CHAPTER 14

Expanding Universe

Edmund Bertschinger & Edwin F. Taylor *

Nothing expands the mind like the expanding universe.

—Richard Dawkins

14.1 ■ DESCRIBING THE UNIVERSE AS A WHOLE

Finding words that correctly describe the unbounded

What is a one-sentence summary of our Universe? Try this:

One-sentence description of our Universe

> **Our visible Universe** consists of hundreds of billions of galaxies, each containing roughly one hundred billion stars, scattered more or less uniformly through a volume about 28 billion light years across.

A one-sentence description of *anything* is bound to be inadequate as a predictor of observed details; this and the following chapter expand(!) and correct this one-sentence description.

Figure 1 shows a small example of our visible Universe, which illustrates our assertion that galaxies are "scattered more or less uniformly." If so, this radically simplifies our model of the Universe: We describe the part we can see, and—in the absence of evidence to the contrary—assume the place we live is not unique but the same as any other location in the Universe. As a first—and it turns out, accurate—approximation, we look for metrics that describe curvature caused by a uniform distribution of mass. Make no assumption about how far this distribution extends. Instead, first, examine all possibilities consistent with general relativity; second, compute their predictions; third, let astronomical observations select the "correct" model or models.

Assume a uniform Universe and that our location is not unique.

Restrict attention to metrics that are uniform in space? Why not also uniform in time—a Universe that remains unchanged as the eons roll? In the absence of evidence to the contrary this would be the simplest hypothesis.

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*
Copyright © 2017 Edmund Bertschinger, Edwin F. Taylor, & John Archibald Wheeler. All rights reserved. This draft may be duplicated for personal and class use.

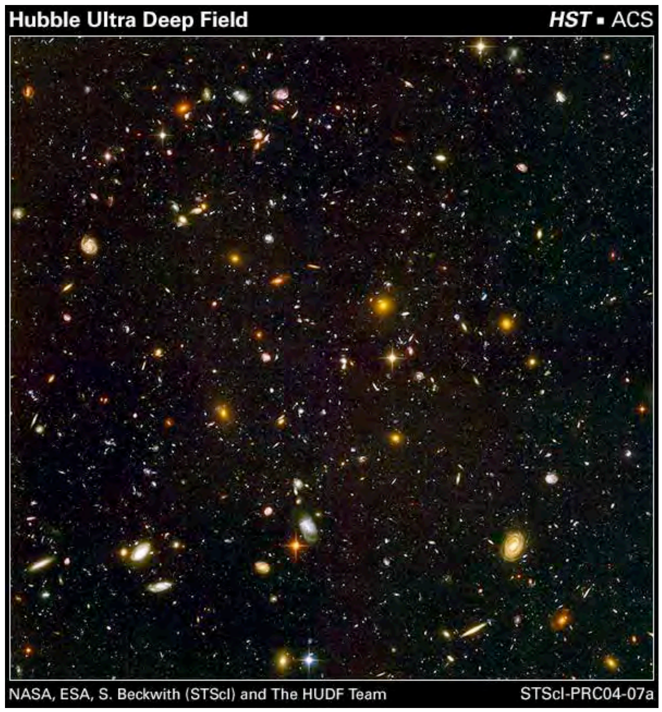


FIGURE 1 "Ultra deep field" image from the Hubble Space Telescope, named after astronomer Edwin P. Hubble. Every dot and every smear in this image is a galaxy, with the exception of a few nearby stars in our local galaxy. (Can you distinguish these exceptions?)

Cosmological constant comes, goes . . . then comes back again!

Indeed, in his 1917 cosmological model inspired by general relativity, Einstein looked for metrics that described a static Universe filled smoothly with mass. He found that no static metric was compatible with his newly-invented field equations unless he introduced a new term into those equations, a term that he called the **cosmological constant** and denoted by the Greek capital letter lambda, Λ. Later, after acknowledging Hubble's discovery that galaxies are flying away from one another, Einstein regretted the addition of Λ to his field equations. Astonishingly, today we know that there is something very similar, if not identical, to Λ at work in the Universe, as described in Chapter 15, Cosmology.

Brief history of the Universe

We know far more about the Universe than Einstein did a century ago. We know that the Universe is not static, but evolving. We know that approximately 14 billion years ago all matter/energy was concentrated in a much smaller structure. We know that this concentration expanded and thinned, from a moment we call the Big Bang, with galaxies forming during the initial expansion.

Box 1. Is this the only Universe?

Are there multiple universes, parallel universes, or baby universes? General relativity theorists write about all these and more. In this book we investigate the simplest model Universe consistent with observations—a single simply-connected spacetime.

Cosmologists often distinguish between "the observable universe" and all that there is or might be, citing plausible arguments that spacetime could be very different trillions of light years away. Here we restrict discussion to the simplest generalization of the observable universe, one—*the*—Universe that is everywhere similar to what we see in our vicinity.

Wait. Isn't science supposed to tell us what exists? Not at all! Science struggles to create theories that we can verify—or disprove—with observation and measurement.

How do we know?

How do we know these things? And how do we describe an evolving, expanding Universe? The present chapter assembles tools for this description, beginning with the metric of a spatially uniform, static Universe, then generalizes the metric to include general features of development with the t-coordinate. However, a detailed prediction of t-development requires a knowledge of the constituents of the Universe. Chapter 15, Cosmology provides this, then applies the tools assembled in the present chapter to analyze the past and predict alternative futures for our Universe.

14.2 ■ SPACE METRICS FOR A STATIC UNIVERSE

Describing a uniform space

Space metric for uniform space curvature

A Universe filled uniformly with mass and energy has—on average—*uniform space curvature* everywhere. In this book we deal mainly with two space dimensions plus a global t-coordinate. In one popular global map coordinate system, the most general constant-curvature *space* metric has the following form on the r, ϕ plane:

$$ds^2 = \frac{dr^2}{1 - Kr^2} + r^2 d\phi^2 \tag{1}$$

Flat, closed, and open spaces

The value of the parameter K determines the shape of the space, which in turn determines the range of r:

$$\text{for } K = 0, \qquad 0 \leq r < \infty \qquad \text{(Case I: flat space)} \tag{2}$$

$$\text{for } K > 0, \qquad 0 \leq r \leq \frac{1}{K^{1/2}} \qquad \text{(Case II: closed space)} \tag{3}$$

$$\text{for } K < 0, \qquad 0 \leq r < \infty \qquad \text{(Case III: open space)} \tag{4}$$

Flat plane, sphere, and saddle

Preview: We easily visualize Case I, flat space—equation (2). Next we visualize Case II, closed space, as a sphere—equation (3) and Figure 2. Finally Case III, open space has the shape of a saddle—equation (4) and Figure 3.

To describe the expansion of the Universe, it is helpful to separate its scale or size, symbolized by a **scale factor** R, from its curvature described by a space metric that uses the unitless coordinate χ ("chi," rhymes with "high"), the lower-case Greek letter that corresponds to the Roman x.

Traveling in flat space

Case I: flat space. For flat space, equation (2) tells us that $K = 0$ in (1). For this case the r-coordinate is simply the product of the scale factor R and the unitless coordinate χ:

$$r = R\chi \qquad \text{so that} \qquad dr = Rd\chi \qquad (\text{flat space}, \ 0 \leq \chi < \infty) \tag{5}$$

This leads to the metric for flat space:

$$ds^2 = R^2\left(d\chi^2 + \chi^2 d\phi^2\right) \qquad (\text{flat space, } K = 0 \text{ and } 0 \leq \chi < \infty) \tag{6}$$

If you start walking "straight in the χ-direction" in a flat space, you do not return to your starting point.

Variable χ automatically satisfies limits.

Case II: closed space. Limits on the r-coordinate in (3) for a closed space can be automatically satisfied with a coordinate transformation. Let

$$r \equiv \frac{1}{K^{1/2}} \sin\chi \qquad (K > 0 \text{ and } 0 \leq \chi \leq \pi) \tag{7}$$

The sine function automatically limits the range of r to that given in (3). The coordinate r is a troublemaker; it has the same value in the two hemispheres of the sphere (Figure 2). But we use the coordinate χ, which does not have this problem; it is single-valued.

The differential dr is

$$dr = \frac{1}{K^{1/2}} \cos\chi d\chi \qquad (K > 0 \text{ and } 0 \leq \chi \leq \pi) \tag{8}$$

With these transformations the metric for the closed, constant-curvature space (1) and (3) becomes

$$ds^2 = \frac{1}{K}\left(d\chi^2 + \sin^2\chi \, d\phi^2\right) \qquad (\text{closed space, } K > 0 \text{ and } 0 \leq \chi \leq \pi) \tag{9}$$

Equation (9) is equivalent to the space metric for the surface of Earth, equation (3), Section 2.3:

$$ds^2 = R^2(d\lambda^2 + \cos^2\lambda \, d\phi^2) \qquad (\textit{space metric}: \text{Earth's surface}) \tag{10}$$

Expressions in parentheses on the right sides of both (9) and (10) refer to the unit sphere. In Chapter 2 we used the latitude λ rather than the colatitude χ. The two are related by the following equation, illustrated in Figure 2:

$$\chi \equiv \frac{\pi}{2} - \lambda \tag{11}$$

Transformation (11) replaces the sine in (9) with the cosine in (10).

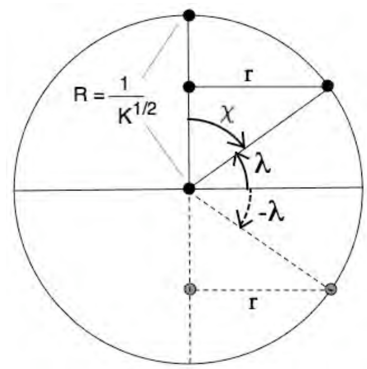


FIGURE 2 Relation between latitude λ and colatitude χ to determine the north-south coordinate on the sphere with $R = 1/K^{1/2}$ in Euclidean space. Latitude λ ranges over the values $-\pi/2 \leq \lambda \leq +\pi/2$, whereas colatitude χ ranges over $0 \leq \chi \leq \pi$. Equation (11) gives the relation between χ and λ, while (7) gives the relation between χ and r. This figure also shows that r is a "bad" coordinate, since it is double-valued, failing to distinguish between northern and southern latitude. In contrast, χ is single-valued from $\chi = 0$ (north pole) to $\chi = \pi$ (south pole).

Describing closed space

Thus for $K > 0$ the shape of constant-curvature space is that of a spherical surface with a scale factor R whose square is equal to $1/K$. The space represented by the surface of the sphere is homogeneous and isotropic: the same everywhere and in all directions. Same shape in this model means same physical experience in its predictions. In addition, if you start walking "straight in the χ-direction" in this closed space, you return eventually to your starting point.

When we use R instead of K, equation (9) becomes

$$ds^2 = R^2 \left(d\chi^2 + \sin^2\chi\, d\phi^2\right) \qquad \text{(closed space, } 0 \leq \chi \leq \pi\text{)} \tag{12}$$

where the expression in the parenthesis on the right side also embodies the shape of the unit sphere.

Comment 1. Scale factor R?
In Figure 2, R is the radius of a sphere in Euclidean space. In equation (12) R is a scale factor in curved spacetime. Euclid does not describe curved spacetime, so what does "scale factor" mean for the description of our Universe? We cannot answer this question until we know what the Universe contains, the subject of the following chapter. In the meantime we continue to play the dangerous analogy between points in flat space and events in curved spacetime begun in Chapter 2.

Describing open space

Case III: open space. Values $K < 0$ in metric (1) lead to an *open* space, as shown by the alternative transformation:

$$r \equiv R \sinh\chi \qquad \text{(open space, } 0 \leq \chi < \infty\text{)} \tag{13}$$

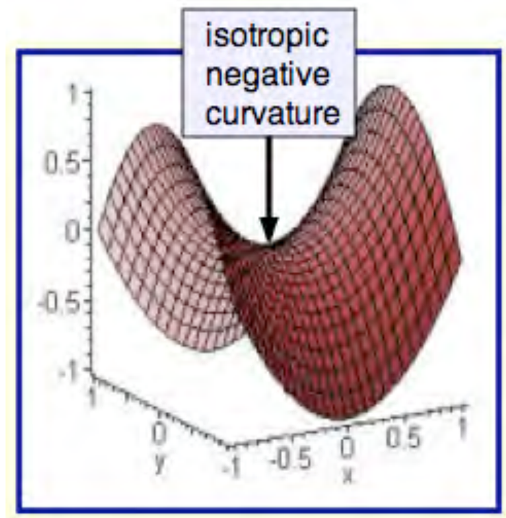


FIGURE 3 The saddle shape has intrinsic negative curvature. Only in the neighborhood of a single (central) point, however, is the negative curvature the same in all directions. Elsewhere on the surface the curvature is negative but varies from place to place and is different in different directions. (It is mathematically impossible to embed in three spatial dimensions a two-dimensional surface that has uniform negative curvature everywhere.)

where $R^2 = -1/K$ and sinh is the hyperbolic sine. The hyperbolic sine and cosine are defined by the equations

$$\sinh\chi \equiv \frac{e^{\chi} - e^{-\chi}}{2} \qquad \text{and} \qquad \cosh\chi \equiv \frac{e^{\chi} + e^{-\chi}}{2} \tag{14}$$

Equation (13) shows r to be a monotonically increasing function of χ, so there is no worry about a single value of r representing more than one location. The differential dr is

$$dr = R\cosh\chi\, d\chi \qquad (\text{open space},\ \ 0 \le \chi < \infty) \tag{15}$$

and the corresponding space metric is

$$ds^2 = R^2\left(d\chi^2 + \sinh^2\chi d\phi^2\right) \qquad (\text{open space, } K < 0 \text{ and } 0 \le \chi < \infty) \tag{16}$$

The expression in the parentheses on the right side of this equation embodies an open space that has a uniform negative curvature. The saddle surface shown in Figure 3 has a single central point whose curvature is negative and the same in all directions. That is the *only* point on the surface with the same curvature in all directions. Unfortunately it is not possible to embed in three spatial dimensions a two-dimensional surface that has uniform negative

Box 2. What does the Universe expand into?

A common misconception is that the Universe expands in the same way that a balloon expands or a firecracker explodes: into a pre-existing three-dimensional space. That is wrong: Spacetime comes into existence with the Big Bang and develops with t.

If you stick with the image of the expanding balloon for the closed Universe, the model correctly requires you to assume that the surface of the balloon is all that exists. Galaxies are scattered across its surface and human observers are surface creatures who view nothing but what lies on that surface. At the beginning of expansion, the surface evolves from a point-event that is also the beginning of time—the so-called **Big Bang**. During the subsequent expansion, every surface creature sees other points on the balloon move away from him, and points farther from him move away faster. In this model, the balloon does not expand *into* space, it represents *all* of space.

curvature everywhere. The best we can do is the saddle shape, with its single point of isotropic negative curvature.

14.3 ■ ROBERTSON-WALKER GLOBAL METRIC

A Universe that expands

“Expands” means $R(\text{constant}) \to R(t)$

We hear that the Universe “expands with time.” What does that mean? Space metric (12) describes the surface of Earth, with R equal to Earth’s radius. Suppose we inflate the Earth like a balloon. Then R increases with t while its property of uniform space curvature remains. By analogy, to describe a Universe that expands while keeping the same shape, we replace the static scale factor R in equations (12), (16), and (6) with a scale factor $R(t)$ that increases with t. In the 1930s, Howard Percy Robertson and Arthur Geoffrey Walker proved that the *only* spacetime metric that describes an evolving, spatially uniform Universe takes the form:

$$\boxed{d\tau^2 = dt^2 - R^2(t)\left[d\chi^2 + S^2(\chi)d\phi^2\right] \qquad \text{(Robertson-Walker metric)} \qquad (17)}$$

Robertson-Walker metric

To describe different shapes of the Universe, we modify the function $S(\chi)$ by generalizing equations (5), (7), and (13) respectively:

$$S(\chi) = \chi \qquad (\text{flat Universe}, \quad 0 \le \chi < \infty) \qquad (18)$$

$$S(\chi) = \sin\chi \qquad (\text{closed Universe}, \quad 0 \le \chi \le \pi) \qquad (19)$$

$$S(\chi) = \sinh\chi \qquad (\text{open Universe}, \quad 0 \le \chi < \infty) \qquad (20)$$

Comoving coordinates

Coordinates χ and ϕ are called **comoving coordinates** because a galaxy with fixed χ and ϕ simply “rides along” as the scale function $R(t)$ increases.

For a closed Universe, $R(t)$ might be interpreted loosely as the “radius of the Universe.” However, for flat or open Universes, $R(t)$ has no such simple interpretation. We simply call R the **scale function of the Universe.**

Box 3. Is a static, uniform Universe possible?

The Robertson-Walker metric (17) is more general than general relativity. Whether or not the Robertson-Walker metric satisfies Einstein's field equations depends on variation of the scale function $R(t)$ with the global t coordinate. At any value of t, the function $R(t)$ depends on what the Universe is made of and how much of each constituent is present at that t and was present at smaller t. Chapter 15, Cosmology, examines the presence and density of the constituents of the Universe at different global t-coordinates, then displays the resulting functions $R(t)$ that satisfy Einstein's equations, and finally traces the consequences for our current model of the development of the Universe. In the present chapter we simply assume that $R(t)$ starts with value zero at the Big Bang and thereafter increases monotonically.

In 1917 Einstein thought that the Universe was not only uniform in space, but also unchanging in t. Such a spacetime has the spacetime metric (17) with R a constant. Is this a valid metric for the Universe?

Einstein showed that metric (17) with R = constant does not satisfy his field equations for a Universe uniformly filled with matter. However, by adding the cosmological constant Λ to his field equations, he obtained a unique solution for a closed Universe, the case described by (19). The effect of Λ is to create a cosmic repulsion that keeps galaxies from being drawn together by gravity. Chapter 15, Cosmology, shows that something very much like Λ—now called $dark\ energy$—repels galaxies, so at the present stage of the Universe distant galaxies fly away from our own galaxy with increasing speed.

YOU ARE AT THE "CENTER OF THE UNIVERSE."

For all three models of the Universe described by (18) through (20), the location $\chi = 0$ appears to be a favored point, for example the north pole for the closed Universe or the center of the saddle for the open Universe or an origin anywhere in the flat Universe. Because the Universe is assumed to be completely uniform, however, we can choose *any* point as $\chi = 0$ (and as the origin of ϕ). That arbitrary point then becomes the north pole or the center of the saddle or the origin in flat space. The mathematical model permits every observer to assume that s/he is at the center of the Universe. (Talk about ego!)

Global t on wristwatch of comoving observer

The squared t-differential dt^2 in (17) has the coefficient one; in Robertson-Walker map coordinates, t has no warpage. Indeed, for $d\chi = d\phi = 0$, passage of coordinate t tracks the passage of wristwatch time τ. The interpretation is simple: coordinate t is that recorded on comoving clocks, those that ride along "at rest" with respect to the space coordinates of the expanding Universe.

Space and time exist only for $t > 0$.

We should also give a range for coordinate t in order to complete the definition of the spacetime region described by equations (17) through (20). However we cannot specify a range of t until we know details of the scale function $R(t)$. For Big Bang models of the Universe—expansion from an initial singularity—the scale function starts with $R(t) = 0$ at $t = 0$. In this book we examine Big Bang models, for which spacetime exists only for $t > 0$.

14.4 ■ REDSHIFT

Light we receive from far away increases in wavelength in an expanding Universe.

Choose the center of the Universe to be at *my location*, and t_0 to be *now*.

We are free to choose the center of the Universe at our location, that is at $\chi = 0$ and to assume that we stay at the center permanently. Then every current observation that we make is an event that takes place at $\chi = 0$ and *now*, which we will call $t = t_0$.

$$\text{Observation NOW on Earth has map coordinates } t \equiv t_0,\ \chi \equiv 0 \qquad (21)$$

Suppose that a distant star is fixed in comoving coordinates χ and ϕ, so it rides along as the scale function $R(t)$ increases. Let the star emit a light flash at $(t_{\text{emit}}, \chi_{\text{emit}})$, which we observe on Earth at $(t_0, 0)$.

For light, $d\tau = 0$ and for radial motion $d\phi = 0$ in metric (17). Write the resulting metric with t and space terms on opposite sides of the equation, take the square root of both sides, and integrate each one:

$$\int_{t_{\text{emit}}}^{t_0} \frac{dt}{R(t)} = \int_0^{\chi_{\text{emit}}} d\chi = \chi_{\text{emit}} \qquad (\text{light, d}\phi = 0) \qquad (22)$$

Emit and detect two light flashes.

Think of a second light flash emitted from the same star at event $(t_{\text{emit}} + \Delta t_{\text{emit}}, \chi_{\text{emit}})$ and observed by us at $(t_0 + \Delta t_0, 0)$. The two flashes can represent two sequential positive peaks in a continuous wave. We assume that the emitter is located at constant χ, so the second flash travels the same χ-coordinate difference as the first. Hence the right-hand integral has the same value for both flashes. Therefore

$$\int_{t_{\text{emit}}+\Delta t_{\text{emit}}}^{t_0+\Delta t_0} \frac{dt}{R(t)} = \chi_{\text{emit}} \qquad (\text{light}) \qquad (23)$$

Compare the t-limits of the integrals on the left sides of (22) and (23). The integration in (23) starts later by Δt_{emit} and ends later by Δt_0. In consequence, when we subtract the two sides of equation (22) from the corresponding sides of equation (23), the result is:

$$\int_{t_0}^{t_0+\Delta t_0} \frac{dt}{R(t)} - \int_{t_{\text{emit}}}^{t_{\text{emit}}+\Delta t_{\text{emit}}} \frac{dt}{R(t)} = 0 \qquad (\text{light}) \qquad (24)$$

Approximate this equation to first order in Δt_{emit} and Δt_0, leading to

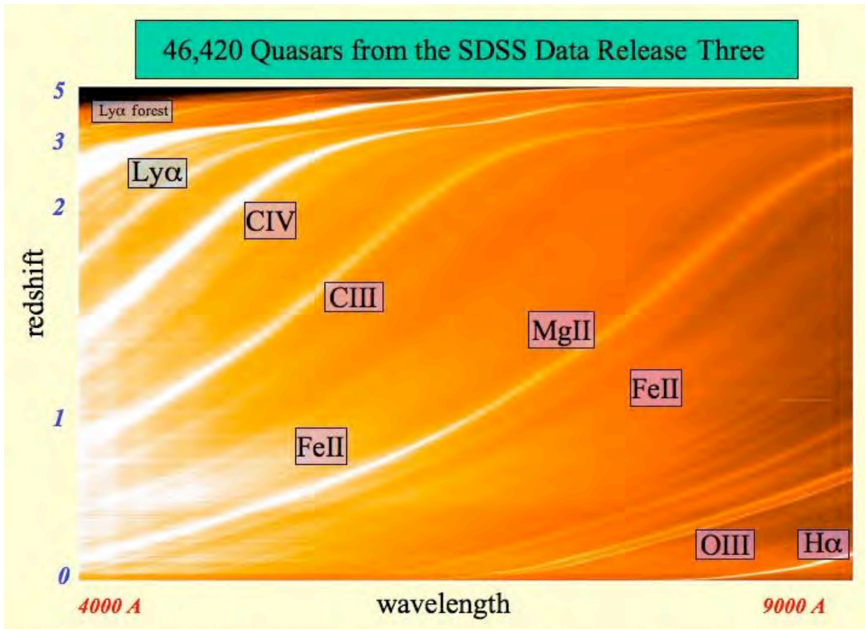
$$\frac{\Delta t_0}{R(t_0)} \approx \frac{\Delta t_{\text{emit}}}{R(t_{\text{emit}})} \qquad (\text{light}) \qquad (25)$$

Let the two flashes represent two sequential peaks in a continuous wave. Then the lapse in t between flashes in meters that each observer measures equals the wavelength in meters.

$$\frac{\Delta t_0}{\Delta t_{\text{emit}}} = \frac{\lambda_0}{\lambda_{\text{emit}}} = \frac{R(t_0)}{R(t_{\text{emit}})} \qquad (\text{light}) \qquad (26)$$

Redshift z

In this equation an equality sign replaces the approximately equal sign in (25) because one wavelength of light λ is truly infinitesimal compared with the scale function $R(t)$ of the Universe. It is customary to measure the fractional



In addition to images, the SDSS has measured the spectra of light from more than a million celestial sources. The spectrum of an object shows the intensity of its light as a function of wavelength. This picture shows the spectra of more than 46,000 quasars from the SDSS 3rd data release; each spectrum has been converted to a single horizontal line, and they are stacked one above the other with the closest quasars at the bottom and the most distant quasars at the top. Bright bands show the emission produced by specific ions of hydrogen, carbon, oxygen, magnesium, and iron. For more distant quasars, these emission lines are shifted to longer wavelengths by the expansion of the universe. This redshift of spectral lines is what the SDSS measures to determine the distances to quasars and galaxies.
Credit: X. Fan and the Sloan Digital Sky Survey.

FIGURE 4 A remarkable plot of the redshifts z of the spectra from more than 46 thousand quasars taken by the Sloan Digital Sky Survey (SDSS). The spectrum of each quasar lies along a single horizontal line at a vertical position corresponding to its redshift z. Some prominent spectral lines from different atoms are labeled: Lyα is the Lyman alpha line of hydrogen. Roman numeral I following an element is the neutral atom; Roman numeral II is the singly ionized atom, and so forth. Thus MgII is singly ionized magnesium and CIV is triply ionized carbon. The observed wavelength λ_0 increases with increasing z. (The redshift scale is nonlinear so the bands are not straight lines.)

change in wavelength using a dimensionless parameter z, called the **redshift**, defined by the equation

$$\lambda_0 \equiv (1+z)\lambda_{\text{emit}} \qquad \text{(light)} \qquad (27)$$

Stretch factor: $1+z$

where we call $1+z$ the **stretch factor**. Then equation (26) can be written

$$1+z \equiv \frac{\lambda_0}{\lambda_{\text{emit}}} = \frac{R(t_0)}{R(t_{\text{emit}})} \qquad \text{(stretch factor)} \qquad (28)$$

In other words, when we train our telescopes on a source with redshift z, we observe light emitted at the t-coordinate when the Universe scale function $R(t)$ was a factor $1/(1+z)$ the size it is today.

Cosmological redshift

The change in wavelength described by equation (28) is called the **cosmological redshift**. The observation t_0 is greater than the emission t_{emit}, and for an expanding universe $R(t_0) > R(t_{\mathrm{emit}})$. Therefore the observed light has a longer wavelength than the emitted light; the color of light visible to our eyes shifts toward the red end of the spectrum, hence the term "redshift." The same fractional increase in wavelength occurs for electromagnetic radiation of any frequency, so the term *redshift* applies to microwaves, infrared, ultraviolet, x-rays, and gamma rays.

Redshift a Doppler shift?

Equation (27) appears not to describe a Doppler shift in the special relativity sense. Both emitter and observer are *at rest* in their comoving coordinate χ; nevertheless, they observe the light to have different wavelengths. In a sense the expansion of the Universe "stretches out" the wavelength of the light as it propagates. In another sense, however, the cosmological redshift is a cumulative redshift, because a star at fixed χ is at an $R(t)\chi$ that grows with t. In other words, it moves away from us. Section 14.7 shows that for $z \ll 1$, the cosmological redshift *is* a Doppler shift.

Redshift deduced from laboratory spectra

When we see light of a given frequency that has been emitted from a distant galaxy, how do we know that it has been redshifted? With what do we compare it? From laboratory experiments on Earth, we know the discrete spectrum of radiation frequencies emitted by a particular atom or molecule. Then the identical *ratios* of frequencies of light received from a distant star tell us what element or molecule we are observing in that star. And from the value of the shift at any one frequency we can deduce the redshift for all frequencies. Figure 4 shows redshifted spectral lines (bright: emission lines; dark: absorption lines) of light from many different atoms in distant quasars.

Astronomers use z for t_{emit}.

Because it is easy to measure a galaxy's redshift z, astronomers use z as a proxy for t_{emit} in equation (26)—Figure 5. Whenever you read a news article about a galaxy formed during the first billion years of the Universe, remember that astronomers do not measure t; they measure redshift. The distant galaxies in the news have $z > 6$: in the process of traveling to us, the wavelength of their light has been stretched by a factor more than 7! Light in our visual spectrum has been redshifted to the infrared. This is why the James Webb Space Telescope—the successor to the Hubble Space Telescope—looks in the infrared region of the spectrum for light from the most distant galaxies, those that appeared earliest in the evolution of the Universe.

14.5 ■ HOW DO GALAXIES MOVE?

Apply the Principle of Maximal Aging to the motion of a galaxy.

Transverse galaxy motion is difficult to detect.

We have a disability in viewing the distant Universe: we are limited to effectively a single point, the Earth and its solar system. The redshift of light from distant galaxies gives us a handle on their radial recession. However,

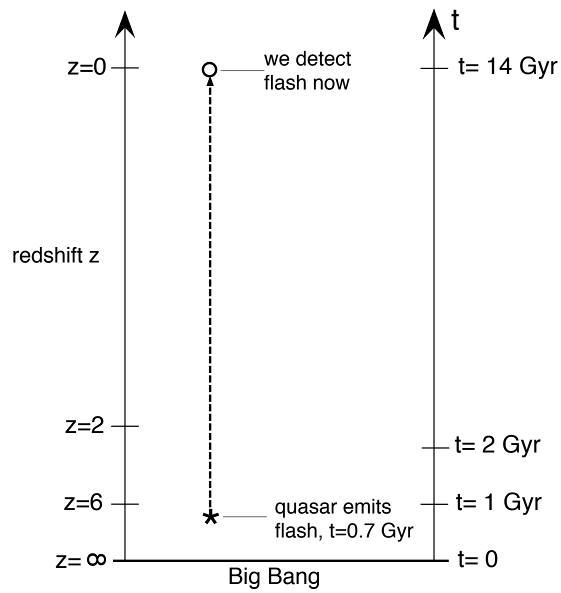


FIGURE 5 Schematic diagram comparing redshift z with cosmic t, in units of Gigayears (10^9 years). Calibration of the scale at the right of the figure depends on the t-development of the Universe, through $R(t)$, based on our current model. Astronomers use redshift as a proxy for t, both because it is directly measurable and also because it does not change as we revise our scale of cosmic t. The flash emission and detection is the case analyzed in Box 4.

transverse motion of a remote source is too small to detect directly in a human lifetime. (See the exercises.) In this and following sections, however, we limit attention to sources that move radially away from us.

Limit attention to radial motion.

How do galaxies move in the global coordinate system of metric (17)? As usual, the metric tells us about the structure of spacetime but does not determine the motion of a stone—or a galaxy. For that we need the Principle of Maximal Aging, which requires that total wristwatch time be a maximum along the worldline of a free galaxy that crosses adjoining flat patches.

Galaxy motion from Principle of Maximal Aging

For radial motion, the metric (17) becomes:

$$d\tau^2 = dt^2 - R^2(t)d\chi^2 \qquad (d\phi = 0) \tag{31}$$

This metric is valid for any function $S(\chi)$ in (17), whether for a flat, closed, or open model Universe. By just looking at this metric, can we anticipate constants of motion? One metric coefficient depends explicitly on t through the function $R(t)$. All our earlier derivations of map energy as a constant of motion required that no metric coefficient be an explicit function of t. Therefore metric (17) tells us that energy will *not* be conserved in the motion of galaxies. However, for radial motion ($d\phi = 0$) the metric coefficients do not

Seek a conserved quantity.

Box 4. How far away (now) is the most distant galaxy that we see (now)?

We see *now* the most distant galaxies as they *were* when they emitted the light: at, say, $t_{\mathrm{emit}} = 0.7$ billion years after the Big Bang (Figure 5). The current age of the Universe is $t_0 \approx 14$ billion years, so $t_0 - t_{\mathrm{emit}} \approx 13.3$ billion years. Naively, then, we might expect that these galaxies lie about 13 billion light years from us. However, this is false; they must lie much further away at the present day. Why? Because these galaxies have moved farther away from us during the 13.3 billion years that it took for their light to reach us. How much farther? What is the "true" map distance *now* between us and a galaxy formed at $t_{\mathrm{emit}} = 0.7$ billion years ago? In this case the word "true" has meaning only through the metric.

Use the Robertson-Walker metric (17) with $d\tau = 0$ to obtain the map distance between the emitting galaxy (at $\chi = \chi_{\mathrm{emit}}$) and Earth (at $\chi = 0$) at any particular t. This map distance is given simply by $R(t)\chi_{\mathrm{emit}}$, since the emitter continually "rides along" at the constant comoving coordinate χ_{emit}. The present separation $d_0 \equiv \sigma_0$ is then just $R(t_0)\chi_{\mathrm{emit}}$ with χ_{emit} given by (22).

$$d_0 = R(t_0)\chi_{\mathrm{emit}} = R(t_0)\int_{t_{\mathrm{emit}}}^{t_0} \frac{dt}{R(t)} \qquad (29)$$

We cannot complete this calculation until we know how the scale function $R(t)$ increases with t. That is the task of Chapter 15. For a rough estimate of the present map distance d_0, assume that the scale function increases uniformly with t: $R(t)/R(t_0) = t/t_0$. Then the integral in (29) can be carried out using $t_{\mathrm{emit}} = 0.7$ billion years and the present $t_0 = 14$ billion years:

$$\begin{aligned} d_0 &= t_0 \int_{t_{\mathrm{emit}}}^{t_0} \frac{dt}{t} = t_0 \ln \frac{t_0}{t_{\mathrm{emit}}} \qquad (30) \\ &= t_0 \ln \frac{14}{0.7} = 14 \times 3.0 = 42 \end{aligned}$$

in billions of light-years. We call d_0 the **look-back distance**. According to this rough model, look-back distances of galaxies that emitted light 13 billion years ago are something like $d_0 = 42$ billion light years. This is their calculated map distance away from us now. We can refine this estimate by using a more accurate scale function $R(t)$; the present look-back distance to these remote galaxies is almost certainly larger than 42 billion light years.

depend explicitly on χ, so there will be a conserved quantity related to motion in χ, a kind of radial momentum.

The galaxy crosses two adjoining patches (Figure 6). Label A and B the segments of its path across the respective patches. Consider three events: Two at the opposite edges of the patches and one where they join. To find momentum as a constant of motion, we fix the t of all three events and fix the locations of the two events at the outer ends of the two segments. Then we vary the χ-coordinate of the connecting event (and the boundary between patches) in order to maximize total wristwatch time.

Over one patch, $R(t)$ is treated as being constant, so each patch is flat. Define

$$R_{\mathrm{A}} \equiv R(\bar{t}_{\mathrm{A}}) \qquad \text{and} \qquad R_{\mathrm{B}} \equiv R(\bar{t}_{\mathrm{B}}) \qquad (32)$$

where $\bar{t}_{\mathrm{A}}$ and $\bar{t}_{\mathrm{B}}$ are the average t-values when the galaxy crosses patch A and B, respectively. Define t for the galaxy to cross each patch as:

$$\begin{aligned} t_{\mathrm{A}} &\equiv t_{\mathrm{middle}} - t_{\mathrm{start}} \qquad (33) \\ t_{\mathrm{B}} &\equiv t_{\mathrm{end}} - t_{\mathrm{middle}} \end{aligned}$$

Let χ_{A} be the *change* in coordinate χ across segment A and χ_{B} be the corresponding change across segment B. Then $R_{\mathrm{A}}\chi_{\mathrm{A}}$ is the radial separation

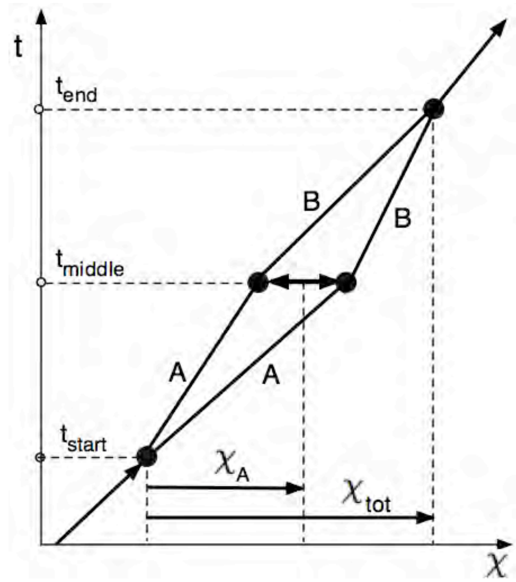


FIGURE 6 Greatly magnified picture of alternative worldlines across incremental segments A and B used in the derivation of the constant of motion (38). We vary the position $\chi_{\rm A}$ of the middle event between segments A and B and demand that the total wristwatch time across both segments be maximum. The origin of this diagram is NOT necessarily at the zero of either t or radial position.

across segment A and $R_{\rm B}(\chi_{\rm tot} - \chi_{\rm A})$ the radial separation across segment B, with $\chi_{\rm A}$ variable. Then the metric (31) across the two patches becomes:

$$\tau_{\rm A} = \left[t_{\rm A}^2 - R_{\rm A}^2\,\chi_{\rm A}^2\right]^{1/2} \tag{34}$$

and

$$\tau_{\rm B} = \left[t_{\rm B}^2 - R_{\rm B}^2\,(\chi_{\rm tot} - \chi_{\rm A})^2\right]^{1/2} \tag{35}$$

Fix $t_{\rm start}$, $t_{\rm middle}$, and $t_{\rm end}$ at the edges of the two segments. This fixes the values of $t_{\rm A}$, $t_{\rm B}$, $R_{\rm A}$, and $R_{\rm B}$ through equations (32) through (35).

Now vary $\chi_{\rm A}$ to maximize the total wristwatch time $\tau_{\rm tot} = \tau_{\rm A} + \tau_{\rm B}$ across both segments:

$$\begin{aligned}\frac{d\tau_{\rm tot}}{d\chi_{\rm A}} &= \frac{d\tau_{\rm A}}{d\chi_{\rm A}} + \frac{d\tau_{\rm B}}{d\chi_{\rm A}} \\ &= -\frac{R_{\rm A}^2\,\chi_{\rm A}}{\tau_{\rm A}} + \frac{R_{\rm B}^2\,(\chi_{\rm tot} - \chi_{\rm A})}{\tau_{\rm B}} \\ &= -\frac{R_{\rm A}^2\,\chi_{\rm A}}{\tau_{\rm A}} + \frac{R_{\rm B}^2\,\chi_{\rm B}}{\tau_{\rm B}} = 0\end{aligned} \tag{36}$$

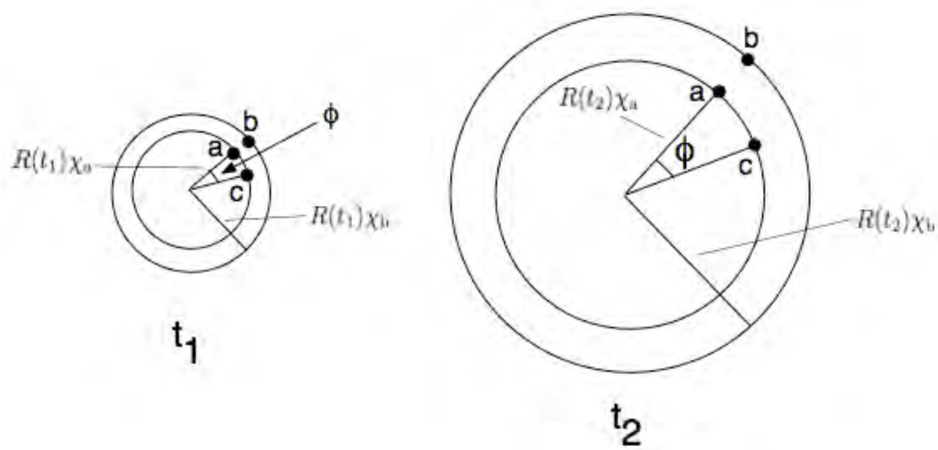


FIGURE 7 One possible radial motion for a galaxy is to remain at rest in the comoving coordinate χ and ϕ and ride outward, following $R(t)$, as the Universe expands. This figure shows the result for a flat Universe. All separations increase by the same ratio, so every observer can analyze galaxy motion with himself at the center and galaxies expanding away from him.

or

$$\frac{R_{\mathrm{B}}^2\,\chi_{\mathrm{B}}}{\tau_{\mathrm{B}}} = \frac{R_{\mathrm{A}}^2\,\chi_{\mathrm{A}}}{\tau_{\mathrm{A}}} \tag{37}$$

Now the usual argument: The left side of (37) refers to parameters of segment B alone, the right side to parameters of segment A alone. We have found a quantity that has the same value for each segment—that is, a constant of motion. Restore differentials and define a constant of motion Q_{r}.

Constant Q_{r} for radial motion only

$$Q_{\mathrm{r}} \equiv mR^2\frac{d\chi}{d\tau} = R\left(\frac{mRd\chi}{d\tau}\right) \equiv Rp_{\mathrm{r}} \quad \text{is a constant of motion} \tag{38}$$

Constant of motion for galaxy or light

where (38) provides a definition of local radial momentum p_{r} because $Rd\chi$ is a measured distance, from (17). Here m is the mass of a stone—or of a galaxy! Let the motion be radial only, so $p_{\mathrm{r}} = p$. Then (38) is still valid as $m \to 0$ for a photon, with $p = E$. In other words $R(t)E$ is constant for light, which means that as $R(t)$ increases, the energy E of photons decreases—another example of cosmological redshift.

Two possible radial motions

We can distinguish two possible radial motions of a galaxy that leave Q_{r} constant. In the first, χ remains constant as t increases, so $d\chi/d\tau = 0$ and $Q_{\mathrm{r}} = p_{\mathrm{r}} = 0$. Each such "comoving" galaxy rides outward with $R(t)$; two galaxies at different values of χ move apart as $R(t)$ increases with t. For flat space ($S = \chi$) one can think of a set of concentric rings of galaxies fixed in the comoving coordinate χ. As t increases, the radius of each ring increases with $R(t)$. Figure 7 shows that radial separations $R(t)\chi$ and tangential separations $R(t)\chi\phi$ both increase proportionally to $R(t)$. This is true for every observer. There is no unique center; every observer can plot the expansion of the Universe in global coordinates with himself at the center.

In the second possible radial motion that leaves Q_{r} constant, a galaxy moves radially with respect to comoving coordinate χ. (Most galaxies have at least a slightly non-zero Q_{r} because of local gravity from spatial inhomogeneities.) Or one can think of a stone thrown radially out of a comoving galaxy. For such motion one can rewrite (38) as:

$$p_{\mathrm{r}} = \frac{Q_{\mathrm{r}}}{R(t)} \tag{39}$$

Q_r remains constant and $R(t)$ increases, so p_{r} decreases. This is called the "cosmological redshift of momentum." The high speed limit on (39) applies to a photon:

$$E = p \propto \frac{1}{R(t)} \qquad \text{(light)} \tag{40}$$

Constant Q_ϕ for *any* motion

We can derive another constant of motion, one that is valid for *any* free motion in Robertson-Walker global coordinates. Apply the Principle of Maximal Aging to two patches separated in ϕ-coordinate instead of χ-coordinate. The result is

$$Q_\phi \equiv mR^2S^2\frac{d\phi}{d\tau} = RS\left(\frac{mRSd\phi}{d\tau}\right) \equiv RSp_\phi \tag{41}$$

(constant for *any* free motion)

Equation (41) provides a definition of local tangential momentum p_ϕ because $RSd\phi$ is a measured distance, from metric (17).

14.6 ■ MEASURING DISTANCE

Extending a ruler from one lonely outpost.

Problems with our observations

So much for the theory of how galaxies move in the expanding Universe. What predictions does theory make about observations? On Earth we describe motion by plotting distance vs. time. Life in the Universe is more complicated. There are two problems: We cannot directly measure distances to objects outside our galaxy, and we cannot directly measure times longer than a few centuries. What hope can we have, therefore, to measure billions of years and billions of light years in the Universe?

First we give up trying to measure time. Instead we measure distance and velocity, both through indirect means. Section 14.7 discusses velocity measurements through redshift of spectral lines; here we focus on distance.

Box 5. Edwin P. Hubble

FIGURE 8 Edwin P. Hubble on the cover of Time Magazine, 1948.

Edwin P. Hubble was as important to astronomy as Copernicus. He expanded our view of the Universe from a single home galaxy to many galaxies that are rushing away from one another.

Hubble was born in 1889. In his youth he was an outstanding athlete and one of the first Rhodes Scholars at Oxford University, England. After returning to the United States he taught Spanish, physics, and mathematics in high school. He served in World War I, after which he earned a Ph.D. at the Yerkes Observatory of the University of Chicago.

In 1919 Hubble took up a position at Mount Wilson Observatory where he used the new 100-inch Hooker reflecting telescope, with which he discovered and analyzed redshifts of light from what were called "nebulae." At that time the prevailing view was that the Universe consisted entirely of our galaxy. Hubble showed that nebulae are not objects within our galaxy but galaxies themselves, in motion away from our galaxy. The nearby galaxies he studied recede from us at speeds proportional to their map separation from us (Figure 11).

Before his death in 1953, Hubble made observations with the 200-inch telescope installed on Mount Palomar, California in 1948.

Comment 2. "Distance" and "time"? Look out!
Review Section 2.7, titled *Goodbye "Distance." Goodbye "Time"*, which first asserted that we cannot apply the concepts of distance and time to our observations of the Universe. The present chapter deeply embodies that assertion.

Determine "distance" with a "standard candle."

We cannot use laser ranging or classical surveying methods to measure distances outside our galaxy. The most widely used method employs what is called a **standard candle**, a light source whose intrinsic brightness is known. From that intrinsic brightness (more precisely, luminosity) and the apparent brightness (more precisely, flux density) of the object viewed on Earth, we can determine a distance. However, the expanding Universe complicates the analysis, as detailed in Box 4.

Cepheid variables: standard candles

When Hubble did his observations, the major standard candle was one form of the so-called *Cepheid variable* stars. These are stars whose emitted power varies periodically. Their rate of pulsation depends on their emitted power: the longer the pulsation period, the greater the emitted power of the star.

Hubble found Cepheid variable stars in nearby galaxies (but he could not detect them in distant galaxies). To find their approximate distances he classified different galaxies, found the intrinsic brightness of galaxies of a given type that were near enough to allow detection of Cepheid variables they contained, then assumed the same intrinsic brightness for more distant (but still nearby) galaxies of the same type.

Hubble's "island universes" = our galaxies.

Hubble's observations in 1923-1924 showed that most spiral nebulae (for him, fuzzy patches of light in the sky) are much farther away than the limits of our galaxy; they are indeed separate "island universes," or what we now call "galaxies." He also classified "elliptical," "lenticular," and "irregular" galaxies, so-called because of their appearance. All lie outside our own Milky Way galaxy. (Interesting fact: Both "galaxy" and "lactose" come from the Greek and Latin words for milk.) In summary: The Universe extends far beyond our galaxy.

Modern standard candle: Type Ia supernova

Cepheid variable stars are too faint to be seen at distances more than a hundred million light years. For more distant sources, the standard candle of choice is a Type Ia supernova. A Type Ia supernova results when a small, dense white dwarf star gradually accretes mass from a binary companion star, finally reaching a mass at which the white dwarf becomes unstable, collapses, and explodes into a supernova. The "slow fuse" on the gradual accretion process can lead to an explosion of almost the same size on each such occasion, giving us a "standard candle" of the same intrinsic brightness. The brightness of the explosion as seen from Earth provides a measure of the distance to the supernova. The cosmological redshift of light tells us how fast the supernova is receding (Section 14.4). Because supernovae (plural of supernova) are so bright, they can be seen at a very great distance, which brings us information about the Universe most of the way back to the Big Bang.

For astronomers, M and m are magnitudes.

Astronomers plot a quantity called *distance modulus* $m - M$ (also called the *effective magnitude*) where m is the apparent magnitude and M is the absolute magnitude (also called the **intrinsic magnitude**). This difference is related to luminosity distance d_{L} (Box 6) by the equation

$$m - M = 5\log_{10}\left(\frac{d_{\mathrm{L}}}{10\,\mathrm{pc}}\right) \qquad (m \text{ and } M \text{ are magnitudes}) \qquad (42)$$

where pc stands for *parsec*, a unit of distance equal to 3.26 light years. Why this peculiar formula? Blame the ancient Greeks, who first quantified the brightness of stars. The key is the realization that M is known (or knowable) for Type Ia supernovae, so measurements of apparent magnitude m, the distance modulus, allow us to solve equation (42) for d_{L}.

Hubble Diagram

A graph of effective magnitude vs. redshift is called a **Hubble Diagram.** Figure 9 shows the Hubble Diagram for Type Ia supernovae. The thin spread of the curve in the vertical direction confirms that Type Ia supernovae are good standard candles—they all have the same M (when small corrections are applied to raw measurements) so that apparent magnitude m can be used to measure distance.

Expansion speeding up

What are the implications of this analysis? First the obvious: Redshift increases with distance. The next section gives an interpretation of this as a result of cosmological expansion. The more subtle and surprising result is that this expansion is speeding up with t. Chapter 15, Cosmology, elaborates on this second point.

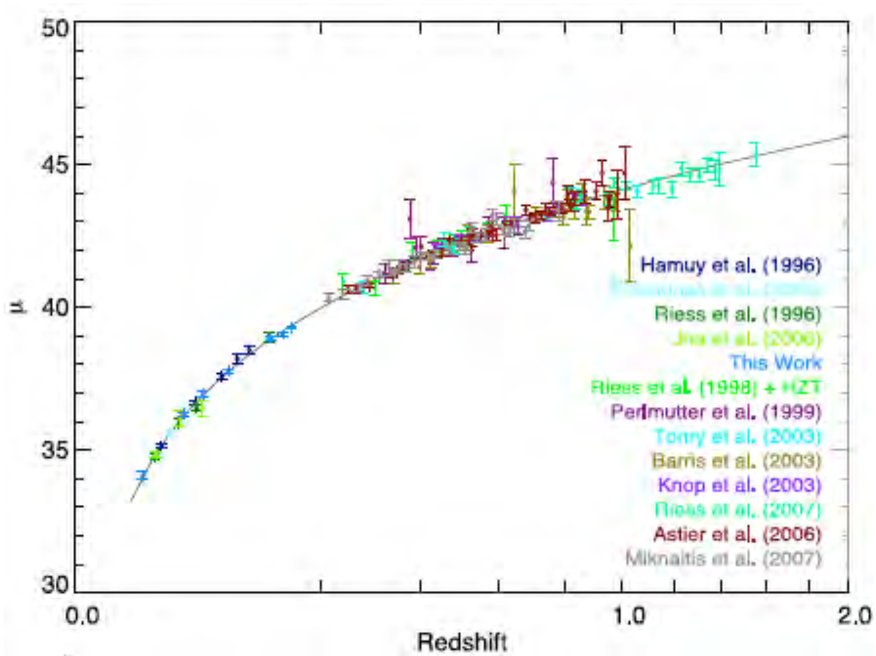


FIGURE 9 Effective magnitude of Type Ia supernovae as a function of their redshift z. The vertical axis is $\mu = m - M$, the difference between apparent magnitude and intrinsic magnitude.

In the future, a second way to measure distances may prove useful in cosmology. From metric (17), objects of known transverse size D at radial coordinate distance χ extend across an angle

$$\theta \approx \frac{D}{S(\chi)R(t_{\rm emit})} \qquad (|\theta| \ll 1) \tag{43}$$

In flat spacetime the distance would be $d = D/\theta$ if $\theta \ll 1$. In the expanding Universe, cosmologists define the **angular diameter distance** as:

$$d_{\rm A} \equiv \frac{D}{\theta} = S(\chi)R(t_{\rm emit}) = \frac{S(\chi)R(t_0)}{1+z} \tag{44}$$

Standard rulers

where we used equation (28). Objects of known transverse size D are called **standard rulers**. Comparing (44) with (52), you can show that $d_{\rm A} = d_{\rm L}/(1+z)^2$. Thus, measurements of standard candles and standard rulers for an object of known z yield the same information. The difficulty lies in determining the intrinsic size and luminosities of objects billions of light years away.

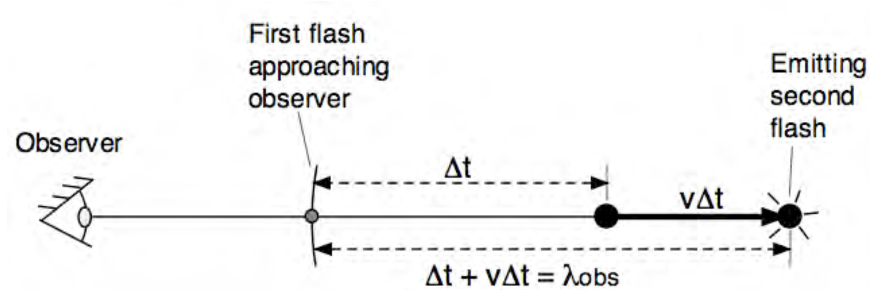


FIGURE 10 Doppler effect observed in a single inertial frame of special relativity, used by Hubble to analyze the speed of receding nearby galaxies.

14.7 ■ LAWS OF RECESSION

Recession rate proportional to "distance"—at least for nearby galaxies.

When Edwin P. Hubble arrived at the Mount Wilson Observatory in California USA in 1919 and began to use the new 100-inch telescope, many astronomers believed that the entire Universe consisted of stars in the Milky Way, what we now call "our galaxy." A disturbing feature of this model of the Universe was the behavior of some of the objects they called **nebulae**. We now know that some nebulae are within our galaxy but most are separate galaxies distant from our own. As early as 1912 Vesto Melvin Slipher had shown that light from many nebulae had significant redshifts, implying that they were moving away from us at high speed. But were these nebulae dim objects in our own galaxy or bright objects outside our galaxy? To answer this question, Hubble needed, first, a relation between redshift and recession velocity. Second, he needed a measure of the distance of these nebulae from us. We examine these tasks in turn.

Velocity vs. Redshift

Hubble used special relativity Doppler shift.

Slipher and Hubble used the Doppler shift of light to find a relation between redshift z and velocity of recession v. They were astronomers, not general relativists. (General relativity theory did not exist when Slipher began his work.) For them the nebulae were speeding away from us in static flat space, and the redshift was a Doppler effect that could be analyzed using special relativity. We will show that this simple analysis gives correct results for nearby nebulae receding from us at relative speeds much less than that of light.

Hubble uses special relativity Doppler shift

Figure 10 introduces the Doppler shift for special relativity. Earlier than the t shown in this figure an object emitted one flash, then moved $v\Delta t$ farther away from the observer, and is emitting the second flash at the instant shown. During that t-lapse the initial flash moved Δt closer to the observer. Let the lapse in t between the two flashes represent one period of a continuous wave. Then the wavelength λ_{obs} detected by the observer has the value shown in the figure. According to Newton, in the rest frame of the source the emitted wavelength would be $\lambda_{\rm source} = \Delta t$. However, we must apply a relativistic correction to Newton's result, because of time stretching.

The t-lapse between flash emissions in the rest frame of the source is different from Δt in the frame of the observer. We say that "the emitting clock runs slow," according to the equation

$$(1-v^2)^{1/2}\Delta t = \Delta t_{\rm source} = \lambda_{\rm source} \qquad \text{(special relativity)} \qquad (45)$$

The ratio of observed wavelength to the wavelength in the frame of the source is

$$\frac{\lambda_{\rm obs}}{\lambda_{\rm source}} = \frac{(1+v)\Delta t}{(1-v^2)^{1/2}\Delta t} = \left(\frac{1+v}{1-v}\right)^{1/2} = 1+z \qquad \text{(special relativity)} \; (46)$$

where we have inserted the definition of redshift z from (28). Nearby galaxies are not moving away from us very fast; for them we may make the approximation:

$$1+z = (1+v)^{1/2}(1-v)^{-1/2} \approx \left(1+\frac{v}{2}\right)^2 \approx 1+v \qquad (v \ll 1) \quad (47)$$

so for slow-moving galaxies the redshift z is equal to the velocity of recession v.

$$v = z \qquad (v \ll 1) \qquad (48)$$

Doppler OK for small z

This Doppler interpretation of the cosmological redshift is valid for $z \ll 1$, because spacetime over such a "small distance" is well approximated by a single flat patch, on which general relativity reduces to special relativity.

Measuring Distance with a "Standard Candle"

Equation (48) gives the velocity of recession. Hubble also needed to know how far away the emitting star is, $\sigma_{\rm now}$. To determine distance we use what is called a **standard candle**, that is, a star whose intrinsic brightness is known. From that intrinsic brightness and the apparent brightness of this star at Earth, one can then determine its distance. However, the expanding Universe complicates this analysis, as detailed in Box 6.

Hubble's Law of Recession

Hubble's law of recession

From the redshift of different galaxies, Hubble now knew from (48) their recession velocities. From the intrinsic brightness of Cephied variable stars and a galaxy of a given type, he could calculate its distance. He found a direct proportion between the average recession velocity of a star and its distance (Figure 11). He called this result the Redshift-Distance Law. We call it **Hubble's Law**, one of the major results of cosmology in the twentieth century:

Box 6. Finding the distance (which distance?) to a standard candle

Consider a star that emits electromagnetic power L (energy per unit time), called **luminosity**, as viewed in its rest frame. We assume that this emission is isotropic, the same in all directions. Place this star at the center of coordinates, $\chi = 0$. Place an observer at a comoving coordinate χ away from the star. In special relativity the power per unit area, also called **flux density** F, reaching an observer at this distant location is:

$$F = \frac{L}{4\pi d^2} \qquad \text{(flat spacetime)} \qquad (49)$$

where d is the distance between star and observer. Now, astronomers cannot measure d directly, so they define a **luminosity distance** $d_{\rm L}$ by the equation

$$d_{\rm L} = \left(\frac{L}{4\pi F}\right)^{1/2} \qquad (50)$$

and report the value of $d_{\rm L}$ for a given star. The luminosity distance $d_{\rm L}$ is the distance from an emitter of power L at which it would produce a flux density F *in flat spacetime*.

In an expanding Universe, F is modified in several ways. First, the metric contains no distance d, but rather a map coodinate χ and an angular factor $S(\chi)$. Second, the energy reaching the observer is reduced by a factor $(1+z)$ due to the cosmological redshift. Third, the lapse in t that this light takes to arrive at the observer is stretched out by another factor $(1+z)$. The result is

$$F = \frac{L}{4\pi(1+z)^2 R^2(t_0) S^2(\chi)} \qquad (51)$$

We can measure F and z. Suppose we also know the intrinsic power L of the emitter and, for a specific model of the Universe, the cosmic scale function $R(t_0)$. We can then obtain a measure of the distance from the emitter using (50):

$$S(\chi) = \frac{d_{\rm L}}{(1+z)R(t_0)} \qquad (52)$$

The quantities $d_{\rm L}$ and $S(\chi)$ are measures of distance to our standard candle of luminosity L. You should convince yourself that (50) and (52) taken together imply (51).

$$v = H_0 d_{\rm L} \qquad \text{(nearby galaxy)} \qquad (53)$$

Hubble constant H_0

Here H_0 is called the **Hubble constant** and refers to its value at the present age of the Universe. The current value of the Hubble constant in units used by astronomers is

$$H_0 = 73 \pm 2 \ \frac{\text{kilometer/second}}{\text{Megaparsec}} \qquad (54)$$

where one Megaparsec equals 3.26 million light years. Expressed in geometric units, this has the value:

$$H_0 = (8.0 \pm 0.2) \times 10^{-27} \text{ meter}^{-1} \qquad (55)$$

Robertson-Walker Law of Recession

Recession at great distance and great speed

What happens when we do not make the assumption that emitting galaxies are nearby? We use the Robertson-Walker metric to answer this question. Write the spacelike form of (17) for fixed ϕ-coordinate.

$$d\sigma^2 = R^2(t)d\chi^2 - dt^2 = ds^2 - dt^2 \qquad (d\phi = 0) \qquad (56)$$

At fixed t_1 this equation can be integrated to give the distance d:

$$d_1 = R(t_1)\chi \qquad (dt = 0) \qquad (57)$$

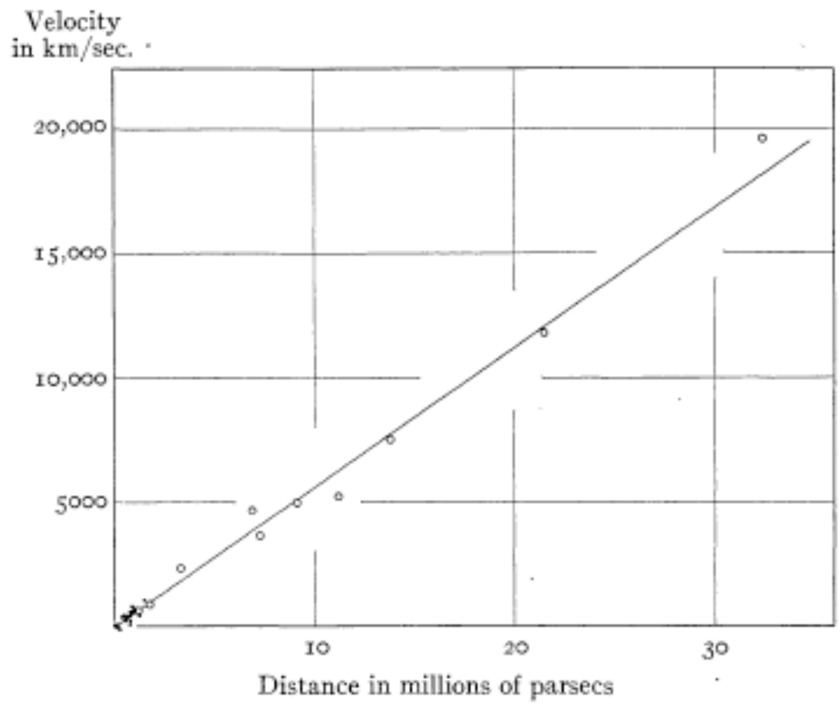


FIGURE 11 A plot of recession velocity as a function of distance by Hubble and Milton Humason (1931). Open circles represent averages of groups of galaxies; solid dots near the origin show individual galaxies from an earlier paper by Hubble. A parsec equals 3.26 light-years, so the most distant group of galaxies is approximately 100 million light-years distant—"nearby" by modern standards. The Hubble constant derived from the slope of the line in this figure is different from the current value, equation (54); see the exercises.

Assume that a distant galaxy is at rest in comoving coordinates χ (and ϕ), so that χ remains constant. Then at a later t_2, the galaxy is at distance

$$d_2 = R(t_2)\chi \qquad (dt = 0) \tag{58}$$

The recession speed at t is expressed using elementary calculus:

$$\begin{aligned} v_{\rm r} &= \lim_{t_2 \rightarrow\ t_1} \frac{d_2 - d_1}{t_2 - t_1} = \lim_{t_2 \rightarrow\ t_1} \frac{R(t_2) - R(t_1)}{t_2 - t_1}\chi \\ &\equiv \dot{R}\chi = \left(\frac{\dot{R}}{R}\right) R\chi \equiv H(t)d \end{aligned} \tag{59}$$

Hubble parameter

where the **Hubble parameter** $H(t)$ is defined as

$$H(t) \equiv \frac{\dot{R}(t)}{R(t)} \qquad \text{(Hubble parameter)} \tag{60}$$

We can expect the Hubble parameter to have different values at different t-values during the evolution of the Universe. Its current value is given the symbol $H_0 \equiv H(t_0)$.

As noted in Section 14.6, astronomers cannot measure d directly. Instead they measure d_{L} or d_{A}. When either of these is plotted against redshift z, the resulting relation is linear only for $z \ll 1$. At high redshift the behavior depends on the detailed form of the scale function $R(t)$.

We need radial function $R(t)$.

We have milked about as much information out of the Robertson-Walker metric as we can without knowing the t-development of the scale function $R(t)$, which derives from the constituents of the Universe as it expands. The following Chapter 15, Cosmology, develops this scale function from a combination of observed redshifts (28) using standard candles at different distances and further solutions of Einstein's equations. The result provides our current picture of the history of the Universe and gives us insight into its possible futures.

14.8 ■ EXERCISES

1. Tangential Momentum

Carry out the full derivation of the tangential momentum Q_ϕ in equation (41), including equations similar to (32) through (38) and a figure similar to Figure 7.

2. Energy not a Constant of Motion

Show that a derivation of the energy as a constant of motion is not possible. Begin by varying only the t-value of the central event in Figure 7. What derails this derivation, making it impossible to complete?

3. Transverse Motion

A galaxy is five billion light-years distant. The most sensitive microwave array can detect a displacement angle as small as 50 microarcseconds transverse to the radial direction of sight. (One second of arc is 1/3600 of a degree.) With what transverse speed, as a fraction (or multiple) of the speed of light, must the distant source move in order that its transverse motion be detected in a 100-year human lifetime? Assume the Universe is flat.

5. Hubble's Error

Compare the value of the slope in Figure 11 with the modern value of Hubble's constant given in equations (54) and (55). By what factor was Hubble's result different from the current value of the Hubble constant?

6. 'Distance' and 'velocity' in Hubble's Law

Section 14.7 states that Hubble *found a direct proportion between the average recession velocity of a star and its distance,* which violates our rule to avoid words like *distance* when we describe observations in curved spacetime.

A. Review Section 14.7 and explain why the word *distance* does not have a unique meaning in this case.
B. Explain why the word *velocity* does not have a unique meaning.
C. Does the relative velocity of two *distant* objects have a unique meaning in curved spacetime? in flat spacetime?
D. Rewrite the Section 14.7 statement of Item A to avoid difficulties of words like *velocity* and *distance*.

14.9 ■ REFERENCES

Richard Dawkins quoted in *A Universe from Nothing* by Lawrence M. Krauss, New York, Atria Paperback, 2012, page 187.

Final cartoon by Jack Ziegler in the New Yorker Magazine July 13, 1998. IF we use it, we need formal permission.

Figure 1 from:
http://thinkexist.com/quotes/like/once_you_can_accept_the_universe_as_being/33

Figure 4 from From the Sloan Digital Sky Survey:
http://www.sdss.org/includes/sideimages/quasar_stack.html

Picture of Edwin Hubble from the cover of Time Magazine, February 9, 1948

Figure 11 from "The Velocity-Distance Relation Among Extra-Galactic Nebulae," by Edwin Hubble and Milton L. Humason, Astrophysical Journal, Volume 74, 1931, pages 43 to 80.

For a history of non-Euclean geometry, see Steven Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*, New York, John Wiley & Sons, 1972, Chapter 1.

For an account of Hubble's discovery of distant galaxies and Hubble's Law, see *Edwin Hubble, Mariner of the Nebulae* by Gale E. Christianson, 1995 University of Chicago Press and *Edwin Hubble, The Discoverer of the Big Bang Universe* by Alexander S. Sharov and Igor D. Novikov, translated by Vitaly Kisin, 1993 Cambridge University Press.

Download file name: Ch14ExpandUniverse170331v1.pdf

Chapter 15. Cosmology

- *What does our Universe contain, beyond what we see with visible light?*
- *What is "dark matter"? Why is it called "dark"? How do we know it is there? Where do we find it concentrated?*
- *What is "dark energy"? How is it different from "dark matter"? Does it accumulate in specific locations?*
- *Does light itself, and radiation of all energies, affect the development of the Universe?*
- *The Universe is expanding, right? Is this expansion slowing down or speeding up?*
- *Will the Universe continue to expand, or recontract into a "Big Crunch"?*

Download file name: Ch15Cosmology170510v1.pdf

CHAPTER 15

Cosmology

Edmund Bertschinger & Edwin F. Taylor *

Some say the world will end in fire,
Some say in ice.
From what I've tasted of desire
I hold with those who favor fire.
But if it had to perish twice,
I think I know enough of hate
To say that for destruction ice
Is also great
And would suffice.

—Robert Frost, "Fire and Ice"

15.1 ■ CURRENT COSMOLOGY

Summary of current cosmology.

Will the Universe end at all? If it ends, will it end in fire: a high-temperature Big Crunch? Or will it end in ice: the relentless separation of galaxies that drift out of view for our freezing descendents? Both the poet and the citizen are interested in these questions.

Our golden age of cosmology

Cosmology is the study of the content, structure, and development of the Universe. We live in a golden age of astrophysics and cosmology: Observations pour down from satellites above Earth's atmosphere that scan the electromagnetic spectrum—from microwaves through gamma rays. These observations combine with ground-based observations in the visible and radio portions of the spectrum to yield a flood of images and data that fuel advances in theory and arouse public interest. For the first time in human history, data and testable models inform our view of the Universe almost all the way back to its beginning. We run these models forward to evaluate alternative predictions of our distant future.

Box 1 summarizes briefly the development of those parts of the Universe that we see. In recent decades we have been surprised by the observation that

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*
Copyright © 2017 Edmund Bertschinger, Edwin F. Taylor, & John Archibald Wheeler. All rights reserved. This draft may be duplicated for personal and class use.

Box 1. Fantasy: Present at the Creation

Want to create a fantasy? Immerse yourself in the expanding "quark soup" created at the Big Bang. This quark soup is so hot that nothing we observe today can survive: not an atom, not a nucleus, not even a proton or neutron—and certainly not you! Ignore this impossibility and take a look around.

Components of the quark soup move away from one another at many times the speed of light. How can this be? The speed limit of light is measured *in spacetime*, but spacetime itself expands after the Big Bang. No limit on that speed!

Where are you located? Then and now every observer thinks s/he is at the center of the Universe. So the early Universe inflates in all directions away from you.

The temperature of the fireball drops; the ambient energy of the soup goes down. Quarks begin to "freeze out" (condense) into elementary particles such as protons and neutrons. Later a few protons and neutrons freeze out into the **deuteron** the proton-neutron nucleus of heavy hydrogen; still later a relatively small number of helium nuclei form (two protons and one neutron). Anti-protons and anti-neutrons are created too; they annihilate with protons and neutrons, respectively, to emit gamma rays. (Why are there more protons than anti-protons in our current Universe? We do not know!)

The state of the fireball—free electrons in a soup of high-speed protons, heavy hydrogen and helium nuclei—is an example of a **plasma**. The plasma fireball is still opaque to light, because a photon cannot move freely through it; free electrons absorb photons, then re-emit them in random directions.

About 300,000 years after the Big Bang, the temperature drops to the point that electrons cascade down the energy levels of hydrogen, deuterium, and helium to form atoms.

At this moment the Universe "suddenly" (during a few tens of thousands of years on your wristwatch) becomes transparent, which releases light to move freely.

From your point of view—still at your own "center of the Universe"—the surrounding Universe does not become transparent instantaneously; light from a distant source still reaches you after some lapse in t. Instead you see the wall of plasma moving away from you at the speed of light. How can plasma move with light speed? The plasma wall is moving *through* the plasma, which is riding at rest in expanding spacetime. The "wall of plasma" is not a *thing*; at sequential instants you see light emitted sequentially from electrons farther and farther from you as these electrons drop into nuclei to form neutral atoms.

As the firewall recedes from you, you see it cooling down. Why? Because atoms in the firewall are moving away from you; the farther the light has to travel to you, the faster the emitting atoms moved when they emitted the light that you see now. Greater time on your wristwatch means longer wavelength (lower frequency) of the background radiation surrounding you.

Fast forward to the present. Looking outward in any direction, you still see the firewall receding from you as it passes through the recombining plasma at the speed of light, but now Doppler down-shifted in temperature to 2.725 degrees Kelvin in your location. Welcome to our current Universe!

Dark matter and dark energy

only about four percent of the Universe is visible to us. Rotation and relative motion of galaxies, along with expansion of the Universe itself, appear to show that 23 percent of our Universe consists of **dark matter** that interacts with visible matter only through gravitation. Moreover, the present Universe appears to be increasing its rate of expansion due to a so-far mysterious **dark energy** that composes 73 percent of the Universe. If current cosmological models are correct, the accelerating expansion will continue indefinitely. The present chapter further analyzes this apparently crazy prediction.

Study constituents of the Universe

Major goals of current astrophysics research are (1) to find more accurate values of quantities that make up the Universe as a whole, (2) to explore the nature of dark matter, which evidently accounts for about 23 percent of the mass-energy in the Universe, and (3) to explore the nature of dark energy, which makes up about 73 percent. Everything we are made of and can see and touch accounts for only four percent of the mass of the Universe. This consists of protons and neutrons in the form of atoms and their associated electrons—called **baryonic matter** because its nuclei are made of protons and neutrons, which are called baryons.

Einstein's general relativity fail?

In this chapter we continue to apply Einstein's general relativity theory to cosmological models. It is possible that Einstein's theory fails over the vast cosmological distances of the Universe and during its extended lifetime. If so, dark matter and dark energy may turn out to be fictions of this outmoded theory. But so far Einstein's theory has not failed a clear test of its correctness. Therefore we continue to use it as the theoretical structure for our rapidly-developing story about the history, present state, and future of the Universe.

15.2 ■ FRIEDMANN-ROBERTSON-WALKER (FRW) MODEL OF THE UNIVERSE

Einstein's equations tell us how the Universe develops in t.

How does $R(t)$ vary with t?

Answer with scale factor $a(t)$.

Chapter 14 introduced the Robertson-Walker metric, expressed in co-moving coordinates χ and ϕ, and the set of functions $S(\chi)$ that embody the curvature of spacetime. We assumed this spacetime curvature to be uniform—on average—throughout the Universe. The Robertson-Walker metric contains the undetermined t-dependent $R(t)$ and cannot provide a cosmological model until we know how $R(t)$ develops with t. Our task in the present chapter is to find an equation for $R(t)$ and to use it to describe the past history and to evaluate possible alternative futures of the Universe. In order to simplify the algebra that follows, we introduce a dimensionless **scale factor** $a(t)$ equal to the function $R(t)$ at any t divided by its value $R(t_0)$ at present, t_0:

$$a(t) \equiv \frac{R(t)}{R(t_0)} \qquad (\textbf{scale factor: } t_0 \equiv \text{ now on Earth}) \qquad (1)$$

Friedmann equation

FRW cosmology

In 1922 Alexander Alexandrovich Friedmann combined the Robertson-Walker metric with Einstein's field equations to obtain what we now call the **Friedmann equation**, which relates the rate of change of the scale factor to the total mass-energy density ρ_{tot}, assumed to be uniform on average, throughout the Universe. Even though uniform in space, the mass-energy density is a function of the t-coordinate, $\rho_{\text{tot}}(t)$. The resulting model of the Universe is called the **Friedmann-Robertson-Walker model** or simply the **FRW cosmology**. The Friedmann equation is:

$$\boxed{H^2(t) \equiv \left[\frac{\dot{R}(t)}{R(t)}\right]^2 \equiv \frac{\dot{a}^2(t)}{a^2(t)} = \frac{8\pi\rho_{\text{tot}}(t)}{3} - \frac{K}{a^2(t)} \qquad \text{(Friedmann equation)} \qquad (2)}$$

where K is the constant parameter in the Robertson-Walker space metric of Chapter 14, with the values $K > 0$, $K = 0$, or $K < 0$ for a closed, flat, or open Universe, respectively. A dot over a symbol indicates a derivative with respect to the t-coordinate, in this case the t-coordinate read directly on the wristwatches of co-moving galaxies. In the present chapter we describe the different constituents that add up to the total $\rho_{\rm tot}(t)$.

Hubble parameter $H(t)$ varies with t.

The Friedmann equation (2) also contains a definition of the Hubble parameter $H(t)$, introduced in Chapter 14. The Hubble parameter changes as the scale factor $a(t)$ evolves with t. Remember: *When you see H, it means $H(t)$.* In this chapter we almost always use the value of H at the present t_0 and give it the symbol H_0.

$H(t_0) \equiv H_0$ is its value now

$$H_0 \equiv H(t_0) \qquad \text{(Hubble parameter, now on Earth)} \qquad (3)$$

Comment 1. An aside on units

In the Friedmann equation (2), R, t, and mass are all measured in meters; $a(t)$ is dimensionless, its t-derivative $\dot{a}(t)$ has the unit meter^{-1}, and density $\rho_{\rm tot}$ has the units of (meters of mass)/meter^3 = meter^{-2}. If you choose to express everything in conventional units, such as mass in kilograms, then the Friedmann equation becomes (using conversion factors inside the front cover):

$$H^2(t) \equiv \frac{\dot{a}^2(t)}{a^2(t)} = \frac{8\pi G}{3}\rho_{\rm tot}(t) - \frac{Kc^2}{a^2(t)} \qquad (4)$$

(Friedmann equation, conventional units)

For simplicity we use equation (2) in what follows.

Write equation (2) in a form that shows how expansion (that stretches space, described by H) fights with density (that curves spacetime due to $\rho_{\rm tot}$) to determine the value of K.

$$K = a^2(t)\left[\frac{8\pi}{3}\rho_{\rm tot}(t) - H^2(t)\right] \qquad (5)$$

Einstein links geometry with energy.

A large density $\rho_{\rm tot}$ in (5) tends to increase the value of K, increasing positive curvature of the Universe. In contrast, a large expansion rate H tends to lower the value of K, decreasing the positive curvature of the Universe. In all cases, $\rho_{\rm tot}(t)$ and $H(t)$ vary together so as to make K independent of t. This remarkable coincidence reflects the local conservation of energy: $(Ha)^2$ is proportional to the "kinetic energy" of a co-moving object in an expanding Universe, while the term proportional to density in equation (5) is proportional to minus the "gravitational potential energy" of that object. Thus the Einstein field equations link geometry and energy.

Critical density $\rho_{\rm crit}$ yields flat spacetime

We need a benchmark value for the density $\rho_{\rm tot}$, something with which to compare observed values. A useful reference density is the **critical density** $\rho_{\rm crit}(t)$, which is the total density for which spacetime is flat, a condition described by the value $K = 0$. For densities greater than the critical density ($\rho_{\rm tot} > \rho_{\rm crit}$) the Universe has a closed geometry ($K > 0$). For densities less than the critical density ($\rho_{\rm tot} < \rho_{\rm crit}$) the Universe has an open geometry ($K < 0$). The Friedmann equation (2) shows that the Hubble parameter H is a function of t. Therefore the critical density also changes with t. We define the **critical density now** as $\rho_{\rm crit,\,0}$, determined by the Hubble constant H_0, the present value of the Hubble parameter. Substitute this value and $K = 0$ into the Friedmann equation (2) to obtain:

$$\rho_{\mathrm{crit},\,0} \equiv \frac{3H_0^2}{8\pi} \qquad \text{(critical density for flat spacetime, now on Earth)} \qquad (6)$$

The ratio of total density to critical density (for flat spacetime) now on Earth is a parameter used widely in cosmology. We give this parameter the Greek symbol capital omega, Ω:

$$\Omega_{\mathrm{tot},\,0} \equiv \frac{\rho_{\mathrm{tot}}(t_0)}{\rho_{\mathrm{crit},\,0}} \qquad (7)$$

Throughout this chapter, we retain the subscript zero as a reminder that we mean the density measured now relative to the critical value now on Earth. Combining equations (5), (6), and (7) now (when $a(t_0) \equiv 1$) gives a simple relation between the curvature parameter K and density parameter $\Omega_{\mathrm{tot},\,0}$:

$$K = H_0^2(\Omega_{\mathrm{tot},\,0} - 1) \qquad \text{(now on Earth)} \qquad (8)$$

QUERY 1. Value of the critical density now on Earth

A. Estimate the numerical value of the critical density in equation (6) in units of (meters of mass)/meter^3 = meter^{-2}. For the value of H_0 see equation (28) and equations later in this chapter.

B. Express your estimate of the value of the critical density in kilograms per cubic meter.

C. Express your estimate of the value of the critical density as a fraction of the density of water (one gram per cubic centimeter).

D. Express your estimate of the value of the critical density in units of hydrogen atoms (effectively, protons) per cubic meter.

Find t-variation of density components.

The Friedmann equation (2) relates the rate of change of the scale factor $a(t)$ to the contents of the Universe. Before we can solve this equation for $a(t)$, we need to list the contributions to the total density ρ_{tot} and determine the t-dependence of each. Section 15.3 catalogs the different contents of the Universe and describes how each of them varies with scale factor $a(t)$. After further analysis, Section 15.7 returns to observations that detail estimated amounts of these different components.

15.3 ■ CONTENTS OF THE UNIVERSE I: HOW DENSITY COMPONENTS VARY WITH SCALE FACTOR $a(t)$

Matter, radiation, and dark energy.

The Friedmann-Robertson-Walker model of the Universe has been widely accepted for 40 years, but recent observations have significantly modified our picture of the contents of the Universe. Such is the excitement of being at the research edge of so large a subject.

Universe composed of matter, radiation, and dark energy.

We group the contents of the Universe into three broad categories: matter, radiation, and dark energy. Each category is chosen because of the way its contribution to the total density changes as the Universe expands. We describe these changes in terms of the scale factor $a(t)$, leaving until later (Section 15.6) the derivation of the way this scale factor changes with t.

Matter

Matter: stars, gas, neutrinos, and dark matter.

The first category we refer to as **matter**. By matter we mean particles or nonrelativistic objects with mass much greater than the mass-equivalent of their kinetic energy. Objects in this category are:

- **STARS**, including white dwarfs, neutron stars, and black holes.
- **GAS**, mostly hydrogen, with a smattering of other elements and dust.
- **NEUTRINOS**, very light particles recently determined to have a small mass. Neutrinos are produced, among other ways, by the decay of free neutrons.
- **DARK MATTER**, the non-luminous stuff, as yet unidentified, that makes up most of the matter in the Universe.

Stars and gas: mostly protons & neutrons.

Stars, interstellar gas, and dust are made of atoms. Cosmologists sometimes call atomic matter **baryonic** matter because most of the mass is made of baryons—largely protons and neutrons. The mass of an electron is negligible compared to the mass of an atomic nucleus, so even though the electron is not technically a baryon (its technical classification: **lepton**), this distinction is unimportant when counting mass.

What we see: 4% of Universe.

Current observations lead to the estimate that **luminous matter**, the stars we can see, make up about one percent of the density of the Universe, with stars and gas together totaling four percent. What a surprise that all the stars, individually and in galaxies and groups of galaxies, taken together, have only a minor influence on the development of the Universe! Yet observation forces us to this conclusion.

Neutrino mass is negligible.

Cosmic background neutrinos have not been directly detected, but their presence is inferred from our understanding of nuclear physics in the early Universe. They contribute at most a small fraction of one percent to all the mass in the Universe.

Dark matter is currently estimated to account for approximately 23 percent of the mass-energy of the Universe. What is dark matter? And how do we know that it contributes so large a fraction? We do not know what dark matter is, but from observations we infer its density and some of its properties. From the **rotation curves** of galaxies (the tangential velocities of gas as a function of R—Figure 5) we can derive the magnitude of gravitational forces needed to keep the galaxies from flying apart, and, by implication, the amount and distribution of matter in galaxies. The results (Section 15.8) show that luminous matter in a galaxy, which of course is all that we can observe directly, typically provides only a few percent of the mass required to bind the galaxy together. Dark matter was originally postulated in the 1970s to complete the total needed to hold each galaxy together as it rotates. Observations on the dynamics of galaxy clusters—first made in the 1930s and greatly refined in the 1980s and 1990s—provide further evidence for the presence of dark matter.

Dark matter holds galaxies together.

The energy density nE of a gas of particles (whether particles of baryonic matter or dark matter) is the number density n of the particles times the energy E per particle. For nonrelativistic matter, the energy per particle is well approximated by its mass m, so the energy density of matter becomes $\rho_{\mathrm{mat}} = nm$. The mass of the particle is a constant (independent of the expansion of the Universe). However, the number density n, the number of particles per unit volume, drops as the volume increases, varying with the scale factor as $a^{-3}(t)$, since volume is proportional to the cube of the linear dimension. By the definition in equation (1), the scale factor $a(t)$ has the value unity at the present age of the Universe t_0. Call $\rho_{\mathrm{mat},\,0}$ the value of the energy density of matter now. Then at any t we predict:

Energy density of matter varies as $a(t)^{-3}$.

$$\rho_{\mathrm{mat}}(t) = \rho_{\mathrm{mat},\,0}\, a^{-3}(t) \tag{9}$$

Equation (9) tells us that if we know the matter density today and the scale factor $a(t)$ as a function of t, we can determine the value of the energy density of matter at any other t, past or future. (Thus far we still have not found the t-dependence of $a(t)$.)

Radiation

Particles whose mass is much less than their energy earn the name **radiation**. Today the category *radiation* consists almost exclusively of photons. At much earlier times, neutrinos—relativistic particles with kinetic energy much greater than their mass—were a significant part of the radiation component.

Radiation: mass much less than energy.

At the present stage of the Universe, radiation is a whisper, but it used to be a shout. Shortly after the Big Bang, radiation contributed the dominant fraction of the mass-energy density of the Universe. In the hot ionized plasma of the early Universe, radiation and matter were tightly coupled: photons continually scattered from free electrons, so photons could not move in straight lines and escape. About 300 000 years after the Big Bang, however, the Universe cooled to a temperature of about 3 000 K, at which electrons combined with protons to create hydrogen gas (with some helium and a trace amount of lithium). This period is called **recombination**, even though the stable electron-nucleus combination was taking place for the first time. At

Recombination: Universe becomes transparent.

recombination, the Universe became transparent to radiation, and photons were essentially decoupled from matter, free to stream across the Universe unimpeded. The cosmic microwave background radiation that we observe in all directions is a view of that early transition from opaque to transparent, with later expansion lowering our observed temperature to 2.725 degrees Kelvin. It is remarkable that the low-energy photons we detect as background radiation between the stars have been streaming freely for billions of years, not interacting with anything until they enter our detectors.

Universe expansion reduces photon energy as well as density . . .

The number of photons emitted by all the stars in the history of the Universe is tiny compared with the number of photons created in the hot Big Bang. In the early Universe these photons were continually being emitted, absorbed, and scattered, but the *number* of photons remains approximately constant as the Universe expands. Therefore the *number of photons per unit volume* varies inversely as the scale factor cubed, or as $a^{-3}(t)$, just as the number of matter particles do. But there is an additional effect for photons. The equation $E = hf = hc/\lambda$ connects the energy E of a photon to the frequency f and wavelength λ of the corresponding electromagnetic wave. The symbol h stands for **Planck's constant**, with the value $h = 6.63 \times 10^{-34}$ kilogram-meter^2/second in conventional units. As this wave propagates through an expanding space, its wavelength increases in proportion to $a(t)$. This increased wavelength is observed as the redshift of light from distant galaxies. An increasing wavelength implies a *decrease* in the energy of each photon, an energy that varies as $a^{-1}(t)$. This leads to an extra (inverse) power of $a(t)$ compared with that for matter in equation (9) because of the drop in energy of each photon as the Universe expands. Let $\rho_{\text{rad},\,0}$ represent the energy density of radiation at t_0, the present age of the Universe. Then we predict that the radiation density obeys the equation

. . . so radiation energy density varies as $a(t)^{-4}$.

$$\rho_{\text{rad}}(t) = \rho_{\text{rad},\,0}\, a^{-4}(t) \qquad (10)$$

Dark Energy

After matter and radiation, the remaining contribution to the contents of the Universe is rather bizarre stuff which we call **dark energy**. Dark energy is entirely unrelated to *dark matter*, the major component of *matter*. Dark energy is detected only indirectly, through its effects on cosmic expansion. Its composition is unknown. Dark energy is the component of the total energy density that accounts for the observed (and surprising) current increase in the rate of expansion of the Universe. Observations described in Sections 15.7 and 15.8 lead to the estimate that approximately 73 percent of the mass-energy of the Universe is in the form of dark energy.

Dark energy composition is unknown.

QUERY 2. Energy density of radiation

The cosmic microwave background radiation has a nearly perfect blackbody spectrum with current temperature $T_0 = 2.725$ K. The temperature decreases as the Universe expands (Box 1).

$$T = T_0\, a^{-1}(t) \tag{11}$$

The energy density $u_{\rm rad}$ (energy/volume) of blackbody radiation in conventional units is given by the equation

$$u_{\rm rad} = \frac{\pi^2}{15}\frac{(k_{\rm B}T)^4}{(c\hbar)^3} \equiv a_{\rm rad}T^4 \tag{12}$$

Here $k_{\rm B}$ is the Boltzmann constant, c is the speed of light, and $\hbar \equiv h/2\pi$ where h is the Planck constant. The quantity $a_{\rm rad}$ is called the **radiation constant**.

A. Show that equations (11) and (12) are consistent with equation (10).
B. Find the present value of the energy density that corresponds to the cosmic background radiation, in kilograms per cubic meter. (We assume that the complete equivalence of energy and mass is by now second nature for you.)
C. Express your answer to part B as a fraction or multiple of the critical density, $\rho_{\rm crit,\,0}$.
D. Take the average energy of a photon in the gas of cosmic background radiation surrounding us to be $k_{\rm B}T$. Estimate the present-day number of photons per cubic meter. Compare your result with the critical mass density expressed in the number of hydrogen atoms (effectively, protons) per cubic meter.
E. At what absolute temperature T will blackbody radiation energy density be equal to the value of the critical density $\rho_{\rm crit,\,0}$ now on Earth?

Dark energy = vacuum energy?

Dark energy is a generic term which encompasses all of the various possibilities for its composition. One possibility is the so-called **vacuum energy**. We often think of the vacuum as "nothing," but that is not the picture offered by modern physics through quantum field theory, which defines the vacuum to be the state of lowest possible energy. As the Universe expands, this lowest possible vacuum energy density does not drop, but rather remains constant. Of what does vacuum energy consist? One can think of the vacuum as containing **virtual particles** that are continually being created and rapidly annihilated, according to quantum field theory. The presence of virtual particles is a well-known and well-tested consequence of the standard model of particle physics. For example, virtual particles in the surrounding vacuum have a small but detectable effect on the energy levels of hydrogen. Virtual particles surely have gravitational effects, but it has proved very difficult to correctly estimate the magnitude of these effects.

Einstein's cosmological constant

Cosmological effects of vacuum energy are described using the **cosmological constant** symbolized by the capital Greek lambda, Λ. In 1917 Einstein added this cosmological constant to his original field equations in order to make the Universe static, that is to keep it from collapsing from what he assumed must be an everlasting constant state. Einstein later removed the cosmological constant from the field equations when Hubble showed in 1929 that the Universe is expanding, but the cosmological constant continues to pop up in different theories of cosmology, as it does here as a possible source of dark energy. The presence of the cosmological constant in modern theory does *not* imply a static Universe. In the 1960s, Yakov Borisovich Zel'dovich and Erast B. Gliner showed that vacuum energy is equivalent to the cosmological constant.

Other more complicated candidates for dark energy could lead to a time-dependent energy density, but there is no current consensus about these possibilities. A full description of dark energy may have to await the development of a complete theory of quantum gravity, which does not yet exist. In this chapter we assume that dark energy does not change with t.

Assume dark energy density remains constant.

IF vacuum energy accounts for dark energy, THEN as the Universe expands the density of dark energy remains constant. We use the subscript Λ for dark energy to remind ourselves of our assumption that vacuum energy accounts for dark energy, and take ρ_Λ to be the symbol for constant dark energy:

$$\rho_{\text{dark energy}}(t) \equiv \rho_\Lambda = \text{constant} \qquad (13)$$

?

Objection 1. *Equation (13) says that the density of dark energy remains constant as the Universe expands. Result: Total dark energy increases as the Universe expands. This violates the law of conservation of energy.*

!

The law of conservation of energy says that total energy is conserved for an *isolated* system. But the term *isolated* does not apply to the Universe as a whole. By definition, the Universe contains all observable particles; it is not isolated from anything. *Result:* The law of conservation of energy does not apply to the Universe as a whole.

Table 1 summarizes the contents of the Universe and the scale factor dependence of each component. The t-independent density of dark (vacuum) energy contrasts with the density of matter, proportional to $a^{-3}(t)$, and the energy density of radiation, proportional to $a^{-4}(t)$, both of which decrease as the Universe expands. As a result, dark energy influences the development of the Universe more and more as t increases.

Variation of the total density with the scale factor $a(t)$

t-variation of total density.

We can now write an expression for the t-dependence of total density from equations (9), (10), and (13),

$$\rho_{\text{tot}}(t) = \frac{\rho_{\text{mat},0}}{a^3(t)} + \frac{\rho_{\text{rad},0}}{a^4(t)} + \rho_\Lambda \qquad (14)$$

Divide through by the critical density at the present t, equation (6), to express the result as fractions of the present critical density, as in equation (7):

TABLE 15.1 Contents of the Universe. (Subscript 0 means now.)

Contents	Consisting of	Scale variation with t
Matter	stars, gas, dark matter, (neutrinos: negligible)	$\rho_{\text{mat},0}\, a^{-3}(t)$
Radiation	photons, (earlier: neutrinos)	$\rho_{\text{rad},0}\, a^{-4}(t)$
Dark energy	cosmological constant?	$\rho_\Lambda = \text{constant}$

$$\frac{\rho_{\text{tot}}(t)}{\rho_{\text{crit},0}} = \frac{\rho_{\text{mat},0}}{\rho_{\text{crit},0}} a^{-3}(t) + \frac{\rho_{\text{rad},0}}{\rho_{\text{crit},0}} a^{-4}(t) + \frac{\rho_\Lambda}{\rho_{\text{crit},0}} \qquad (15)$$

Fractional densities Ω

We want to plot equation (15) as a function of the scale factor $a(t)$. To do this we need numerical values for the three fractional densities in that equation. These fractional densities also define contributions to the total density parameter Ω defined in equation (7).

In Section 15.7 we describe current observations that yield the approximate values:

$$\Omega_{\text{mat},0} \equiv \frac{\rho_{\text{mat},0}}{\rho_{\text{crit},0}} = 0.27 \pm 0.03 \qquad (16)$$

$$\Omega_{\Lambda,0} \equiv \frac{\rho_\Lambda}{\rho_{\text{crit},0}} = 0.73 \pm 0.03 \qquad (17)$$

In Query 9 you showed that currently on Earth the background radiation yields an energy density of approximately 5×10^{-5} times the critical density. The assumption that neutrinos have zero mass and move with the speed of light would increase this by 68% implying

$$\Omega_{\text{rad},0} \equiv \frac{\rho_{\text{rad},0}}{\rho_{\text{crit},0}} \approx 8.4 \times 10^{-5} \qquad (18)$$

We know now that neutrinos are nonrelativistic—that is, with mass—so this is not the correct value; nonetheless, their contribution to the density today is so small that the error made in equation (18) by assuming massless neutrinos is negligible.

We live between matter domination and vacuum energy domination.

Figure 1 plots equation (15) with numerical values given in equations (16) through (18). Because each of the individual quantities is proportional to a power of $a(t)$, when one component dominates the total density, ρ versus $a(t)$ is a straight line on the log-log graph. Figure 1 shows that the radiation contribution has little effect at present, but was dominant at early stages because of the multiplier a^{-4} in equation (15). For a while after the radiation-dominated era, matter had the greatest influence on the evolution of

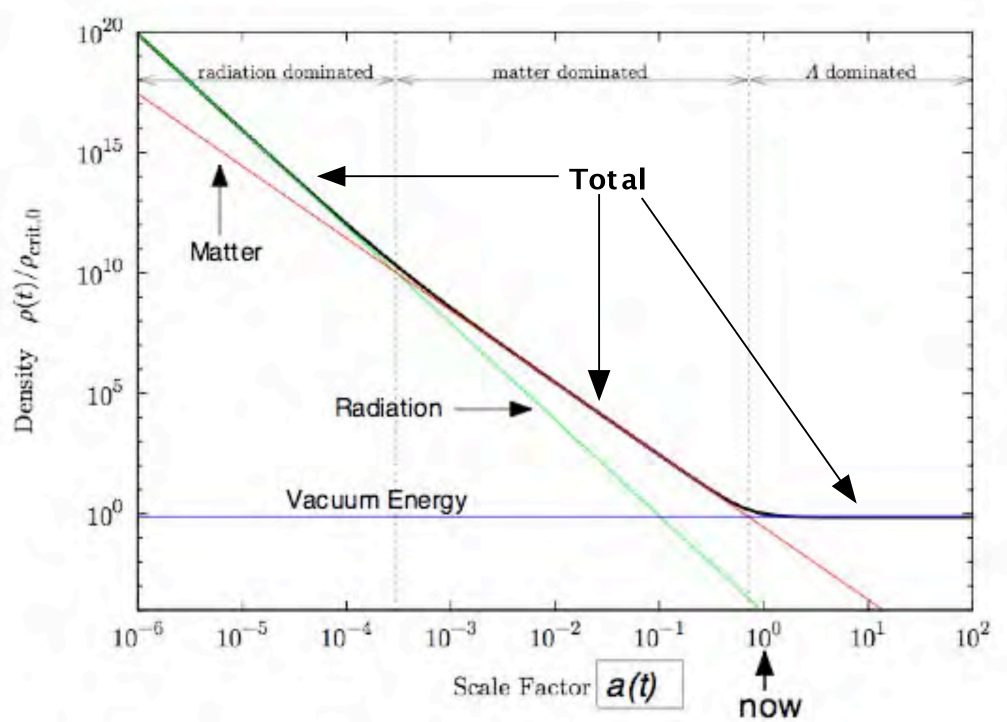


FIGURE 1 Total mass-energy density of the Universe (heavy line) in units of the present critical value as a function of the expansion scale factor. The vertical dashed lines denote transitions between the radiation-dominated early phase, the matter-dominated middle era, and the vacuum-energy-dominated late stage of the Universe. (We assume here that dark energy is vacuum energy.)

the Universe. But the influence of matter is also fading by now because of the multiplier a^{-3}. The contribution of dark energy was negligible in the distant past but has an increasing effect at the present and later stages of expansion, because its density remains constant, while densities of matter and radiation decay away with the increase in $a(t)$. If the data and assumptions behind Figure 1 are correct, we are at the beginning of the era dominated by dark energy.

QUERY 3. Contributions to the Density

A. Use equation (15) to find the approximate values of $\rho_{\text{tot}}(t)/\rho_{\text{crit},0}$ at the following times:

- at the end of the radiation-dominated era (that is, when radiation and matter make approximately equal contributions)

- at the end of the matter-dominated era (that is, when matter and dark energy make approximately equal contributions)
- now on Earth
- when $a(t) = 10^2$.

Check that your results agree with the main curve (heavy line) in Figure 1.

B. What additional information do you need in order to answer the question: How many billions of years ago did the radiation-dominated era end?

?

Objection 2. *It seems an odd coincidence that at the present moment—now in Figure 1—we are at the transition between the matter-dominated Universe and one shaped by vacuum energy. Is there a deep reason for this? Could life have developed on Earth at a different t-coordinate on the curves of Figure 1?*

!

Deep questions indeed, which we encourage you to pursue. We do not see how to answer these questions with the limited range of skills developed in this book. Also, we do not see how to move past speculation to scientific verification, mainly because we have only one Universe in which this “experiment” is taking place. We cannot (yet? ever?) do a statistical study that compares several or many Universes!

15.4 ■ UNIVERSES WITH DIFFERENT CURVATURES

Effective potential for the Universe

We can use the Friedmann equation (2), to analyze the development of alternative model Universes with different assumptions for the curvature K. To put the Friedmann equation in a more useful form, divide it through by H_0^2 and substitute for the critical density from equation (6):

$$\left(\frac{H}{H_0}\right)^2 = \left(\frac{\dot{a}}{H_0 a}\right)^2 = \frac{\rho_{\rm tot}}{\rho_{\rm crit,\,0}} - \frac{K}{H_0^2 a^2} \qquad (19)$$

t-development of the Universe

where, remember, a dot over a symbol means its derivative with respect to t. Re-express equation (19) in terms of the components of $\Omega_{\rm tot}$ defined in equations (8), (16), (17), and (18):

$$\left(\frac{H}{H_0}\right)^2 \equiv \left(\frac{\dot{a}}{H_0\, a}\right)^2 = \Omega_{\rm mat,\,0}\, a^{-3} + \Omega_{\rm rad,\,0}\, a^{-4} + \Omega_{\Lambda,\,0} - \frac{K}{H_0^2\, a^2} \qquad (20)$$

For the present, t_0, when $a(t_0) = 1$, we can write equation (20) in the very simple form:

$$1 = \Omega_{\rm mat,\,0} + \Omega_{\rm rad,\,0} + \Omega_{\Lambda,\,0} - \frac{K}{H_0^2} \qquad \text{(now, on Earth)} \qquad (21)$$

This equation allows us to determine the curvature parameter K from current measurements of $\Omega_{\mathrm{mat},0}$, $\Omega_{\mathrm{rad},0}$, and $\Omega_{\Lambda,0}$. Compare it with equation (8). Current observations lead to the conclusion that, within measurement uncertainties of about 2% in $\Omega_{\mathrm{tot},0}$, the Universe is flat ($K = 0$), in agreement with equations (16) through (18).

For any arbitrary t, we can arrange equation (20) to read:

$$\dot{a}^2 - H_0^2\left[\Omega_{\mathrm{mat},0}a^{-1} + \Omega_{\mathrm{rad},0}a^{-2} + \Omega_{\Lambda,0}a^2\right] = -K \tag{22}$$

Compare equation (22) with the corresponding Newtonian expression derived from the conservation of energy for a particle moving in the x-direction subject to a potential $V(x)$:

$$\dot{x}^2 + \frac{2V(x)}{m} = \frac{2E_{\mathrm{total}}}{m} \qquad \text{(Newton)} \tag{23}$$

In the Newtonian case we can get a qualitative feel for the particle motion by plotting $V(x)$ as a function of position and drawing a straight line at the value of E_{total}. We use equation (22) for a similar purpose, to get a qualitative feel for the evolution of the Universe. Rewrite equation (22) as:

$$\dot{a}^2 + V_{\mathrm{eff}}(a) = -K \tag{24}$$

Here the $-K$ on the right takes the place of total energy, and $V_{\mathrm{eff}}(a)$ is an effective potential given by the equation

$$V_{\mathrm{eff}}(a) \equiv -H_0^2\left[\Omega_{\mathrm{mat},0}a^{-1} + \Omega_{\mathrm{rad},0}a^{-2} + \Omega_{\Lambda,0}a^2\right] \tag{25}$$

Isn't it remarkable that effective potentials appear when we analyze orbits of a stone (Chapter 9), trajectories of light (Chapter 12), and expansion of the Universe (present chapter)?

We summarize here the assumptions on which equations (22), (24), and (25) are based.

ASSUMPTIONS FOR THE DEPENDENCE OF $\dot{a}$ ON $a(t)$

Assumptions

1. The Universe is homogeneous (on average the same in all locations).
2. The Universe is isotropic (on average the same as viewed in all directions).
3. Dark energy is vacuum energy and therefore its density is constant, independent of $a(t)$.
4. *Background assumptions*: There are no other forms of mass-energy in the Universe; spacetime has four dimensions; general relativity is correct; the Standard Model of particle theory is correct, and so on.

Figure 2 plots V_{eff}/H_0^2 as a function of $a(t)$, using the values of the densities given in equations (16), (17), and (18). For the range of $a(t)$ plotted,

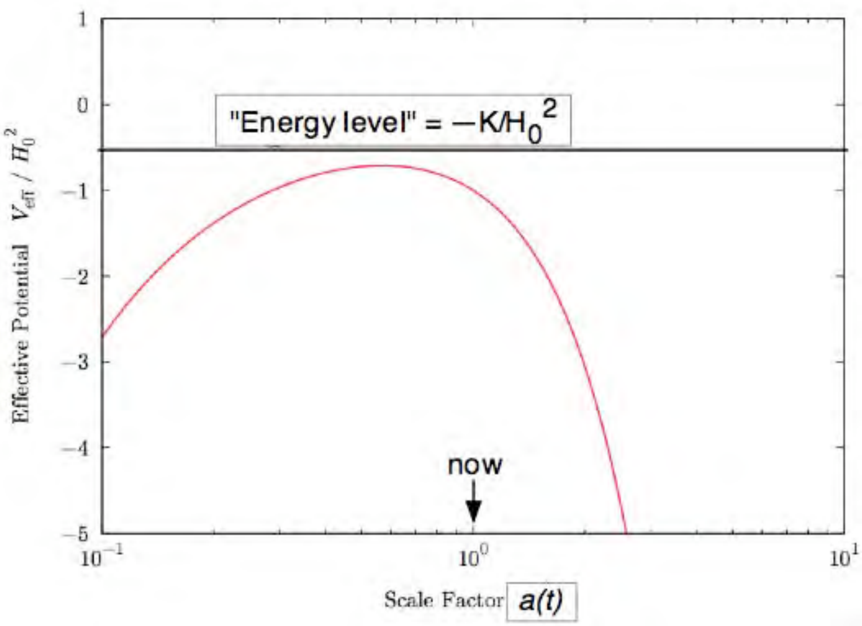


FIGURE 2 Effective potential governing the evolution of $a(t)$ according to equations (24) and (25). The "energy level" is set by $V_{\mathrm{eff}}/H_0^2 = -K/H_0^2$. The figure shows an example of a closed Universe that expands endlessly. Our Universe has $K = 0$ to a good approximation and will apparently expand without limit.

"Effective potential" for the Universe

radiation has negligible effect. Figure 2 carries a lot of information about the history and alternative futures of the Universe according to different values of K. In the Newtonian analogy, an effective potential with a *positive* slope yields a force tending to slow down positive motion along the horizontal axis, while the portion of the effective potential with a *negative* slope yields a force tending to speed up positive motion along the horizontal axis. These two conditions occur, respectively, to the left and the right of the peak at $a(t) \approx 0.57$. By analogy, then, $a(t)$ decelerates to the left of $a(t) \approx 0.57$ and accelerates to the right of $a(t) \approx 0.57$. This acceleration is due to dark energy. (*Caution:* Cosmological models described in older textbooks, written before dark energy was shown to be significant in the observed expansion of the Universe—say, before 1999—effectively assume that $\Omega_{\Lambda,0} = 0$ so the expansion does not accelerate.)

QUERY 4. The Friedmann-Robertson-Walker Universe
Figure 2 enables us to deduce many things about the history of the Universe. Answer the following questions about the predictions of this model under the assumption that the Universe begins with a Big Bang. Make a reasonable assumption about the qualitative influence of radiation on $V_{\mathrm{eff}}(a)$ for small $a(t)$.

1. True or false: The descending curve to the right of $a(t) \approx 0.57$ says that the Universe is contracting after $a(t)$ reaches this value.
2. Can the Universe be closed and expand endlessly?
3. Can the Universe be closed and recontract?
4. Can the Universe be open and expand endlessly?
5. Can the Universe be open and recontract?
6. Can the Universe be flat and expand endlessly?
7. Can the Universe be flat and recontract?
8. Describe qualitatively the evolution of a flat Universe ($K = 0$). Be specific about the evolution of $a(t)$ in the region to the right of the peak in the curve of V_{eff}/H_0^2.
9. What point on the graph of Figure 2 corresponds to a value of K that would lead to a static Universe? How could the Universe arrive at this configuration starting from a Big Bang? Is this static configuration stable or unstable, and what are the physical meanings of the terms *stable* and *unstable*?

15.5 ■ SOLVING FOR THE SCALE FACTOR

Integrating $\dot{a}(t)$

Our ignorance is stuffed into $a(t)$.

Thus far we have stuffed all our ignorance about the time development of the Universe into the scale factor $a(t)$, as given in equation (14) and plotted along the horizontal axes in Figures 1 and 2. We need to determine how $a(t)$ itself develops with time. To do this we integrate the Friedmann equation (2) as modified in equation (22). Using equations (8) and (21), rearrange (22) to read:

$$\frac{da}{dt} = H_0[\Omega_{\text{mat},0}(a^{-1} - 1) + \Omega_{\text{rad},0}(a^{-2} - 1) + \Omega_{\Lambda,0}(a^2 - 1) + 1]^{1/2} \tag{26}$$

Integrate da/dt.

By eliminating the curvature we have shown that the components of Ω_0 completely determine the expansion of the Universe—they are *important!* Now invert this equation and derive an integral with the limits from now $(a(t_0) = 1)$ to any arbitrary $a(t)$:

$$t - t_0 = \frac{1}{H_0}\int_1^a \frac{da'}{[\Omega_{\text{mat},0}(a'^{-1} - 1) + \Omega_{\text{rad},0}(a'^{-2} - 1) + \Omega_{\Lambda,0}(a'^2 - 1) + 1]^{1/2}} \tag{27}$$

Here a' is the dummy variable of integration. We can integrate equation (27) numerically from the present t_0 to either a future t ($a > 1$) or to an earlier t ($a < 1$). The Big Bang occurred when $a = 0$.

In order to carry out the integration in (27), we need to put into the integral all of our t-variations of the Ω functions. Before doing this, however, we express the constituents of (27) in convenient units. Recall that the scale factor $a(t)$ is unitless and is defined to have the value unity at present, equation (1). If we choose to express t in years, then the t-derivative $\dot{a}(t)$ will

Present value of Hubble constant

have the units years^{-1}. Then, the current value of the Hubble constant H_0 will also be expressed in the unit of years^{-1}. This is a different unit than those conventional in the field. Recent observations yield the following approximate value for H_0 in conventional units:

$$H_0 = 72 \pm 3 \ \frac{\text{kilometers/second}}{\text{Megaparsec}} \tag{28}$$

QUERY 5. Hubble parameter H_0 in years^{-1}

Use conversion factors inside the front cover to convert the units of (28) to years^{-1}. Verify that the resulting value is:

$$H_0 \approx 7.37 \times 10^{-11}\,\text{year}^{-1} \tag{29}$$

Approximate age of the Universe: $t_0 \approx H_0^{-1}$

It is not a coincidence that the quantity $H_0^{-1} = 1.36 \times 10^{10}$ years in equation (29) approximates the estimated age of the Universe: $t_0 \approx 14$ billion years. If $a(t)$ represented a linear expansion, then we would have $a = At$ for some constant A, and because $a = a(t_0) = 1$ today, the age of the Universe would be $t_0 = A^{-1}$. The Hubble constant is $H_0 \equiv \dot{a}(t_0)/a(t_0) = A$. So, for the case of linear expansion, $t_0 = H_0^{-1}$. Although the solution $a(t)$ is not linear in our Universe, $a(t_0)/t_0$ is close to $\dot{a}(t_0) = H_0$ because the Universe has recently made the transition from deceleration to acceleration. Therefore the age of the Universe approximately equals the **Hubble time** H_0^{-1}.

QUERY 6. Various kinds of Universes

Integrate equation (27) in three simplifying cases, under the assumption that spacetime is flat ($K = 0$).

A. Assume the Universe contains only matter and that $\Omega_{\text{mat},0} = 1$. Find an expression for $a(t)$ and the corresponding value of $H_0 t_0$.

B. Assume the Universe contains only radiation and that $\Omega_{\text{rad},0} = 1$. Find an expression for $a(t)$ and the corresponding value of $H_0 t_0$.

C. Assume that the Universe contains only dark energy and that $\Omega_{\Lambda,0} = 1$. Find an expression for $a(t)$.

D. *Optional.* Discuss the validity of your results for parts A, B, and C for $t < t_0$ and in particular for $t = 0$.

Integrating equation (27) requires that we know the values of the components of the total density. Remember that the total density parameter Ω_{tot} determines the curvature parameter according to equation (21). Therefore (27) has been integrated numerically for several cases, as shown in Figure 3. The model with dark energy present clearly undergoes accelerated expansion at late times.

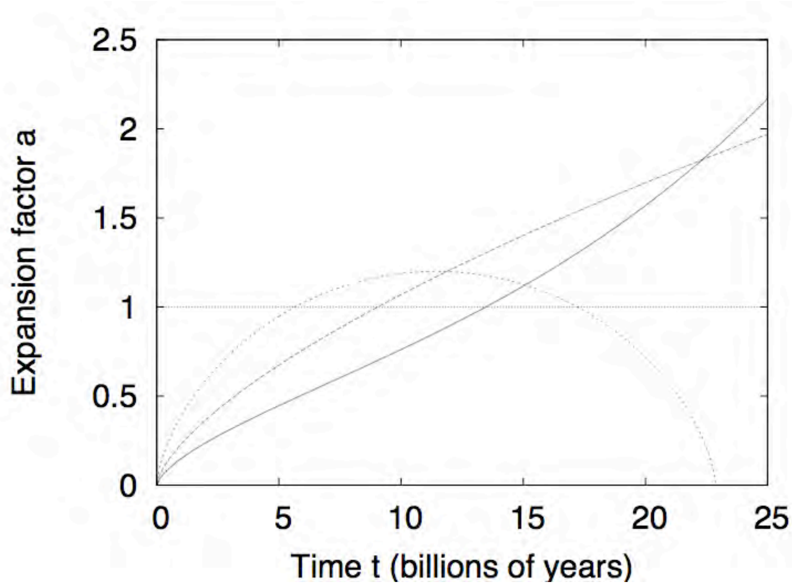


FIGURE 3 Expansion scale factor $a(t)$ versus t for three different models. The solid curve is the favored model with $\Omega_{\mathrm{mat},\,0} = 0.27$ and $\Omega_{\Lambda,0} = 0.73$. The two dotted curves show alternative models with no dark energy, $\Omega_{\mathrm{mat},\,0} = 1$ and $\Omega_{\mathrm{rad},\,0} = 1$. Can you tell which is which? The curves all have the same slope where they cross $a = 1$, because that slope is the measured current value H_0 of the Hubble constant, equations (28) and (29).

15.6 ■ LOOK-BACK DISTANCE AS A FUNCTION OF REDSHIFT

Where are earlier emitters now?

Box 4 in Section 14.5 shows that the calculated look-back distance now to an object that emitted light at t and is observed by us now is (when expressed using the scale factor)

$$d_0(t) = \int_t^{t_0} \frac{dt'}{a(t')} \qquad \text{(look-back distance, now on Earth)} \qquad (30)$$

where t' is a dummy variable. We call d_0 the **look-back distance**. In Box 4 in Section 14.4 we approximated $a(t) \approx H_0 t$ to deduce that $d_0 = 40+$ billion light years for $t = 0.7$ billion years after the Big Bang as the t-coordinate of emission. We can now improve on this estimate, using our new understanding of $a(t)$.

? **Objection 3.** *Wait! With what observations do we verify the current look-back distance of 40+ billion light years to an object that emitted light 0.7 billion years after the Big Bang?*

!

We cannot verify the current look-back distance with observation. The speed of light is finite. Right now we see the emitting object as it was 0.7 billion years after the Big Bang. We have no direct information about its condition since then. The 40+ billion light year present look-back distance is our projection under a set of assumptions about the motion of this emitter in the approximately 13 billion years since it emitted the light we see now.

QUERY 7. Look-back distance d_0 in terms of redshift z.
Because astronomers measure redshift z, not t, we rewrite (30) using the relation between redshift and expansion, equation (28) of Section 14.4, which now becomes

$$1 + z(t) = \frac{1}{a(t)} \tag{31}$$

A. Differentiate both sides of (31) and use equation (2) to write H as a function of z:

$$H(z) = -a(t)\frac{dz}{dt} \tag{32}$$

B. Substitute the result into equation (30) and show that

$$d_0(z) = \int_0^z \frac{dz'}{H(z')} \tag{33}$$

where z' is a dummy variable of integration.

Look-back d_0 *vs* z

Now we can numerically integrate equation (33) using the best-fit FRW model. Figure 4 shows the calculated "look-back" (present) distance to a galaxy with observed redshift z.

15.7 ■ WHY IS THE RATE OF EXPANSION OF THE UNIVERSE INCREASING?

Negative pressure pushes!

Acceleration $\ddot{a}(t)$ of scale factor $a(t)$

Figure 3 displays changes in the scale factor $a(t)$ of the Universe as a function of t. The slope of the curve at any point is the rate of expansion $\dot{a}(t)$ then. Changes in the slope correspond to changes in this expansion rate. We can call the rate of change of the expansion rate the *acceleration of the scale factor, symbolized by a double dot:* $\ddot{a}(t)$. Why does the Universe change its rate of expansion?

Matter-dominated era: expansion slows down.

For the matter-dominated era, one can understand that matter mutually attracts and "holds back" or "slows down" the expansion, as shown in the left-hand portion of Figure 4. But the expansion in the dark-energy-dominated era clearly violates this explanation, since the rate of expansion increases

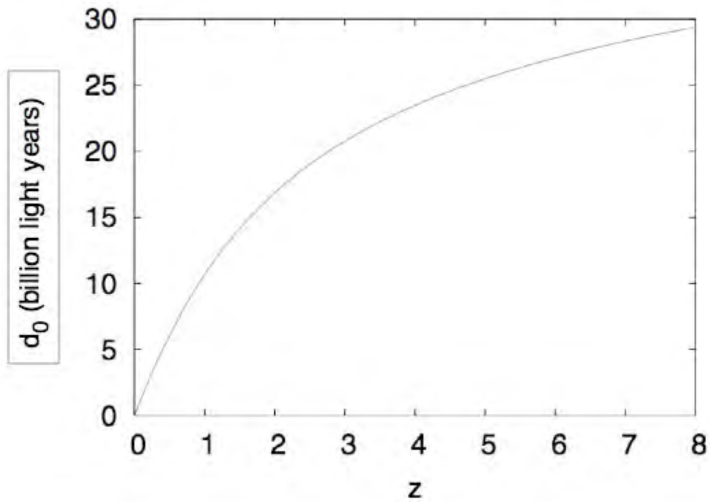


FIGURE 4 The present-day "look-back distance" d_0 to objects at redshift z. As explained in Box 4 in Section 14.4, in an expanding Universe an object that we see now (at its earlier position) is at present much farther away from us in light-years than the age of the Universe in years.

there. What is the physical reason for this increased expansion rate? This question is the subject of the present section.

Thermodynamics

Begin with some basic thermodynamics. The first law of thermodynamics says that as the volume of a box of gas increases by dV, the energy of the gas inside it decreases by an amount PdV where P is the pressure of the gas, as long as no heat flows into or out of the box. The energy change PdV goes into the work done by the gas due to its pressure acting on the outward-moving wall of the box. The energy of the gas is simply the volume that it occupies times its energy density. However, we measure energy in units of mass, so the energy density is just the mass density ρ_{tot}. Therefore we have

$$d(\rho_{\text{tot}}V) = -P_{\text{tot}}dV \tag{34}$$

Expanding gas cools.

It turns out that this relation holds whenever the volume of a gas changes, regardless of the shape of the box. It even holds when there are no walls at all! It implies a general result: an expanding gas cools.

$\ddot{a}(t)$ depends on pressure.

In Query 8 you show that the second time derivative, the acceleration $\ddot{a}(t)$ of the scale factor, depends not only on the density ρ_{tot} but also on pressure. Pressure, along with total density, appears in Einstein's field equations. In special relativity, pressure and energy density transform into each other under Lorentz transformations in a way analogous to (but not the same as) electric and magnetic fields. Energy density in one inertial frame implies pressure in another. Since the Einstein field equations are written to be valid in any frame, pressure must make a contribution to gravity (spacetime curvature). Positive pressure has an attractive gravitational effect similar to positive energy density.

Positive pressure slows expansion.

The gravitational effect of pressure may seem paradoxical: the *greater* the positive pressure, the *more negative* the value of $\ddot{a}$, the acceleration of the scale factor. We are used to watching pressure expand things like a bicycle tire. The stretching surface of an expanding balloon is often used as an analogy to the expansion of our Universe. These images can carry the incorrect implication that positive pressure is what makes the Universe expand. A balloon is expanded by pressure *differences*: the pressure inside the balloon is higher than the pressure outside combined with the balloon surface tension. Pressure differences produce mechanical forces. By contrast, we are considering a homogeneous pressure, the same everywhere—there is no "outside" of the Universe for it to expand into. There is no mechanical force of pressure in this case, only a gravitational force.

QUERY 8. Acceleration of the Scale Factor

A. Divide the energy conservation equation (34) through by dt (in other words, consider the differential energy change in an increment dt) and apply it to a local volume V that has the current value V_0 and expands (or possibly contracts) with the Universe according to the equation $V = V_0\, a^3(t)$. Show that

$$\dot{\rho}_{\rm tot} = -3\frac{\dot{a}}{a}\left(\rho_{\rm tot} + P_{\rm tot}\right) \qquad (35)$$

B. Rewrite the Friedmann equation (2) as

$$\dot{a}^2 = \frac{8\pi}{3}\rho_{\rm tot}a^2 - K \qquad (36)$$

Take the t-derivative of both sides of (36) and substitute equation (35) to obtain the equation for the acceleration of the cosmic scale factor:

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}\left(\rho_{\rm tot} + 3P_{\rm tot}\right) \qquad (37)$$

This equation predicts that for a positive total density and positive total pressure, the scale factor will decelerate with t.

Negative pressure speeds up expansion.

Here comes the big surprise. In Query 9 you show that dark energy leads to *negative* pressure. In contrast to positive pressure, negative pressure tends to *increase* the rate of expansion of the Universe. Recent observations bring evidence that we live in a Universe whose rate of expansion is increasing, not decreasing as our model would predict if only matter and radiation were present. Now for the details.

QUERY 9. Pressure from Different Sources

A. Solve equation (35) for $P_{\rm tot}$ and show that the result is:

$$P_{\rm tot} = -\frac{a}{3\dot{a}}\dot{\rho}_{\rm tot} - \rho_{\rm tot} \tag{38}$$

Equation (38) is linear in $\dot{\rho}_{\rm tot}$ and $\rho_{\rm tot}$. Therefore we can apply it separately to the different components of which $\rho_{\rm tot}$ and $P_{\rm tot}$ are composed. In parts B through D below, apply equation (38) to each component of the density to find the individual pressures due to matter, dark energy, and radiation.

B. Apply equation (38) to nonrelativistic matter for which $\rho_{\rm mat}(t) = \rho_{\rm mat,\,0}\,a^{-3}(t)$. What is the pressure $P_{\rm mat}(t)$?

C. Apply equation (38) to dark energy for which $\rho_\Lambda(t) = \text{constant} = \rho_{\Lambda,\,0}$. What is the pressure $P_\Lambda(t)$? This surprising result leads to an unavoidable fate for the Universe.

D. Finally, apply equation (38) to a gas of photons. Though we can neglect $\rho_{\rm rad}$ in describing how the Universe behaves today, Figure 1 shows that in the early Universe $\rho_{\rm rad}$ was larger than the corresponding matter term $\rho_{\rm mat}$ and could not be neglected. For radiation, $\rho_{\rm rad}(t) = \rho_{\rm rad,\,0}\,a^{-4}(t)$. What is the pressure of radiation $P_{\rm rad}(t)$?

E. Substitute your results of parts B through D into equation (37) to find an expression for $\ddot{a}$ as a function of $a(t)$:

$$\ddot{a} = -\frac{4\pi}{3}\left[\rho_{\rm mat,\,0}\,a^{-2} + 2\rho_{\rm rad,\,0}\,a^{-3} - 2\rho_{\Lambda,\,0}\,a\right] \tag{39}$$

F. Assuming that $\rho_{\rm rad,\,0}$ is negligible, show that the condition for acceleration today ($a = 1$) is

$$\Omega_{\Lambda,0} > \frac{1}{2}\Omega_{\rm mat,0} \tag{40}$$

Negative pressure? OK. Negative mass? No.

The result of part C of Query 9 tells us that the pressure of the vacuum is negative, a result unfamiliar in elementary thermodynamics. However, it is perfectly physical—neither the energy density nor the pressure of the vacuum arise from physical particles. The vacuum has constant energy density produced by quantum fluctuations. Conservation of energy—represented by equation (35)—then implies that the pressure must be negative. Negative pressure—but not negative mass density—is physically allowed.

History of changes in expansion

Equation (39) gives a history of the changes in expansion rate since the Big Bang. Early in the expansion, when the dimensionless scale factor $a(t)$ was very small, the dominant term on the right side of (39) was due to radiation, because a^{-3} was large. As $a(t)$ increased, the matter term, proportional to a^{-2}, came to dominate. These radiation and matter terms in (39) resulted in negative acceleration of $a(t)$, that is a *decrease* in the expansion rate $\dot{a}$. More recently, as $a(t)$ approached its current value one, the negative dark energy term, proportional to a, has become more and more important. At the present

MATTER: positive mass and zero pressure

age of the Universe, the net result is a positive value of the acceleration $\ddot{a}(t)$, that is an *increase* in the expansion rate $\dot{a}(t)$.

RADIATION: positive energy density and positive pressure.

DARK ENERGY: positive mass, but *negative* pressure.

What is the *physical reason* for these changes in acceleration of the dimensionless scale factor $a(t)$? Simply that matter has mass and zero pressure, while radiation energy density and pressure are both positive. Both mass and positive pressure contribute to a deceleration of $a(t)$, a decrease of $\dot{a}(t)$, as seen in (39). In contrast, dark energy contributes positive mass but *negative* pressure. The same equation shows us that negative pressure of dark energy contributes to an acceleration of $a(t)$, that is an increase in $\dot{a}(t)$, an effect that dominates as $a(t)$ becomes large.

QUERY 10. Einstein's Static Universe

Einstein introduced the cosmological constant Λ to make the Universe static according to general relativity. This constant Λ is related to ρ_Λ by

$$\rho_\Lambda = \frac{\Lambda}{8\pi G} \qquad \text{(conventional units)} \tag{41}$$

To change to units of meters, use the usual shortcut, setting $G = 1$. Then

$$\rho_\Lambda = \frac{\Lambda}{8\pi} \qquad \text{(units of meters)} \tag{42}$$

Einstein's model included only matter ρ_{mat} and the cosmological constant Λ.

A. From (36), show that $\dot{a} = 0$ and $a = 1$ (Universe always has the same scale factor as now) imply

$$K = \frac{8\pi}{3}\left(\rho_{\text{mat}} + \rho_\Lambda\right) \qquad (\dot{a} = 0) \tag{43}$$

B. From (39), show that $\ddot{a} = 0$ implies

$$\rho_{\text{mat}} - 2\rho_\Lambda = 0 \qquad (\ddot{a} = 0) \tag{44}$$

C. Combine these to deduce that Einstein's static Universe is closed, with spatial curvature

$$K = \Lambda = 8\pi\rho_\Lambda = 4\pi\rho_{\text{mat}} \qquad \text{(Einstein's static Universe)} \tag{45}$$

D. From Figure 1, show that Einstein's model is unstable. That is, any slight displacement from the maximum leads to a runaway Universe that either expands or contracts.

E. Suppose $\Lambda < 0$. Is a static Universe possible then?

Now that we have a model for the t-development of the Universe, we need to validate the assumptions that went into it, namely the values of $\Omega_{\text{mat},0}$ and $\Omega_{\Lambda,0}$ given in equations (16) and (17) along with the value of $\Omega_{\text{rad},0}$ given in equation (18). For that validation we turn to observations.

15.8 ■ CONTENTS OF THE UNIVERSE II: OBSERVATIONS

Galaxy rotation and cosmic background radiation

In this section we examine observational evidence for the quantitative amounts of the different components of our Universe: matter (visible baryonic plus dark matter), dark energy, radiation. This will allow us, in Section 15.10, to draw numerical conclusions about our Universe now and to use our present model to project these results into the past and future.

Galaxy Rotation: Evidence for Dark Matter

Evidence for dark matter.

How do we know that dark matter exists around and within galaxies? The most direct evidence comes from observing the orbits of stars or gas around a galaxy. Spiral galaxies are perfect for this exercise—their rotating disks contain neutral hydrogen gas that emits radiation with a rest wavelength of 21 centimeters. If we see the galaxy edge on, then as gas orbits the galaxy it moves directly towards us on one side of the galaxy and directly away from us on the other side. We then use the Doppler effect to measure the speed of the gas as a function of its R-value from the center of the galaxy. The result is a **rotation curve**.

Galaxy rotation curve

Figure 5 shows the rotation curve of a nearby edge-on spiral galaxy. It is quite different from a graph of the orbital speeds of planets in the Solar System, which decrease with increasing R-value from the Sun according to Kepler's Third Law. Spiral galaxies by contrast almost always have nearly-constant rotation curves at radii outside of their dense centers.

Surface luminosity density

Evidence for dark matter appears when we ask what one would *expect* the rotation curve to be if the gravitating mass were composed of only the observed stars and gas. Now think of the galaxy face-on, like a dinner plate held at arm's length, with stars rotating in circular paths at R from the center of the disk. Optical measurements of spiral galaxies show that the **surface luminosity density**, $\Sigma(R)$, varies exponentially from the center to the edge to a very good approximation:

$$\Sigma(R) = \Sigma_0 \exp(-R/h) \tag{46}$$

The surface luminosity density is defined as the total luminosity emitted along a column perpendicular to the galactic disk, taken to be the direction toward us. In this equation, sigma Σ (Greek capital S) in the function $\Sigma(R)$ simply means "surface" and is not a summation sign. The constant Σ_0 is surface luminosity density at the center of the galaxy. We assume that the galaxy is sparse enough so that light from the stars across the thickness of the disk simply adds in the direction toward us. Surface luminosity density has units of luminosity (typically watts or solar luminosities, L_{Sun}) per unit area (typically square meters or square parsecs).

The form of equation (46) has two constants: Σ_0, the central surface luminosity density, and h, the disk's scale length. For NGC 3198 the approximate values for these parameters are

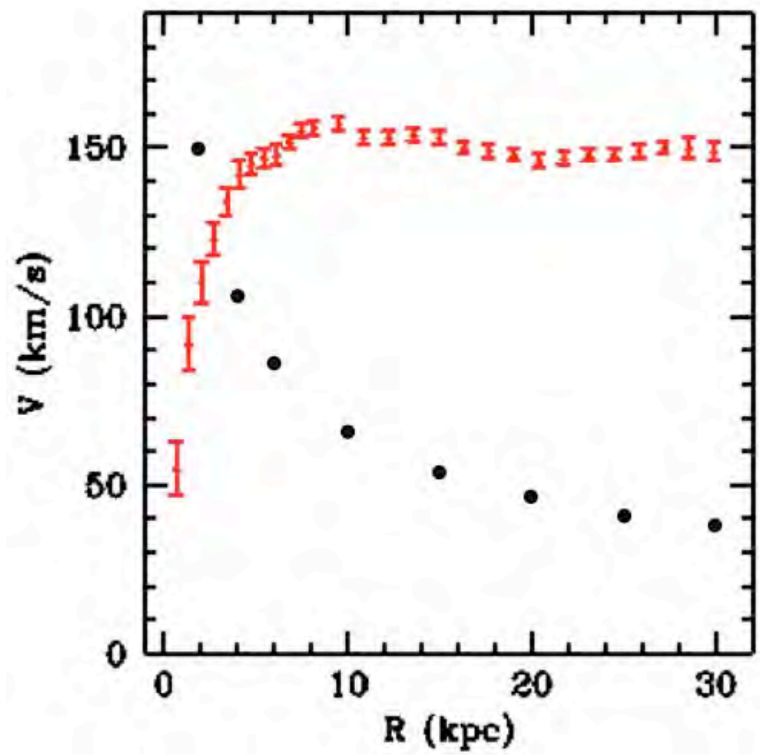


FIGURE 5 Upper plot: Rotation curve for spiral galaxy NGC 3198, from Begeman 1989, *Astronomy and Astrophysics*, 223, 47. Filled dots: Points showing the shape of a rotation curve if the attractive mass were concentrated at the center, for example in our solar system. The vertical position of the filled-dot curve depends on the value of the central mass, but the shape of the curve does not.

$$\Sigma_0 = 100\ L_{\text{Sun}}/\text{parsec}^2\ , \quad h = 2.725 \text{ kiloparsec} \tag{47}$$

One solar luminosity (L_{Sun}) is the amount of power emitted by the sun in optical light. To get the luminosity dL emitted between radii R and $R + dR$ of the galactic disk, multiply by the area of the annulus: $dL = \Sigma(R)\, 2\pi R\, dR$. The total light emitted out to R follows immediately by integration.

Mass *vs* lumnosity

To predict the rotation curve arising from luminous matter we need to know how much *mass* there is, not how much *light* the stars emit. If the luminous matter in galaxies is mainly stars like the sun, then the light in solar luminosities, L_{Sun}, equals approximately the mass in solar masses, M_{Sun}. In other words, if the total light emitted from the center out to R is $L(R)$, then the total luminous mass (stars and gas) is $M(R) = \Upsilon L(R)$ where capital Greek upsilon Υ is a factor called the **mass-to-light ratio** and whose units are solar mass per solar luminosity, that is $M_{\text{Sun}}/L_{\text{Sun}}$. If all stars in the galaxy were identical to our sun, then Υ would have the value unity. However, not all stars have the same mass-to-light ratio. A reasonable range for spiral galaxies is $0.5 < \Upsilon < 5$.

In Query 11 you apply these ingredients to show that NGC 3198 contains substantial amounts of dark matter. Make the following assumptions:

1. To describe motion of stars, assume mass density of the galaxy is spherically symmetric, but a function of R. (The tangential speed of stars in the disk has approximately the same value regardless of whether the mass is distributed in a thin disk or in a more spherical halo.)
2. Motion of stars in a galaxy can be described using Newtonian mechanics, including Newton's result that total mass inside a spherically symmetric distribution leads to a gravitational force equivalent to the force due to that total mass concentrated at the center of the sphere.
3. Stars in the galaxy move in circular orbits at a speed V that is a function of R.
4. The surface mass density follows the same function as the surface luminosity density, implying that the mass enclosed in a sphere of R is

$$M(R) = \Upsilon \int_0^R \Sigma_0 e^{-r/b}\, 2\pi r\, dr \tag{48}$$

In Query 11 you show that assumption 4 is incorrect; the galaxy contains more mass than that of its stars.

QUERY 11. Dark Matter from a Rotation Curve

With the following outline, combine Figure 5 with the surface luminosity density of equation (46), to show that the galaxy contains far more mass than can be accounted for by the stars.

A. Set up the Newtonian equation of motion and use it to find an expression for the circular speed V as a function of R, in terms of the enclosed mass $M(R)$

B. Carry out the integration in equation (48) and use it to obtain a prediction for $V(R)$. Qualitatively describe the predicted $V(R)$. Does it have a maximum value? Does it approach a nonzero constant as $R \to \infty$? If not, how does it behave for $R \gg h$, where b is in the integrand of (48)? Also, how does it behave for $R \ll h$?

C. The observed rotation curve will exceed the predicted one if there is dark matter present, which is not accounted for by equation (48). Use Figure 5 and assume that the luminous matter predominates for $R < 5$ kpc, what is the maximum mass-to-light ratio Υ for the luminous matter in NGC 3198?

D. From the results of the previous parts together with Figure 5, determine the ratio of total mass to luminous mass contained within 30 kpc from the center of NGC 3198.

Increasingly sophisticated measurements of dark matter in and around galaxies have led to a consensus range $0.2 < \Omega_{\text{mat},0} < 0.35$.

Cosmic Microwave Background Radiation

The Universe is filled with a nearly uniform glow of microwaves called the cosmic microwave background (CMB) radiation. This radiation has a **blackbody spectrum**, whose intensity as a function of frequency f is given by the Planck law, discovered in 1900 by Max Planck:

$$I(f) = \frac{2hf^3}{c^3}\,\frac{1}{e^{hf/k_\mathrm{B}T} - 1} \tag{49}$$

Radiation that has this spectrum (this dependence on frequency) is produced by an opaque medium with temperature T. The microwave background radiation fits the Planck law stunningly well—the *CO*smic *B*ackground *E*xplorer (COBE) satellite measured the spectrum to match the Planck Law to about 1 part in 10^4 in the early 1990s. Figure 6 shows the measured spectrum; the estimate of the best-fit temperature has increased by 0.001 K to $T_0 = 2.725$ K since this figure was made in 1998, where, remember T_0 is the temperature now.

Why blackbody spectrum?

At first glance, the microwave background radiation is absurd—the Universe is not opaque, and the matter that emitted the radiation was much hotter than 3 degrees above absolute zero. However, the microwave background radiation is a messenger from the early Universe, and it has aged and become stretched out during the trip. Remarkably, the form of the Planck law—the shape of the function (49) for different temperatures—is preserved by the cosmic redshift (Section 14.4). As the Universe expands, the frequency of every light wave and the temperature of the radiation decrease in proportion to $1/a(t)$. In other words, at redshift z—defined in equation (27) of Section 14.4—the radiation temperature was higher. Using equations (11) and (31), we find:

$$T(z) = (1+z)T_0 \tag{50}$$

This is an example of the way cosmologists use redshift as a proxy for increase in t since the Big Bang.

Redshift at recombination.

Most of the gas filling the Universe is hydrogen. Neutral atomic hydrogen gas is transparent to microwaves, to infrared light, and to optical light—only when the photon energy becomes large enough to ionize hydrogen does the gas become opaque. For the conditions prevailing in the Universe, hydrogen gas ionizes at a temperature comparable to that of the surface layer of cool stars, $T \approx 3000$ K. Conclusion: the microwave background radiation was produced at a redshift $z \approx 3000/2.725 = 1100$. We call the value of t at which this occurred the *recombination time* (even though it is the t-value at which electrons and protons *first* combined to make hydrogen).

Our earliest view of the Universe.

The age of the Universe at the t-value when hydrogen became transparent, t_CMB, follows from $a(t_\mathrm{CMB}) \approx 2.725/3000$. A rather complicated argument

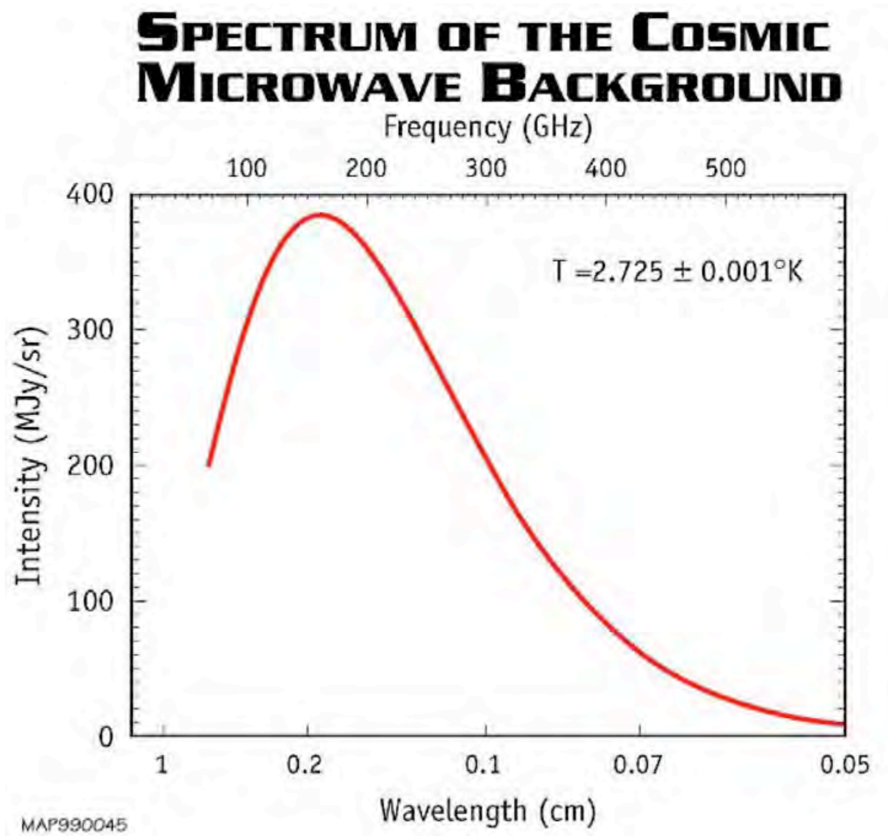


FIGURE 6 Spectrum of the cosmic microwave background radiation measured in the 1990s by the FIRAS instrument aboard the COBE satellite. (From the WMAP website.)

leads to the value $t_{\rm CMB} \approx 300\,000$ years. The CMB radiation gives us a picture of the Universe nearly 14 billion years ago. Currently this is our earliest view of the Universe; only neutrinos and gravitational waves could have penetrated the primordial plasma to bring us information from farther back toward the t-value of the Big Bang.

Objection 4. *This is hard to visualize. From where is the cosmic microwave background originating? From the direction of the center of the Universe? What direction is that?*

There is no unique center of the Universe; every observer has the impression of being at the center, as explained in Chapter 14. Looking

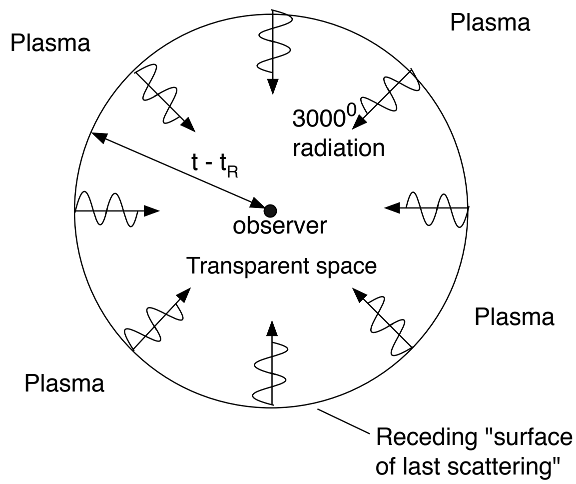


FIGURE 7 Observer's view of a non-expanding model Universe at $t - t_{\rm R}$, where $t_{\rm R}$ is the t-value, at which entire Universe becomes transparent to radiation. Looking outward, the observer cannot receive a signal from the entire Universe, but will see radiation released earlier from receding "surface of last scattering" a map distance $t - t_{\rm R}$ away. In a static Universe, this radiation would be at the recombination temperature of $\approx 3000^0$ Kelvin, approximately that of the surface of our Sun. However, in our expanding Universe (not pictured here), this radiation has been down-shifted to a temperature of 2.725^0 Kelvin, forming the cosmic microwave background radiation.

outward in every direction, we see radiation from the receding **surface of last scattering** that has been down-shifted to a temperature of 2.725^0 Kelvin, as illustrated in Figure 6.

COBE satellite

What do we see when we look at microwave radiation from the early Universe? The spectrum tells only part of the story. To see the rest, we can look at images of the sky in microwaves. The first sensitive all-sky maps of the microwave background radiation were made in the early 1990s by the COBE satellite. In 2001 a new microwave telescope called the *W*ilkinson *M*icrowave *A*nisotropy *P*robe (WMAP) was launched into orbit. It has greatly refined our picture of the early Universe.

WMAP satellite

Figure 8 shows an image of the microwave brightness around the sky made by WMAP. The Planck law is an excellent fit to the spectrum in a fixed direction of the sky; however, the temperature varies slightly in different directions. The temperature varies by a few parts in 10^5 from place to place in the early Universe. These fluctuations are, we believe, the seeds from which galaxies, stars, and all cosmic structures formed during the past 13 billion years.

Map of fluctuations: fingerprint of early Universe

In this chapter we focus on the average properties of the Universe rather than the fluctuations. However, the map of fluctuations is also a treasure trove

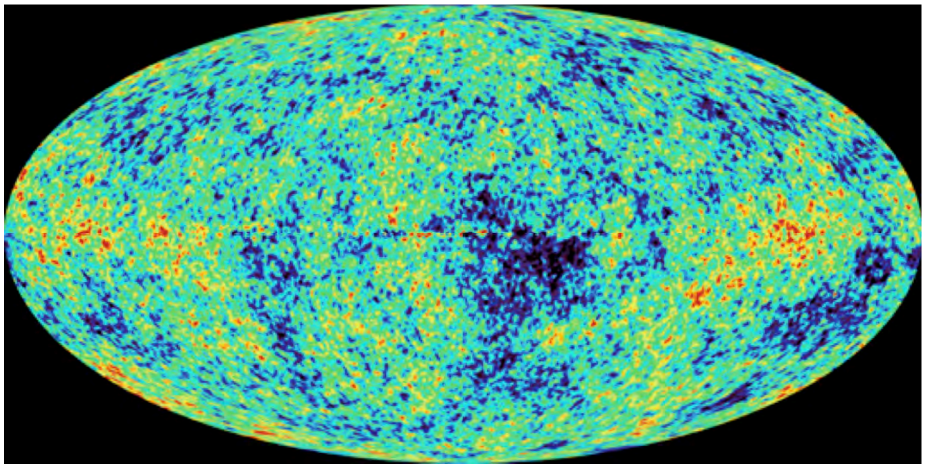


FIGURE 8 An all-sky map of the cosmic microwave background radiation at high contrast made by the WMAP satellite, with radiation from the nearby milky way stars removed. The oval is a projection of the entire sky onto the page. The colors in the original are "false colors" that indicate the temperature of the radiation ranging from $T_0 - 2 \times 10^{-4}$ K (black) to $T_0 + 2 \times 10^{-4}$ K (red) where T_0 is the average temperature. The early Universe had slight temperature variations. (Image courtesy of the WMAP Science Team, from the WMAP website.)

of information about the cosmic parameters, because the pattern of fluctuations in the sky provides a kind of fingerprint of the early Universe. For example, Figure 8 shows that the fluctuations have a characteristic angular size of about one degree.

Fluctuations due to sound waves.

The one degree scale has a direct physical significance and can be used to measure the curvature of the Universe. The fluctuations in temperature are due to sound waves in the hot gas of the early Universe: the Universe was filled with a super low frequency static created in the aftermath of the Big Bang. Sound waves compressed and rarefied the gas, changing its temperature. Sound waves oscillated in t but they also oscillated in amplitude at a given t-coordinate. The temperature fluctuations we see in the microwave background give a snapshot of the spatial variation of these sound waves 400 000 years after the Big Bang!

Sound waves: "standard ruler."

The one degree scale is a measure of how far those sound waves could travel from their creation at the big bang until $t = 400\ 000$ years, when they were revealed to us as fluctuations in the cosmic microwave background radiation. This gives us a *standard ruler*. IF we know the size of this standard ruler in meters and the distance the released radiation has since travelled to reach our telescopes—AND we know the spatial geometry (open, closed, or flat)—THEN we can predict the angular size of the fluctuations. In practice, we measure the angular size and other quantities enabling us to determine accurately the standard ruler size and the distance travelled. This method is called "baryon acoustic oscillations" (BAO). See Figure 8. The details are beyond the level of this book, but the result is not: The angular size measurement implies that the cosmic spatial curvature K is very small, consistent with zero. The spatial geometry of the Universe appears to be the simplest one possible: flat space. On the other hand, dark matter and dark energy curve space*time* in such a way that the cosmic expansion accelerates. What a strange Universe we live in!

15.9 ■ EXPANSION HISTORY FROM STANDARD CANDLES

Finding t from redshift z

To find t, measure z.

Astronomers do not directly measure $a(t)$. As discussed in Chapter 14, they measure redshift z and luminosity distance $d_{\mathrm{L}}(z)$. The observable redshift is used as a proxy for the unobservable cosmic t via equation (31). The goal here is to determine t from redshift z. From equations (31) and (32)

$$\frac{dz}{dt} = -(1+z)H(z) \tag{51}$$

where H is the Hubble parameter at t related to redshift z by equation (31). In an expanding Universe, $(1+z)H > 0$, so redshift increases looking backwards in t. If astronomers could measure $H(z)$ directly, we could integrate (51) to get $t(z)$:

$$t_0 - t(z) = \int_0^z \frac{dz}{(1+z)H(z)} \tag{52}$$

Unfortunately, $H(z)$ is very difficult to measure directly. The luminosity distance d_{L} is much easier, especially since the refinement of Type Ia supernovas as standard candles (Section 14.6). The relation between d_{L} and z can be found starting from results of Chapter 14. Along a light ray ($d\tau = 0$) coming from a distant supernova to our telescope, equation (17) of Chapter 14 gives

$$dt = -R(t)d\chi \tag{53}$$

which implies

$$\begin{aligned} R(t_0)\chi &= -R(t_0)\int_{t_0}^{t} \frac{dt'}{R(t')} \\ &= -\int_{t_0^t} \frac{dt'}{a(t')} \end{aligned} \tag{54}$$

Equation (44) of Section 14.6, with $d_{\mathrm{A}} = d_{\mathrm{L}}/(1+z)^2$ tells us that

$$\frac{d_{\mathrm{L}}(z)}{1+z} = R(t_0)S(\chi) \tag{55}$$

where $S(\chi)$ is given by equations (18) to (20) of Section 14.3. Therefore, in a flat Universe ($K = 0$, implying $S = \chi$),

$$\frac{d_\mathrm{L}(z)}{1+z} = -\int_{t_0}^{t} \frac{dt'}{a(t')} \qquad \text{(flat Universe)} \qquad (56)$$

Thus, if $d_\mathrm{L}(z)$ is measured at many different redshifts, one can determine $t(z)$ by differentiating (56) and re-integrating it again. Differentiating:

$$\frac{d}{dz}\left\{\frac{d_\mathrm{L}(z)}{1+z}\right\} = -\frac{1}{a(t)}\frac{dt}{dz} = -(1+z)\frac{dt}{dz} \qquad \text{(flat Universe)} \qquad (57)$$

then reintegrating:

$$t_0 - t(z) = \int_0^z \left[\frac{d}{dz}\left\{\frac{d_\mathrm{L}(z)}{1+z}\right\}\right]\frac{dz}{1+z} \qquad \text{(flat Universe)} \qquad (58)$$

Alternative model universes

which must be integrated numerically. More complicated formulas are required if $K \neq 0$, but the idea is similar. In practice, measurements are too imprecise to determine $d_\mathrm{L}(z)$ with enough accuracy so that equation (58) can be used directly. Instead, astronomers construct different model Friedmann-Robertson-Walker universes by adopting choices for parameters $\Omega_{\mathrm{mat},\,0}$ and $\Omega_{\Lambda,\,0}$. They integrate equation (26) to get $a(t)$, then substitute into (56) (or its generalization for a non-flat Universe) to predict $d_\mathrm{L}(z)$.

15.10 ■ THE UNIVERSE NOW: THE OMEGA DIAGRAM

Squeeze the Universe model from all sides.

Observational data from supernovas and the microwave background radiation constrain the values of $\Omega_{\mathrm{mat},0}$ and $\Omega_{\Lambda,0}$. We have already seen that radiation contributes very little to the critical density today. The major contributors are thus matter (dark matter plus baryons) and dark energy, which we model as a cosmological constant.

Squeezing the parameters

During recent years, our knowledge of the density parameter values has gone from shadowy outline to measurements of 10% accuracy. Figure 9 illustrates our current knowledge about the key parameters based on observations of Type Ia supernovas (SNe), the cosmic microwave background radiation (CMB), and the Baryon Acoustic Oscillations (BAO). The microwave background data clearly show that the Universe is close to flat, perhaps exactly so. They also imply a nonzero dark energy contribution, especially when combined with the baryon acoustic oscillations. The latter measurement is most sensitive to $\Omega_{\mathrm{mat},0}$ and indicates that there is too little matter to close the Universe. Microwave background and BAO data independently support the radical claim made by the supernova observers in 1998 that the Universe is accelerating. We found out earlier that the expansion accelerates if $\Omega_{\Lambda,0} > \frac{1}{2}\Omega_{\mathrm{mat},0}$.

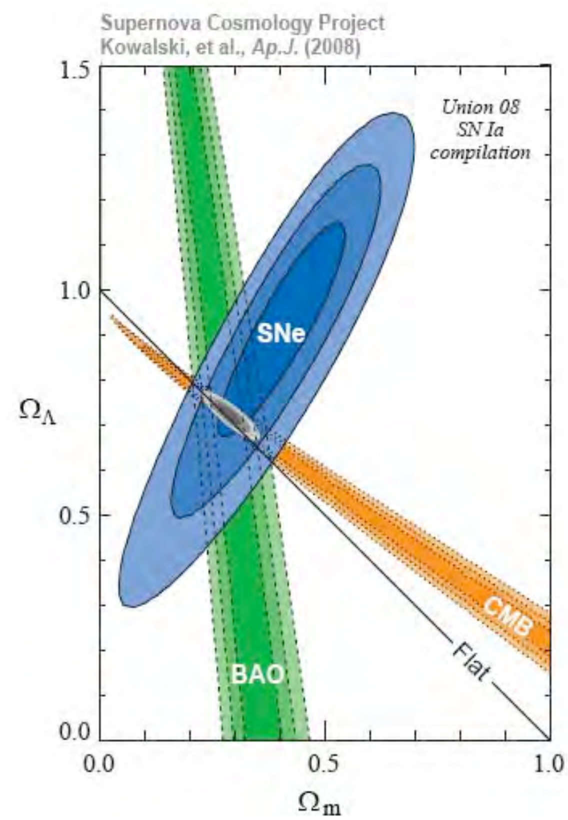


FIGURE 9 The Omega Diagram. Parameters Ω_m and Ω_Λ are called $\Omega_{mat,0}$ and $\Omega_{\Lambda,0}$ in this chapter. Relative amounts of matter and vacuum energy in the universe at present corresponds to the relatively tiny region of intersection of three sets of measurements: Type Ia supernovas (SNe), the cosmic microwave background radiation (CMB), and "baryon acoustic oscillations" (BAO). Darkest regions represent a statistical 68% confidence level and the lighter two represent statistical 95% and 99.78% confidence levels, respectively. The straight line represents conditions for a flat Universe.

Figure 9 does not include all of the constraints on the Omegas. When they are applied, the result is equations (16) and (17). Future satellite missions should shrink the uncertainties in the Omegas to less than 0.01. Once they do, we may still be left with two outstanding mysteries: What are dark matter and dark energy?

QUERY 12. No Big Bang?
Are all points on the Omega diagram allowable? Some can be excluded because they have no hot dense phase. In other words, some regions correspond to "No Big Bang."

A. Consider a FRW Universe with $\Omega_{\mathrm{mat},0} = 1$ and $\Omega_{\Lambda,0} = 3$. Neglect radiation. What are $V_{\mathrm{eff}}(a)/H_0^2$ and $-K/H_0^2$ for this case?

B. Sketch $V_{\mathrm{eff}}(a)/H_0^2$ similar to Figure 2 for the parameters of part A. Show that the Universe has a turning point in the past, so that it could not start from $a = 0$ (the Big Bang) and get to $a = 1$ (today) in this model.

C. Consider models with $\Omega_{\mathrm{mat},0} = 0$ and only dark energy with $\Omega_{\Lambda,0} > 0$. Show that these models also have a turning point at $a > 0$.

D. Show that a given model *cannot* have a Big Bang if there exists a solution $a = a_{\mathrm{min}}$ of the equation:

$$V_{\mathrm{eff}}(a) + K = 0 \qquad \text{where } 0 < a_{\mathrm{min}} < 1 \tag{59}$$

E. Show that the Universe will recollapse if there exists a solution $a = a_{\mathrm{max}}$ of (59) with $a_{\mathrm{max}} > 1$.

15.11 ■ FIRE OR ICE?

You predict the fate of the Universe.

Will the Universe end in fire or in ice? You choose the answer to this question:

Fire or ice? **You** predict.

ANSWER 1: FIRE if the temperature $T \to \infty$ for large t-values. This requires $a(t) \to 0$ for large t-values in equation (11). This happened, in effect, at the Big Bang. It will happen again if the expansion reverses, leading to a Big Crunch, that is $a \to 0$ in the future (Part E of Query 11).

ANSWER 2: ICE if the temperature $T \to 0$ for large t-values, or $a \to \infty$ as $t \to \infty$. What does Figure 2 imply for this case?

DECIDE: You are now an informed cosmologist. Choose one of Robert Frost's alternatives in his poem that began this chapter: Will the Universe end in fire or in ice?

Download file name: Ch15Cosmology170510v1.pdf

Chapter 16. Gravitational Waves

- *What are gravitational waves?*
- *How do gravitational waves differ from ocean waves?*
- *How do gravitational waves differ from light waves?*
- *What is the source (or sources) of gravitational waves?*
- *Why has it taken us so long to detect gravitational radiation?*
- *Why is the Laser Interferometer Gravitational-Wave Observatory (LIGO) so big?*
- *Why are LIGOs located all over the Earth?*
- *What will the next generation of gravitational wave detectors look like?*

Download file name: Ch16GravitationalWaves171018v1.pdf

CHAPTER 16

Gravitational Waves

Edmund Bertschinger & Edwin F. Taylor *

If you ask me whether there are gravitational waves or not, I must answer that I do not know. But it is a highly interesting problem.

—Albert Einstein

16.1 ■ THE PREDICTION AND DISCOVERY OF GRAVITATIONAL WAVES

Gravitational wave: a tidal acceleration that propagates through spacetime.

General relativity predicts black holes with properties utterly foreign to Newtonian and quantum physics. And general relativity predicts gravitational waves, also foreign to Newtonian and quantum physics.

Newton: Gravity propagates instantaneously.

Without quite saying so, Newton assumed that gravitational interaction propagates instantaneously: When the Earth moves around the Sun, the Earth's gravitational field changes all at once everywhere. When Einstein formulated special relativity and recognized its requirement that no information can travel faster than the speed of light in a vacuum, he realized that Newtonian gravity would have to be modified. Not only would static gravitational effects differ from the Newtonian prediction in the vicinity of compact masses, but also gravitational effects would propagate as waves that move with the speed of light.

Einstein: No signal propagates faster than light.

Einstein's conceptual prototype for gravitational waves was electromagnetic radiation. In 1873 James Clerk Maxwell demonstrated that the laws of electricity and magnetism predicted electromagnetic radiation. Einstein was born in 1879. Heinrich Hertz demonstrated electromagnetic waves experimentally in 1888. The adult Einstein realized that a general relativity theory would not look like Maxwell's electromagnetic theory, but he and others were able to formulate the corresponding gravitational wave equations.

Compare gravitational waves to electromagnetic waves.

What do we mean by a "gravitational wave"? The gravitational wave is a tidal acceleration that propagates; that is all it is. As a gravitational wave passes over you, you are alternately stretched and compressed in ways that

Gravitational waves propagate tidal accelerations.

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*
Copyright © 2017 Edmund Bertschinger, Edwin F. Taylor, & John Archibald Wheeler. All rights reserved. This draft may be duplicated for personal and class use.

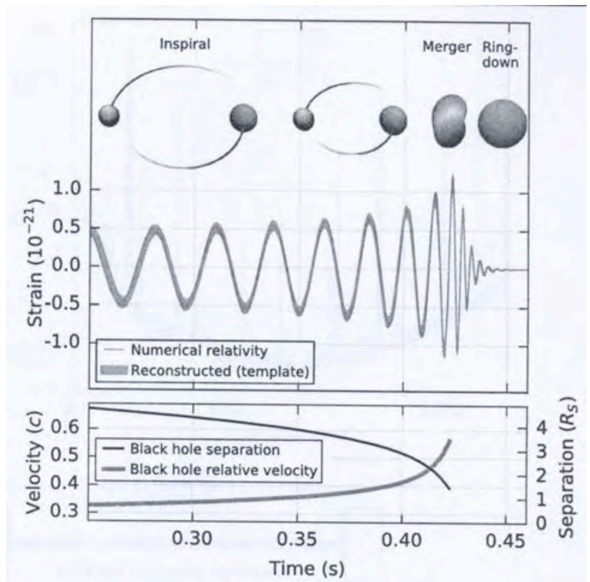


FIGURE 1 Predicted "chirp" of the gravitational wave as two black holes in a binary system merge. Frequency and amplitude increase, followed by a "ring down" due to oscillation of the merged black hole. The present chapter explains details of this figure.

depend on the particular form of the wave. In principle there is no limit to the amplitude of a gravitational wave. In the vicinity of the coalescence, gravity-wave-induced tidal forces would be lethal. Far from such a source, gravitational waves are tiny, which makes them difficult to detect.

In 2015, the most sensitive gravitational wave detector is the **Laser Interferometer Gravitational Wave Observatory**, or **LIGO** for short. Gravitational waves were first detected on 14 September 2015 with two LIGO detectors, one at Hanford, Washington state USA, the other at Livingston, Louisiana state. These detections give us confidence that gravitational waves from various sources continually sweep over us on Earth. Sections 16.3 and 16.7 describe some of these sources.

Gravitational wave on Earth: An extremely small traveling tidal effect

Basically we observe gravitational waves by detecting changes in separation between two test masses suspended near to one another—changes in gravitational-wave tidal effects. Changes in this separation are *extremely* small for gravitational waves detected on Earth.

Gravitational wave detectors are interferometers.

Current gravitational wave detectors on Earth are interferometers in which light reflects back and forth between "free" test masses (mirrors) positioned at the ends of two perpendicular vacuum chambers. A passing gravitational wave changes the relative number of wavelengths along each leg, with a resulting change in interference between the two returning waves. The "free" test masses are hung from wires that are in turn supported on elaborate shock-absorbers to minimize the vibrations from passing trucks and even ocean waves crashing on a distant shore. The pendulum-like motions of these test masses are free enough to permit measurement of their change in separation due to tidal effects of a passing gravitational wave, caused by some remote gigantic distant event such as the coalescence of two black holes modeled in Figure **??**.

Objection 1. *Does the change in separation induced by gravitational waves affect everything, for example a meter stick or the concrete slab on which a gravitational wave detector rests?*

The structure of a meter stick and a concrete slab are determined by electromagnetic forces mediated by quantum mechanics. The two ends of a meter stick are not freely-floating test masses. The tidal force of a passing gravitational wave is much weaker than the internal forces that maintain the shape of a meter stick—or the concrete slab supporting the vacuum chamber of a gravitational-wave observatory; these are stiff enough to be negligibly affected by a passing gravitational wave.

Comment 1. Why not "gravity wave"?

Why do we use the five-syllable *gravitational* to describe this waves, and not the three-syllable *gravity*? Because the term *gravity wave* is already taken. *Gravity wave* describes the disturbance at an interface—for example between the sea and the atmosphere—where gravity provides the restoring force.

16.2 ■ GRAVITATIONAL WAVE METRIC

Tiny but significant departure from the inertial metric

Our analysis examines effects of a particular gravitational wave: a plane wave from a distant source that moves in the z-direction. Every gravitational wave we discuss in this chapter (except those shown in Figure **??**) represents a very small deviation from flat spacetime. Here is the metric for a gravitational plane wave that propagates along the z-axis.

Gravitational wave metric

$$d\tau^2 = dt^2 - (1+h)dx^2 - (1-h)dy^2 - dz^2 \qquad (h \ll 1) \tag{1}$$

First, for light $d\tau = 0$. Then, as usual, no experiment or observation is global; every one is local. At the LIGO detector the local metric has the form:

$$\begin{aligned} 0 &\approx \Delta t_{\mathrm{LIGO}}^2 - [(1+h)^{1/2}\Delta x_{\mathrm{LIGO}}]^2 - [(1-h)^{1/2}\Delta y_{\mathrm{LIGO}}]^2 - \Delta z_{\mathrm{LIGO}}^2 \\ &\approx \Delta t_{\mathrm{LIGO}}^2 - [(1+h/2)\Delta x_{\mathrm{LIGO}}]^2 - [(1-h/2)\Delta y_{\mathrm{LIGO}}]^2 - \Delta z_{\mathrm{LIGO}}^2 \qquad (h \ll 1) \end{aligned} \tag{2}$$

In this metric $h/2$ is the tiny fractional deviation from the flat-spacetime coefficients of dx^2 and dy^2. The technical name for fractional deviation of length is **strain**, so $h/2$ is also called the **gravitational wave strain**. Metric (1) describes a transverse wave, since h is a perturbation in the x and y directions transverse to the z-direction of propagation. The metric guarantees that t will vary, along with x and y.

$h/2 =$ gravitational wave strain

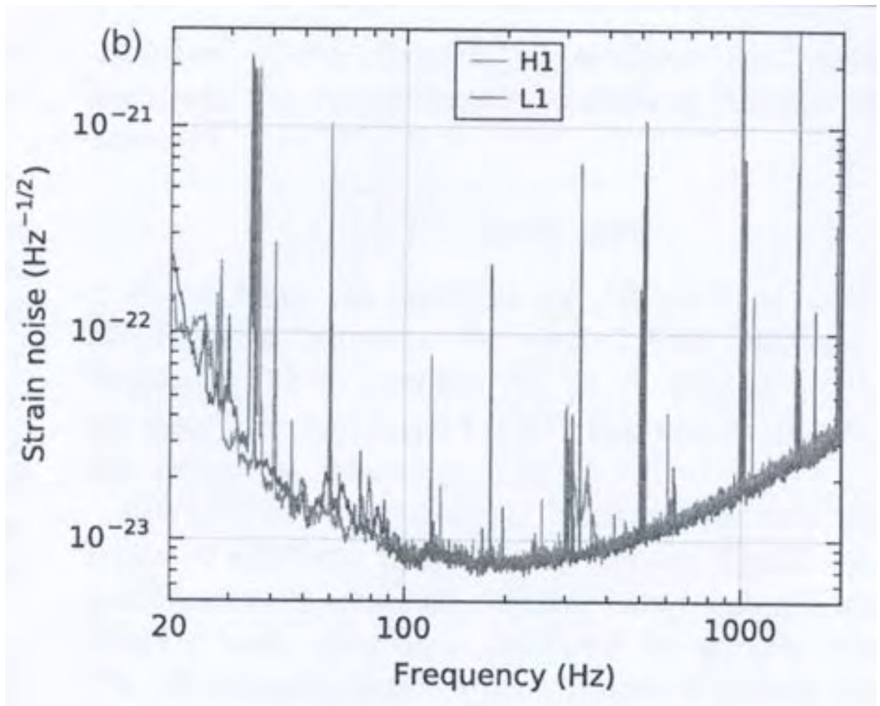


FIGURE 2 Strain noise of LIGO detectors at Hanford, Washington state (curve H1) and at Livingston, Louisiana state (curve L1) at the first detection of a gravitational wave on 14 September, 2015. On the vertical axis $h = 10^{-23}$, for example, means a fractional change in separation of 10^{-23} between test masses. Spikes occur at frequencies of electrical or acoustical noise. To be detectable, gravitational wave signals must cause fractional change above these noise curves.

Let two free test masses be at rest D apart in the x or y direction. When a z-directed gravitational wave passes over them, the change in their separation, called the **displacement**, equals $h/2 \times D$, which follows from the definition of $h/2$ as a "fractional deviation."

Objection 2. *Awkward! Why define the strain as $h/2$ instead of simply h?*

Response: This results from squared values of separation in both global and local metrics. We could use $(1-2h)$ instead of $(1-h)$ in global metric (1), but that would be awkward in another way. As usual, we get to choose the awkwardness, but cannot eliminate or ignore it!

Einstein's field equations yield predictions about the magnitude of the function h in equation (1) for various kinds of astronomical phenomena. Current gravity wave detectors use laser interferometry and go by the full name **Laser Interferometer Gravitational Wave Observatory**, or **LIGO** for short.

LIGO gravity wave detector

Figure 2 shows the noise spectrum of the two LIGO instruments that were the first to detect a gravitational wave. The displacement sensitivity is expressed in the units of meter/(hertz)$^{1/2}$ because the amount of noise limiting the measurement grows with the frequency range being sampled. Note that the instruments are designed to be most sensitive near 150 hertz. This frequency is determined by the different kinds of noise faced by experimenters: Quantum noise ("shot noise") limits the sensitivity at high frequencies, while seismic noise (shaking of the Earth) is the largest problem at low frequencies. If the range of sampled frequencies—*bandwidth*—is 100 hertz, then LIGO's best sensitivity is about $10^{-21} \times 100^{1/2} = 10^{-23}$. This means that along a length of 4 kilometers $= 4 \times 10^3$ meters, the change in length is approximately $10^{-21} \times 4 \times 10^3 = 4 \times 10^{-18}$ meters, which is one thousandth the size of a proton, or a hundred million times smaller than a single atom!

Various kinds of noise

LIGO sensitivity

Objection 3. *Your gravitational wave detector sits on Earth's surface, but equation (1) says nothing about curved spacetime described, for example, by the Schwarzschild metric. The expression $2M/r$ measures departure from flatness in the Schwarzschild metric. At Earth's surface, $2M/r \approx 1.4 \times 10^{-9}$, which is 10^{13}—ten million million!—times greater than the corresponding gravitational wave factor $h \sim 10^{-22}$. Why doesn't the quantity $2M/r$—which is much larger than h—appear in (1)?*

The factor $2M/r$ is essentially constant across the structure of LIGO, so we can ignore its change as the gravitational wave sweeps over it. LIGO is totally insensitive to the *static* curvature introduced by the factor $2M/r$ at Earth's surface. Indeed, the LIGO detector is "tuned" to detect gravitational wave frequencies near 150 hertz. For this reason, we simply omit static curvature factors from equation (1), effectively describing gravitational waves "in free space" for the predicted $h \ll 1$.

In flat spacetime and for small values of h, Einstein's field equations reduce to a wave equation for h. For the most general case, this wave has the

Einstein's equations become a wave equation.

form $h = h(t, x, y, z)$. When t, x, y, z are all expressed in meters, this wave equation takes the form:

$$\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} + \frac{\partial^2 h}{\partial z^2} = \frac{\partial^2 h}{\partial t^2} \qquad \text{(flat spacetime and } h \ll 1\text{)} \qquad (3)$$

For simplicity, think of a plane wave moving along the z-axis. The most general solution to the wave equation under these circumstances is

$$h = h_{+z}(z - t) + h_{-z}(z + t) \qquad (4)$$

Assume gravity wave moves in $+z$ direction.

The expression $h_{+z}(z - t)$ means a function h of the single variable $z - t$. The function $h_{+z}(z - t)$ describes a wave moving in the positive z-direction and the function $h_{-z}(z + t)$ describes a wave moving in the negative z-direction. In this chapter we deal only with a gravitational wave propagating in the positive z-direction (Figure 5) and hereafter set

$$h \equiv h(z - t) \equiv h_{+z}(z - t) \qquad \text{(wave moves in } +z \text{ direction)} \qquad (5)$$

The argument $z - t$ means that h is a function of *only* the combined variable $z - t$. Indeed, h can be *any function whatsoever* of the variable $(z - t)$. The form of this variable tells us that, whatever the profile of the gravitational wave, that profile displaces itself in the positive z-direction with the speed of light (local light speed = one in our units).

LIGO sensitive 75 to 500 hertz

Figure 2 shows that the LIGO gravitational wave detector has maximum sensitivity for frequencies between 75 and 500 hertz, with a peak sensitivity at around 150 hertz. Even at 500 hertz, the wavelength of the gravitational wave is very much longer than the overall 4-kilometer dimensions of the LIGO detector. Therefore *we can assume in the following that the value of h is spatially uniform over the entire LIGO detector.*

QUERY 1. Uniform h?

Using numerical values, verify the claim in the preceding paragraph that h is effectively uniform over the LIGO detector.

Analogy: draw global map coordinates on rubber sheet.

It is important to understand that coordinates in metric (1) are global and to recall that global coordinates are arbitrary; we choose them to help us visualize important aspects of spacetime. For $h \neq 0$, these global coordinates are invariably distorted. Think of the three mutually perpendicular planes formed by (x, y), (y, z), and (z, x) pairs. Draw a grid of lines on a rubber sheet lying in each corresponding plane. By analogy, the passing gravitational wave distorts these rubber sheets.

Gravitational wave distorts rubber sheet.

Glue map clocks to intersections of these grid lines on a rubber sheet so that they move as the rubber sheet distorts. A gravitational wave moving in the $+z$ direction (Figure 3) passes through a rubber sheet and acts in different

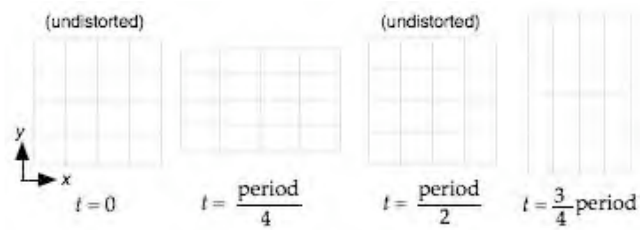


FIGURE 3 Change in shape (greatly exaggerated!) of the map coordinate grid at the same x, y location at four sequential t-values as a periodic gravitational wave passes through in the z-direction (perpendicular to the page). NOTE carefully: The x-axis is stretched while the y-axis is compressed and vice versa. The areas of the panels remain the same.

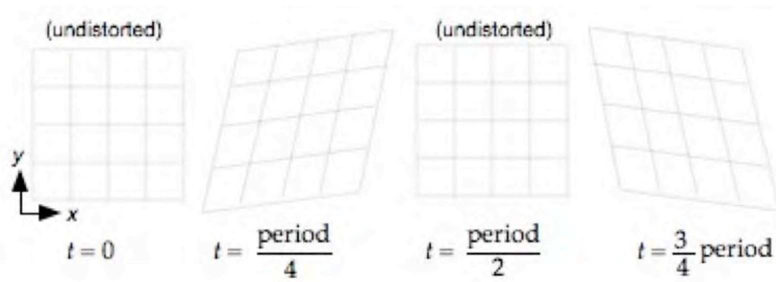


FIGURE 4 Effects of a periodic gravitational wave with polarization "orthogonal" to that of Figure 3 on the map grid in the xy plane. Note that the axes of compression and expansion are at 45 degrees from the x and y axes. All grids stay in the xy plane as they distort. As in Figure 3, the areas of the panels are all the same.

directions within the plane of the sheet (Figures 3 and 4). The map clocks glued at intersections of map coordinate grid lines ride along with the grid as the sheet distorts, so the map coordinates of any clock do not change.

Map t read on clocks glued to the rubber sheet.

Think of two ticks on a single map clock. Between ticks the map coordinates of the clock do not change: $dx = dy = dz = 0$. Therefore metric (1) tells us that the wristwatch time $d\tau$ between two ticks is also map dt between ticks. Map t corresponds to the time measured on the clocks glued to the rubber sheet, even when the strain $h/2$ varies at their locations.

Figure 3 represents the map distortion of the rubber sheet with t at a given location due to a particular polarization of the gravitational wave. Although gravitational waves are transverse like electromagnetic waves, the polarization forms of gravitational waves are different from those of electromagnetic waves. Figure 4 shows the distortion caused by a polarization "orthogonal" to that shown in Figure 3.

16.3 ■ SOURCES OF GRAVITATIONAL WAVES

Many sources; only one type leads to a clear prediction

No linear "antenna" for gravitational waves

Sources of gravitational waves include collapsing stars, exploding stars, stars in orbit around one another, and the Big Bang itself. Neither an electromagnetic wave nor a gravitational wave results from a spherically symmetric distribution of charge (for electromagnetic waves) or matter (for gravitational waves), even when that spherical distribution pulses symmetrically in and out (Birkhoff's Theorem, Section 6.5). Therefore, a *symmetric* collapse or explosion emits no waves, either electromagnetic or gravitational. The most efficient source of electromagnetic radiation, for example along an antenna, is oscillating pairs of electric charges of opposite sign moving back and forth along the antenna, the resulting waves technically called **dipole radiation**. But mass has only one "polarity" (there is no negative mass), so there is no gravity dipole radiation from masses that oscillate back and forth along a line. Emission of gravitational waves requires *asymmetric* movement or oscillation; the technical name for the simplest result is **quadrupole radiation**. Happily, most collapses and explosions are asymmetric; even the motion in a binary system is sufficiently asymmetric to emit gravitational waves.

Binary system emits gravity waves . . .

We study here gravitational waves emitted by a binary system consisting of two black holes orbiting about one another (Section 16.7). The pair whose gravitational waves were detected are a billion light-years distant, so are not visible to us. As the two objects orbit, they emit gravitational waves, so the orbiting objects gradually spiral in toward one another. These orbits are well described by Newtonian mechanics until about one millisecond before the two objects coalesce.

. . . whose amplitude is predictable.

Emitted gravitational waves are nearly periodic during the Newtonian phase of orbital motion. As a result, these particular gravitational waves are easy to predict and hence to search for. When the two objects coalesce, they emit a burst of gravitational waves (Figures **??** and 1). After coalescence the resulting black hole vibrates ("rings down"), emitting additional gravitational waves as it settles into its final state.

Comment 2. Amplitude, not intensity of gravitational waves
The gravitational wave detector measures the *amplitude* of the wave. The wave amplitude received from a small source decreases as the inverse r-separation. In contrast, our eyes and other detectors of light respond to its *intensity,* which is proportional to the square of its amplitude, so the received intensity of light decreases as the inverse r-separation.

QUERY 2. Increased volume containing detectable sources
If LIGO sensitivity is increased by a factor of two, what is the increased volume ratio from which it can detect sources?

From other sources: hard to predict.

Binary coalescence is the only source for which we can currently make a clear prediction of the signal. Other possible sources include supernovae and the collapse of a massive star to form a black hole—the event that triggers a so-called **gamma-ray burst**. We can only speculate about how far away any of these can be and still be detectable by LIGO.

Comment 3. Detectors do not affect gravitational waves
We know well that metal structures can distort or reduce the amplitude of electromagnetic waves passing across them. Even the presence of a receiving antenna can distort an electromagnetic wave in its vicinity. The same is not true of gravitational waves, whose generation requires massive moving structures. Gravitational wave detectors have negligible effect on the waves they detect.

QUERY 3. Electromagnetic waves *vs.* gravitational waves. Discussion.
What property of electromagnetic waves makes their interaction with conductors so huge compared with the interaction of gravitational waves with matter of any kind?

16.4 ■ MOTION OF LIGHT IN MAP COORDINATES

Light reflected back and forth between mirrored test masses

LIGO is an interferometer.

Currently the LIGO detector system consists of two *interferometers* that employ mirrors mounted on "test masses" suspended at rest at the ends of an L-shaped vacuum cavity. The length of each leg $L = 4$ kilometers for interferometers located in the United States. Gravitational wave detection measures the changing interference of light waves round-trip *time delays*sent down the two legs of the detector.

Suppose that a gravitational wave of the polarization illustrated in Figure 3 moves in the z-direction as shown in Figure 5 and that one leg of the detector along the x-direction and the other leg along the y-direction. In order to analyze the operation of LIGO, we need to know (a) how light propagates along the x and y legs of the interferometer and (b) how the test masses at the ends of the legs move when the z-directed gravitational wave passes over them.

Motion of light in map coordinates.

With what map speed does light move in the x-direction in the presence of a gravitational wave implied by metric (1)? To answer this question, set $dy = dz = 0$ in that equation, yielding

$$d\tau^2 = dt^2 - (1+h)dx^2 \tag{6}$$

As always, the wristwatch time is zero between two adjacent events on the worldline of a light pulse. Set $d\tau = 0$ to find the map speed of light in the x-direction.

$$\frac{dx}{dt} = \pm(1+h)^{-1/2} \qquad \text{(light moving in } x \text{ direction)} \tag{7}$$

The plus and minus signs correspond to a pulse traveling in the positive or negative x-direction, respectively—that is, in the plane of LIGO in Figure 5.

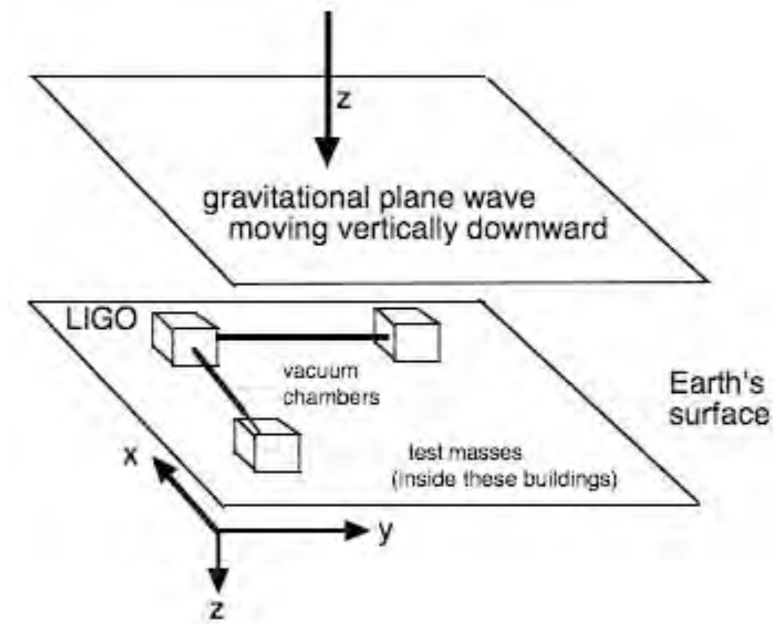


FIGURE 5 Perspective drawing of the relative orientation of legs of the LIGO interferometer lying in the x and y directions on the surface of Earth and the z-direction of the incident gravitational wave descending vertically. [Illustrator: Rotate lower plate and contents CCW 90 degrees, so corner box is above the origin of the coordinate system. Same for Figure 10.]

Remember that the magnitude of h is very much smaller than one, so we use the approximation inside the front cover. To first order:

$$(1+\epsilon)^n \approx 1 + n\epsilon \qquad |\epsilon| \ll 1 \text{ and } |n\epsilon| \ll 1 \tag{8}$$

Apply this approximation to (7) to obtain

$$\frac{dx}{dt} \approx \pm(1 - \frac{h}{2}) \qquad \text{(light moving in } x \text{ direction)} \tag{9}$$

Gravitational wave modifies map speed of light.

In words, the map speed of light changes (slightly!) in the presence of our gravitational wave. Since h is a function of t as well as x and y, the map speed of light in the x-direction is not constant, but varies as the wave passes through. (Should we worry that the speed in (9) does not have the standard value one? No! This is a *map speed*—a mythical beast—measured directly by no one.)

By similar arguments, the map speeds of light in the y and z directions for the wave described by the metric (1) are:

$$\frac{dy}{dt} \approx \pm(1 + \frac{h}{2}) \qquad \text{(light moving in } y \text{ direction)} \tag{10}$$

$$\frac{dz}{dt} = \pm 1 \qquad \text{(light moving in } z \text{ direction)} \tag{11}$$

16.5 ■ ZERO MOTION OF LIGO TEST MASSES IN MAP COORDINATES

"Obey the Principle of Maximal Aging!"

Consider two test masses with mirrors suspended at opposite ends of the x-leg of the detector. The signal of the interferometer due to the motion of light along this leg will be influenced only by the x-motion of the test masses due to the gravitational wave. In this case the metric is the same as (6).

How does the test mass move?

How does a test mass move as the gravitational wave passes over it? As always, to answer this question we use the Principle of Maximal Aging to maximize the wristwatch time of the test mass across two adjoining segments of its worldline between fixed end-events. In what follows we verify the surprising result, anticipated in Section 16.2, that a test mass initially at rest in map coordinates rides with the expanding and contracting map coordinates drawn on the rubber sheet, so this test mass does not move with respect to map coordinates as a gravitational wave passes over it. This result comes from showing that an out-and-back jog in the vertical worldline in map coordinates leads to smaller aging and therefore does not occur for a free test mass.

Idealized case: Linear jogs out and back.

Figure 6 pictures the simplest possible round-trip excursion: an incremental linear deviation from a vertical worldline from origin 0 to the event at $t = 2t_0$. Along Segment A the displacement x increases linearly with t: $x = v_0 t$, where v_0 is a constant. Along segment B the displacement returns to zero at the same constant rate. Twice the strain h has average values $\bar{h}_\text{A}$ and $\bar{h}_\text{B}$ along segments A and B respectively. We use the Principle of Maximal Aging to find the value of the speed v_0 that maximizes the wristwatch time along this worldline. We will find that $v_0 = 0$. In other words, the free test mass initially at rest in map coordinates stays at rest in map coordinates; it does not deviate from the vertical worldline in Figure 6. Now for the details.

Write the metric (6) in approximate form for one of the segments:

$$\Delta\tau^2 \approx \Delta t^2 - (1 + \bar{h})\Delta x^2 \tag{12}$$

where $\bar{h}$ is an average value of h across that segment. Apply (12) first to Segment A in Figure 6, then to Segment B. We are going to take derivatives of these expressions, which will look awkward applied to Δ symbols. Therefore we temporarily ignore the Δ symbols in (12) and let τ stand for $\Delta\tau$, t for Δt, and x for Δx, holding in mind that these symbols represent increments, so equations in which they appear are approximations.
With these substitutions, equation (12) becomes, for the two adjoining worldline segments:

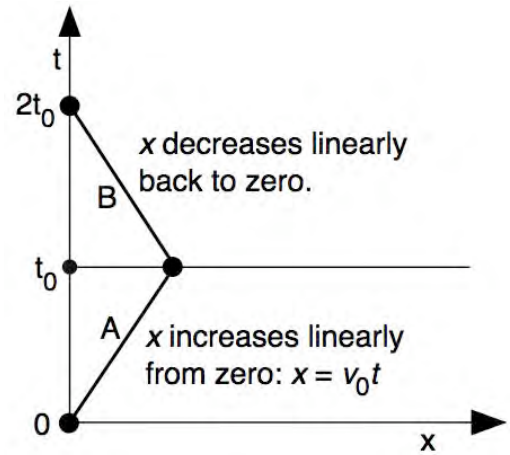


FIGURE 6 Trial worldline for a test mass; incremental departure from vertical line of a particle at rest. Segments A and B are very short.

$$\tau_{\rm A} \approx \left[t_0^2 - \left(1+\bar{h}_{\rm A}\right)\left(v_0 t_0\right)^2\right]^{1/2} \qquad \text{Segment A} \tag{13}$$

$$\tau_{\rm B} \approx \left[t_0^2 - \left(1+\bar{h}_{\rm B}\right)\left(v_0 t_0\right)^2\right]^{1/2} \qquad \text{Segment B}$$

so that the total wristwatch time along the bent worldline from $t = 0$ to $t = 2t_0$ is the sum of the right sides of equations (13).

We want to know what value of v_0 (the out-and-back speed of the test mass) will lead to a maximal value of the total wristwatch time. To find this, take the derivative with respect to v_0 of the sum of individual wristwatch times and set the result equal to zero.

$$\frac{d\tau_{\rm A}}{dv_0} + \frac{d\tau_{\rm B}}{dv_0} \approx -\frac{(1+\bar{h}_{\rm A})v_0 t_0^2}{\tau_{\rm A}} - \frac{(1+\bar{h}_{\rm B})v_0 t_0^2}{\tau_{\rm B}} = 0 \tag{14}$$

so that

$$\frac{(1+\bar{h}_{\rm A})v_0 t_0^2}{\tau_{\rm A}} = -\frac{(1+\bar{h}_{\rm B})v_0 t_0^2}{\tau_{\rm B}} \tag{15}$$

Initially at rest in map coordinates? Then stays at rest in map coordinates.

Worldline segments A and B in Figure 6 are identical except in the direction of motion in x. In equation (15), v_0 is our proposed speed in global coordinates, a positive quantity. The only way that (15) can be satisfied is if $v_0 = 0$. *The test mass initially at rest does not change its map x-coordinate as the gravitational wave passes over.*

Our result seems rather specialized in two senses: First, it treats only the vertical worldline in Figure 6 traced out by a test mass at rest. Second, it deals only with a very short segment of the worldline, along which $\bar{h}$ is considered to be nearly constant. Concerning the second point, you can think of (14) as a tiny out-and-back "jog" *anywhere* on a much longer vertical worldline. Then our result implies that *any* jog in the vertical worldline does not lead to an increased value of the wristwatch time, even if h varies a lot over a longer stretch of the worldline.

Not at rest in map coordinates? Maybe kink in map worldline.

The first specialization, the vertical worldline in Figure 6, is important: The gravitational wave does not cause a kink in a *vertical* map worldline. The same is typically *not* true for a particle that is moving in map coordinates before the gravitational wave arrives. (We say "typically" because the kink may not appear for some directions of motion of the test mass and for some polarization forms and directions of propagation of the gravitational wave.) In this more general case, a kink in the worldline corresponds to a change of velocity. In other words, a passing gravitational wave can change the map velocity of a moving particle just as if it were a velocity-dependent force. If the particle velocity is zero, then the force is zero: a particle at rest in map coordinates remains at rest.

QUERY 4. Disproof of relativity? (optional)

"Aha!" exclaims Kristin Burgess. "Now I can disprove relativity once and for all. If the test mass *moves*, a passing gravitational wave can cause a kink in the worldline of the test mass as observed in the local inertial Earth frame. No kink appears in its worldline if the test mass is at rest. But if a worldline has a kink in it as observed in one inertial frame, it will have a kink in it as observed in all overlapping relatively-moving inertial frames. An observer in any such frame can detect this kink. So the *absence* of a kink tells me *and every other inertial observer* that the test mass is 'at rest'? We have found a way to determine absolute rest using a local experiment. Goodbye relativity!" Is Kristin right? (A detailed answer is beyond the scope of this book, but you can use some relevant generalizations drawn from what we already know to think about this paradox. As an analogy from flat-spacetime electromagnetism, think of a charged particle at rest in a purely magnetic field: The particle experiences no magnetic force. In contrast, when the same charged particle moves in the same frame, it may experience a magnetic force for some directions of motion.)

At rest in map coordinates? Still can move in Earth coordinates.

In this book we make every measurement in a local inertial frame, not using differences in global map coordinates. So of what possible use is our result that a particle at rest in global coordinates does not move in those coordinates when a gravitational wave passes over it? Answer: Just because something is at rest in map coordinates does not mean that it is at rest in local inertial Earth coordinates. In the following section we find that a gravitational wave *does* move a test mass as observed in the Earth coordinates. LIGO—attached to the Earth—can detect gravitational waves!

16.6 ■ DETECTION OF A GRAVITATIONAL WAVE BY LIGO

Make measurement in the local Earth frame.

Suppose that the gravitational wave that satisfies metric (1) passes over the LIGO detector oriented as in Figure 5. We know how the test masses at the two ends of the legs of the detector respond to the gravitational wave: they remain at rest in map coordinates (Section 16.5). We know how light propagates along both legs: as the gravitational wave passes through, the map speed of light varies slightly from the value one, as given by equations (9) through (11) in Section 16.4.

Earth frame tied to LIGO slab

The trouble with map coordinates is that they are arbitrary and typically do not correspond to what an observer measures. Recall that we require all measurements to take place in a local inertial frame. So think of a local inertial frame anchored to the concrete slab on which LIGO rests. (Section 16.1 insisted that the gravitational wave has essentially no effect on this slab.) Call the coordinates in the resulting local coordinate system **Earth coordinates**. Earth coordinates are analogous to shell coordinates for the Schwarzschild black hole: useful only locally but yielding the numbers that predict results of measurements. The metric for the local inertial frame then has the form:

$$\Delta\tau^2 \approx \Delta t_{\mathrm{Earth}}^2 - \Delta x_{\mathrm{Earth}}^2 - \Delta y_{\mathrm{Earth}}^2 - \Delta z_{\mathrm{Earth}}^2 \qquad (16)$$

Compare this with the approximate version of (1):

$$\Delta\tau^2 \approx \Delta t^2 - (1+h)\Delta x^2 - (1-h)\Delta y^2 - \Delta z^2 \qquad (h \ll 1) \qquad (17)$$

Earth frame coordinate differences

Legalistically, in order to make the coefficients in (17) constants we should use the symbol $\bar{h}$, with a bar over the h, to indicate the average value of the gravitational wave amplitude over the detector. However, in Query 1 you showed that for the frequencies at which LIGO is sensitive, the wavelength is very much greater than the dimensions of the detector, so the amplitude h of the gravitational wave is effectively uniform across the LIGO detector. Therefore it is not necessary to take an average, and we use the symbol h without a superscript bar.

Compare (16) with (17) to yield:

$$\boxed{\begin{aligned}
\Delta t_{\mathrm{Earth}} &= \Delta t && && (18)\\
\Delta x_{\mathrm{Earth}} &= (1+h)^{1/2}\Delta x \approx \left(1+\frac{h}{2}\right)\Delta x && h \ll 1 && (19)\\
\Delta y_{\mathrm{Earth}} &= (1-h)^{1/2}\Delta y \approx \left(1-\frac{h}{2}\right)\Delta y && h \ll 1 && (20)\\
\Delta z_{\mathrm{Earth}} &= \Delta z && && (21)
\end{aligned}}$$

where we use approximation (8). Notice, first, that the lapse $\Delta t_{\mathrm{Earth}}$ between two events is identical to their lapse Δt and the z component of their separation in Earth coordinates, $\Delta z_{\mathrm{Earth}}$, is identical to the z component of their separation in map coordinates, Δz.

Now for the differences! Let Δx be the map x-coordinate separation between the pair of mirrors in the x-leg of the LIGO interferometer and Δy be the map separation between the corresponding pair of mirrors in the y-leg. As the z-directed wave passes through the LIGO detector, the test masses at rest at the ends of the legs stay at rest in map coordinates, as Section 16.5 showed. Therefore the value of Δx remains the same during this passage, as does the value of Δy. But the presence of varying $h(t)$ in (19) and (20) tell us that these test masses move when observed in Earth coordinates. *More:* When Δx_{Earth} between test masses increases (say) along the Earth x-axis, it decreases along the perpendicular Δy_{Earth}; and vice versa. Perfect for detection of a gravitational wave by an interferometer!

Test masses move in Earth coordinates.

Earth metric (16) is that of an inertial frame in which the speed of light has the value one in whatever direction it moves. With light we have the opposite weirdness to that of the motion of test masses initially at rest: In map coordinates light moves at map speeds different from unity in the presence of this gravitational wave—equations (9) through (11)—but in Earth coordinates light moves with speed one. This is reminiscent of the corresponding case near a Schwarzschild black hole: In Schwarzschild map coordinates light moves at speeds different from unity, but in local inertial shell coordinates light moves at speed one.

Light speed $= 1$ in local Earth frame.

In summary the situation is this: As the gravitational wave passes over the LIGO detector, the speed of light propagating down the two legs of the detector has the usual value one as measured by the Earth observer. However, for the Earth observer the separations between the test masses along the x-leg and the y-leg change: one increases while the other decreases, as given by equations (19) and (20). The result is a t-difference in the round-trip of light along the two legs. It is this difference that LIGO is designed to measure and thereby to detect the gravitational wave.

Different Earth times along different legs

What will be the value of this difference in round-trip t between light propagation along the two legs? Let D be the Earth-measured length of each leg in the absence of the gravitational wave. The round-trip t is twice this length divided by the speed of light, which has the value one in Earth coordinates. Equations (19) and (20) tell us that the difference in round-trip t between light propagated along the two legs is

$$\Delta t_{\text{Earth}} = 2D\left(\frac{h}{2}+\frac{h}{2}\right) = 2Dh \qquad \text{(one round trip of light)} \tag{22}$$

Time difference after N round trips.

Using the latest interferometer techniques, LIGO reflects the light back and forth down each leg approximately $N = 300$ times. That is, light executes approximately 300 round trips, which multiplies the detected delay, increasing the sensitivity of the detector by the same factor. Equation (22) becomes

$$\Delta t_{\text{Earth}} = 2NDh \qquad (N \text{ round trips of light}) \tag{23}$$

Quantities N and h have no units, so the unit of Δt_{Earth} in (23) is the same as the unit of D, for example meters.

QUERY 5. LIGO fast enough?
Do the 300 round trips of light take place much faster than one period of the gravitational wave being detected? (If it does not, then LIGO detection is not fast enough to track the *change* in h.)

QUERY 6. Application to LIGO.
Each leg of the LIGO interferometer is of length $D = 4$ kilometers. Assume that the laser emits light of wavelength 1064 nanometer, $\approx 10^{-6}$ meter (infrared light from a NdYAG laser). Suppose that we want LIGO to reach a sensitivity of $h = 10^{-23}$. For $N = 300$, find the corresponding value of Δt_{Earth}. Express your answer as a decimal fraction of the period T of the laser light used in the experiment.

QUERY 7. Faster derivation?
In this book we insist that global map coordinates are arbitrary human choices and do not treat map coordinate differences as measurable quantities. However, the value of h in (1) is so small that the metric differs only slightly from an inertial metric. This once, therefore, we treat map coordinates as directly measurable and ask you to redo the derivation of equations (22) and (23) using only map coordinates.

Remember that test masses initially at rest in map coordinates do not change their coordinates as the gravitational wave passes over them (Section 16.4), but the gravitational wave alters the map speeds of light, differently in the x-direction, equation (9), and in the y-direction, equation (10). Assume that each leg of the interferometer has the length D_{map} in map coordinates.

A. Find an expression for the difference Δt between the two legs for one round trip of the light.

B, How great do you expect the difference to be between Δt and Δt_{Earth} and the difference between D (in Earth coordinates) and D_{map}? Taken together, will these differences be great enough so that the result of your prediction and that of equation (23) can be distinguished experimentally?

QUERY 8. Different directions of propagation of the gravitational wave
Thus far we have assumed that the gravitational plane wave of the polarization described by equation (1) descends vertically onto the LIGO detector, as shown in Figure 5. Of course the observers cannot prearrange in what direction an incident gravitational wave will move. Suppose that the wave propagates along the direction of, say, the y-leg of the interferometer, while the x-direction lies along the other leg, as before. What is the equation that replaces (23) in this case?

QUERY 9. LIGO fails to detect a gravitational wave?

Think of various directions of propagation of the gravitational wave pictured in Figure 3, together with different directions of x and y in equation (1) with respect to the LIGO detector. Give the name **orientation** to a given set of directions x and y—the transverse directions in (1)—plus z (the direction of propagation) in (1) relative to the LIGO detector. How many orientations are there for which LIGO will detect *no signal whatever*, even when its sensitivity is 10 times better than that needed to detect the wave arriving in the orientation shown in Figure 5? Are there zero such orientations? one? two? three? some other number less than 10? an unlimited number?

16.7 ■ BINARY SYSTEM AS A SOURCE OF GRAVITATIONAL WAVES

"Newtonian" source of gravitational waves

Unequal masses, each in circular orbit

The gravitational wave detected on 15 September 2015 came from the merging of two black holes; assume that each is initially in a circular orbit around their center of mass. The binary system is the only known example for which we can explicitly calculate the emitted gravitational waves. Let the M_1 and M_2 represent the masses of these two black holes that initially orbit at a value r apart, as shown in Figure 7.

Energy of the system.

The basic parameters of the orbit are adequately computed using Newtonian mechanics, according to which the energy of the system in conventional units is given by the expression:

$$E_{\text{conv}} = -\frac{GM_{1,\text{kg}}M_{2,\text{kg}}}{2r} \qquad \text{(Newtonian circular orbits)} \tag{24}$$

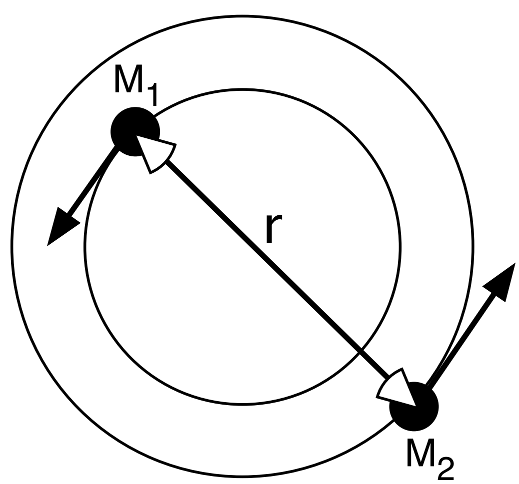


FIGURE 7 A binary system with each object in a circular path.

Rate of energy loss . . .

As these black holes orbit, they generate gravitational waves. General relativity predicts the rate at which the orbital energy is lost to this radiation. In conventional units, this rate is:

$$\frac{dE_{\text{conv}}}{dt_{\text{conv}}} = -\frac{32G^4}{5c^5r^5}\left(M_{1,\text{kg}}M_{2,\text{kg}}\right)^2\left(M_{1,\text{kg}}+M_{2,\text{kg}}\right) \qquad \text{(Newtonian circular orbits)} \tag{25}$$

. . . derived from Einstein's equations.

Equation (25) assumes that the two orbiting black holes are separated by much more than the r-values of their event horizons and that they move at nonrelativistic speeds. Deriving equation (25) involves a lengthy and difficult calculation starting from Einstein's field equations. The same is true for the derivation of the metric (1) for a gravitational wave. These are two of only three equations in this chapter that we simply quote from a more advanced treatment.

QUERY 10. Energy and rate of energy loss

Convert Newton's equations (24) and (25) to units of meters to be consistent with our notation and to get rid of the constants G and c. Use the sloppy professional shortcut, "Let $G = c = 1$."

A. Show that (24) and (25) become:

$$E = -\frac{M_1M_2}{2r} \qquad \text{(Newton: units of meters)} \tag{26}$$

$$\frac{dE}{dt} = -\frac{32}{5r^5}\left(M_1M_2\right)^2\left(M_1+M_2\right) \qquad \text{(Newton: units of meters)} \tag{27}$$

B. Verify that in both of these equations E has the unit of length.

C. Suppose you are given the value of E in meters. Show how you would convert this value first to kilograms and then to joules.

QUERY 11. Rate of change of radius

Derive a Newtonian expression for the rate at which the radius changes as a result of this energy loss. Show that the result is:

$$\frac{dr}{dt} = -\frac{64}{5r^3}M_1M_2\left(M_1+M_2\right) \qquad \text{(Newton: circular orbits)} \tag{28}$$

16.8 ■ GRAVITATIONAL WAVE AT EARTH DUE TO DISTANT BINARY SYSTEM

How far away from a binary system can we detect its emitted gravitational waves?

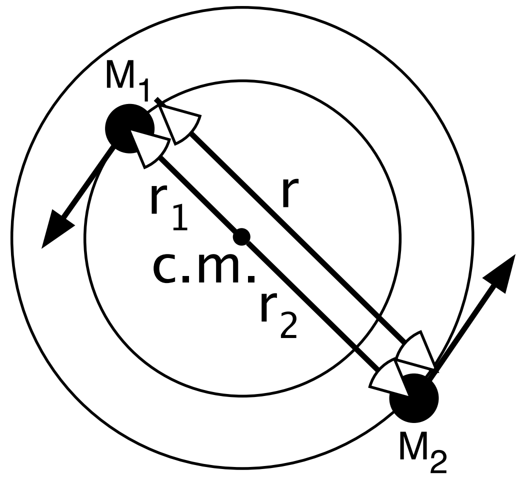


FIGURE 8 Figure 7 augmented to show the center of mass (c.m.) and orbital r-values of individual masses in the binary system.

LIGO on Earth's surface detects the gravitational waves emitted by the distant binary system of two black holes of Figure 7, augmented in Figure 8 to show the center of mass and individual r_1 and r_2 of the two black holes.

Gravitational waveform .

What is the amplitude of gravitational waves from this source measured on Earth? Here is the third and final result of general relativity quoted without proof in this chapter. The function $h(z,t)$ is given by the equation (in conventional units)

$$h(z,t) = -\frac{4G^2 M_1 M_2}{c^4 r z} \cos\left[\frac{2\pi f(z-ct)}{c}\right] \qquad \text{(conventional units)} \qquad (29)$$

where r is the separation of orbiters in Figures 7 through 9. Here z is the separation between source to detector, and—surprisingly—f is twice the frequency of the binary orbit (see Query 15). Convert (29) to units of meters by setting $G = c = 1$. Note that $h(z,t)$ is a function of z and t.

Figure 9 schematically displays the notation of equation (29), along with relative orientations and relative magnitudes assumed in the equation. This equation makes the Newtonian assumptions that

(a) the r separation between two the circulating black holes is much larger than either Schwarzschild r-value, and

(b) they move at nonrelativistic speeds.

Additional assumptions are:

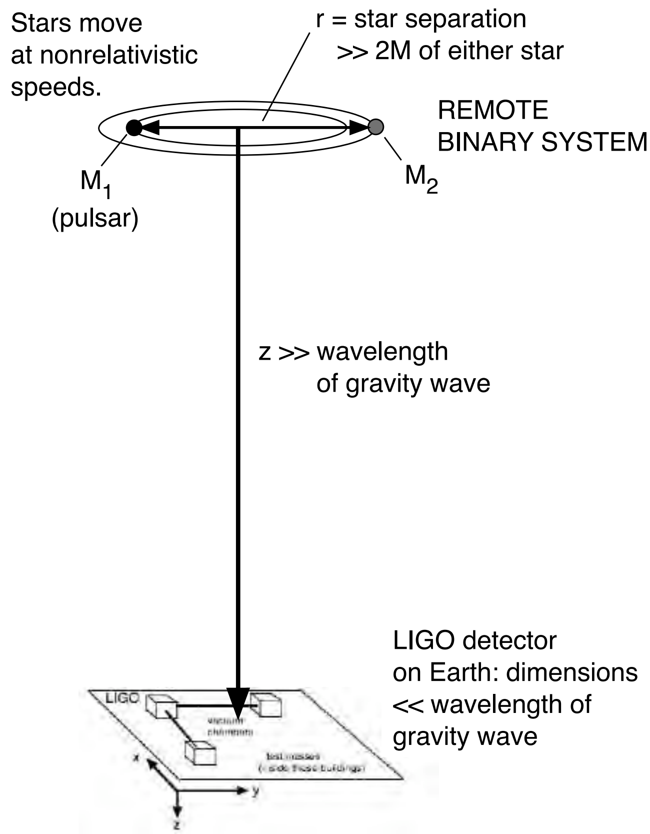


FIGURE 9 Schematic diagram, *not to scale*, showing notation and relative magnitudes for equation (29). The binary system and the LIGO detector lie in parallel planes.[Illustrator: See note in caption to Figure 5.]

(c) Separation z between the binary system and Earth is very much greater than a wavelength of the gravitational wave. This assumption assures that the radiation at Earth constitutes the so-called "far radiation field" where it assumes the form of a plane wave given in equation (5).

(d) The wavelength of the gravitational wave is much longer than the dimensions of the LIGO detector.

(e) The binary stars are orbiting in the xy plane, so that from Earth the orbits would appear as circles if we could see them (which we cannot).

. . . for one case

Equation (29) describes only one linear polarization at Earth, the one generated by metric (1) and shown in Figure 3. The orthogonal polarization

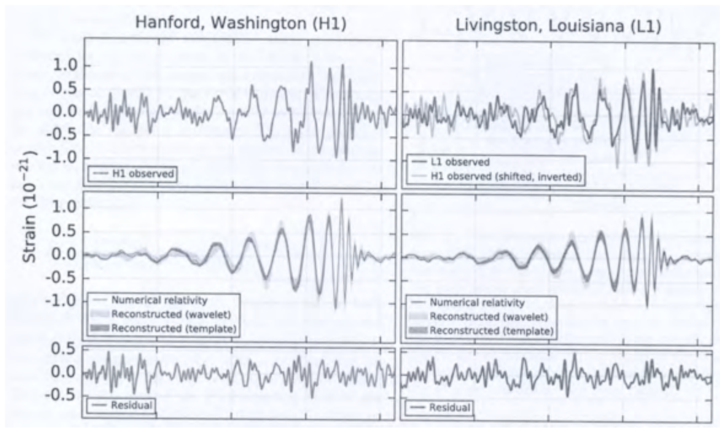


FIGURE 10 Detected "chirps" of the gravitational wave at two locations. The top row shows detected waveforms (superposed in the right-hand panel). The second row shows the cleaned-up image (again superposed). The bottom row displays "residuals," the noise deducted from images in the first row.

shown in Figure 4 is also transverse and equally strong, with components proportional to $(1 \pm h)$. The formula for the magnitude of h in that orthogonally polarized wave is identical to (29) with a sine function replacing the cosine function. We have not displayed the metric for that orthogonal polarization.

Detection requirements

In order for LIGO to detect a gravitational wave, two conditions must be met: (a) the amplitude h of the gravitational wave must be sufficiently large, and (b) the frequency of the wave must be in the range in which LIGO is most sensitive (100 to 400 hertz). Query 14 deals with the amplitude of the wave. The frequency of gravitational waves, discussed in Query 15, contains a surprise.

QUERY 12. Amplitude of gravitational wave at Earth

A. Use (29) to calculate the maximum amplitude of h at Earth due to the radiation from our "idealized circular-orbit" binary system.

B. Can LIGO detect the gravitational waves whose amplitude is given in part A?

C. What is the maximum amplitude of h at Earth just before coalescence, when the orbiting black holes are separated by $r = 20$ kilometers (but with orbits still described approximately by Newtonian mechanics)?

QUERY 13. Frequency of emitted gravitational waves

A. In order LIGO to detect the gravitational waves whose amplitude is given in Query 14, the frequency of the gravitational wave must be in the range 100 to 400 hertz. In Figure 9 the point

C. M. is the stationary center of mass of the pulsar system. Using the symbols in this figure, fill in the steps to complete the following derivation.

$$\frac{v_1^2}{r_1} = \frac{GM_1}{r_1^2} \qquad \text{(for } M_1\text{, Newton, conventional units)} \tag{30}$$

$$\frac{v_2^2}{r_1} = \frac{GM_2}{r_2^2} \qquad \text{(for } M_2\text{, Newton, conventional units)} \tag{31}$$

$$M_1 r_1 = M_2 r_2 \qquad \text{(center-of-mass condition)} \tag{32}$$

$$f_{\text{orbit}} \equiv \frac{1}{T_{\text{orbit}}} = \frac{v_1}{2\pi r_1} = \frac{v_2}{2\pi r_2} \qquad \text{(common orbital frequency)} \tag{33}$$

where f_{orbit} and T_{orbit} are the frequency and period of the orbit, respectively. From these equations, show that for $r \equiv r_1 + r_2$ the frequency of the orbit is

$$f_{\text{orbit}} = \frac{1}{2\pi}\left[\frac{G\left(M_1+M_2\right)}{r^3}\right]^{1/2} \qquad \text{(conventional units)} \tag{34}$$

$$= \frac{1}{2\pi}\left[\frac{M_1+M_2}{r^3}\right]^{1/2} \qquad \text{(metric units)} \tag{35}$$

B. Next is a surprise: The frequency f of the gravitational wave generated by this binary pair and appearing in (29) is twice the orbital frequency.

$$f_{\text{gravity wave}} = 2 f_{\text{orbit}} \tag{36}$$

Why this doubling? Essentially it is because gravitational waves are waves of tides. Just as there are two high tides and two low tides per day caused by the moon's gravity acting on the Earth, there are two peaks and two troughs of gravitational waves generated per binary orbit.

C. Approximate the average of the component masses in (34) by the value $M = 30M_{\text{Sun}}$. Find the r-value between the binary stars when the orbital frequency is 75 hertz, so that the frequency of the gravitational wave is 150 hertz.

D. Use results quoted earlier in this chapter to find an approximate expression for the time for the binary system to decay from the current radial separation to the radial separation calculated in part C.
ANS: $t_2 - t_1 = 5(r_2^4 - r_1^4)/(256M^3)$, every symbol in unit meter.

“Chirp” at coalescence

Newtonian mechanics predicts the motion of the binary system surprisingly accurately until the two components touch, a few milliseconds before they coalesce. Newton tells us that as the separation r between the orbiting masses decreases, their orbiting frequency increases. As a result the gravitational wave sweeps upward in both frequency and amplitude in what is called a **chirp**. Figure 1 is the predicted wave form for such a chirp.

16.9 ■ RESULTS FROM GRAVITATIONAL WAVE DETECTION; FUTURE PLANS

Unexpected details

Investigators milked a surprising amount of information from the first detection of gravitational waves. For example:

1. The initial binary system consisted of two black holes of mass $M_1 = (36 + 5/-4)M_{\text{Sun}}$ (that is, uncertainty of $+5M_{\text{Sun}}$ and $-4M_{\text{Sun}}$) and $M_2 = (29 \pm 4)M_{\text{Sun}}$.
2. The mass of the final black hole was $(62 \pm 4)M_{\text{Sun}}$.
3. Items 1 and 2 mean that the total energy of emitted gravitational radiation was about $3M_{\text{Sun}}$. A cataclysmic event indeed!
4. The two detection locations are separated by 10 milliseconds of light-travel time, or 3000 kilometers.
5. The signals were separated by $6.9 + 0.5/-0.4$ milliseconds, which means that they did not come from overhead.

How did observations lead to these results?

Item 1 derives from two equations in two unknowns (27) and (34), with validation in the small separation r-value at which merging takes place.
Item 2 follows from the frequency of ringing in the merged black hole.
Item 3 follows from Item 2.
Item 4 results from standard surveying.
Item 5 follows from direct comparison of synchronized clocks.

What are plans for future gravitational wave detections?

A. Increased sensitivity of each LIGO system
B. Increased number of LIGO detectors across the Earth, to measure the source direction more accurately.
C. Installation of LISA (Laser Interferometer Space Antenna Project) in space, which removes seismic noise at low frequencies in Figure 2).

16.10 ■ REFERENCES

Initial quote: Janna Levin, *Black Hole Blues*, New York, Alfred A. Knopf, 2016, page 36.

"Observation of Gravitational Waves from a Binary Black Hole Merger," Physical Review Letters, Volume 116, 12 February 2016, with 1000 authors!

Einstein's first paper on gravitational waves: Albert Einstein, "Über Gravitationswellen," Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, Berlin (1918), 154–167.

Correcting an error in Einstein's first paper: Albert Einstein, "Näherungsweise Integration der Feldgleichungen der Gravitation", Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, Berlin (1916), 688–696.

Download file name: Ch16GravitationalWaves171018v1.pdf