Chapter 10

Cosmology

Cosmology is the study of the universe, or cosmos, regarded as a whole. Some questions addressed by cosmologists are "What is the universe made of? Is it finite or infinite in spatial extent? Did it have a beginning at some time in the past? Will it come to an end at some time in the future?"

In addition ot dealing with Very Big Things, cosmology also deals with very small things. Early in its history, as we'll see later on, the universe was very hot in addition to being very dense, and interesting particle physics phenomena were occurring. Thus, a brief review of elementary particle physics will be useful as a preface to this chapter. Particle physicists tend to measure energy in units of electron volts (eV).[1] The conversion between electron volts and joules is $1\,\mathrm{eV} = 1.60 \times 10^{-19}\,\mathrm{J}$.

The most cosmologically important particles are listed in Table 10.1. The

[1]And in multiples thereof, such as keV (10^3 eV), MeV (10^6 eV), GeV (10^9 eV), and TeV (10^{12} eV).

Table 10.1: Particle Properties

particle	symbol	rest energy (MeV)	charge
proton	p	938.3	+1
neutron	n	939.6	0
electron	e^-	0.511	-1
neutrino	ν_e,ν_μ,ν_τ	$< 2 \times 10^{-6}$	0
photon	γ	0	0
dark matter	?	?	0

objects that surround us in everyday life are made of protons, neutrons, and electrons. Protons and neutrons are both examples of **baryons**, where a baryon is defined as a particle made of three quarks.[2] A proton (p) contains two "up" quarks, each with a charge $q = +2/3$, and a "down" quark, with a charge of $q = -1/3$. A neutron (n) contains one "up" quark and two "down" quarks. A proton has a mass (or equivalently, a rest energy) that is 0.1% less than that of a neutron. A free neutron is is unstable, decaying into a proton with a decay time of $\tau_n = 940\,\mathrm{s}$, about a quarter of an hour.

Electrons (e^-) are examples of **leptons**, a class of elementary particles that are not made of quarks.[3] The mass of an electron is small compared to that of a proton or neutron; the electric charge of an electron is equal in magnitude, but opposite in sign, to that of a proton. On large scales, the universe seems to be electrically neutral, with equal numbers of protons and electrons. The component of the universe made of atoms, molecules, and ions is called **baryonic matter**, since only the baryons contribute significantly to the mass density.

Neutrinos (ν) are also leptons. Neutrinos have no electric charge, and interact with other particles only through the weak nuclear force or gravity. There are three types, or flavors, of neutrinos: electron neutrinos (ν_e), muon neutrinos (ν_μ), and tau neutrinos (ν_τ). Although recent experiments indicate that the different neutrino types have different masses, those masses must be small compared to the electron mass, with $m_\nu c^2 < 2\,\mathrm{eV}$ being the approximate upper limit on the rest energy.

A particle known to be massless is the photon (γ). Unlike neutrinos, photons interact readily with electrons, protons, and neutrons. Although photons are massless, they have an energy $E = hc/\lambda$, where λ is the wavelength.

The most mysterious component of the universe is the dark matter. As discussed in section 6.2, some of the dark matter may be baryonic (in the form of brown dwarfs or other dense, dim MACHOs). Some of the dark matter, but not much, is contributed by the lightweight neutrinos. It is likely that some of the dark matter is contributed by WIMPs, weakly interacting massive particles that are far more massive than neutrinos.

---

[2] "Baryon" comes from the Greek root "barus", meaning heavy or weighty. A *barometer* measures the weight of the atmosphere, and the *barycenter* of the Local Group (see Figure 9.2) is the center of gravity, or center of mass.

[3] "Lepton" comes from the Greek root "leptos", meaning small or thin. In Greece, the euro cent (1/100 of a euro) is called a *lepton*, since it is the smallest coin minted.

10.1 Basic Cosmological Observations

Observations of the universe around us have led cosmologists to adopt the **Hot Big Bang** model, which states that the universe has expanded from an initial hot and dense state to its current cooler and lower-density state, and that the expansion is continuing today. Several observations have contributed to the acceptance of the Hot Big Bang model. Many of these observations are recent, and depend on sophisticated technology. However, the first observation on which the Hot Big Bang is based is very ancient, and requires nothing more sophisticated than your own eyes.

The first observation underpinning modern cosmology is this: **The night sky is dark.** When you go outside on a clear night and look upward, you see scattered stars on a dark background. The fact that the night sky is dark at visible wavelengths, rather than being uniformly bright with starlight, is known as Olbers' Paradox, after the astronomer Heinrich Olbers, who wrote a paper on the subject in the year 1826.[4] Olbers was not actually the first person to think about Olbers' Paradox; as early as 1576, Thomas Digges was worrying in print about the darkness of the night sky.

Why should the darkness of the night sky be paradoxical? First, consider the light from a single star of luminosity L at a distance r. The flux from the star is given by the inverse square law:

$$f = \frac{L}{4\pi r^2} \ . \tag{10.1}$$

The solid angle subtended by the star is also inversely proportional to its distance; if the star's radius is $r_\star$, its angular area, in steradians, is

$$d\Omega = \frac{\pi r_\star^2}{r^2} \ . \tag{10.2}$$

This means that the surface brightness $\Sigma_\star$ of the star, in watts per square meter per steradian, is independent of distance:

$$\Sigma_\star = \frac{f}{d\Omega} = \frac{L}{4\pi^2 r_\star^2} = 2.0 \times 10^7 \,\mathrm{W\,m^{-2}\,ster^{-1}} \left(\frac{L}{L_\odot}\right)\left(\frac{r_\star}{r_\odot}\right)^{-2} \ . \tag{10.3}$$

For a sun-like star, this corresponds to $\Sigma_\star \sim 0.5\,\mathrm{mW\,m^{-2}\,arcsec^{-2}}$. Since even nearby stars, like those of the Alpha Centauri system, have angular

[4]Closer to home, Olbers is also known as the discoverer of the asteroids Pallas and Vesta.

areas $d\Omega < 10^{-5}\,\mathrm{arcsec}^2$, any individual star will cover only a tiny fraction of the celestial sphere and contribute only a tiny flux here at Earth. But what if the universe stretches to infinity in all directions?

Let $n_\star$ be the average number density of stars in the universe, and let L and $r_\star$ be the average stellar luminosity and radius. Consider a thin spherical shell of radius r and thickness dr centered on the Earth (Figure 10.1). The

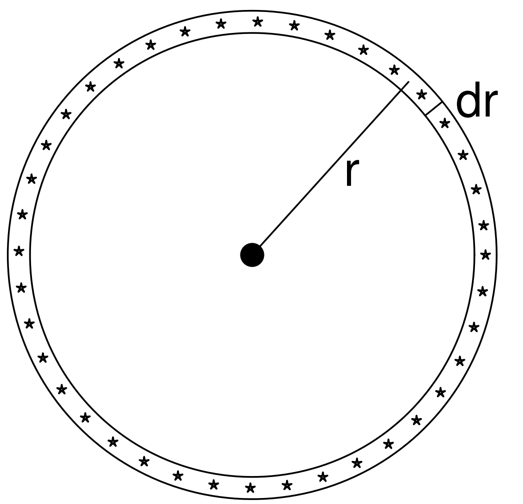

Figure 10.1: A star-filled spherical shell.

total number of stars in the shell will be

$$dN_\star = n_\star 4\pi r^2 dr \ . \tag{10.4}$$

Since each star covers an angular area $d\Omega = \pi r_\star^2/r^2$, the fraction of the shell's area covered with stars will be

$$dF = \frac{dN_\star d\Omega}{4\pi} = n_\star \pi r_\star^2 dr \ , \tag{10.5}$$

independent of the radius r of the shell. The total fraction of the sky covered with stars within a distance r of us will then be

$$F = \int_0^r dF = n_\star \pi r_\star^2 \int_0^r dr = n_\star \pi r_\star^2 r \ . \tag{10.6}$$

The coverage becomes complete when $F \approx 1$, corresponding to a distance

$$r_{\mathrm{olbers}} \approx \frac{1}{n_\star \pi r_\star^2} \ . \tag{10.7}$$

Thus, if the universe extends for a distance $r \geq r_{\rm olbers}$, the sky must be uniformly bright, with a surface brightness equal to that of a typical star.

Obviously, the sky is *not* uniformly bright; at least one of the assumptions that went into our calculation must be wrong. One assumption we made was that $n_\star$ and L were independent of distance. This might be wrong. Distant stars might be less luminous or less numerous than nearby stars.

A second assumption is that the universe is bigger than $r_{\rm olbers}$. This might be wrong. If the universe only stretches to a distance $r_0 < r_{\rm olbers}$, then the fraction of the night sky covered by stars will be

$$F \approx n_\star \pi r_\star^2 r_0 < 1 \ , \tag{10.8}$$

and the average surface brightness of the sky will be

$$\Sigma_{\rm sky} = F\Sigma_\star \approx n_\star \pi r_\star^2 r_0 \frac{L}{4\pi^2 r_\star^2} \approx \frac{n_\star L r_0}{4\pi} \ . \tag{10.9}$$

This result will also be found if the universe is infinitely large, but empty of star beyond a distance r_0.

A third assumption, slightly more subtle, is that the universe is infinitely old. This might be wrong. If the universe has a finite age t_0, then the greatest distance we can see is $r_0 \approx ct_0$, and the average surface brightness of the sky will be

$$\Sigma_{\rm sky} \approx \frac{n_\star L c t_0}{4\pi} \ . \tag{10.10}$$

This result will also be found if the universe is eternally old, but has only contained stars for a finite time t_0.

A fourth assumption made in computing the surface brightness is that the flux of stars is given by the inverse square law of equation (10.1). This might be wrong. The assumption that $f \propto r^{-2}$ follows directly from Euclid's laws of geometry. However, on large scales, the universe is under no obligation to be Euclidean. In some non-Euclidean geometries, the flux falls off more rapidly than an inverse square law.[5]

The darkness of the night sky caused astronomers to question many of their assumptions; an infinitely large, infinitely old, Euclidean universe can't stand up to close scrutiny.

[5] Of course, in other non-Euclidean geometries, the flux falls off less rapidly than an inverse square law, which will only increase the problem.

The second observation on which modern cosmology is based is the Hubble law: **Galaxies show a redshift proportional to distance.** As noted in section 7.4, the Hubble law is a natural consequence of homogeneous, isotropic expansion. If the expansion is perfectly homogeneous and isotropic, then the distance $r(t)$ between any two points can be written in the form

$$r(t) = a(t) r_0 \ , \tag{10.11}$$

where $r_0 \equiv r(t_0)$ is the separation at the current time t_0, and $a(t)$ is a dimensionless function known as the **scale factor**.[6] The homogeneity and isotropy of the expansion imply that $a(t)$ is not a function of position or direction, but only of the time t. The distance between the two points will increase at the rate

$$v(r) = \dot{a} r_0 = \frac{\dot{a}}{a(t)}[a(t) r_0] = \frac{\dot{a}}{a(t)} r(t) \ . \tag{10.12}$$

Thus, the velocity-distance relation takes the form of the Hubble law: $v(t) = H(t) r(t)$, where $H(t) \equiv \dot{a}/a$. The function $H(t)$ is called the **Hubble parameter**. Its value at the present day, $H_0 \equiv H(t_0)$, is called the **Hubble constant**.

Note how the Hubble law ties in with Olbers' Paradox. If the universe is of finite age, $t_0 \sim H_0^{-1}$, then we expect that the **horizon distance**, the maximum distance from which light has had time to reach us, will be of order $t_0 \sim ct_0 \sim c/H_0$. The luminosity density of starlight in the universe, computed in section 9.3, is

$$n_\star L = \rho_L = 2.3 \times 10^8 L_\odot \,\mathrm{Mpc}^{-3} \ . \tag{10.13}$$

The average surface brightness of the sky should then be approximately

$$\begin{aligned} \Sigma_{\rm sky} \ &\sim \ \frac{n_\star L}{4\pi} \frac{c}{H_0} \sim \frac{(2.3 \times 10^8 L_\odot \,\mathrm{Mpc}^{-3})(4300 \,\mathrm{Mpc})}{4\pi} \\ &\sim \ 8 \times 10^{10} L_\odot \,\mathrm{Mpc}^{-2} \,\mathrm{ster}^{-1} \sim 3 \times 10^{-8} \,\mathrm{W\,m}^{-2} \,\mathrm{ster}^{-1} \ . \end{aligned} \tag{10.14}$$

When we compare this to the surface brightness of a sun-like star,

$$\Sigma_\star \approx 2.0 \times 10^7 \,\mathrm{W\,m}^{-2} \,\mathrm{ster}^{-1} \ , \tag{10.15}$$

[6]Zeilik & Gregory use the letter R to designate the scale factor.

we find that $\Sigma_\star \sim 6 \times 10^{14}\Sigma_{\rm sky}$. Thus, for the entire sky to have a surface brightness as great as the Sun's, the universe would have to be 600 trillion times older than it is – *and* you'd have to keep the stars shining during all that time.

The primary resolution to Olbers' Paradox is that the universe has a finite age. Stars beyond the horizon distance are invisible to us because their light hasn't had enough time to reach us. A secondary contribution to the darkness of the night sky is the redshift of distant light sources, close to the horizon, which reduces their flux as measured from Earth.

A third observation on which modern cosmology is based was made in the year 1965: **The universe is filled with a Cosmic Microwave Background.** The discovery of the Cosmic Microwave Background (CMB) by Arno Penzias and Robert Wilson has entered cosmological folklore. Using a microwave antenna at Bell Labs, they discovered a slightly stronger signal than they expected from the sky. The extra signal was isotropic and independent of time. After removing all sources of noise that they could (Figure 10.2), they realized that they were truly detecting an isotropic background of microwave radiation. More recently, the Cosmic Background Ex-

Figure 10.2: Wilson (left) and Penzias (right), scrubbing pigeon droppings from their antenna.

plorer satellite revealed that the Cosmic Microwave Background has a Planck spectrum (*BA*, sec. 6.7),

$$I_\nu = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/kT_0} - 1} , \tag{10.16}$$

with a temperature

$$T_0 = 2.725 \pm 0.001\,\mathrm{K} . \tag{10.17}$$

That is, the CMB is just what we would see if we were inside a hollow blackbody at a temperature of 2.725 K. The current energy density of the Cosmic Microwave Background is

$$u_0 = \frac{4\sigma}{c}T_0^4 = 4.17 \times 10^{-14}\,\mathrm{J\,m^{-3}} = 0.260\,\mathrm{MeV\,m^{-3}} , \tag{10.18}$$

where σ is the Stefan-Boltzmann constant. The average energy of a single CMB photon, integrating over the complete Planck spectrum, is

$$\varepsilon_0 = 2.7kT_0 = 6.34 \times 10^{-4}\,\mathrm{eV} . \tag{10.19}$$

The average photon energy ε_0 corresponds to a wavelength $\lambda_0 = hc/\varepsilon_0 \approx 2\,\mathrm{mm}$, in the microwave range of the electromagnetic spectrum (hence the name Cosmic *Microwave* Background). The number density of CMB photons is

$$n_0 = \frac{u_0}{\varepsilon_0} = \frac{2.60 \times 10^5\,\mathrm{eV\,m^{-3}}}{6.34 \times 10^{-4}\,\mathrm{eV}} = 4.11 \times 10^8\,\mathrm{m^{-3}} . \tag{10.20}$$

In an expanding Big Bang universe, cosmic background radiation arises naturally if the universe was initially very hot in addition to being very dense. Suppose the initial temperature was $T \gg 10^4\,\mathrm{K}$. At such high temperatures, the baryonic matter in the universe was completely ionized (Figure 10.3), and scattering of photons from the free electrons rendered the universe opaque. A dense, hot, opaque medium produces blackbody radiation, with a Planck spectrum. However, as the universe expanded, it cooled. When the temperature dropped to $T \sim 3000\,\mathrm{K}$, ions and free electrons combined to form neutral atoms. When the universe no longer contained a significant number of free electrons, the liberated blackbody photons started streaming through the universe, without further scattering.

At the time the universe became transparent, the temperature of the background radiation was $T \sim 3000\,\mathrm{K}$, about the temperature of an M star's photosphere. The temperature of the background radiation today is $T_0 =$

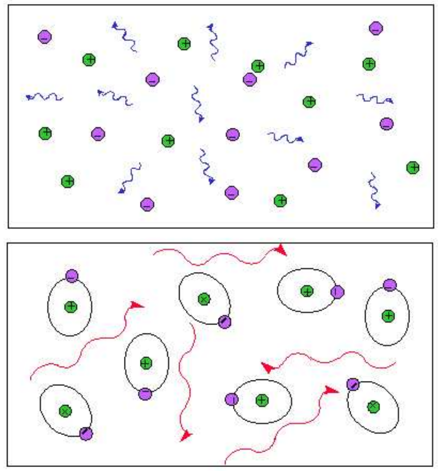

Figure 10.3: Above: the opaque universe before recombination. Below: the transparent universe after recombination.

2.725 K, a factor of 1100 lower. Why has the background radiation cooled? It's a consequence of the expansion of the universe.

Consider a region of volume V that expands along with the universe, so that $V(t) \propto a(t)^3$, where $a(t)$ is the scale factor. The blackbody radiation within this volume can be thought of as a photon gas with energy density $u = (4\sigma/c)T^4$ and pressure $P = u/3$. The photon gas within our volume obeys the first law of thermodynamics:

$$dQ = dE + PdV \ , \tag{10.21}$$

where dQ is the amount of heat flowing into or out of the volume, and dE is the change in the internal energy of the photon gas. In a homogeneous and isotropic universe, there is no flow of heat, since everything is at the same temperature; thus, $dQ = 0$. The first law of thermodynamics, applied to a gas in an expanding universe, then becomes

$$\frac{dE}{dt} = -P(t)\frac{dV}{dt} \ . \tag{10.22}$$

For the photons of the Cosmic Microwave Background, the internal energy is $E(t) = u(t)V(t) = (4\sigma/c)T(t)^4V(t)$ and the pressure is $P(t) = (1/3)u(t) = (4\sigma/3c)T(t)^4$. Equation (10.22) then becomes

$$\frac{4\sigma}{c}\left(4T^3\frac{dT}{dt}V + T^4\frac{dV}{dt}\right) = -\frac{4\sigma}{3c}T^4\frac{dV}{dt} \ , \tag{10.23}$$

or, with a little algebraic manipulation,

$$\frac{1}{T}\frac{dT}{dt} = -\frac{1}{3V}\frac{dV}{dt} \ . \tag{10.24}$$

Since $V(t) \propto a(t)^3$ as the universe expands, equation (10.24) can be rewritten in the form

$$\frac{1}{T}\frac{dT}{dt} = -\frac{1}{a}\frac{da}{dt} \ , \tag{10.25}$$

or

$$\frac{d}{dt}(\ln T) = -\frac{d}{dt}(\ln a) \ . \tag{10.26}$$

This implies the simple relation $T(t) \propto a(t)^{-1}$; the temperature of the Cosmic Microwave Background drops as the universe expands. Note that it also implies $\varepsilon(t) \propto a(t)^{-1}$ for the average photon energy and $\lambda(t) \propto a(t)$ for the average photon wavelength. The background radiation has dropped in temperature by a factor of 1100 since the universe became transparent because the scale factor has grown by a factor of 1100 since then.

The observations we have noted so far – the dark night sky, the Hubble law, and the Cosmic Microwave Background – all fit neatly within the framework of the **Hot Big Bang** model for the universe, in which the universe was initially very hot and dense, but has since cooled as it expanded. An exact treatment of how the universe expands requires knowledge of General Relativity. If you happen to suffer from relativity-phobia, don't panic, though. Many of the most important aspects of the expanding universe can be explained using purely Newtonian dynamics.

10.2 Cosmology à la Newton

Let's compute, using Newton's Law of Gravity and Second Law of Motion, how the scale factor $a(t)$ depends on time. Consider a homogeneous sphere of matter, with fixed total mass M. The sphere is expanding (or contracting) homogeneously, so that its radius $r(t)$ is changing with time (Figure 10.4). Place a test mass, of infinitesimal mass m, at the surface of the sphere. The gravitational acceleration of the test mass will be

$$\frac{d^2r}{dt^2} = -\frac{GM}{r(t)^2} \ . \tag{10.27}$$

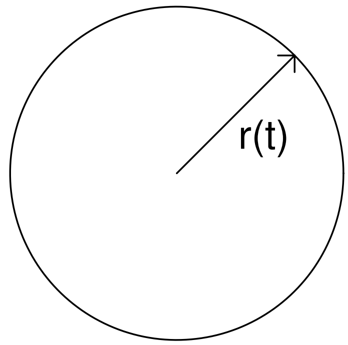

Figure 10.4: A sphere of fixed mass M and variable radius $r(t)$.

If we multiply each side of equation (10.27) by dr/dt and integrate over time, we find

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 = \frac{GM}{r(t)} + k \ , \tag{10.28}$$

where k is the constant of integration. Equation (10.28) is an energy conservation statement. The sum of the kinetic energy per unit mass and the gravitational potential energy per unit mass is a constant (k) for a bit of mass at the sphere's surface.[7]

The future of an expanding, self-gravitating sphere falls into one of three classes, depending on the sign of the constant k. First, consider the case $k > 0$. In this case, the right hand side of equation (10.28) is always positive. Therefore, the left hand side of the equation never goes to zero, and the expansion continues forever. Second, consider the case $k < 0$. In this case, the right hand side of equation (10.28) goes to zero at a maximum radius $r_{\rm max} = GM/k$, and the expansion stops. However, at the maximum radius, the acceleration, given by equation (10.27), is still negative, so the sphere will then contract. Third and last, consider the case $k = 0$. This is the boundary case in which dr/dt asymptotically approaches zero as $t \to \infty$.

The three possible fates of an expanding sphere in a Newtonian universe are analogous to the three possible fates of a ball thrown upward from the

[7]Note also that the expansion velocity, dr/dt, enters equation (10.28) only as its square. This means that a contracting sphere ($dr/dt < 0$) is simply a time reversal of an expanding sphere ($dr/dt > 0$).

Earth's surface. First, the ball can be thrown upward with a speed greater than the escape velocity $v_{\rm esc}$. In this case, the ball goes upward forever. Second, the ball can be thrown upward with a speed less than the escape velocity. In this case, the ball reaches a maximum height, then falls back down. Third and last, the ball can be thrown upward with a speed exactly equal to $v_{\rm esc}$. In this case, the speed of the ball asymptotically approaches zero as $t \to \infty$.

Equation (10.28), describing an expanding (or contracting) sphere, can be rewritten in such a way that it applies to a sphere of arbitrary radius and mass. The mass M, which is constant, can be written in the form

$$M = \frac{4\pi}{3}\rho(t)r(t)^3 \ . \tag{10.29}$$

Since the expansion is isotropic about the center of the sphere, we can write

$$r(t) = a(t)r_0 \ , \tag{10.30}$$

where $a(t)$ is the dimensionless scale factor and r_0 is the current radius of the sphere. Using these relations, equation (10.28) can be written in the form

$$\frac{1}{2}r_0^2\dot{a}^2 = \frac{4\pi}{3}Gr_0^2\rho(t)a(t)^2 + k \ , \tag{10.31}$$

or, dividing each side of the equation by $r_0^2 a^2/2$,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(t) + \frac{2k}{r_0^2}\frac{1}{a(t)^2} \ . \tag{10.32}$$

The left hand side of equation (10.32) is the square of the Hubble parameter, $H(t) \equiv \dot{a}/a$. Thus, we now have an equation that links the expansion rate of the universe to its mass density ρ. Equation (10.32) is called the **Friedmann equation**, after the Russian cosmologist Alexander Friedmann, who first found it (using a the relativistically correct derivation) in the early 20th century.

For a given value of the Hubble parameter, $H(t)$, there is a **critical density** $\rho_c(t)$ for which $k = 0$, and the universe is exactly on the boundary between eternally expanding ($k > 0$) and eventually recollapsing ($k < 0$). The value of the critical density is, from equation (10.32),

$$\rho_c(t) = \frac{3H(t)^2}{8\pi G} \ . \tag{10.33}$$

At the present moment in the real universe, $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and the value of the critical density is

$$\rho_{c,0} = \frac{3H_0^2}{8\pi G} = 9.2 \times 10^{-27}\,\mathrm{kg\,m^{-3}} = 1.4 \times 10^{11} M_\odot\,\mathrm{Mpc^{-3}}\ . \qquad (10.34)$$

If the average density of the universe is greater than this value, then (if our Newtonian analysis is adequate) the universe will eventually collapse in a "Big Crunch". If the average density is less than or equal to this value, then it will expand forever in an increasingly tenuous "Big Chill". Is the average density greater than or less than $\rho_{c,0}$? It's not immediately obvious. Although $\rho_{c,0}$ is equivalent to a density of one hydrogen atom per 200 liters – much more tenuous than even the lowest density coronal gas in the interstellar medium – you must remember that most of the universe consists of very low density voids.

Scientists sometimes joke that they are searching for a theory of the universe that can fit on the front of a T-shirt. If our universe is strictly Newtonian, an appropriate T-shirt slogan is: "DENSITY IS DESTINY" (Figure 10.5). Like all terse summaries of complex concepts, the slogan "Den-

Figure 10.5: Newtonian teddy bear.

sity is destiny" requires a qualifying footnote. In this case, the footnote is

"*Applies to Newtonian universes only; void if a *cosmological constant* is present." A cosmological constant is an entity that provides a positive acceleration ($\ddot{a} > 0$) to the expansion of the universe. The cosmological constant was introduced by Einstein in the context of general relativity. Since the Newtonian view is that gravity is always an attractive force ($\ddot{a} < 0$), it will be necessary for us to dabble in general relativity in order to understand the cosmological constant and the possibility of an accelerating universe.

10.3 Cosmology à la Einstein

In Newton's view of the universe, space is static, unchanging, and Euclidean. In Euclidean, or "flat", space, all the axioms and theorems of plane geometry (as codified by Euclid in the third century BC) hold true. In Newton's view, an object with no net force acting on it moves through this Euclidean space with a constant velocity. However, when we look at real celestial objects (comets, planets, asteroids, and so forth) we find that their velocity is not constant; they move on curved lines with continuously changing speeds. Why is this? Newton would say, "Their velocities are changing because there is a force acting on them; the force called *gravity*."

Newton derived a useful formula for computing the gravitational force between two objects. Every object in the universe, said Newton, has a property that we may call the "gravitational mass". Let the gravitational mass of two spherical objects be m_g and M_g, and let the distance between their centers be r.. The gravitational force acting between the objects is

$$F_{\rm grav} = -\frac{G M_g m_g}{r^2} \ , \tag{10.35}$$

where G is the Newtonian gravitational constant. The gravitational mass of an object is a non-negative number, so the Newtonian gravitational force is always attractive, with $F_{\rm grav} \leq 0$. Newton also provided us with a useful formula that tells us how objects move in response to a force. Every object in the universe, said Newton, has a property that we may call the "inertial force". If the inertial mass of an object is m_i, then if a net force F is applied to it, Newton's second law of motion tells us that its acceleration will be

$$a = F/m_i \ . \tag{10.36}$$

In equations (10.35) and (10.36), we have used different subscripts to distinguish between the gravitational mass m_g and the inertial mass m_i. One of the fundamental principles of physics (a rather remarkable one, if you stop to think about it) is that the gravitational mass and the inertial mass of an object are identical:

$$m_g = m_i \ . \tag{10.37}$$

The equality of gravitational and inertial mass is known as the **equivalence principle**. The gravitational acceleration a of an object under the influence of a sphere of mass M_g will generally be

$$a = \frac{F_{\rm grav}}{m_i} = -\frac{GM_g}{r^2}\left(\frac{m_g}{m_i}\right) \ . \tag{10.38}$$

If the equivalence principle didn't hold true, then different objects would fall at different rates in the Earth's gravitational field. The observation that $a = -9.8\,\mathrm{m\,s^{-2}}$ for all objects near the Earth's surface is supporting evidence that the equivalence principle holds true.

It is the equivalence principle that led Einstein to devise his theory of general relativity. To see why, let's do a thought experiment. Suppose you wake up one morning to find that you've been sealed inside a small, opaque, soundproof box. You are so startled by this, you drop your teddy bear. Observing the falling bear, you find that it falls toward the floor with an acceleration $a = -9.8\,\mathrm{m\,s^{-2}}$. "Whew!" you say with relief. "At least I am still on the Earth's surface, and not being abducted by space aliens." At that moment, a window in the side of the box opens to reveal you are in an alien spacecraft that is being accelerated at $a = 9.8\,\mathrm{m\,s^{-2}}$ by a rocket engine. When you drop a teddy bear, or any other object, in a small sealed box, the equivalence principle allows two possible interpretations, illustrated in Figure 10.6. (1) The bear is moving at a constant velocity, and the box is being accelerated upward by a constant non-gravitational force. (2) The box is moving at a constant velocity (which may be zero), and the bear is being accelerated downward by a constant gravitational force. The observed behavior of the bear is the same in each case.

Now suppose you are still in the sealed box, being accelerated through space by a rocket at $a = 9.8\,\mathrm{m\,s^{-2}}$. You grab the flashlight you keep on the bedside table, and shine a beam of light perpendicular to the acceleration vector (Figure 10.7). Since the box is accelerating upward, the path of the light will appear to you to be bent downward toward the floor, as the floor of the box accelerates upward to meet the photons. However, thanks to the equivalence principle, we can replace the accelerated box with a stationary

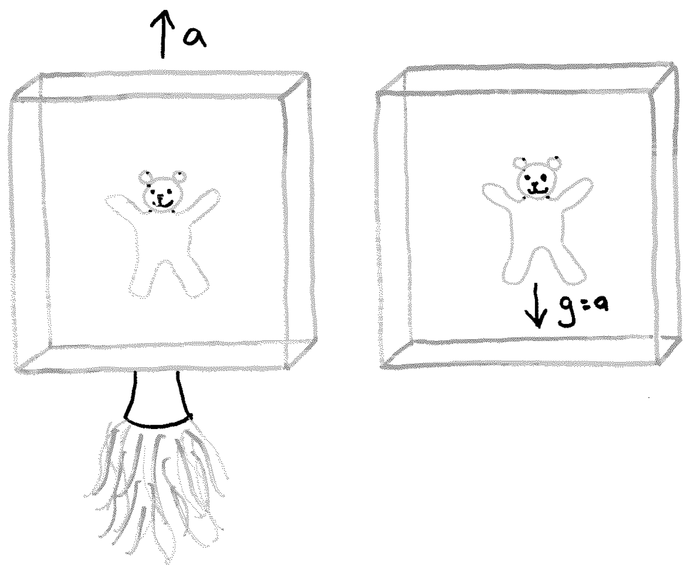

Figure 10.6: Equivalence principle (teddy bear version).

box experiencing a constant gravitational acceleration. Since there's no way to distinguish between these two cases, we are led to the conclusion that the paths of photons will be *curved* in the presence of a gravitational field. Gravity affects photons, Einstein concluded, even though they have no mass.

Contemplating the curved path of the photons, Einstein had another insight. A fundamental principle of optics is Fermat's Principle, which states that light travels between two points along a path that minimizes the travel time.[8] In a vacuum, where the speed of light is constant, this translates into the requirement that light takes the shortest path between two points. In Euclidean space, the shortest distance between two points is a straight line. In the presence of gravity, however, the path taken by light in a vacuum is a *curved* line. This led Einstein to conclude that space is non-Euclidean.

The presence of mass, in Einstein's view, causes space to be curved. More broadly, in the theory of general relativity, mass and energy (which Newton thought of as very different things) are interchangeable, via the equation $E = mc^2$. Moreover, space and time (which Newton thought of as very

[8]More precisely, Fermat's Principle requires that the travel time be an extremum. Under most circumstances, the path minimizes travel time rather than maximizes it.

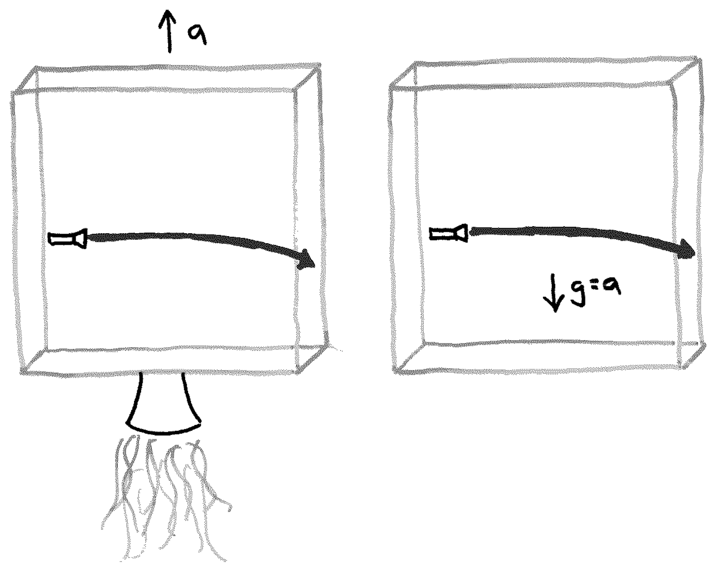

Figure 10.7: Equivalence principle (flashlight version).

different things) form a four-dimensional **space-time**. A more complete summary of Einstein's viewpoint, then, is that the presence of mass-energy causes space-time to be curved. This gives us a third way of thinking about the motion of the teddy bear in the box: (3) No forces are acting on the bear; it is simply following a **geodesic** in curved space-time.[9]

In general, computing the curvature of space-time is a complicated problem. Since the distribution of mass and energy is very inhomogeneous on small scales, the curvature of space and time is also very inhomogeneous, with strong curvature near black holes and neutron stars, and weak curvature in intergalactic voids. However, on scales bigger than 100 Mpc¡ the spatial distribution of mass and energy appears homogeneous and isotropic. Thus, we conclude that the curvature of space is also homogeneous and isotropic on large scales. The assumption of homogeneity and isotropy vastly simplifies the problem. There are only three basic geometries that space can have under such restrictive conditions. Since picturing the curvature of three-dimensional space is difficult, we'll start by considering the curvature of two-dimensional spaces, whose pictures can be neatly drawn on paper; later, we'll

[9]The word "geodesic", in this context, is shorthand for "the shortest distance between two points.

generalize to three dimensions.

First of all, space could be **flat**, or Euclidean. A picture of a flat two-dimensional space, otherwise known as a plane, is given in Figure 10.8. In

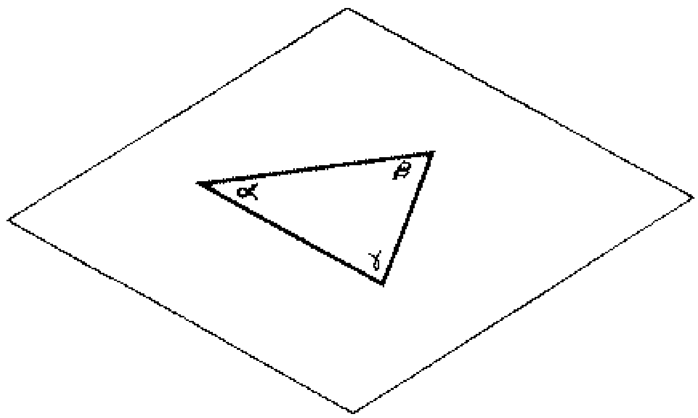

Figure 10.8: A flat two-dimensional space (plane).

flat space, all of Euclidean geometry holds true. For instance, in flat space, a geodesic is a straight line. If a triangle is constructed in flat space by connecting three points with geodesics, the angles at the vertices (α, β, and γ in Figure 10.8) must obey the relation

$$\alpha + \beta + \gamma = \pi \ , \tag{10.39}$$

when the angles are measured in radians. A plane has an infinite area,[10] and has no edge or boundary.

Another two-dimensional space with homogeneous, isotropic curvature is the surface of a sphere, as illustrated in Figure 10.9. On a sphere, a geodesic is a portion of a great circle.[11] If a triangle is constructed on the surface of a sphere by connecting three points with geodesics, the angles at its vertices (α, β, and γ in Figure 10.9) must obey the relation

$$\alpha + \beta + \gamma = \pi + A/r_c^2 \ , \tag{10.40}$$

where A is the area of the triangle and r_c is the radius of the sphere. Spaces in which $\alpha + \beta + \gamma > \pi$ are called **positively curved** spaces. A sphere has a finite area, $4\pi r_c^2$, but no edge or boundary.

[10]Figure 10.8, of course, only shows a portion of a plane.

[11]If the Earth is approximated as a sphere, a line of constant longitude falls along a great circle. The equator is a great circle, but other lines of constant latitude are not.

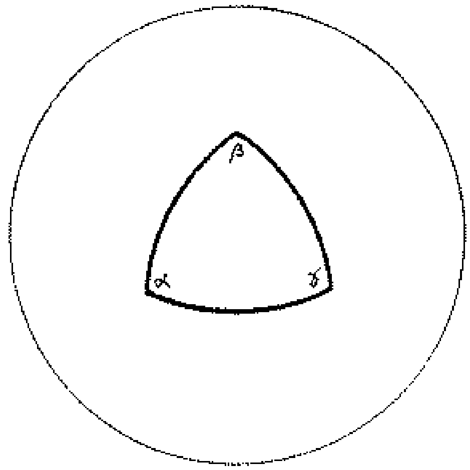

Figure 10.9: A positively curved two-dimensional space (sphere).

In addition to flat spaces and positively curved spaces, there exist **negatively curved** spaces. An example of a negatively curved two-dimensional space is the hyperboloid, or "saddle shape", shown in Figure 10.10. Consider a two-dimensional space of constant negative curvature, with radius of curvature r_c. If a triangle is constructed on this surface by connecting three points with geodesics, the angles at its vertices (α, β, and γ in Figure 10.10) must obey the relation

$$\alpha + \beta + \gamma = \pi - A/r_c^2 \ , \tag{10.41}$$

where A is the area of the triangle. A surface of constant negative curvature has infinite area, just as a plane does.

If you want a two-dimensional surface to have homogeneous, isotropic curvature, only three cases fit the bill: it can be uniformly flat, it can have uniform positive curvature, or it can have uniform negative curvature. The same holds true for three-dimensional spaces. Thus, the curvature of homogeneous, isotropic space can be specified by just two numbers, κ and r_c. The number κ, called the **curvature constant**, is $\kappa = 0$ for flat space, $\kappa = +1$ for positively curved space, and $\kappa = -1$ for negatively curved space. If κ is not zero, then r_c, which has dimensions of length, is the **radius of curvature** of the space. Generally, $r_c(t)$ is a function of time, with $r_c(t) = a(t)r_{c,0}$ if the space is to remain homogeneous and isotropic.

So what is the curvature of the universe – positive, negative, or flat? As

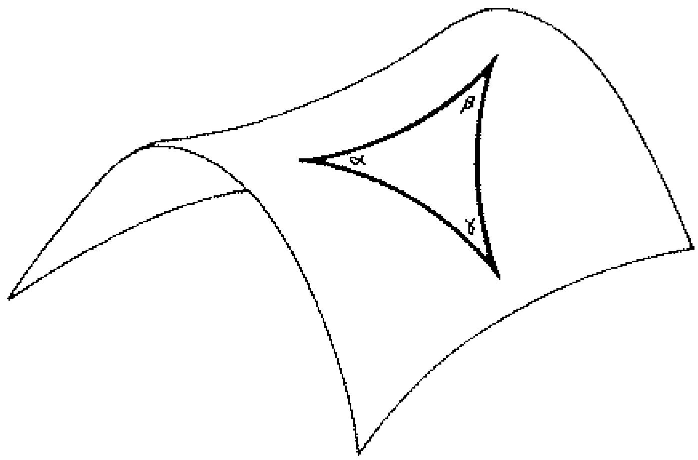

Figure 10.10: A negatively curved two-dimensional space (hyperboloid).

early as the year 1829, long before Einstein's parents were twinkles in his grandparents' eyes, the mathematician Nikolai Ivanovich Lobachevski, one of the founders of non-Euclidean geometry, proposed observational tests to determine the curvature of the universe. In principle, measuring the curvature is simple. Just draw a triangle, then measure its area A and the angles α, β, and γ at its vertices. From equations (10.39), (10.40), and (10.41), we know that

$$\alpha + \beta + \gamma = \pi + \frac{\kappa A}{r_{c,0}^2} \ , \tag{10.42}$$

where κ is the curvature constant and $r_{c,0}$ is the present radius of curvature. Thus, we can compute

$$\frac{\kappa}{r_{c,0}^2} = \frac{\alpha + \beta + \gamma - \pi}{A} \ . \tag{10.43}$$

Unfortunately for this elegant plan, the deviation of $\alpha + \beta\gamma$ from π radians is tiny unless the area of the triangle is comparable to $r_{c,0}^2$. Really, really big triangles are required.

We can conclude that if the universe is curved, with $\kappa = \pm 1$, the radius of curvature cannot be much smaller than the Hubble distance, $c/H_0 \approx 4300\,\text{Mpc}$. To see why this is true, consider looking at a galaxy of diameter D that is at a distance d from the Earth (Figure 10.11). In a flat universe, in the limit $D \ll d$, we can use the small angle formula to compute the angular

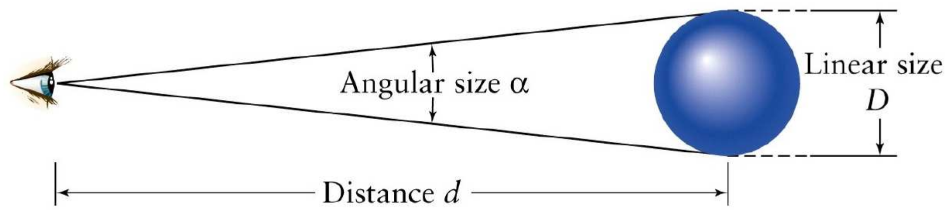

Figure 10.11: Angular size of a distant galaxy.

size α of the galaxy:

$$\alpha_{\text{flat}} = \frac{D}{d} , \tag{10.44}$$

where the angle α is in radians. However, in positively or negatively curved space, the angular size is no longer proportional to $1/d$.

In a space with uniform positive curvature, the angular size is

$$\alpha_{\text{pos}} = \frac{D}{r_{c,0}\sin(d/r_{c,0})} > \frac{D}{d} . \tag{10.45}$$

In a positively curved universe, the mass-energy content acts as a magnifying gravitational lens, making galaxies appear larger than they would in flat space. There are two interesting consequences of equation (10.45). First, the angular size blows up when $d = \pi r_{c,0}$; that is, when a galaxy is at a distance equal to half the circumference of the universe, it fills the entire sky.[12] No such enormous, sky-filling galaxies are seen. Second, since the universe has a finite circumference $C_0 = 2\pi r_{c,0}$, an object seen at a distance d will also be seen, with the same angular size, at a distance $d + C_0$, and at a distance $d + 2C_0$, and at a distance $d + 3C_0$, and so forth, ad nauseum. No such periodic galaxy images are seen. If the universe is positively curved, its radius of curvature must therefore be comparable to or greater than the Hubble distance.

In a space with uniform negative curvature, the angular size of a galaxy

[12] As a two-dimensional analogy, suppose that you were at the north pole of the Earth and a light source were at the south pole. If the light were constrained to follow great circles, it would flow along all the lines of longitude stretching away from the south pole, and converge on your position at the north pole. No matter which way you turned, you would see the south pole beacon.

is

$$\alpha_{\rm neg} = \frac{D}{r_{c,0}\sinh(d/r_{c,0})} < \frac{D}{d} \ . \tag{10.46}$$

A negatively curved universe thus acts as a de-magnifying lens.[13] If a galaxy is at a distance $d \gg r_{c,0}$, we can use the approximation $\sinh x \approx e^x/2$ when $x \gg 1$. With this approximation,

$$\alpha_{\rm neg} \approx \frac{2D}{r_{c,0}} e^{-d/r_{c,0}} \ . \tag{10.47}$$

In a negatively curved universe, objects at a distance much greater than the radius of curvature will appear exponentially tiny. Since galaxies are resolved in angular size, with $\alpha > 1\,$arcsec, out to distances comparable to the Hubble distance, we conclude that if the universe is negatively curved, its radius of curvature must be comparable to or greater than the Hubble distance.

The conclusion of cosmologists, using geometrical arguments like the ones given above, is that the universe is consistent with being *flat* ($\kappa = 0$). Although we cannot rule out the possibility of slight positive or negative curvature, the radius of curvature in that case would be bigger than the Hubble distance, and would have negligible effects on the small bit of the universe within a Hubble distance of us.[14] To make life simpler, we will assume, in many of the following equations, that the universe is perfectly flat.

10.4 Metrics of Space-Time

Astronomers study events that are widely spread out in space, and also widely spread out in time. Thus, it is useful for them to be able to compute the distance between two events in a four-dimensional space-time. Computing the distance between two points in a flat three-dimensional space is easy. If one point is at (x, y, z) and the other is at $(x+dx, y+dy, z+dz)$, the distance $d\ell$ between them is given by the formula

$$d\ell^2 = dx^2 + dy^2 + dz^2 \ . \tag{10.48}$$

[13] Or a de-magnifying rear view mirror: "Objects in mirror are closer than they appear."

[14] Similarly, a small bit of the Earth's curved surface is reasonably well described by a flat map. A flat map of the entire Earth results in distortions of size or shape (think of Greenland or Antarctica in a Mercator projection), but a flat map of Ohio doesn't have perceptible distortions.

A formula such as equation (10.48) that gives the distance between two points is known as a **metric**. Equation (10.48) uses the convention, common among relativists, that $d\ell^2 = (d\ell)^2$, not $d(\ell^2)$. Omitting the parentheses reduces the visual clutter. The metric of flat space appears different when different coordinate systems are used. For instance, in spherical coordinates, the metric of flat space is

$$d\ell^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)\ . \tag{10.49}$$

By extension, we can compute the four-dimensional space-time distance between two events, one at (t, x, y, z) and the other at $(t+dt, x+dx, y+dy, z+dz)$. According to special relativity, the space-time distance between these events is

$$\begin{aligned} d\ell^2 &= -c^2dt^2 + dx^2 + dy^2 + dz^2 \qquad (10.50) \\ &= -c^2dt^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)\ . \end{aligned}$$

The metric given in equation (10.51) is called the **Minkowski metric**, and the space-time in which it holds true is called Minkowski space-time. Note that the sign of the term involving time ($-c^2dt^2$) is opposite to that of the terms involving the spatial coordinates.[15] The Minkowski metric applies only in the context of *special* relativity, which deals with the special case in which space-time is not distorted by the presence of mass or energy. Thus, the Minkowski metric represents a static, empty, spatially flat universe.

In an expanding (or contracting) universe, the metric we use to measure space-time distances is called the **Robertson-Walker metric**. If space is flat, then the Robertson-Walker metric takes the form

$$d\ell^2 = -c^2dt^2 + a(t)^2[dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)]\ . \tag{10.51}$$

Notice how the spatial component of the Robertson-Walker metric is scaled by the square of the scale factor $a(t)$. The time variable t in the Robertson-Walker metric is the **cosmic time**, which is the time measured by an observer who sees the universe expanding uniformly around him or her. The spatial variables (r, θ, ϕ) in the Robertson-Walker metric are the **comoving coordinates** of a point in space. If the expansion of the universe is perfectly

[15] Zeilik and Gregory use the opposite sign convention: $d\ell^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$. This is purely a formal convention, and has no physical meaning. It's like the arbitrary pronouncement that electrons have negative charge and protons have positive; physics would be unchanged if we assigned + to electrons and − to protons.

homogeneous and isotropic, then the comoving coordinates of any point remain constant with time.[16]

Suppose you are observing a distant galaxy, and want to know how far away it is. Since we are in an expanding universe, when we assign a distance ℓ between two objects (such as an astronomer and a galaxy), we must specify the time t at which that distance is correct. For convenience, let's set up a coordinate system in which you are at the origin and the galaxy is at a comoving coordinate position (r, θ, ϕ), as shown in Figure 10.12. The **proper**

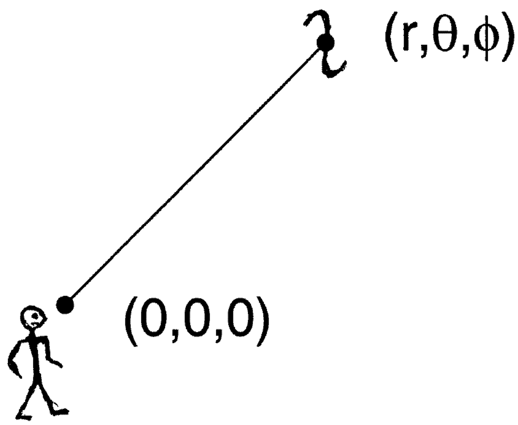

Figure 10.12: An observer looks at a galaxy.

distance $\ell_p(t)$ between two points is the length of the geodesic between them when the cosmic time is fixed at the value t, and the scale factor is thus fixed at the value $a(t)$. The proper distance between and observer and galaxy in a flat universe can be found by using the Robertson-Walker metric of equation (10.51) at fixed time t:

$$d\ell^2 = a(t)^2[dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)] \ . \tag{10.52}$$

Along the geodesic between the galaxy and observer, the angle (θ, ϕ) is constant, and thus

$$d\ell = a(t)dr \ . \tag{10.53}$$

The proper distance $\ell_p(t)$ is found by integrating over the radial comoving coordinate r:

$$\ell_p(t) = a(t)\int_0^r dr = a(t)r \ . \tag{10.54}$$

[16]Similarly, if the Earth were uniformly expanding or contracting with time, the latitude and longitude of any point would remain constant with time.

The normalization $a(t_0) = 1$ for the scale factor means that the comoving coordinate r is simply the current proper distance to the galaxy: $r = \ell_p(t_0)$.

Unfortunately, the proper distance $\ell_p(t_0)$ to a distant galaxy is impossible to measure, since we don't have gigaparsec long tape measures that can be extended infinitely rapidly. As astronomers, we are condemned to a passive role; we learn what we can about the galaxy in Figure 10.12 by gathering up the photons that it emits. A photon that we collect at time t_0 was emitted at an earlier time $t_e < t_0$. Photons travel on geodesics through space-time; more precisely, they travel on **null geodesics**. A null geodesic is a geodesic for which $d\ell = 0$ along every infinitesimal section of its path. Given equation (10.51), a photon must satisfy the relation

$$c^2 dt^2 = a(t)^2 [dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)] \tag{10.55}$$

as it travels through an expanding, spatially flat universe. A photon traveling from the galaxy at (r, θ, ϕ) to an observer at the origin follows a beeline with θ and ϕ constant. This implies

$$c^2 dt^2 = a(t)^2 dr^2 \tag{10.56}$$

along every infinitesimal segment of the photon's radial path. Rearranging equation (10.56), we find

$$c\frac{dt}{a(t)} = dr \ , \tag{10.57}$$

in which the left hand side depends only on t and the right hand side depends only on r. Integrating along the photon's path,

$$c\int_{t_e}^{t_0} \frac{dt}{a(t)} = \int_0^r dr = r \ . \tag{10.58}$$

Since the comoving distance r is equal to the current proper distance $\ell_p(t_0)$, this implies that the proper distance is related to the scale factor by the relation

$$\ell_p(t_0) = c\int_{t_e}^{t_0} \frac{dt}{a(t)} \ . \tag{10.59}$$

In a static universe, where $a(t) = 1$ for all time, equation (10.59) states that the proper distance to a galaxy is equal to the speed of light times the photon's travel time: $\ell_p = c(t_0 - t_e)$. If the universe has been steadily expanding since t_e and t_0, then $a(t)$ was smaller than the past than it is now, and thus $\ell_p(t_0) > c(t_0 - t_e)$. In general, although the current proper distance $\ell_p(t_0)$ isn't something we can measure, it's something we can compute if we know $a(t)$.

Although we can't directly measure the current proper distance of a galaxy, there is a consolation prize; we can measure the galaxy's *redshift*. The redshift z tells us something useful: the scale factor $a(t_e)$ at the time the observed light was emitted. When we considered the cooling of the Cosmic Microwave Background, we learned that the wavelength of light expands along with the expansion of the universe: $\lambda(t) \propto a(t)$. This applies to all photons, not just CMB photons. If we observe a galaxy's emission line with wavelength λ_0 at time t_0, it was emitted with a shorter wavelength λ_e at an earlier time t_e. The relation between observed wavelength λ_0 and emitted wavelength λ_e is

$$\frac{\lambda_e}{a(t_e)} = \frac{\lambda_0}{a(t_0)} \ . \tag{10.60}$$

Using the definition of redshift,

$$z = \frac{\lambda_0 - \lambda_e}{\lambda_e} \ , \tag{10.61}$$

we find that the redshift is simply related to the scale factor at the time of emission:

$$1 + z = \frac{\lambda_e}{\lambda_0} = \frac{a(t_0)}{a(t_e)} = \frac{1}{a(t_e)} \ . \tag{10.62}$$

If we observe a quasar with $z = 6.4$, we are observing it as it was when the universe had a scale factor $a(t_e) = 1/7.4 = 0.135$.

The most distant objects we can see, in theory, are those for which the light emitted at time $t = 0$ is just now reaching us at $t = t_0$. The proper distance to such an object is called the **horizon distance**. In the limit $t_e \to 0$, equation (10.59) tells us that the current horizon distance is

$$\ell_{\rm hor}(t_0) = c \int_0^{t_0} \frac{dt}{a(t)} \ . \tag{10.63}$$

As an example, let's suppose that the scale factor is a power law, with $a(t) = (t/t_0)^n$. If $n < 1$, the horizon distance is finite, with

$$\ell_{\rm hor}(t_0) = c \int_0^{t_0} \frac{dt}{(t/t_0)^n} = \frac{ct_0}{1-n} \ . \tag{10.64}$$

Since the Hubble constant is

$$H_0 = \left(\frac{\dot{a}}{a}\right)_{t=t_0} = \frac{n}{t_0} , \tag{10.65}$$

the horizon distance can also be written in the form

$$\ell_{\rm hor}(t_0) = \frac{n}{1-n}\frac{c}{H_0} , \tag{10.66}$$

when $0 < n < 1$. Thus, if we want to know the exact relation between the Hubble distance c/H_0 and the horizon distance, we need to know the functional form of $a(t)$.

10.5 The Friedmann Equation

In the context of general relativity, the form of $a(t)$, as well as the curvature constant κ and radius of curvature $r_{c,0}$, are dictated by Einstein's field equations. In general relativity, the field equations link the curvature of space-time at any point to the energy density and pressure at that point.[17] The equation that links $a(t)$, κ, and $r_{c,0}$ is the **Friedmann equation**, named after Alexander Friedmann, who first derived it in the year 1922. We have already seen the Newtonian version of the Friedmann equation; it's the energy conservation equation for the expanding sphere (Equation (10.32).

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(t) + \frac{2k}{r_0^2}\frac{1}{a(t)^2} . \tag{10.67}$$

The relativistically correct form of the Friedmann equation is

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3c^2}u(t) - \frac{\kappa c^2}{r_{c,0}^2}\frac{1}{a(t)^2} + \frac{\Lambda}{3} . \tag{10.68}$$

Equation (10.68) is offered without proof. (A derivation should only be done by a highly trained relativist; please do not try it at home!)

Consider the changes made in going from the Newtonian form of the Friedmann equation to the relativistically correct form. First, the mass density ρ has been replaced by an energy density u. Relativistic particles, such

[17] The field equations are the relativistic equivalent of Poisson's equation, which links the gravitational potential at any point to the mass density ρ at that point.

as photons, have an energy $\epsilon = hc/\lambda$ that contributes to the energy density. Not only do photons respond to the curvature of space-time, they also contribute to it.

Second, in going from the Newtonian to the relativistic form, we make the substitution

$$\frac{2k}{r_0^2} \to -\frac{\kappa c^2}{r_{c,0}^2} \ . \tag{10.69}$$

In the Newtonian model, the constant k told us whether the universe was gravitationally bound ($k < 0$) or unbound ($k > 0$). In the relativistic model, the constant κ tells us whether the universe is positively curved ($\kappa > 0$) or negatively curved ($\kappa < 0$).[18]

Third and last, in going from the Newtonian to the relativistic form, we add a new term, $\Lambda/3$ to the right hand side of the equation. The Greek letter Λ is the symbol for the famous (or perhaps infamous) **cosmological constant**. The cosmological constant has a checkered history, going back to the year 1917, when Einstein published his first paper on the cosmological implications of general relativity. In a formal mathematical sense, Λ is a constant of integration resulting from solving Einstein's field equations, which are a set of differential equations. In addition, however, the cosmological component can be given a physical meaning.[19]

A close look at the Friedmann equation (eq. 10.68) shows that adding the Λ term is equivalent to adding a new component to the universe that has a constant energy density

$$u_\Lambda = \frac{c^2\Lambda}{8\pi G} \ . \tag{10.70}$$

Thus, any component of the universe whose energy density is constant with time will play the part of a cosmological constant. One such component is the **vacuum energy**. In quantum physics, a vacuum is not a sterile void. The Heisenberg uncertainty principle allows particle/antiparticle pairs to spontaneously appear and then annihilate in an otherwise empty space. Just as there is an energy density u associated with real particles, there's an energy

[18]Throw away your t-shirt that says "Density is destiny", and buy one that says "Density is curvature".

[19]Pedantic footnote: I am using the convention that Λ has units of $1/[\text{time}]^2$. Other authors, such as Zeilik and Gregory, use a value of Λ that differs by a factor $1/c^2$, and thus has units of $1/[\text{length}]^2$. Just a warning, in case you want to go browsing through the cosmological literature.

density $u_{\rm vac}$ associated with the virtual particles and antiparticles. The vacuum energy density $u_{\rm vac}$ is a small scale quantum effect that is unaffected by the large scale expansion of the universe; hence $u_{\rm vac}$ remains constant as the universe expands. (Unfortunately, quantum field theory cannot tell us the expected numerical value of $u_{\rm vac}$.)

Let's rewrite the Friedmann equation in terms of the energy density of the universe, including the energy density u_Λ associated with the cosmological constant:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3c^2}[u_r(t) + u_m(t) + u_\Lambda] - \frac{\kappa c^2}{r_{c,0}^2}\frac{1}{a(t)^2} \ . \tag{10.71}$$

We've subdivided the energy density into three categories. First, the **radiation density** u_r is the energy density contributed by relativistic particles, such as photons. Second, the **matter density** u_m is the energy density contributed by non-relativistic particles such as protons, neutrons, electrons, and WIMPs. For non-relativistic particles, $u_m = \rho_m c^2$. Finally, the **lambda density**, alias the vacuum density, is the constant energy density provided by the cosmological constant Λ.

The fact that our universe is flat (or very close to it) means that the total energy density is equal to the **critical density** (or very close to it). For perfect flatness ($\kappa = 0$),

$$u_r + u_m + u_\Lambda = u_c \ , \tag{10.72}$$

where the critical energy density is

$$u_c = \rho_c c^2 = \frac{3H(t)^2 c^2}{8\pi G} \ . \tag{10.73}$$

Since the Hubble parameter is currently $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$, this translates into a current critical density

$$u_{c,0} = \frac{3H_0^2 c^2}{8\pi G} = 8.3\times 10^{-10}\,{\rm J\,m^{-3}} = 5200\,{\rm MeV\,m^{-3}} \ . \tag{10.74}$$

This is one of the more fascinating results of general relativity. Because the universe is flat on large scales, we know the average energy density of the universe! Even if we don't know how much is contributed by each component, we know that the total must come to $5200\,{\rm MeV\,m^{-3}}$.[20]

[20] That's the calorie content of a standard candy bar spread over a million cubic meters.

Since the critical density $u_c(t)$ is vital to an understanding of the curvature and expansion of the universe, cosmologists frequently express the energy density of the universe in terms of the dimensionless **density parameter**

$$\Omega(t) \equiv \frac{u(t)}{u_c(t)} . \tag{10.75}$$

If $\Omega < 1$, the universe is negatively curved; if $\Omega > 1$, the universe is positively curved. Saying "The universe is flat" is equivalent to saying "Omega equals one." Density is curvature. By extension, we can write down a density parameter for each component of the universe:

$$\Omega_r(t) \equiv \frac{u_r(t)}{u_c(t)} , \ \Omega_m(t) \equiv \frac{u_m(t)}{u_c(t)} , \ \Omega_\Lambda(t) \equiv \frac{u_\Lambda}{u_c(t)} . \tag{10.76}$$

Knowing how the universe expands with time requires knowing how much energy density is in radiation, matter, and the cosmological constant today, and knowing how the energy density of radiation and matter evolves with time. (There are various exotic cosmologies that contain other components, like cosmic strings and domain walls and various funky types of dark energy, but for simplicity, we'll stick to a universe with just radiation, non-relativistic matter, and a cosmological constant.)