}},t) = \vec{\mathbf{v}}_0(\vec{\mathbf{x}}) + \delta\vec{\mathbf{v}}_1(\vec{\mathbf{x}},t) \tag{3.120a}$$
$$\rho(\vec{\mathbf{x}},t) = \rho_0(\vec{\mathbf{x}}) + \delta\rho_1(\vec{\mathbf{x}},t) \tag{3.120b}$$
$$p(\vec{\mathbf{x}},t) = p_0(\vec{\mathbf{x}}) + \delta p_1(\vec{\mathbf{x}},t) \tag{3.120c}$$
$$\Phi(\vec{\mathbf{x}},t) = \Phi_0(\vec{\mathbf{x}}) + \delta\Phi_1(\vec{\mathbf{x}},t). \tag{3.120d}$$

Then the perturbation terms give

$$\frac{d}{dt}\left(\frac{\delta\rho}{\rho_0}\right) = -\vec{\nabla}\cdot\delta\vec{\mathbf{v}}, \tag{3.121}$$

$$\frac{d}{dt}(\delta\vec{\mathbf{v}}) = (\delta\vec{\mathbf{v}}\cdot\vec{\nabla})\vec{\mathbf{v}}_0 = -\frac{\delta p}{\rho_0} - \vec{\nabla}\delta\Phi, \tag{3.122}$$

and

$$\nabla^2\delta\Phi = 4\pi G\delta\rho. \tag{3.123}$$

In comoving (Lagrangian) coordinates,

$$\vec{\mathbf{r}} \equiv \frac{\vec{\mathbf{x}}}{a(t)}, \tag{3.124a}$$
$$\vec{\mathbf{u}} \equiv \frac{d\vec{\mathbf{r}}}{dt} \tag{3.124b}$$

are the comoving positions and velocities, related to the "absolute" position $\vec{\mathbf{x}}$ and $\vec{\mathbf{v}}$ by

$$\delta\vec{\mathbf{x}} = \vec{\mathbf{r}}\delta a + a\delta\vec{\mathbf{r}} \tag{3.125}$$

and

$$\vec{\mathbf{v}} = \frac{d\vec{\mathbf{x}}}{dt} = \vec{\mathbf{r}}\frac{da}{dt} + a\frac{d\vec{\mathbf{r}}}{dt}. \tag{3.126}$$

The comoving gradient $\vec{\nabla}_c$ is given by

$$\vec{\nabla} = \frac{1}{a}\vec{\nabla}_c = \frac{1}{a}\frac{\partial}{\partial\vec{\mathbf{r}}}. \tag{3.127}$$

Then equation (3.122) becomes

$$\frac{d}{dt}\vec{\mathbf{u}} + 2\frac{\dot{a}}{a}\vec{\mathbf{u}} = -\frac{1}{\rho_0 a^2}\vec{\nabla}_c\delta p - \frac{1}{a^2}\vec{\nabla}_c\delta\Phi \tag{3.128}$$

and taking the comoving divergence,

$$\frac{d}{dt}(\vec{\nabla}_c\cdot\vec{\mathbf{u}}) + 2\frac{\dot{a}}{a}(\vec{\nabla}_c\cdot\vec{\mathbf{u}}) = -\frac{1}{\rho_0 a^2}\nabla_c^2\delta p - \frac{1}{a^2}\nabla_c^2\delta\Phi. \tag{3.129}$$

The continuity equation in comoving coordinates gives

$$\vec{\nabla}_c\cdot\vec{\mathbf{u}} = -\frac{d}{dt}\left(\frac{\delta\rho}{\rho_0}\right), \tag{3.130}$$

and assuming adiabatic expansion, we have

$$\delta p = c_s^2 \delta\rho. \tag{3.131}$$

Bringing it all together, we get a second-order wave equation

$$\frac{d^2}{dt^2}\left(\frac{\delta\rho}{\rho_0}\right) + 2\frac{\dot{a}}{a}\frac{d}{dt}\left(\frac{\delta\rho}{\rho_0}\right) = c_s^2\nabla_c^2\frac{\delta\rho}{\rho_0} + 4\pi G\delta\rho. \tag{3.132}$$

Just as in the case of a static fluid undergoing gravitational collapse, we can look at the wave equation in the Fourier domain. For single-mode perturbations with amplitude Δ and comoving wave-vector $\vec{\mathbf{k_c}}$, the density fluctuations can be written as

$$\frac{\delta\rho}{\rho_0} = \Delta e^{i\vec{\mathbf{k_c}}\cdot\vec{\mathbf{r}}}. \tag{3.133}$$

Then the dispersion relation from equation (3.132) is

$$\frac{d^2}{dt^2}\Delta + 2\frac{\dot{a}}{a}\frac{d}{dt}\Delta = \Delta(4\pi G\rho_0 - a^2k_c^2c_s^2). \tag{3.134}$$

Setting $\dot{a} = 0$ for a static universe, we recover the Jeans equation with $\lambda_J \approx c_s(\pi/G\rho_0)^{1/2}$.

3.10 Growth of perturbations in the early universe

At early times after the big bang, the temperature of the universe was high enough for the leptons and baryons to be ultra-relativistic, behaving with an equation of state similar to that of pure radiation: $p = c^2\rho/3$. However, since the massive particles do not lose energy with redshift like the photons do, the dynamics of the expansion *do* discriminate between matter-dominated ($\Omega_m = 1, \Omega_r = 0$) and radiation-dominated ($\Omega_m = 0, \Omega_r = 1$) universes. For the Einstein-deSitter cosmology ($\Omega_m = 1, \Omega_r = 0, \Omega_\Lambda = 0$), recall

$$\frac{\dot{a}}{a} = \frac{2}{3}\frac{1}{t}, \qquad t = \left(\frac{1}{6\pi G\rho}\right)^{1/2}.$$

Then large-scale perturbations ($k^{-1} = \lambda \gg \lambda_J$) will grow according to

$$\frac{d^2}{dt^2}\Delta + \frac{4}{3t}\frac{d}{dt}\Delta - \frac{2}{3t^2}\Delta = 0. \tag{3.135}$$

Trying solutions of the form $\Delta \propto t^n$ gives

$$n(n-1) + \frac{4}{3}n - \frac{2}{3} = 0, \tag{3.136}$$

with solutions $n = (2/3, -1)$. The negative solution gives a decaying mode, while the $n = 2/3$ solution corresponds to perturbations growing with the same rate as the background expansion:

$$\Delta \sim t^{2/3} \sim \frac{1}{1+z}. \tag{3.137}$$

For an empty or very low density universe, $\dot{a}/a \sim t^{-1}$ and the power-law growth solutions have $n = (0, -1)$, in other words, no growth. In general, there will be positive growth when $\Omega_m \sim 1$ and stasis when $\Omega_m \sim 0$. For the case of a small density parameter at the current epoch ($\Omega_{m0} \ll 1$), the perturbation growth stopped at a redshift of

$$1 + z \approx \frac{1}{\Omega_{m0}}. \tag{3.138}$$

For the case of a flat universe with a cosmological constant and $\Omega_{m0} + \Omega_{\Lambda 0} = 1$, growth stopped around a redshift of

$$1 + z \approx \frac{1}{\Omega_{m0}^{1/3}}. \tag{3.139}$$

For a fluid in hydrostatic equilibrium (no pressure gradients), the peculiar velocities evolve according to

$$\frac{d}{dt}\vec{\mathbf{u}} + 2\frac{\dot{a}}{a}\vec{\mathbf{u}} = -\frac{1}{a^2}\vec{\nabla}_c\Phi, \tag{3.140}$$

so for a uniform density distribution ($\vec{\nabla}_c\Phi = 0$), peculiar velocities decrease with the expansion: $u \sim a^{-2}$.

In the case where the density gradient is not zero, we can write the peculiar velocity as

$$\vec{\mathbf{u}} = \vec{\mathbf{u}}_{\parallel} + \vec{\mathbf{u}}_{\perp}, \tag{3.141}$$

where $\vec{\mathbf{u}}_{\parallel}$ is the component of the velocity parallel to $\vec{\nabla}_c\delta\rho$. In this form, we see that velocity perturbations can grow in the direction of density perturbations, while the perpendicular components are damped. Since the curl of the gradient of the density field is zero, the overall vorticity of the velocity field decays with expansion proportional to a^{-1}.

Returning to a general fluid of radiation and relativistic particles, we write down the modified relativistic continuity, Euler, and Poisson equations:

$$\frac{\partial\rho}{\partial t} + \vec{\nabla}\cdot\left[\left(\rho + \frac{p}{c^2}\right)\vec{\mathbf{v}}\right] = 0, \tag{3.142}$$

$$\left(\rho + \frac{p}{c^2}\right)\frac{\partial\vec{\mathbf{v}}}{\partial t} = -\left(\rho + \frac{p}{c^2}\right)\vec{\nabla}\Phi - \vec{\nabla}P, \tag{3.143}$$

and

$$\nabla^2\Phi = 4\pi G\left(\rho + \frac{3p}{c^2}\right). \tag{3.144}$$

Note how the hydrodynamic $(\rho + p/c^2)$ and gravitational $(\rho + 3p/c^3)$ terms are different with the addition of relativistic corrections. The resulting wave equation in Fourier space is

$$\frac{d^2}{dt^2}\Delta + 2\frac{\dot{a}}{a}\frac{d}{dt}\Delta = \Delta\left(\frac{32\pi G\rho_0}{3} - a^2k_c^2c_s^2\right). \tag{3.145}$$

Neglecting the pressure due to the massive particles, the sound speed c_s is given by

$$c_s^2 = \frac{\left(\frac{\partial p}{\partial T}\right)_{\rm rad}}{\left(\frac{\partial\rho}{\partial T}\right)_{\rm rad} + \left(\frac{\partial\rho}{\partial T}\right)_{\rm mat}} = \frac{1}{3}c^2\frac{\rho_{\rm rad}}{\rho_{\rm rad} + \frac{4}{3}\rho_{\rm mat}}. \tag{3.146}$$

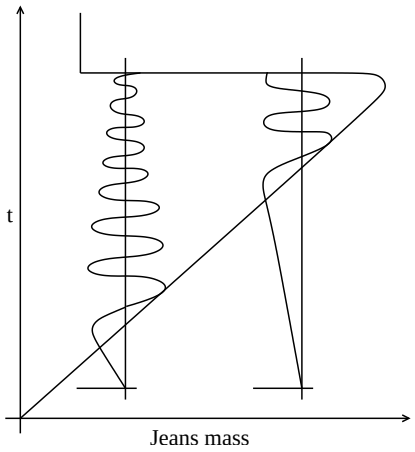

Figure 3.14: The growth of the Jeans mass as a function of time. Also Shown is the growth of two perturbations of different masses. They grow only when their masses are larger than the Jeans mass, oscillating when not.

For a radiation dominated expansion, the dynamics are similar to those calculated for the Einstein-deSitter universe:

$$\frac{\dot{a}}{a} = \frac{1}{2t}, \tag{3.147a}$$

$$t = \left(\frac{3}{32\pi G\rho}\right)^{1/2}. \tag{3.147b}$$

Again looking for solutions of the form $\Delta \propto t^n$ gives $n = (1, -1)$, or growing modes with

$$\Delta \sim t \sim a^2. \tag{3.148}$$

In the radiation era, the Jeans mass grows in time as

$$M_J \sim \lambda_J^3 \rho_{\rm mat} \sim t^{3/2}, \tag{3.149}$$

with all smaller perturbations undergoing stable oscillations before the period of recombination, when the sound speed falls off dramatically, leading to the unstable collapse of all fluctuations larger than the suddenly reduced Jeans mass (see Fig. 3.14). In fact, as we saw in Section 3.3, a more careful treatment of stable oscillations in the radiation era shows that perturbations with masses smaller than the Silk mass ($M_S \sim 10^{12} M_\odot$) are actually damped out before recombination, leaving only very large mass perturbations to collapse. Furthermore, from the fluctuations in the cosmic microwave background radiation, we have direct measurements of the amplitude of radiation and baryonic density fluctuations at the time of recombination, and these amplitudes are not large enough to have seeded the structures observed today. So then where do the typical galaxies with $M_{\rm gal} \ll M_S$ come from?

Zeldovich (1967, 1970) studied perturbations of the form

$$\frac{\delta\rho_{\rm mat}}{\rho_{\rm mat}} = \frac{3}{4}\frac{\delta\rho_{\rm rad}}{\rho_{\rm rad}}, \tag{3.150}$$

called *adiabatic* perturbations, as a possible solution to the galaxy problem. This "pancake," or "top-down" scenario seemed to predict more small-mass objects, but they generally do not form early enough, and there was already significant observational evidence for galaxies at large redshifts. Another option was *isothermal* perturbations, with

$$\frac{\delta\rho_{\rm rad}}{\rho_{\rm rad}} = 0, \tag{3.151}$$

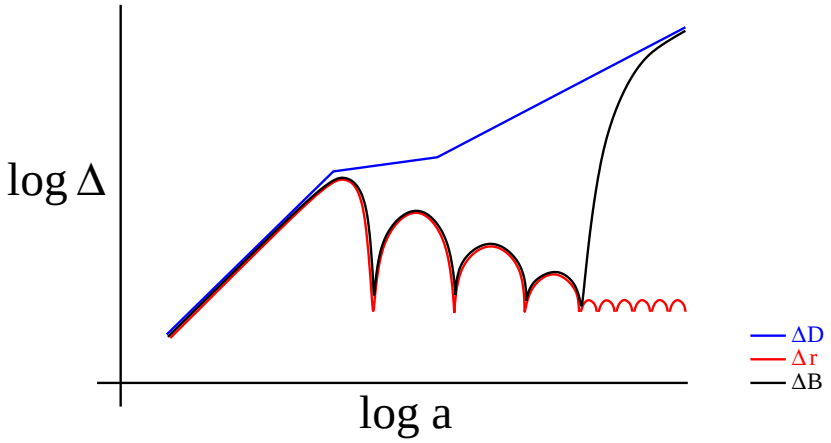

Figure 3.15: The growth of perturbation amplitude as a function of expansion factor a. The radiation and baryonic components cease to grow when the perturbation passes through the horizon. The dark matter component grows more rapidly once radiation no longer dominates. The baryonic perturbations grow rapidly after recombination.

but these provided little more hope than the adiabatic approach.

The current explanation for the source of galactic-size mass perturbations is based on dark matter. Assuming an initial perturbation power spectrum that is identical for radiation, baryons, and dark matter, fluctuations will grow according to equation (3.148) with the expansion until the horizon is comparable to the mode wavelength, at which point the baryonic and radiation perturbations undergo Silk damping. But by virtue of its being transparent, the dark matter experiences no such pressure forces, and dark matter perturbations keep on growing through recombination. After recombination, with little or no radiation pressure, the baryonic matter falls into the potential wells formed by the dark matter fluctuations (see Fig. 3.15). The smaller masses collapse first to form galaxies, then larger masses merge to form hierarchical structures, including the super-clusters that we see today.

3.11 Statistical measures of cosmic structure

Recall our general distinction in cosmological structure formation between field perturbations and discrete objects. Roughly speaking, if the density of the region in question is $\bar{\rho}$ and the average density of the background universe is ρ_u, then these approaches are categorized by

$$\begin{aligned} \ln\frac{\bar{\rho}}{\rho_u} &< 1 \qquad \text{fields} \\ \ln\frac{\bar{\rho}}{\rho_u} &> \ln\frac{9\pi^2}{16} \qquad \text{objects.} \end{aligned} \tag{3.152}$$

In this section we are going to be concerned primarily with *objects* and their statistical distribution in space. These statistics are usually expressed in one of the following forms:

- counts in cells (1-point correlation function)

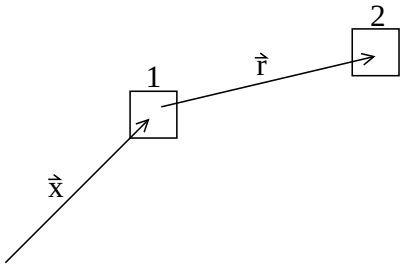

Figure 3.16: The spatial correlation coefficient ξ(r) measures the extra probability of finding a galaxy in volume dV_2 given a galaxy in volume dV_1.

- 2-point correlation function
- power spectrum (Fourier transform of the 2-point correlation function)

The one-point correlation function is simply the probability of finding a galaxy (or quasar or cluster or any given type of object) within a small volume element dV. If the average number density of that type of object is N_0, then the probability is just $P(dV) = N_0 dV$.

The two-point correlation function is the probability of finding an object in a volume dV_1 and also another object in a volume dV_2, where the two volume elements are separated by $\vec{\mathbf{r}}$. For an average number density N_0, this distribution can be written

$$dN_{\rm pair}(\vec{\mathbf{r}}) = N_0^2 dV_1 dV_2[1 + \xi(\vec{\mathbf{r}})], \tag{3.153}$$

where $\xi(\vec{\mathbf{r}})$ is called the spatial correlation function. For isotropic distributions, $\xi(\vec{\mathbf{r}}) = \xi(r)$. While these relations have been defined for discrete objects, it is conceptually straightforward to define a correlation function for continuous distributions as well. For example, with a density field defined by

$$\rho(\vec{\mathbf{x}}) = \rho_0[1 + \Delta(\vec{\mathbf{x}})], \tag{3.154}$$

the probability of matching "pairs" with a separation $\vec{\mathbf{r}}$ (see Fig. 3.16) is given by

$$dN_{\rm cont}(\vec{\mathbf{r}}) = \rho_0^2[1 + \Delta(\vec{\mathbf{x}})][1 + \Delta(\vec{\mathbf{x}} + \vec{\mathbf{r}})], \tag{3.155}$$

giving a spatial correlation function

$$\xi(\vec{\mathbf{r}}) = \langle \Delta(\vec{\mathbf{x}})\Delta(\vec{\mathbf{x}} + \vec{\mathbf{r}})\rangle_{\vec{\mathbf{x}}} = \frac{1}{V}\int \Delta(\vec{\mathbf{x}})\Delta(\vec{\mathbf{x}} + \vec{\mathbf{r}}) d^3\vec{\mathbf{x}}. \tag{3.156}$$

Since, from an observer's point of view, the sky is two-dimensional, it also makes sense to define an angular correlation function. Replacing the volume elements with solid angles $d\Omega_1$ and $d\Omega_2$ and the density with n_0 the angular correlation function $\omega(\vec{\theta})$ is defined by

$$dN_{\rm pair}(\vec{\theta}) = n_0^2 d\Omega_1 d\Omega_2[1 + \omega(\vec{\theta})]. \tag{3.157}$$

Typically, $\omega(\vec{\theta})$ is smaller in magnitude than $\xi(\vec{\mathbf{r}})$ because foreground and background dilution of sources makes the universe look more uniform when projected onto the two-dimensional surface of the visible sky.

The last third of the 20th century saw increasingly large samples of galaxies used for measurement of the correlation function (and the power spectrum). The counts of Shane and Wirtaanen gave a

reliable angular correlation function. Redshift surveys carried out one galaxy at a time produced thousands of redshifts, giving spatial correlation functions. Surveys carried out using hundreds of optical fibers strategically positioned in the focal plane of the telescope have produced tens of thousands of redshifts (at Las Campanas) and hundreds of thousands of redshifts (at the Anglo-Australian Telescope and with the Sloan Sky Survey Telescope) with a cumulative total that will soon pass the million galaxy mark.

To good approximation, the galaxy-galaxy spatial correlation function is a power law,

$$\xi(r) \approx \left(\frac{r}{r_0}\right)^{-1.8} \quad , \tag{3.158}$$

with scale length $r_0 \approx 5h^{-1}$Mpc. A straightforward consequence of this power law is that if you see a galaxy, any nearby comparable galaxy is probably at the same redshift. As one goes to more distant galaxies the angle governing the definition of "nearby" becomes smaller. With the largest surveys of galaxies, it has been possible to establish deviations from the above power law. In particular the 2-point correlation function shows an inflection at the transition between pairs of galaxies that are members of the same virialized group and galaxies that are memebers of different but correlated groups.

A useful alternative to the 2-point correlation function is the *power spectrum*, which can be compared readily with analytic theories of structure formation, as well as the results from N-body computer simulations. From a technical standpoint it has the advantage that it may be calculated more robustly, particularly at large scales. Using the deviation from uniform density $\Delta(\vec{\mathbf{x}})$ from equation (3.154), we can express it and its Fourier pair $\tilde{\Delta}(\vec{\mathbf{k}})$ as

$$\begin{aligned} \Delta(\vec{\mathbf{x}}) &= \frac{V}{(2\pi)^3}\int \tilde{\Delta}(\vec{\mathbf{k}})e^{-i\vec{\mathbf{k}}\cdot\vec{\mathbf{x}}}d^3\vec{\mathbf{k}} \\ \tilde{\Delta}(\vec{\mathbf{k}}) &= \frac{1}{V}\int \Delta(\vec{\mathbf{x}})e^{i\vec{\mathbf{k}}\cdot\vec{\mathbf{x}}}d^3\vec{\mathbf{x}}. \end{aligned} \tag{3.159}$$

Then equation (3.156) can be written

$$\begin{aligned} \xi(\vec{\mathbf{r}}) &= \frac{1}{V}\int \Delta(\vec{\mathbf{x}})\Delta(\vec{\mathbf{x}}+\vec{\mathbf{r}})d^3\vec{\mathbf{x}} \\ &= \frac{1}{V}\int d^3\vec{\mathbf{x}}\int \frac{V}{(2\pi)^3}\tilde{\Delta}(\vec{\mathbf{k}})e^{-i\vec{\mathbf{k}}\cdot\vec{\mathbf{x}}}d^3\vec{\mathbf{k}}\int \frac{V}{(2\pi)^3}\tilde{\Delta}^*(\vec{\mathbf{k}}')e^{i\vec{\mathbf{k}}'\cdot(\vec{\mathbf{x}}+\vec{\mathbf{r}})}d^3\vec{\mathbf{k}}' \\ &= \frac{V}{(2\pi)^3}\int |\tilde{\Delta}(\vec{\mathbf{k}})|^2 e^{i\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}}d^3\vec{\mathbf{k}}. \end{aligned} \tag{3.160}$$

In the last line we have used the orthogonality relation

$$\frac{1}{(2\pi)^3}\int e^{i(\vec{\mathbf{k}}-\vec{\mathbf{k}}')\cdot\vec{\mathbf{x}}}d^3\vec{\mathbf{k}} = \delta_3(\vec{\mathbf{k}}-\vec{\mathbf{k}}'). \tag{3.161}$$

Then the power spectrum is defined by

$$P(\vec{\mathbf{k}}) \equiv |\tilde{\Delta}(\vec{\mathbf{k}})|^2. \tag{3.162}$$

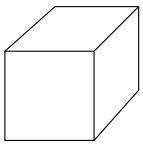
Figure 3.17: A reference volume V over which the power spectrum is computed.

The power spectrum and the 2-point spatial correlation function are a Fourier pairs, each the Fourier transform of the other. Assuming isotropy and averaging over angle, the correlation function is given by

$$\xi(r) = \frac{V}{2\pi^2} \int P(k) \left(\frac{\sin(kr)}{kr} \right) k^2 dk, \tag{3.163}$$

where the factor $\sin(kr)/kr$, often called the window function, shows that the longer wavelengths with $kr < 1$ make the dominant contribution to the correlation function.

3.12 Cosmic Microwave Background Fluctuations

Recall from our discussion of cosmological perturbation theory the basic sequence of matter evolution in the early universe:

- Fluctuations with $M > M_{\rm horizon}$ grow;
- Fluctuations with $M < M_{\rm horizon}$ are schizophrenic, with the dark matter component growing slowly while the photon-baryon fluid oscillates;
- regions with (dark matter) $\rho_m > \rho_r$ grow faster than when radiation dominates;
- post-recombination, the baryons fall into the gravitational potential wells formed by dark matter structure.

To model the stasis/growth phase, we need to include four fluid components (dark matter, baryons, photons, and neutrinos) interacting under gravity and electromagnetic forces for the photons and baryons. To constrain these models, we have direct observations of the baryon-baryon and dark matter-dark matter correlation functions ξ_{bb} and ξ_{dd} today as well as the photon-photon correlation function $\xi_{\gamma\gamma}$ from the recombination epoch of $z_{\rm rec} \approx 1100$. ξ_{bb} is easily observed with galaxy surveys and ξ_{dd} can be inferred (at least in principle) from weak lensing. They are related with the "bias" parameter b:

$$\xi_{bb} = b^2 \xi_{dd} \quad . \tag{3.164}$$

The idea behind biasing is illustrated in Figure 3.18. Galaxy sized perturbations will grow more or less rapidly depending upon whether they are embedded in large scale perturbations with positive or negative amplitudes. It would be a surprise of the properties of galaxies were the same in larger regions that were more and less dense. Remarkably it has proven difficult to identify major differences between the galaxies in clusters and voids.

The dark matter can be correlated with the baryonic matter with a cross-correlation function ξ_{bd} defined in an analogous manner to equations (3.153) and (3.156). The "correlation" parameter r is determined from

$$\xi_{bd} = rb\xi_{dd}, \tag{3.165}$$

where a value of $r = 1$ indicates non-stochastic biasing.

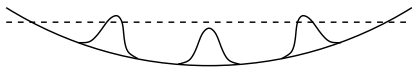

Figure 3.18: Three small scale (galaxy sized) positive dark matter perturbations embedded within a larger scale (cluster sized) negative perturbation. The dashed line indicates the critical density. Only two of the three perturbations can collapse. They will find themselves in a large scale region relatively devoid of galaxies.

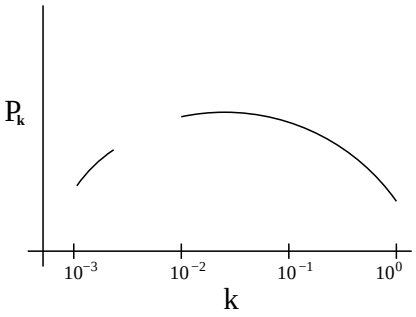

Figure 3.19: A schematic of the power spectrum as observed from the CMB (at low wavenumber) with COBE and from galaxies (at high wavenumber). More recent CMB results extend to higher wavenumbers and now overlap with the galaxy results. The results are consistent with an initial Harrison-Zeldovich, constant curvature spectrum of perturbations with $n = 1$. The CMB and the galaxies are observed at very different epochs. Either one must be evolved forward or the other backward.

The power spectrum corresponding of the photon anisotropies at recombination gives an remarkable snapshot of the early universe. Since the temperature fluctuations in the background radiation are straightforwardly related to the spatial density and velocity fluctuations in the baryon-dark matter fluid, cosmic microwave background (CMB) measurements serve as a direct statistical determination of the initial conditions for the calculations of perturbation growth. In the past ten years, unprecedented accuracy has been achieved in CMB observations and the future promises to give even better precision on finer and finer scales. The basic observable that all these experiments are trying to measure is the power spectrum $P(k)$ in the photon temperature anisotropy. Figure 3.19 shows the dark matter power spectrum $P(k)$ as obtained from the CMB (at low wavenumber) and $z \sim 0$ galaxies (at large wavenumber. Note that (to first order) there are no wiggles in the dark matter power spectrum – these are present only photon-baryon fluid. The comoving wavenumber k in given in units of Mpc^{-1}.

While the CMB was discovered in 1965, its anisotropies have only been measurable since the 1990s, starting with the COBE satellite, proceeding through a series of groundbased experiments and balloon experiments to the Wilkinson Microwave Anisotropy Probe (WMAB). The fluctuations are very small – of order 10^{-5} times the CMB temperature, and require remarkably stable instruments to measure them. The technical advances making these measurements possible include HEMT amplifiers at 30 GHz and lower frequencies and cryogenic bolometric arrays at 100 GHz and higher frequencies.

Considering the direct observable of the blackbody temperature as a function of angular position in the sky $\Theta(\theta, \phi)$, we can decompose the signal into spherical harmonic modes defined by

$$\Theta_{lm} = \int Y^*_{lm}(\theta, \phi)\Theta(\theta, \phi)d\Omega. \tag{3.166}$$

In the limit of large l, the variance of the perturbation field is

$$\sigma^2 = \frac{1}{(2\pi)^2}\int d^2 l C_l, \tag{3.167}$$

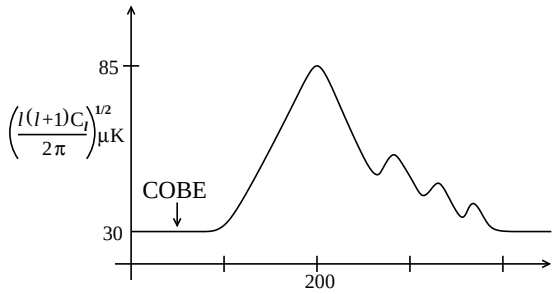

Figure 3.20: A schematic of the cosmic microwave background spectrum. The COBE satellite measured the spectrum in the linear regime.

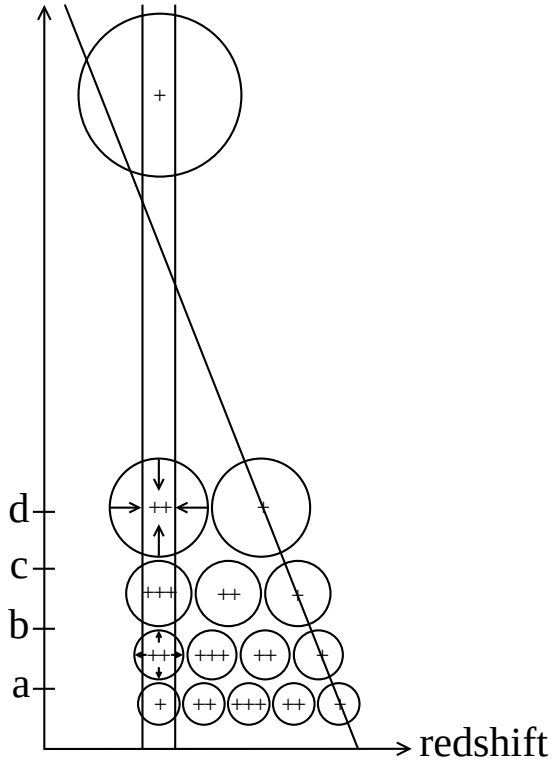

Figure 3.21: This figure has been adapted from Lineweaver's paper "CMBology," published in "Gravitational Lensing: Recent Progress and Future Goals" but most easily found at http://www.arxiv.org/pdf/astro-ph/9909301. The two vertical lines mark the beginning and end of recombination. The circles represent perturbations on increasingly large scales toward the top of the figure. The diagonal line marks the epoch at which the horizon passes through perturbations of a given scale. All of the perturbations have a positive density contrast at the moment they pass through the horizon, and all of continue to grow more positive after horizon crossing. Though the density of the photon-baryon fluid oscillates, the density perturbation is always positive, with amplitude indicated by the number of + signs. A similar diagram can be drawn for negative perturbations.

so for this reason, the temperature power spectrum is usually plotted in terms of Δ_T, where

$$\Delta_T^2 = \frac{l(l+1)}{2\pi} C_l T^2. \tag{3.168}$$

Figure 3.20 shows this power spectrum as a function of multipole moment l. The scale of micro-Kelvins shows just how impressive the precision of these measurements is.

The most prominent feature in the CMB power spectrum is the maximum near $l \approx 200$, called the "first acoustic peak." This peak can be understood qualitatively by estimating the distance a sound wave could travel in the early photon-baryon fluid from the time of the big bang to the era of recombination. For a flat universe, this appears at $l \approx 200$, with a corresponding angular scale of 1.8°. This factors of 3-10 larger than the perturbations that give rise to today's clusters of galaxies.

In addition to the first acoustic peak, there are also a number of significant harmonic acoustic peaks at higher wave numbers. These can be understood with the help of a *Lineweaver diagram* (shown in Fig. 3.21), a schematic that tracks the onset of perturbation growth as the causal horizon expands in time. As the horizon becomes larger than a perturbation the perturbations continue to collapse, but the radiation pressure of the photon fluid resists the collapse, causing acoustic

oscillations in the photon-baryon fluid. The dark matter does not interact with the photon-baryon fluid, except gravitationally, and continues to collapse. The phase of the photon-baryon oscillations at the time of recombination determines how they affect the CMB spectrum. Those wavelengths that are at maximum or minimum densities contribute maxmimally to the power. Those modes observed at the peak velocity phase of their harmonic oscillations add a smaller, out of phase contribution to the CMB power.

The following features in the *Lineweaver diagram* are noteable:

bullet The pertubations at scale c are at maximum compression at recombination. The first peak in the CMB power spectrum is seen at this scale.

The perturbations at scale a reach maximum compression and re-expand. They are at maximum re-expansion at recombination. The second peak in the CMB power spectrum is seen at this scale.

bullet The perturbations at scale b reach maximum compression and re-expand. They have maximum outward velocity at recombination. They produce a "doppler" peak in the CMB power spectrum that partially fills in the valley between the first and second peaks

bullet The perturbations at scale d have maximum inward velocity at recombination. They produces a "doppler" peak in the CMB power spectrum shortward of the first peak that adds a shoulder to it.

bullet At scale b the baryon-photon fluid is at maximum compression. While most of the binding energy comes from the dark matter, the baryons contribute, causing the compression to be larger than it would have been had there been no baryons. The opposite is true at scale d. Hence odd-numbered peaks are stronger and even numbered peaks are weaker. This phenomenon is called "baryon-loading."

The addition of baryonic matter into the fluid equations causes the differential amplification of the odd and even acoustic peaks. The sign of the baryonic perturbations at the even peaks are opposite to those of the dark matter perturbations, suppressing the amplitude of the temperature fluctuations. The location and amplitude of these acoustic peaks are influenced by a number of different cosmological paramaters, including relative amounts of baryons, dark matter, and curvature in the universe. The first acoustic peak in particular has already given very strong restraints on the total density parameter $\Omega_0 \approx 1$, with a best fit around $(\Omega_{m0} = 0.3, \Omega_{\Lambda 0} = 0.7)$. These results nicely compliment the high-redshift supernova data, which give consistent values for the density parameters, with an error ellipse approximately orthogonal to that shown in Figure 3.13.

Future of CMB anisotropy experiments will expand our range to higher mode numbers l and higher accuracy in the individual peaks. Equally exciting is the possibility of measuring the *polarization* of the CMB fluctuations, which will give us yet another window into the physics of the early universe as well as the "dark ages" - the time between recombination and galaxy formation.

3.13 Big Bang Nucleosynthesis

In 1925, Cecilia Payne first determined the chemical makeup of the sun. The observed mass fractions in hydrogen, X, helium Y, and the heavier elements, Z, was found to be $X : Y : Z = 0.70 : 0.28 : 0.02$. This result was most remarkable in that the Sun had a very different makeup than the Earth. But ratio of hydrogen to helium was itself remarkable. Spectral observations of other stars yield, on rare occasion, values for Z that are smaller by factors as large as 10^3, but the lowest values for Y are still quite significant: $Y_{\rm min} \sim 0.235 - 0.245$. Nuclear burning theory predicts that from a pure hydrogen fuel, the relative amounts of helium produced compared to the heavier elements could not be more than $\Delta Y/\Delta Z \sim 2 - 3$, far below the smallest astrophysical values of $\Delta Y/\Delta Z \geq 14$. The conclusion of this discrepancy is that the majority of the helium in the universe must *not* have been made in stars.

If the helium doesn't come from hydrogen fusion in the center of stars, where does it come from? The only other environment that was hot and dense enough for fusion to occur was the Big Bang itself. Turn back the cosmic clock to the point at which the universe was hot enough for hydrogen to fuse, yet cool enough for the helium products to be stable. At these temperatures, the background fluid energy density is dominated by relativistic particles – photons, neutrinos, anti-neutrinos, electrons, and positrons. Assuming three flavors of neutrino (electron-, muon-, and tau-neutrinos), the relativistic fluid energy density is

$$\rho_{\rm rel} = \rho_\gamma + \rho_{\nu_e} + \rho_{\bar{\nu}_e} + \rho_{\nu_\mu} + \rho_{\bar{\nu}_\mu} + \rho_{\nu_\tau} + \rho_{\bar{\nu}_\tau} + \rho_e + \rho_{\bar{e}} \tag{3.169}$$

(note that the photon, a boson, is its own anti-particle). The energy density of each species i is given by integrating over the momentum phase space, weighting the density by the statistical occupation values:

$$\rho_i = \frac{g_i}{h^3} \int_0^\infty \frac{4\pi p^2 E(p) dp}{e^{E(p)/kT} \pm 1} \tag{3.170}$$

where g_i is the quantum spin degeneracy of the particle species ($g = 1$ for neutrinos and anti-neutrinos, $g = 2$ for photons, electrons, and positrons). T is the black-body temperature of the fluid and $E(p) = p/c$ is the energy for relativistic particles. The $\pm$ sign in the denominator corresponds to the type of statistics used: $+1$ for fermions and -1 for bosons.

An interesting simplification can be made for evaluating the integral when we see

$$\int \frac{x^3 dx}{e^x - 1} - \int \frac{x^3 dx}{e^x + 1} = \int \frac{2x^3 dx}{e^{2x} - 1} = \frac{1}{8} \int \frac{(2x)^3 d(2x)}{e^{(2x)} - 1}. \tag{3.171}$$

Thus the Fermi-Dirac integral and the Bose-Einstein integral are closely related:

$$I_{BE} - I_{FD} = \frac{1}{8} I_{BE} \Rightarrow I_{FD} = \frac{7}{8} I_{BE}. \tag{3.172}$$

Now the evaluation of equation (3.170) is easy because we know that the photon energy density for a black-body flux is

$$\rho_\gamma(T) = aT^4 \qquad \left(a \equiv \frac{\pi^2 k^4}{15 c^3 \hbar^3} \right), \tag{3.173}$$

so if all the particle species are at the same temperature, (3.169) becomes

$$\rho_{\rm rel}(T) = aT^4\left[1+\frac{7}{16}(6)+\frac{7}{8}(2)\right]. \tag{3.174}$$

Neutrinos couple to electrons and photons only through the weak interaction. While they are in thermal equilibrium at the highest temperatures (at the earliest times), they decouple from the rest of the background fluid:

$$\gamma+\gamma \rightleftharpoons \underbrace{e^-+e^+ \rightleftharpoons \nu_e+\bar{\nu_e}}_{\text{weak interaction}}. \tag{3.175}$$

At a temperature of 1 MeV, the electrons and positrons remain in thermal equilibrium with the photon background, but are no longer in equilibrium with the neutrinos. Then, as the temperature drops further below the rest mass energy of an electron, 0.511 MeV, the electrons and positrons annihilate each other, depositing their energy directly into the radiation field. The result is a higher background temperature for the photons than for the neutrinos. This temperature difference is relatively straightforward to calculate.

We assume that before annihilations, all particle species are at the same temperature, $T_{\rm before}$. The neutrino temperature T_ν evolves decreases linearly with redshift,

$$T_\nu = T_{\rm before}\frac{(1+z)}{(1+z_{\rm before})}, \tag{3.176}$$

while the photons temperature jumps from $T_{\rm before}$ to $T_{\rm after}$ and then proceeds to evolve with redshift

$$T_\gamma = T_{\rm after}\frac{(1+z)}{(1+z_{\rm after})}. \tag{3.177}$$

Now throughout the annihilation phase, entropy must be conserved. The entropy density of a relativistic photon-electron-positron gas in a volume R^3 at temperature $T_{\rm before}$ is

$$\begin{aligned} s &= \frac{R^3}{T_{\rm before}}(\rho_{e^-}+\rho_{e^+}+\rho_\gamma+p_{e^-}+p_{e^+}+p_\gamma) \\ &= \frac{4R^3}{3T_{\rm before}}(\rho_{e^-}+\rho_{e^+}+\rho_\gamma) = \frac{11}{3}a(T_{\rm before}R)^3. \end{aligned} \tag{3.178}$$

Then the entropy after annihilation, when only photons remain, is

$$s = \frac{4R^3}{3T_{\rm after}}(\rho_\gamma) = \frac{4}{3}a(T_{\rm after}R)^3. \tag{3.179}$$

Again, since the total entropy is conserved, equating (3.178) and (3.179) gives

$$T_{\rm after} = \left(\frac{11}{3}\right)^{1/3}T_{\rm before} \quad \Rightarrow \quad T_\gamma = \left(\frac{11}{3}\right)^{1/3}T_\nu. \tag{3.180}$$

Near the time of neutrino freeze-out, the predominant weak nuclear reactions taking place are

$$\nu_e+n \rightleftharpoons p+e^-, \qquad E_e-E_\nu = Q \tag{3.181a}$$

$$e^++n \rightleftharpoons p+\bar{\nu}_e, \qquad E_\nu-E_e = Q \tag{3.181b}$$

$$n \rightleftharpoons p+e^-+\bar{\nu}_e, \qquad E_\nu+E_e = Q \tag{3.181c}$$

where the excess mass energy $Q = (m_n - m_p)c^2$ is deposited in the form of the kinetic energy of the products. Just as in chemical reactions in equilibrium, we expect the ratio of the reactant concentrations to be

$$\left[\frac{n_n}{n_p}\right]_{eq} = e^{-Q/kT}. \tag{3.182}$$

If the total forward ($n \to p$) reaction rate is $\lambda_{np} = \lambda_a + \lambda_b + \lambda_c$ and the backwards rate is λ_{pn}, then the neutron concentration should evolve as

$$\frac{dn_n}{dt} = -\lambda_{np} n_n + \lambda_{pn} n_p, \tag{3.183}$$

where at these temperatures, the reaction rate scales as

$$\lambda_{np} = \lambda_0 \left(\frac{T}{T_0}\right)^5. \tag{3.184}$$

Here our notation departs from that used in the remainder of the chapter, with the subscript 0 referring to the time of neutron freeze-out rather than to the present epoch. With the expansion of the universe and the corresponding decrease in temperature, the reaction will eventually fall out of equilibrium. The criterion for equilibrium is

$$-\frac{d}{dt}\ln\left[\frac{n_n}{n_p}\right]_{eq} < \frac{1}{n_n}\frac{d}{dt}[n_n] = -\lambda_0 \left(\frac{T}{T_0}\right)^5. \tag{3.185}$$

From equation (3.182), we have

$$\frac{d}{dT}\ln\left[\frac{n_n}{n_p}\right]_{eq} = \frac{Q}{kT^2} \tag{3.186}$$

and for the evolution of a radiation dominated universe (cf. eqn. 3.147b),

$$t = \left[\frac{3c^2}{32\pi G\left(\frac{43}{8}aT^4\right)}\right]^{1/2}, \tag{3.187}$$

or

$$\frac{dt}{dT} = -\frac{2}{T^3}\left(\frac{3c^2}{172\pi G a}\right)^{1/2}. \tag{3.188}$$

Then the reaction rate λ_0 at the time of freeze-out t_0 is

$$\lambda_0 \equiv \lambda_{np}(t_0, T_0) = \frac{Q}{kT_0}\frac{1}{2t_0}. \tag{3.189}$$

During the radiation era, the time scales as $t \propto T^{-2}$ so

$$\frac{dt}{dT} = -2t_0\frac{T_0^2}{T^3} \tag{3.190}$$

and then (3.185) becomes

$$\frac{d}{dT}\ln n_n = \lambda_0 \left(\frac{T}{T_0}\right)^5 \left(2t_0\frac{T_0^2}{T^3}\right) = \frac{Q}{kT_0}\left(\frac{T}{T_0}\right)^2. \tag{3.191}$$

Solving equations (3.186) and (3.191), we get

$$\ln\left(\frac{n_n}{n_p}\right) = \frac{Q}{kT_0}\left[\frac{1}{3}\left(\frac{T}{T_0}\right)^3 - \frac{4}{3}\right]. \tag{3.192}$$

Then near the time of freeze-out, the neutron to proton ratio should be

$$\left[\frac{n_n}{n_p}\right]_0 = \exp\left(-\frac{4}{3}\frac{Q}{kT_0}\right). \tag{3.193}$$

To determine the actual freeze-out temperature T_0, we need to know the reaction rate in order to solve the equation

$$\lambda_{np} = \frac{Q}{kT^2}\left(\frac{172\pi GaT^4}{3c^2}\right)^{1/2}\frac{T}{2}, \tag{3.194}$$

along with a working cosmology to give $a(T)$. The cosmological model is relatively easy since at these very early times, the universe is radiation dominated, any cosmological constant is negligible, and the expansion is not sensitive to the relative baryon/dark matter density parameters.

A detailed calculation of the reaction rates relies on some quantum mechanics, namely "Fermi's Golden Rule," which says that the reaction rate of a system transitioning between two states i and j in a continuum energy band is given by

$$\lambda = \frac{2\pi}{\hbar}|\langle H_{ij}\rangle|^2\rho(E_j), \tag{3.195}$$

where H is the Hamiltonian and $\rho(E_j)$ is the density of states around the final energy. For our reactions, we have to sum over all possible transitions, i.e. all possible energies of the incoming and outgoing particles. Because of the energy constraints on E_e, E_ν, and Q, we can substitute a Dirac delta function for the density of states and get

$$\lambda_a = \sum_{\substack{\nu_e\,(in)\\ e^-\,(out)}} \frac{2\pi}{\hbar}|\langle H_{n,p}\rangle|^2\delta(Q + E_\nu - E_e) \tag{3.196a}$$

$$\lambda_b = \sum_{\substack{e^+(in)\\ \bar{\nu_e}\,(out)}} \frac{2\pi}{\hbar}|\langle H_{n,p}\rangle|^2\delta(Q + E_e - E_\nu) \tag{3.196b}$$

$$\lambda_c = \sum_{\substack{e^-\,(out)\\ \bar{\nu_e}\,(out)}} \frac{2\pi}{\hbar}|\langle H_{n,p}\rangle|^2\delta(Q - E_\nu - E_e). \tag{3.196c}$$

Considering the continuum of possible electron and neutrino energies, we can change the sum to an

integral over momentum weighted by phase space density $h^{-3}vd^3p$:

$$\lambda_a = \left[\frac{2\pi}{\hbar}(4\pi)^2|\langle H_{n,p}\rangle|^2\frac{g_e g_\nu}{h^6}v^2\right]\int_0^\infty \frac{v_e}{c^4}\frac{p_\nu^2 dp_\nu}{(e^{E_\nu/kT}+1)}\frac{E_e^2}{(e^{-E_e/kT}+1)}, \quad (E_e = E_\nu + Q) \tag{3.197a}$$

$$\lambda_b = \left[\frac{2\pi}{\hbar}(4\pi)^2|\langle H_{n,p}\rangle|^2\frac{g_e g_\nu}{h^6}v^2\right]\int_{\frac{Q}{c}+m_e c}^\infty \frac{v_e}{c^4}\frac{p_\nu^2 dp_\nu}{(e^{-E_\nu/kT}+1)}\frac{E_e^2}{(e^{E_e/kT}+1)}, \quad (E_e = E_\nu - Q) \tag{3.197b}$$

$$\lambda_c = \left[\frac{2\pi}{\hbar}(4\pi)^2|\langle H_{n,p}\rangle|^2\frac{g_e g_\nu}{h^6}v^2\right]\int_0^{\frac{Q}{c}-m_e c} \frac{v_e}{c^4}\frac{p_\nu^2 dp_\nu}{(e^{-E_\nu/kT}+1)}\frac{E_e^2}{(e^{-E_e/kT}+1)}. \quad (E_e = Q - E_\nu) \tag{3.197c}$$

For high temperatures ($kT \gg Q, m_e c^2$), define a dimensionless momentum scaled by temperature, $q \equiv \frac{p_\nu c}{kT}$ and we get

$$\lambda_{np} = \lambda_a + \lambda_b + \lambda_c = \left[\frac{2\pi}{\hbar}(4\pi)^2|\langle H_{n,p}\rangle|^2\frac{g_e g_\nu}{h^6}v^2\right]\frac{(kT)^5}{c^6}\int_{-\infty}^\infty \frac{q^4 dq}{(e^q+1)(e^{-q}+1)}, \tag{3.198}$$

where the integral can be evaluated analytically to give $I_1 = \frac{7}{15}\pi^4$. For later times, when ($kT \ll Q, m_e c^2$), the reaction rate can be calculated by defining a dimensionless momentum scaled by the rest energy of an electron, $r \equiv \frac{p_\nu c}{m_e c^2}$ and a corresponding dimensionless energy, $R \equiv \frac{Q}{m_e c^2}$ and then

$$\lambda_{np} = \left[\frac{2\pi}{\hbar}(4\pi)^2|\langle H_{n,p}\rangle|^2\frac{g_e g_\nu}{h^6}v^2\right]\frac{(m_e c^2)^5}{c^6}\int_0^{R-1} r^2(R-r)^2 dr\left[1 - \frac{1}{(R-r)^2}\right]^{1/2}, \tag{3.199}$$

where this integral can be evaluated numerically to give $I_2 \approx 1.636$. Now the biggest missing piece is the leading Fermi factor,

$$\left[\frac{2\pi}{\hbar}(4\pi)^2|\langle H_{n,p}\rangle|^2\frac{g_e g_\nu}{h^6}v^2\right].$$

But this can be determined by measuring the reaction rate (3.199) experimentally, which is also quite straightforward since the low-temperature reaction rate is just the inverse of the free neutron half-life time:

$$\lambda_{np}(kT << Q) = \frac{1}{\tau_n} = \frac{1}{887\text{sec}}. \tag{3.200}$$

Combining equations (3.194, 3.198, 3.199, and 3.200), the freeze-out temperature can be determined from

$$\frac{Q}{kT_0} = \left[\frac{1}{\tau_n}\frac{I_1}{I_2}\left(\frac{3c^2}{172\pi G a}\right)^{1/2}\frac{k^2 Q^3}{(m_e c^2)^5}\right]^{1/4}. \tag{3.201}$$

Plugging into (3.193), the theoretical neutron-to-proton ratio at the end of the annihilation era is

$$\frac{n_n}{n_p} = \exp\left(-\frac{4}{3}\frac{Q}{kT_0}\right) \approx 0.176. \tag{3.202}$$

So what happens to all these free neutrons flying around? Before most of them get a chance to decay on their own, they most encounter a proton forming deuterium, most of which, in turn, goes

on to form helium:

$$\begin{aligned} n + p &\rightarrow d + \gamma \\ d + d &\rightarrow t + p \\ d + d &\rightarrow {}^3\mathrm{He} + n \\ t + d &\rightarrow {}^4\mathrm{He} + n \\ {}^3\mathrm{He} + d &\rightarrow {}^4\mathrm{He} + p. \end{aligned} \tag{3.203}$$

Some small fraction of the deuterium survives to the present time, but for the most part, all the primordial neutrons end up in ^{4}He. The observed mass fractions in the universe of $X : Y \approx 0.75 : 0.25$ corresponds to an initial ratio of $n_n/n_p = 1/7$, rather close to the result derived above.

3.14 Quasar Absorption Systems

This section draws heavily on concepts already introduced in different contexts. From Sections 2.7 and 2.8 we recall the production and emission of radiation from accretion disks around active galactic nuclei (AGN or quasars). From Sections 3.2 we recall the redshift of photons emitted by receeding sources, and from Section 3.6 we use the basic dynamics of an expanding universe. The results of our discussion of quasar absorption lines will give new insight into the cosmic structures introduced in Section 3.10.

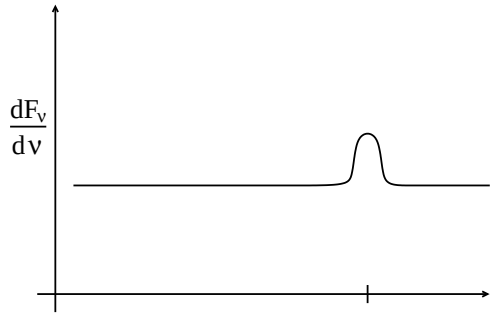

Figure 3.22: The spectrum of a quasar is roughly flat over many decades with a broad emission peak around 1216 Å in the rest frame of the quasar or $1216(1 + z_{\mathrm{em}})$ Å in the rest frame of the observer.

The natural place to begin with is the source: the compact region of superheated gas surrounding the supermassive black hole at the center of a distant galaxy. Due to the many different radiative processes taking place in the accretion disk and surrounding nuclear bulge, including synchrotron radiation, magnetic reconnection, scattering, and absorption, the rest-frame spectrum of the quasar is rather complicated. However, for our purposes it can be approximated as roughly flat around the visible and UV, with a large broad peak around the Lyα (pronounced "Lyman-alpha") emission line at 1216 Å in the emitter's frame (see Fig. 3.22). This line corresponds to the $(2p \rightarrow 1s)$ transition in atomic hydrogen.

The quasar is sufficiently compact that it can accurately be modeled as a single point source at redshift z_{em}. The photons travel from this source through the expanding universe for billions of light years until they finally get detected by a spectrometer on Earth. In this section we are not so interested in what happens at the quasar where the photons are emitted, but rather what happens during their trip through the vast reaches of intergalactic space. More specifically, what is the chance

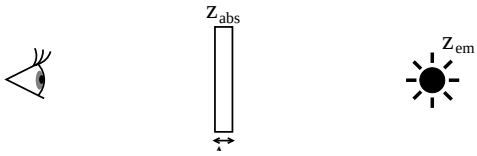

Figure 3.23: The schematic geometry of a quasar at redshift $(1 + z_{\rm em})$, shining through an slab of intevening gas at $(1+z_{\rm abs})$ and detected by a spectrometry at $z = 0$.

of a photon at a given wavelength to be absorbed by an intervening cloud of gas at redshift $z_{\rm abs}$, and what does this tell us? The basic geometry of this problem is shown in Figure 3.23.

The probability of observing such a photon is given by

$$P_{\rm obs} = e^{-\tau}, \tag{3.204}$$

where τ is the total optical depth from the quasar to the observer. In all likelihood, there will be multiple absorbing clouds of gas along the photons path, in which case we just add the optical depths together. Suppose that a photon is absorbed by a hydrogen atom in an intervening cloud at redshift $(1 + z_{\rm abs})$. While absorbed at 1216 Åin the rest frame of the absorbed, it was emitted by the quasar at a *shorter* wavelength. It appears to us as an absorption feature at $\lambda_{\rm obs} = 1216(1 + z_{\rm abs})$ Å (see Fig. 3.24).

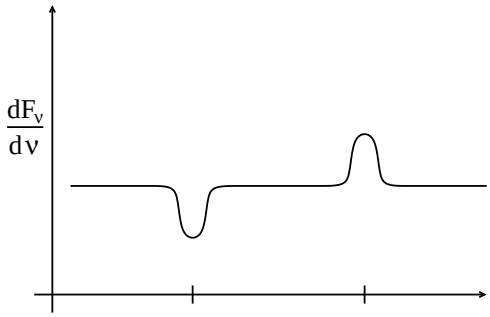

Figure 3.24: A photon emitted at the quasar with $\lambda_{\rm em} < 1216$ Å can get absorbed by a cloud of gas at redshift $(1 + z_{\rm abs})$, appearing to an observer as an absorption feature in the flat quasar spectrum.

For simplicity, we will treat each gas cloud as a plane-parallel slab with constant temperature and density at a single redshift. The differential optical depth is given by

$$d\tau = n_{\rm HI}(z_{\rm abs})\sigma_{\rm HI} dl, \tag{3.205}$$

where $n_{\rm HI}$ is the number density of neutral hydrogen, $\sigma_{\rm HI}$ is the absorption cross section of a single neutral hydrogen atom, and dl is the differential path length in the cloud's rest frame. While it is physically accurate to think of the photon getting absorbed by the hydrogen atom, in reality, the excited atom can only "hold" the photon for about one nanosecond before emitting it again, but since the randomly directed emission is most likely not to be in our line of sight, the absorption cross section can just as easily be thought of as a scattering cross section.

As shown in Figure 3.25, this cross section will actually be a function of wavelength with a Lorentzian shape, peaked at 1216 Å, but exhibiting a finite width due to the quantum mechanical uncertainty of the absorption processes. For the typical lifetime Δt of an excited state, the width of the peak $\Delta\lambda$ scales as

$$\Delta\lambda \sim (\Delta t)^{-1}. \tag{3.206}$$

We can estimate this line width by invoking the energy form of Heisenberg's uncertainty principle:

$$\Delta E \Delta t \geq \hbar. \tag{3.207}$$

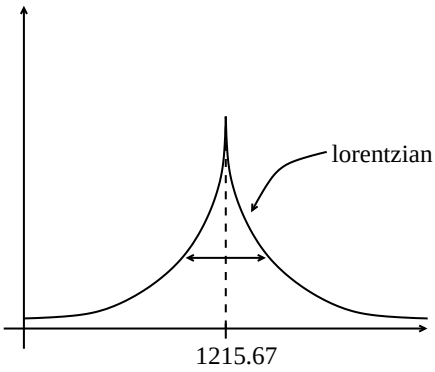

Figure 3.25: In the rest frame of the emitter (or absorber), the Lyα transition line has a Lorentzian profile centered around 1216 Å.

Taking the lifetime from the decay constant of the hydrogen $2p$ state, $\Delta t \approx 1.6 \times 10^{-9}$ sec, so the effective Doppler width of the Lyα absorption line is

$$\Delta v = 76 \text{ m/s}$$

or an energy resolution of

$$R \equiv \frac{\Delta E}{E} \approx 3 \times 10^{-7}. \tag{3.208}$$

For sufficiently cold clouds the dominant effect on the spectral line widths is from this Lorentzian broadening. At finite gas temperature the absorbers have a Gaussian distribution of velocities – lines are therefore Doppler broadened as well. Convolving both effects gives the well-known *Voigt profile*, which is a Lorentzian in the zero temperature limit and a Gaussian in the limit of no atomic oscillator damping.

It should be noted that clouds may be finite in extent, in which case the cross section $\sigma_{\rm HI}$ averaged over variations in temperature and density within the cloud.

From basic thermodynamics, we know the energy of a gas at temperature T is $E = \frac{1}{2}kT$ per degree of freedom in the gas. Since we are assuming a plane parallel system, we only care about the radial velocity dispersion σ_v (not to be confused with the cross section!). For intergalactic hydrogen, typical temperatures are $T_0 \sim 10^4$ °K, which is a minimum of the simultaneous heating and cooling curves (see Spitzer, *Physical Processes in the Interstellar Medium*), corresponding to

$$\sigma_v \approx 10 \left(\frac{T_0}{10^4 \text{K}} \right) \text{ km/s}. \tag{3.209}$$

Now to include the cosmological expansion alluded to above, we rewrite the differential optical depth as

$$d\tau = n_{\rm HI}(z_{\rm abs}, T_0) \sigma_{\rm HI}(\Delta\lambda, \sigma_v) \left(\frac{dl}{dz} \right) dz \quad , \tag{3.210}$$

where we have made explicit the dependence of the cross section upon both the intrinsic width of the line and upon the Doppler broadening. Writing $dl = cdt$, we have the following from equation (3.98):

$$\frac{dl}{dz} = \frac{c}{H(z)} = \frac{c}{(1+z)H_0} \left[\Omega_{m0}(1+z)^3 + \Omega_{k0}(1+z)^2 + \Omega_{\Lambda 0} \right]^{-1/2}. \tag{3.211}$$

By measuring the optical depth between z and $z + dz$ we get a single number that depends upon the the temperature (σ_v) and density $n_{\rm HI}$ in the vicinity of z.

Of course, there are some technical complications. At any given wavelength (redshift), the spectrum has contributions from the thermal broadening and Lorentzian damping from the "wings" of adjacent slabs. So in practice, we often take the optical depth averaging over cross sections, assuming constant temperature and density:

$$\bar{\tau}(z) = \bar{n}_{\rm HI}(\int \sigma_{\rm HI} dz)\frac{dl}{dz}. \tag{3.212}$$

For a fraction of neutral hydrogen $\chi_{\rm HI}$, the average number density of HI at a redshift $(1 + z)$ is simply given by

$$\bar{n}_{\rm HI} \approx 0.8 n_{\rm baryon}\chi_{\rm HI}(1 + z)^3. \tag{3.213}$$

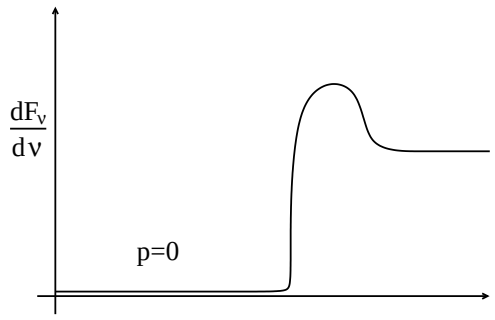

Figure 3.26: The theoretical spectrum of a quasar at a redshift greater than that of reionization $(1 + z_{\rm reion})$. The optical depth through the surrounding neutral hydrogen is much greater than unity.

The immediate problem, as pointed out by Gunn & Peterson and by Bahcall & Saltpeter, is that if the universe were neutral ($x_{\rm HI} \approx 1$), the optical depth of Lyα absorption to most quasars is of the order $\tau_{\rm HI} \sim 10^4$. The entire spectrum below $\lambda = 1216$ Å would be completely absrobed (see Fig. 3.26), which is not at all what we observe.[1] More typical values of the absorption probability are $P_{\rm obs} \sim 0.5$, corresponding to a neutral fraction of only $\chi_{\rm HI} \approx 10^{-4}$.[2] The implication is that at some redshift greater than that of most quasars, the plasma that "recombined" at recombination has since been reionized. There is evidence from the CMB that reionization may have begun as early as $z \sim 20$. There is also evidence from the highest redshift quasars that it may not have ended until $z \sim 6$. Reionization is thought to be caused by the earliest generation of stars or by the first quasars, both of which would send a bath of ionizing ultraviolet radiation through the universe, ionizing the hydrogen that had been neutral through the "dark ages."

What determines this photoionization rate? We can think of the equilibrium as a source-sink problem, where the source of neutral hydrogen comes from free electrons and protons and the sink comes from the ionizing photons. Then we can write the rate of change in local density of neutral hydrogen according to

$$\frac{dn_{\rm HI}}{dt} = \alpha n_e n_{\rm HII} - \Gamma n_{\rm HI} \tag{3.214}$$

or

$$\frac{d}{dt}\ln n_{\rm HI} = \alpha\frac{n_e n_{\rm HII}}{n_{\rm HI}} - \Gamma, \tag{3.215}$$

[1] While this estimate assumes enough baryons to close the universe, even with the present day measurements of $\Omega_b = 0.047$, the optical depth to absorption would still be much greater than unity.

[2] This absorption manifests itself not as a uniform depression shortward of Lyα but as a "forest" of lines very much denser than the metallic lines observed longward of Lyα.

where α is the recombination coefficent (in units of cm^3/s) and Γ is the integrated ionization rate:

$$\Gamma = 4\pi \int_{h\nu > 13.6 \text{ eV}}^{\infty} \frac{J_\nu}{h\nu} \sigma(\nu) d\nu \sim 10^{-12} \text{ s}^{-1}, \tag{3.216}$$

with J_ν the specific intensity of the background ionizing radiation.

For the temperatures and densities we are interested in, $\alpha \sim 4 \times 10^{-13}$ and $n_e \sim 10^{-5}$, so we get an equilibrium solution in equation (3.215) as long as the neutral hydrogen fraction is small: $\chi_{\rm HI} \sim 10^{-5} - 10^{-4}$, as we saw above. There is a problem, however, in that, for $z > 3$, the quasar density drops off sharply, and the UV photons from O and B type stars have a relatively difficult time getting out of the dense star-forming regions. So it is still an open question as to what caused reionization at $z > 6$.

Finally it should be noted that while we focused our discussion here on the Lyα transition of neutral hydrogen, most quasar spectra actually show an large collection of different absorption lines, from many other elements in addition to hydrogen. Coupled with the fact that we have multiple gas clouds at multiple redshifts, degeneracies arise that can significantly complicate the analysis. Yet at the same time these other lines give us a great deal of information about the composition, density, and structure of the universe as it evolved over many billions of years.

3.15 Past and Present Superposed

Astronomers tend to take for granted the remarkable fact that when we look at the sky we see many different historical epochs simultaneously. Our archeologist and geologist colleagues must think us lucky, and perhaps a bit dimwitted for not having understood more, having been handed the universe on a silver platter.

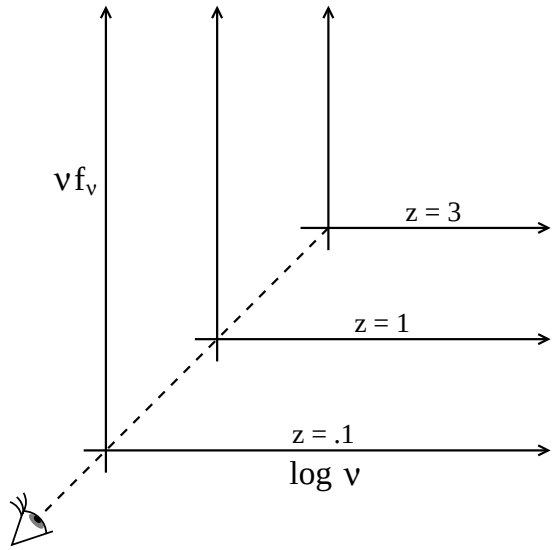

Figure 3.27: When we look at the sky we see the universe at very different times. The spectrum will be quite different depending on the ages of the stars formed, the absorption and re-emission by dust, and the relative importance of AGNs.

One minor problem is that all these different epochs are superposed (as seen in Fig. 3.27), and we need a method of deprojecting them. We have that in the form of Doppler shifts, if an object has measurable lines and if the object in question is sufficiently bright that we can afford to disperse its

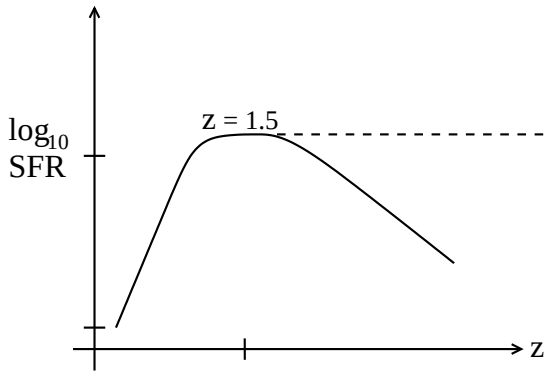

Figure 3.28: Star formation rate as a function of redshift. Observations at optical wavelengths indicate a peak, but these may need correction for absorption by dust.

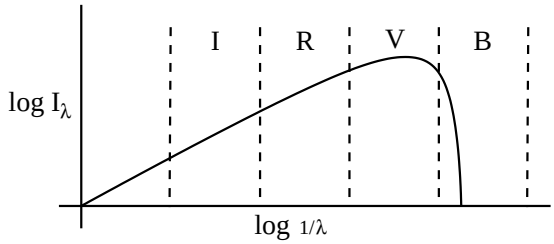

Figure 3.29: By observing a blackbody in different filters, one can determine its temperature.

photons. Until now Doppler shifts have been obtained almost exclusively at optical and near infrared wavelengths.

In the course of our efforts we have understood a number of ways in which the past was different from the present. There was less clustering of galaxies in the past and there were many more active galactic nuclei, quasars and radio galaxies. There are other ways as well.

The fundamental plane for elliptical galaxies (Section 1.3) was in the past at higher surface brightness for fixed effective radius r_e and velocity dispersion σ. The observations are consistent with the very early formation of the stars in ellipticals and the slow fading of the stellar component as stars evolve off the main sequence.

The luminosity function for galaxies has likewise evolved with time, somewhat differently for red and blue galaxies. The changes have been modest for the redder bulge dominated systems more dramatic for the bluer disk dominated systems. In particular, the number of faint blue galaxies increases dramatically looking backward in time. The obvious interpretation is an increased rate of star formation (SFR), rising steadily backward in time until redshift $z = 1.5$ (see Fig. 3.28). What happens at yet higher redshift is still a matter of debate. A fraction of the optical and UV light from star formation is absorbed by dust and reradiated in the mid-infrared. Whether the star formation rate falls at $z > 1.5$ or continues to rise or flattens out depends upon corrections for the absorbed flux.

Prior to the mid-1990s observations of galaxies were largely limited to redshifts less than unity. Two independent developments pushed those limits out to $z = 6.5$. The first was the development by Steidel of the "Lyman break" technique. If the spectra of stars were black bodies, we could measure their temperatures by observing them with broadband filters – for example Johnson's $U, B, V, R,$ & I – computing the relative fluxes in the different filters as shown in Figure 3.29. If we knew their intrinsic temperatures, we would be able to compute a "photometric" redshift. While the spectral energy distributions of galaxies are not, in general, those of black bodies, they nonetheless have features that permit such photometric redshift estimates, with increasing accuracy as the number of

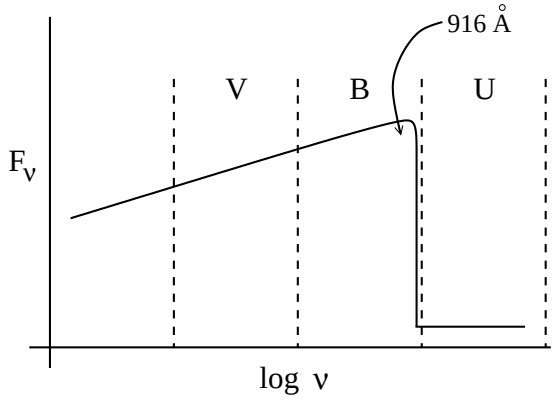

Figure 3.30: The spectra of hot stars show a precipitous drop in flux at the Lyman limit due to the increasing opacity at $h\nu > 13.6$ eV. This suggests a scheme for finding high redshift galaxies.

filters increases.

It is our good luck that stellar spectra deviate from those of black bodies in a way that makes redshift determination even easier. The stellar opacity shortward of the Lyman limit at 912 Å is very high, causing a sharp drop in the flux. The U filter is centered at 3600 Å. For objects at $z = 3$ and slightly higher redshift there is little or no flux in the U filter but plenty of flux in the B and V filters as seen in Figure 3.30. The galaxies are called "UV dropouts." The same technique can be used at yet higher redshifts by using triplets of filters that lie further to the red. Spectra have been taken of the brighter UV dropouts, showing absorption lines typical of stellar winds in O and B stars. These are evidently galaxies with substantial SFRs.

The second development was sociological. There have always been unspoken rules about how much telescope time was too much time to ask for. In 1995 the then director of the Hubble Space Telescope Science Institute, Bob Williams, decided to devote 300 orbits to four filter photometry of what was called the Hubble Deep Field. Photometric redshifts were determined for thousands of galaxies. Ironically, only now, with yet deeper surveys, have we been able to obtain large samples of galaxies in the range $1 < z < 3$. The strength of the Lyman break made it possible to leapfrog this range into higher redshifts.

There is some concern that Lyman break surveys are missing those galaxies that are *not* forming stars as well as those galaxies that are forming stars only in the depths of dark clouds, where their light is absorbed by dust and re-emitted in the thermal infrared. Averaging over nearby galaxies, at most a third of their bolometric luminosity is in the mid-infrared. But most of this comes from a small number of IR bright galaxies star forming galaxies. There is reason to think that a larger fraction of the galaxies at high redshift would be IR bright. Most starlight may never escape from the galaxies in which it is emitted. Starting in the late 1990s the detectors on submillimeter telescopes became sufficiently sensitive to detect $z \sim 3$ galaxies at an observed wavelength of 850 μm. These objects have very high star formation rates and seem to fill in the decline in SFR inferred from optically selected galaxies, although the numbers are still quite small. Objects found in the mid-infrared with the Spitzer infrared telescope may have a similar effect.

A major shortcoming of the submillimeter observations is their low resolution. Diffraction limits the positional accuracy to roughly 10 arcseconds, producing error boxes that include large numbers of optical sources. This can be improved by looking for submillimeter sources behind lensing clusters of galaxies. One can also use X-ray and radio observations to identify which of the many candidates in an error box is the submillimeter galaxy.

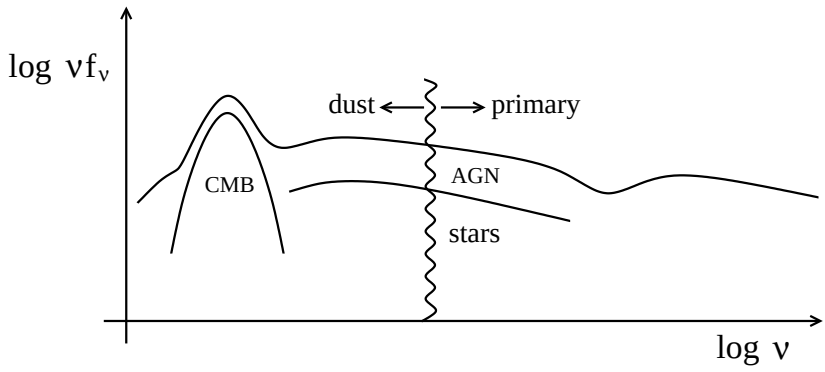

Figure 3.31: A schematic spectrum of the night sky decomposed into contributions from stars, AGN, and the CMB. The CMB dominates by a factor of 100 at its peak. AGN contributes one tenth as much as stars except at x-ray frequencies.

Interestingly, deep X-ray searches have produced relatively few high redshift objects. The redshifts for the AGN in these surveys are mostly less than unity.

While astronomy had its origins at optical wavelengths we now carry out deep surveys of the extragalactic sky over almost every octave of the electromagnetic spectrum. Averaging over the entire sky we get a spectrum that looks something like the cartoon in Figure 3.31. We plot the quantity νI_ν which has the property that it is flat if the flux from the sky is the same over every decade of wavelength. The integrated brightness of the extragalactic sky is dominated by the cosmic microwave background (much more so than in the figure). At infrared and optical wavelengths we have $\nu I_\nu \sim 20$ nanoWatts per steradian.

We have attempted to account, at least roughly, for the relative contributions of stars and active galactic nuclei. The contribution from stars dominates that from AGN at most wavelengths longward of the X-ray region. It is unclear how the balance between stars and AGN changes with redshift. At wavelengths longward of 3 μm the light is re-radiated by dust. At shorter wavelengths it comes to us directly from the atmospheres of stars and quasars.

In constructing Figure 3.31 we have plotted the extragalactic sky brightness (ESB). There are two components to this – resolved sources and an unresolved background (the extragalactic background light, acronymized as EBL). At different wavelengths the unresolved background makes very different contributions – almost none at optical wavelengths and almost everything at mid-IR wavelengths. The measurements in the mid-IR are mostly from the DIRBE and FIRAS experiments on the COBE satellite, which had almost no resolution. Most of the mid-IR light probably comes from higher redshift than the optical light. The near UV again comes from higher redshift than the optical, the result of higher star formation rates at higher redshift.