Chapter 27

Cosmology

Version 0027.3, May 2001

27.1 Overview

General Relativity is an indispensable foundation for understanding the large scale structure and evolution of the universe (*cosmology*), but it is only one foundation out of many. The crudest of understandings can be achieved with general relativity and little else; but more detailed and deeper understandings require combining general relativity with quantum field theory, nuclear and atomic physics, thermodynamics, fluid mechanics, and large bodies of astrophysical lore.

In this chapter we shall explore aspects of cosmology which are sufficiently crude that general relativity, augmented by only bits and pieces of other physics, can provide an adequate foundation. Our exploration will simultaneously illustrate key aspects of general relativity and give the reader an overview of modern cosmology.

We shall begin in Sec. 27.2 by discussing the observational data that suggest our universe is homogeneous and isotropic when averaged over regions of space huge compared to clusters of galaxies, and we then shall construct a spacetime metric for an idealized homogeneous, isotropic model for the universe. In Sec. 27.3 we shall construct a stress-energy tensor that describes, approximately, the total, averaged energy and pressure of the universe's matter and radiation; and we shall insert that stress-energy tensor and the metric of Sec. 27.2 into the Einstein field equation, thereby deducing a set of equations that govern the evolution of the universe. In Sec. 27.4 we shall study the predictions that those evolution equations make for the rate of expansion of the universe and the manner in which the expansion changes with time, and we shall describe the most important physical processes that have occured in the universe during its evolution into its present state. As we shall see, the details of the expansion are determined by the values of seven parameters that can be measured today—with the caveate that there may be some big surprises associated with an "exotic" form of matter that is ill-understood. In Sec. 27.5 we shall describe the astronomical observations by which the universe's seven parameters are being measured, and the growing evidence for exotic matter (also called "dark energy"). In Sec. 27.6 we shall discuss the big-bang singularity in which the universe probably began, and shall discuss the fact that this singularity, like singularities inside black holes, is a signal that general relativity breaks down and must be replaced by a quantum theory of gravity which (hopefully) will not predict singular behavior. We shall also examine a few features that the quantum theory of gravity is likely to exhibit. Finally, in Sec. 27.7 we shall discuss the "inflationary" epoch that the universe may have undergone immediately after the quantum gravity, big-bang epoch.

27.2 Homogeneity and Isotropy of the Universe; Robertson-Walker Line Element

The universe obviously is not homogeneous or isotropic in our neighborhood: In our solar system (size $\sim 10^{14}$ cm) almost all the mass is concentrated in the sun and planets, with a great void in between. Looking beyond the solar system, one sees the Milky Way Galaxy (size $\sim 10^{23}$ cm $\sim 10^5$ light years, or equivalently 3×10^4 parsecs),[1] with its mass concentrated toward the center and its density falling off roughly as $1/(\text{distance})^2$ as one moves out past the sun and into the Galaxy's outer reaches. Beyond the Galaxy is the emptiness of intergalactic space; then other galaxies congregated into our own "local group" (size $\sim 10^6$ parsecs). The local group is in the outer reaches of a cluster of several thousand galaxies called the Virgo cluster (size $\sim 10^7$ parsecs), beyond which is the void of intercluster space, and then other clusters at distances $\gtrsim 10^8$ parsecs.

Despite all this structure, the universe appears to be nearly homogeneous and isotropic on scales $\gtrsim 10^8$ parsecs, i.e., $\gtrsim 3 \times 10^8$ light years: On such scales one can regard galaxies and clusters of galaxies as "atoms" of a homogeneous, isotropic "gas" that fills the universe.

On scales far larger than clusters of galaxies, our best information about homogeneity and isotropy comes from the cosmic microwave background radiation ("CMB"). As we shall see in Secs. 27.4 and 27.5 below, this radiation, emitted by hot, primordial gas long before galaxies formed, comes to us from distances of order 3×10^9 parsecs (1×10^{10} light years)—a scale 100 times larger than a rich cluster of galaxies (i.e., than a "supercluster"), and the largest scale on which observations can be made. Remarkably, this microwave radiation has a black body spectrum with a temperature that is the same, in all directions on the sky, to within about three parts in 10^5. This means that the temperature of the primordial gas was homogeneous on large scales to within this impressive accuracy.

These observational data justify a procedure in modeling the universe which was adopted by Einstein (1917) and others, in the early days of relativistic cosmology, with little more than philosophical justification: Like Einstein, we shall assume, as a zero-order approximation, that the universe is precisely homogeneous and isotropic. Later we shall briefly discuss galaxies and clusters of galaxies as first-order corrections to the homogeneous and isotropic structure.

Our assumption of homogeneity and isotropy can be stated more precisely as follows: *There exists a family of slices of simultaneity (3-dimensional spacelike hypersurfaces), which completely covers spacetime, with the special property that on a given slice of simultaneity*

[1] One parsec is 3.262 light years, i.e. 3.086×10^{18} cm. It is defined as the distance of a star whose apparent motion on the sky, induced by the Earth's orbital motion, is a circle with radius one arc second.

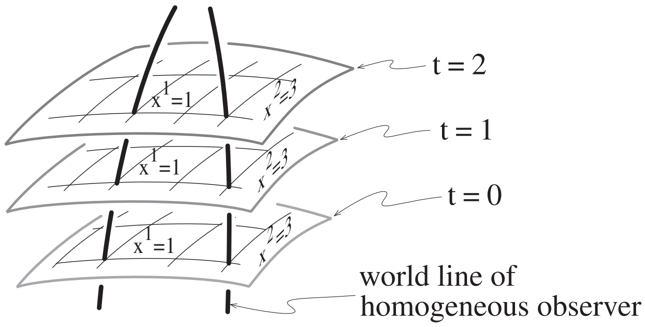

Fig. 27.1: The synchronous coordinate system for a homogeneous, isotropic model of the universe.

(i) *no two points are distinguishable from each other ("homogeneity"), and* (ii) *at a given point no one spatial direction is distinguishable from any other ("isotropy").*

Whenever, as here, the physical (geometrical) structure of a system has special symmetries, it is useful to introduce a coordinate system which shares and exhibits those symmetries. In the case of a spherical black hole, we introduced spherical coordinate systems. Here we shall introduce a coordinate system in which the homogeneous and isotropic hypersurfaces are slices of constant coordinate time t [Fig. (27.1)]:

Recall (Sec. 25.6) the special role of observers whose world lines are orthogonal to the homogeneous and isotropic hypersurfaces: if they define simultaneity locally (on small scales) by the Einstein light-ray synchronization process, they will regard the homogeneous hypersurfaces as their own slices of simultaneity. Correspondingly, we shall call them the "homogeneous observers."

We shall define our time coordinate t to be equal to the proper time τ as measured by these homogeneous observers, with the arbitrary additive constant in t so adjusted that one of the homogeneous hypersurfaces (the "initial" hypersurface) has $t = 0$ everywhere on it. Stated differently, but equivalently, we select arbitrarily the initial hypersurface and set $t = 0$ throughout it; and we then define t along the world line of a homogeneous observer to be the proper time that the observer's clock has ticked since the observer passed through the initial hypersurface.

This definition of t has an important consequence: Since the points at which each of the observers pass through the initial hypersurface are all equivalent (all indistinguishable; "homogeneity"), the observers' subsequent explorations of the homogeneous universe must be indistinguishable; and, correspondingly, they must all reach any specific homogeneous hypersurface at the same proper time τ, and thence at the same coordinate time $t = \tau$. Thus, the hypersurfaces of constant coordinate time t are the same as the homogeneous hypersurfaces.

Turn, next, to the three spatial coordinates x^j. We shall define them in an arbitrary manner on the initial hypersurface, but shall insist that the homogeneous observers carry them forward (and backward) in time along their world lines, so that each homogeneous observer's world line is a curve of constant x^1, x^2, and x^3; cf. Fig. 27.1.

In this $\{t, x^j\}$ coordinate system the spacetime metric, described as a line element, will take the generic form

$$ds^2 = g_{tt}dt^2 + 2g_{tj}dt\,dx^j + g_{jk}dx^j dx^k \ . \tag{27.1}$$

Since x^j are constant along a homogeneous observer's world line, the basis vector $(\partial/\partial t)_{x^j}$ is tangent to the world line; and since t is constant in a homogeneous hypersurface, the basis vector $(\partial/\partial x^j)_t$ lies in the hypersurface. These facts, plus the orthogonality of the homogeneous observer's world line to the homogeneous hypersurface, imply that

$$g_{tj} \equiv \mathbf{g}\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial x^j}\right) = \frac{\partial}{\partial t} \cdot \frac{\partial}{\partial x^j} = 0 \ . \tag{27.2}$$

Moreover, since the proper time along a homogeneous observer's world line (line of constant x^j) is $d\tau = \sqrt{-g_{tt}dt^2}$, and since by construction dt is equal to that proper time, it must be that

$$g_{tt} = -1 \ . \tag{27.3}$$

By combining Eqs. (27.1)–(27.3) we obtain for the line element in our very special coordinate system

$$ds^2 = -dt^2 + g_{jk}dx^j dx^k \ . \tag{27.4}$$

Because our spatial coordinates, thus far, are arbitrary (i.e., they do not yet mold themselves in any special way to the homogeneous hypersurfaces), the spatial metric coefficients g_{jk} must be functions of the spatial coordinates x^i as well as of time t.

[*Side Remark*: Any coordinate system in which the line element takes the form (27.4) is called a *synchronous coordinate system*. This is true whether the hypersurfaces $t =$ const are homogeneous and isotropic or not. The key features of synchronous coordinates are that they mold themselves to the world lines of a special family of observers in such a way that t is proper time along the family's world lines, and the slices of constant t are orthogonal to those world lines (and thus are light-ray-synchronization-defined simultaneities for those observers). Since the introduction of synchronous coordinates involves a specialization of precisely four metric coefficients ($g_{tt} = -1$, $g_{tj} = 0$), by a careful specialization of the four coordinates one can construct a synchronous coordinate system in any and every spacetime. On the other hand, one cannot pick an arbitrary family of observers and use them as the basis of synchronous coordinates: The observers must move freely; i.e., their world lines must be geodesics. This one can see by computing $u^\alpha{}_{;\beta}u^\beta$ for the vector field $\vec{u} \equiv \partial/\partial t$, which represents the 4-velocities of the synchronous coordinate system's special observers; a straightforward calculation [Exercise 26.1] gives $u^\alpha{}_{;\beta}u^\beta = 0$, in accord with geodesic motion. Thus it is that the static observers (observers with constant r, θ, ϕ) outside a black hole cannot be used as the foundation for synchronous coordinates. Those observers must accelerate to prevent themselves from falling into the hole; and correspondingly, the closest thing to a synchronous coordinate system that one can achieve, using for $x^j =$ const the world lines of the static observers, is the Schwarzschild coordinate system, which has $g_{tj} = 0$ (the slices of constant t are simultaneities as measured by the static observers), but $g_{tt} = -(1 - 2M/r) \neq -1$ (the proper time between two adjacent simultaneities depends on the radius at which the static observer resides).]

Returning to cosmology, we shall now specialize our spatial coordinates so they mold themselves nicely to the homogeneity and isotropy of the slices of constant t. One might have hoped this could be done in such a way that the metric coefficients are independent of all three coordinates x^j. Not so. The surface of a sphere is a good example in one lower dimension: it is homogeneous and isotropic, but the most symmetric coordinates one can find for it, spherical polar coordinates, produce a line element ${}^{(2)}ds^2 = a^2 d\theta^2 + a^2 \sin^2\theta d\phi^2$ with a metric coefficient $g_{\phi\phi} = a^2\sin^2\theta$ that depends on θ. The deep, underlying reason is that the vector field $(\partial/\partial\phi)_\theta$ that "generates" rotations about the polar axis (z-axis) does not commute with the vector field that generates rotations about any other axis; and, correspondingly, those two vector fields cannot simultaneously be made the basis vectors of any coordinate system, and the metric coefficients cannot be made independent of two angular coordinates simultaneously. (For further detail see Secs. 25.2 and 25.3 of MTW, and especially Exercise 25.8.)

Similarly, on our cosmological homogeneous hypersurfaces the most symmetric coordinate system possible entails metric coefficients that are independent of only one coordinate, not all three. In order to construct that most-symmetric coordinate system, we choose arbitrarily on the hypersurface $t = \text{const}$ an origin of coordinates. Isotropy about that origin (all directions indistinguishable) is equivalent to spherical symmetry, which motivates our introducing spherical polar coordinates θ, ϕ, and a radial coordinate that we shall call χ. In this coordinate system the line element of the hypersurface will take the form

$$a^2[d\chi^2 + \Sigma^2(d\theta^2 + \sin^2\theta d\phi^2)] \, , \tag{27.5}$$

where a multiplicative constant (*scale factor*) a has been factored out for future convenience (it could equally well have been absorbed into χ and Σ), and where Σ is an unknown function of the radial coordinate χ. Correspondingly, the 4-dimensional line element of spacetime (27.4) will take the form

$$ds^2 = -dt^2 + a^2[d\chi^2 + \Sigma^2(d\theta^2 + \sin^2\theta d\phi^2)] \, , \tag{27.6}$$

where a is now a function of time t (i.e., it varies from hypersurface to hypersurface).

Our next task is to figure out what functions $\Sigma(\chi)$ are compatible with homogeneity and isotropy of the hypersurfaces. There are elegant, group-theoretic ways to figure this out; see, e.g., Ryan and Shepley (1975). A more straightforward but tedious way is to note that, because the 3-dimensional Riemann curvature tensor of the hypersurface must be homogeneous and isotropic, it must be algebraically expressible in terms of (i) constants, and the only tensors that pick out no preferred locations or directions: (ii) the metric tensor g_{jk} and (iii) the Levi-Civita tensor ϵ_{ijk}. Trial and error shows that the only combination of these three quantities which has the correct number of slots and the correct symmetries is

$$R_{ijkl} = K(g_{ik}g_{jl} - g_{il}g_{jk}) \, , \tag{27.7}$$

where K is a constant. By computing, for the 3-dimensional metric (27.5), the components of the 3-dimensional Riemann tensor and comparing with (27.7), one can show that there are three possibilities for the function $\Sigma(\chi)$ in the metric, and three corresponding possibilities for the constant K in the three-dimensional Riemann tensor. These three possibilities are nicely parametrized by a quantity k which takes on the values $+1$, 0, and -1:

$$k = +1 \quad : \qquad \Sigma = \sin\chi \, , \qquad K = \frac{k}{a^2} = +\frac{1}{a^2}$$

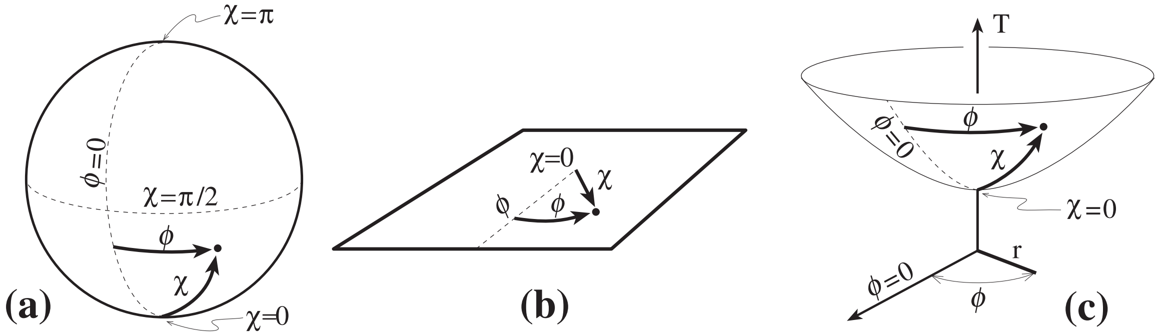

Fig. 27.2: Embedding diagrams for the homogeneous hypersurfaces of (a) a closed, $k = +1$, cosmological model; (b) a flat, $k = 0$, model; and (c) an open, $k = -1$ model.

$$
\begin{aligned}
k = 0 \quad &: \qquad \Sigma = \chi\ , \qquad K = \frac{k}{a^2} = 0\ , \\
k = -1 \quad &: \qquad \Sigma = \sinh\chi\ , \quad K = \frac{k}{a^2} = -\frac{1}{a^2}\ .
\end{aligned}
\tag{27.8}
$$

We shall discuss each of these three possibilities in turn:

Closed universe $[k = +1]$: For $k = +1$ the geometry of the homogeneous hypersurfaces,

$$
{}^{(3)}ds^2 = a^2[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2)]\ , \tag{27.9}
$$

is that of a 3-sphere, i.e., an ordinary sphere generalized to one higher dimension. One can verify this, for example, by showing (Ex. 27.2) that in a 4-dimensional Euclidean space with Cartesian coordinates (w, x, y, z)

(27.10)

the 3-sphere

$$
w^2 + x^2 + y^2 + z^2 = a^2 \tag{27.11}
$$

has the same metric (27.9) as our cosmological, homogeneous hypersurface. Figure 27.2(a) is an embedding diagram for an equatorial slice, $\theta = \pi/2$, through the homogeneous hypersurface [2-geometry ${}^{(2)}ds^2 = a^2(d\chi^2 + \sin^2\chi d\phi^2)$; cf. Eq. (27.9)]. Of course, the embedded surface is a 2-sphere. As radius χ increases, the circumference $2\pi a\sin\chi$ around the spatial origin at first increases, then reaches a maximum at $\chi = \pi/2$, then decreases again to zero at $\chi = \pi$. Clearly, the homogeneous hypersurface is topologically "closed" and has a finite volume. For this reason a $k = +1$ cosmological model is often called a "closed universe". The universe's 3-volume, $V = 2\pi^2 a^3$ (Ex. 27.2).

Flat universe $[k = 0]$: For $k = 0$ the geometry of the homogeneous hypersurfaces,

$$
{}^{(3)}ds^2 = a^2[d\chi^2 + \chi^2(d\theta^2 + \sin^2\theta d\phi^2)]\ , \tag{27.12}
$$

is that of a flat, 3-dimensional Euclidean space—as one can easily see by setting $r = a\chi$ and thereby converting (27.12) into the standard spherical-polar line element for Euclidean space. Correspondingly, this cosmological model is said to represent a "flat universe." Note, however, that this universe is only spatially flat: the Riemann curvature tensor of its 3-dimensional homogeneous hypersurfaces vanishes; but, as we shall discuss below, because of the time evolution of the expansion factor a, the Riemann curvature of the full 4-dimensional spacetime does not vanish. The volumes of the homogeneous hypersurfaces are infinite, so one cannot talk of the universe's total volume changing with time. However, the volume ΔV of a box in which resides a specific set of homogeneous observers will change as the expansion factor a evolves. For example, the volume could be a box with sides $\Delta\chi$, $\Delta\theta$, $\Delta\phi$, so

$$\Delta V = \epsilon_{\chi\theta\phi}\Delta\chi\Delta\theta\Delta\phi = a^3\chi^2 \sin\theta\Delta\chi\Delta\theta\Delta\phi \ , \tag{27.13}$$

where $\epsilon_{\chi\theta\phi}$ is a component of the Levi Civita tensor).

Open universe $[k = -1]$: For $k = -1$ the geometry of the homogeneous hypersurfaces,

$$^{(3)}ds^2 = a^2[d\chi^2 + \sinh^2\chi(d\theta^2 + \sin^2\theta d\phi^2)] \ , \tag{27.14}$$

is different from geometries with which we have ordinary experience: The equatorial plane $\theta = \pi/2$ is a 2-surface whose circumference $2\pi a \sinh\chi$ increases with growing radius $a\chi$ faster than is permitted for any 2-surface that can ever reside in a 3-dimensional Euclidean space:

$$\frac{d(\text{circumference})}{d(\text{radius})} = 2\pi \cosh[(\text{radius})/a] > 2\pi \ . \tag{27.15}$$

Correspondingly, any attempt to embed that equatorial plane in a Euclidean 3-space is doomed to failure. As an alternative, we can embed it in a flat, Minkowskii 3-space with line element

$$^{(3)}ds^2 = -dT^2 + dr^2 + r^2 d\phi^2 \ . \tag{27.16}$$

The result is the hyperboloid of revolution,

$$T^2 - r^2 = a^2 \ , \tag{27.17}$$

which is shown pictorially in Fig. 27.2(c). By analogy it is reasonable to expect, and one easily can verify, that the full homogeneous hypersurface [metric (27.14)] has the same 3-geometry as the 3-dimensional hyperboloid

$$T^2 - r^2 = a^2 \tag{27.18}$$

in the 4-dimensional Minkowskii space

$$^{(4)}ds^2 = -dT^2 + dr^2 + r^2(d\theta^2 + \sin^2 d\phi^2) \ . \tag{27.19}$$

That this hyperboloid is, indeed, homogeneous and isotropic one can show by verifying that Lorentz transformations in the T, r, θ, ϕ 4-space can move any desired point on the hyperboloid into the origin, and can rotate the hyperboloid about the origin by an arbitrary angle. Note that the T, r, θ, ϕ space has no relationship whatsoever to the physical spacetime of our homogeneous, isotropic universe. It merely happens that both spaces possess 3-dimensional hypersurfaces with the same 3-geometry (27.14). Because these hypersurfaces are topologically open, with infinite volume, the $k = -1$ cosmological model is often called an "open universe."

[*Side remark*: Although homogeneity and isotropy force the cosmological model's hypersurfaces to have one of the three metrics (27.9), (27.12), (27.14), the topologies of those hypersurfaces need not be the obvious ones adopted and described above. For example, a flat model could have a closed topology with finite volume rather than an open topology with infinite volume. This could come about if, for example, in a Cartesian coordinate system $\{x = \chi \sin\theta\cos\phi,\ y = \chi\sin\theta\sin\phi,\ z = \chi\cos\theta\}$ the 2-surface $x = -L/2$ were identical to $x = +L/2$ (so x, like ϕ in spherical polar coordinates, is periodic), and if similarly $y = -L/2$ were identical to $y = +L/2$ and $z = -L/2$ were identical to $z = +L/2$. The resulting universe would have volume a^3L^3; and if one were to travel outward far enough, one would find oneself (as on the surface of the earth) returning to where one started. This and other unconventional choices for the topology of the standard cosmological models are kept in mind by cosmologists, just in case observational data someday should give evidence for them; but in the absence of such evidence, cosmologists assume the simplest choices of topology: those made above.]

Historically, the three possible choices for the geometry of a homogeneous, isotropic cosmological model were discovered by Alexander Alexandrovich Friedmann (1922), a Russian mathematician in Saint Petersburg; and, correspondingly, the specific solutions to the Einstein field equations which Friedmann constructed using those geometries are called *Friedmann cosmological models*. The first proof that these three choices are the only possibilities for the geometry of a homogeneous, isotropic cosmological model was given independently by Howard Percy Robertson (1935), who was a professor at Caltech, and by Arthur Geoffrey Walker (1936), who was a young researcher at the Royal College of Science in London; and, correspondingly, the general line element (27.6) with $\Sigma = \sin\chi$, χ, or $\sinh\chi$ is called the *Robertson-Walker line element*.

****

EXERCISES

Exercise 27.1 *Example: The Observers of a Synchronous Coordinate System*
Show that any observer who is at rest in a synchronous coordinate system [Eq. (27.4)] is freely falling, i.e., moves along a geodesic of spacetime.

Exercise 27.2 *Example: The 3-Sphere Geometry of a Closed Universe*

(a) Show, by construction, that there exist coordinates χ, θ, ϕ on the 3-sphere (27.11) [which resides in the the 4-dimensional Euclidean space of Eq. (27.10)] such that the 3-sphere's line element assumes the same form (27.9) as that of a homogeneous hypersurface in a closed, $k = +1$, universe.

(b) Show that the total 3-volume of this 3-sphere is $V = 2\pi^2 a^3$.

****

27.3 The Stress-energy Tensor and the Einstein Field Equation

The expansion factor, $a(t)$, of our zero-order, homogeneous, isotropic cosmological model is governed by the Einstein field equation $\mathbf{G} = 8\pi\mathbf{T}$. In order to evaluate that equation we shall need a mathematical expression for the stress-energy tensor, $\mathbf{T}$.

We shall deduce an expression for $\mathbf{T}$ in two different ways: by mathematical arguments, and by physical considerations; and the two ways will give the same answer. Mathematically, we note that because the spacetime geometry is homogeneous and isotropic, the Einstein curvature tensor must be homogeneous and isotropic, and thence the Einstein equation forces the stress-energy tensor to be homogeneous and isotropic. In the local Lorentz frame of a homogeneous observer, which has basis vectors

$$\vec{e}_{\hat{0}} = \frac{\partial}{\partial t} \; , \quad \vec{e}_{\hat{\chi}} = \frac{1}{a}\frac{\partial}{\partial \chi} \; , \quad \vec{e}_{\hat{\theta}} = \frac{1}{a\Sigma}\frac{\partial}{\partial \theta} \; , \quad \vec{e}_{\hat{\phi}} = \frac{1}{a\Sigma \sin\theta}\frac{\partial}{\partial \phi} \; , \tag{27.20}$$

the components of the stress-energy tensor are $T^{\hat{0}\hat{0}}$ =(energy density measured by homogeneous observer), $T^{\hat{0}\hat{j}}$ =(momentum density), $T^{\hat{j}\hat{k}}$ =(stress). Isotropy requires that the momentum density (a 3-dimensional vector in the homogeneous hypersurface) vanish; if it did not vanish, its direction would break the isotropy. Isotropy also requires that the stress, a symmetric-second rank 3-tensor residing in the homogeneous hypersurface, not pick out any preferred directions; and this is possible if and only if the stress is proportional to the metric tensor of the hypersurface. Thus, the components of the stress-energy tensor in the observer's local Lorentz frame must have the form

$$T^{\hat{0}\hat{0}} \equiv \rho \; , \quad T^{\hat{0}\hat{j}} = 0, \quad T^{\hat{j}\hat{k}} = P\delta^{jk} \; , \tag{27.21}$$

where ρ is just a new notation for the energy density, and P is the isotropic pressure. This is precisely the stress-energy tensor of a *perfect fluid* which is at rest with respect to the homogeneous observer. Reexpressed in geometric, frame-independent form, this stress-energy tensor is

$$\mathbf{T} = (\rho + P)\vec{u} \otimes \vec{u} + P\mathbf{g} \; , \tag{27.22}$$

where $\vec{u}$ is the common 4-velocity of the fluid and of the homogeneous observers

$$\vec{u} = \vec{e}_{\hat{0}} = \frac{\partial}{\partial t} \; . \tag{27.23}$$

Physical considerations lead to this same stress-energy tensor: The desired stress-energy tensor must be that of our own universe, coarse-grain-averaged over scales large compared to a cluster of galaxies, i.e., averaged over scales $\sim 10^8$ parsecs. The contributors to that stress-energy tensor will be (i) the galaxy clusters themselves, which like the atoms of a gas will produce a perfect-fluid stress-energy with ρ equal to their smeared-out mass density and P equal to $1/3$ times ρ times their mean square velocity relative to the homogeneous observers; (ii) the intercluster gas, which (one can convince oneself by astrophysical and observational arguments) is a perfect fluid, nearly at rest in the frame of the homogeneous observers;

(*iii*) the cosmic microwave radiation, which, being highly isotropic, has the stress-energy tensor of a perfect fluid with rest frame the same as that of the homogeneous observers; (*iv*) as-yet undetected cosmological backgrounds of other fundamental particles such as neutrinos, gravitons, axions, neutralinos, ..., which are expected on theoretical grounds to be homogeneous and isotropic when coarse-grain averaged, with the same rest frame as the homogeneous observers; and (v) a possibly nonzero stress-energy tensor of the vacuum, which we shall discuss in Sec. 27.4 below, and which also has the perfect-fluid form . Thus, all the contributors are perfect fluids, and their energy densities and pressures add up to give a stress-energy tensor of the form (27.22).

As in our analysis of relativistic stars (Sec. 25.3), so also here, before evaluating the Einstein field equation we shall study the local law of conservation of 4-momentum, $\vec{\nabla}\cdot\mathbf{T} = 0$. (That conservation law is always easier to evaluate than the field equation, and by virtue of the contracted Bianchi identity it is equivalent to some combination of components of the field equation.)

The quantity $\vec{\nabla}\cdot\mathbf{T}$, which appears in the law of 4-momentum conservation, is a vector. Since $\mathbf{T}$ has already been forced to be spatially isotropic, the spatial, 3-vector part of $\vec{\nabla}\cdot\mathbf{T}$, i.e., the projection of this quantity into a homogeneous hypersurface, is guaranteed already to vanish. Thus, only the projection orthogonal to the hypersurface, i.e., along $\vec{e}_{\hat{0}} = \vec{u} = \partial/\partial t$, will give us any information. This projection is viewed by a homogeneous observer, or equivalently by the perfect fluid, as the law of energy conservation. Evaluation of it, i.e., computation of $T_{\hat{0}}{}^{\hat{\mu}}{}_{;\hat{\mu}} = 0$ with $\mathbf{T}$ given by (27.21) and the metric given by (27.6), yields (Exercise 26.3)

$$\frac{d(\rho a^3)}{dt} = -P\frac{da^3}{dt} \,. \tag{27.24}$$

This is precisely the first law of thermodynamics for a perfect fluid, as one can see by the following calculation: Imagine a rectangular parallelopiped of fluid contained in the spatial region between χ and $\chi+\Delta\chi$, between θ and $\theta+\Delta\theta$, and between ϕ and $\phi+\Delta\phi$. As time passes the "walls" of this parallelopiped remain fixed relative to the homogeneous observers (since the walls and the observers both keep x^j fixed as t passes), and correspondingly the walls remain fixed in the fluid's rest frame. The volume of this fluid element is $\Delta V = a^3\Sigma^2\sin\theta\Delta\chi\Delta\theta\Delta\phi$, and the total mass-energy contained in it is $E = \rho V$. Correspondingly, the first law of thermodynamics for the fluid element, $dE/dt = -PdV/dt$ says

$$\frac{\partial(\rho a^3\Sigma^2\sin\theta\Delta\chi\Delta\theta\Delta\phi)}{\partial t} = -P\frac{\partial(a^3\Sigma^2\sin\theta\Delta\chi\Delta\theta\Delta\phi)}{\partial t} \,. \tag{27.25}$$

By dividing out the coordinate volume $\Sigma^2\sin\theta\Delta\chi\Delta\theta\Delta\phi$ (which is time independent), and then replacing the partial derivative by an ordinary derivative (because ρ and a depend only on t), we obtain the local law of energy conservation (27.24).

The fact that the local law of energy conservation, $T_{\hat{0}}{}^{\hat{\mu}}{}_{;\hat{\mu}}$ is identical to the first law of thermodynamics should not be surprising. Into our stress-energy tensor we put only the contribution of a perfect fluid, so energy conservation for it, in the fluid's local rest frame, must reduce to energy conservation for a perfect fluid, which *is* the first law of thermodynamics. If we had put other contributions into the stress-energy tensor, we would have obtained from energy conservation corresponding contributions to the first law; for example (as we shall see in Part IV, when we study fluid mechanics), if we had put viscous stresses into the stress-energy tensor, we would have obtained the first law in the form $d(\rho V) = -PdV + TdS$, including an explicit expression for the entropy increase dS due to viscous heating.

Turn, next, to the components of the Einstein equation $\mathbf{G} = 8\pi\mathbf{T}$. Because the metric has already been forced to be homogeneous and isotropic, the Einstein tensor is guaranteed already to have the homogeneous, isotropic form $G^{\hat{0}\hat{0}} \neq 0$, $G^{\hat{0}\hat{j}} = 0$, $G^{\hat{j}\hat{k}} \propto \delta^{jk}$, i.e., the same form as the stress-energy tensor (27.21). Correspondingly, there are only two nontrivial components of the Einstein field equation, the time-time component and the isotropic (proportional to δ^{jk}) space-space component. Moreover, the contracted Bianchi identity guarantees that some combination of these two components will be equivalent to our nontrivial law of energy conservation, thereby leaving only one new piece of information to be extracted from the Einstein equation. We shall extract that information from the time-time component, $G^{\hat{0}\hat{0}} = 8\pi T^{\hat{0}\hat{0}}$. A straightforward but tedious evaluation of $G^{\hat{0}\hat{0}} = G^{tt}$ for the Robertson-Walker line element (27.6), and insertion into the field equation along with $T^{\hat{0}\hat{0}} = T^{tt} = \rho$ gives

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi}{3}\rho \; , \tag{27.26}$$

where the dot represents a derivative with respect to the homogeneous observers' proper time t.

To verify that no errors have been made, one can evaluate the remaining nontrivial component of the field equation, $G^{\hat{x}\hat{x}} = 8\pi T^{\hat{x}\hat{x}}$ (or the $\hat{\theta}\hat{\theta}$ or $\hat{\phi}\hat{\phi}$ component; they are all equivalent since $G^{\hat{j}\hat{k}}$ and $T^{\hat{j}\hat{k}}$ are both proportional to δ^{jk}). The result,

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = -8\pi P \; , \tag{27.27}$$

is, as expected, a consequence of the first of the Einstein components (27.26) together with the law of energy conservation (27.24): by differentiating (27.26) and then using (27.24) to eliminate $\dot{\rho}$, one obtains (27.27).

The task of computing the time evolution of our zero-order cosmological model now takes the following form: (i) Specify an equation of state

$$P = P(\rho) \tag{27.28}$$

for the cosmological perfect fluid; (ii) integrate the first law of thermodynamics

$$\frac{d\rho}{da} = -3\frac{(\rho + P)}{a} \tag{27.29}$$

[Eq. (27.24), rearranged] to obtain the density ρ [and via Eq. (27.28)] the pressure P as functions of the expansion factor a; (iii) evolve the expansion factor forward in time using the field equation

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi}{3}\rho \tag{27.30}$$

[Eq. (27.26)].

****

EXERCISES

Exercise 27.3 *Practice: Energy Conservation for a Perfect Fluid*

Consider a perfect fluid, with the standard stress-energy tensor $T^{\alpha\beta} = (\rho + P)u^\alpha u^\beta + Pg^{\alpha\beta}$. Assume that the fluid resides in an arbitrary spacetime, not necessarily our homogeneous, isotropic cosmological model.

(a) Explain why the law of energy conservation, as seen by the fluid, is given by $u_\alpha T^{\alpha\beta}{}_{;\beta} = 0$.

(b) Show that this law of energy conservation reduces to

$$\frac{d\rho}{d\tau} = -(\rho + P)\vec{\nabla} \cdot \vec{u} \ , \tag{27.31}$$

where τ is proper time as measured by the fluid.

(c) Show that for a fluid at rest with respect to the homogeneous observers in our homogeneous, isotropic cosmological model, (27.31) reduces to the first law of thermodynamics (27.24). *Note:* as a tool in this calculation, you might want to derive and use the following formulas, which are valid in any *coordinate* basis:

$$\Gamma^\alpha{}_{\mu\alpha} = \frac{1}{\sqrt{-g}}(\sqrt{-g})_{,\mu} \ , \quad A^\alpha{}_{;\alpha} = \frac{1}{\sqrt{-g}}(\sqrt{-g}A^\alpha)_{,\alpha} \ . \tag{27.32}$$

Here, g denotes the determinant of the covariant components of the metric.

$$g \equiv \det||g_{ij}|| \ . \tag{27.33}$$

****

27.4 Evolution of the Universe

27.4.1 Constituents of the universe: Cold matter, radiation, and exotic matter

The evolution of our zero-order cosmological model is highly dependent on the equation of state $P(\rho)$; and that equation of state, in turn, depends on the types of matter and fields that fill the universe—i.e., the universe's *constituents*.

The constituents can be divided into three classes: (i) *Cold matter*, i.e. material whose pressure is neglible compared to its total density of mass-energy so the equation of state can be idealized as $P_M = 0$ (subscript M for "matter"). The cold matter includes the *baryonic matter* of which people, planets, stars, galaxies, and intergalactic gas are made, as well as so-called *cold, dark matter* which is known to exist in profusion and might be predominantly fundamental particles (axions or neutralinos). (ii) *Radiation*, i.e. material with equation of state $P_R = \rho_R/3$. This includes the CMB (primordial photons), primordial gravitons, primordial neutrinos when their temperatures exceed their rest masses, and other finite-rest-mass particles when the temperature is sufficiently high (i.e., very early in the universe). (iii) *Exotic matter* or *dark energy* with very negative pressure, $P_X \lesssim -\frac{1}{2}\rho_X$. As we shall see in Sec. 27.5, observations give strong evidence that such matter is present today in profusion. We do not yet know for sure its nature or its equation of state, but the most likely candidate is a nonzero stress-energy tensor associated with the vacuum, for which the equation of state is $P_X = -\rho_X$.

27.4.2 The vacuum stress-energy tensor

Let us digress, briefly, to discuss the vacuum: The stress-energy tensors of quantum fields are formally divergent, and must be renormalized to make them finite. In the early decades of quantum field theory, it was assumed that (in the absence of boundaries such as those of highly electrically conducting plates) the renormalized vacuum stress-energy tensor $\mathbf{T}_{\rm vac}$ would vanish. In 1968 Yakov Borisovich Zel'dovich initiated speculations that $\mathbf{T}_{\rm vac}$ might, in fact, be nonzero, and those speculations became fashionable in the 1980s in connection with inflationary models for the very early universe (Sec. 27.7). It was presumed in the 1980s and 90s that a phase transition in the early universe had driven quantum fields into a new vacuum state, for which $\mathbf{T}_{\rm vac}$ vanishes; but in the late 1990s, much to physicists' amazement, observational evidence began to mount that our universe today is filled with a profusion of exotic matter, most likely in the form of a nonzero $\mathbf{T}_{\rm vac}$; and by 2000 that evidence was getting compellingly strong.

If $\mathbf{T}_{\rm vac}$ is nonzero, what form can it take? It must be a second-rank symmetric tensor, and it would be very surprising if that tensor broke the local homogeneity and isotropy of spacetime or picked out any preferred reference frames. In order not to break those local symmetries, $\mathbf{T}_{\rm vac}$ must be proportional to the metric tensor, with its proportionality factor independent of location in spacetime:

$$\mathbf{T}_{\rm vac} = -\rho_X \mathbf{g} \; , \quad \text{i.e.,} \quad T^{\hat 0\hat 0}_{\rm vac} = \rho_X \; , \quad T^{\hat j\hat k}_{\rm vac} = -\rho_X \; . \tag{27.34}$$

This is a perfect-fluid equation of state with $P_X = -\rho_X$.

If there is no significant transfer of energy or momentum between the vacuum and other constituents of the universe, then energy-momentum conservation requires that $\mathbf{T}_{\rm vac}$ be divergence free. This, together with the vanishing divergence of the metric tensor, implies that ρ_X is constant, despite the expansion of the universe! This constancy can be understood in terms of the first law of thermodynamics (27.24): As the universe expands, the expansion does work against the vacuum's tension $-P_X = \rho_X$ at just the right rate as to replenish the vacuum's otherwise-decreasing energy density. For further insight into $\mathbf{T}_{\rm vac}$, see Box 27.1.

27.4.3 Evolution of the densities

In order to integrate the Einstein equation backward in time and thereby deduce the universe's past evolution, we need to know how much radiation, cold matter, and exotic matter

Box 27.1
The Cosmological Constant

Soon after formulating general relativity, Einstein discovered that his field equation, together with then plausible equations of state $P(\rho)$, demanded that the universe be either expanding or contracting; it could not be static. Firmly gripped by the mind-set of his era, Einstein regarded a nonstatic universe as implausible, and thus thought his field equation incompatible with the way the universe ought to behave; and so he modified his field equation. There were very few possibilities for the modification, since (i) it seemed clear that the source of curvature should still be the stress-energy tensor, and accordingly the field equations should say $\mathbf{E} = 8\pi\mathbf{T}$ where $\mathbf{E}$ is a tensor (evidently *not* the Einstein tensor) which characterizes gravity; and (ii) in order that the field equation leave four of the metric coefficients arbitrary (so they could be adjusted by coordinate freedom) the tensor $\mathbf{E}$ should have an automatically vanishing divergence. Of the various possibilities for $\mathbf{E}$, one stood out as far simpler than all the rest: $\mathbf{E} = \mathbf{G} + \Lambda\mathbf{g}$, where Λ is a "cosmological constant." To Einstein's great satisfaction, by choosing Λ negative he was able, from his modified field equation

$$\mathbf{G} + \Lambda\mathbf{g} = 8\pi\mathbf{T}\ , \tag{27.35}$$

to obtain a forever-static, homogeneous and isotropic cosmological model; see Ex. 27.5.

In 1929 Edwin Powell Hubble (1929), at the Mount Wilson Observatory, discovered that the universe was expanding. What a shock this was to Einstein! After visiting Mount Wilson and discussing Hubble's observations with him, Einstein (1931) formally renounced the cosmological constant and returned to his original, 1915, field equation $\mathbf{G} = 8\pi\mathbf{T}$. In his later years, Einstein described the cosmological constant as the greatest mistake of his life. Had he stuck to his original field equation, the expansion of the universe might have been regarded as the greatest of all the predictions made by his general relativity.

Remarkably, the cosmological-constant term $\Lambda\mathbf{g}$ in Einstein's modified field equation is identical to the modern vacuum contribution to the stress-energy tensor. More specifically, if we define $\rho_X \equiv \Lambda/8\pi$ so $T_{\rm vac} = -\rho_X\mathbf{g} = -(\Lambda/8\pi)\mathbf{g}$, then $\mathbf{G} + \Lambda\mathbf{g} = 8\pi\mathbf{T}$ becomes $\mathbf{G} = 8\pi(\mathbf{T} + \mathbf{T}_{\rm vac})$. Thus, the modern conclusion that there is probably a nonzero vacuum stress-energy tensor is actually a return to Einstein's modified field equation.

the universe contains today. Those amounts are generally expressed as fractions of the *critical energy density* that marks the dividing line between a closed universe and an open one. By asking that k/a^2 be zero, we find from the Einstein equation (27.26) that

$$\rho_{\rm crit} = \frac{3}{8\pi}\left(\frac{\dot a_o}{a_o}\right)^2 \simeq 8\times 10^{-30}{\rm g/cm}^3\ . \tag{27.36}$$

Here we have used a numerical value of $\dot{a}_o/a_o$ (the value of $\dot{a}/a$ today) that is discussed in Sec. 27.5 below. The energy density today in units of the critical density is denoted

$$\Omega \equiv \frac{\rho_o}{\rho_{\rm crit}} , \tag{27.37}$$

and observations give values

$$\Omega_R \sim 10^{-4} , \quad \Omega_M \sim 0.4 , \quad \Omega_X \sim 0.6 , \quad \Omega \equiv \Omega_R + \Omega_M + \Omega_X \sim 1 . \tag{27.38}$$

We shall discuss these numbers and the observational error bars on them in Sec. 27.5.

The evolution of the universe could be influenced by energy transfer among its three constituents. However, that transfer was small during the epoch from $a/a_o \sim 3 \times 10^{-13}$ to today; see Box 27.2. This means that the first law of thermodynamics (27.24) must hold true for each of the three constituents individually: $d(\rho a^3) = -P da^3$. By combining this law with the constituents' equations of state, $P_M = 0$, $P_R = \rho_R/3$, and (assuming the exotic matter is vacuum) $P_X = -\rho_X$, we obtain

$$\rho_R = \rho_{Ro}\frac{a_o^4}{a^4}, \quad \rho_M = \rho_{Mo}\frac{a_o^3}{a^3} , \quad \rho_X = \text{const} . \tag{27.39}$$

These relations are plotted in Fig. 27.4 below, which we shall discuss later.

The qualitative evolution of our zero-order cosmological model is easily deduced by inserting Eqs. (27.39) into Einstein's equation (27.26) and rewriting the result in the standard form for the motion of a particle in a potential well:

$$\frac{1}{2}\dot{a}^2 + V(a) = \frac{-k}{2} , \tag{27.40}$$

where

$$V(a) = -\frac{4\pi}{3}a^2\rho = -\frac{4\pi}{3}\rho_{\rm crit}a_o^2\left(\Omega_R\frac{a_o^2}{a^2} + \Omega_M\frac{a_o}{a} + \Omega_X\frac{a^2}{a_o^2}\right) . \tag{27.41}$$

Note that a/a_o is the ratio of the linear size of the universe at some time in the past, to the size of the universe today. Each volume element, comoving with the homogeneous observers, expands in length by a_o/a from then until now, and expands in volume by $(a_o/a)^3$.

The shape of the effective potential $V(a)$ is shown in Fig. 27.3: It increases monotonically from $-\infty$ at $a=0$ to about $4\rho_{\rm crit}a_o^2$ at $a/a_o \simeq -0.7$, and then, as the universe nears our own era, it begins decreasing. The universe is radiation dominated at $a/a_o \lesssim 10^{-4}$ (Fig. 27.4), it is cold-matter dominated between $a/a_o \sim 10^{-4}$ and $a/a_o \sim 1$, and the maxing-out of the effective potential and reversal to plunge is triggered by a modern-era ($a/a_o \sim 1$) transition to exotic-matter dominance.

The implications of this effective potential for the past evolution of our universe should be clear from one's experience with particle-in-potential problems: The universe must have expanded at an ever decreasing rate $\dot{a}$ from an age $\sim$ a small fraction of second, when our equations of state became valid, until nearly the present epoch, $a/a_o \sim 0.7$, and then the universe's vacuum tension must have triggered an acceleration of the expansion.

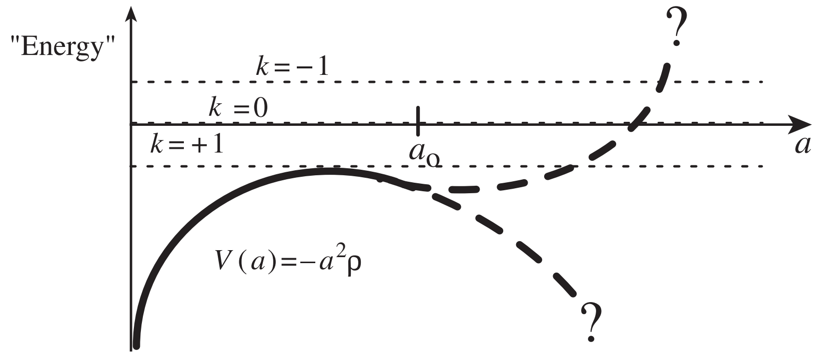

Fig. 27.3: The "particle-in-a-potential" depiction of the evolutionary equation (27.40) for the expansion factor a of the universe. Plotted horizontally is the expansion factor, which plays the role of the position of a particle. Plotted vertically is the "particle's" potential energy $V(a)$ (thick curve) and its total energy $-k/2$ (thin dotted line). The difference in height between the dotted line and the thick curve is the particle's kinetic energy $\frac{1}{2}\dot{a}^2$, which must be positive. The form of $V(a)$ in the past ($a \leq a_o$) is shown solid. The form in the future is unknown because we do not know the nature of the exotic matter. If the form is that of the upper thick dashed curve, the universe will reach a maximum size and then recontract, independently of whether it is spatially open $k = -1$, flat $k = 0$ or closed $k = +1$. If the form is that of the lower dashed curve, the univeral expansion will continue to accelerate forever.

It seems strange that the universe should switch over to acceleration in just the epoch that we are alive, rather than far earlier or far later or not at all. The reasons for this are unknown. It is a big surprise, revealed by recent observations.

If P_X/ρ_X is independent of time, then the universe's past evolution is *not* very sensitive to the precise value of P_X/ρ_X. For $P_X/\rho_X = -\frac{1}{2}$ (the least negative pressure allowed by current observations), as for $P_X/\rho_X = -1$ (vacuum), the exotic matter begins to influence $V(a)$ significantly only in the modern era, and its influence is to accelerate the expansion in accord with observation. It is of no dynamical importance earlier.

However, nothing requires that P_X/ρ_X be constant. It is possible, in principle for P_X/ρ_X to evolve in a wide variety of ways that, in principle, could have had a strong influence on the universe's early evolution. That this probably did *not* occur we know from observational data which show that the exotic matter cannot have had a very significant influence on the universal expansion at several key epochs in the past: (i) during the nucleosynthesis of light elements when the universe was about 1 minute old, (ii) during recombination of the primordial plasma (conversion from ionization to neutrality) when the universe was about 10^6 years old, and (iii) during early stages of galaxy formation when the universe was about 1 billion years old. Nevertheless, we are so ignorant, today, of the precise nature of the exotic matter, that we must be prepared for new surprises.

By contrast, the evolution of the exotic matter in the future and the resulting evolution of the universe are unconstrained by observation and are unknown. Until we learn for sure the nature and dynamics of the exotic matter, we cannot predict the universe's future evolution.

Box 27.2
Interaction of Radiation and Matter

In the present epoch there is negligible radiation/matter interaction: the radiation propagates freely, with negligible absorption by interstellar or intergalactic gas. However, much earlier, when the matter was much denser and the radiation much hotter than today and galaxies had not yet formed, the interaction must have been so strong as to keep the photons and matter in thermodynamic equilibrium with each other. In this early epoch the matter temperature, left to its own devices, would have liked to drop as $1/\text{Volume}^{2/3}$, i.e., as $1/a^2$, while the radiation temperature, left to its own devices would have dropped as $1/\text{Volume}^{1/3}$, i.e., as $1/a$. To keep their temperatures equal, the photons had to feed energy into matter. This feeding was not at all a serious drain on the photons' energy supply, however: Today the ratio of the number density of background photons to the number density of baryons (i.e., protons and neutrons) is

$$\frac{n_{Ro}}{n_{Mo}} = \frac{a_R {T_{Ro}}^4/(2.8kT_{Ro})}{\rho_{mo}/m_p} \sim 10^8 \; , \tag{1}$$

where m_p is the proton mass, k is Boltzman's constant and $2.8kT_{Ro}$ is the average energy of a black-body photon at the CMB temperature $T_{Ro} = 2.728$ K. Because, aside from a factor of order unity, this ratio is the entropy per baryon in the universe (Chaps. 3 and 4), and because the entropy per baryon was (nearly) conserved during the (nearly) adiabatic expansion of the universe, this ratio was about the same in the early era of thermal equilibrium as it is today. Since the specific heat per photon and that per baryon are both of order Boltzman's constant k, the specific heat of a unit volume of background radiation in the early era exceeded that of a unit volume of matter by eight orders of magnitude. Consequently, the radiation could keep the matter's temperature equal to its own with little effort; and accordingly, despite their interaction, the radiation by itself satisfied energy conservation to high accuracy.

This remained true, going backward in time, until the temperature reached $T \sim \frac{1}{5}m_e/k \sim 10^9$ K, at which point electron-positron pairs formed in profusion and sucked roughly half the photon energy density out of the photons. This pair formation, going backward in time, or pair annihilation going forward, occurred when the universe was several 10's of seconds old (Fig. 27.4) and can be regarded as converting one form of radiation (photons) into another (relativistic pairs). Going further backward in time, at $T \sim m_p/k \sim 10^{13}$ K, the neutrons and protons (baryons) became relativistic, so cold matter ceased to exist—which means that, going forward in time, cold matter formed at $T \sim m_p/k \sim 10^{13}$ K.

As is shown in the text, the exotic matter only became significant in the modern era; so its interaction with cold matter and radiation (if any, and there presumably is very little) cannot have been important during the universe's past evolution.

27.4.4 Evolution in time and redshift

Since the exotic matter cannot have had an important dynamical role, we shall ignore it in the remainder of this section and shall idealize the universe as containing only cold matter, with density $\rho_M = \rho_{Mo}(a_o/a)^3$ and radiation with density $\rho_R = \rho_{Ro}(a_o/a)^3$.

The radiation includes the cosmic background photons (which today are in the microwave frequency band), plus gravitons, and plus those neutrinos whose rest masses are much less than their thermal energies. In order for the observed abundances of the light elements to agree with the theory of their nucleosynthesis, it is necessary that the neutrino and graviton contributions to ρ_R be less than or of order the photon contributions, so we shall neglect the gravitons and neutrinos and build an approximate model containing only photons and cold matter.

The photons were in thermal equilibrium with other forms of matter in the early universe and thus had a Planckian spectrum. Since black-body radiation has energy density $\rho_R \propto T_R^4$ and the density decreases with expansion as $\rho_R \propto 1/a^4$, the photon temperature must have been redshifted during this early era as $T_R \propto 1/a$. When the temperature dropped below $\sim 10^4$ K, the electrons dropped into bound states around atomic nuclei, making the matter neutral, and its opacity negligible, so the photons were librated from interaction with the matter and began to propagate freely. Kinetic theory (Box 27.3) tells us that during this free propagation, the photons retained their Planckian spectrum, and their temperature continued to be redshifted as $T_R \propto 1/a$. In accord with this prediction, the spectrum of the photons today is measured to be Planckian to very high accuracy; its temperature is $T_{Ro} = 2.728$ K.

To recapitulate: the matter and radiation densities and temperatures must have evolved as

$$\rho_M = \rho_{Mo}\frac{a_o^3}{a^3} , \quad \rho_R = \rho_{Ro}\frac{a_o^4}{a^4} , \quad T_R = T_{Ro}\frac{a_o}{a} , \tag{27.42}$$

throughout the entire epoch from $a/a_o \sim 3 \times 10^{-13}$ until today.

The density and temperature evolutions (27.42) are depicted as functions of the universe's expansion factor a/a_o in Fig. 27.4. A second way to express the evolution is in terms of *cosmological redshift*: Imagine photons emitted in some chosen spectral line of some chosen type of atom (e.g., the Lyman alpha line of atomic hydrogen), at some chosen epoch during the universe's evolution. Let the atoms be at rest in the mean rest frame of the matter and radiation, i.e., in the rest frame of a homogeneous observer, so they move orthogonally to the homogeneous hypersurfaces. Focus attention on specific photons that manage to propagate to earth without any interaction whatsoever with matter. Then they will arrive with a wavelength, as measured on earth today, which is much larger than that with which they were emitted: The expansion of the universe has increased their wavelength, i.e., has redshifted them. As is shown in Exercise 27.5 below, if the expansion factor was a at the time of their emission, and if their wavelength at emission as measured by the emitter was λ, then at reception on earth as measured by an astronomer, they will have wavelength λ_o given by

$$\frac{\lambda_o}{\lambda} = \frac{a_o}{a} ; \tag{27.43}$$

Box 27.3
Kinetic Theory of Photons in General Relativity

The kinetic theory of photons and other particles (Chap. 2) can be lifted from special relativity into general relativity using the equivalence principle:

In any local Lorentz frame in curved spacetime, the number density in phase space is given by the special relativity expression (2.11): $\mathcal{N}(\mathcal{P}, \vec{p}) = dN/d\mathcal{V}_x d\mathcal{V}_p$. Here $\mathcal{P}$ is the location of the observer in spacetime, $\vec{p}$ is the momentum of some chosen "fiducial" photon, $d\mathcal{V}_x$ is a small 3-volume at $\mathcal{P}$ in the physical space of the observer's local Lorentz frame, $d\mathcal{V}_p$ is a small 3-volume in the momentum space of the observer's local Lorentz frame, centered on $\vec{p}$, and dN is the number of photons in $d\mathcal{V}_x$ and $d\mathcal{V}_p$. For a homogeneous observer, we can choose $d\mathcal{V}_x = a^3\Sigma^2 \sin\theta \, d\chi \, d\theta \, d\phi$, $d\mathcal{V}_p = dp^{\hat{\chi}} dp^{\hat{\theta}} dp^{\hat{\phi}}$, where the hats denote components on the unit vectors $\vec{e}_{\hat{\chi}}$, $\vec{e}_{\hat{\theta}}$, $\vec{e}_{\hat{\phi}}$.

The equivalence principle guarantees that, just as in flat spacetime, so also in curved spacetime, (i) the number density in phase space, $\mathcal{N}$, is independent of the velocity of the local Lorentz frame in which it is measured (with all the frames presumed to be passing through the event $\mathcal{P}$); and (ii) if the photons do not interact with matter, then $\mathcal{N}$ is constant along the world line of any chosen (fiducial) photon as it moves through spacetime and as its 4-momentum $\vec{p}$ evolves in the free-particle (geodesic) manner. [In asserting this constancy of $\mathcal{N}$, one must examine carefully the issue of curvature coupling; Sec. 24.7. Because the volume element $d\mathcal{V}_x$ involved in the definition of $\mathcal{N}$ has some finite, though tiny size, spacetime curvature will produce geodesic deviation between photons on opposite sides of $d\mathcal{V}_x$. One can show fairly easily, however, that this geodesic deviation merely skews the phase-space volume element along its momentum directions in a manner analogous to Fig. 2.6(b), while leaving the product $d\mathcal{V}_x \, d\mathcal{V}_p$ fixed and thereby leaving $\mathcal{N}$ unchanged; cf. Sec. 2.7.]

The equivalence principle also guarantees that in curved spacetime, as in flat, the number density in phase space can be expressed in terms of the specific intensity I_ν and the frequency of the chosen photon ν (as measured in any local Lorentz frame): $\mathcal{N} = h^{-4} I_\nu/\nu^3$ [Eq. (2.18)]. If the spectrum is Planckian with temperature T as measured in this Lorentz frame, then $\mathcal{N}$ will have the form

$$\mathcal{N} = \frac{2}{h^3} \frac{1}{e^{h\nu/kT} - 1} \,. \tag{1}$$

The Lorentz-invariance and conservation of $\mathcal{N}$, together with the fact that this $\mathcal{N}$ depends only on the ratio ν/T, implies that, (i) a spectrum that is Planckian in one reference frame will be Planckian in all reference frames, with the temperature T getting doppler shifted in precisely the same manner as the photon frequency ν; and (ii) an initially Planckian spectrum will remain always Planckian (under free propagation), with its temperature experiencing the same cosmological redshift, gravitational redshift, or other redshift as the frequencies of its individual photons.

For the CMB as measured by homogeneous observers, the frequencies of individual photons get redshifted by the expansion as $\nu \propto 1/a$, so the photon temperature also gets redshifted as $T \propto 1/a$.

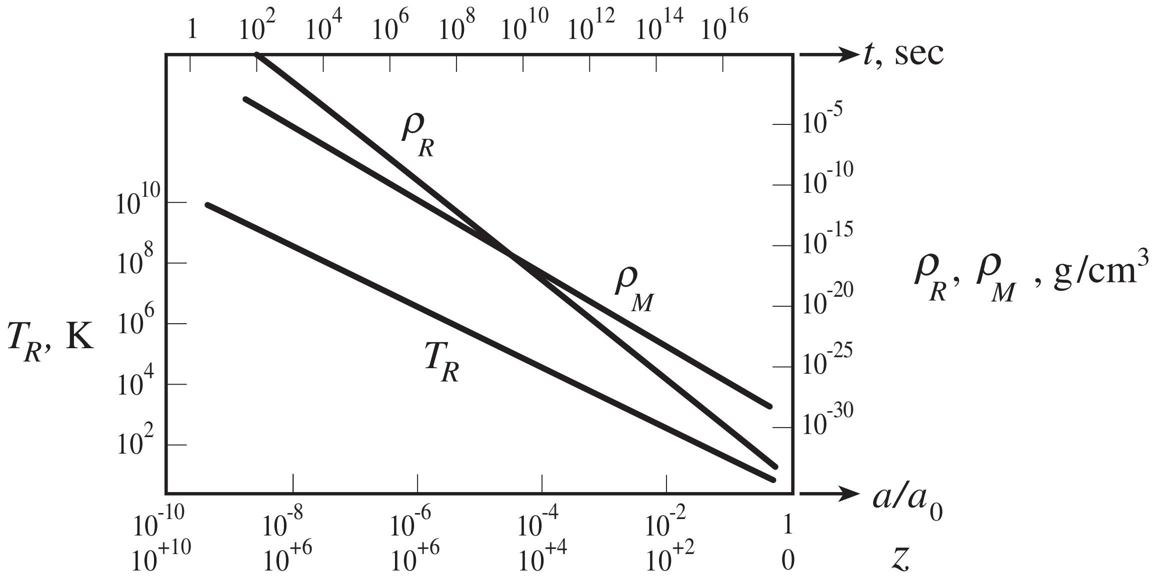

Fig. 27.4: The evolution of the total mass-energy densities ρ_M and ρ_R in matter and in radiation and the radiation's photon temperature T_R, as functions of the expansion factor a of the universe, the cosmological redshift z and the proper time t (in the mean rest frame of the matter and radiation) since the "big bang."

i.e., *the photons' wavelength is redshifted in direct proportion to the expansion factor of the universe.* It is conventional to speak of the redshift z as not the ratio of the wavelength today to that when emitted, but rather as the fractional change in wavelength, so

$$z \equiv \frac{\lambda_o - \lambda}{\lambda} = \frac{a_o}{a} - 1 \, . \tag{27.44}$$

In Fig. 27.4's depiction of the density evolution of the universe, the horizontal axis at the bottom is marked off in units of z.

It is also instructive to examine the density evolution in terms of proper time, t, as measured in the mean rest frame of the matter and radiation; i.e., as measured by clocks carried by the homogeneous observers. At redshifts $z > \rho_{Mo}/\rho_{Ro} \sim 10^4$, when the energy in radiation dominated over that in matter, a as a function of time was governed by the Einstein field equation

$$\left(\frac{\dot a}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi}{3}\rho_R = \frac{8\pi}{3}\rho_{Ro}\left(\frac{a_o}{a}\right)^4 \, . \tag{27.45}$$

[Eq. (27.30)]. As we shall see in Sec. 27.5 below, the present epoch of the universe's expansion is an early enough one that, if it is closed or open, the evolution has only recently begun to depart significantly from that associated with a flat, $k = 0$ model. Correspondingly, in the early, radiation-dominated era, the evolution was independent of k to high precision, i.e., the factor k/a^2 in the evolution equation was negligible. Ignoring that factor and integrating

Eq. (27.30), then setting $\rho \cong \rho_R = \rho_{Ro}(a_o/a)^4$, we obtain

$$\rho \cong \rho_R = \frac{3}{32\pi t^2} \; , \quad \frac{a}{a_o} = \left(\frac{32\pi}{3} \rho_{Ro} t^2 \right)^{1/4} \quad \text{when } \frac{a}{a_o} < \frac{\rho_{Ro}}{\rho_{Mo}} \sim 10^{-4} \; . \tag{27.46}$$

Here the origin of time, $t = 0$, is taken to be at the moment when the expansion of the universe began: the "big-bang."

This early, *radiation-dominated era* ended at a *cross-over time*

$$t_c = \left[\frac{3}{32\pi \rho_{Ro}} \left(\frac{\rho_{Ro}}{\rho_{Mo}} \right)^4 \right]^{1/2} \sim \left[\frac{3}{32\pi \times (5 \times 10^{-34}\mathrm{g/cm}^3)} 10^{-16} \right]^{1/2} \times$$
$$\times \left[\frac{1\,\mathrm{g}}{0.742 \times 10^{-28}\mathrm{cm}} \right]^{1/2} \times \frac{1\,\mathrm{year}}{0.946 \times 10^{18}\mathrm{cm}} \sim 1 \times 10^5 \,\mathrm{years}\; . \tag{27.47}$$

In this calculation the first two factors on the second line are introduced to convert from geometrized units to conventional units. After the crossover time the solution to the Einstein equation is that for pressure-free matter. The precise details of the time evolution will depend on whether the universe is open, closed, or flat; but the three possibilities will agree up to the present epoch to within a few tens of per cent (see Sec. 27.5). Ignoring the differences between open, closed, and flat, we can adopt the $k = 0$, pressure-free evolution as given by Friedmann's flat model, Eqs. (27.49), (27.50), (27.51), and (27.52) —but with the origin of time adjusted to match onto the radiation-dominated solution (27.46) at the cross-over time:

$$\rho \cong \rho_M = \frac{1}{6\pi(t + t_c/3)^2} \; , \quad \frac{a}{a_o} = \left[6\pi\rho_{Mo} \left(t + \frac{t_c}{3} \right)^2 \right]^{1/3} \quad \text{when } \frac{a}{a_o} > \frac{\rho_{Ro}}{\rho_{Mo}} \sim 10^{-4} \; . \tag{27.48}$$

The present age of the universe, as evaluated by setting $\rho_M = \rho_{Mo}$ in this formula and converting to cgs units, is of order 10^{10} years. We shall evaluate the age with higher precision in section 27.5.8 below. In Fig. 27.4's depiction of the evolution, the time t since the big bang, as computed from Eqs. (27.46), (27.47), and (27.48), is marked along the top axis.

27.4.5 Physical processes in the expanding universe

The evolution laws (27.42) and (27.46), (27.47), and (27.48) for ρ_M and ρ_R are a powerful foundation for deducing the main features of the physical evolution of the universe from the very early epoch, $a/a_o \sim 3 \times 10^{-13}$, i.e. $z \sim 3 \times 10^{12}$ and $t \sim 10^{-5}$ sec up to the present. For detailed, pedagogical presentations of those features see, e.g., Peebles (1971) and Zel'dovich and Novikov (1983). Here we shall just present a very short summary.

Some key physical events that one deduces during the evolution from $z = 3 \times 10^{12}$ to the present, $z = 0$, are these:

(i) At redshift $z \sim 3 \times sim 10^{12}$, baryon-antibaryon pairs annihilated and the thermal energies of neutrons and protons became much smaller than their rest-mass energies. This was the epoch of formation of baryonic cold matter. (ii) At redshifts $z \sim 10^9$ when the universe was of order a second old, the photons ceased being energetic enough to make electron-positron pairs; the pairs, which earlier had been plentiful, annihilated, feeding their energy into photons; and with their annihilation the primordial gas suddenly became transparent to neutrinos. Since then the neutrinos, born in thermodynamic equilibrium at $z > 10^9$, should have propagated freely.

(*iii*) At redshifts $z \sim 3 \times 10^8$, when the universe was a few minutes old, ρ_R was roughly 1 g/cm^3, and the temperature was $T_R = T_M \sim 10^9$ K, nuclear burning took place. Going into this epoch of *primordial nucleosynthesis* the matter consisted of equal numbers of protons, neutrons, and electrons all in thermodynamic equilibrium with each other. Coming out, according to evolutionary calculations for the relevant nuclear reactions, it consisted of about 75 per cent protons (by mass), 25 per cent alpha particles (^{4}He nuclei), and tiny ($< 10^{-6}$), but observationally important amounts of deuterium, ^{3}He, lithium, beryllium, and boron. [The agreement of these predictions with observation constitutes strong evidence that cosmologists are on the right track in their deductions about the early universe.] All the elements heavier than boron were almost certainly made in stars when the universe was billions of years old.

(*iv*) At the redshift $z \sim 10^4$, when the universe was about 10^5 years old and $T_R = T_M$ was about 3×10^4 K, came the cross-over from radiation dominance to matter dominance; i.e., $\rho_R = \rho_M \sim 10^{-17}\mathrm{g/cm}^3$.

(*v*) At the redshift $z \sim 10^3$, when the universe was $\sim 10^6$ years old and its temperature had dropped to roughly 3000 K and its density to $\sim 10^{-20}\mathrm{g/cm}^3$, the electrons in the primordial plasma were captured by the protons and alpha particles to form neutral hydrogen and helium. Before this *epoch of recombination* the matter was highly ionized and opaque to radiation; afterward it was highly neutral and transparent.

(*vi*) Before recombination, if any matter tried to condense into stars or galaxies, it would get adiabatically heated as it condensed and the rising pressure of radiation trapped inside it would prevent the condensation from proceeding. After recombination, radiation was no longer trapped inside a growing condensation. Now, for the first time, stars, galaxies, and clusters of galaxies could begin to form. Measured anisotopies of the CMB, however, tells us that the size of the density fluctuations at recombination was $\Delta\rho_M/\rho_M \sim 10^{-4}$, which is just the right size to grow, by gravitational condensation, to $\Delta\rho/\rho \sim 1$ at $z \sim 10$. Thus it is that the *epoch of galaxy formation* probably began around a redshift $z \sim 10$, when the universe was already about two billion years old compared to its present age of roughly 14 billion years.

(*vi*) In galaxies such as ours there has been, since formation, a continuing history of stellar births, nucleosynthesis, and deaths. The unravelling of our Galaxy's nucleosynthesis history and the resulting understanding of the origin of all the elements heavier than boron, was achieved, in large measure in the 1950s, 60s and 70s, by nuclear astrophysicists under the leadership of Caltech's William A. Fowler. Our own sun is of the second generation or later: By measuring in meteorites the relative abundances of unstable atomic nuclei and their decay products, Caltech's Gerald J. Wasserburg and his colleagues have deduced that the solar system is only 4.58 billion years old—i.e., it formed when our Milky Way Galaxy was already ~ 5 billion years old.

Before recombination the radiation was kept thermalized by interactions with matter. However, since recombination the radiation has propagated freely, interacting only with the gravitation (spacetime curvature) of our universe.

****

EXERCISES

Exercise 27.4 *Example: Friedman's Cosmological Models*
Consider a model universe of the type envisioned by Alexander Friedman (1922)—one with zero pressure (i.e., containing only cold matter), so its density is $\rho_M \propto 1/a^3$; cf. Eq. (27.39). Write this density in the form

$$\rho = \frac{3}{8\pi}\frac{a_m}{a^3} , \tag{27.49}$$

where a_m is a constant whose normalization (the factor $3/8\pi$) is chosen for later convenience.

(a) Draw the effective potential $V(a)$ for this model universe, and from it discuss qualitatively the evolution in the three cases $k = 0, \pm 1$.

(b) Show that for a closed, $k = +1$ universe with zero pressure, the expansion factor a evolves as follows:

$$a = \frac{a_m}{2}(1 - \cos\eta) , \quad t = \frac{a_m}{2}(\eta - \sin\eta) . \tag{27.50}$$

Here η is a parameter which we shall use as a time coordinate in Sec. 27.5 [Eq. (27.79)]below. This $a(t)$ [depicted in Fig. 27.5], is a *cycloid*.

(c) Show that for a flat, $k = 0$ universe, the evolution is given by

$$a = \left(\frac{9a_m}{4}\right)^{1/3} t^{2/3} , \quad \rho = \frac{1}{6\pi t^2} . \tag{27.51}$$

(d) Show that for an open, $k = -1$ universe, the evolution is given by

$$a = \frac{a_m}{2}(\cosh\eta - 1) , \quad t = \frac{a_m}{2}(\sinh\eta - \eta) . \tag{27.52}$$

Note, as shown in Fig. 27.5, that for small expansion factors, $a \ll a_m$, the evolutions of the three models are almost identical.

Exercise 27.5 *Example: Einstein's Static Universe*
Consider a model universe of the sort that Einstein (1917) envisioned: one with a nonzero, positive cosmological constant and containing matter with $P = 0$. Reinterpret this in modern language as a universe with cold matter and a nonzero vacuum stress-energy.

(a) Show that, if the effective potential $V(a)$ has a vanishing derivative today, $V'(a_o) = 0$, then this model universe is in a static state, neither expanding nor contracting, and (in the absence of small perturbations) can remain in that state indefinitely.

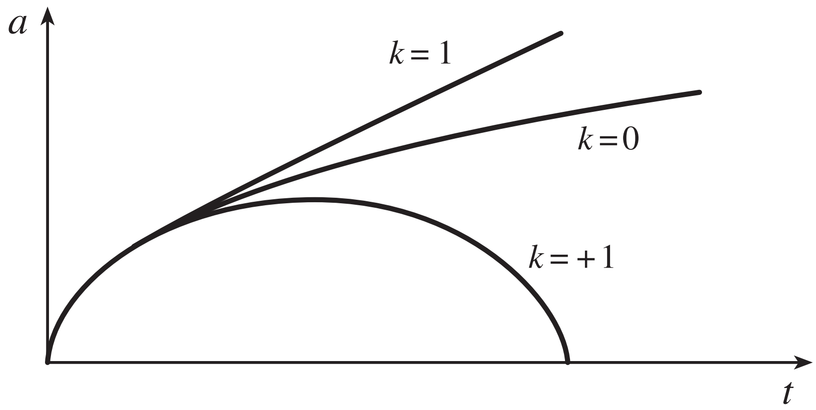

Fig. 27.5: Time evolution of the expansion factor $a(t)$ for the zero-pressure, Friedmann cosmological models (Exercise 27.4). The three curves correspond to the closed, $k = +1$ model; the flat, $k = 0$, model; and the open, $k = -1$ model.

(b) Assuming that the cold-matter density today is $\rho_{Mo} \sim 10^{-31}$ g/cm^3 and the size of the universe today is $a_o \gtrsim 10^{10}$ pc, what are the values of the vacuum energy density ρ_X and the cosmological constant Λ?

(c) Draw the effective potential $V(a)$ for this model universe. Using that diagram, show that this universe is unstable against gravitational collapse.

Exercise 27.6 *Example: Cosmological Redshift*

Consider a particle, with finite rest mass or zero, that travels freely through a homogeneous, isotropic universe. Let the particle have energy E as measured by a homogeneous observer side-by-side with it, when it starts its travels, at some early epoch in the universe's evolution; and denote by E_o its energy as measured by a homogeneous observer at its location, near earth, today. Denote by

$$p = \sqrt{E^2 - m^2}\ , \quad p_o = \sqrt{{E_o}^2 - m^2} \tag{27.53}$$

the momentum of the particle as measured in the early epoch and today. In this problem you will evaluate the ratio of the momentum today to the momentum in the early epoch, p_o/p, and will deduce some consequences of that ratio.

(a) Place the spatial origin, $\chi = 0$, of the spatial coordinates of a Robertson-Walker coordinate system [Eq. (27.6)] at the point where the particle started its travel. (Homogeneity guarantees we can put the spatial origin anywhere we wish,) Orient the coordinates so the particle starts out moving along the "equatorial plane" of the coordinate system, $\theta = \pi/2$ and along $\phi = 0$. (Isotropy guarantees we can orient our spherical coordinates about their origin in any way we wish.) Then spherical symmetry about the origin guarantees the particle will continue always to travel radially, with $\theta = \pi/2$ and $\phi = 0$ all along its world line; in other words, the only nonvanishing contravariant components of its 4-momentum are $p^t = dt/d\zeta$ and $p^\chi = d\chi/d\zeta$; and, since the metric is diagonal, the lowering of components shows that the only nonvanishing covariant components are p_t and p_χ. Show that the quantity p_χ is conserved along the particle's world line.

(b) Express the momentum p measured by the homogeneous observer at the starting point, and that p_o measured near earth today in terms of p_χ. [Hint: The local Lorentz frame of a homogeneous observer has the basis vectors (27.20).] Show that

$$\frac{p_o}{p} = \frac{a}{a_o} = \frac{1}{1+z} , \tag{27.54}$$

where z is the cosmological redshift at the starting point.

(c) Show that if the particle is a photon, then its wavelength is redshifted in accord with Eqs. (27.43) and (27.44).

(d) Show that if the particle has finite rest mass and has speed $v \ll 1$ at its starting point, as measured by a homogeneous observer there, then its velocity today as measured by the near-earth homogeneous observer will be

$$v_o = v\frac{a}{a_o} = \frac{v}{1+z} . \tag{27.55}$$

Exercise 27.7 *Practice: Cosmic Microwave Radiation in an Anisotropic Cosmological Model*

Consider a cosmological model with the spacetime metric

$$ds^2 = -dt^2 + a^2(t)dx^2 + b^2(t)dy^2 + c^2(t)dz^2 . \tag{27.56}$$

The quantities a, b, and c (not to be confused with the speed of light which is unity in this chapter) are expansion factors for the evolution of the universe along its x, y, and z axes. The Einstein field equation governs the time evolution of these expansion factors; but the details of that evolution will not be important to us in this problem.

(a) Show that the space slices $t =$ const in this model have Euclidean geometry, so the model is spatially flat and homogeneous. Show that the observers who see these slices as hypersurfaces of simultaneity, i.e., the homogeneous observers, have world lines of constant x, y, and z, and their proper time is equal to the coordinate time t.

(b) At time t_e when the expansion factors were a_e, b_e, and c_e the universe was filled with isotropic black-body photons with temperature T_e, as measured by homogeneous observers. Define $p_x \equiv \vec{p}\cdot\partial/\partial x$, $p_y \equiv \vec{p}\cdot\partial/\partial y$, $p_z \equiv \vec{p}\cdot\partial/\partial z$ for each photon. Show that in terms of these quantities the photon distribution function at time t_e is

$$\mathcal{N} = \frac{2}{h^3}\frac{1}{e^{E/kT_e}-1} , \quad \text{where } E = \left[\left(\frac{p_x}{a_e}\right)^2 + \left(\frac{p_y}{b_e}\right)^2 + \left(\frac{p_z}{c_e}\right)^2\right]^{1/2} . \tag{27.57}$$

(c) After time t_e each photon moves freely through spacetime (no emission, absorption, or scattering). Explain why p_x, p_y, and p_z are constants of the motion along the phase-space trajectory of each photon.

(d) Explain why $\mathcal{N}$, expressed in terms of p_x, p_y, p_z, retains precisely the form (27.57) for all times $t > t_e$.

(e) At time $t_o > t_e$, when the expansion factors are a_o, b_o, c_o, what are the basis vectors $\vec{e}_{\hat{0}}$, $\vec{e}_{\hat{x}}$, $\vec{e}_{\hat{y}}$, $\vec{e}_{\hat{z}}$ of the local Lorentz frame of a homogeneous observer?

(f) Suppose that such an observer looks at the photons coming in from a direction $\mathbf{n} = n_{\hat{x}}\vec{e}_{\hat{x}} + n_{\hat{y}}\vec{e}_{\hat{y}} + n_{\hat{z}}\vec{e}_{\hat{z}}$ on the sky. Show that she sees a precisely Planck frequency distribution with temperature T_o that depends on the direction $\mathbf{n}$ that she looks:

$$T_o = T_e \left[\left(\frac{a_o}{a_e} n_{\hat{x}}\right)^2 + \left(\frac{b_o}{b_e} n_{\hat{y}}\right)^2 + \left(\frac{c_o}{c_e} n_{\hat{z}}\right)^2 \right]^{-1/2} . \tag{27.58}$$

(g) In the case of isotropic expansion, $a = b = c$, show that T_o is isotropic and is redshifted by the same factor, $1 + z$, as the frequency of each photon [Eqs. (27.43) and (27.44)]:

$$\frac{T_o}{T_e} = \frac{1}{1+z} = \frac{a_e}{a_o} . \tag{27.59}$$

[The redshift z must not be confused with the coordinate z of Eq. (27.56).]

****

27.5 Observational Cosmology

27.5.1 Parameters characterizing the universe

Our zero-order (homogeneous and isotropic) model of the universe is characterized, today, by the following parameters: (i) The quantity $H_o \equiv \dot{a}_o/\dot{a}$, which is called the Hubble expansion rate, and which determines the critical density $\rho_{\rm crit} = (3/8\pi)H_o^2$ [Eq. (27.36)]. (ii) The density of cold matter measured in units of the critical density, $\Omega_M = \rho_{Mo}/\rho_{\rm crit}$ [Eq. (27.37)]. (iii) The split of Ω_M into two parts, $\Omega_M = \Omega_B + \Omega_{\rm CDM}$. Here Ω_B is that portion due to "baryonic matter," the type of matter (protons, neutrons, electrons, and atoms and molecules made from them) of which stars, galaxies, and interstellar gas are made; and $\Omega_{\rm CDM}$ is the portion due to bon-baryonic, "cold, dark matter" (probably axions and/or neutralinos and/or other types of weakly interacting, massive particles produced in the big bang). (iv) The temperature T_{Ro} of the CMB. (v) The density of radiation in units of the critical density, $\Omega_R = \rho_{Ro}/\rho_{\rm crit}$. (vi) Ω_X, the density of exotic matter in units of the critical density. (vii) P_X/ρ_X, the ratio of the exotic matter's pressure to its density (equal to -1 if the exotic matter is a nonzero stress-energy of the vacuum).

The time-time component of the Einstein field equation, Eq. (27.26), translated into the notation of our seven parameters, says

$$\frac{k}{a_o^2} = H_o^2(\Omega - 1) , \quad \text{where} \quad \Omega = \Omega_M + \Omega_R + \Omega_X \simeq \Omega_M + \Omega_X \tag{27.60}$$

is the total density in units of the critical density. In most of the older literature (before ~ 1995), much attention is paid to the dimensionless *deceleration parameter* of the universe, defined as

$$q_o \equiv \frac{-\ddot{a}_o/a_o}{H_o^2} = \frac{\Omega_M}{2} + \frac{P_X}{\rho_X}\Omega_X \ . \tag{27.61}$$

Here the second equality follows from the space-space component of the Einstein field equation, Eq. (27.27), translated into the language of our parameters, together with the fact that today the only significant pressure is that of exotic matter, P_X. We shall not use q_o in this book.

Remarkably, the values of our seven independent parameters H_o, Ω_B, Ω_M, T_{Ro}, Ω_R, Ω_X, and P_X/ρ_X are all fairly well known today (spring 2000), thanks largely to major observational progress in the past several years. In this section we shall discuss the observations that have been most effective in determining these parameters. For greater detail see, e.g., the review article by Turner (1999).

27.5.2 Local Lorentz frame of homogeneous observers near Earth

As a foundation for discussing some of the observations, we shall construct the local Lorentz frame of a homogeneous observer near Earth. (The Earth moves relative to this frame with a speed $v = 630 \pm 20$ km s^{-1}, as revealed by a dipole anisotropy in the temperature distribution of the CMB on the Earth's sky.)

Homogeneous observers can have local Lorentz frames because they moves freely (along timelike geodesics) through spacetime; cf. Ex. 27.1. For ease of analysis, we place the spatial origin of the Robertson-Walker $\{t, \chi, \theta, \phi\}$ coordinate system at the location of a near-earth homogeneous observer. Then that observer's Lorentz coordinates are

$$\hat{t} = t + \frac{1}{2}\chi^2 a\dot{a} \ , \quad \hat{x} \equiv a\chi \sin\theta\cos\phi \ , \quad \hat{y} \equiv a\chi\sin\theta\sin\phi \ , \quad \hat{z} \equiv a\chi\cos\theta \ ; \tag{27.62}$$

cf. Eq. (24.12) and associated discussion. Note that only at second-order in the distance away from the origin of the local Lorentz frame does the Lorentz time $\hat{t}$ differ from the Robertson-Walker time coordinate t. This second-order difference will never be important for anything we compute, so henceforth we will ignore it and set $\hat{t} = t$.

The near-earth local Lorentz frame, like any local Lorentz frame, must be kept spatially small compared to the radius of curvature of spacetime. That radius of curvature is related to the Lorentz-frame components of spacetime's Riemann curvature tensor by

$$\mathcal{R} \sim \frac{1}{|R_{\hat{\alpha}\hat{\beta}\hat{\gamma}\hat{\delta}}|^{1/2}} \ . \tag{27.63}$$

More precisely, since it is the largest components of Riemann that have the biggest physical effects, we should use the largest components of Riemann when evaluating $\mathcal{R}$. These largest components of Riemann today turn out to be $\sim \dot{a}_o^2/a_o^2 = H_o^2$ and $\sim k/a_o^2$. Observations discussed below reveal that $\Omega \sim 1$, so $k/a_o^2 \lesssim H_o^2$ [Eq. (27.60)]. Therefore, the universe's coarse-grain-averaged radius of spacetime curvature today is

$$\mathcal{R} \sim \frac{1}{H_o} \ ; \tag{27.64}$$

and the demand that the local Lorentz frame be small compared to this radius of curvature is equivalent to the demand that we confine the local-Lorentz spatial coordinates to the region

$$H_o r \ll 1 \ , \quad \text{where } r \equiv \sqrt{\hat{x}^2 + \hat{y}^2 + \hat{z}^2} = a\chi + \mathrm{O}(a\chi^3) \ . \tag{27.65}$$

(Below we shall neglect the tiny $a\chi^3$ correction.)

27.5.3 Hubble expansion rate

Consider a homogeneous observer near enough to earth to be in the near-earth local Lorentz frame, but not at its origin. Such an observer has fixed Robertson-Walker radius χ and local-Lorentz radius $r = a\chi$, and thus moves away from the origin of the local Lorentz frame with a velocity, as measured by the frame's rods and clocks, given by $v = dr/dt = \dot{a}\chi$; i.e., evaluating that velocity today, $v = \dot{a}_o\chi$. Correspondingly, special relativity insists that light emitted by this homogeneous observer at local Lorentz radius r and received today by the homogeneous observer at $r = 0$ should be Doppler shifted by an amount $\Delta\lambda/\lambda \cong v = \dot{a}_o\chi$. Note that this Doppler shift is proportional to the distance between the homogeneous observers, with the proportionality factor equal to the Hubble constant:

$$z \equiv \frac{\Delta\lambda}{\lambda} = v = H_o r \ . \tag{27.66}$$

This Doppler shift is actually nothing but the cosmological redshift, looked at from a new viewpoint: When specialized to emitters and receivers that are near each other, so they can be covered by a single local Lorentz frame, the cosmological redshift formula (27.44) reduces to

$$z = \frac{a_o}{a} - 1 = \frac{1}{a_o}(a_o - a) \cong \frac{1}{a_o}\dot{a}_o\Delta t = H_o\Delta t \ , \tag{27.67}$$

where Δt is the time required for the light to travel from emitter to receiver. Since the light travels at unit speed as measured in the local Lorentz frame, Δt is equal to the distance r between emitter and receiver, and the cosmological redshift becomes $z = H_o r$, in agreement with the Doppler shift (27.66).

To the extent that the galaxies which astronomers study are at rest with respect to homogeneous observers, they should exhibit the *distance-redshift relation* (27.66). In reality, because of the gravitational attractions of other, nearby galaxies, typical galaxies are not at rest relative to homogeneous observers, i.e., not at rest relative to *the "Hubble flow"*. Their velocities relative to local homogeneous observers are called *peculiar velocities* and have magnitudes that are typically $v_{\rm pec} \sim 300$ km/sec $\sim 10^{-3}$, and can be as large as $v_{\rm pec} \sim 1000$ km/sec. In order to extract the Hubble constant from measurements of galactic distances and redshifts, astronomers must determine and correct for these peculiar motions. That correction task is rather difficult when one is dealing with fairly nearby galaxies, say with $z \lesssim 0.003$ so $v \lesssim 1000$ km/sec. On the other hand, when one is dealing with more distant galaxies, the task of determining the distance r is difficult . As a result, the measurement of the Hubble constant has been a long, arduous task, involving hundreds of astronomers working for 2/3 of a century. Today this effort has finally paid off, with a number of somewhat independent measurements that give

$$H_o = (65 \pm 5)\,\mathrm{km\ sec^{-1}Mpc^{-1}} \ , \tag{27.68}$$

where the unit of velocity is km s $^{-1}$ and the unit of distance is 1 Mpc = 1 megaparsec = 10^6 pc. Converting into units of length and time (using $c = 1$), the inverse Hubble constant is

$$\frac{1}{H_o} = (4.6 \pm 0.4) \text{ Gpc} = (15.0 \pm 1.3) \times 10^9 \text{ years} . \tag{27.69}$$

Correspondingly, the critical density to close the universe, Eq. (27.36), is

$$\rho_{\rm crit} = (7.9 \pm 1.2) \times 10^{-30} \text{g/cm}^3 . \tag{27.70}$$

In the cosmology literature one often meets the "Hubble parameter," whose definition and measured value are

$$h \equiv \frac{H_o}{100 \text{km s}^{-1} \text{Mpc}^{-1}} = 0.65 \pm 0.05 . \tag{27.71}$$

27.5.4 Big-bang nucleosynthesis

When the universe was about a minute old and had temperature $T_R \sim 10^9$ K, nuclear burning converted free protons and neutrons into the light isotopes deuterium $\equiv {}^2$H, ^{3}He, ^{4}He, and ^{7}Li. Over the past four decades astronomers have worked hard to achieve precision measurements of the primordial abundances of these isotopes. Those measurements, when compared with nucleosynthesis calculations based on models for the universal expansion, produce remarkably good agreement—but only when (i) the number of species of neutrinos (which contribute to the radiation density and via the Einstein equation to the expansion rate during the burning) is no greater than three (electron, muon, and tau neutrinos); and (ii) exotic matter has negligible influence on the universe's expansion except in and near the modern era, and possibly before nucleosynthesis; and (iii) the normalized baryon density is

$$\Omega_B = (0.045 \pm 0.0057) \left(\frac{0.65}{h}\right)^2 . \tag{27.72}$$

Here 0.0057 is the 95 per cent confidence limit. This is a remarkably accurate measurement of the density of baryonic matter—and it shows that ρ_B is only about 5 per cent of the critical density.

27.5.5 Density of Cold Dark Matter

The only kind of matter that can condense, gravitationally, is that with pressure $P \ll \rho$, i.e., cold matter. The pressures of the universe's other constituents (radiation and exotic matter) prevent them from condensing significantly; they must be spread rather smoothly throughout the universe. The total density of cold matter, Ω_M, can be inferred from the gravitationally-measured masses M of large clusters of galaxies. Those masses are measured in three ways: (i) by applying the virial theorem to the motions of their individual galaxies, (ii) by applying the equation of hydrostatic equilibrium to the distributions of hot, X-ray emitting gas in the clusters, and (iii) by studying the influence of the clusters as gravitational lenses for more distant objects. The results of the three methods agree reasonably well and yield a total density of cold matter

$$\Omega_M = 0.4 \pm 0.1 \; ; \tag{27.73}$$

see Turner (1999) for details and references.

This $\Omega_M \simeq 0.4$ is much larger than the density of baryonic matter $\Omega_B \simeq 0.05$. Their difference,

$$\Omega_{\rm CDM} = \Omega_M - \Omega_B = 0.35 \pm 0.1 \tag{27.74}$$

is the density of cold, dark matter.

27.5.6 Radiation Temperature and Density

The temperature of the CMB has been measured, from its Planckian spectrum, to be

$$T_R = 2.728 \pm 0.002 \mathrm{K} \tag{27.75}$$

This temperature tells us with excellent accuracy the contribution of photons to the radiation density

$$\Omega_\gamma = (0.5870 \pm 0.005) \left(\frac{h}{0.65}\right)^2 \times 10^{-4} \; . \tag{27.76}$$

Those species of neutrinos with rest masses $m_\nu \ll 10^{-4}$ eV move with relativistic velocities and thus contribute to the radiation energy density. Statistical equilibrium in the early universe implies that the radiation density today for each such species is $(7/8)(4/11)^{4/2}\Omega_\gamma$. Gravitons could also have an energy density of order Ω_γ, but no larger as otherwise they would have produced an uncacceptably large influence on the expansion rate of the universe during big-bang nucleosynthesis and thereby would have distorted the nuclear abundances measureably. The same is true of other hypothetical forms of radiation. Therefore, the net radiation density must lie in the range

$$0.6 \times 10^{-4} \lesssim \Omega_R \lesssim 3 \times 10^{-4} \; . \tag{27.77}$$

27.5.7 Anisotropy of the CMB: Measurements of the Doppler Peaks

Consider an object with physical diameter D that resides at a distance r from Earth, and neglect the earth's motion and object's motion relative to homogeneous observers. Then the object's angular diameter Θ as observed from Earth will be $\Theta = D/r$, if $r \ll 1/H_o$ so the effects of spacetime curvature are negligible. For greater distances, $r \sim 1/H_o$, the ratio

$$r_{\rm AD} \equiv \frac{D}{\Theta} \tag{27.78}$$

(called the object's *angular-diameter distance*) will be strongly influenced by the spacetime curvature—and thence by the cosmological parameters H_o, Ω_M, Ω_X, P_X/ρ_X that influence the curvature significantly. In Ex. 27.8 formulas are derived for $r_{\rm AD}$ as a function of these parameters and the object's cosmological redshift z (a measure of its distance).

Astronomers searched for many decades for objects on the sky (*standard yardsticks*), whose physical diameter D could be known with high confidence. By combining the known D's with the yardsticks' measured angular diameters Θ to get their $r_{\rm AD} = \Theta/D$ and by measuring the redshifts z of their spectral lines, the astronomers hoped thereby to infer the cosmological parameters from the theoretical relation $r_{\rm AD}(z$, cosmological parameters).

This effort produced little of value in the era ∼1930 to ∼1990, when astronomers were focusing on familiar astronomical objects such as galaxies. No way could be found to determine, reliably, the physical diameter D of any such object.

Finally, in 1994, Marc Kamionkowski, David Spergel and Naoshi Sugiyama (1994) identified an object of a very different sort, whose physical diameter D could be known with high confidence: the *cosmological horizon* in the era when the primordial plasma was recombining and matter and radiation were decoupling from each other. This was the long-sought standard yardstick.

This cosmological horizon is *not* the same thing as the horizon of a black hole, but it is analogous. It is the distance between objects that are just barely able to communicate with each other via light signals. To discuss this concept quantitatively, it is useful to introduce a new time coordinate η for the Robertson-Walker line element (27.6)

$$\eta = \int \frac{dt}{a}\ ; \quad \text{so} \quad d\eta = \frac{dt}{da}\ . \tag{27.79}$$

Then the line element becomes

$$ds^2 = a^2[-d\eta^2 + d\chi^2 + \Sigma^2(d\theta^2 + \sin^2\theta d\phi^2)]\ . \tag{27.80}$$

By setting $\eta = 0$ at the beginning of the expansion and $\eta = \eta_{\rm rec}$ at the era of recombination, and noting that light travels in the χ direction with coordinate speed $d\chi/d\eta = 0$, we see that the diameter of the horizon at recombination is

$$D_{\rm rec} = \eta_{\rm rec}\, a_{\rm rec}\ , \quad \text{where} \quad \eta_{\rm rec} = \int_0^{t_{\rm rec}} \frac{dt}{a} \tag{27.81}$$

and $a_{\rm rec}$ is the value of a at recombination. Two objects separated by a distance greater than $D_{\rm rec}$ were unable to communicate with each other at recombination, because there had not been sufficient time since the birth of the universe for light to travel from one to the other. In this sense, they were outside each others' cosmological horizon. Objects with separations less than $D_{\rm rec}$ could communicate, at recombination; i.e., they were inside each others' cosmological horizon.

As the universe expands, the cosmological horizon expands; objects that are outside each others' horizons in the early universe come inside those horizons at some later time, and can then begin to communicate.

Kamionkowski, Spergel and Sugiyama realized that the universe provides us with markers on the sky that delineate the horizon diameter at recombination, $D_{\rm rec}$. These markers are anisotropies of the CMB, produced by the same density and temperature inhomogeneities as would later grow to form galaxies.

The inhomogeneities are known, observationally, to have been perturbations of the density with fixed, homogeneous entropy per baryon, i.e. with fixed T_R^3/ρ_M, and with amplitudes, as they came inside the horizon,

$$3\frac{\Delta T_R}{T_R} = \frac{\Delta\rho_M}{\rho_M} \sim 1 \times 10^{-4} . \tag{27.82}$$

We can resolve the perturbations $\Delta T_R/T_R$ at recombination into spatial Fourier components characterized by wave number k, or equivalently by reduced wavelength $\lambda\!\!\!^- = 1/k$. Observers on Earth find it more convenient to resolve the perturbations into spherical harmonics on the sky. Since order $\ell = 1$ corresponds to a perturbation with angular wavelength 360 degrees $= 2\pi$ radians, order ℓ must be a perturbation with angular wavelength $2\pi/\ell$ and thence angular reduced wavelength $\Theta = 1/\ell$. The ratio $\lambda\!\!\!^-/\Theta$ of physical reduced wavelength to angular reduced wavelength is the angular-diameter distance over which the CMB photons have traveled since recombination:

$$r_{\rm AD}^{\rm rec} = \frac{\lambda\!\!\!^-}{\Theta} = \frac{\ell}{k} . \tag{27.83}$$

Now, consider perturbations with spatial scale small enough that a reduced wavelength $\lambda\!\!\!^-$ came inside the horizon ("crossed the horizon") somewhat earlier than recombination. Before $\lambda\!\!\!^-$ crossed the horizon, each high temperature region was unaware of a neighboring low temperature region, so the two evolved independently, in such a way that their fractional temperature difference grew as

$$\frac{\Delta T_R}{T_R} \propto a \propto t^{2/3} \tag{27.84}$$

[Ex. 27.10]. When $\lambda\!\!\!^-$ crossed the horizon (i.e., when the horizon expanded to become larger than $\lambda\!\!\!^-$), the neighboring regions began to communicate. The high-T_R, high-ρ_M region pushed out against its low-T_R, low-ρ_M neighbor, generating sound waves; and correspondingly, the growth of $\Delta T_R/T_R$ changed into acoustic oscillations.

For perturbations with some specific physical size $\lambda\!\!\!^-_{1/4}$ (angular size $\Theta_{1/4}$, the acoustic oscillations had completed one quarter cyle at the time of recombination, so their temperature contrast was reduced, at recombination, to zero. For pertubations a little smaller, $\Theta_{1/2}$, the oscillations had completed a half cycle at recombination, so the hot regions and cold regions were reversed, and the temperature contrast was roughly as large as at horizon crossing. Perturbations still smaller, $\Theta_{3/4}$, had completed $3/4$ of a cycle at recombination, so their density contrast was momentarily zero. Perturbations smaller still, Θ_1, had completed a full cycle of oscillation at recombination and so had a large density contrast; and so forth.

The result is the pattern of temperature anisotropy as a function of Θ or equivalently $\ell = 1/\Theta$ shown in Fig. 27.6. The first peak in the pattern is for perturbations whose reduced wavelength $\lambda\!\!\!^-_0$ had only recently come inside the horizon at recombination, so $\lambda\!\!\!^-_0 = r_{\rm AD}\Theta_0$ is equal to the diameter $D_{\rm rec}$ of the horizon at recombination, aside from a small difference that can be computed with confidence. This is the standard yardstick that astronomers had sought for decades.

The basic structure of the pattern of anisotropy oscillations shown in Fig. 27.6 is in accord with the above description of acoustic oscillations, but the precise details are modestly

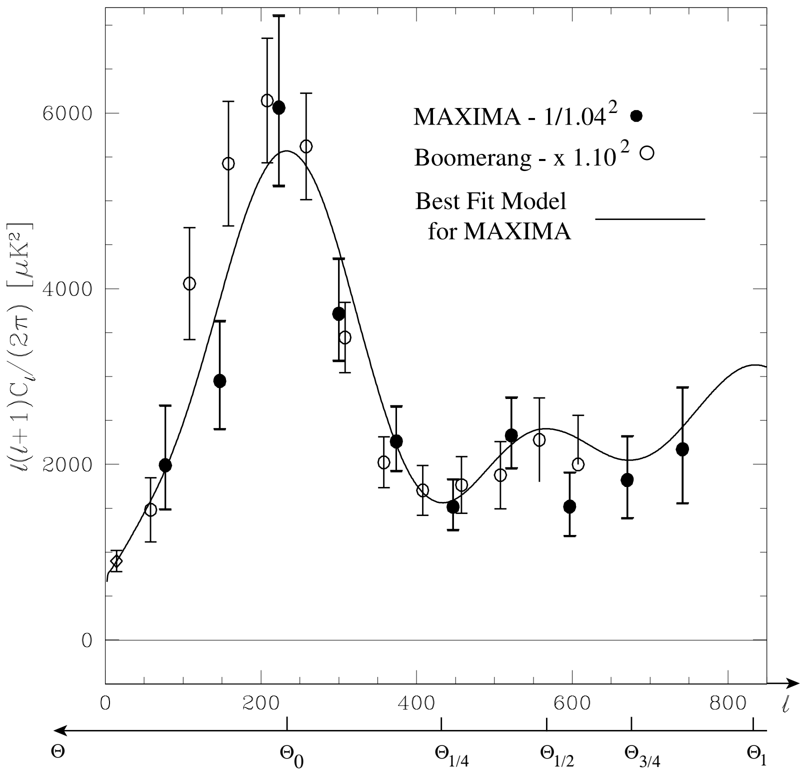

Fig. 27.6: Anisotropy of the CMB as measured by MAXIMA (Balbi et. al. 2000) and Boomerang (Lange et. al. 2000), compared with a theoretical model with total mass density $\Omega = 0.9$ (solid curve). Plotted vertically is the mean square temperature fluctuation; plotted horizontally is the fluctuation's spherical harmonic order ℓ (top scale) and angular scale Θ (bottom scale). The temperature scales for the MAXIMA and Boomerang data have been adjusted by the indicated amounts to produce agreement; the discrepancies were presumably due to errors in temperature calibration. This figure is adapted from Balbi et. al. (2000).

different because the initial distribution of inhomogeneities is statistical (i.e. is a random process in the sense of Chap. 5), and the physics of the oscillations is somewhat complex. Examples of the complexities are: (i) ΔT_R does not go to zero at the minima $\Theta_{1/3}$ and $\Theta_{3/4}$ because the emitting matter has acquired inhomogeneous velocities relative to Earth by falling into the oscillations' gravitational potential wells, and these velocities produce doppler shifts that smear out the minima. (ii) This same infall makes the density and temperature contrasts smaller at the half-cycle point $\Theta_{1/2}$ than at the full-cycle points Θ_0 and Θ_1.

Despite these complexities and statistical effects, the shapes of the acoustic oscillations can be computed with high confidence, once one has chosen values for the cosmological parameters. The reason for the confidence is that the amplitude of the oscillations is very small, so nonlinear effects are negligible. The pattern of the temperature oscillations at recombination is computed as a function of physical length λ, with results that depend modestly on some of the cosmological parameters; and then the physical pattern is converted into a pattern as seen on Earth's sky, using the angular-diameter distance $r_{AD} = \lambda/\Theta$ that the CMB photons have traveled since recombination, which depends very strongly on the cosmological parameters.

Remarkably, the positions Θ_0, $\Theta_{1/2}$, Θ_1, ... of the oscillation peaks (called *doppler peaks* for no good reason[2]), depend more strongly on the total density $\Omega \simeq \Omega_M + \Omega_X$ than on other parameters. By comparison of the theoretical models with measurements reported in May 2000 from the MAXIMA Project (Balbi et. al. 2000), and the Boomerang Project (Lange et. al. 2000), the experimenters infer

$$\begin{aligned} \Omega &= 0.90 \pm 0.15 \quad \text{from MAXIMA}, \\ \Omega &= 1.1 \pm 0.2 \quad \text{from Boomerang}; \end{aligned} \tag{27.85}$$

see Fig. 27.6. The errors are 2σ (95% confidence), and the two experiments obviously agree well with each other. We shall summarize their results as

$$\Omega = 1.0 \pm 0.2\ . \tag{27.86}$$

This near-unity value of Ω implies that the universe is close to being spatially flat; see Eq. (27.60).

The Boomerang and MAXIMA measurements also reveal that — aside from the oscillations — the spectral density of the temperature perturbations decays as $G_{T_R}(k) \propto k^{-1.0\pm0.1}$, so the rms amplitude of the fluctuations, $\Delta T_R^{\rm rms} = \sqrt{(k/2\pi)G_{T_R}(k)}$ [Eq. (5.64)] is independent of wave number, i.e. independent of ℓ. This is in accord with predictions from "inflationary" models for the production of the perturbations (Sec. 27.7 below.)

27.5.8 Age of the universe: Constraint on the exotic matter

The total mass density $\Omega = 1.0 \pm 0.2$ from CMB anisotropy and the cold-matter mass density $\Omega_M = 0.4 \pm 0.1$ leave a *missing mass density* $\Omega_X = 0.6 \pm 0.2$, which must be in some exotic form that does not condense, gravitationally, along with the cold matter, and that therefore must have a pressure $|P_X| \sim \rho_X$. This exotic matter must have had a negligible density at the time of recombination and at the time of nucleosynthesis; otherwise, it would have disturbed the shapes of the doppler peaks and distorted the nuclear abundances. In order that it be significant now and small earlier, compared to cold matter, it must have a negative pressure $P_X < 0$. One handle on how negative comes from the age of the universe.

Assuming, that $P_X/\rho_X \equiv w_X$ was constant or approximately so during most of the age of the universe (i.e., back to redshift $z \sim 10$) and that the exotic matter did not exchange significant energy with other constituents of the universe during that recent epoch, then the first law of thermodynamics implies that $\rho_X = \rho_{\rm crit}\Omega_X(a_o/a)^{3(1+w_X)}$. Inserting this,

[2]Yakov B. Zel'dovich and Rashid A. Sunyaev (1970), who first predicted the existence of these peaks, later gave them a name that has a little more justification: They called them *Sakharov oscillations* because Andrei D. Sakharov (1965) was the first to predict the sound waves that give rise to the peaks. Zel'dovich and Sunyaev introduced this name at a time when their close friend Andrei Sakharov was being attacked by the Soviet government; they hoped that this would call attention to Sakharov's international eminence and help protect him. It seemed not to help.

$\rho_M = \rho_{\rm crit}\Omega_M(a_o/a)^3$, and $\rho = \rho_M + \rho_X$ into the Einstein equation (27.26), solving for $dt = da/\dot a$, and integrating, we obtain for the product of the current age of the universe t_o and the Hubble expansion rate H_o:

$$H_o t_o = \int_0^1 \frac{dv}{\sqrt{1 - \Omega_M - \Omega_X + \frac{1}{v}\left(\Omega_M + \Omega_X v^{-3w_X}\right)}} \; . \tag{27.87}$$

The more negative is $P_X/\rho_X = w_X$, the larger is the integral, and thus the larger is $H_o t_o$.

By comparing the observed properties of the oldest stars in our galaxy with the theory of stellar evolution, and estimating the age of the universe at galaxy formation, astronomers arrive at an estimate

$$t_o = (14 \pm 1.5) \times 10^9 \mathrm{yr} \tag{27.88}$$

for the age of the universe. Combining this with the measured value of the Hubble constant, Eq. (27.69), we obtain

$$H_o t_o = 0.93 \pm 0.13 \; . \tag{27.89}$$

For comparison, Eq. (27.87) with $\Omega_M = 0.4$, $\Omega_X = 0.6$ (in accord with the measurements) gives $H_o t_o = 0.89$ for $w_X \equiv P_X/\rho_X = -1$ (vacuum), 0.82 for $w_X = -\frac{1}{2}$, 0.67 for $w_X = 0$, and even smaller values for positive w_X. Thus, *the measured ages of the oldest stars imply that*

$$w_X \equiv \frac{P_X}{\rho_X} \lesssim -\frac{1}{2} \; ; \tag{27.90}$$

and those ages are nicely compatible with the exotic matter being vacuum stress-energy, $w_X = -1$.

27.5.9 Magnitude-Redshift relation for type Ia supernovae: Confirmation that the universe is decelerating

The constraint $P_X/\rho_X < -1/2$ on the exotic matter has a profound consequence for the expansion rate of the universe. In the "particle-in-a-potential" analysis [Eqs. (27.40), (27.41) and Fig. 27.3] the contribution of the exotic matter to the potential is $V_X(a) = -(4\pi/3)a^2\rho_X \propto a^n$ where $n > 1/2$, which grows stronger with increasing a. Correspondingly, in the present era, the "potential energy" is becoming more negative, which means that the universe's "kinetic energy" $\frac{1}{2}\dot a^2$ must be increasing: the universe has recently made the transition from a decelerating expansion to an *accelerating expansion.*

In 1998 two independent groups of astronomers reported a direct observational confirmation of this acceleration (Riess et. al. 1998, Perlmutter et. al. 1999). This confirmation is based on systematic observations of the apparent brightness of *type Ia supernovae* as a function of the supernovae's redshift-measured distances. If the universal expansion is, indeed, accelerating, then distant objects (including supernovae), which we see when the universe was much younger than today, would have experienced a slower universal expansion than we experience today, so their observed redshifts z should be lower than in a universe with constant or decelerating expansion rate. These lowered redshifts should show up as a leftward displacement of the supernovae's locations in a diagram plotting the supernovae's

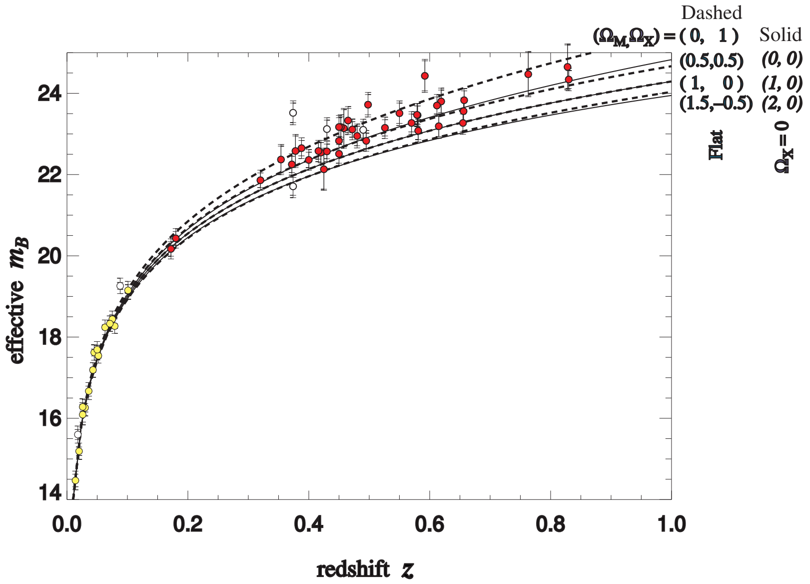

Fig. 27.7: Magnitude-redshift diagram for type Ia supernovae based on observations by Perlmutter et. al. (1998) and others. [Adapted from Perlmutter et. al. (1998).]

redshift horizontally and their brightness (a measure of their distance) vertically—a so-called *magnitude-redshift diagram.*

Such a diagram is shown in Fig. 27.7. The measure of brightness used in this diagram is the supernova's *apparent magnitude*

$$m \equiv -2.5 \log_{10}(\mathcal{F}/2.5 \times 10^{-8}\,\mathrm{WM}^{-2}) \ , \tag{27.91}$$

where $\mathcal{F}$ is the flux received at Earth. The sign is chosen so that the dimmer the supernova, the larger the magnitude. A series of theoretical curves is plotted in the diagram, based on assumed values for Ω_M and Ω_X, and on the presumption that the exotic matter is vacuum stress-energy so $P_X/\rho_X = -1$. The formulae for these curves are derived in Ex. 27.9. The solid curves are for *no* exotic matter, $\Omega_X = 0$. The exotic matter, which converts the universal deceleration $\ddot{a} < 0$ into acceleration $\ddot{a} > 0$, pushes the curves leftward for distant supernovae (upper right-hand region), as described above. The dashed curves are for a mixture of exotic matter and cold matter that sums to the critical density, $\Omega_X + \Omega_M = 1$.

A detailed analysis of the data by Perlmutter et. al. (1999) gives (assuming $P_X/\rho_X = -1$)

$$\Omega_X = \frac{1}{3}(4\Omega_M + 1) \pm \frac{1}{6} \ . \tag{27.92}$$

Combining with $\Omega_M = 0.4 \pm 0.1$, this implies

$$\Omega_X = 0.85 \pm 0.2 \ . \tag{27.93}$$

To recapitulate: A variety of observations all point in the same direction. They agree that our universe is close to spatially flat, with $\Omega_X \simeq 0.6$, $\Omega_M \sim 0.4$, and $\Omega_R \sim 10^{-4}$.

****

EXERCISES

Exercise 27.8 *Example: Angular-Diameter Distance*
Consider an electromagnetic emitter at rest relative in the cosmological fluid, i.e. at rest relative to homogeneous observers, and let the emitter's radiation be observed at Earth. Neglect the Earth's motion relative to homogeneous observers. Let the cosmological redshift of the emitted radiation be $z = \Delta\lambda/\lambda$.

(a) Show that the emitter's angular-diameter distance is

$$r_{\rm AD} = \frac{R}{1+z} , \tag{27.94}$$

where

$$R = a_o \Sigma(\Delta\chi) . \tag{27.95}$$

Here a_o is the universe's expansion factor today, Σ is the function defined in Eq. (27.8), and $\Delta\chi$ is the coordinate distance that light must travel in going from emitter to Earth, if its path has constant θ and ϕ. [*Hint:* Place the Earth at $\chi = 0$ of the Robertson-Walker coordinate system, and the emitter at $\chi = \Delta\chi$, and use the line element (27.80).]

(b) Assuming that the exotic matter is vacuum stress-energy so $P_X = -\rho_X$, show that the quantity $\Delta\chi$ appearing in Eq. (27.95) is

$$\Delta\chi = |1 - \Omega_M - \Omega_X|^{1/2} \int_1^{1+z} \frac{du}{\sqrt{\Omega_M u^3 + (1 - \Omega_M - \Omega_X)u^2 + \Omega_X}} \tag{27.96}$$

[*Hint:* Use Eqs. (27.79) and (27.80) to deduce that

$$\Delta\chi = \int_{t_e}^{t_o} \frac{dt}{a} = \int_1^{1+z} \frac{a}{a_o}\frac{dt}{da} d\frac{a_o}{a} \tag{27.97}$$

and use the Einstein equation for da/dt.]

(c) Plot $r_{\rm AD}(z)$ for several representative values of Ω_M and Ω_X.

Exercise 27.9 *Example: Magnitude-Redshift Relation*
Consider a supernova which emits a total luminosity L as measured in its own local Lorentz frame in an epoch when the expansion factor is a and the cosmological redshift is $z = a_o/a - 1$. Assume that the supernova and the Earth are both at rest relative to homogeneous observers at their locations. Place the origin $\chi = 0$ of a Robertson-Walker coordinate system at the supernova's location, orient the coordinate axes so the earth lies at $\theta = \pi/2$ and $\phi = 0$, and denote by $\Delta\chi$ the earth's radial coordinate location. Show that the flux of energy received from the supernova at earth today is given by

$$\mathcal{F} = \frac{L}{4\pi R^2 (1+z)^2} , \tag{27.98}$$

where R is the same function as appears in the angular-diameter distance, Eqs. (27.95), (27.96).

Exercise 27.10 *Challenge: Growth of perturbations before they cross the horizon*
Show that in a matter-dominated universe, $\rho_M \gg \rho_R$ and $\rho_M \gg \rho_X$, the fractional density difference of two neighboring regions that are outside each others' cosmological horizons grows as $\Delta\rho/\rho \propto a \propto t^{2/3}$. [Hint: the difference of spatial curvature between the two regions is of importance. For a solution, see, e.g., pp. 108–110 of Peebles (1993).]

****

27.6 The Big-Bang Singularity, Quantum Gravity, and the Initial Conditions of the Universe

Although we do not know for sure the correct equation of state at redshifts $z \gg 3 \times 10^{12}$, where the thermal energy of each baryon exceeds its rest mass-energy, the "particle-in-a-potential" form (27.40) of the evolution equation tells us that, so long as $dV/da \geq 0$ at small a, the universe must have begun its expansion in a state of vanishing expansion factor $a = 0$, nonzero $\dot{a}$, and infinite $\dot{a}/a$. Since some of the components of the Riemann curvature tensor in the local Lorentz frame of a homogeneous observer are of order $\dot{a}^2/a^2$, this means the expansion began in a state of infinite spacetime curvature, i.e., infinite tidal gravity, i.e., in a *"big-bang" singularity*. From the form $V = -(4\pi/3)a^2\rho$ for the effective potential [Eq. (27.40)] and the first law of thermodynamics (27.29), we see that the sufficient condition $dV/da \geq 0$ for the universe to have begun with a singularity is

$$\rho + 3P > 0 \, . \tag{27.99}$$

Cold matter and radiation satisfy this condition, but exotic matter violates it. As we have seen, exotic matter seems to have become important in our universe only recently, so it might not have been important at the universe's beginning. In this section we shall assume that it was not, and that the energy condition $\rho + 3P > 0$ was satisfied in the early universe. In the next section we shall discuss some consequences of a possible early-universe violation of $\rho + 3P > 0$.

The conclusion that the universe, if homogeneous and isotropic (and if $\rho + 3P > 0$), must have begun in a big-bang singularity, drove Yevgeny Lifshitz and Isaak Khalatnikov, students of Lev Landau in Moscow, to begin pondering in the late 1930s the issue of whether deviations from homogeneity and isotropy might have permitted the universe to avoid the singularity. A few events (the imprisonment of Landau for a year during Stalin's purges, then World War II, then the effort to rebuild Moscow, a nuclear weapons race with the United States, and other more urgent physics research) intervened, preventing the Lifshitz-Khalatnikov studies from reaching fruition until the early 1960s. However, after a great push in 1959–1961, Lifshitz and Khalatnikov reached the preliminary conclusion that early anisotropy and inhomogeneity could have saved the universe from an initial singularity: Perhaps the universe contracted from an early, large-scale state, then rebounded at finite size and finite curvature as a result of inhomogeneities and anisotropies. For a pedagogical presentation of the analysis which produced this conclusion see Landau and Lifshitz (1962).

The Lifshitz-Khalatnikov analysis was based on the mathematics of tensor analysis (differential geometry). In 1964 Roger Penrose (1965), a young faculty member at Kings College in London, introduced into general relativity an entire new body of mathematical techniques, those of differential topology, and used them to prove a remarkable theorem: that no matter how inhomogeneous and anisotropic an imploding star may be, if it implodes so far as to form a horizon, then it necessarily will produce a singularity of infinite spacetime curvature inside that horizon. Stephen Hawking and George Ellis (1968), at first graduate students and then research fellows at Cambridge University, by picking up Penrose's techniques and applying them to the early universe, proved that Lifshitz and Khalatnikov had to be wrong: The presently observed state of the universe plus reasonable constraints on its early equation of state imply that, regardless of any inhomogeneity or anisotropy, there must have been a singularity of infinite curvature. In response to these differential-topology analyses, Lifshitz, Khalatnikov, and their student Vladimir Belinsky reexamined their differential-geometry analyses, found an error, and discovered a possible structure for generic spacetime singularities. In this so-called *mixmaster structure*, as a freely moving observer approaches the singularity, inhomogeneities and anisotropies drive the tidal gravity (spacetime curvature) to oscillate in such a way that the observer feels an infinite, chaotic sequence of oscillations with ever growing amplitude, ever shortening period, and finite total proper-time duration. This is an example of the chaotic behavior which occurs frequently in nonlinear physics.

John Archibald Wheeler, a professor at Princeton University, realized in the mid 1950s that the singularities which began the big bang and terminate the implosion of a star cannot be classical: as one nears them, one must encounter a breakdown in classical general relativity, and new physics governed by the laws of *quantum gravity* (Wheeler 1957). Wheeler devised a simple argument to show that this is so, and to determine the critical radius of curvature of spacetime at which the transition to quantum gravity occurs:

Quantum theory insists that every field possess a half quantum of flucutational zero-point energy in each of its modes. Moreover, if one wishes to measure the average value of the field in a spacetime region with 4-volume L^4 (a region with side L along each of its 4 dimensions), one's measurements will be sensitive to the zero-point fluctuations of the modes that have wavelengths $\sim L$, but not to any others.

Now, so long as gravity is weak over the scale L, one can introduce a nearly Lorentz frame in the region L^4 and regard the deviations $\delta g_{\mu\nu} \equiv g_{\mu\nu} - \eta_{\mu\nu}$ of the metric coefficients $g_{\mu\nu}$ from the flat metric $\eta_{\mu\nu}$ as a nearly linear field that lives in nearly flat spacetime. This field must be just as subject to the laws of quantum mechanics as any other field. Its gravitational-wave modes with wavelength L have an energy density of order the square of the gradient of $\delta g_{\mu\nu}$, i.e., $\sim (\delta g_{\mu\nu}/L)^2$, and thus for these modes to contain a half quantum of unpredictable, fluctuational energy, they must have unpredictable fluctuations of the metric given by

$$\left(\frac{\delta g_{\mu\nu}}{L}\right)^2 L^3 \sim \frac{\hbar}{2L}\,. \tag{27.100}$$

Here the first term is the fluctuational energy density, L^3 is the 3-dimensional volume of the mode, and $\hbar/2L$ is its total fluctuational energy. Correspondingly, the mode's metric fluctuations are

$$\delta g_{\mu\nu} \sim \frac{\sqrt{\hbar}}{L}\,. \tag{27.101}$$

These fluctuations (which we have evaluated in the closest thing there is to a local Lorentz frame in our region L^4) guarantee that, whenever we try to measure a length L, we will make unavoidable errors with magnitude

$$\frac{\delta L}{L} \sim \delta g_{\mu\nu} \sim \frac{\sqrt{\hbar}}{L} . \tag{27.102}$$

The smaller is L, the larger are these fractional errors. When L is made smaller than $\sqrt{\hbar}$, the fractional errors exceed unity, there is no hope of measuring L at all (and our analysis also breaks down because we cannot introduce a nearly Lorentz frame throughout the region L^4). Thus, for a lengthscale L to be measureable, it must lie in the regime

$$L \gtrsim L_{PW} \ , \quad \text{where } L_{PW} \equiv \sqrt{\hbar} = \left(\frac{G\hbar}{c^3}\right)^{1/2} = 1.616 \times 10^{-33} \text{ cm} . \tag{27.103}$$

The critical lengthscale L_{PW} is called the Planck-Wheeler length. It is the shortest length that can possibly be measured with any accuracy. Thus, it is the smallest length that can be subjected to the classical laws of physics. Since gravity is characterized, classically, by the geometry of spacetime, classical gravity (i.e., general relativity) must break down on lengthscales shorter than L_{PW}. This should be true in the small in ordinary, nearly flat spacetime; and it also should be true near singularities: Near a singularity, when the radius of curvature of spacetime as predicted by classical general relativity becomes shorter than L_{PW}, general relativity must break down and be replaced by the correct quantum theory of gravity. And when quantum gravity comes into play, it may very well convert the singularity into something nonsingular.

Thus, to understand the true outcome of the gravitational implosion of a star, deep inside the horizon, one must understand quantum gravity; and to understand the initial conditions of the universe, one must understand quantum gravity.

The attempt to construct a quantum theory of gravity which unifies gravity with the strong, electromagnetic, and weak forces in an elegant and mutually consistent way is one of the "holy grails" of current theoretical physics research.

27.7 Inflationary Cosmology

If $\rho + 3P > 0$, then the universe is guaranteed to have cosmological horizons of the sort that we met when discussing acoustic oscillations in the era of recombination (Sec. 27.5.7).

The background radiation received at earth today last interacted with matter (at a red-shift $z \sim 10^3$) quite near our cosmological horizon. Two observers at the locations of that last interaction, one on our north celestial pole (i.e., directly above the north pole of the earth) and the other on our south celestial pole (i.e., directly above the south pole of the earth), are today far outside each others' cosmological horizons; and at the moment of that last interaction, they were enormously far outside each others' horizons. It is a great mystery how two regions of the universe, so far outside each others' horizons (i.e. with no possibility for causal contact since the big bang) could have the same temperatures at the time of that last interaction, to within the measured accuracy of $\Delta T/T \sim 10^{-4}$.

One solution to this mystery is to assume that the universe emerged from the *Planck-Wheeler era* of quantum gravity in a very nearly homogeneous and isotropic state (but one with enough inhomogeneities to seed galaxy formation). This "solution" leaves to a future theory of quantum gravity the task of explaining why this state was nearly homogeneous and isotropic. An alternative solution, proposed by Alan Guth (1981), then a postdoctoral fellow at Stanford University, is *inflation*.

Suppose, Guth suggests, that the universe emerged from the Planck-Wheeler, quantum-gravity era, with its fields in a vacuum state for which $\mathbf{T}_{\rm vac} = -\rho_{\rm vac}\mathbf{g}$ was nonzero and perhaps was even as large in magnitude as $\rho_{\rm vac} \sim \hbar^{-2} \sim 10^{93}{\rm g/cm}^3$.

The expansion factor a presumably will have been of order L_{PW} when the universe emerged from the Planck-Wheeler era; and the evolution equation (27.30) predicts that it subsequently will expand classically in accord with the law

$$a = L_{PW} \exp\left[\left(\frac{4\pi}{3}\rho_\Lambda\right)^{1/2} t\right] = L_{PW} \exp\left(\frac{t}{\mu L_{PW}}\right) , \tag{27.104}$$

where μ is a dimensionless constant that might be of order unity. This exponential expansion under the action of vacuum stress-energy is called "inflation;" and if it lasted long enough, that means our entire universe was so small in the early stages of inflation that it could easily communicate with itself, producing homogeneity and isotropy.

Of course, inflation at this enormous rate could not have lasted forever; it surely is not continuing today. If it occurred at all, it must have turned off at some point as a result of the fields undergoing a phase transition from the original vacuum state (sometimes called the "false vacuum") to a new vacuum state in which $\rho_{\rm vac}$ is zero, or perhaps equal to the tiny ρ_X that we observe today.

Although these ideas seem speculative, they have been made quite plausible by two factors: (*i*) they fit naturally into present ideas about the physics of the grand unification of all forces; and (*ii*) they successfully explain a number of mysterious features of the universe in which we live, including the high degree of isotropy of the background radiation (Ex. 27.11), and the flat (wavelength-independent) spectrum of rms density fluctuations that ultimately condensed into galaxies. For details see, e.g., Kolb and Turner (1994).

****

EXERCISES

Exercise 27.11 *Practice: Inflationary explanation for the isotropy of the cosmological background radiation*
Consider an inflationary cosmological model in which (i) the expansion factor inflates as $a = L_{PW} \exp(t/\mu L_{PW})$ until it has e-folded $N \gg 1$ times, i.e., from time $t = 0$ (when it emerges from the Planck-Wheeler era) to time $t = N\mu L_{PW}$, and then (ii) a phase transition drives a into the standard expansion produced by radiation with $P = \rho/3$: an expansion with $a \propto t^{1/2}$ [Eq. (27.46)]. Show that in this model, if the number of e-folding times during inflation is $N \gg 70$, then the north-celestial-pole and south-celestial-pole regions which emit the background radiation we see today are actually inside each others' cosmological horizons: They were able to communicate with each other (and thereby, the inflationary scenario suggests, were able to homogenize to the same temperature) during the inflationary era. *Hint*: the number of e-foldings required is given analytically by

$$N \gg \ln\left[\frac{{H_o}^{-1}}{L_{PW}}\left(\frac{\rho_o}{\Lambda}\right)^{1/4}\right] \simeq 70 \,. \tag{27.105}$$

****

Bibliography

Balbi, A., et. al., 2000. "Constraints on cosmological parameters from MAXIMA-1," *Astrophysical Journal Letters*, submitted. astro-ph/0005124.

Dicke, Robert H., Peebles, P. James E., Roll, Peter G., and Wilkinson, David T., 1965. "Cosmic-black-body radiation," *Astrophysical Journal*, **142**, 414–419.

Einstein, Albert, 1917. "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie," *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, **1917 volume**, 142–152.

Einstein, Albert, 1931. "Zum kosmologischen Problem der allgemeinen Relativitätstheorie," *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, phys.-math. Kl.*, **1931 volume**, 235–237.

Friedman, Alexander Alexandrovichi, 1922. "Über die Krümmung des Raumes," *Zeitschrift für Physik*, **10**, 377–386.

Guth, Alan H., 1981. "Inflationary universe: A possible solution to the horizon and flatness problems," *Physical Review*, **23**, 347–356.

Hawking, Stephen W. and Ellis, George F. R., 1968. "The cosmic black body radiation and the existence of singularities in our universe," *Astrophysical Journal*, **152**25–36.

Hubble, Edwin Powell, 1929. "A relation between distance and radial velocity among extragalactic nebulae," *Proceedings of the National Academy of Sciences*, **15**, 169–173.

Kolb, Edward W., Turner, Michael S., 1994. *The Early Universe*, Reading: Addison-Wesley.

Landau, Lev Davidovich, and Lifshitz, Yevgeny Michailovich, 1962. *The Classical Theory of Fields*, Addison Wesley, Reading, MA.

Lange, Andrew E. et. al., 2000. "First estimations of cosmological parameters from Boomerang," *Physical Review Letters*, submitted. astro-ph/000504.

Penrose, Roger, 1965. *Gravitational collapse and space-time singularities*, *Physical Review Letters*, **14**, 57–59.

Penzias, Arno A., and Wilson, Robert W, 1965. "A measurement of excess antenna temperature at 4080 Mc/s," *Astrophysical Journal*, **142**, 419–421.

Peebles, P. J. E., 1993. *Principles of Physical Cosmology*, Princeton: Princeton University Press.

Perlmutter, S. et. al., "Measurements of Ω and Λ from 42 high-redshift supernovae," *Astrophysical Journal*, **517**, 565–586 (1999).

Riess, A. G. et. al., 1998. "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *Astronomical Journal*, **116**, 1009.

Ryan, Michael, and Shepley, Lawrencei, 1975. *Homogeneous, Relativistic Cosmology*, Princeton University Press, Princeton, NJ.

Robertson, Howard Percy, 1935. "Kinematics and World Structure," *Astrophysical Journal*, **82**, 248–301; and **83**, 287–201 & 257–271.

Sakharov, Andrei D., 1965. *Zhurnal Eksperimentalnoi i Teoreticheskii Fisika*, **49**, 345.

Sunyaev, Rashid A., and Zel'dovich, Yakov B., 1970. "Small-scale fluctuations of relic radiation," *Astrophysics and Space Science*, **7**, 3–19.

Turner, Michael S., 1999. in *Proceedings of Particle Physics and the Universe (Cosmo-98)*, ed. D. O. Caldwell, AIP, Woodbury, NY; astro-ph/9904051.

Walker, Arthur Geoffrey, 1936. "On Milne's theory of world-structure," *Proceedings of the London Mathematical Society*, **42**, 90–127.

Weinberg, S., 1972. *Gravitation and Cosmology*, New York: John Wiley.

Wheeler, John Archibald, 1957, "On the nature of quantum geometrodynamics," *Annals of Physics*, **2**, 604–614.

Zel'dovich, Yakov Borisovich, 1968. "The cosmological constant and the theory of elementary particles," *Soviet Physics—Uspekhi*, **11**, 381–393.

Zel'dovich, Yakov Borisovich, and Novikov, Igor Dmitrivich, 1983. *Relativistic Astrophysics Volume 2: The Structure and Evolution of the Universe*, University of Chicago Press, Chicago.