CHAPTER 4

Gravitation

4.1 ■ PHYSICS IN CURVED SPACETIME

Having paid our mathematical dues, we are now prepared to examine the physics of gravitation as described by general relativity. This subject falls naturally into two pieces: how the gravitational field influences the behavior of matter, and how matter determines the gravitational field. In Newtonian gravity, these two elements consist of the expression for the acceleration of a body in a gravitational potential Φ,

$$\mathbf{a} = -\nabla\Phi, \tag{4.1}$$

and Poisson's differential equation for the potential in terms of the matter density ρ and Newton's gravitational constant G:

$$\nabla^2\Phi = 4\pi G\rho. \tag{4.2}$$

In general relativity, the analogous statements will describe how the curvature of spacetime acts on matter to manifest itself as gravity, and how energy and momentum influence spacetime to create curvature. In either case it would be legitimate to start at the top, by stating outright the laws governing physics in curved spacetime and working out their consequences. Instead, we will try to be a little more motivational, starting with basic physical principles and attempting to argue that these lead naturally to an almost unique physical theory.

In Chapter 2 we motivated our discussion of manifolds by introducing the Einstein Equivalence Principle, or EEP: "In small enough regions of spacetime, the laws of physics reduce to those of special relativity; it is impossible to detect the existence of a gravitational field by means of local experiments." The EEP arises from the idea that gravity is *universal*; it affects all particles (and indeed all forms of energy-momentum) in the same way. This feature of universality led Einstein to propose that what we experience as gravity is a manifestation of the curvature of spacetime. The idea is simply that something so universal as gravitation could be most easily described as a fundamental feature of the background on which matter fields propagate, as opposed to as a conventional force. At the same time, the identification of spacetime as a curved manifold is supported by the similarity between the undetectability of gravity in local regions and our ability to find locally inertial coordinates ($g_{\hat{\mu}\hat{\nu}} = \eta_{\hat{\mu}\hat{\nu}}$, $\partial_{\hat{\sigma}} g_{\hat{\mu}\hat{\nu}} = 0$ at a point p) on a manifold.

Best of all, this abstract philosophizing translates directly into a simple recipe for generalizing laws of physics to the curved-spacetime context, known as the **minimal-coupling principle**. In its baldest form, this recipe may be stated as follows:

1. Take a law of physics, valid in inertial coordinates in flat spacetime.
2. Write it in a coordinate-invariant (tensorial) form.
3. Assert that the resulting law remains true in curved spacetime.

It may seem somewhat melodramatic to take such a simple idea and spread it out into a three-part procedure. We hope only to make clear that there is nothing very complicated going on. Operationally, this recipe usually amounts to taking an agreed-upon law in flat space and replacing the Minkowski metric $\eta_{\mu\nu}$ by the more general metric $g_{\mu\nu}$, and replacing partial derivatives ∂_μ by covariant derivatives ∇_μ. For this reason, this recipe is sometimes known as the "Comma-Goes-to-Semicolon Rule," by those who use commas and semicolons to denote partial and covariant derivatives.

As a straightforward example, we can consider the motion of freely-falling (unaccelerated) particles. In flat space such particles move in straight lines; in equations, this is expressed as the vanishing of the second derivative of the parameterized path $x^\mu(\lambda)$:

$$\frac{d^2x^\mu}{d\lambda^2} = 0. \tag{4.3}$$

This is not, in general coordinates, a tensorial equation; although $dx^\mu/d\lambda$ are the components of a well-defined vector, the second derivative components $d^2x^\mu/d\lambda^2$ are not. You might really think that this is a tensorial-looking equation; however, you can readily check that it's not even true in polar coordinates, unless you expect free particles to move in circles. We can use the chain rule to write

$$\frac{d^2x^\mu}{d\lambda^2} = \frac{dx^\nu}{d\lambda}\partial_\nu\frac{dx^\mu}{d\lambda}. \tag{4.4}$$

Now it is clear how to generalize this to curved space—simply replace the partial derivative by a covariant one,

$$\frac{dx^\nu}{d\lambda}\partial_\nu\frac{dx^\mu}{d\lambda} \quad\to\quad \frac{dx^\nu}{d\lambda}\nabla_\nu\frac{dx^\mu}{d\lambda} = \frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda}. \tag{4.5}$$

We recognize, then, that the appropriate general-relativistic version of the Newtonian relation (4.3) is simply the geodesic equation,

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda} = 0. \tag{4.6}$$

In general relativity, therefore, free particles move along geodesics; we have mentioned this before, but now you have a slightly better idea why it is true.

As an even more straightforward example, and one that we have referred to already, we have the law of energy-momentum conservation in flat spacetime:

$$\partial_\mu T^{\mu\nu} = 0. \tag{4.7}$$

Plugging into our recipe reveals the appropriate generalization to curved spacetime:

$$\nabla_\mu T^{\mu\nu} = 0. \tag{4.8}$$

It really is just that simple—sufficiently so that we felt quite comfortable using this equation in Chapter 3, without any detailed justification. Of course, this simplicity should not detract from the profound consequences of the generalization to curved spacetime, as illustrated in the example of the expanding universe.

It is one thing to generalize an equation from flat to curved spacetime; it is something altogether different to argue that the result describes gravity. To do so, we can show how the usual results of Newtonian gravity fit into the picture. We define the Newtonian limit by three requirements: the particles are moving slowly (with respect to the speed of light), the gravitational field is weak (so that it can be considered as a perturbation of flat space), and the field is also static (unchanging with time). Let us see what these assumptions do to the geodesic equation, taking the proper time τ as an affine parameter. "Moving slowly" means that

$$\frac{dx^i}{d\tau} \ll \frac{dt}{d\tau}, \tag{4.9}$$

so the geodesic equation becomes

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{00}\left(\frac{dt}{d\tau}\right)^2 = 0. \tag{4.10}$$

Since the field is static ($\partial_0 g_{\mu\nu} = 0$), the relevant Christoffel symbols Γ^μ_{00} simplify:

$$\begin{aligned} \Gamma^\mu_{00} &= \tfrac{1}{2} g^{\mu\lambda}(\partial_0 g_{\lambda 0} + \partial_0 g_{0\lambda} - \partial_\lambda g_{00}) \\ &= -\tfrac{1}{2} g^{\mu\lambda}\partial_\lambda g_{00}. \end{aligned} \tag{4.11}$$

Finally, the weakness of the gravitational field allows us to decompose the metric into the Minkowski form plus a small perturbation:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1. \tag{4.12}$$

We are working in inertial coordinates, so $\eta_{\mu\nu}$ is the canonical form of the metric. The "smallness condition" on the metric perturbation $h_{\mu\nu}$ doesn't really make sense in arbitrary coordinates. From the definition of the inverse metric, $g^{\mu\nu}g_{\nu\sigma} = \delta^\mu_\sigma$, we find that to first order in h,

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}, \tag{4.13}$$

where $h^{\mu\nu} = \eta^{\mu\rho}\eta^{\nu\sigma}h_{\rho\sigma}$. In fact, we can use the Minkowski metric to raise and lower indices on an object of any definite order in h, since the corrections would only contribute at higher orders. If you like, think of $h_{\mu\nu}$ as a symmetric $(0, 2)$ tensor field propagating in Minkowski space and interacting with other fields.

Putting it all together, to first order in $h_{\mu\nu}$ we find

$$\Gamma^{\mu}_{00} = -\tfrac{1}{2}\eta^{\mu\lambda}\partial_\lambda h_{00}. \tag{4.14}$$

The geodesic equation (4.10) is therefore

$$\frac{d^2x^\mu}{d\tau^2} = \frac{1}{2}\eta^{\mu\lambda}\partial_\lambda h_{00}\left(\frac{dt}{d\tau}\right)^2. \tag{4.15}$$

Using $\partial_0 h_{00} = 0$, the $\mu = 0$ component of this is just

$$\frac{d^2t}{d\tau^2} = 0. \tag{4.16}$$

That is, $dt/d\tau$ is constant. To examine the spacelike components of (4.15), recall that the spacelike components of $\eta^{\mu\nu}$ are just those of a 3×3 identity matrix. We therefore have

$$\frac{d^2x^i}{d\tau^2} = \frac{1}{2}\left(\frac{dt}{d\tau}\right)^2 \partial_i h_{00}. \tag{4.17}$$

Dividing both sides by $(dt/d\tau)^2$ has the effect of converting the derivative on the left-hand side from τ to t, leaving us with

$$\frac{d^2x^i}{dt^2} = \frac{1}{2}\partial_i h_{00}. \tag{4.18}$$

This begins to look a great deal like Newton's theory of gravitation. In fact, if we compare this equation to (4.1), we find that they are the same once we identify

$$h_{00} = -2\Phi, \tag{4.19}$$

or in other words

$$g_{00} = -(1 + 2\Phi). \tag{4.20}$$

Therefore, we have shown that the curvature of spacetime is indeed sufficient to describe gravity in the Newtonian limit, as long as the metric takes the form (4.20). It remains, of course, to find field equations for the metric that imply this is the form taken, and that for a single gravitating body we recover the Newtonian formula

$$\Phi = -\frac{GM}{r}, \tag{4.21}$$

but that will come soon enough.

The straightforward procedure we have outlined for generalizing laws of physics to curved spacetime does have some subtleties, which we address in Section 4.7. But it's more than good enough for our present purposes, so let's not delay our pursuit of the second half of our task, obtaining the field equation for the metric in general relativity.

4.2 ■ EINSTEIN'S EQUATION

Just as Maxwell's equations govern how the electric and magnetic fields respond to charges and currents, Einstein's field equation governs how the metric responds to energy and momentum. Ultimately the field equation must be postulated and tested against experiment, not derived from any bedrock principles; however, we can motivate it on the basis of plausibility arguments. We will actually do this in two ways: first by some informal reasoning by analogy, close to what Einstein himself was thinking, and then by starting with an action and deriving the corresponding equations of motion.

The informal argument begins with the realization that we would like to find an equation that supersedes the Poisson equation for the Newtonian potential:

$$\nabla^2\Phi = 4\pi G\rho, \tag{4.22}$$

where $\nabla^2 = \delta^{ij}\partial_i\partial_j$ is the Laplacian in space and ρ is the mass density. [The explicit form of Φ given in (4.21) is one solution of (4.22), for the case of a pointlike mass distribution.] What characteristics should our sought-after equation possess? On the left-hand side of (4.22) we have a second-order differential operator acting on the gravitational potential, and on the right-hand side a measure of the mass distribution. A relativistic generalization should take the form of an equation between tensors. We know what the tensor generalization of the mass density is; it's the energy-momentum tensor $T_{\mu\nu}$. The gravitational potential, meanwhile, should get replaced by the metric tensor, because in (4.20) we had to relate a perturbation of the metric to the Newtonian potential to successfully reproduce gravity. We might therefore guess that our new equation will have $T_{\mu\nu}$ set proportional to some tensor, which is second-order in derivatives of the metric; something along the lines of

$$[\nabla^2 g]_{\mu\nu} \propto T_{\mu\nu}, \tag{4.23}$$

but of course we want it to be completely tensorial.

The left-hand side of (4.23) is not a sensible tensor; it's just a suggestive notation to indicate that we would like a symmetric (0, 2) tensor that is second-order in derivatives of the metric. The first choice might be to act the d'Alembertian $\Box = \nabla^\mu\nabla_\mu$ on the metric $g_{\mu\nu}$, but this is automatically zero by metric compatibility. Fortunately, there is an obvious quantity which is not zero and is constructed from second derivatives (and first derivatives) of the metric: the Riemann tensor $R^\rho{}_{\sigma\mu\nu}$. Recall that the Riemann tensor is constructed from the Christoffel symbols and their first derivatives, and the Christoffel symbols are constructed from the metric and its first derivatives, so $R^{\rho}{}_{\sigma\mu\nu}$ contains second derivatives of $g_{\mu\nu}$. It doesn't have the right number of indices, but we can contract it to form the Ricci tensor $R_{\mu\nu}$, which does (and is symmetric to boot). It is therefore tempting to guess that the gravitational field equations are

$$R_{\mu\nu} = \kappa T_{\mu\nu}, \tag{4.24}$$

for some constant κ. In fact, Einstein did suggest this equation at one point. There is a problem, unfortunately, with conservation of energy. If we want to preserve

$$\nabla^{\mu} T_{\mu\nu} = 0, \tag{4.25}$$

by (4.24) we would have

$$\nabla^{\mu} R_{\mu\nu} = 0. \tag{4.26}$$

This is certainly not true in an arbitrary geometry; we have seen from the Bianchi identity (3.150) that

$$\nabla^{\mu} R_{\mu\nu} = \tfrac{1}{2}\nabla_{\nu} R. \tag{4.27}$$

But our proposed field equation implies that $R = \kappa g^{\mu\nu} T_{\mu\nu} = \kappa T$, so taking these together we have

$$\nabla_{\mu} T = 0. \tag{4.28}$$

The covariant derivative of a scalar is just the partial derivative, so (4.28) is telling us that T is constant throughout spacetime. This is highly implausible, since $T = 0$ in vacuum while $T \neq 0$ in matter. We have to try harder.

Of course we don't have to try much harder, since we already know of a symmetric (0, 2) tensor, constructed from the Ricci tensor, which is automatically conserved: the Einstein tensor

$$G_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu}, \tag{4.29}$$

which always obeys $\nabla^{\mu} G_{\mu\nu} = 0$. We are therefore led to propose

$$G_{\mu\nu} = \kappa T_{\mu\nu} \tag{4.30}$$

as a field equation for the metric. (Actually it is probably more common to write out $R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}$, rather than use the abbreviation $G_{\mu\nu}$.) This equation satisfies all of the obvious requirements: the right-hand side is a covariant expression of the energy and momentum density in the form of a symmetric and conserved (0, 2) tensor, while the left-hand side is a symmetric and conserved (0, 2) tensor constructed from the metric and its first and second derivatives. It only remains to fix the proportionality constant κ, and to see whether the result actually reproduces gravity as we know it. In other words, does this equation predict the Poisson equation for the gravitational potential in the Newtonian limit?

To answer this, note that contracting both sides of (4.30) yields (in four dimensions)

$$R = -\kappa T, \tag{4.31}$$

and using this we can rewrite (4.30) as

$$R_{\mu\nu} = \kappa(T_{\mu\nu} - \tfrac{1}{2}Tg_{\mu\nu}). \tag{4.32}$$

This is the same equation, just written slightly differently. We would like to see if it predicts Newtonian gravity in the weak-field, time-independent, slowly-moving-particles limit. We consider a perfect-fluid source of energy-momentum, for which

$$T_{\mu\nu} = (\rho + p)U_\mu U_\nu + pg_{\mu\nu}, \tag{4.33}$$

where U^μ is the fluid four-velocity and ρ and p are the rest-frame energy and momentum densities. In fact for the Newtonian limit we may neglect the pressure; roughly speaking, the pressure of a body becomes important when its constituent particles are traveling at speeds close to that of light, which we exclude from the Newtonian limit by hypothesis. So we are actually considering the energy-momentum tensor of dust:

$$T_{\mu\nu} = \rho U_\mu U_\nu. \tag{4.34}$$

The "fluid" we are considering is some massive body, such as the Earth or the Sun. We will work in the fluid rest frame, in which

$$U^\mu = (U^0, 0, 0, 0). \tag{4.35}$$

The timelike component can be fixed by appealing to the normalization condition $g_{\mu\nu}U^\mu U^\nu = -1$. In the weak-field limit we write, in accordance with (4.12) and (4.13),

$$\begin{aligned} g_{00} &= -1 + h_{00}, \\ g^{00} &= -1 - h_{00}. \end{aligned} \tag{4.36}$$

Then to first order in $h_{\mu\nu}$ we get

$$U^0 = 1 + \tfrac{1}{2}h_{00}. \tag{4.37}$$

In fact, however, this is needlessly careful, as we are going to plug the four-velocity into (4.34), and the energy density ρ is already considered small (spacetime will be flat as ρ is taken to zero). So to our level of approximation, we can simply take $U^0 = 1$, and correspondingly $U_0 = -1$. Then

$$T_{00} = \rho, \tag{4.38}$$

and all other components vanish. In this limit the rest energy $\rho = T_{00}$ will be much larger than the other terms in $T_{\mu\nu}$, so we want to focus on the $\mu = 0$, $\nu = 0$ component of (4.32). The trace, to lowest nontrivial order, is

$$T = g^{00}T_{00} = -T_{00} = -\rho. \tag{4.39}$$

We plug this into the 00 component of our proposed gravitational field equation (4.32), to get

$$R_{00} = \tfrac{1}{2}\kappa\rho. \tag{4.40}$$

This is an equation relating derivatives of the metric to the energy density. To find the explicit expression in terms of the metric, we need to evaluate $R_{00} = R^{\lambda}{}_{0\lambda 0}$. In fact we only need $R^{i}{}_{0i0}$, since $R^{0}{}_{000} = 0$. We have

$$R^{i}{}_{0j0} = \partial_j\Gamma^{i}_{00} - \partial_0\Gamma^{i}_{j0} + \Gamma^{i}_{j\lambda}\Gamma^{\lambda}_{00} - \Gamma^{i}_{0\lambda}\Gamma^{\lambda}_{j0}. \tag{4.41}$$

The second term here is a time derivative, which vanishes for static fields. The third and fourth terms are of the form $(\Gamma)^2$, and since Γ is first-order in the metric perturbation these contribute only at second order, and can be neglected. We are left with $R^{i}{}_{0j0} = \partial_j\Gamma^{i}_{00}$. From this we get

$$\begin{aligned} R_{00} &= R^{i}{}_{0i0} \\ &= \partial_i\left[\tfrac{1}{2}g^{i\lambda}(\partial_0 g_{\lambda 0} + \partial_0 g_{0\lambda} - \partial_\lambda g_{00})\right] \\ &= -\tfrac{1}{2}\delta^{ij}\partial_i\partial_j h_{00} \\ &= -\tfrac{1}{2}\nabla^2 h_{00}. \end{aligned} \tag{4.42}$$

Comparing to (4.40), we see that the 00 component of (4.30) in the Newtonian limit predicts

$$\nabla^2 h_{00} = -\kappa\rho. \tag{4.43}$$

Since (4.19) sets $h_{00} = -2\Phi$, this is precisely the Poisson equation (4.22), if we set $\kappa = 8\pi G$.

So our guess, (4.30), seems to have worked out. With the normalization chosen so as to correctly recover the Newtonian limit, we can present **Einstein's equation** for general relativity:

$$\boxed{R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}.} \tag{4.44}$$

This tells us how the curvature of spacetime reacts to the presence of energy-momentum. G is of course Newton's constant of gravitation; it has nothing to do with the trace of $G_{\mu\nu}$. Einstein, you may have heard, thought that the left-hand side was nice and geometrical, while the right-hand side was somewhat less compelling.

It is sometimes useful to rewrite Einstein's equation in a slightly different form. Following (4.31) and (4.32), we can take the trace of (4.44) to find that $R = -8\pi GT$. Plugging this in and moving the trace term to the right-hand side, we obtain

$$\boxed{R_{\mu\nu} = 8\pi G\left(T_{\mu\nu} - \tfrac{1}{2}Tg_{\mu\nu}\right).} \tag{4.45}$$

The difference between this and (4.44) is purely cosmetic; in substance they are precisely the same. We will often be interested in the Einstein's equation in vacuum, where $T_{\mu\nu} = 0$ (for example, outside a star or planet). Then of course the right-hand side of (4.45) vanishes. Therefore the vacuum Einstein equation is simply

$$\boxed{R_{\mu\nu} = 0.} \tag{4.46}$$

This is both slightly less formidable, and of considerable physical usefulness.

4.3 ■ LAGRANGIAN FORMULATION

An alternative route to Einstein's equation is through the principle of least action, as we discussed for classical field theories in flat spacetime at the end of Chapter 1. Let's spend a moment to generalize those results to curved spacetime, and then see what kind of Lagrangian is appropriate for general relativity. We'll work in n dimensions, since our results will not depend on the dimensionality; we will, however, assume that our metric has Lorentzian signature.

Consider a field theory in which the dynamical variables are a set of fields $\Phi^i(x)$. The classical solutions to such a theory will be those that are critical points of an action S, generally expressed as an integral over space of a Lagrange density $\mathcal{L}$,

$$S = \int \mathcal{L}(\Phi^i, \nabla_\mu \Phi^i)\, d^n x. \tag{4.47}$$

Note that we are now imagining that the Lagrangian is a function of the fields and their covariant (rather than partial) derivatives, as is appropriate in curved space. Note also that, since $d^n x$ is a density rather than a tensor, $\mathcal{L}$ is also a density (since their product must be a well-defined tensor); we typically write

$$\mathcal{L} = \sqrt{-g}\,\widehat{\mathcal{L}}, \tag{4.48}$$

where $\widehat{\mathcal{L}}$ is indeed a scalar. You might think it would be sensible to forget about what we are calling $\mathcal{L}$ and just focus on $\widehat{\mathcal{L}}$, but in fact both quantities are useful in different circumstances; it is $\mathcal{L}$ that will matter whenever we are varying with respect to the metric itself. The associated Euler–Lagrange equations make use of the scalar $\widehat{\mathcal{L}}$, and are otherwise like those in flat space, but with covariant instead of partial derivatives:

$$\frac{\partial\widehat{\mathcal{L}}}{\partial\Phi} - \nabla_\mu\left(\frac{\partial\widehat{\mathcal{L}}}{\partial(\nabla_\mu\Phi)}\right) = 0. \tag{4.49}$$

In deriving these equations, we make use of Stokes's theorem (3.35),

$$\int_\Sigma \nabla_\mu V^\mu \sqrt{|g|}\, d^n x = \int_{\partial\Sigma} n_\mu V^\mu \sqrt{|\gamma|}\, d^{n-1}x, \tag{4.50}$$

and set the variation equal to zero at infinity (the boundary). Integration by parts therefore takes the form

$$\int A^\mu(\nabla_\mu B)\sqrt{-g}\, d^n x = -\int (\nabla_\mu A^\mu) B\sqrt{-g}\, d^n x + \text{boundary terms}. \tag{4.51}$$

For example, the curved-spacetime generalization of the action for a single scalar field ϕ considered in Chapter 1 would be

$$S_\phi = \int \left[-\frac{1}{2}g^{\mu\nu}(\nabla_\mu\phi)(\nabla_\nu\phi) - V(\phi)\right]\sqrt{-g}\, d^n x, \tag{4.52}$$

which would lead to an equation of motion

$$\Box\phi - \frac{dV}{d\phi} = 0, \tag{4.53}$$

where the covariant d'Alembertian is

$$\Box = \nabla^\mu\nabla_\mu = g^{\mu\nu}\nabla_\mu\nabla_\nu. \tag{4.54}$$

Just as in flat spacetime, the combination $g^{\mu\nu}(\nabla_\mu\phi)(\nabla_\nu\phi)$ is often abbreviated as $(\nabla\phi)^2$. Of course, the covariant derivatives are equivalent to partial derivatives when acting on scalars, but it is wise to use the ∇_μ notation still; you never know when you might integrate by parts and suddenly be acting on a vector.

With that as a warm-up, we turn to the construction of an action for general relativity. Our dynamical variable is now the metric $g_{\mu\nu}$; what scalars can we make out of the metric to serve as a Lagrangian? Since we know that the metric can be set equal to its canonical form and its first derivatives set to zero at any one point, any nontrivial scalar must involve at least second derivatives of the metric. The Riemann tensor is of course made from second derivatives of the metric, and we argued earlier that the only independent scalar we could construct from the Riemann tensor was the Ricci scalar R. What we did not show, but is nevertheless true, is that any nontrivial tensor made from products of the metric and its first and second derivatives can be expressed in terms of the metric and the Riemann tensor. Therefore, the *only* independent scalar constructed from the metric, which is no higher than second order in its derivatives, is the Ricci scalar. Hilbert figured that this was therefore the simplest possible choice for a Lagrangian, and proposed

$$S_H = \int \sqrt{-g} R \, d^n x, \tag{4.55}$$

known as the **Hilbert action** (or sometimes the Einstein–Hilbert action). As we shall see, he was right.

The equation of motion should come from varying the action with respect to the metric. Unfortunately the action isn't quite in the form (4.47), since it can't be written in terms of covariant derivatives of $g_{\mu\nu}$ (which would simply vanish). Therefore, instead of simply plugging into the Euler–Lagrange equations, we will consider directly the behavior of S_H under small variations of the metric. In fact it is more convenient to vary with respect to the inverse metric $g^{\mu\nu}$. Since $g^{\mu\lambda} g_{\lambda\nu} = \delta^\mu_\nu$, and the Kronecker delta is unchanged under any variation, it is straightforward to express variations of the metric and inverse metric in terms of each other:

$$\delta g_{\mu\nu} = -g_{\mu\rho} g_{\nu\sigma} \delta g^{\rho\sigma}, \tag{4.56}$$

so stationary points with respect to variations in $g^{\mu\nu}$ are equivalent to those with respect to variations in $g_{\mu\nu}$. Using $R = g^{\mu\nu} R_{\mu\nu}$, we have

$$\delta S_H = (\delta S)_1 + (\delta S)_2 + (\delta S)_3, \tag{4.57}$$

where

$$\begin{aligned}
(\delta S)_1 &= \int d^n x \sqrt{-g} g^{\mu\nu} \delta R_{\mu\nu} \\
(\delta S)_2 &= \int d^n x \sqrt{-g} R_{\mu\nu} \delta g^{\mu\nu} \\
(\delta S)_3 &= \int d^n x \, R \delta \sqrt{-g}.
\end{aligned} \tag{4.58}$$

The second term $(\delta S)_2$ is already in the form of some expression multiplied by $\delta g^{\mu\nu}$; let's examine the others more closely.

Recall that the Ricci tensor is the contraction of the Riemann tensor, which is given by

$$R^\rho{}_{\mu\lambda\nu} = \partial_\lambda \Gamma^\rho_{\nu\mu} + \Gamma^\rho_{\lambda\sigma} \Gamma^\sigma_{\nu\mu} - (\lambda \leftrightarrow \nu). \tag{4.59}$$

The variation of the Riemann tensor with respect to the metric can be found by first varying the connection with respect to the metric, and then substituting into this expression. However, let us consider arbitrary variations of the connection by replacing

$$\Gamma^\rho_{\nu\mu} \to \Gamma^\rho_{\nu\mu} + \delta \Gamma^\rho_{\nu\mu}. \tag{4.60}$$

The variation $\delta\Gamma^{\rho}_{\nu\mu}$ is the difference of two connections, and therefore is itself a tensor. We can thus take its covariant derivative,

$$\nabla_\lambda(\delta\Gamma^{\rho}_{\nu\mu}) = \partial_\lambda(\delta\Gamma^{\rho}_{\nu\mu}) + \Gamma^{\rho}_{\lambda\sigma}\delta\Gamma^{\sigma}_{\nu\mu} - \Gamma^{\sigma}_{\lambda\nu}\delta\Gamma^{\rho}_{\sigma\mu} - \Gamma^{\sigma}_{\lambda\mu}\delta\Gamma^{\rho}_{\nu\sigma}. \tag{4.61}$$

Here and elsewhere, the covariant derivatives are taken with respect to $g_{\mu\nu}$, not $g_{\mu\nu} + \delta g_{\mu\nu}$. Given this expression and a small amount of labor, it is easy to show that, to first order in the variation,

$$\delta R^{\rho}{}_{\mu\lambda\nu} = \nabla_\lambda(\delta\Gamma^{\rho}_{\nu\mu}) - \nabla_\nu(\delta\Gamma^{\rho}_{\lambda\mu}). \tag{4.62}$$

You are encouraged check this yourself. Therefore, the contribution of the first term in (4.58) to δS can be written

$$\begin{aligned}(\delta S)_1 &= \int d^n x\sqrt{-g}\; g^{\mu\nu}\left[\nabla_\lambda(\delta\Gamma^{\lambda}_{\nu\mu}) - \nabla_\nu(\delta\Gamma^{\lambda}_{\lambda\mu})\right] \\ &= \int d^n x\sqrt{-g}\; \nabla_\sigma\left[g^{\mu\nu}(\delta\Gamma^{\sigma}_{\mu\nu}) - g^{\mu\sigma}(\delta\Gamma^{\lambda}_{\lambda\mu})\right],\end{aligned} \tag{4.63}$$

where we have used metric compatibility and relabeled some dummy indices. We can now plug in the expression for $\delta\Gamma^{\sigma}_{\mu\nu}$ in terms of $\delta g^{\mu\nu}$, which works out to be

$$\delta\Gamma^{\sigma}_{\mu\nu} = -\tfrac{1}{2}\left[g_{\lambda\mu}\nabla_\nu(\delta g^{\lambda\sigma}) + g_{\lambda\nu}\nabla_\mu(\delta g^{\lambda\sigma}) - g_{\mu\alpha}g_{\nu\beta}\nabla^{\sigma}(\delta g^{\alpha\beta})\right], \tag{4.64}$$

leading to

$$(\delta S)_1 = \int d^n x\sqrt{-g}\; \nabla_\sigma[g_{\mu\nu}\nabla^{\sigma}(\delta g^{\mu\nu}) - \nabla_\lambda(\delta g^{\sigma\lambda})], \tag{4.65}$$

as you are also welcome to check. But (4.63) [or (4.65)] is an integral with respect to the natural volume element of the covariant divergence of a vector; by Stokes's theorem, this is equal to a boundary contribution at infinity, which we can set to zero by making the variation vanish at infinity. Therefore this term contributes nothing to the total variation. Although to be honest, we have cheated. The boundary term will include not only the metric variation, but also its first derivative, which is not traditionally set to zero. For our present purposes it doesn't matter, but in principle we might care about what happens at the boundary, and would have to include an additional term in the action to take care of this subtlety.

To make sense of the $(\delta S)_3$ term we need to use the following fact, true for any square matrix M with nonvanishing determinant:

$$\ln(\det M) = \mathrm{Tr}(\ln M). \tag{4.66}$$

Here, $\ln M$ is defined by $\exp(\ln M) = M$. For numbers this is obvious, for matrices it's a little less straightforward. The variation of this identity yields

$$\frac{1}{\det M}\delta(\det M) = \mathrm{Tr}(M^{-1}\delta M). \tag{4.67}$$

We have used the cyclic property of the trace to allow us to ignore the fact that M^{-1} and δM may not commute. Taking the matrix M to be the metric $g_{\mu\nu}$, so that $\det M = \det g_{\mu\nu} = g$, we get

$$\begin{aligned}\delta g &= g(g^{\mu\nu}\delta g_{\mu\nu}) \\ &= -g(g_{\mu\nu}\delta g^{\mu\nu}).\end{aligned} \tag{4.68}$$

In the last step we converted from $\delta g_{\mu\nu}$ to $\delta g^{\mu\nu}$ using (4.56). Now we can just plug in to get

$$\begin{aligned}\delta\sqrt{-g} &= -\frac{1}{2\sqrt{-g}}\delta g \\ &= \frac{1}{2}\frac{g}{\sqrt{-g}}g_{\mu\nu}\delta g^{\mu\nu} \\ &= -\frac{1}{2}\sqrt{-g}\,g_{\mu\nu}\delta g^{\mu\nu}.\end{aligned} \tag{4.69}$$

Hearkening back to (4.58), and remembering that $(\delta S)_1$ does not contribute, we find

$$\delta S_H = \int d^n x\sqrt{-g}\left[R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}\right]\delta g^{\mu\nu}. \tag{4.70}$$

Recall that the functional derivative of the action satisfies

$$\delta S = \int \sum_i \left(\frac{\delta S}{\delta\Phi^i}\,\delta\Phi^i\right)\,d^n x, \tag{4.71}$$

where $\{\Phi^i\}$ is a complete set of fields being varied (in our case, it's just $g^{\mu\nu}$). Stationary points are those for which each $\delta S/\delta\Phi^i = 0$, so we recover Einstein's equation in vacuum:

$$\frac{1}{\sqrt{-g}}\frac{\delta S_H}{\delta g^{\mu\nu}} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 0. \tag{4.72}$$

The advantage of the Lagrangian approach is manifested by the fact that our very first guess (which was practically unique) gave the right answer, in contrast with our previous trial-and-error method. This is a reflection of two elegant features of this technique: First, the Lagrangian is a scalar, rather than a tensor, and therefore more restricted; second, the symmetries of the theory are straightforwardly imposed (in this case, we automatically derived a tensor with vanishing divergence, which is related to diffeomorphism invariance, as discussed in Appendix B).

We derived Einstein's equation "in vacuum" because we only included the gravitational part of the action, not additional terms for matter fields. What we would really like, however, is to get the nonvacuum field equation as well. That means we consider an action of the form

$$S = \frac{1}{16\pi G} S_H + S_M, \tag{4.73}$$

where S_M is the action for matter, and we have presciently normalized the gravitational action so that we get the right answer. Following through the same procedure as above leads to

$$\frac{1}{\sqrt{-g}} \frac{\delta S}{\delta g^{\mu\nu}} = \frac{1}{16\pi G}\left(R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}\right) + \frac{1}{\sqrt{-g}} \frac{\delta S_M}{\delta g^{\mu\nu}} = 0. \tag{4.74}$$

We now boldly define the energy-momentum tensor to be

$$T_{\mu\nu} = -2 \frac{1}{\sqrt{-g}} \frac{\delta S_M}{\delta g^{\mu\nu}}. \tag{4.75}$$

This allows us to recover the complete Einstein's equation,

$$R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{4.76}$$

or equivalently, $G_{\mu\nu} = 8\pi G T_{\mu\nu}$.

Why should we think that (4.75) is really the energy-momentum tensor? In some sense it is only because it is a symmetric, conserved, (0, 2) tensor with dimensions of energy density; if you prefer to call it by some other name, go ahead. But it also accords with our preconceived expectations. Consider again the action for a scalar field, (4.52). Now vary this action with respect, not to ϕ, but to the inverse metric:

$$\delta S_\phi = \int d^n x \left[\sqrt{-g}\left(-\frac{1}{2} \delta g^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi\right) + \delta\sqrt{-g}\left(-\frac{1}{2} g^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi - V(\phi)\right)\right] \tag{4.77}$$

$$= \int d^n x \sqrt{-g}\, \delta g^{\mu\nu} \left[-\frac{1}{2} \nabla_\mu \phi \nabla_\nu \phi + \left(-\frac{1}{2} g_{\mu\nu}\right)\left(-\frac{1}{2} g^{\rho\sigma} \nabla_\rho \phi \nabla_\sigma \phi - V(\phi)\right)\right]. \tag{4.78}$$

We therefore have

$$\begin{aligned} T^{(\phi)}_{\mu\nu} &= -2 \frac{1}{\sqrt{-g}} \frac{\delta S_\phi}{\delta g^{\mu\nu}} \\ &= \nabla_\mu \phi \nabla_\nu \phi - \frac{1}{2} g_{\mu\nu} g^{\rho\sigma} \nabla_\rho \phi \nabla_\sigma \phi - g_{\mu\nu} V(\phi). \end{aligned} \tag{4.79}$$

In flat spacetime this reduces to what we had asserted, in Chapter 1, was the correct energy-momentum tensor for a scalar field.

On the other hand, in Minkowski space there is an alternative definition for the energy-momentum tensor, which is sometimes given in books on electromagnetism or field theory. In this context energy-momentum conservation arises as a consequence of symmetry of the Lagrangian under spacetime translations. **Noether's theorem** states that every symmetry of a Lagrangian implies the existence of a conservation law; invariance under the four spacetime translations leads to a tensor $S^{\mu\nu}$, which obeys $\partial_\mu S^{\mu\nu} = 0$ (four relations, one for each value of ν). The details can be found in Wald (1984) or Peskin and Schroeder (1995). Applying Noether's procedure to a Lagrangian that depends on some fields Φ^i and their first derivatives $\partial_\mu \Phi^i$ (in flat spacetime), we obtain

$$S^{\mu\nu} = \frac{\delta\mathcal{L}}{\delta(\partial_\mu \Phi^i)}\partial^\nu \Phi^i - \eta^{\mu\nu}\mathcal{L}, \tag{4.80}$$

where a sum over i is implied. You can check that this tensor is conserved by virtue of the equations of motion of the matter fields. $S^{\mu\nu}$ often goes by the name "canonical energy-momentum tensor"; however, there are a number of reasons why it is more convenient for us to use (4.75). First, (4.75) is in fact what appears on the right hand side of Einstein's equation when it is derived from an action, and it is not always possible to generalize (4.80) to curved spacetime. But even in flat space (4.75) has its advantages; it is manifestly symmetric, and also guaranteed to be gauge invariant, neither of which is true for (4.80). We will therefore stick with (4.75) as the definition of the energy-momentum tensor.

Now that Einstein's equation has been derived, the rest of this chapter is devoted to exploring some of its properties. These discussions are fascinating but not strictly necessary; if you like, you can jump right to the applications discussed in subsequent chapters.

4.4 ■ PROPERTIES OF EINSTEIN'S EQUATION

Einstein's equation may be thought of as a set of second-order differential equations for the metric tensor field $g_{\mu\nu}$. There are really ten independent equations (since both sides are symmetric two-index tensors), which seems to be exactly right for the ten unknown functions of the metric components. However, the Bianchi identity $\nabla^\mu G_{\mu\nu} = 0$ represents four constraints on the functions $R_{\mu\nu}(x)$, so there are only six truly independent equations in (4.44). In fact this is appropriate, since if a metric is a solution to Einstein's equation in one coordinate system x^μ it should also be a solution in any other coordinate system $x^{\mu'}$. This means that there are four unphysical degrees of freedom in $g_{\mu\nu}$, represented by the four functions $x^{\mu'}(x^\mu)$, and we should expect that Einstein's equation only constrains the six coordinate-independent degrees of freedom.

As differential equations, these are extremely complicated; the Ricci scalar and tensor are contractions of the Riemann tensor, which involves derivatives and products of the Christoffel symbols, which in turn involve the inverse metric and derivatives of the metric. Furthermore, the energy-momentum tensor $T_{\mu\nu}$ will generally involve the metric as well. The equations are also nonlinear, so that two known solutions cannot be superposed to find a third. It is therefore very difficult to solve Einstein's equation in any sort of generality, and it is usually necessary to make some simplifying assumptions. Even in vacuum, where we set the energy-momentum tensor to zero, the resulting equation (4.46) can be very difficult to solve. The most popular sort of simplifying assumption is that the metric has a significant degree of symmetry, and we will see later how isometries make life easier.

The nonlinearity of general relativity is worth a remark. In Newtonian gravity the potential due to two point masses is simply the sum of the potentials for each mass, but clearly this does not carry over to general relativity outside the weak-field limit. There is a physical reason for this, namely that in GR the gravitational field couples to itself. This can be thought of as a consequence of the equivalence principle—if gravitation did not couple to itself, a gravitational atom (two particles bound by their mutual gravitational attraction) would have a different inertial mass than gravitational mass (due to the negative binding energy). The nonlinearity of Einstein's equation is a reflection of the back-reaction of gravity on itself.

A nice way to think about this is provided by Feynman diagrams. These are used in quantum field theory to calculate the amplitudes for scattering processes, which can be obtained by summing the various contributions from different interactions, each represented by its own diagram. Even if we don't go so far as to quantize gravity and calculate scattering cross-sections (see the end of this section), we can still draw Feynman diagrams as a simple way of keeping track of which interactions exist and which do not. A simple example is provided by the electromagnetic interaction between two electrons; this can be thought of as due to exchange of a virtual photon, as shown in Figure 4.1.

In contrast, there is no diagram in which two photons exchange another photon between themselves, because electromagnetism is linear (there is no back-reaction). The gravitational interaction, meanwhile, can be thought of as deriving from the exchange of a virtual graviton (a quantized perturbation of the metric). The nonlinearity manifests itself as the fact that both electrons and gravi-

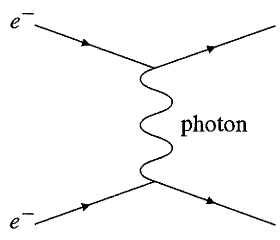

FIGURE 4.1 A Feynman diagram for electromagnetism. In quantum field theory, such diagrams are used to calculate amplitudes for scattering processes; here, just think of it as a cartoon representing a certain interaction. The point of this particular diagram is that the coupling of photons to electrons is what causes the electromagnetic interaction between them. In contrast, there is no coupling of photons to other photons, and no analogous diagram in which photons interact.

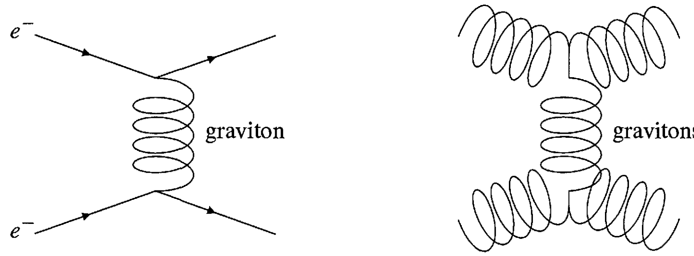

FIGURE 4.2 Feynman diagrams for gravity. Upon quantization, Einstein's equation predicts spin-two particles called gravitons. We don't know how to carry out such a quantization consistently, but the existence of gravitons is sufficiently robust that it is expected to be a feature of any well-defined scheme. Since gravity couples to energy-momentum, gravitons interact with every kind of particle, including other gravitons. This provides a way of thinking about the nonlinearity of Einstein's theory.

tons can exchange virtual gravitons, and therefore exert a gravitational force, as shown in Figure 4.2. There is nothing unique about this feature of gravity; it is shared by most gauge theories, such as quantum chromodynamics, the theory of the strong interactions. Electromagnetism is actually the exception; the linearity can be traced to the fact that the relevant gauge group, U(1), is abelian. But nonlinearity does represent a departure from the Newtonian theory. This difference is experimentally detectable; the reason why (as we shall see) the orbit of Mercury is different in GR versus Newtonian gravity is that the gravitational field influences itself, and the closer we get to the Sun, the more noticeable that influence is.

Beyond the fact that it is complicated and nonlinear, it is worth thinking a bit about what Einstein's equation is actually telling us. Clearly it relates the energy-momentum distribution to components of the curvature tensor; but from a physical point of view, precisely what kind of gravitational field is generated by a given kind of source? One way to answer this question is to consider the evolution of the *expansion* θ of a family of neighboring timelike geodesics. We imagine a small ball of free test particles moving along geodesics with four-velocities U^μ, and follow their evolution; the expansion $\theta = \nabla_\mu U^\mu$ tells us how the volume of the ball is growing (or shrinking, if $\theta < 0$) at any one moment of time. Clearly the value of the expansion will depend on the initial conditions for our test particles. The effects of gravity, on the other hand, are encoded in the *evolution* of the expansion, which is governed by Raychaudhuri's equation. This equation, discussed in Appendix F, tells us that the derivative of the expansion with respect to the proper time τ along the geodesics is given by the following expression:

$$\frac{d\theta}{d\tau} = 2\omega^2 - 2\sigma^2 - \frac{1}{3}\theta^2 - R_{\mu\nu}U^\mu U^\nu. \tag{4.81}$$

The terms on the right-hand side are explained carefully in Appendix F; ω encodes the rotation of the geodesics, σ encodes the shear, and $R_{\mu\nu}$ is of course the Ricci tensor. Raychaudhuri's equation is a purely geometric relation, making no reference to Einstein's equation. The combination of the two equations, however, can be used to describe how energy-momentum influences the motion of test particles, since Einstein's equation relates $T_{\mu\nu}$ to $R_{\mu\nu}$ and Raychaudhuri's equation relates $R_{\mu\nu}$ to $d\theta/d\tau$.

Let us consider the simplest possible situation, where we start with all of the nearby particles at rest with respect to each other in a small region of spacetime. Then the expansion, rotation, and shear will all vanish at this initial moment. Let us further construct locally inertial coordinates $x^{\hat{\mu}}$, in which $U^{\hat{\mu}}$ is in its rest frame, so that $U^{\hat{\mu}} = (1, 0, 0, 0)$ and $R_{\hat{\mu}\hat{\nu}}U^{\hat{\mu}}U^{\hat{\nu}} = R_{\hat{0}\hat{0}}$. We therefore have (in these coordinates, at this point)

$$\frac{d\theta}{d\tau} = -R_{\hat{0}\hat{0}}. \tag{4.82}$$

Now we can turn to Einstein's equation, in the form

$$R_{\mu\nu} = 8\pi G\left(T_{\mu\nu} - \tfrac{1}{2}Tg_{\mu\nu}\right). \tag{4.83}$$

Since we are in locally inertial coordinates, we have

$$g_{\hat{\mu}\hat{\nu}} = \eta_{\hat{\mu}\hat{\nu}} \tag{4.84}$$

$$T = g^{\hat{\mu}\hat{\nu}}T_{\hat{\mu}\hat{\nu}} = -\rho + p_x + p_y + p_z, \tag{4.85}$$

where $\rho = T_{\hat{0}\hat{0}}$ is the rest-frame energy density and $p_k = T_{\hat{k}\hat{k}}$ is the pressure in the $x^{\hat{k}}$ direction. Thus, (4.82) becomes

$$\frac{d\theta}{d\tau} = -4\pi G(\rho + p_x + p_y + p_z). \tag{4.86}$$

This equation is telling us that energy and pressure create a gravitational field that works to decrease the volume of our initially stationary ball of test particles (if ρ and the p_i's are all positive). In other words, gravity is attractive.

Of course, from (4.86) we see that gravity is not *necessarily* attractive; we could imagine sources for which $\rho+p_x+p_y+p_z$ were a negative number. Clearly, the role of pressure bears noting. For one thing, it represents an unambiguous departure from Newtonian theory, in which the pressure does not influence gravity (it doesn't appear in Poisson's equation, $\nabla^2\Phi = 4\pi G\rho$). The difference is hard to notice in our Solar System, since the pressure in the Sun and planets is much less than the energy density, which is dominated by the rest masses of the constituent particles. For another thing, notice that the *gravitational* effect of the pressure is opposite to that of the *direct* effect with which we are more familiar, namely that positive pressure works to push things apart. In most circumstances the direct effect of pressure is much more noticeable. However, the pressure can only act directly when there is a pressure gradient (for example, a change in pressure between the interior and exterior of a piston), whereas the gravitational effect depends only on the value of the pressure locally. If there were a perfectly smooth pressure, it would only be detectable through its gravitational effect; an example is provided by vacuum energy, discussed in Section 4.5.

As a final comment on (4.86), let's point out that it is completely equivalent to Einstein's equation—they convey identical information. This very specific relation will hold for any set of initially motionless test particles; the only way this can happen is if all of the components of Einstein's equation are true. If we like, then, we can state Einstein's equation in words[1] as follows: "The expansion of the volume of any set of particles initially at rest is proportional to (minus) the sum of the energy density and the three components of pressure."

So Einstein's equation tells us that energy density and pressure affect the Ricci tensor in such a way as to attract particles together when ρ and p are positive. What about the components of the Riemann tensor that are not included in the Ricci tensor? In Chapter 3 we found that these components were described by the Weyl tensor (expressed here in four dimensions),

$$C_{\rho\sigma\mu\nu} = R_{\rho\sigma\mu\nu} + \tfrac{1}{3} g_{\rho[\mu} g_{\nu]\sigma} R - g_{\rho[\mu} R_{\nu]\sigma} + g_{\sigma[\mu} R_{\nu]\rho}. \tag{4.87}$$

The Ricci tensor is the trace of the Riemann tensor, while the Weyl tensor describes the trace-free part; together they provide a complete characterization of the curvature. Clearly, given some specified energy-momentum distribution, there is still some freedom in the choice of Weyl curvature, since there is no analogue of Einstein's equation to relate $C^{\rho}{}_{\sigma\mu\nu}$ algebraically to $T_{\mu\nu}$. This is exactly as it should be. Imagine for example a spacetime that is vacuum everywhere, $R_{\mu\nu} = 0$. Flat Minkowski space is a possible solution in such a case, but so is a gravitational wave propagating through empty spacetime (as we will discuss in Chapter 7).

Since only $R_{\mu\nu}$ enters Einstein's equation, it might appear that the components of $C_{\rho\sigma\mu\nu}$ are completely unconstrained. But recall that we are not permitted to arbitrarily specify the components of the curvature tensor throughout a manifold; they are related by the Bianchi identity,

$$\nabla_{[\lambda} R_{\rho\sigma]\mu\nu} = 0. \tag{4.88}$$

As you showed in Exercise 10 of Chapter 3, this identity implies a differential relation for the Weyl tensor of the form

$$\nabla^{\rho} C_{\rho\sigma\mu\nu} = \nabla_{[\mu} R_{\nu]\sigma} + \tfrac{1}{6} g_{\sigma[\mu} \nabla_{\nu]} R. \tag{4.89}$$

On the right-hand side, the Riemann tensor only appears via its contractions the Ricci scalar and tensor, which can be related to $T_{\mu\nu}$ by Einstein's equation; we therefore have

$$\nabla^{\rho} C_{\rho\sigma\mu\nu} = 8\pi G \left(\nabla_{[\mu} T_{\nu]\sigma} + \tfrac{1}{3} g_{\sigma[\mu} \nabla_{\nu]} T \right). \tag{4.90}$$

So, while $R_{\mu\nu}$ and $T_{\mu\nu}$ are related algebraically through Einstein's equation, $C_{\rho\sigma\mu\nu}$ and $T_{\mu\nu}$ are related by this first-order differential equation. There will be

[1] J.C Baez, "The Meaning of Einstein's Equation," http://arXiv.org/abs/gr-qc/0103044.

a number of possible solutions for a given energy-momentum distribution, each specified by certain boundary conditions. This equation can be thought of as a propagation equation for gravitational waves, in close analogy with Maxwell's equations $\nabla_\mu F^{\nu\mu} = J^\nu$.

Having listed all of these lovely properties of Einstein's equation, it seems only fair that we should mention one distressing feature: the well-known difficulty of reconciling general relativity with quantum mechanics. GR is a classical field theory: the dynamical variable is a field (the metric) defined on spacetime, and coordinate-invariant quantities constructed from this field (such as the curvature scalar) can in principle be specified and measured to arbitrary accuracy. In the case of other field theories, such as electromagnetism, there are well-understood procedures for beginning with the classical theory and quantizing it, to obtain the dynamics of operators acting on wave functions living in a Hilbert space. For GR, the usual procedures run into both technical and conceptual difficulties, a description of which is beyond the scope of this book. One aspect of the technical difficulties is that GR is not "renormalizable" in the way that the Standard Model of particle physics is; when considering higher-order quantum effects, infinities appear that cannot be absorbed in any finite number of parameters. Nonrenormalizability does not mean that theory is fundamentally incorrect, but is a strong suggestion that it should only be taken seriously up to a certain energy scale.

Fortunately, the regime in which observable effects of quantum gravity are expected to become important is far from our everyday experience (or, for that matter, any conditions we can produce in the lab). Way back in 1899 Planck noticed that his constant h, for which nowadays we more often substitute $\hbar = h/2\pi = 1.05 \times 10^{-27}$ cm^2 g/sec, could be combined with Newton's constant $G = 6.67 \times 10^{-8}$cm^3 g^{-1} sec^{-2} and the speed of light $c = 3.00 \times 10^{10}$cm sec^{-1} to form a basic set of dimensionful quantities: the Planck mass,

$$m_{\rm P} = \left(\frac{\hbar c}{G}\right)^{1/2} = 2.18 \times 10^{-5}\,{\rm g}, \tag{4.91}$$

the Planck length,

$$l_{\rm P} = \left(\frac{\hbar G}{c^3}\right)^{1/2} = 1.62 \times 10^{-33}\,{\rm cm}, \tag{4.92}$$

the Planck time,

$$t_{\rm P} = \left(\frac{\hbar G}{c^5}\right)^{1/2} = 5.39 \times 10^{-44}\,{\rm sec}, \tag{4.93}$$

and the Planck energy,

$$E_{\rm P} = \left(\frac{\hbar c^5}{G}\right)^{1/2} = 1.95 \times 10^{16}\,{\rm erg} \tag{4.94}$$

$$= 1.22 \times 10^{19}\,{\rm GeV}. \tag{4.95}$$

A GeV is 10^9 electron volts, a common unit in particle physics, as it is approximately the mass of a proton. We usually set $\hbar = c = 1$, so that these quantities are all indistinguishable in the sense that $m_P = l_P^{-1} = t_P^{-1} = E_P$. You will hear people say things like "the Planck mass is 10^{19} GeV"; or simply refer to "the Planck scale." Another commonly used quantity is the reduced Planck scale, $\bar{m}_P = m_P/\sqrt{8\pi} = 2.43 \times 10^{18}$ GeV, which is often more convenient in equations—note that the coefficient of the curvature scalar in (4.73) is $\bar{m}_P^2/2$. Most likely, quantum gravity does not become important until we consider particle masses greater than m_P, or times shorter than t_P, or lengths smaller than l_P, or energies higher than E_P; at lower scales, classical GR should suffice. Since these are all far removed from observable phenomena, constructing a consistent theory of quantum gravity is more an issue of principle than of practice. On the other hand, quantum effects in curved spacetime might be important in the real world; as we will discuss in Chapter 8, they might lead to density fluctuations in the early universe, which grow into the galaxies and large-scale structure we observe today.

There is a leading contender for a fully quantum theory that would encompass GR in the appropriate limit: string theory. In string theory we imagine that the fundamental objects are not point particles like electrons or photons, but rather small one-dimensional objects called strings, which can be either closed loops or open segments. String theory was originally proposed as a model of the strong nuclear force, but it was soon realized that the theory inevitably predicted a massless spin-two particle: exactly what a quantum theory of gravity would require. String theory seems to be a consistent quantum theory, and it predicts gravity, but there is still a great deal about it that we don't understand. In particular, the way in which a classical spacetime arises out of fundamental strings is somewhat mysterious, and the connection to direct experiments is tenuous at best. Nevertheless, string theory is remarkably rich and robust, and promises to be an important part of theoretical physics for the foreseeable future.

4.5 ■ THE COSMOLOGICAL CONSTANT

A characteristic feature of general relativity is that the source for the gravitational field is the entire energy-momentum tensor. In nongravitational physics, only *changes* in energy from one state to another are measurable; the normalization of the energy is arbitrary. For example, the motion of a particle with potential energy $V(x)$ is precisely the same as that with a potential energy $V(x) + V_0$, for any constant V_0. In gravitation, however, the actual value of the energy matters, not just the differences between states.

This behavior opens up the possibility of **vacuum energy**: an energy density characteristic of empty space. One feature that we might want the vacuum to exhibit is that it not pick out a preferred direction; it will still be possible to have a nonzero energy density if the associated energy-momentum tensor is Lorentz invariant in locally inertial coordinates. Lorentz invariance implies that the corresponding energy-momentum tensor should be proportional to the metric,

$$T^{(\text{vac})}_{\hat{\mu}\hat{\nu}} = -\rho_{\text{vac}}\eta_{\hat{\mu}\hat{\nu}}, \tag{4.96}$$

since $\eta_{\hat{\mu}\hat{\nu}}$ is the only Lorentz invariant (0, 2) tensor. This generalizes straightforwardly from inertial coordinates to arbitrary coordinates as

$$T^{(\text{vac})}_{\mu\nu} = -\rho_{\text{vac}}g_{\mu\nu}. \tag{4.97}$$

Comparing to the perfect-fluid energy-momentum tensor $T_{\mu\nu} = (\rho + p)U_\mu U_\nu + pg_{\mu\nu}$, we find that the vacuum looks like a perfect fluid with an isotropic pressure opposite in sign to the energy density,

$$p_{\text{vac}} = -\rho_{\text{vac}}. \tag{4.98}$$

The energy density should be constant throughout spacetime, since a gradient would not be Lorentz invariant.

If we decompose the energy-momentum tensor into a matter piece $T^{(\text{M})}_{\mu\nu}$ and a vacuum piece $T^{(\text{vac})}_{\mu\nu} = -\rho_{\text{vac}}g_{\mu\nu}$, Einstein's equation is

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G\left(T^{(\text{M})}_{\mu\nu} - \rho_{\text{vac}}g_{\mu\nu}\right). \tag{4.99}$$

Soon after inventing GR, Einstein tried to find a static cosmological model, since that was what astronomical observations of the time seemed to imply. The result was the Einstein static universe, which will be discussed in Chapter 8. In order for this static cosmology to solve the field equation with an ordinary matter source, it was necessary to add a new term called the **cosmological constant**, Λ, which enters as

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu}. \tag{4.100}$$

From comparison with (4.99), we see that the cosmological constant is precisely equivalent to introducing a vacuum energy density

$$\rho_{\text{vac}} = \frac{\Lambda}{8\pi G}. \tag{4.101}$$

The terms "cosmological constant" and "vacuum energy" are essentially interchangeable.

Is a nonzero vacuum energy something we should expect? We arrived at the Hilbert Lagrangian $\widehat{\mathcal{L}}_H = R$ by looking for the simplest possible scalar we could construct from the metric. Of course there is an even simpler one, namely a constant. Using (4.69), it is straightforward to check that

$$S = \int d^4x\,\sqrt{-g}\left[\frac{1}{16\pi G}(R - 2\Lambda) + \widehat{\mathcal{L}}_{\text{M}}\right] \tag{4.102}$$

leads to the modified equation (4.100); alternatively, the vacuum Lagrangian is simply

$$\widehat{\mathcal{L}}_{\mathrm{vac}} = -\rho_{\mathrm{vac}}. \tag{4.103}$$

So it is certainly easy to introduce vacuum energy; however, we have no insight into its expected value, since it enters as an arbitrary constant.

The vacuum energy ultimately is a constant of nature in its own right. (An exception occurs in certain theories where a spacetime symmetry such as supersymmetry or conformal invariance governs the value of the vacuum energy; here we are considering a more generic field theory.) Nevertheless, there are various distinct contributions to the vacuum energy, and it would be strange if the total value were much smaller than the individual contributions. One such contribution comes from zero-point fluctuations—the energies of quantum fields in their vacuum state.

Consider a simple harmonic oscillator, a particle moving in a one-dimensional potential $V(x) = \frac{1}{2}\omega^2 x^2$. Classically, the vacuum for this system is the state in which the particle is motionless and at the minimum of the potential ($x = 0$), for which the energy in this case vanishes. Quantum-mechanically, however, the uncertainty principle forbids us from isolating the particle both in position and momentum, and we find that the lowest energy state has an energy $E_0 = \frac{1}{2}\hbar\omega$ (where we have temporarily reintroduced explicit factors of $\hbar$ for clarity). Of course, in the absence of gravity, either system actually has a vacuum energy that is completely arbitrary; we could add any constant to the potential without changing the theory. But quantum fluctuations have changed the zero-point energy from our classical expectation.

A precisely analogous situation holds in field theory. If we take the Fourier transform of a free quantum field (one where we ignore interactions for simplicity), we find that it becomes an infinite number of harmonic oscillators in momentum space, as we discuss in Chapter 9. The frequency ω of each oscillator is $\omega = \sqrt{m^2 + k^2}$, where m is the mass of the field and k is the magnitude of the wave vector of the mode. If we set the classical vacuum energy to zero, each of these modes contributes a quantum zero-point energy of $\hbar\omega/2$. Formally, adding all of these contributions together yields an infinite result. If, however, we discard the very high-momentum modes on the grounds that we trust our theory only up to a certain ultraviolet momentum cutoff k_{max}, we find that the resulting energy density is of the form

$$\rho_{\mathrm{vac}} \sim \hbar k_{\mathrm{max}}^4. \tag{4.104}$$

This answer could have been guessed by dimensional analysis; the numerical constants that have been neglected will depend on the precise theory under consideration. If we are confident that we can use ordinary quantum field theory all the way up to the reduced Planck scale $\bar{m}_{\mathrm{P}} = (8\pi G)^{-1/2} \sim 10^{18}$ GeV, we expect a contribution of order

$$\rho_{\mathrm{vac}} \sim (10^{18}\ \mathrm{GeV})^4 \sim 10^{112}\ \mathrm{erg/cm}^3. \tag{4.105}$$

Field theory may fail earlier, although quantum gravity is the best reason we have to believe it will fail at any specific scale.

As we will discuss in Chapter 8, cosmological observations imply

$$|\rho_\Lambda^{(\mathrm{obs})}| \le (10^{-12}\,\mathrm{GeV})^4 \sim 10^{-8}\,\mathrm{erg/cm^3}, \tag{4.106}$$

much smaller than the naive expectation just derived. The ratio of (4.105) to (4.106) is the origin of the famous discrepancy of 120 orders of magnitude between the theoretical and observational values of the cosmological constant. We are free to imagine that the bare vacuum energy is adjusted so that the net cosmological constant is consistent with the limit (4.106), except for one problem: we know of no special symmetry that could enforce a vanishing vacuum energy while remaining consistent with the known laws of physics; this conundrum is the "cosmological constant problem." We will discuss the cosmological effects of vacuum energy more in Chapter 8.[2]

4.6 ■ ENERGY CONDITIONS

Sometimes it is useful to think about Einstein's equation without specifying the theory of matter from which $T_{\mu\nu}$ is derived. This leaves us with a great deal of arbitrariness; consider for example the question, What metrics obey Einstein's equation? In the absence of some constraints on $T_{\mu\nu}$, the answer is any metric at all; simply take the metric of your choice, compute the Einstein tensor $G_{\mu\nu}$ for this metric, and then demand that $T_{\mu\nu}$ be equal to $G_{\mu\nu}$. It will automatically be conserved, by the Bianchi identity. Our real concern is with the existence of solutions to Einstein's equation in the presence of "realistic" sources of energy and momentum, whatever that means. One strategy is to consider specific kinds of sources, such as scalar fields, dust, or electromagnetic fields. However, we occasionally wish to understand properties of Einstein's equations that hold for a variety of different sources. In this circumstance it is convenient to impose *energy conditions* that limit the arbitrariness of $T_{\mu\nu}$.

Energy conditions are coordinate-invariant restrictions on the energy-momentum tensor. We must therefore construct scalars from $T_{\mu\nu}$, which is typically accomplished by contracting with arbitrary timelike vectors t^μ or null vectors ℓ^μ. For example, the weak energy condition (WEC) states that $T_{\mu\nu}t^\mu t^\nu \ge 0$ for all timelike vectors t^μ. For purposes of physical intuition, it is useful to consider the special case where the source is a perfect fluid, so that the energy-momentum tensor takes the form

$$T_{\mu\nu} = (\rho + p)U_\mu U_\nu + p g_{\mu\nu}, \tag{4.107}$$

where U^μ is the fluid four-velocity. Let's use this form to translate the WEC into physical terms. Because the pressure is isotropic, $T_{\mu\nu}t^\mu t^\nu$ will be nonnegative

[2]For more on the physics and cosmology of vacuum energy, see S.M. Carroll, *Liv. Rev. Rel.* **4**, 1 (2001), `http://arxiv.org/astro-ph/0004075`.

for all timelike vectors t^μ if both $T_{\mu\nu}U^\mu U^\nu \geq 0$ and $T_{\mu\nu}\ell^\mu\ell^\nu \geq 0$ for some null vector ℓ^μ (convince yourself of this; it's just adding vectors). We therefore evaluate

$$T_{\mu\nu}U^\mu U^\nu = \rho, \quad T_{\mu\nu}\ell^\mu\ell^\nu = (\rho + p)(U_\mu \ell^\mu)^2. \tag{4.108}$$

The WEC therefore implies $\rho \geq 0$ and $\rho + p \geq 0$. These are simply the reasonable-sounding requirements that the energy density be nonnegative and the pressure not be too large compared to the energy density. Of course we need not restrict ourselves to perfect fluids, we merely use them to gain insight into the requirements the energy conditions impose.

There are a number of different energy conditions, appropriate to different circumstances. Some of the most popular are the following:

- The **Weak Energy Condition** or WEC, as just discussed, states that $T_{\mu\nu}t^\mu t^\nu \geq 0$ for all timelike vectors t^μ, or equivalently that $\rho \geq 0$ and $\rho + p \geq 0$.

- The **Null Energy Condition** or NEC states that $T_{\mu\nu}\ell^\mu\ell^\nu \geq 0$ for all null vectors ℓ^μ, or equivalently that $\rho + p \geq 0$. It is a special case of the WEC, with the timelike vector replaced by a null vector. The energy density may now be negative, so long as there is a compensating positive pressure.

- The **Dominant Energy Condition** or DEC includes the WEC ($T_{\mu\nu}t^\mu t^\nu \geq 0$ for all timelike vectors t^μ), as well as the additional requirement that $T^{\mu\nu}t_\mu$ is a nonspacelike vector (namely, that $T_{\mu\nu}T^\nu{}_\lambda t^\mu t^\lambda \leq 0$). For a perfect fluid, these conditions together are equivalent to the simple requirement that $\rho \geq |p|$; the energy density must be nonnegative, and greater than or equal the magnitude of the pressure.

- The **Null Dominant Energy Condition** or NDEC is the DEC condition for null vectors only: for any null vector ℓ^μ, $T_{\mu\nu}\ell^\mu\ell^\nu \geq 0$ and $T^{\mu\nu}\ell_\mu$ is a nonspacelike vector. The allowed density and pressure are the same as for the DEC, except that negative densities are allowed so long as $p = -\rho$. In other words, the NDEC excludes all sources excluded by the DEC, except for a negative vacuum energy.

- The **Strong Energy Condition** or SEC states that $T_{\mu\nu}t^\mu t^\nu \geq \frac{1}{2}T^\lambda{}_\lambda t^\sigma t_\sigma$ for all timelike vectors t^μ, or equivalently that $\rho + p \geq 0$ and $\rho + 3p \geq 0$. Note that the SEC does *not* imply the WEC. It implies the NEC, along with excluding excessively large negative pressures. From (4.86) we see that it is the SEC that implies gravitation is attractive.

These conditions are illustrated in Figure 4.3. In addition we have plotted the constraint $w \geq -1$, where $w = p/\rho$ is called the **equation-of-state parameter**. This is a useful concept in cosmology, where sources often have equations of state $p = w\rho$ with w being a constant (of course, w is defined whether it is constant

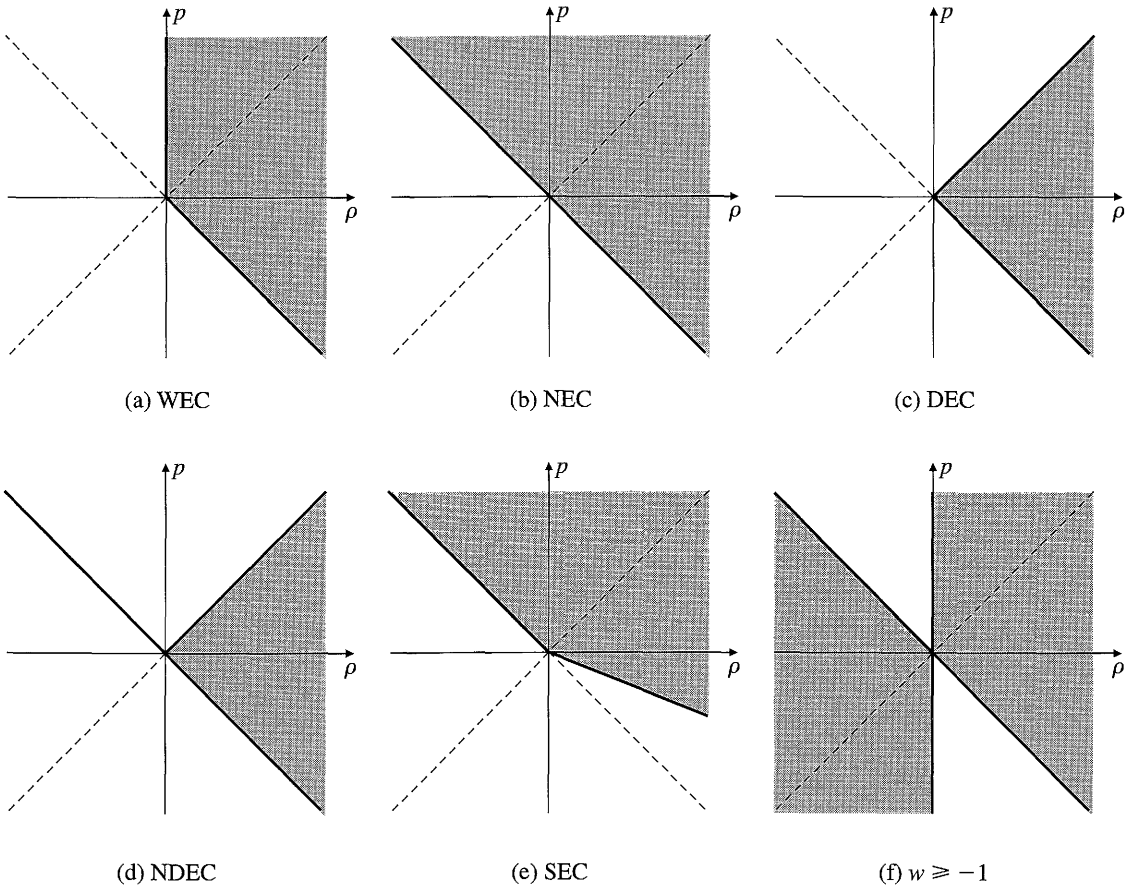

FIGURE 4.3 Energy conditions as applied to perfect fluids, expressed as allowed regions of energy density ρ and pressure p. Illustrated are the Weak Energy Condition (WEC), Null Energy Condition (NEC), Dominant Energy Condition (DEC), Null Dominant Energy Condition (NDEC), and the Strong Energy Condition (SEC). For comparison, we also have illustrated the condition $w \geq -1$, where $w = p/\rho$ is the equation-of-state parameter.

or not). If we restrict ourselves to sources with $\rho \geq 0$, then any of the energy conditions mentioned above will imply $w \geq -1$.

Most ordinary classical forms of matter, including scalar fields and electromagnetic fields, obey the DEC (see Exercises), and hence the less restrictive conditions (WEC, NEC, NDEC). The SEC is useful in the proof of some singularity theorems, but can be violated by certain forms of matter, such as a massive scalar field. It turns out that quantum fields can generically violate any of the energy conditions we have listed; there may, however, be inequalities involving integrals over regions of spacetime that are satisfied even by quantum fields. This is an area of current investigation.

The energy conditions are not, strictly speaking, related to energy conservation; the Bianchi identity guarantees that $\nabla_\mu T^{\mu\nu} = 0$ regardless of whether we impose any additional constraints on $T^{\mu\nu}$. Rather, they serve to prevent other properties that we think of as "unphysical," such as energy propagating faster than the speed of light, or empty space spontaneously decaying into compensating regions of positive and negative energy. In particular, Hawking and Ellis (1973) prove a *conservation theorem*: Essentially, if the energy-momentum tensor obeys the DEC and vanishes in some spacelike region, then it will necessarily vanish everywhere in the future domain of dependence of that region (see Section 2.7 for the definition of the future domain of dependence). Thus, energy cannot spontaneously appear from nothing, nor can it sneak outside the light cone. The theorem does not include the converse statement (that sources violating the DEC are necessarily acausal), so it pays to be careful.

4.7 ■ THE EQUIVALENCE PRINCIPLE REVISITED

In this section we will examine more carefully the underpinnings and consequences of the Principle of Equivalence, which we used in Section 4.1 to motivate the minimal-coupling procedure for generalizing physics to curved spacetime. We will see that the Principle of Equivalence is not a sacred physical law, nor is it even a mathematically rigorous statement; at a more fundamental level, it arises as a consequence of the nature of general relativity as an effective field theory valid at macroscopic distances, and our job is to determine which kinds of couplings between matter and the metric we would expect in such a theory.

In practice, it is common to invoke the Equivalence Principle to justify any of the following four ideas:

1. Laws of physics should be expressed (or at least be expressible) in generally covariant form.
2. There exists a metric on spacetime, the curvature of which we interpret as gravity.
3. There do not exist any other fields that resemble gravity.
4. The interactions of matter fields to curvature are minimal: they do not involve direct couplings to the Riemann tensor or its contractions.

These very different statements each have a very different status: the first is vacuous, the second is both profound and almost certainly true, the third is interesting and testable, and the fourth is just a useful approximation. Let's examine each of them in turn.

The first statement is sometimes called the Principle of Covariance. It is more or less content-free. "Generally covariant" simply means that all of the terms in an equation transform in the same way under a change of coordinates, so that the form of the equation is coordinate-invariant. Due to the universal nature of the tensor transformation law, the most straightforward way of achieving this aim is to make the equation manifestly tensorial. Certainly there is nothing wrong if a law is expressed in a form that is not generally covariant, as long as we know that it is possible to rewrite it in a coordinate-independent way. On the other hand, it is *always* possible to write laws in a coordinate-independent way, if the laws are well-defined to begin with. A physical system acting in a certain way doesn't know which coordinate system you are using to describe it; consequently, anything deserving of the name "law of physics" (as opposed to some particular statement of that law) must be independent of coordinates. An insistence on explicit coordinate-independence says nothing about the adaptation of laws to curved spacetime; as we have seen, manifestly tensorial equations take on the same form regardless of the geometry.

Consider Maxwell's equations in flat spacetime, as we wrote them in Chapter 1:

$$\partial_\mu F^{\nu\mu} = J^\nu. \tag{4.109}$$

The right-hand side is a well-defined tensor, while the left-hand side is not, due to the appearance of the partial derivative. That's okay, since we know that this equation is valid only in inertial coordinates in Minkowski space. A coordinate-invariant way of expressing the same law is

$$\nabla_\mu F^{\nu\mu} = J^\nu. \tag{4.110}$$

No physical principle needs to be invoked to conclude that this is the correct formulation in Minkowski space; it is the *unique* tensorial equation, which is equivalent to (4.109) in inertial coordinates. It is not the unique generalization to curved spacetime, since we could imagine new terms involving products of $F_{\mu\nu}$ and $R^\rho{}_{\sigma\mu\nu}$; the status of such additional terms is directly addressed by the minimal-coupling assumption, point four in the above list. By itself, however, making things "tensorial" or "generally covariant" is a simple matter of logical necessity, not a physical principle that one could imagine disproving by experiment. (Another spin on the same idea is "diffeomorphism invariance," discussed in the Appendix B.)

The second purported consequence of the Equivalence Principle from our list above is much deeper, and by no means obvious. Although he was inspired by the EP, this geometric insight was Einstein's great breakthrough. At the beginning of Chapter 2 we discussed why such an insight was warranted: the EP implies that gravity is universal, which implies in turn that gravitational fields become impossible to measure in small regions of spacetime, a feature which in turn is most directly implemented by identifying gravitation with the effects of spacetime geometry. These steps are well-motivated suggestions, not rigorously derived consequences; once we have the idea that there is a metric whose curvature gives rise to gravity, we can check its usefulness by comparing with experiment. As we've mentioned, it passes with flying colors. An accumulation of evidence (such as the gravitational redshift discussed in Chapter 2) is consistent with the idea that idealized rods and clocks behave as they should if the geometry of spacetime were curved. Still, one should not imagine proving that there really is a metric with the desired properties; we make the hypothesis, test it against ever-more precise experiments, and deduce its range of usefulness. Indeed, the demands of eventually reconciling general relativity with quantum mechanics suggest to many that the metric will ultimately be revealed as a concept derived from a more fundamental collection of degrees of freedom. For our present purposes this ultimate resolution doesn't matter; the idea of a curved metric has proven its usefulness beyond a reasonable doubt, and we work to extend out understanding of its properties until they run up against insurmountable obstacles (either theoretical or empirical).

Given our conviction that the effects of gravitation are best ascribed to the curvature of a metric on spacetime, what would we conclude if experiments were to detect an apparent violation of the Equivalence Principle? For example, we might imagine an experiment that revealed that the acceleration of test bodies in the direction of the Earth or Sun actually did depend, ever so slightly, on the composition of the test body. (The best current limits on such anomalous accelerations constrain them to be less than 10^{-12} times that due to gravity.)[3] In such a circumstance, nobody would really be tempted to declare that general relativity had been completely undermined and it was necessary to start over. Rather, we would return to the definition of "test body," which includes the proviso that the body be uncharged. An electron, for example, would not make a good test body, as it would be buffeted about by ambient electromagnetic fields as well as by gravity. Similarly, by far the most straightforward explanation of any hypothetical anomalous acceleration on purportedly neutral test bodies would be to imagine that we had discovered the existence of a new long-range field, under which our test bodies were actually charged. To have remain undetected thus far, such a field must be either very weakly coupled, or must couple almost universally, so as to mimic the effects of gravity. We could imagine, for example, scalar fields that couple to the trace of the energy-momentum tensor, or vector fields that couple to baryon number. The mass of ordinary test bodies is almost proportional to their baryon number, which counts the number of protons and neutrons in the body. It is therefore sometimes convenient to think of "tests of the Equivalence Principle" as tests of the third of our statements above—that there do not exist any other fields that resemble gravity (where a field resembles gravity if it is long-range and couples almost universally to mass). Again, detecting a violation of this hypothesis would be most directly interpreted as discovery of a new "fifth force" rather than as a repudiation of Einstein's ideas. As to whether we should expect to discover such a new field if we improve upon current experiments, it is hard to say; on the one hand, it is easy to concoct models with new long-range forces, but on the other hand, they would typically be strong enough to already have been detected. At this stage it is still worthwhile to keep an open mind.

Beyond the very existence of the metric, the heart of the Equivalence Principle lies in the fourth of our formulations, that the interactions of matter fields to curvature are minimal: they do not involve direct couplings to the Riemann tensor or its contractions. For example, we could consider the following possible alternative

[3] Y. Su et al., *Phys. Rev.* **D 50**, 3614 (1994).

to the conventional geodesic equation:

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\sigma} \frac{dx^\rho}{d\lambda} \frac{dx^\sigma}{d\lambda} = \alpha \left(\nabla_\sigma R\right) \frac{dx^\mu}{d\lambda} \frac{dx^\sigma}{d\lambda}, \tag{4.111}$$

where R is the Ricci scalar and α is a coupling constant. This equation also reduces to straight-line motion in flat spacetime, but would allow for direct detection of spacetime curvature in small regions by measurement of the coupling to $\nabla_\sigma R$. Why, then, does nature choose the simple geodesic equation? As a first step toward an answer, consider the dimensions of the coupling α. Since $c = 1$ and space and time have the same units, we can use length as our basic dimension. The metric, the inverse metric, and $dx^\mu/d\lambda$ are then dimensionless. The partial derivative operator has units of inverse length, as does the covariant derivative. The Christoffel symbols involve first derivatives of the metric, and thus have dimensions of inverse length; similarly, the Riemann tensor, Ricci tensor, and Ricci scalar have dimensions of inverse length squared:

$$\left[\frac{dx^\mu}{d\lambda}\right] = [g_{\mu\nu}] = [g^{\mu\nu}] = L^0, \quad [\nabla_\mu] = [\Gamma^\mu_{\rho\sigma}] = L^{-1}, \quad [R] = L^{-2}. \tag{4.112}$$

To be consistent, the coupling α must have dimensions of length squared:

$$[\alpha] = L^2. \tag{4.113}$$

The square root of α therefore defines a length scale; what should the length scale be? We don't know for sure, but there is every reason to believe it should be extremely small. There are two arguments for this. One is that, since the coupling represented by α is of gravitational origin, the only reasonable expectation for the relevant length scale is

$$\alpha \sim l_{\mathrm{P}}^2, \tag{4.114}$$

where l_{P} is the Planck length. Another reason is simply a more sophisticated version of this "what else could it be?" rationale. Although general relativity is a classical theory, at a deeper level we expect that it is merely an effective field theory describing an underlying quantum-mechanical structure. Even without knowing what this structure may be, a generic expectation (derived from our experience with quantum field theories we do understand) is that the effective classical limit should contain all possible interactions, but with dimensionful length parameters representing scales at which new degrees of freedom become important (recall our discussion of effective field theory at the end of Chapter 1). Thus, the Fermi theory of the weak interactions contains a length scale, which we now know to correspond to the scale of electroweak symmetry breaking where W and Z bosons become relevant. Since we do not expect new gravitational physics to arise before the Planck scale, the higher-order interactions associated with gravity should be suppressed by appropriate powers of the Planck length.

How much suppression does this represent? One measure would be to compare $l_{\rm P}$ (and thus the likely value of the parameter α) to a typical gravitational length scale near the vicinity of the Earth. The strength of gravity on Earth is characterized by the acceleration due to gravity, $a_g = 980$ cm/sec^2. To construct a quantity with dimensions of length, we define

$$l_\oplus = c^2/a_g \sim 10^{18}\ \text{cm}, \tag{4.115}$$

where the symbol $\oplus$ in this context stands for the Earth (not a direct sum). So the relative strength of higher-order gravitational effects is measured by

$$\frac{l_{\rm P}}{l_\oplus} \sim 10^{-51}. \tag{4.116}$$

In fact, since we expect $\alpha \sim l_{\rm P}^2$, the suppression will be of order 10^{-102}. Consequently, there seems to be little need to worry about the possible role of such couplings. But dramatic departures should be kept in mind; recent ideas about large extra dimensions have opened up the possibility of observing direct gravitational interactions at particle accelerators. Ultimately, there is no way to resolve these problems by pure thought alone; only experiment can decide among the alternatives.

4.8 ■ ALTERNATIVE THEORIES

General relativity has passed a wide variety of experimental tests. Nevertheless, it is always possible that the next experiment we do will reveal a deviation from Einstein's original formulation. Let us therefore briefly consider ways in which general relativity could be modified. There are an uncountable number of such ways, but we will consider four different possibilities:

- gravitational scalar fields
- extra spatial dimensions
- higher-order terms in the action
- nonChristoffel connections

A popular set of alternative models are known as **scalar-tensor theories** of gravity, since they involve both the metric tensor, $g_{\mu\nu}$ and a scalar field, λ. In particular, the scalar field couples directly to the curvature scalar, not simply to the metric (as the Equivalence Principle would seem to imply). The action can be written as a sum of a gravitational piece, a pure-scalar piece, and a matter piece:

$$S = S_{fR} + S_\lambda + S_{\rm M}, \tag{4.117}$$

where

$$S_{fR} = \int d^4x \sqrt{-g}\, f(\lambda) R, \tag{4.118}$$

$$S_\lambda = \int d^4x \sqrt{-g} \left[-\frac{1}{2} h(\lambda) g^{\mu\nu} (\partial_\mu \lambda)(\partial_\nu \lambda) - U(\lambda) \right], \tag{4.119}$$

and

$$S_{\rm M} = \int d^4x \sqrt{-g}\, \widehat{\mathcal{L}}_{\rm M}(g_{\mu\nu}, \psi_i). \tag{4.120}$$

Here, $f(\lambda)$, $h(\lambda)$ and $U(\lambda)$ are functions that define the theory, and the matter Lagrangian $\widehat{\mathcal{L}}_{\rm M}$ depends on the metric and a set of matter fields ψ_i, but not on λ. By change of variables we can always set $h(\lambda) = 1$, but we leave it here to facilitate comparison with models found in the literature.

The equations of motion for this theory include the gravitational equation (from varying with respect to the metric), and the scalar equation (from varying with respect to λ), as well as the appropriate matter equations. Let's start with the gravitational equation, which we can derive by following the same steps as for the ordinary Hilbert action (4.55). We consider perturbations of the metric,

$$g^{\mu\nu} \to g^{\mu\nu} + \delta g^{\mu\nu}. \tag{4.121}$$

Following the procedure from Section 4.3, the variation of the gravitational part of the action is

$$\delta S_{fR} = \int d^4x \sqrt{-g}\, f(\lambda) \Big[\left(R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \right) \delta g^{\mu\nu} + \nabla_\sigma \nabla^\sigma (g_{\mu\nu} \delta g^{\mu\nu}) \\ - \nabla_\mu \nabla_\nu (\delta g^{\mu\nu}) \Big]. \tag{4.122}$$

For the Hilbert action, f is a constant, so the last two terms are total derivatives, which can be converted to surface terms through integration by parts and therefore ignored. Now integration by parts (twice) picks up derivatives of f, and we obtain

$$\delta S_{fR} = \int d^4x \sqrt{-g} \left[f(\lambda) G_{\mu\nu} + g_{\mu\nu} \Box f - \nabla_\mu \nabla_\nu f \right] \delta g^{\mu\nu}, \tag{4.123}$$

where $G_{\mu\nu}$ is the Einstein tensor. We have discarded surface terms as usual, although there are subtleties concerning boundary contributions in this case; see Wald (1984) for a discussion. The gravitational equation of motion, including contributions from S_λ and $S_{\rm M}$, is thus

$$G_{\mu\nu} = f^{-1}(\lambda) \left(\tfrac{1}{2} T^{(\rm M)}_{\mu\nu} + \tfrac{1}{2} T^{(\lambda)}_{\mu\nu} + \nabla_\mu \nabla_\nu f - g_{\mu\nu} \Box f \right), \tag{4.124}$$

where the energy-momentum tensors are $T^{(i)}_{\mu\nu} = -2(-g)^{-1/2}\delta S_i/\delta g^{\mu\nu}$; in particular,

$$T^{(\lambda)}_{\mu\nu} = h(\lambda)\nabla_\mu\lambda\nabla_\nu\lambda - g_{\mu\nu}\left[\tfrac{1}{2}h(\lambda)g^{\rho\sigma}\nabla_\rho\lambda\nabla_\sigma\lambda + U(\lambda)\right]. \tag{4.125}$$

From looking at the coefficient of $T^{(\mathrm{M})}_{\mu\nu}$ in (4.124), we see that when the scalar field is *constant* (or practically so), we may identify $f(\lambda) = 1/(16\pi G)$, as makes sense from the original action (4.118). Meanwhile, if λ varies slightly from point to point in spacetime, it would be interpreted as a spacetime-dependent Newton's constant. The dynamics that control this variation are determined by the equation of motion for λ, which is straightforward to derive as

$$h\,\Box\lambda + \tfrac{1}{2}h'g^{\mu\nu}\nabla_\mu\lambda\nabla_\nu\lambda - U' + f'R = 0, \tag{4.126}$$

where primes denote differentiation with respect to λ. Notice that if we set $h(\lambda) = 1$ to get a conventional kinetic term for the scalar, λ obeys a conventional scalar-field equation of motion, with an additional coupling to the curvature scalar. In the real world, we don't want $f(\lambda)$ to vary too much, as it would have observable consequences in the classic experimental tests of GR in the solar system, and also in cosmological tests such as primordial nucleosynthesis. This can be ensured either by choosing $U(\lambda)$ so that there is a minimum to the potential and λ cannot deviate too far without a large input of energy—in other words, λ has a large mass—or by choosing $f(\lambda)$ and $h(\lambda)$ so that large changes in λ give rise to relatively small changes in the effective value of Newton's constant.

One of the earliest scalar-tensor models is known as Brans–Dicke theory, and corresponds in our notation to the choices

$$f(\lambda) = \frac{\lambda}{16\pi}, \qquad h(\lambda) = \frac{\omega}{8\pi\lambda}, \qquad U(\lambda) = 0. \tag{4.127}$$

where ω is a coupling constant. The scalar-tensor action takes the form

$$S_{\mathrm{BD}} = \int d^4x\sqrt{-g}\left[\frac{\lambda}{16\pi}R - \frac{\omega}{16\pi}g^{\mu\nu}\frac{(\partial_\mu\lambda)(\partial_\nu\lambda)}{\lambda}\right]. \tag{4.128}$$

In the Brans–Dicke theory, the scalar field is massless, but in the $\omega \to \infty$ limit the field becomes nondynamical and ordinary GR is recovered. Current bounds from Solar System tests imply $\omega > 500$, so if there is such a scalar field it must couple only weakly to the Ricci scalar.

A popular approach to dealing with scalar-tensor theories is to perform a conformal transformation to bring the theory in to a form that looks like conventional GR. We define a conformal metric

$$\tilde{g}_{\mu\nu} = 16\pi\tilde{G}f(\lambda)g_{\mu\nu}, \tag{4.129}$$

where $\tilde{G}$ will become Newton's constant in the conformal frame. Using formulae for conformal transformations from the Appendix G, the action S_{fR} from (4.118) becomes

$$\begin{aligned} S_{fR} &= \int d^4x\sqrt{-g}\, f(\lambda)R \\ &= \int d^4x\sqrt{-\tilde{g}}\,(16\pi\tilde{G})^{-1}\left[\tilde{R} - \frac{3}{2}\tilde{g}^{\rho\sigma} f^{-2}\left(\frac{df}{d\lambda}\right)^2 (\tilde{\nabla}_\rho\lambda)(\tilde{\nabla}_\sigma\lambda)\right], \end{aligned} \tag{4.130}$$

where as usual we have integrated by parts and discarded surface terms. In the conformal frame, therefore, the curvature scalar appears by itself, not multiplied by any function of λ. This frame is sometimes called the **Einstein frame**, since Einstein's equations for the conformal metric $\tilde{g}_{\mu\nu}$ take on their conventional form. The original frame with metric $g_{\mu\nu}$ is called the **Jordan frame**, or sometimes the **string frame**. (String theory typically predicts a scalar-tensor theory rather than ordinary GR, and the string worldsheet responds to the metric $g_{\mu\nu}$.)

Before going on with our analysis of the conformally-transformed theory, consider what happens if we choose

$$f(\lambda) = e^{\lambda/\sqrt{3}}, \quad h(\lambda) = 0, \quad U(\lambda) = 0, \tag{4.131}$$

a specific choice for $f(\lambda)$, but turning off the pure scalar terms in S_λ. Then we notice that the Einstein frame action (4.130) actually includes a conventional kinetic term for the scalar, even though it wasn't present in the Jordan frame action (4.118). Even without an explicit kinetic term for λ, the degrees of freedom of this theory include a propagating scalar as well as the metric. This should hopefully become more clear after we examine the degrees of freedom of the gravitational field in Chapter 7. There we will find that the metric $g_{\mu\nu}$ actually includes scalar (spin-0) and vector (spin-1) degrees of freedom as well as the expected tensor (spin-2) degrees of freedom; however, with the standard Hilbert action, these degrees of freedom are constrained rather than freely propagating. What we have just found is that multiplying R by a scalar in the action serves to bring the scalar degree of freedom to life, which is revealed explicitly in the Einstein frame.

If we do choose to include the pure-scalar action S_λ, we obtain

$$S_{fR} + S_\lambda = \int d^4x\sqrt{-\tilde{g}}\left[\frac{\tilde{R}}{16\pi\tilde{G}} - \frac{1}{2}K(\lambda)\tilde{g}^{\rho\sigma}(\tilde{\nabla}_\rho\lambda)(\tilde{\nabla}_\sigma\lambda) - \frac{U(\lambda)}{(16\pi\tilde{G})^2 f^2(\lambda)}\right], \tag{4.132}$$

where

$$K(\lambda) = \frac{1}{16\pi\tilde{G}f^2}\left[fh + 3(f')^2\right]. \tag{4.133}$$

We can make our action look utterly conventional by defining a new scalar field ϕ via

$$\phi = \int K^{1/2}\, d\lambda, \tag{4.134}$$

in terms of which the action becomes

$$S_{fR} + S_\lambda = \int d^4x\sqrt{-\tilde{g}}\left[\frac{\tilde{R}}{16\pi\tilde{G}} - \frac{1}{2}\tilde{g}^{\rho\sigma}(\tilde{\nabla}_\rho\phi)(\tilde{\nabla}_\sigma\phi) - V(\phi)\right], \tag{4.135}$$

where

$$V(\phi) = \frac{U(\lambda(\phi))}{(16\pi\tilde{G})^2 f^2(\lambda(\phi))}. \tag{4.136}$$

Amazingly, in the Einstein frame we have a completely ordinary theory of a scalar field in curved spacetime. So long as $f(\lambda)$ is well-behaved, the variables $(\tilde{g}_{\mu\nu}, \phi)$ can be used instead of $(g_{\mu\nu}, \lambda)$, in the sense that varying with respect to the new variables is equivalent to starting with the original equations of motion (4.124) and (4.126) and then doing the transformations (4.129) and (4.134).

Finally, we add in the matter action (4.120). Varying with respect to $\tilde{g}_{\mu\nu}$ will yield an energy-momentum tensor in the Einstein frame. In the original variables $(g_{\mu\nu}, \lambda)$, we knew that $S_{\rm M}$ was independent of λ, but now it will depend on both of the new variables $(\tilde{g}_{\mu\nu}, \phi)$; we can use the chain rule to characterize this dependence. Let us also assume that $S_{\rm M}$ depends on $g_{\mu\nu}$ only algebraically, not through derivatives. This will hold for ordinary scalar-field or gauge-field matter; things become more complicated for fermions, which we won't discuss here. We obtain

$$\begin{aligned}
\tilde{T}_{\mu\nu} &\equiv -2\frac{1}{\sqrt{-\tilde{g}}}\frac{\delta S_{\rm M}}{\delta\tilde{g}^{\mu\nu}} \\
&= -2\frac{1}{\sqrt{-\tilde{g}}}\frac{\partial g^{\alpha\beta}}{\partial\tilde{g}^{\mu\nu}}\frac{\delta S_{\rm M}}{\delta g^{\alpha\beta}} \\
&= -2(16\pi\tilde{G}f)^{-1}\frac{1}{\sqrt{-g}}\delta^\alpha_\mu\delta^\beta_\nu\frac{\delta S_{\rm M}}{\delta g^{\alpha\beta}} \\
&= (16\pi\tilde{G}f)^{-1}T_{\mu\nu}.
\end{aligned} \tag{4.137}$$

A similar trick works for the coupling of matter to ϕ, which comes from varying $S_{\rm M}$ with respect to ϕ, using $g^{\alpha\beta} = 16\pi\tilde{G}f\tilde{g}^{\alpha\beta}$:

$$\begin{aligned}
\frac{\delta S_{\rm M}}{\delta\phi} &= \frac{\partial g^{\alpha\beta}}{\partial\phi}\frac{\delta S_{\rm M}}{\delta g^{\alpha\beta}} \\
&= \left(16\pi\tilde{G}\frac{df}{d\phi}\tilde{g}^{\alpha\beta}\right)\left(-\frac{1}{2}\sqrt{-g}T^{\rm M}_{\alpha\beta}\right) \\
&= -\frac{1}{2f}\frac{df}{d\phi}\sqrt{-\tilde{g}}\tilde{T}^{\rm M},
\end{aligned} \tag{4.138}$$

where

$$\widetilde{T}^{(\mathrm{M})} = \tilde{g}^{\alpha\beta}\widetilde{T}^{(\mathrm{M})}_{\alpha\beta} = \frac{1}{(16\pi\widetilde{G}f)^2} g^{\alpha\beta} T^{(\mathrm{M})}_{\alpha\beta} \tag{4.139}$$

is the trace of the energy-momentum tensor in the conformal frame.

Varying (4.135) with respect to $\tilde{g}_{\mu\nu}$ and ϕ returns equations of motion equivalent to Einstein's equations and an equation for ϕ. The gravitational equation is

$$\widetilde{G}_{\mu\nu} = 8\pi\widetilde{G}\left(\widetilde{T}^{(\mathrm{M})}_{\mu\nu} + \widetilde{T}^{(\phi)}_{\mu\nu}\right), \tag{4.140}$$

where

$$\widetilde{T}^{(\phi)}_{\mu\nu} = \widetilde{\nabla}_\mu\phi\widetilde{\nabla}_\nu\phi - \tilde{g}_{\mu\nu}\left[\tfrac{1}{2}\tilde{g}^{\rho\sigma}\widetilde{\nabla}_\rho\phi\widetilde{\nabla}_\sigma\phi + V(\phi)\right], \tag{4.141}$$

and the scalar field equation is

$$\widetilde{\Box}\phi - \frac{dV}{d\phi} = \frac{1}{2f}\frac{df}{d\phi}\widetilde{T}^{(\mathrm{M})}. \tag{4.142}$$

Given that (4.140) looks just like Einstein's equation with both matter and scalar-field sources, why should we even bother to call this scalar-tensor theory an alternative to GR? Isn't it the same theory, just in different variables? In fact it is not the same, because of the dependence of S_{M} on ϕ in the Einstein frame. In particular, physical test particles will move along geodesics of $g_{\mu\nu}$, which will not generally coincide with those of $\tilde{g}_{\mu\nu}$. The original metric is the one that test particles "see." So either we work in the original variables $(g_{\mu\nu}, \lambda)$, where the gravitational field equation is altered, or we use the new variables $(\tilde{g}_{\mu\nu}, \phi)$, in which the equations of motion for matter are altered; either way, there will be unambiguously measurable departures (in principle) from ordinary GR.

Another way to modify general relativity is to allow for the existence of extra spatial dimensions; in fact the physical consequences of extra dimensions turn out to be closely related to those of scalar-tensor theories. By extra dimensions we don't simply mean considering GR in higher-dimensional spaces, but rather considering models in which the spacetime appears four-dimensional on large scales even though there are really $4 + d$ total dimensions. The simplest way for this to happen is if the extra d dimensions are "compactified" on some manifold; it is this possibility we consider here.[4] Models of this kind are known as Kaluza-Klein theories.

Let G_{ab} be the metric for a $(4 + d)$-dimensional spacetime with coordinates X^a, where indices a, b run from 0 to $d + 3$.

[4]We follow the analysis of S.M. Carroll, J. Geddes, M. Hoffman, and R.M. Wald, *Phys. Rev. D* **66**, 024036 (2002); `http://arxiv.org/hep-th/0110149`. The original papers on extra dimensions are those by Kaluza and Klein: T. Kaluza, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **K1**, 966 (1921); O. Klein, *Z. Phys.* **37**, 895 (1926) [*Surveys High Energ. Phys.* **5**, 241 (1926)]; O. Klein, *Nature* **118**, 516 (1926).

$$ds^2 = G_{ab}\mathrm{d}X^a\mathrm{d}X^b = g_{\mu\nu}(x)\mathrm{d}x^\mu\mathrm{d}x^\nu + b^2(x)\gamma_{ij}(y)\mathrm{d}y^i\mathrm{d}y^j, \tag{4.143}$$

where the x^μ are coordinates in the four-dimensional spacetime and the y^i are coordinates on the extra-dimensional manifold, taken to be a maximally symmetric space with metric γ_{ij}. Of course the geometry of the extra dimensions is actually something dynamical that should be determined by solving the full equations of motion, but we are going to take (4.143) as a simplifying ansatz. (In a more complete treatment, we would expand the dynamical modes of the compactified geometry as a Fourier series, and show that the modes we are presently neglecting have larger masses than the overall-size mode we are choosing to examine.) The action is the $(4+d)$-dimensional Hilbert action plus a matter term:

$$S = \int d^{4+d}X\sqrt{-G}\left(\frac{1}{16\pi G_{4+d}}R[G_{ab}] + \widehat{\mathcal{L}}_\mathrm{M}\right), \tag{4.144}$$

where $\sqrt{-G}$ is minus the square root of the determinant of G_{ab}, $R[G_{ab}]$ is the Ricci scalar of G_{ab}, and $\widehat{\mathcal{L}}_\mathrm{M}$ is the matter Lagrange density with the metric determinant factored out.

The first step is to dimensionally reduce the action (4.144). By this we mean to actually perform the integral over the extra dimensions, which is possible because we have assumed that the extra-dimensional scale factor b is independent of y^i. Therefore we can express everything in terms of $g_{\mu\nu}$, γ_{IJ}, and $b(x)$, integrate over the extra dimensions, and arrive at an effective four-dimensional theory. From the metric (4.143) we have

$$\sqrt{-G} = b^d\sqrt{-g}\sqrt{\gamma}, \tag{4.145}$$

and we can evaluate the curvature scalar for this metric to obtain

$$\begin{aligned} R[G_{ab}] &= R[g_{\mu\nu}] + b^{-2}R[\gamma_{ij}] - 2db^{-1}g^{\mu\sigma}\nabla_\mu\nabla_\sigma b \\ &\quad - d(d-1)b^{-2}g^{\mu\sigma}(\nabla_\mu b)(\nabla_\sigma b), \end{aligned} \tag{4.146}$$

where ∇_μ is the covariant derivative associated with the four-dimensional metric $g_{\mu\nu}$. We denote by $\mathcal{V}$ the volume of the extra dimensions when $b = 1$; it is given by

$$\mathcal{V} = \int d^d y\sqrt{\gamma}. \tag{4.147}$$

The four-dimensional Newton's constant G_4 is determined by evaluating the coefficient of the curvature scalar in the action; we find that G_4 is related to its higher-dimensional analogue by

$$\frac{1}{16\pi G_4} = \frac{\mathcal{V}}{16\pi G_{4+d}}. \tag{4.148}$$

We are thus left with

$$S = \int d^4x\sqrt{-g}\left\{\frac{1}{16\pi G_4}\left[b^d R[g_{\mu\nu}] + d(d-1)b^{d-2}g^{\mu\nu}(\nabla_\mu b)(\nabla_\nu b)\right.\right.$$
$$\left.\left. + d(d-1)\kappa b^{d-2}\right] + \mathcal{V}b^d\widehat{\mathcal{L}}_{\mathrm{M}}\right\}, \tag{4.149}$$

where we have integrated by parts for convenience, and introduced the curvature parameter κ of γ_{ij}, given by

$$\kappa = \frac{R[\gamma_{ij}]}{d(d-1)}. \tag{4.150}$$

Comparing to (4.117)–(4.120), we see that the dimensionally-reduced action is precisely that of a scalar-tensor theory; the size of the extra dimensions plays the role of the scalar field. We can therefore make it look more conventional by performing a change of variables and a conformal transformation,

$$\begin{aligned} \beta(x) &= \ln b, \\ \tilde{g}_{\mu\nu} &= e^{d\beta}g_{\mu\nu}, \end{aligned} \tag{4.151}$$

which turns the reduced action into that of a scalar field coupled to gravity in the Einstein frame. Following the same procedure as outlined in our discussion of scalar-tensor theories yields

$$S = \int d^4x\sqrt{-\tilde{g}}\left\{\frac{1}{16\pi G_4}\left[R[\tilde{g}_{\mu\nu}] - \frac{1}{2}d(d+2)\tilde{g}^{\mu\nu}(\widetilde{\nabla}_\mu\beta)(\widetilde{\nabla}_\nu\beta)\right.\right.$$
$$\left.\left. + d(d-1)\kappa e^{(d+2)\beta}\right] + \mathcal{V}e^{-d\beta}\widehat{\mathcal{L}}_{\mathrm{M}}\right\}, \tag{4.152}$$

where we have dropped terms that are total derivatives.

To turn β into a canonically normalized scalar field, we make one final change of variables, to

$$\phi = \sqrt{\frac{d(d+2)}{2}}\,\bar{m}_{\mathrm{P}}\beta, \tag{4.153}$$

where the reduced Planck mass is $\bar{m}_{\mathrm{P}} = (8\pi G_4)^{-1/2}$. We are then left with

$$S = \int d^4x\sqrt{-\tilde{g}}\left\{\frac{1}{16\pi G_4}R[\tilde{g}_{\mu\nu}] - \frac{1}{2}\tilde{g}^{\mu\nu}(\widetilde{\nabla}_\mu\phi)(\widetilde{\nabla}_\nu\phi)\right.$$
$$\left. + \frac{1}{2}\kappa d(d-1)\bar{m}_{\mathrm{P}}^2 e^{-\sqrt{2(d+2)/d}\,\phi/\bar{m}_{\mathrm{P}}} + \mathcal{V}e^{-\sqrt{2d/(d+2)}\,\phi/\bar{m}_{\mathrm{P}}}\widehat{\mathcal{L}}_{\mathrm{M}}\right\}. \tag{4.154}$$

The scalar ϕ is known as the **dilaton** or **radion**, and characterizes the size of the extra-dimensional manifold.

The last two terms in (4.154) represent (minus) the potential $V(\phi)$. If we ignore the matter term $\widehat{\mathcal{L}}_{\mathrm{M}}$, the behavior of the dilaton will depend only on the sign of κ. If the extra-dimensional manifold is flat ($\kappa = 0$), the potential vanishes and we simply have a massless scalar field; this possibility runs afoul of the experimental constraints on scalar-tensor theories mentioned above. If there is curvature ($\kappa \neq 0$), the potential has no minimum; for $\kappa > 0$ the field will roll to $-\infty$, while for $\kappa < 0$ the field will roll to $+\infty$. But $\phi \propto \ln b$, so this means the scale factor $b(x)$ of the extra dimensions either shrinks to zero or becomes arbitrarily large, in either case ruining the hope for stable extra dimensions. Stability can be achieved, however, by choosing an appropriate matter Lagrangian, and an appropriate field configuration in the extra dimensions.

Let us now move on to a different kind of alternative theory, those that feature Lagrangians of more than second order in derivatives of the metric. We could imagine an action of the form

$$S = \int d^n x \sqrt{-g}(R + \alpha_1 R^2 + \alpha_2 R_{\mu\nu} R^{\mu\nu} + \alpha_3 g^{\mu\nu} \nabla_\mu R \nabla_\nu R + \cdots), \quad (4.155)$$

where the α's are coupling constants and the dots represent every other scalar we can make from the curvature tensor, its contractions, and its derivatives. Traditionally, such terms have been neglected on the reasonable grounds that they merely complicate a theory that is already both aesthetically pleasing and empirically successful. There is also, classically speaking, a more substantive objection. In conventional form, Einstein's equation leads to a well-posed initial value problem for the metric, in which coordinates and momenta specified at an initial time can be used to predict future evolution. With higher-derivative terms, we would require not only those data, but also some number of derivatives of the momenta; the character of the theory is dramatically altered.

However, there are also good reasons to consider such additional terms. As mentioned in our brief discussion of quantum gravity, one of the technical obstacles to consistent quantization of general relativity is that the theory is non-renormalizable: Inclusion of higher-order quantum effects leads to infinite answers. With the appropriate combination of higher-order Lagrangian terms, it turns out that you can actually render the theory renormalizable, which gives some hope of constructing a consistent quantum theory.[5] Unfortunately, it turns out that renormalizability comes at too high a price; these models generally feature negative-energy field excitations (ghosts). Consequently, the purported vacuum state (empty space) would be unstable to decay into positive- and negative-energy modes, which is inconsistent with both empirical experience and theoretical prejudice.

Nevertheless, the prevailing current view is that GR is an effective theory valid at energies below the Planck scale, and we should actually include all of the pos-

[5] See, for example, K.S. Stelle, *Phys. Rev.* **D16**, 953 (1977).

sible higher-order terms; but they will suppressed by appropriate powers of the Planck scale, just as we argued in our discussion of the Equivalence Principle in Section 4.7. They will therefore only become important when the length scale characteristic of the curvature approaches the Planck scale, which is far from any plausible experiment. Higher-order terms are therefore interesting in principle, but not in practice. On the other hand, similar reasoning would lead us to expect a huge vacuum energy term, since it is lower-order than the Hilbert action, which we know not to be true; so we should keep an open mind.

As a final alternative to general relativity, we should mention the possibility that the connection really is not derived from the metric, but in fact has an independent existence as a fundamental field. As one of the exercises you are asked to show that it is possible to consider the conventional action for general relativity but treat it as a function of both the metric $g_{\mu\nu}$ and a torsion-free connection $\Gamma^{\lambda}_{\rho\sigma}$, and the equations of motion derived from varying such an action with respect to the connection imply that $\Gamma^{\lambda}_{\rho\sigma}$ is actually the Christoffel connection associated with $g_{\mu\nu}$. We could drop the demand that the connection be torsion-free, in which case the torsion tensor could lead to additional propagating degrees of freedom. The basic reason why such theories do not receive much attention is simply because the torsion is itself a tensor; there is nothing to distinguish it from other, nongravitational tensor fields. Thus, we do not really lose any generality by considering theories of torsion-free connections (which lead to GR) plus any number of tensor fields, which we can name what we like. Similar considerations apply when we consider dropping the requirement of metric compatibility—any connection can be written as a metric-compatible connection plus a tensorial correction, so any such theory is equivalent to GR plus extra tensor fields, which wouldn't really deserve to be called an "alternative to general relativity".

4.9 ■ EXERCISES

1. The Lagrange density for electromagnetism in curved space is

$$\mathcal{L} = \sqrt{-g}\left(-\tfrac{1}{4}F^{\mu\nu}F_{\mu\nu} + A_\mu J^\mu\right), \tag{4.156}$$

where J^μ is the conserved current.

(a) Derive the energy-momentum tensor by functional differentiation with respect to the metric.

(b) Consider adding a new term to the Lagrangian,

$$\mathcal{L}' = \beta R^{\mu\nu} g^{\rho\sigma} F_{\mu\rho} F_{\nu\sigma}.$$

How are Maxwell's equations altered in the presence of this term? Einstein's equation? Is the current still conserved?

2. We showed how to derive Einstein's equation by varying the Hilbert action with respect to the metric. They can also be derived by treating the metric and connection as independent degrees of freedom and varying separately with respect to them; this is known as the **Palatini formalism**. That is, we consider the action

$$S = \int d^4x \sqrt{-g}\, g^{\mu\nu} R_{\mu\nu}(\Gamma),$$

where the Ricci tensor is thought of as constructed purely from the connection, not using the metric. Variation with respect to the metric gives the usual Einstein's equations, but for a Ricci tensor constructed from a connection that has no a priori relationship to the metric. Imagining from the start that the connection is symmetric (torsion free), show that variation of this action with respect to the connection coefficients leads to the requirement that the connection be metric compatible, that is, the Christoffel connection. Remember that Stokes's theorem, relating the integral of the covariant divergence of a vector to an integral of the vector over the boundary, does not work for a general covariant derivative. The best strategy is to write the connection coefficients as a sum of the Christoffel symbols $\widetilde{\Gamma}^\lambda_{\mu\nu}$ and a tensor $C^\lambda{}_{\mu\nu}$,

$$\Gamma^\lambda_{\mu\nu} = \widetilde{\Gamma}^\lambda_{\mu\nu} + C^\lambda{}_{\mu\nu},$$

and then show that $C^\lambda{}_{\mu\nu}$ must vanish.

3. The four-dimensional δ-function on a manifold M is defined by

$$\int_M F(x^\mu) \left[\frac{\delta^{(4)}(x^\sigma - y^\sigma)}{\sqrt{-g}}\right] \sqrt{-g}\, d^4x = F(y^\sigma), \tag{4.157}$$

for an arbitrary function $F(x^\mu)$. Meanwhile, the energy-momentum tensor for a pressureless perfect fluid (dust) is

$$T^{\mu\nu} = \rho U^\mu U^\nu, \tag{4.158}$$

where ρ is the energy density and U^μ is the four-velocity. Consider such a fluid that consists of a single particle traveling on a world line $x^\mu(\tau)$, with τ the proper time. The energy-momentum tensor for this fluid is then given by

$$T^{\mu\nu}(y^\sigma) = m \int_M \left[\frac{\delta^{(4)}(y^\sigma - x^\sigma(\tau))}{\sqrt{-g}}\right] \frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} d\tau, \tag{4.159}$$

where m is the rest mass of the particle. Show that covariant conservation of the energy-momentum tensor, $\nabla_\mu T^{\mu\nu} = 0$, implies that $x^\mu(\tau)$ satisfies the geodesic equation.

4. Show that the energy-momentum tensors for electromagnetism and for scalar field theory satisfy the dominant energy condition, and thus also the weak, null, and null dominant conditions. Show that they also satisfy $w \geq -1$.

5. A spacetime is static if there is a timelike Killing vector that is orthogonal to spacelike hypersurfaces. (See the Appendices for more discussion, including a definition of Raychaudhuri's equation.)

   (a) Generally speaking, if a vector field v^μ is orthogonal to a set of hypersurfaces defined by f = constant, then we can write the vector as $v_\mu = h\nabla_\mu f$ (here both f and h are functions). Show that this implies

$$v_{[\sigma}\nabla_\mu v_{\nu]} = 0.$$

(b) Imagine we have a perfect fluid with zero pressure (dust), which generates a solution to Einstein's equations. Show that the metric can be static only if the fluid four-velocity is parallel to the timelike (and hypersurface-orthogonal) Killing vector.

(c) Use Raychaudhuri's equation to prove that there is no static solution to Einstein's equations if the pressure is zero and the energy density is greater than zero.

6. Let K be a Killing vector field. Show that an electromagnetic field with potential $A_\mu = K_\mu$ solves Maxwell's equations if the metric is a vacuum solution to Einstein's equations. This is a slight cheat, since you won't be in vacuum if there is a nonzero electromagnetic field strength, but we assume the field strength is small enough not to dramatically affect the geometry.

CHAPTER

5 The Schwarzschild Solution

5.1 ■ THE SCHWARZSCHILD METRIC

The most obvious application of a theory of gravity is to a spherically symmetric gravitational field. This would be the relevant situation to describe, for example, the field created by the Earth or the Sun (to a good approximation), in which apples fall or planets move. Furthermore, our first concern is with exterior solutions (empty space surrounding a gravitating body), since understanding the motion of test particles outside an object is both easier and more immediately useful than considering the relatively inaccessible interior. In addition to its practical usefulness, the answer to this problem in general relativity will lead us to remarkable solutions describing new phenomena of great interest to physicists and astronomers: black holes. In this chapter we examine the simple case of vacuum solutions with perfect spherical symmetry; in the next chapter we consider features of black holes in more general contexts.

In GR, the unique spherically symmetric vacuum solution is the **Schwarzschild metric**; it is second only to Minkowski space in the list of important spacetimes. In spherical coordinates $\{t, r, \theta, \phi\}$, the metric is given by

$$\boxed{ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\Omega^2,} \tag{5.1}$$

where $d\Omega^2$ is the metric on a unit two-sphere,

$$d\Omega^2 = d\theta^2 + \sin^2\theta \, d\phi^2. \tag{5.2}$$

The constant M is interpreted as the mass of the gravitating object (although some care is required in making this identification). In this section we will derive the Schwarzschild metric by trial and error; in the next section we will be more systematic in both the derivation of the solution and its consequences.

Since we are interested in the solution *outside* a spherical body, we care about Einstein's equation in vacuum,

$$R_{\mu\nu} = 0. \tag{5.3}$$

Our hypothesized source is static (unevolving) and spherically symmetric, so we will look for solutions that also have these properties. Rigorous definitions of both "static" and "spherically symmetric" require some care, due to subtleties of coordinate independence. For now we will interpret static to imply two conditions: that all metric components are independent of the time coordinate, and that there are no time-space cross terms ($dt dx^i + dx^i dt$) in the metric. The latter condition makes sense if we imagine performing a time inversion $t \to -t$; the dt^2 term remains invariant, as do any $dx^i dx^j$ terms, while cross terms would not. Since we hope to find a solution that is independent of time, it should be invariant under time reversal, and we therefore leave cross terms out. To impose spherical symmetry, we begin by writing the metric of Minkowski space (a spherically symmetric spacetime we know something about) in polar coordinates $x^\mu = (t, r, \theta, \phi)$:

$$ds^2_{\text{Minkowski}} = -dt^2 + dr^2 + r^2 d\Omega^2. \tag{5.4}$$

One requirement to preserve spherical symmetry is that we maintain the form of $d\Omega^2$; that is, if we want our spheres to be perfectly round, the coefficient of the $d\phi^2$ term should be $\sin^2\theta$ times that of the $d\theta^2$ term. But we are otherwise free to multiply all of the terms by separate coefficients, so long as they are only functions of the radial coordinate r:

$$ds^2 = -e^{2\alpha(r)} dt^2 + e^{2\beta(r)} dr^2 + e^{2\gamma(r)} r^2 d\Omega^2. \tag{5.5}$$

We've expressed our functions as exponentials so that the signature of the metric doesn't change. In a full treatment, we would allow for complete freedom and see what happens.

We can use our ability to change coordinates to make a slight simplification to the static, spherically-symmetric metric (5.5), even before imposing Einstein's equation. Unlike other theories of physics, in general relativity we simultaneously define coordinates and the metric as a function of those coordinates. In other words, we don't know ahead of time what, for example, the radial coordinate r really is; we can only interpret it once the solution is in our hands. Let us therefore imagine defining a new coordinate $\bar{r}$ via

$$\bar{r} = e^{\gamma(r)} r, \tag{5.6}$$

with an associated basis one-form

$$d\bar{r} = e^\gamma dr + e^\gamma r\, d\gamma = \left(1 + r\frac{d\gamma}{dr}\right) e^\gamma dr. \tag{5.7}$$

In terms of this new variable, the metric (5.5) becomes

$$ds^2 = -e^{2\alpha(r)} dt^2 + \left(1 + r\frac{d\gamma}{dr}\right)^{-2} e^{2\beta(r) - 2\gamma(r)} d\bar{r}^2 + \bar{r}^2 d\Omega^2, \tag{5.8}$$

where each function of r is a function of $\bar{r}$ in the obvious way. But now let us make the following relabelings:

$$\bar{r} \to r \tag{5.9}$$

$$\left(1 + r\frac{d\gamma}{dr}\right)^{-2} e^{2\beta(r)-2\gamma(r)} \to e^{2\beta}. \tag{5.10}$$

There is nothing to stop us from doing this, as they are simply labels, with no independent external definition. If you wish you can continue to use $\bar{r}$, and set (5.10) equal to $e^{2\bar{\beta}}$, but we won't bother. Our metric (5.8) becomes

$$ds^2 = -e^{2\alpha(r)}\,dt^2 + e^{2\beta(r)}\,dr^2 + r^2\,d\Omega^2. \tag{5.11}$$

This looks exactly like (5.5), except that the $e^{2\gamma}$ factor has disappeared. We have not set $e^{2\gamma}$ equal to one, which would be a statement about the geometry; we have simply chosen our radial coordinate such that this factor doesn't exist. Thus, (5.11) is precisely as general as (5.5).

Let's now take this metric and use Einstein's equation to solve for the functions $\alpha(r)$ and $\beta(r)$. We begin by evaluating the Christoffel symbols. If we use labels (t, r, θ, ϕ) for $(0, 1, 2, 3)$ in the usual way, the Christoffel symbols are given by

$$\begin{array}{lll}
\Gamma^t_{tr} = \partial_r\alpha & \Gamma^r_{tt} = e^{2(\alpha-\beta)}\partial_r\alpha & \Gamma^r_{rr} = \partial_r\beta \\
\Gamma^\theta_{r\theta} = \dfrac{1}{r} & \Gamma^r_{\theta\theta} = -re^{-2\beta} & \Gamma^\phi_{r\phi} = \dfrac{1}{r} \\
\Gamma^r_{\phi\phi} = -re^{-2\beta}\sin^2\theta & \Gamma^\theta_{\phi\phi} = -\sin\theta\cos\theta & \Gamma^\phi_{\theta\phi} = \dfrac{\cos\theta}{\sin\theta}.
\end{array} \tag{5.12}$$

Anything not written down explicitly is meant to be zero, or related to what is written by symmetries. From these we get the following nonvanishing components of the Riemann tensor:

$$\begin{aligned}
R^t{}_{rtr} &= \partial_r\alpha\partial_r\beta - \partial_r^2\alpha - (\partial_r\alpha)^2 \\
R^t{}_{\theta t\theta} &= -re^{-2\beta}\partial_r\alpha \\
R^t{}_{\phi t\phi} &= -re^{-2\beta}\sin^2\theta\ \partial_r\alpha \\
R^r{}_{\theta r\theta} &= re^{-2\beta}\partial_r\beta \\
R^r{}_{\phi r\phi} &= re^{-2\beta}\sin^2\theta\ \partial_r\beta \\
R^\theta{}_{\phi\theta\phi} &= (1 - e^{-2\beta})\sin^2\theta.
\end{aligned} \tag{5.13}$$

Taking the contraction as usual yields the Ricci tensor:

$$\begin{aligned}
R_{tt} &= e^{2(\alpha-\beta)}\left[\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\alpha\partial_r\beta + \tfrac{2}{r}\partial_r\alpha\right] \\
R_{rr} &= -\partial_r^2\alpha - (\partial_r\alpha)^2 + \partial_r\alpha\partial_r\beta + \tfrac{2}{r}\partial_r\beta \\
R_{\theta\theta} &= e^{-2\beta}[r(\partial_r\beta - \partial_r\alpha) - 1] + 1 \\
R_{\phi\phi} &= \sin^2\theta\, R_{\theta\theta},
\end{aligned} \tag{5.14}$$

and for future reference we calculate the curvature scalar,

$$R = -2e^{-2\beta}\left[\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\alpha\partial_r\beta + \frac{2}{r}(\partial_r\alpha - \partial_r\beta) + \frac{1}{r^2}(1 - e^{2\beta})\right]. \tag{5.15}$$

With the Ricci tensor calculated, we would like to set it equal to zero. Since R_{tt} and R_{rr} vanish independently, we can write

$$0 = e^{2(\beta-\alpha)}R_{tt} + R_{rr} = \frac{2}{r}(\partial_r\alpha + \partial_r\beta), \tag{5.16}$$

which implies $\alpha = -\beta + c$, where c is some constant. We can set this constant equal to zero by rescaling our time coordinate by $t \to e^{-c}t$, after which we have

$$\alpha = -\beta. \tag{5.17}$$

Next let us turn to $R_{\theta\theta} = 0$, which now reads

$$e^{2\alpha}(2r\partial_r\alpha + 1) = 1. \tag{5.18}$$

This is equivalent to

$$\partial_r(re^{2\alpha}) = 1. \tag{5.19}$$

We can solve this to obtain

$$e^{2\alpha} = 1 - \frac{R_S}{r}, \tag{5.20}$$

where R_S is some undetermined constant. With (5.17) and (5.20), our metric becomes

$$ds^2 = -\left(1 - \frac{R_S}{r}\right)dt^2 + \left(1 - \frac{R_S}{r}\right)^{-1}dr^2 + r^2 d\Omega^2. \tag{5.21}$$

We now have no freedom left except for the single constant R_S, so this form had better solve the remaining equations $R_{tt} = 0$ and $R_{rr} = 0$; it is straightforward to check that it does, for any value of R_S.

The only thing left to do is to interpret the constant R_S, called the **Schwarzschild radius**, in terms of some physical parameter. Nothing could be simpler. In Chapter 4 we found that, in the weak-field limit, the tt component of the metric around a point mass satisfies

$$g_{tt} = -\left(1 - \frac{2GM}{r}\right). \tag{5.22}$$

The Schwarzschild metric should reduce to the weak-field case when $r \gg 2GM$, but for the tt component the forms are already exactly the same; we need only identify

$$R_S = 2GM. \tag{5.23}$$

This can be thought of as the definition of the parameter M.

Our final result is the Schwarzschild metric, (5.1). We have shown that it is a static, spherically symmetric vacuum solution to Einstein's equation; M functions as a parameter, which we happen to know can be interpreted as the conventional Newtonian mass that we would measure by studying orbits at large distances from the gravitating source. It won't simply be the sum of the masses of the constituents of whatever body is curving spacetime, since there will be a contribution from what we might think of as the gravitational binding energy; however, in the weak field limit, the quantities will agree. Note that as $M \to 0$ we recover Minkowski space, which is to be expected. Note also that the metric becomes progressively Minkowskian as $r \to \infty$; this property is known as **asymptotic flatness**. A more technical definition involves matching regions at infinity in a conformal diagram, as discussed in the next chapter.

5.2 ■ BIRKHOFF'S THEOREM

Birkhoff's theorem is the statement that the Schwarzschild metric is the *unique* vacuum solution with spherical symmetry (and in particular, that there are no time-dependent solutions of this form); proving it is an instructive exercise, which consists of three major steps. First, we argue that a spherically symmetric spacetime can be foliated by two-spheres—in other words, that (almost) every point lies on a unique sphere that is left invariant by the generators of spherical symmetry. Second, we show on purely geometric grounds that the metric on such a space can always (at least in a local region) be put in the form

$$ds^2 = d\tau^2(a,b) + r^2(a,b)\, d\Omega^2(\theta,\phi), \tag{5.24}$$

where (a, b) are coordinates transverse to the spheres, and r is a function of these coordinates. Third, we plug this metric into Einstein's equation in vacuum to show that Schwarzschild is the unique solution. We will argue in favor of the first two points at a level of rigor that is likely to be convincing to most physicists, although mathematicians will be uneasy; the third point is straightforward calculation. For a more careful treatment see Hawking and Ellis (1973). We will use a few concepts from Appendix C, which may be useful to read at this point. Of course, if you are more interested in exploring properties of the Schwarzschild solution than in proving its uniqueness, you are welcome to skip right to the next section.

We begin with the concept of a four-dimensional spherically symmetric spacetime M. Spherically symmetric means having the same symmetries as a sphere. (In this chapter the word sphere refers specifically to S^2, not spheres of other dimension.) The symmetries of a sphere are precisely those of ordinary rotations in three-dimensional Euclidean space; in the language of group theory, they comprise the special orthogonal group SO(3). (Recall the discussion of the Lorentz and rotation groups in Chapter 1.) In the case of a metric on a manifold, symmetries are characterized by the existence of Killing vectors. In Section 3.8 we found the three Killing vectors of S^2, labeled (R, S, T); in (θ, ϕ) coordinates they take the form

$$\begin{aligned} R &= \partial_\phi \\ S &= \cos\phi\,\partial_\theta - \cot\theta\sin\phi\,\partial_\phi \\ T &= -\sin\phi\,\partial_\theta - \cot\theta\cos\phi\,\partial_\phi. \end{aligned} \tag{5.25}$$

A spherically symmetric manifold is one that has three Killing vector fields that are the same as those on S^2. But how do we know, in a coordinate-independent way, that a set of Killing vectors on one manifold is the same as that on some other manifold? The structure of a set of symmetry transformations is given by the commutation relations of the transformations, which express the difference between performing two infinitesimal transformations in one order versus the reversed order. In group theory these are expressed by the Lie algebra of the symmetry generators, while in differential geometry they are expressed by the commutators of the Killing vector fields. There is a deep connection here, which we don't have time to pursue; see Schutz (1980). In the Exercises for Chapter 3 you verified that the commutators of the rotational Killing vectors (R, S, T) satisfied

$$\begin{aligned} [R, S] &= T \\ [S, T] &= R \\ [T, R] &= S. \end{aligned} \tag{5.26}$$

This algebra of Killing vectors fully characterizes the kind of symmetry we have. A manifold will be said to possess **spherical symmetry** if and only if there are three Killing fields satisfying (5.26).

In Appendix C we discuss Frobenius's theorem, which states that if you have a set of vector fields whose commutator closes—the commutator of any two fields in the set is a linear combination of other fields in the set—then the integral curves of these vector fields fit together to describe submanifolds of the manifold on which they are all defined. The dimensionality of the submanifold may be smaller than the number of vectors, or it could be equal, but obviously not larger. Vector fields that obey (5.26) will of course form 2-spheres. Since the vector fields stretch throughout the space, every point will be on exactly one of these spheres. (Actually, it's almost every point—we will show below how it can fail to be absolutely every point.) Thus, we say that a spherically symmetric manifold can be foliated into spheres.

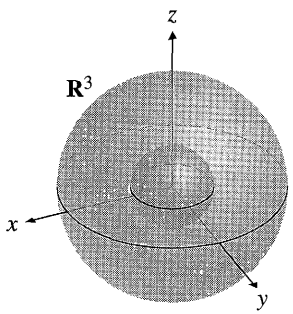

FIGURE 5.1 Foliating $\mathbf{R}^3$ (minus the origin) by two-spheres.

Let's consider some examples to bring this down to earth. The simplest example is flat three-dimensional Euclidean space. If we pick an origin, then $\mathbf{R}^3$ is clearly spherically symmetric with respect to rotations around this origin. Under such rotations (that is, under the flow of the Killing vector fields), points move into each other, but each point stays on an S^2 at a fixed distance from the origin.

These spheres foliate $\mathbf{R}^3$, as depicted in Figure 5.1. Of course, they don't really foliate all of the space, since the origin itself just stays put under rotations—it

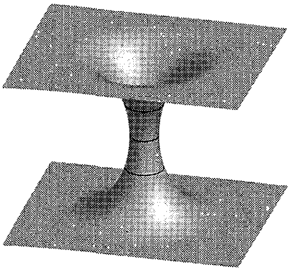

FIGURE 5.2 Foliation of a wormhole by two-spheres.

doesn't move around on some two-sphere. But it should be clear that almost all of the space is properly foliated, and this will turn out to be enough for us.

We can also have spherical symmetry without an origin to rotate things around. An example is provided by a wormhole, with topology $\mathbf{R} \times S^2$. If we suppress a dimension and draw our two-spheres as circles, such a space might look like Figure 5.2. In this case the entire manifold can be foliated by two-spheres.

Given that manifolds with SO(3) symmetry may be foliated by spheres, our second step is to show that the metric on M can be put into the form (5.24). The set of all the spheres forms a two-dimensional space (since a four-dimensional spacetime is being foliated with two-dimensional spheres). You might hope we could simply put coordinates (θ, ϕ) on each sphere, and coordinates (a, b) on the set of all spheres, for a complete set of coordinates (a, b, θ, ϕ) on M. Then each sphere is specified by $a =$ constant, $b =$ constant. We know that the metric on a round sphere is $d\Omega^2$, so this strategy would be sufficient to guarantee that the metric restricted to any fixed values $a = a_0$ and $b = b_0$ (so that $\mathrm{d}a = \mathrm{d}b = 0$) takes the form

$$ds^2(a_0, b_0, \theta, \phi) = f(a_0, b_0)\, d\Omega^2. \tag{5.27}$$

In particular, the function f must be independent of θ and ϕ, or the sphere would be lumpy rather than round. Furthermore, it's equally clear that the metric restricted to any fixed values $\theta = \theta_0$ and $\phi = \phi_0$ (so that $\mathrm{d}\theta = \mathrm{d}\phi = 0$) takes the form

$$ds^2(a, b, \theta_0, \phi_0) = d\tau^2(a, b). \tag{5.28}$$

Again, any dependence on θ or ϕ would destroy the symmetry; it would mean that the geometry transverse to the spheres depended on where you were on the sphere.

However, we have been too reckless by slapping down these coordinates, since we cannot rule out cross terms of the form $\mathrm{d}a\mathrm{d}\theta + \mathrm{d}\theta\mathrm{d}a$ and so on. In other words, we must be careful to line up our spheres appropriately, so that travel along a curve that is perpendicular to one of the spheres keeps us at constant θ and ϕ. To guarantee this we need to be more careful in setting up our coordinates. Begin by considering a single point q lying on a sphere S_q (note that q must not be a degenerate point at which all of the Killing vectors vanish). Put coordinates (θ, ϕ) on this particular sphere only, not yet through the manifold. At each point p on S_q, there will be a two-dimensional orthogonal subspace O_p, consisting of points along geodesics emanating from p whose tangent vectors at p are orthogonal to S_q. Note that there will be a one-dimensional subgroup R_p of rotations that leave p fixed; indeed, these rotations keep fixed any direction perpendicular to S_q at p, and hence the entire two-surface O_p is left invariant by R_p.

Consider a point r that is not on S_q, but on some other sphere S_r in the foliation, and that lies in the two-surface O_p orthogonal to S_q at p. Since p is arbitrary, this includes any possible point r in a neighborhood of S_q. Note that O_p will be orthogonal to S_r as well as to S_q. To see this, consider the two-dimensional plane

V_r of vectors in the tangent space $T_r M$ that are orthogonal to the two-surface O_p. Since O_p is left invariant by the rotations R_p, these rotations must take V_r into itself, because they are an isometry, and hence preserve orthogonality. But R_p also takes the set of vectors tangent to S_q into itself, since these rotations leave the spheres invariant. In four dimensions, two planes that are both orthogonal to a given plane at the same point must be the same plane; hence, the vectors tangent to S_r must be orthogonal to O_p.

There will be a unique geodesic that is orthogonal to S_q and connects p to r. Traveling down such geodesics provides a map $f : S_q \to S_r$, which is both one-to-one and onto (at least in a neighborhood of the original sphere). We use this map to define coordinates on S_r (and, similarly, on any other sphere) by assigning the same values of (θ, ϕ) to $r \in S_r$ that were the coordinates at $p \in S_q$. We have therefore defined (θ, ϕ) throughout the manifold. Now to define coordinates (a, b), choose two basis vectors S, T for the subspace of $T_q M$ that generates the orthogonal space O_q. Any other sphere will be connected to q by a unique orthogonal geodesic, with tangent vector $aS + bT \in T_q M$. Assign those components (a, b) as coordinates everywhere on that sphere. This defines the full set of coordinates (a, b, θ, ϕ) throughout the manifold.

The metric in these coordinates satisfies (5.27) and (5.28); it remains to be shown that there are no cross terms between directions along the spheres and those transverse. This means, for example, that the vector field ∂_a should be orthogonal to ∂_θ, and so on; it is straightforward to verify that this is so. First, consider ∂_θ at some point $r \in S_r$; this vector is the directional derivative along a curve of the form $x^\mu(\theta) = (a_r, b_r, \theta, \phi_r)$. Since a and b are constant along the curve, the entire curve remains in the sphere S_r, so that ∂_θ is tangent to the sphere. Meanwhile, ∂_a is a derivative along $x^\mu(a) = (a, b_r, \theta_r, \phi_r)$. Since this curve remains in the orthogonal subspace O_r, ∂_a will be orthogonal to S_r, and hence to ∂_θ. Similar arguments guarantee that there will be no cross terms between (a, b) and (θ, ϕ).

We have thus succeeded in putting the metric on a spherically symmetric spacetime in the form

$$ds^2 = g_{aa}(a, b)\, da^2 + g_{ab}(a, b)(da\,db + db\,da) + g_{bb}(a, b)\, db^2 + r^2(a, b)\, d\Omega^2. \tag{5.29}$$

Here $r(a, b)$ is some as-yet-undetermined function, to which we have merely given a suggestive label. There is nothing to stop us, however, from changing coordinates from (a, b) to (a, r) by inverting $r(a, b)$, unless r were a function of a alone; in this case we could just as easily switch to (b, r), so we will not consider this situation separately. The metric is then

$$ds^2 = g_{aa}(a, r)\, da^2 + g_{ar}(a, r)(da\, dr + dr\, da) + g_{rr}(a, r)\, dr^2 + r^2\, d\Omega^2. \tag{5.30}$$

Our next step is to find a function $t(a, r)$ such that, in the (t, r) coordinate system, there are no cross terms $dt\,dr + dr\,dt$ in the metric. Notice that

$$dt = \frac{\partial t}{\partial a}\, da + \frac{\partial t}{\partial r}\, dr, \tag{5.31}$$

so

$$dt^2 = \left(\frac{\partial t}{\partial a}\right)^2 da^2 + \left(\frac{\partial t}{\partial a}\right)\left(\frac{\partial t}{\partial r}\right)(da\, dr + dr\, da) + \left(\frac{\partial t}{\partial r}\right)^2 dr^2. \tag{5.32}$$

We would like to replace the first three terms in the metric (5.30) by

$$m dt^2 + n dr^2, \tag{5.33}$$

for some functions m and n. This is equivalent to the requirements

$$m\left(\frac{\partial t}{\partial a}\right)^2 = g_{aa}, \tag{5.34}$$

$$n + m\left(\frac{\partial t}{\partial r}\right)^2 = g_{rr}, \tag{5.35}$$

and

$$m\left(\frac{\partial t}{\partial a}\right)\left(\frac{\partial t}{\partial r}\right) = g_{ar}. \tag{5.36}$$

We therefore have three equations for the three unknowns $t(a, r)$, $m(a, r)$, and $n(a, r)$, just enough to determine them precisely, up to initial conditions for t. (Of course, they are "determined" in terms of the unknown functions g_{aa}, g_{ar}, and g_{rr}, so in this sense they are still undetermined.) We can therefore put our metric in the form

$$ds^2 = m(t, r)\, dt^2 + n(t, r)\, dr^2 + r^2\, d\Omega^2. \tag{5.37}$$

To this point the only difference between the two coordinates t and r is that we have chosen r to be the one that multiplies the metric for the two-sphere. This choice was motivated by what we know about the metric for flat Minkowski space, which can be written $ds^2 = -dt^2 + dr^2 + r^2\, d\Omega^2$. We know that the spacetime under consideration is Lorentzian, so either m or n will have to be negative. Let us choose m, the coefficient of dt^2, to be negative. This is not a choice we are simply allowed to make, and in fact we will see later that it can go wrong; but we will assume it for now. The assumption is not completely unreasonable, since we know that Minkowski space is itself spherically symmetric, and will therefore be described by (5.37). With this choice we can trade in the functions m and n for new functions α and β, such that

$$ds^2 = -e^{2\alpha(t,r)}\, dt^2 + e^{2\beta(t,r)}\, dr^2 + r^2\, d\Omega^2. \tag{5.38}$$

This is the best we can do using only geometry; spherical symmetry is certainly not enough to say anything substantive about the functions $\alpha(t, r)$ and $\beta(t, r)$. Our next step is therefore to actually solve Einstein's equation; the steps follow closely along those of Section 5.1, in which we considered a metric similar to (5.38) but with the additional assumption of time-independence. Here we will see that this assumption was unnecessary, as the solution will necessarily be static.

The nonvanishing Christoffel symbols for (5.38) are

$$\begin{aligned}
\Gamma^t_{tt} &= \partial_t\alpha & \Gamma^t_{tr} &= \partial_r\alpha & \Gamma^t_{rr} &= e^{2(\beta-\alpha)}\partial_t\beta \\
\Gamma^r_{tt} &= e^{2(\alpha-\beta)}\partial_r\alpha & \Gamma^r_{tr} &= \partial_t\beta & \Gamma^r_{rr} &= \partial_r\beta \\
\Gamma^\theta_{r\theta} &= \frac{1}{r} & \Gamma^r_{\theta\theta} &= -re^{-2\beta} & \Gamma^\phi_{r\phi} &= \frac{1}{r} \\
\Gamma^r_{\phi\phi} &= -re^{-2\beta}\sin^2\theta & \Gamma^\theta_{\phi\phi} &= -\sin\theta\cos\theta & \Gamma^\phi_{\theta\phi} &= \frac{\cos\theta}{\sin\theta},
\end{aligned} \tag{5.39}$$

the following nonvanishing components of the Riemann tensor are

$$\begin{aligned}
R^t{}_{rtr} &= e^{2(\beta-\alpha)}[\partial_t^2\beta + (\partial_t\beta)^2 - \partial_t\alpha\partial_t\beta] + [\partial_r\alpha\partial_r\beta - \partial_r^2\alpha - (\partial_r\alpha)^2] \\
R^t{}_{\theta t\theta} &= -re^{-2\beta}\partial_r\alpha \\
R^t{}_{\phi t\phi} &= -re^{-2\beta}\sin^2\theta\,\partial_r\alpha \\
R^t{}_{\theta r\theta} &= -re^{-2\alpha}\partial_t\beta \\
R^t{}_{\phi r\phi} &= -re^{-2\alpha}\sin^2\theta\,\partial_t\beta \\
R^r{}_{\theta r\theta} &= re^{-2\beta}\partial_r\beta \\
R^r{}_{\phi r\phi} &= re^{-2\beta}\sin^2\theta\,\partial_r\beta \\
R^\theta{}_{\phi\theta\phi} &= (1-e^{-2\beta})\sin^2\theta,
\end{aligned} \tag{5.40}$$

and the Ricci tensor is

$$\begin{aligned}
R_{tt} &= \left[\partial_t^2\beta + (\partial_t\beta)^2 - \partial_t\alpha\partial_t\beta\right] + e^{2(\alpha-\beta)}\left[\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\alpha\partial_r\beta + \frac{2}{r}\partial_r\alpha\right] \\
R_{rr} &= -\left[\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\alpha\partial_r\beta - \frac{2}{r}\partial_r\beta\right] \\
&\quad + e^{2(\beta-\alpha)}\left[\partial_t^2\beta + (\partial_t\beta)^2 - \partial_t\alpha\partial_t\beta\right] \\
R_{tr} &= \frac{2}{r}\partial_t\beta \\
R_{\theta\theta} &= e^{-2\beta}[r(\partial_r\beta - \partial_r\alpha) - 1] + 1 \\
R_{\phi\phi} &= R_{\theta\theta}\sin^2\theta.
\end{aligned} \tag{5.41}$$

Our job is to solve Einstein's equation in vacuum, $R_{\mu\nu} = 0$. From $R_{tr} = 0$ we get

$$\partial_t\beta = 0. \tag{5.42}$$

If we consider taking the time derivative of $R_{\theta\theta} = 0$ and using $\partial_t \beta = 0$, we get

$$\partial_t \partial_r \alpha = 0. \tag{5.43}$$

We can therefore write

$$\begin{aligned} \beta &= \beta(r) \\ \alpha &= f(r) + g(t). \end{aligned} \tag{5.44}$$

The first term in the metric (5.38) is thus $-e^{2f(r)}e^{2g(t)}\,\mathrm{d}t^2$. But we can always simply redefine our time coordinate by replacing $\mathrm{d}t \to e^{-g(t)}\,\mathrm{d}t$; in other words, we are free to choose t such that $g(t) = 0$, whence $\alpha(t,r) = f(r)$. We therefore have

$$ds^2 = -e^{2\alpha(r)}\,\mathrm{d}t^2 + e^{2\beta(r)}\,\mathrm{d}r^2 + r^2\,d\Omega^2. \tag{5.45}$$

All of the metric components are independent of the coordinate t. We have therefore proven a crucial result: *any spherically symmetric vacuum metric possesses a timelike Killing vector.*

This property is so interesting that it gets its own name: a metric that possesses a Killing vector that is timelike near infinity is called **stationary**. (Often, including in Schwarzschild, the Killing vector that is timelike at infinity will become spacelike somewhere in the interior.) In a stationary metric we can choose coordinates (t, x^1, x^2, x^3) in which the Killing vector is ∂_t and the metric components are independent of t; the general form of a stationary metric in these coordinates is thus

$$ds^2 = g_{00}(\vec{x})\,\mathrm{d}t^2 + g_{0i}(\vec{x})(\mathrm{d}t\mathrm{d}x^i + \mathrm{d}x^i\,\mathrm{d}t) + g_{ij}(\vec{x})\,\mathrm{d}x^i\mathrm{d}x^j. \tag{5.46}$$

There is also a more restrictive property: a metric is called **static** if it possesses a timelike Killing vector that is orthogonal to a family of hypersurfaces. (For more details on hypersurfaces, see Appendix D.) In the Exercises for Chapter 4 you showed that a hypersurface-orthogonal vector field v^μ obeys

$$v_{[\mu}\nabla_\nu v_{\sigma]} = 0. \tag{5.47}$$

But there is a simpler diagnostic; if we have adapted coordinates so that the components $g_{\mu\nu}$ are all independent of t, the surfaces to which the Killing vector will be orthogonal are defined by the condition $t =$ constant. Operationally, this means that the time-space cross terms in (5.46) will be absent; the general static metric can be written

$$ds^2 = g_{00}(\vec{x})\,\mathrm{d}t^2 + g_{ij}(\vec{x})\,\mathrm{d}x^i\mathrm{d}x^j. \tag{5.48}$$

We notice that only even powers of the time coordinate t appear in this form; thus, an alternative definition of "static" is "stationary, and invariant under time reversal ($t \to -t$)." The metric (5.45) is clearly static. You should think of stationary as meaning "doing exactly the same thing at every time," while static means "not doing anything at all." For example, the static spherically symmetric metric (5.45) will describe nonrotating stars or black holes, while rotating systems that keep rotating in the same way at all times will be described by metrics that are stationary but not static.

Notice that (5.45) is precisely the same as (5.11), the metric we originally used to derive the Schwarzschild solution in Section 5.1. We have therefore proven Birkhoff's theorem, that the unique spherically symmetric vacuum solution is the Schwarzschild metric,

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\Omega^2, \tag{5.49}$$

as promised.

We did not say anything about the source of the Schwarzschild metric, except that it be spherically symmetric. Specifically, we did not demand that the source itself be static; it could be a collapsing star, as long as the collapse is symmetric. Therefore a process such as a supernova explosion would generate very little gravitational radiation (in comparison to the amount of energy released through other channels) if it were close to spherically symmetric, which a realistic supernova may or may not be depending on its origin. This is the same result we would have obtained in electromagnetism, where the electromagnetic fields around a spherical charge distribution do not depend on the radial distribution of the charges.

5.3 ■ SINGULARITIES

Before exploring the behavior of test particles in the Schwarzschild geometry, we should say something about singularities. From the form of (5.1), the metric coefficients become infinite at $r = 0$ and $r = 2GM$—an apparent sign that something is going wrong. The metric coefficients, of course, are coordinate-dependent quantities, and as such we should not make too much of their values; it is certainly possible to have a coordinate singularity that results from a breakdown of a specific coordinate system rather than the underlying manifold. An example occurs at the origin of polar coordinates in the plane, where the metric $ds^2 = dr^2 + r^2 d\theta^2$ becomes degenerate and the component $g^{\theta\theta} = r^{-2}$ of the inverse metric blows up, even though that point of the manifold is no different from any other.

What kind of coordinate-independent signal should we look for as a warning that something about the geometry is out of control? This turns out to be a difficult question to answer, and entire books have been written about the nature of singularities in general relativity. We won't go into this issue in detail, but rather turn to one simple criterion for when something has gone wrong—when the curvature becomes infinite. The curvature is measured by the Riemann tensor, and it is hard to say when a tensor becomes infinite, since its components are coordinate-dependent. But from the curvature we can construct various scalar quantities, and since scalars are coordinate-independent it is meaningful to say that they become infinite. The simplest such scalar is the Ricci scalar

$R = g^{\mu\nu}R_{\mu\nu}$, but we can also construct higher-order scalars such as $R^{\mu\nu}R_{\mu\nu}$, $R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma}$, $R_{\mu\nu\rho\sigma}R^{\rho\sigma\lambda\tau}R_{\lambda\tau}{}^{\mu\nu}$, and so on. If any of these scalars (but not necessarily all of them) goes to infinity as we approach some point, we regard that point as a singularity of the curvature. We should also check that the point is not infinitely far away; that is, that it can be reached by traveling a finite distance along a curve.

We therefore have a sufficient condition for a point to be considered a singularity. It is not a necessary condition, however, and it is generally harder to show that a given point is nonsingular; for our purposes we will simply test to see if geodesics are well-behaved at the point in question, and if so then we will consider the point nonsingular. In the case of the Schwarzschild metric (5.1), direct calculation reveals that

$$R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma} = \frac{48G^2M^2}{r^6}. \tag{5.50}$$

This is enough to convince us that $r = 0$ represents an honest singularity.

The other trouble spot is $r = 2GM$, the Schwarzschild radius. You could check that none of the curvature invariants blows up there. We therefore begin to think that it is actually not singular, and we have simply chosen a bad coordinate system. The best thing to do is to transform to more appropriate coordinates if possible. We will soon see that in this case it is in fact possible, and the surface $r = 2GM$ is very well-behaved (although interesting) in the Schwarzschild metric—it demarcates the event horizon of a black hole.

Having worried a little about singularities, we should point out that the behavior of the Schwarzschild metric inside the Schwarzschild radius is of little day-to-day consequence. The solution we derived is valid only in vacuum, and we expect it to hold outside a spherical body such as a star. However, in the case of the Sun we are dealing with a body that extends to a radius of

$$R_\odot = 10^6 GM_\odot. \tag{5.51}$$

Thus, $r = 2GM_\odot$ is far inside the solar interior, where we do not expect the Schwarzschild metric to apply. In fact, realistic stellar interior solutions consist of matching the exterior Schwarzschild metric to an interior metric that is perfectly smooth at the origin. Nevertheless, there are objects for which the full Schwarzschild metric is required—black holes—and therefore we will let our imaginations roam far outside the solar system in this chapter.

5.4 ■ GEODESICS OF SCHWARZSCHILD

The first step we will take to understand the Schwarzschild metric more fully is to consider the behavior of geodesics. We need the nonzero Christoffel symbols for Schwarzschild:

$$
\begin{array}{lll}
\Gamma^r_{tt} = \dfrac{GM}{r^3}(r-2GM) & \Gamma^r_{rr} = \dfrac{-GM}{r(r-2GM)} & \Gamma^t_{tr} = \dfrac{GM}{r(r-2GM)} \\
\Gamma^\theta_{r\theta} = \dfrac{1}{r} & \Gamma^r_{\theta\theta} = -(r-2GM) & \Gamma^\phi_{r\phi} = \dfrac{1}{r} \\
\Gamma^r_{\phi\phi} = -(r-2GM)\sin^2\theta & \Gamma^\theta_{\phi\phi} = -\sin\theta\cos\theta & \Gamma^\phi_{\theta\phi} = \dfrac{\cos\theta}{\sin\theta}.
\end{array}
\tag{5.52}
$$

The geodesic equation therefore turns into the following four equations, where λ is an affine parameter:

$$
\begin{aligned}
\frac{d^2t}{d\lambda^2} + \frac{2GM}{r(r-2GM)}\frac{dr}{d\lambda}\frac{dt}{d\lambda} &= 0, \\
\frac{d^2r}{d\lambda^2} + \frac{GM}{r^3}(r-2GM)\left(\frac{dt}{d\lambda}\right)^2 - \frac{GM}{r(r-2GM)}\left(\frac{dr}{d\lambda}\right)^2 & \\
-(r-2GM)\left[\left(\frac{d\theta}{d\lambda}\right)^2 + \sin^2\theta\left(\frac{d\phi}{d\lambda}\right)^2\right] &= 0, \\
\frac{d^2\theta}{d\lambda^2} + \frac{2}{r}\frac{d\theta}{d\lambda}\frac{dr}{d\lambda} - \sin\theta\cos\theta\left(\frac{d\phi}{d\lambda}\right)^2 &= 0, \\
\frac{d^2\phi}{d\lambda^2} + \frac{2}{r}\frac{d\phi}{d\lambda}\frac{dr}{d\lambda} + 2\frac{\cos\theta}{\sin\theta}\frac{d\theta}{d\lambda}\frac{d\phi}{d\lambda} &= 0.
\end{aligned}
\tag{5.53}
$$

There does not seem to be much hope for simply solving this set of coupled equations by inspection. Fortunately our task is greatly simplified by the high degree of symmetry of the Schwarzschild metric. We know that there are four Killing vectors: three for the spherical symmetry, and one for time translations. Each of these will lead to a constant of the motion for a free particle. If K^μ is a Killing vector, we know that

$$
K_\mu \frac{dx^\mu}{d\lambda} = \text{constant}. \tag{5.54}
$$

In addition, we always have another constant of the motion for geodesics: the geodesic equation (together with metric compatibility) implies that the quantity

$$
\epsilon = -g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} \tag{5.55}
$$

is constant along the path. (For any trajectory we can choose the parameter λ such that ϵ is a constant; we are simply noting that this is compatible with affine parameterization along a geodesic.) Of course, for a massive particle we typically choose $\lambda = \tau$, and this relation simply becomes $\epsilon = -g_{\mu\nu}U^\mu U^\nu = +1$. For massless particles, which move along null trajectories, we always have $\epsilon = 0$, and this equation does not fix the parameter λ. As discussed in Section 3.4, it is convenient to normalize λ along null geodesics such that the four-momentum and four-velocity are equal, $p^\mu = dx^\mu/d\lambda$. We might also be concerned with spacelike geodesics (even though they do not correspond to paths of particles), for which we will choose $\epsilon = -1$.

Rather than immediately writing out explicit expressions for the four conserved quantities associated with Killing vectors, let's think about what they are telling us. Notice that the symmetries they represent are also present in flat spacetime, where the conserved quantities they lead to are very familiar. Invariance under time translations leads to conservation of energy, while invariance under spatial rotations leads to conservation of the three components of angular momentum. Essentially the same applies to the Schwarzschild metric. We can think of the angular momentum as a three-vector with a magnitude (one component) and direction (two components). Conservation of the direction of angular momentum means that the particle will move in a plane. We can choose this to be the equatorial plane of our coordinate system; if the particle is not in this plane, we can rotate coordinates until it is. Thus, the two Killing vectors that lead to conservation of the direction of angular momentum imply that, for a single particle, we can choose

$$\theta = \frac{\pi}{2}. \tag{5.56}$$

The two remaining Killing vectors correspond to energy and the magnitude of angular momentum. The energy arises from the timelike Killing vector

$$K^\mu = (\partial_t)^\mu = (1, 0, 0, 0). \tag{5.57}$$

The Killing vector whose conserved quantity is the magnitude of the angular momentum is

$$R^\mu = (\partial_\phi)^\mu = (0, 0, 0, 1). \tag{5.58}$$

In both cases it is convenient to lower the index to obtain

$$K_\mu = \left(-\left(1 - \frac{2GM}{r}\right), 0, 0, 0\right) \tag{5.59}$$

and

$$R_\mu = \left(0, 0, 0, r^2 \sin^2\theta\right). \tag{5.60}$$

Since (5.56) implies that $\sin\theta = 1$ along the geodesics of interest to us, the two conserved quantities are

$$E = -K_\mu \frac{dx^\mu}{d\lambda} = \left(1 - \frac{2GM}{r}\right)\frac{dt}{d\lambda} \tag{5.61}$$

and

$$L = R_\mu \frac{dx^\mu}{d\lambda} = r^2 \frac{d\phi}{d\lambda}. \tag{5.62}$$

For massless particles, these can be thought of as the conserved energy and angular momentum, while for massive particles they are the conserved energy and angular momentum per unit mass of the particle. In the discussion of rotating black holes in the next chapter, we will use E and L to refer to the actual energy and angular momentum, not "per unit mass"; the meaning should be clear from context. Note that the constancy of (5.62) is the GR equivalent of Kepler's second law—equal areas are swept out in equal times.

Recall that in Section 3.4 we claimed that the energy of a particle with four-momentum p^μ, as measured by an observer with four-velocity U^μ, would be $-p_\mu U^\mu$. This is *not* equal, or even proportional, to (5.61), even if the observer is taken to be stationary ($U^i = 0$). Mathematically, this is because the four-velocity is normalized to $U_\mu U^\mu = -1$, which the Killing vector K^μ is not: If we tried to normalize it in that way, it would no longer solve Killing's equation. At a slightly deeper level, $-p_\mu U^\mu$ may be thought of as the inertial/kinetic energy of the particle, while $-p_\mu K^\mu$ is the total conserved energy, including the potential energy due to the gravitational field. The notion of gravitational potential energy is not always well-defined, but the total energy is well-defined in the presence of a timelike Killing vector. We will presently use E to help characterize geodesics of Schwarzschild; later we will also use $-p_\mu U^\mu$ for massless particles, where it can be thought of as the observed frequency of a photon, to describe gravitational redshift.

Together the conserved quantities E and L provide a convenient way to understand the orbits of particles in the Schwarzschild geometry. Let us expand the expression (5.55) for ϵ to obtain

$$-\left(1 - \frac{2GM}{r}\right)\left(\frac{dt}{d\lambda}\right)^2 + \left(1 - \frac{2GM}{r}\right)^{-1}\left(\frac{dr}{d\lambda}\right)^2 + r^2\left(\frac{d\phi}{d\lambda}\right)^2 = -\epsilon. \tag{5.63}$$

If we multiply this by $(1 - 2GM/r)$ and use our expressions for E and L, we obtain

$$-E^2 + \left(\frac{dr}{d\lambda}\right)^2 + \left(1 - \frac{2GM}{r}\right)\left(\frac{L^2}{r^2} + \epsilon\right) = 0. \tag{5.64}$$

This is certainly progress, since we have taken a messy system of coupled equations and obtained a single equation for $r(\lambda)$. It looks even nicer if we rewrite it as

$$\frac{1}{2}\left(\frac{dr}{d\lambda}\right)^2 + V(r) = \mathcal{E}, \tag{5.65}$$

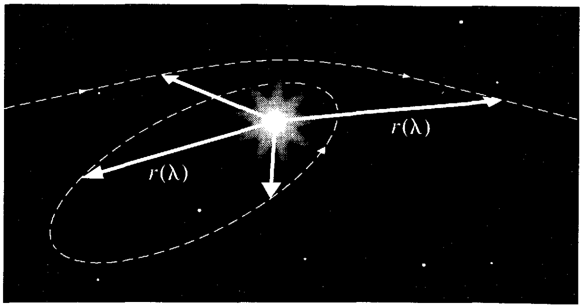

FIGURE 5.3 Orbits around a star are characterized by giving the radius r as a function of a parameter λ.

where

$$V(r) = \frac{1}{2}\epsilon - \epsilon\frac{GM}{r} + \frac{L^2}{2r^2} - \frac{GML^2}{r^3} \tag{5.66}$$

and

$$\mathcal{E} = \frac{1}{2}E^2. \tag{5.67}$$

In (5.65) we have precisely the equation for a classical particle of unit mass and "energy" $\mathcal{E}$ moving in a one-dimensional potential given by $V(r)$. It's a little confusing, but not too bad: the conserved energy per unit mass is E, but the effective potential for the coordinate r responds to $\mathcal{E} = E^2/2$.

Of course, our physical situation is quite different from a classical particle moving in one dimension; the trajectories under consideration are orbits around a star or other object, as shown in Figure 5.3. The quantities of interest to us are not only $r(\lambda)$, but also $t(\lambda)$ and $\phi(\lambda)$. Nevertheless, we can go a long way toward understanding all of the orbits by understanding their radial behavior, and it is a great help to reduce this behavior to a problem we know how to solve.

A similar analysis of orbits in Newtonian gravity would have produced a similar result; the general equation (5.65) would have been the same, but the effective potential (5.66) would not have had the last term. (Note that this equation is not a power series in $1/r$, it is exact.) In the potential (5.66) the first term is just a constant, the second term corresponds exactly to the Newtonian gravitational potential, and the third term is a contribution from angular momentum that takes the same form in Newtonian gravity and general relativity. The last term, the GR contribution, will turn out to make a great deal of difference, especially at small r.

Let us examine the effective potentials for different kinds of possible orbits, as illustrated in Figures 5.4 and 5.5. There are different curves $V(r)$ for different values of L; for any one of these curves, the behavior of the orbit can be judged by comparing $\mathcal{E}$ to $V(r)$. The general behavior of the particle will be to move in the potential until it reaches a "turning point" where $V(r) = \mathcal{E}$, when it will begin

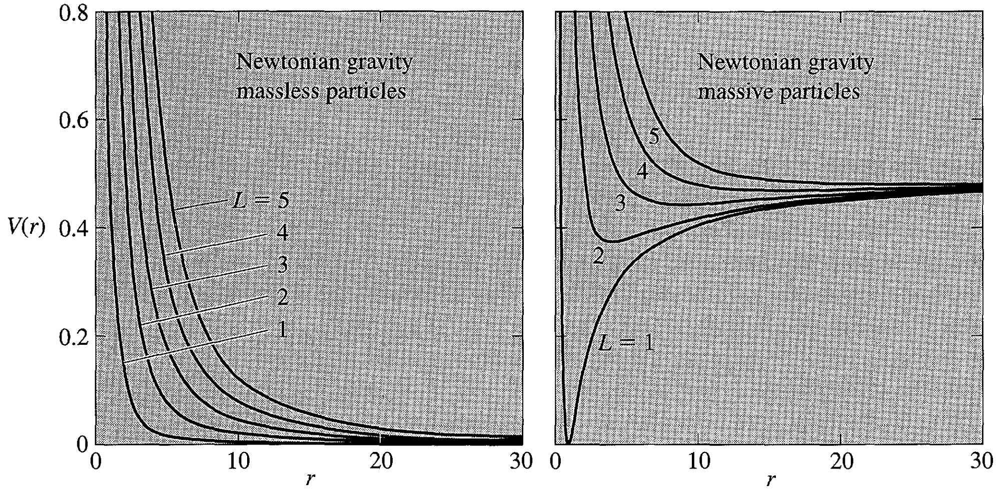

FIGURE 5.4 Effective potentials in Newtonian gravity. Five curves are shown, corresponding to the listed values of the angular momentum (per unit mass) L, and we have chosen $GM = 1$. Note that, for large enough energy, every orbit reaches a turning point and returns to infinity.

moving in the other direction. Sometimes there may be no turning point to hit, in which case the particle just keeps going. In other cases the particle may simply move in a circular orbit at radius $r_c =$ constant; this can happen at points where the potential is flat, $dV/dr = 0$. Differentiating (5.66), we find that the circular orbits occur when

$$\epsilon GMr_c^2 - L^2 r_c + 3GML^2\gamma = 0, \tag{5.68}$$

where $\gamma = 0$ in Newtonian gravity and $\gamma = 1$ in general relativity. Circular orbits will be stable if they correspond to a minimum of the potential, and unstable if they correspond to a maximum. Bound orbits that are not circular will oscillate around the radius of the stable circular orbit.

Turning to Newtonian gravity, we find that circular orbits appear at

$$r_c = \frac{L^2}{\epsilon GM}. \tag{5.69}$$

For massless particles, $\epsilon = 0$, and there are no circular orbits; this is consistent with the first plot in Figure 5.4, which illustrates that there are no bound orbits of any sort. Although it is somewhat obscured in polar coordinates, massless particles actually move in a straight line, since the Newtonian gravitational force on a massless particle is zero. Of course the standing of massless particles in Newtonian theory is somewhat problematic, so you can get different answers depending on what assumptions you make. In terms of the effective potential, a photon with a given energy E will come in from $r = \infty$ and gradually slow down (actually $dr/d\lambda$ will decrease, but the speed of light isn't changing) until it reaches the

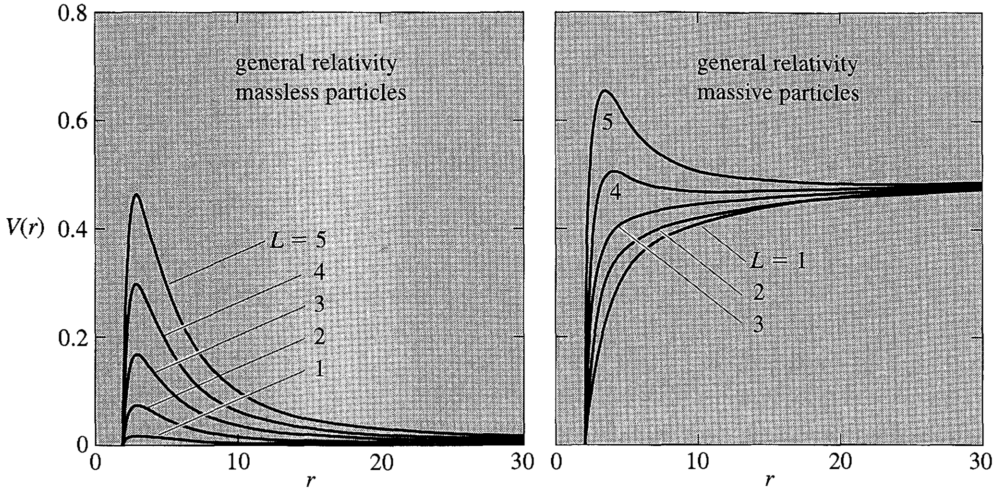

FIGURE 5.5 Effective potentials in general relativity. Again, five curves are shown, corresponding to the listed values of the angular momentum (per unit mass) L, and we have chosen $GM = 1$. In GR there is an innermost circular orbit greater than or equal to $3GM$, and any orbit that falls inside this radius continues to $r = 0$ (for particles on geodesics).

turning point, when it will start moving away back to $r = \infty$. The lower values of L, for which the photon will come closer before it starts moving away, are simply those trajectories that are initially aimed closer to the gravitating body. For massive particles there will be stable circular orbits at the radius (5.69), as well as bound orbits that oscillate around this radius. If the energy is greater than the asymptotic value $E = 1$, the orbits will be unbound, describing a particle that approaches the star and then recedes. We know that the orbits in Newton's theory are conic sections—bound orbits are either circles or ellipses, while unbound ones are either parabolas or hyperbolas—although we won't show that here.

In general relativity the situation is different, but only for r sufficiently small. Since the difference resides in the term $-GML^2/r^3$, as $r \to \infty$ the behaviors are identical in the two theories. But as $r \to 0$ the potential goes to $-\infty$ rather than $+\infty$ as in the Newtonian case. At $r = 2GM$ the potential is always zero; inside this radius is the black hole, which we will discuss more thoroughly later. For massless particles there is always a barrier (except for $L = 0$, for which the potential vanishes identically), but a sufficiently energetic photon will nevertheless go over the barrier and be dragged inexorably down to the center. Note that "sufficiently energetic" means "in comparison to its angular momentum"—in fact the frequency of the photon is immaterial, only the direction in which it is pointing. At the top of the barrier are unstable circular orbits. For $\epsilon = 0$, $\gamma = 1$, we can easily solve (5.68) to obtain

$$r_c = 3GM. \tag{5.70}$$

This is borne out by the first part of Figure 5.5, which shows a maximum of $V(r)$ at $r = 3GM$ for every L. This means that a photon can orbit forever in a circle at this radius, but any perturbation will cause it to fly away either to $r = 0$ or $r = \infty$.

For massive particles there are once again different regimes depending on the angular momentum. The circular orbits are at

$$r_c = \frac{L^2 \pm \sqrt{L^4 - 12G^2M^2L^2}}{2GM}. \tag{5.71}$$

For large L there will be two circular orbits, one stable and one unstable. In the $L \to \infty$ limit their radii are given by

$$r_c = \frac{L^2 \pm L^2(1 - 6G^2M^2/L^2)}{2GM} = \left(\frac{L^2}{GM}, 3GM\right). \tag{5.72}$$

In this limit the stable circular orbit becomes farther away, while the unstable one approaches $3GM$, behavior that parallels the massless case. As we decrease L, the two circular orbits come closer together; they coincide when the discriminant in (5.71) vanishes, which is at

$$L = \sqrt{12}GM, \tag{5.73}$$

for which

$$r_c = 6GM, \tag{5.74}$$

and they disappear entirely for smaller L. Thus $6GM$ is the smallest possible radius of a stable circular orbit in the Schwarzschild metric. There are also unbound orbits, which come in from infinity and turn around, and bound but noncircular orbits, which oscillate around the stable circular radius. Note that such orbits, which would describe exact conic sections in Newtonian gravity, will not do so in GR, although we would have to solve the equation for $d\phi/d\lambda$ to demonstrate it. Finally, there are orbits that come in from infinity and continue all the way in to $r = 0$; this can happen either if the energy is higher than the barrier, or for $L < \sqrt{12}GM$, when the barrier goes away entirely.

We have therefore found that the Schwarzschild solution possesses stable circular orbits for $r > 6GM$ and unstable circular orbits for $3GM < r < 6GM$. It's important to remember that these are only the geodesics; there is nothing to stop an accelerating particle from dipping below $r = 3GM$ and emerging, as long as it stays beyond $r = 2GM$.

5.5 ■ EXPERIMENTAL TESTS

Most experimental tests of general relativity involve the motion of test particles in the solar system, and hence geodesics of the Schwarzschild metric. Einstein

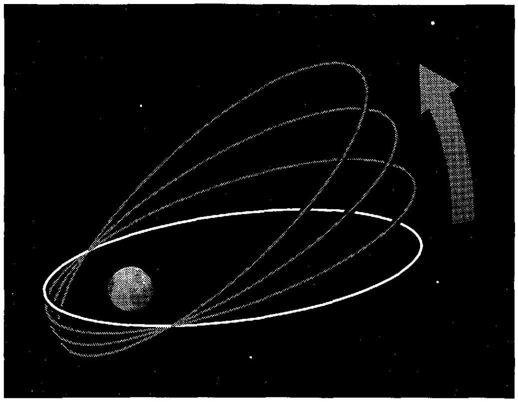

FIGURE 5.6 Orbits in general relativity describe precessing ellipses.

suggested three tests: the deflection of light, the precession of perihelia, and gravitational redshift. The deflection of light is observable in the weak-field limit, and is therefore discussed in Chapter 7. In this section we will discuss the precession of perihelia and the gravitational redshift. (The perihelion of an elliptical orbit is its point of closest approach to the Sun; orbits around the Earth or a star would have perigee or periastron, respectively.)

The precession of perihelia reflects the fact that noncircular orbits in GR are not perfect closed ellipses; to a good approximation they are ellipses that precess, describing a flower pattern as shown in Figure 5.6. Despite its conceptual simplicity, the rate of perihelion precession is somewhat cumbersome to calculate; here we follow d'Inverno (1992). The strategy is to describe the evolution of the radial coordinate r as a function of the angular coordinate ϕ; for a perfect ellipse, $r(\phi)$ would be periodic with period 2π, reflecting the fact that perihelion occurred at the same angular position each orbit. Using perturbation theory we can show how GR introduces a slight alteration of the period, giving rise to precession.

We start with our radial equation of motion of a massive particle in a Schwarzschild metric (5.65). To get an equation for $dr/d\phi$ we multiply by

$$\left(\frac{d\phi}{d\lambda}\right)^{-2} = \frac{r^4}{L^2}, \tag{5.75}$$

which yields

$$\left(\frac{dr}{d\phi}\right)^2 + \frac{1}{L^2}r^4 - \frac{2GM}{L^2}r^3 + r^2 - 2GMr = \frac{2\mathcal{E}}{L^2}r^4. \tag{5.76}$$

Two tricks are useful in solving this equation. The first trick is to define a new variable

$$x = \frac{L^2}{GMr}. \tag{5.77}$$

From (5.69) we see that $x = 1$ at a Newtonian circular orbit. Our equation of motion (5.76) becomes

$$\left(\frac{dx}{d\phi}\right)^2 + \frac{L^2}{G^2M^2} - 2x + x^2 - \frac{2G^2M^2}{L^2}x^3 = \frac{2\mathcal{E}L^2}{G^2M^2}. \tag{5.78}$$

The second trick is to differentiate this with respect to ϕ, obtaining a second-order equation for $x(\phi)$:

$$\frac{d^2x}{d\phi^2} - 1 + x = \frac{3G^2M^2}{L^2}x^2. \tag{5.79}$$

In a Newtonian calculation, the last term would be absent, and we could solve for x exactly; here, we can treat it as a perturbation.

We expand x into a Newtonian solution plus a small deviation,

$$x = x_0 + x_1. \tag{5.80}$$

The zeroth-order part of (5.79) is then

$$\frac{d^2x_0}{d\phi^2} - 1 + x_0 = 0 \tag{5.81}$$

and the first-order part is

$$\frac{d^2x_1}{d\phi^2} + x_1 = \frac{3G^2M^2}{L^2}x_0^2. \tag{5.82}$$

The solution for the zeroth-order equation can be written

$$x_0 = 1 + e\cos\phi. \tag{5.83}$$

This is the standard result of Newton or Kepler; it describes a perfect ellipse, with e the eccentricity. An ellipse is specified by the semi-major axis a, the distance from the center to the farthest point on the ellipse, and the semi-minor axis b, the distance from the center to the closest point. The eccentricity satisfies $e^2 = 1 - b^2/a^2$.

Plugging the Newtonian solution into the first-order equation (5.82), we obtain

$$\begin{aligned}\frac{d^2x_1}{d\phi^2} + x_1 &= \frac{3G^2M^2}{L^2}(1 + e\cos\phi)^2\\ &= \frac{3G^2M^2}{L^2}\left[\left(1 + \frac{1}{2}e^2\right) + 2e\cos\phi + \frac{1}{2}e^2\cos 2\phi\right].\end{aligned} \tag{5.84}$$

To solve this equation, notice that

$$\frac{d^2}{d\phi^2}(\phi \sin \phi) + \phi \sin \phi = 2 \cos \phi \tag{5.85}$$

and

$$\frac{d^2}{d\phi^2}(\cos 2\phi) + \cos 2\phi = -3 \cos 2\phi. \tag{5.86}$$

Comparing these to (5.84), we see that a solution is provided by

$$x_1 = \frac{3G^2M^2}{L^2}\left[\left(1 + \frac{1}{2}e^2\right) + e\phi \sin \phi - \frac{1}{6}e^2 \cos 2\phi\right], \tag{5.87}$$

as you are welcome to check. The three terms here have different characters. The first is simply a constant displacement, while the third oscillates around zero. The important effect is thus contained in the second term, which accumulates over successive orbits. We therefore combine this term with the zeroth-order solution to write

$$x = 1 + e \cos \phi + \frac{3G^2M^2e}{L^2}\phi \sin \phi. \tag{5.88}$$

This is not a full solution, even to the perturbed equation, but it encapsulates the part that we care about. In particular, this expression for x can be conveniently rewritten as the equation for an ellipse with an angular period that is not quite 2π:

$$x = 1 + e \cos\left[(1 - \alpha)\phi\right], \tag{5.89}$$

where we have introduced

$$\alpha = \frac{3G^2M^2}{L^2}. \tag{5.90}$$

The equivalence of (5.88) and (5.89) can be seen by expanding $\cos[(1 - \alpha)\phi]$ as a power series in the small parameter α:

$$\begin{aligned}\cos\left[(1 - \alpha)\phi\right] &= \cos \phi + \alpha \frac{d}{d\alpha} \cos\left[(1 - \alpha)\phi\right]_{\alpha=0} \\ &= \cos \phi + \alpha\phi \sin \phi.\end{aligned} \tag{5.91}$$

We have therefore found that, during each orbit of the planet, perihelion advances by an angle

$$\Delta\phi = 2\pi\alpha = \frac{6\pi G^2M^2}{L^2}. \tag{5.92}$$

To convert from the angular momentum L to more conventional quantities, we may use expressions valid for Newtonian orbits, since the quantity we're looking at is already a small perturbation. An ordinary ellipse satisfies

$$r = \frac{(1-e^2)a}{1+e\cos\phi}, \tag{5.93}$$

where a is the semi-major axis. Comparing to our zeroth-order solution (5.83) and the definition (5.77) of x, we see that

$$L^2 \approx GM(1-e^2)a. \tag{5.94}$$

This is an approximation, valid if the orbit were a perfect closed ellipse. Plugging this into (5.92) and restoring explicit factors of the speed of light, we obtain

$$\Delta\phi = \frac{6\pi GM}{c^2(1-e^2)a}. \tag{5.95}$$

Historically, the precession of Mercury was the first test of GR. In fact it was known before Einstein invented GR that there was an apparent discrepancy in Mercury's orbit, and a number of solutions had been proposed (including "dark matter" in the inner Solar System). Einstein knew of the discrepancy, and one of his first tasks after formulating GR was to show that it correctly accounted for Mercury's perihelion precession. For the motion of Mercury around the Sun, the relevant orbital parameters are

$$\begin{aligned} \frac{GM_\odot}{c^2} &= 1.48 \times 10^5 \text{ cm}, \\ a &= 5.79 \times 10^{12} \text{ cm} \\ e &= 0.2056, \end{aligned} \tag{5.96}$$

and of course $c = 3.00 \times 10^{10}$ cm/sec. This gives

$$\Delta\phi_{\text{Mercury}} = 5.01 \times 10^{-7} \text{ radians/orbit} = 0.103''/\text{orbit}, \tag{5.97}$$

where $''$ stands for arcseconds. It is more conventional to express this in terms of precession per century; Mercury orbits once every 88 days, yielding

$$\Delta\phi_{\text{Mercury}} = 43.0''/\text{century}. \tag{5.98}$$

So the major axis of Mercury's orbit precesses at a rate of 43.0 arcsecs every 100 years. The observed value is 5601 arcsecs/100 years. However, much of that is due to the precession of equinoxes in our geocentric coordinate system; 5025 arcsecs/100 years, to be precise. The gravitational perturbations of the other planets contribute an additional 532 arcsecs/100 years, leaving 43 arcsecs/100 years to be explained by GR, which it does quite well. You can imagine that Einstein must have been very pleased when he first figured this out.

In Chapter 2 we discussed the gravitational redshift of photons as a consequence of the Principle of Equivalence. The Schwarzschild metric is an exact solution of GR, and should therefore predict a redshift that reduces to the EP prediction in small regions of spacetime. Let's see how that works.

Consider an observer with four-velocity U^μ, who is stationary in the Schwarzschild coordinates ($U^i = 0$). We could allow the observer to be moving, but that would merely superimpose a conventional Doppler shift over the gravitational effect. The four-velocity satisfies $U_\mu U^\mu = -1$, which for a stationary observer in Schwarzschild implies

$$U^0 = \left(1 - \frac{2GM}{r}\right)^{-1/2}. \tag{5.99}$$

Any such observer measures the frequency of a photon following along a null geodesic $x^\mu(\lambda)$ to be

$$\omega = -g_{\mu\nu} U^\mu \frac{dx^\nu}{d\lambda}. \tag{5.100}$$

Indeed, this relation defines the normalization of λ. We therefore have

$$\omega = \left(1 - \frac{2GM}{r}\right)^{1/2} \frac{dt}{d\lambda} \tag{5.101}$$

$$= \left(1 - \frac{2GM}{r}\right)^{-1/2} E, \tag{5.102}$$

where E is defined by (5.61), applied to the photon trajectory. E is conserved, so ω will clearly take on different values when measured at different radial distances. For a photon emitted at r_1 and observed at r_2, the observed frequencies will be related by

$$\frac{\omega_2}{\omega_1} = \left(\frac{1 - 2GM/r_1}{1 - 2GM/r_2}\right)^{1/2}. \tag{5.103}$$

This is an exact result for the frequency shift; in the limit $r \gg 2GM$ we have

$$\begin{aligned}\frac{\omega_2}{\omega_1} &= 1 - \frac{GM}{r_1} + \frac{GM}{r_2} \\ &= 1 + \Phi_1 - \Phi_2, \end{aligned} \tag{5.104}$$

where $\Phi = -GM/r$ is the Newtonian potential. This tells us that the frequency goes down as Φ increases, which happens as we climb out of a gravitational field: thus, a redshift. (Photons that fall toward a gravitating body are blueshifted.) We see that the $r \gg 2GM$ result agrees with the calculation based on the Equivalence Principle.

The gravitational redshift was first detected in 1960 by Pound and Rebka, using gamma rays traveling upward a distance of only 72 feet (the height of the physics building at Harvard). Subsequent tests have become increasingly precise, often making use of artificial spacecraft or atomic clocks carried aboard airplanes. The agreement with Einstein's predictions has been excellent in all cases.

Since Einstein's proposal of the three classic tests, further tests of GR have been proposed. The most famous is of course the binary pulsar, to be discussed in Chapter 7. Another is the gravitational time delay, discovered and observed by Shapiro, also discussed in Chapter 7. In a very different context, Big-Bang nucleosynthesis provides a cosmological test of GR at an epoch when the universe was only seconds old, as discussed in Chapter 8. Modern advances have also introduced a host of new tests; for a comprehensive introduction see Will (1981).

5.6 ■ SCHWARZSCHILD BLACK HOLES

We now know something about the behavior of geodesics outside the troublesome radius $r = 2GM$, which is the regime of interest for the solar system and most other astrophysical situations. We next turn to the study of objects that are described by the Schwarzschild solution even at radii smaller than $2GM$—black holes. (We'll use the term "black hole" for the moment, even though we haven't introduced a precise meaning for such an object.)

One way to understand the geometry of a spacetime is to explore its causal structure, as defined by the light cones. We therefore consider radial null curves, those for which θ and ϕ are constant and $ds^2 = 0$:

$$ds^2 = 0 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2, \tag{5.105}$$

from which we see that

$$\frac{dt}{dr} = \pm\left(1 - \frac{2GM}{r}\right)^{-1}. \tag{5.106}$$

This of course measures the slope of the light cones on a spacetime diagram of the t-r plane. For large r the slope is ± 1, as it would be in flat space, while as we approach $r = 2GM$ we get $dt/dr \to \pm\infty$, and the light cones "close up," as shown in Figure 5.7. Thus a light ray that approaches $r = 2GM$ never seems to get there, at least in this coordinate system; instead it seems to asymptote to this radius.

As we will see, the apparent inability to get to $r = 2GM$ is an illusion, and the light ray (or a massive particle) actually has no trouble reaching this radius. But an observer far away would never be able to tell. If we stayed outside while an intrepid observational general relativist dove into the black hole, sending back signals all the time, we would simply see the signals reach us more and more slowly, as portrayed in Figure 5.8. In the Exercises you are asked to look at this phenomenon more carefully. As an infalling observer approaches $r = 2GM$, any fixed interval $\Delta\tau_1$ of their proper time corresponds to a longer and longer interval $\Delta\tau_2$ from our point of view. This continues forever; we would never see

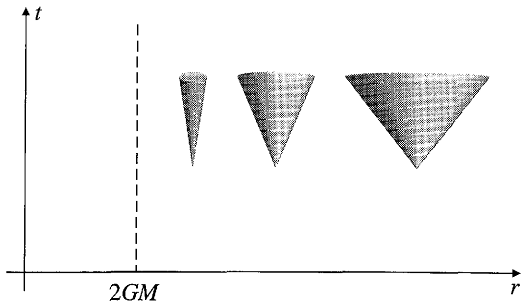

FIGURE 5.7 In Schwarzschild coordinates, light cones appear to close up as we approach $r = 2GM$.

the observer cross $r = 2GM$, we would just see them move more and more slowly (and become redder and redder, as if embarrassed to have done something as stupid as diving into a black hole).

The fact that we never see the infalling observer reach $r = 2GM$ is a meaningful statement, but the fact that their trajectory in the t-r plane never reaches there is not. It is highly dependent on our coordinate system, and we would like to ask a more coordinate-independent question (such as, "Does the observer reach this radius in a finite amount of their proper time?"). The best way to do this is to change coordinates to a system that is better behaved at $r = 2GM$. We now set out to find an appropriate set of such coordinates. There is no way to "derive" a coordinate transformation, of course, we just say what the new coordinates are and plug in the formulas. But we will develop these coordinates in several steps, in hopes of making the choices seem somewhat motivated.

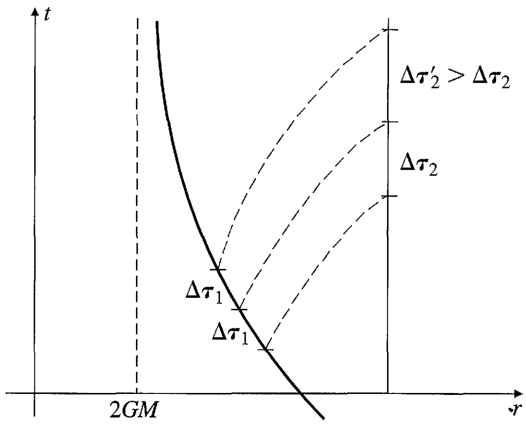

FIGURE 5.8 A beacon falling freely into a black hole emits signals at intervals of constant proper time $\Delta\tau_1$. An observer at fixed r receives the signals at successively longer time intervals $\Delta\tau_2$.

The problem with our current coordinates is that $dt/dr \to \infty$ along radial null geodesics that approach $r = 2GM$; progress in the r direction becomes slower and slower with respect to the coordinate time t. We can try to fix this problem by replacing t with a coordinate that moves more slowly along null geodesics. First notice that we can explicitly solve the condition (5.106) characterizing radial null curves to obtain

$$t = \pm r^* + \text{constant}, \tag{5.107}$$

where the **tortoise coordinate** r^* is defined by

$$r^* = r + 2GM \ln\left(\frac{r}{2GM} - 1\right). \tag{5.108}$$

(The tortoise coordinate is only sensibly related to r when $r \geq 2GM$, but beyond there our coordinates aren't very good anyway.) In terms of the tortoise coordinate the Schwarzschild metric becomes

$$ds^2 = \left(1 - \frac{2GM}{r}\right)\left(-dt^2 + dr^{*2}\right) + r^2\, d\Omega^2, \tag{5.109}$$

where r is thought of as a function of r^*. This represents some progress, since the light cones now don't seem to close up, as shown in Figure 5.9; furthermore, none of the metric coefficients becomes infinite at $r = 2GM$ (although both g_{tt} and $g_{r^*r^*}$ become zero). The price we pay, however, is that the surface of interest at $r = 2GM$ has just been pushed to infinity.

Our next move is to define coordinates that are naturally adapted to the null geodesics. If we let

$$\begin{aligned} v &= t + r^* \\ u &= t - r^*, \end{aligned} \tag{5.110}$$

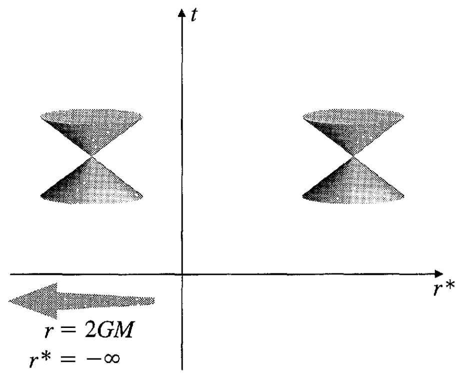

FIGURE 5.9 Schwarzschild light cones in tortoise coordinates, equation (5.109). Light cones remain nondegenerate, but the surface $r = 2GM$ has been pushed to infinity.

then infalling radial null geodesics are characterized by $v =$ constant, while the outgoing ones satisfy $u =$ constant. Now consider going back to the original radial coordinate r, but replacing the timelike coordinate t with the new coordinate v. These are known as **Eddington–Finkelstein coordinates**. In terms of these coordinates the metric is

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dv^2 + (dv\,dr + dr\,dv) + r^2\,d\Omega^2. \tag{5.111}$$

Here we see our first sign of real progress. Even though the metric coefficient g_{vv} vanishes at $r = 2GM$, there is no real degeneracy; the determinant of the metric is

$$g = -r^4 \sin^2\theta, \tag{5.112}$$

which is perfectly regular at $r = 2GM$. Therefore the metric is invertible, and we see once and for all that $r = 2GM$ is simply a coordinate singularity in our original (t, r, θ, ϕ) system. In the Eddington–Finkelstein coordinates the condition for radial null curves is solved by

$$\frac{dv}{dr} = \begin{cases} 0, & \text{(infalling)} \\ 2\left(1 - \frac{2GM}{r}\right)^{-1}. & \text{(outgoing)} \end{cases} \tag{5.113}$$

We can therefore see what has happened: In this coordinate system the light cones remain well-behaved at $r = 2GM$, and this surface is at a finite coordinate value. There is no problem in tracing the paths of null or timelike particles past the surface. On the other hand, something interesting is certainly going on. Although the light cones don't close up, they do tilt over, such that for $r < 2GM$ all future-directed paths are in the direction of decreasing r, as shown in Figure 5.10.

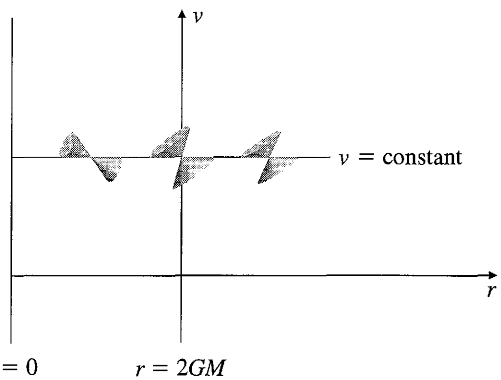

FIGURE 5.10 Schwarzschild light cones in the (v, r) coordinates of (5.111). In these coordinates we can follow future-directed timelike paths past $r = 2GM$.

The surface $r = 2GM$, while being locally perfectly regular, globally functions as a point of no return—once a test particle dips below it, it can never come back. We define an **event horizon** to be a surface past which particles can never escape to infinity; in Schwarzschild the event horizon is located at $r = 2GM$. (This is a rough definition; we will be somewhat more precise in the next chapter.) Despite being located at fixed radial coordinate, the event horizon is a null surface rather than a timelike one, so it is really the causal structure of spacetime itself that makes it impossible to cross the horizon in an outward-going direction. Since nothing can escape the event horizon, it is impossible for us to see inside—thus the name **black hole**. A black hole is simply a region of spacetime separated from infinity by an event horizon. The notion of an event horizon is a global one; the location of the horizon is a statement about the spacetime as a whole, not something you could determine just by knowing the geometry at that location. This will continue to be true in more general spacetimes.

We should mention a couple of features of black holes that sometimes get confused in the popular imagination. First, the external geometry of a black hole is the same Schwarzschild solution that we would have outside a star or planet. In particular, a black hole does not suck in everything around it any more than the Sun does; a particle well outside $r = 2GM$ behaves in exactly the same way regardless of whether the gravitating source is a black hole or not. Second, there is a misleading Newtonian analogy for black holes. The Newtonian escape velocity of a particle at distance r from a gravitating body of mass M is

$$v_{\rm esc} = \sqrt{\frac{2GM}{r}}. \tag{5.114}$$

If we naively ask where the Newtonian escape velocity equals the velocity of light, we find exactly $r = 2GM$. Despite the fact that the speed of light plays no fundamental role in Newtonian theory, it might seem provocative that light, thought of as inertial particles moving at a velocity c, is seemingly not able to escape from a body with mass M and radius less than $2GM$. But there is a profound difference between this case and what we see in GR. The escape velocity is the velocity that a particle would initially need to have in order to escape from a gravitating source on a free trajectory. But nothing stops us from considering accelerated trajectories; for example, one could imagine an acceleration chosen such that the particle moved steadily away from the massive body at some constant velocity. Therefore, a purported Newtonian black hole would not have the crucial property that *nothing* can escape; whereas in GR, arbitrary timelike paths must stay inside their light cones, and hence never escape the event horizon.

5.7 ■ THE MAXIMALLY EXTENDED SCHWARZSCHILD SOLUTION

Let's review what we have done. Acting under the suspicion that our coordinates may not have been good for the entire manifold, we have changed from our original coordinate t to the new one v, which has the nice property that if we decrease

r along a radial null curve $v =$ constant, we go right through the event horizon without any problems. Indeed, a local observer actually making the trip would not necessarily know when the event horizon had been crossed—the local geometry is no different from anywhere else. We therefore conclude that our suspicion was correct and our initial coordinate system didn't do a good job of covering the entire manifold. The region $r \leq 2GM$ should certainly be included in our spacetime, since physical particles can easily reach there and pass through. However, there is no guarantee that we are finished; perhaps we can extend our manifold in other directions.

In fact there are other directions. In the (v, r) coordinate system we can cross the event horizon on future-directed paths, but not on past-directed ones. This seems unreasonable, since we started with a time-independent solution. But we could have chosen u instead of v, in which case the metric would have been

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) du^2 - (du dr + dr du) + r^2 d\Omega^2. \tag{5.115}$$

Now we can once again pass through the event horizon, but this time only along past-directed curves, as shown in Figure 5.11.

This is perhaps a surprise: we can consistently follow either future-directed or past-directed curves through $r = 2GM$, but we arrive at different places. It was actually to be expected, since from the definitions (5.110), if we keep v constant and decrease r we must have $t \to +\infty$, while if we keep u constant and decrease r we must have $t \to -\infty$. (The tortoise coordinate r^* goes to $-\infty$ as $r \to 2GM$.) So we have extended spacetime in two different directions, one to the future and one to the past.

The next step would be to follow spacelike geodesics to see if we would uncover still more regions. The answer is yes, we would reach yet another piece of the spacetime, but let's shortcut the process by defining coordinates that are good all over. A first guess might be to use both u and v at once (in place of t and r),

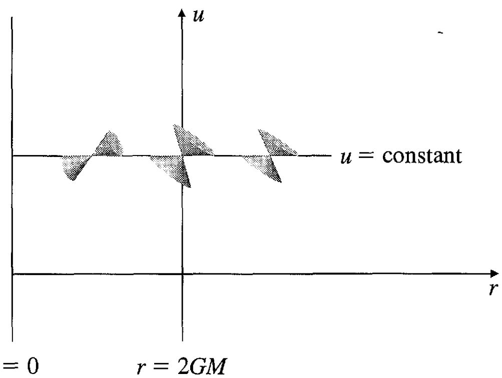

FIGURE 5.11 Schwarzschild light cones in the (u, r) coordinates of (5.115). In these coordinates we can follow past-directed timelike paths past $r = 2GM$.

which leads to

$$ds^2 = -\frac{1}{2}\left(1 - \frac{2GM}{r}\right)(dv du + du\, dv) + r^2 d\Omega^2, \tag{5.116}$$

with r defined implicitly in terms of v and u by

$$\frac{1}{2}(v - u) = r + 2GM \ln\left(\frac{r}{2GM} - 1\right). \tag{5.117}$$

We have actually reintroduced the degeneracy with which we started out; in these coordinates $r = 2GM$ is "infinitely far away" (at either $v = -\infty$ or $u = +\infty$). The thing to do is to change to coordinates that pull these points into finite coordinate values; a good choice is

$$\begin{aligned} v' &= e^{v/4GM} \\ u' &= -e^{-u/4GM}, \end{aligned} \tag{5.118}$$

which in terms of our original (t, r) system is

$$\begin{aligned} v' &= \left(\frac{r}{2GM} - 1\right)^{1/2} e^{(r+t)/4GM} \\ u' &= -\left(\frac{r}{2GM} - 1\right)^{1/2} e^{(r-t)/4GM}. \end{aligned} \tag{5.119}$$

In the (v', u', θ, ϕ) system the Schwarzschild metric is

$$ds^2 = -\frac{16G^3M^3}{r} e^{-r/2GM}(dv' du' + du' dv') + r^2 d\Omega^2. \tag{5.120}$$

Finally the nonsingular nature of $r = 2GM$ becomes completely manifest; in this form none of the metric coefficients behaves in any special way at the event horizon.

Both v' and u' are null coordinates, in the sense that their partial derivatives $\partial/\partial v'$ and $\partial/\partial u'$ are null vectors. There is nothing wrong with this, since the collection of four partial derivative vectors (two null and two spacelike) in this system serve as a perfectly good basis for the tangent space. Nevertheless, we are somewhat more comfortable working in a system where one coordinate is timelike and the rest are spacelike. We therefore define

$$T = \frac{1}{2}(v' + u') = \left(\frac{r}{2GM} - 1\right)^{1/2} e^{r/4GM} \sinh\left(\frac{t}{4GM}\right) \tag{5.121}$$

and

$$R = \frac{1}{2}(v' - u') = \left(\frac{r}{2GM} - 1\right)^{1/2} e^{r/4GM} \cosh\left(\frac{t}{4GM}\right), \tag{5.122}$$

in terms of which the metric becomes

$$\boxed{ds^2 = \frac{32G^3M^3}{r} e^{-r/2GM}(-\mathrm{d}T^2 + \mathrm{d}R^2) + r^2\, d\Omega^2,} \tag{5.123}$$

where r is defined implicitly from

$$T^2 - R^2 = \left(1 - \frac{r}{2GM}\right) e^{r/2GM}. \tag{5.124}$$

The coordinates (T, R, θ, ϕ) are known as **Kruskal coordinates**, or sometimes Kruskal–Szekres coordinates.

The Kruskal coordinates have a number of miraculous properties. Like the (t, r^*) coordinates, the radial null curves look like they do in flat space:

$$T = \pm R + \text{constant}. \tag{5.125}$$

Unlike the (t, r^*) coordinates, however, the event horizon $r = 2GM$ is not infinitely far away; in fact it is defined by

$$T = \pm R, \tag{5.126}$$

consistent with it being a null surface. More generally, we can consider the surfaces $r = \text{constant}$. From (5.124) these satisfy

$$T^2 - R^2 = \text{constant}. \tag{5.127}$$

Thus, they appear as hyperbolae in the R-T plane. Furthermore, the surfaces of constant t are given by

$$\frac{T}{R} = \tanh\left(\frac{t}{4GM}\right), \tag{5.128}$$

which defines straight lines through the origin with slope $\tanh(t/4GM)$. Note that as $t \to \pm\infty$ (5.128) becomes the same as (5.126); therefore $t = \pm\infty$ represents the same surface as $r = 2GM$.

Our coordinates (T, R) should be allowed to range over every value they can take without hitting the real singularity at $r = 0$; the allowed region is therefore

$$\begin{aligned} -\infty \le R \le \infty \\ T^2 < R^2 + 1. \end{aligned} \tag{5.129}$$

From (5.121) and (5.122), T and R seem to become imaginary for $r < 2GM$, but this is an illusion; in that region the (r, t) coordinates are no good (specifically, $|t| > \infty$). We can now draw a spacetime diagram in the T-R plane (with θ and ϕ suppressed), known as a **Kruskal diagram**, shown in Figure 5.12. Each point on the diagram is a two-sphere. This diagram represents the maximal extension

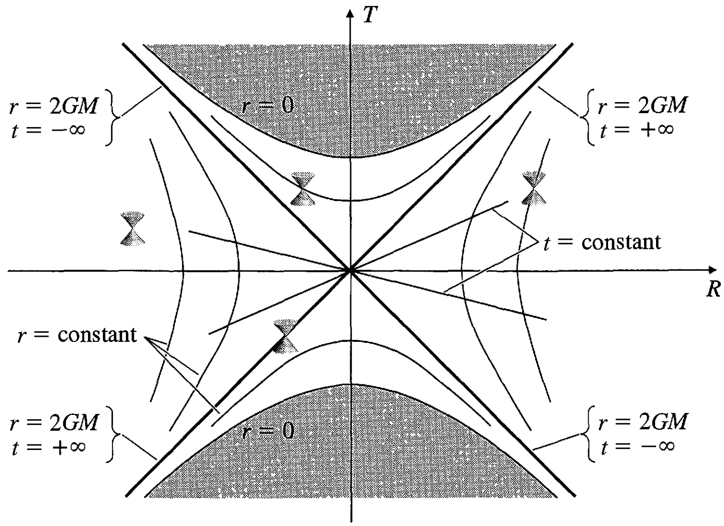

FIGURE 5.12 The Kruskal diagram—the Schwarzschild solution in Kruskal coordinates, where all light cones are at ±45°.

of the Schwarzschild geometry; the coordinates cover what we should think of as the entire manifold described by this solution.

The original Schwarzschild coordinates (t, r) were good for $r > 2GM$, which is only a part of the manifold portrayed on the Kruskal diagram. It is convenient to divide the diagram into four regions, as shown in Figure 5.13. Region I corresponds to $r > 2GM$, the patch in which our original coordinates were well-defined. By following future-directed null rays we reach region II, and by following past-directed null rays we reach region III. If we had explored spacelike geodesics, we would have been led to region IV. The definitions (5.121) and (5.122), which relate (T, R) to (t, r), are really only good in region I; in the other regions it is necessary to introduce appropriate minus signs to prevent the coordinates from becoming imaginary.

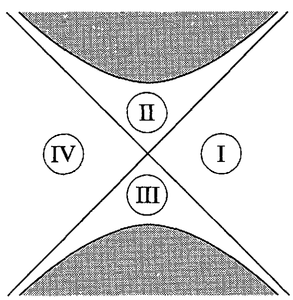

FIGURE 5.13 Regions of the Kruskal diagram.

Having extended the Schwarzschild geometry as far as it will go, we have described a remarkable spacetime. Region II, of course, is what we think of as the black hole. Once anything travels from region I into II, it can never return. In fact, every future-directed path in region II ends up hitting the singularity at $r = 0$; once you enter the event horizon, you are utterly doomed. This is worth stressing; not only can you not escape back to region I, you cannot even stop yourself from moving in the direction of decreasing r, since this is simply the timelike direction. This could have been seen in our original coordinate system; for $r < 2GM$, t becomes spacelike and r becomes timelike. Thus you can no more stop moving toward the singularity than you can stop getting older. Since proper time is maximized along a geodesic, you will live the longest if you don't struggle, but just relax as you approach the singularity. Not that you will have long to relax, nor will the voyage be very relaxing; as you approach the singularity the tidal forces become infinite. As you fall toward the singularity your feet and head will be pulled apart from each other, while your torso is squeezed to infinitesimal thinness. The grisly demise of an astrophysicist falling into a black hole is detailed in Misner, Thorne, and Wheeler (1973), Section 32.6. Note that they use orthonormal frames, as we discuss in Appendix J (not that it makes the trip any more enjoyable).

Regions III and IV might be somewhat unexpected. Region III is simply the time-reverse of region II, a part of spacetime from which things can escape to us, while we can never get there. It can be thought of as a **white hole**. There is a singularity in the past, out of which the universe appears to spring. The boundary of region III is the past event horizon, while the boundary of region II is the future event horizon. Region IV, meanwhile, cannot be reached from our region I either forward or backward in time, nor can anybody from over there reach us. It is another asymptotically flat region of spacetime, a mirror image of ours. It can be thought of as being connected to region I by a wormhole (or Einstein–Rosen bridge), a neck-like configuration joining two distinct regions. Consider slicing up the Kruskal diagram into spacelike surfaces of constant T, as shown in Figure 5.14. Now we can draw pictures of each slice, restoring one of the angular coordinates for clarity, as in Figure 5.15. In this way of slicing, the Schwarzschild geometry describes two asymptotically flat regions that reach toward each other, join together via a wormhole for a while, and then disconnect. But the wormhole closes up too quickly for any timelike observer to cross it from one region into the next.

As pleasing as the Kruskal diagram is, it is often even more useful to collapse the Schwarzschild solution into a finite region by constructing its conformal diagram. The idea of a conformal diagram is discussed in Appendix H; it is a crucial tool for analyzing spacetimes in general relativity, and you are encouraged to review that discussion now. We will not go through the manipulations necessary to construct the conformal diagram of Schwarzschild in full detail, since they parallel the Minkowski case with considerable additional algebraic complexity. We would start with the null version of the Kruskal coordinates, in which the metric

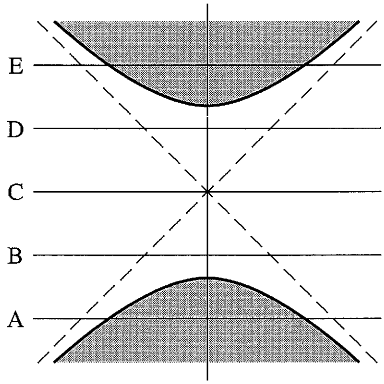

FIGURE 5.14 Spacelike slices in Kruskal coordinates.

takes the form

$$ds^2 = -\frac{16G^3M^3}{r}e^{-r/2GM}(dv'du' + du'dv') + r^2\,d\Omega^2, \tag{5.130}$$

where r is defined implicitly via

$$v'u' = -\left(\frac{r}{2GM} - 1\right)e^{r/2GM}. \tag{5.131}$$

Then essentially the same transformation used in the flat spacetime case suffices to bring infinity into finite coordinate values:

$$\begin{aligned} v'' &= \arctan\left(\frac{v'}{\sqrt{2GM}}\right) \\ u'' &= \arctan\left(\frac{u'}{\sqrt{2GM}}\right), \end{aligned} \tag{5.132}$$

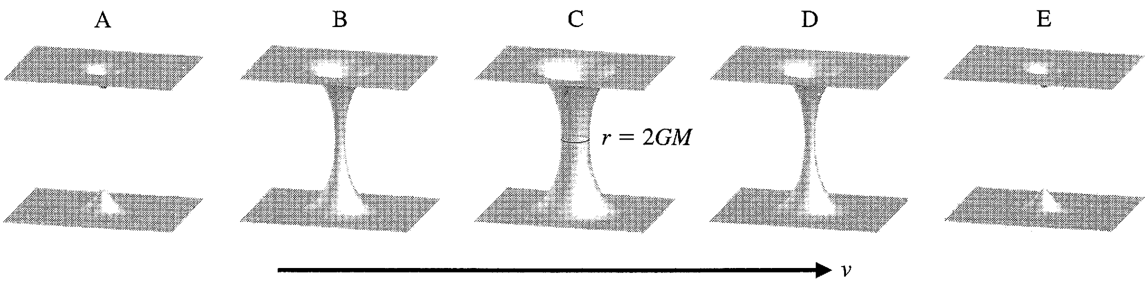

FIGURE 5.15 Geometry of the spacelike slices in Figure 5.14.

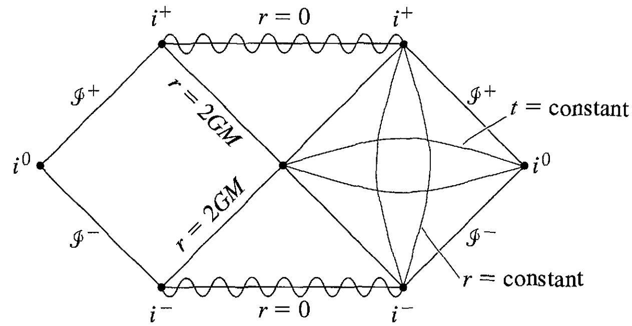

FIGURE 5.16 Conformal diagram for Schwarzschild spacetime.

with ranges

$$-\frac{\pi}{2} < v'' < +\frac{\pi}{2}$$
$$-\frac{\pi}{2} < u'' < +\frac{\pi}{2}$$
$$-\frac{\pi}{2} < v'' + u'' < \frac{\pi}{2}.$$

The $(v,'' u'')$ part of the metric (that is, at constant angular coordinates) is now conformally related to Minkowski space. In the new coordinates the singularities at $r = 0$ are straight lines that stretch from timelike infinity in one asymptotic region to timelike infinity in the other.

The conformal diagram for the maximally extended Schwarzschild solution thus looks like Figure 5.16. The only real subtlety about this diagram is the necessity to understand that i^+ and i^- (future and past infinity) are distinct from $r = 0$—there are plenty of timelike paths that do not hit the singularity. As in the Kruskal diagram, light cones in the conformal diagram are at 45°; the major difference is that the entire spacetime is represented in a finite region. Notice also that the structure of conformal infinity is just like that of Minkowski space, consistent with the claim that Schwarzschild is asymptotically flat.

5.8 ■ STARS AND BLACK HOLES

The maximally extended Schwarzschild solution we have just constructed tells a remarkable story, including not only the sought-after black hole, but also a white hole and an additional asymptotically flat region, connected to our universe by a wormhole. It would be premature, however, to imagine that such features are common in the real world. The Schwarzschild solution represents a highly idealized situation: not only spherically symmetric, but completely free of energy-momentum throughout spacetime. Birkhoff's theorem implies that any vacuum *region* of a spherically symmetric spacetime will be described by *part of* the Schwarzschild metric, but the existence of matter somewhere in the universe may dramatically alter the global picture.

A static spherical object—let's call it a star for definiteness—with radius larger than $2GM$ will be Schwarzschild in the exterior, but there won't be any singularities or horizons, and the global structure will actually be very similar to Minkowski spacetime. Of course, real stars evolve, and it may happen that a star eventually collapses under its own gravitational pull, shrinking down to below $r = 2GM$ and further into a singularity, resulting in a black hole. There is no need for a white hole, however, because the past of such a spacetime looks nothing like that of the full Schwarzschild solution. A conformal diagram describing stellar collapse would look like Figure 5.17. The interior shaded region is nonvacuum, so is not described by Schwarzschild; in particular, there is no wormhole connecting to another universe. It is asymptotically Minkowskian, except for a future region giving rise to an event horizon. We see that a realistic black hole may share the singularity and future horizon with the maximally extended Schwarzschild solution, without any white hole, past horizon, or separate asymptotic region.

We believe that gravitational collapse of this kind is by no means a necessary endpoint of stellar evolution, but will occur under certain conditions. General relativity places rigorous limits on the kind of stars that can resist gravitational collapse; for any given sort of matter, enough mass will always lead to the collapse to a black hole. Furthermore, from astrophysical observations we have excellent evidence that black holes exist in our universe.

To understand gravitational collapse to a black hole, we should first understand static configurations describing the interiors of spherically symmetric stars. We won't delve into this subject in detail, only enough to get a feeling for the basic features of interior solutions. Consider the general static, spherically symmetric

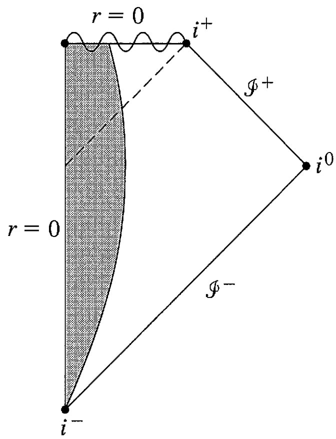

FIGURE 5.17 Conformal diagram for a black hole formed from a collapsing star. The shaded region contains matter, and will be described by an appropriate dynamical interior solution; the exterior region is Schwarzschild.

metric from (5.11):

$$ds^2 = -e^{2\alpha(r)}\,dt^2 + e^{2\beta(r)}\,dr^2 + r^2\,d\Omega^2. \tag{5.133}$$

We are now looking for nonvacuum solutions, so we turn to the full Einstein equation,

$$G_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{5.134}$$

The Einstein tensor follows from the Ricci tensor (5.14) and curvature scalar (5.15),

$$\begin{aligned}
G_{tt} &= \frac{1}{r^2}e^{2(\alpha-\beta)}\left(2r\partial_r\beta - 1 + e^{2\beta}\right)\\
G_{rr} &= \frac{1}{r^2}\left(2r\partial_r\alpha + 1 - e^{2\beta}\right)\\
G_{\theta\theta} &= r^2 e^{-2\beta}\left[\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\alpha\partial_r\beta + \frac{1}{r}(\partial_r\alpha - \partial_r\beta)\right]\\
G_{\phi\phi} &= \sin^2\theta\, G_{\theta\theta}.
\end{aligned} \tag{5.135}$$

We model the star itself as a perfect fluid, with energy-momentum tensor

$$T_{\mu\nu} = (\rho + p)U_\mu U_\nu + pg_{\mu\nu}. \tag{5.136}$$

The energy density ρ and pressure p will be functions of r alone. Since we seek static solutions, we can take the four-velocity to be pointing in the timelike direction. Normalized to $U^\mu U_\mu = -1$, it becomes

$$U_\mu = (e^\alpha, 0, 0, 0), \tag{5.137}$$

so that the components of the energy-momentum tensor are

$$T_{\mu\nu} = \begin{pmatrix} e^{2\alpha}\rho & & & \\ & e^{2\beta}p & & \\ & & r^2 p & \\ & & & r^2(\sin^2\theta)p \end{pmatrix}. \tag{5.138}$$

We therefore have three independent components of Einstein's equation: the tt component,

$$\frac{1}{r^2}e^{-2\beta}\left(2r\partial_r\beta - 1 + e^{2\beta}\right) = 8\pi G\rho, \tag{5.139}$$

the rr component,

$$\frac{1}{r^2}e^{-2\beta}\left(2r\partial_r\alpha + 1 - e^{2\beta}\right) = 8\pi Gp, \tag{5.140}$$

and the $\theta\theta$ component,

$$e^{-2\beta}\left[\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\alpha\partial_r\beta + \frac{1}{r}(\partial_r\alpha - \partial_r\beta)\right] = 8\pi G p. \tag{5.141}$$

The $\phi\phi$ equation is proportional to the $\theta\theta$ equation, so there is no need to consider it separately.

We notice that the tt equation (5.139) involves only β and ρ. It is convenient to replace $\beta(r)$ with a new function $m(r)$, given by

$$m(r) = \frac{1}{2G}(r - re^{-2\beta}), \tag{5.142}$$

or equivalently

$$e^{2\beta} = \left[1 - \frac{2Gm(r)}{r}\right]^{-1}, \tag{5.143}$$

so that

$$ds^2 = -e^{2\alpha(r)}\, dt^2 + \left[1 - \frac{2Gm(r)}{r}\right]^{-1} dr^2 + r^2\, d\Omega^2. \tag{5.144}$$

The metric component g_{rr} is an obvious generalization of the Schwarzschild case, but this will not be true for g_{tt}. The tt equation (5.139) becomes

$$\frac{dm}{dr} = 4\pi r^2 \rho, \tag{5.145}$$

which can be integrated to obtain

$$m(r) = 4\pi \int_0^r \rho(r') r'^2 dr'. \tag{5.146}$$

Let's imagine that our star extends to a radius R, after which we are in vacuum and described by Schwarzschild. In order that the metrics match at this radius, the Schwarzschild mass M must be given by

$$M = m(R) = 4\pi \int_0^R \rho(r) r^2 dr. \tag{5.147}$$

It looks like $m(r)$ is simply the integral of the energy density over the stellar interior, and can be interpreted as the mass within a radius r.

There is one subtlety with interpreting $m(r)$ as the integrated energy density; in a proper spatial integral, the volume element should be

$$\sqrt{\gamma}\, d^3x = e^{\beta} r^2 \sin\theta\, dr d\theta d\phi, \tag{5.148}$$

where

$$\gamma_{ij}\, dx^i dx^j = e^{2\beta}\, dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta\, d\phi^2 \tag{5.149}$$

is the spatial metric. The true integrated energy density is therefore

$$\begin{aligned}\bar{M} &= 4\pi \int_0^R \rho(r) r^2 e^{\beta(r)}\, dr \\ &= 4\pi \int_0^R \frac{\rho(r) r^2}{\left[1 - \frac{2Gm(r)}{r}\right]^{1/2}}\, dr. \end{aligned} \tag{5.150}$$

The difference, of course, arises because there is a binding energy due to the mutual gravitational attraction of the fluid elements in the star, which is given by

$$E_{\rm B} = \bar{M} - M > 0. \tag{5.151}$$

The binding energy is the amount of energy that would be required to disperse the matter in the star to infinity. It is not always a well-defined notion in general relativity, but makes sense for spherical stars.

In terms of $m(r)$, the rr equation (5.140) can be written

$$\frac{d\alpha}{dr} = \frac{Gm(r) + 4\pi G r^3 p}{r[r - 2Gm(r)]}. \tag{5.152}$$

It is convenient not to use the $\theta\theta$ equation directly, but instead appeal to energy-momentum conservation, $\nabla_\mu T^{\mu\nu} = 0$. For our metric (5.144), it is straightforward to derive that $\nu = r$ is the only nontrivial component, and it gives

$$(\rho + p)\frac{d\alpha}{dr} = -\frac{dp}{dr}. \tag{5.153}$$

Combining this with (5.152) allows us to eliminate $\alpha(r)$ to obtain

$$\frac{dp}{dr} = -\frac{(\rho + p)[Gm(r) + 4\pi G r^3 p]}{r[r - 2Gm(r)]}. \tag{5.154}$$

This is the **Tolman–Oppenheimer–Volkoff equation**, or simply the equation of hydrostatic equilibrium. Since $m(r)$ is related to $\rho(r)$ via (5.146), this equation relates $p(r)$ to $\rho(r)$. To get a closed system of equations, we need one more relation: the equation of state. In general this will give the pressure in terms of the energy density and specific entropy, $p = p(\rho, S)$. Often we care about situations in which the entropy is very small, and can be neglected; the equation of state then takes the form

$$p = p(\rho). \tag{5.155}$$

Astrophysical systems often obey a polytropic equation of state, $p = K\rho^\gamma$ for some constants K and γ.

A simple and semi-realistic model of a star comes from assuming that the fluid is incompressible: the density is a constant ρ_* out to the surface of the star, after which it vanishes,

$$\rho(r) = \begin{cases} \rho_*, & r < R \\ 0, & r > R. \end{cases} \tag{5.156}$$

Specifying $\rho(r)$ explicitly takes the place of an equation of state, since $p(r)$ can be determined from hydrostatic equilibrium. It is then straightforward to integrate (5.146) to get

$$m(r) = \begin{cases} \frac{4}{3}\pi r^3 \rho_*, & r < R \\ \frac{4}{3}\pi R^3 \rho_* = M, & r > R. \end{cases} \tag{5.157}$$

Integrating the equation of hydrostatic equilibrium yields

$$p(r) = \rho_* \left[\frac{R\sqrt{R - 2GM} - \sqrt{R^3 - 2GMr^2}}{\sqrt{R^3 - 2GMr^2} - 3R\sqrt{R - 2GM}} \right]. \tag{5.158}$$

Finally we can get the metric component $g_{tt} = -e^{2\alpha(r)}$ from (5.152); we find that

$$e^{\alpha(r)} = \frac{3}{2}\left(1 - \frac{2GM}{R}\right)^{1/2} - \frac{1}{2}\left(1 - \frac{2GMr^2}{R^3}\right)^{1/2}, \quad r < R. \tag{5.159}$$

The pressure increases near the core of the star, as one would expect. Indeed, for a star of fixed radius R, the central pressure $p(0)$ will need to be greater than infinity if the mass exceeds

$$M_{\text{max}} = \frac{4}{9G} R. \tag{5.160}$$

Thus, if we try to squeeze a greater mass than this inside a radius R, general relativity admits no static solutions; a star that shrinks to such a size must inevitably keep shrinking, eventually forming a black hole. We derived this result from the rather strong assumption that the density is constant, but it continues to hold when that assumption considerably weakened; **Buchdahl's theorem** states that any reasonable static, spherically symmetric interior solution has $M < 4R/9G$. Although a careful proof requires more work, this result makes sense; if we imagine that there is some maximum sustainable density in nature, the most massive object we could in principle make would have that density everywhere, which is the specific case we considered.

Of course, this still doesn't mean that realistic astrophysical objects will always ultimately collapse to black holes. An ordinary planet, supported by material pressures, will persist essentially forever (apart from some fantastically unlikely quantum tunneling from a planet to something very different, or the possibility of eventual proton decay). But massive stars are a different story. The pressure supporting a star comes from the heat produced by fusion of light nuclei into heavier ones. When the nuclear fuel is used up, the temperature declines and the star begins to shrink under the influence of gravity. The collapse may eventually be halted by Fermi degeneracy pressure: Electrons are pushed so close together that they resist further compression simply on the basis of the Pauli exclusion principle (no two fermions can be in the same state). A stellar remnant supported by electron degeneracy pressure is called a **white dwarf**; a typical white dwarf is comparable in size to the Earth. Lower-mass particles become degenerate at lower number densities than high-mass particles, so nucleons do not contribute appreciably to the pressure in a white dwarf. White dwarfs are the end state for most stars, and are extremely common throughout the universe.

If the total mass is sufficiently high, however, the star will reach the **Chandrasekhar limit**, where even the electron degeneracy pressure is not enough to resist the pull of gravity. Calculations put the Chandrasekhar limit at about $1.4\ M_\odot$, where $M_\odot = 2 \times 10^{33}$ g is the mass of the Sun. When it is reached, the star is forced to collapse to an even smaller radius. At this point electrons combine with protons to make neutrons and neutrinos (inverse beta decay), and the neutrinos simply fly away. The result is a **neutron star**, with a typical radius of about 10 km. Neutron stars have a low total luminosity, but often are rapidly spinning and possess strong magnetic fields. This combination gives rise to **pulsars**, which accelerate particles in jets emanating from the magnetic poles, appearing to rapidly flash as the neutron star spins. Pulsars were discovered by Bell in 1967; after a brief speculation that they might represent signals from an extraterrestrial civilization, the more prosaic astrophysical explanation was settled on.

Since the conditions at the center of a neutron star are very different from those on Earth, we do not have a perfect understanding of the equation of state. Nevertheless, we believe that a sufficiently massive neutron star will itself be unable to resist the pull of gravity, and will continue to collapse; current estimates of the maximum possible neutron-star mass are around 3–4 $M_\odot$, the **Oppenheimer-Volkoff limit**. Since a fluid of neutrons is the densest material we know about (apart from some very speculative suggestions), it is believed that the outcome of such a collapse is a black hole.

How would we know if there were a black hole? The fundamental obstacle to direct detection is, of course, blackness: a black hole will not itself give off any radiation (neglecting Hawking radiation, which is a very small effect to be discussed in Chapter 9). But black holes will feature extremely strong gravitational fields, so we can hope to detect them indirectly by observing matter being influenced by these fields. As matter falls into a black hole, it will heat up and emit X-rays, which we can detect with satellite observatories. A large number of black-hole candidates have been detected by this method, and the case for real black holes in our universe is extremely strong.[1] The large majority of candidates fall into one of two classes. There are black holes with masses of order a solar mass or somewhat higher; these are thought to be the endpoints of evolution for very massive stars. The other category describes supermassive black holes, be-

[1]For a review on astrophysical evidence for black holes, see A. Celotti, J.C. Miller, and D.W. Sciama (1999), *Class. Quant. Grav.* **16**, A3; `http://arxiv.org/abs/astro-ph/9912186`.

tween 10^6 and 10^9 solar masses. These are found at the centers of galaxies, and are thought to be the engines that powered quasars in the early era of galaxy formation. Our own Milky Way galaxy contains an object (Sgr A*) that is believed to be a black hole of at least $2 \times 10^6 M_{\odot}$. The precise history of the formation of these supermassive holes is not well understood. Other possibilities include very small primordial black holes produced in the very early universe, and so-called "middleweight" black holes of order a thousand solar masses.

As matter falls into a black hole, it tends to settle into a rotating accretion disk, and both energy and angular momentum are gradually fed into the hole. As a result, the black holes we expect to see in astrophysical situations should be spinning, and indeed observations are consistent with very high spin rates for observed black holes. In this chapter we have excluded the possibility of black hole spin by focusing on the spherically symmetric Schwarzschild solution; in the next chapter we turn to more general types of black holes.

5.9 ■ EXERCISES

1. A space monkey is happily orbiting a Schwarzschild black hole in a circular geodesic orbit. An evil baboon, far from the black hole, tries to send the monkey to its death inside the black hole by dropping a carefully timed coconut radially toward the black hole, knowing that the monkey can't resist catching the falling coconut. Given the monkey's mass and initial orbital radius and the mass of the coconut, explain how you would go about solving the problem (but do not do the calculation). What are the possible fates for our intrepid space monkey?

2. Consider a perfect fluid in a static, circularly symmetric $(2+1)$-dimensional spacetime, equivalently, a cylindrical configuration in $(3+1)$ dimensions with perfect rotational symmetry.

   (a) Derive the analogue of the Tolman–Oppenhiemer–Volkov (TOV) equation for $(2+1)$ dimensions.

   (b) Show that the vacuum solution can be written as

   $$ds^2 = -dt^2 + \frac{1}{1-8GM}dr^2 + r^2 d\theta^2$$

   Here M is a constant.

   (c) Show that another way to write the same solution is

   $$ds^2 = -d\tau^2 + d\xi^2 + \xi^2 d\phi^2$$

   where $\phi \in [0, 2\pi(1-8GM)^{1/2}]$.

   (d) Solve the $(2+1)$ TOV equation for a constant density star. Find $p(r)$ and solve for the metric.

   (e) Solve the $(2+1)$ TOV equation for a star with equation of state $p = \kappa\rho^{3/2}$. Find $p(r)$ and solve for the metric.

   (f) Find the mass $M(R) = \int_0^{2\pi}\int_0^R \rho\, dr\, d\theta$ and the proper mass $\bar{M}(R) = \int_0^{2\pi}\int_0^R \rho\sqrt{-g}\, dr\, d\theta$ for the solutions in parts (d) and (e).

3. Consider a particle (not necessarily on a geodesic) that has fallen inside the event horizon, $r < 2GM$. Use the ordinary Schwarzschild coordinates (t, r, θ, ϕ). Show that the radial coordinate must decrease at a minimum rate given by

$$\left|\frac{dr}{d\tau}\right| \geq \sqrt{\frac{2GM}{r} - 1}.$$

Calculate the maximum lifetime for a particle along a trajectory from $r = 2GM$ to $r = 0$. Express this in seconds for a black hole with mass measured in solar masses. Show that this maximum proper time is achieved by falling freely with $E \rightarrow 0$.

4. Consider Einstein's equations in vacuum, but with a cosmological constant, $G_{\mu\nu} + \Lambda g_{\mu\nu} = 0$.

(a) Solve for the most general spherically symmetric metric, in coordinates (t, r) that reduce to the ordinary Schwarzschild coordinates when $\Lambda = 0$.

(b) Write down the equation of motion for radial geodesics in terms of an effective potential, as in (5.66). Sketch the effective potential for massive particles.

5. Consider a comoving observer sitting at constant spatial coordinates (r_*, θ_*, ϕ_*), around a Schwarzschild black hole of mass M. The observer drops a beacon into the black hole (straight down, along a radial trajectory). The beacon emits radiation at a constant wavelength $\lambda_{\rm em}$ (in the beacon rest frame).

(a) Calculate the coordinate speed dr/dt of the beacon, as a function of r.

(b) Calculate the proper speed of the beacon. That is, imagine there is a comoving observer at fixed r, with a locally inertial coordinate system set up as the beacon passes by, and calculate the speed as measured by the comoving observer. What is it at $r = 2GM$?

(c) Calculate the wavelength $\lambda_{\rm obs}$, measured by the observer at r_*, as a function of the radius $r_{\rm em}$ at which the radiation was emitted.

(d) Calculate the time $t_{\rm obs}$ at which a beam emitted by the beacon at radius $r_{\rm em}$ will be observed at r_*.

(e) Show that at late times, the redshift grows exponentially: $\lambda_{\rm obs}/\lambda_{\rm em} \propto e^{t_{\rm obs}/T}$. Give an expression for the time constant T in terms of the black hole mass M.

CHAPTER 6

More General Black Holes

6.1 ■ THE BLACK HOLE ZOO

Birkhoff's theorem ensures that the Schwarzschild metric is the only spherically symmetric vacuum solution to general relativity. This shouldn't be too surprising, as it is reminiscent of the situation in electromagnetism, where the only spherically symmetric field configuration in a region free of charges will be a Coulomb field. Moving beyond spherical symmetry, there is an unlimited variety of possible gravitational fields. For a planet like the Earth, for example, the external field will depend on the density and profile of all the various mountain ranges and valleys on the surface. We could imagine decomposing the metric into multipole moments, and an infinite number of coefficients would have to be specified to describe the field exactly.

It might therefore come as something of a surprise that black holes do not share this property. Only a small number of stationary black-hole solutions exist, described by a small number of parameters. The specific set of parameters will depend on what matter fields we include in our theory; if electromagnetism is the only long-range nongravitational field, we have a **no-hair theorem**:

> Stationary, asymptotically flat black hole solutions to general relativity coupled to electromagnetism that are nonsingular outside the event horizon are fully characterized by the parameters of mass, electric and magnetic charge, and angular momentum.

Stationary solutions are of special interest because we expect them to be the end states of gravitational collapse. The alternative might be some sort of oscillating configuration, but oscillations will ultimately be damped as energy is lost through the emission of gravitational radiation; in fact, typical evolutions will evolve quite rapidly to a stationary configuration.

We speak of "a" no-hair theorem, rather than "the" no-hair theorem, because the result depends not only on general relativity, but also on the matter content of our theory. In the Standard Model of particle physics, electromagnetism is the only long-range field, and the above theorem applies; but for different kinds of fields there might be other sorts of hair.[1] Examples have even been found of static (nonrotating) black holes that are axisymmetric but not completely spherically

[1]For a discussion see M. Heusler, "Stationary Black Holes: Uniqueness and Beyond," Living Rev. Relativity **1**, (1998), 6; `http://www.livingreviews.org/Articles/Volume1/1998-6heusler/`.

symmetric. The central point, however, remains unaltered: black hole solutions are characterized by a very small number of parameters, rather than the potentially infinite set of parameters characterizing, say, a planet.

As we will discuss at the end of this chapter and again in Chapter 9, the no-hair property leads to a puzzling situation. In most physical theories, we hope to have a well-defined initial value problem, so that information about a state at any one moment of time can be used to predict (or retrodict) the state at any other moment of time. As a consequence, any two states that are connected by a solution to the equations of motion should require the same amount of information to be specified. But in GR, it seems, we can take a very complicated collection of matter, collapse it into a black hole, and end up with a configuration described completely by mass, charge, and spin. In classical GR this might not bother us so much, since the information can be thought of as hidden behind the event horizon rather than truly being lost. But when quantum field theory is taken into account, we find that black holes evaporate and eventually disappear, and the information seems to be truly lost. Conceivably, the outgoing Hawking radiation responsible for the evaporation somehow encodes information about what state was originally used to make a black hole, but how that could happen is completely unclear. Understanding this "information loss paradox" is considered by many to be a crucial step in building a sensible theory of quantum gravity.

In this chapter, however, we will stick to considerations of classical GR. We begin with some general discussion of black hole properties, especially those of event horizons and Killing horizons. This subject can be subtle and technical, and our philosophy here will be to try to convey the main ideas without being rigorous about definitions or proofs of theorems. We then discuss the specific solutions corresponding to charged (Reissner–Nordström) and spinning (Kerr) black holes; consistent with our approach, we will not carefully go through the coordinate redefinitions necessary to construct the maximally extended spacetimes, but instead simply draw the associated conformal diagrams. The reader interested in further details should consult the review article by Townsend,[2] or the books by Hawking and Ellis (1973) and Wald (1984), all of which we draw on heavily in this chapter.

6.2 ■ EVENT HORIZONS

Black holes are characterized by the fact that you can enter them, but never exit. Thus, their most important feature is actually not the singularity at the center, but the event horizon at the boundary. An event horizon is a hypersurface separating those spacetime points that are connected to infinity by a timelike path from those that are not. To understand what this means in practice, we should think a little more carefully about what we mean by "infinity." In general relativity, the global structure of spacetime can take many different forms, with correspondingly different notions of infinity. But to think about black holes in the real universe, we

[2] P.K. Townsend, "Black Holes: Lecture Notes," `http://arxiv.org/abs/gr-qc/9707012`.

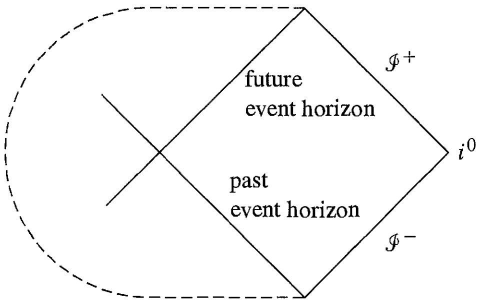

FIGURE 6.1 An asymptotically flat spacetime is one for which infinity in a conformal diagram matches that of Minkowski spacetime, with future null infinity $\mathscr{I}^+$, spacelike infinity i^0, and past null infinity $\mathscr{I}^-$. The future event horizon is the boundary of the past of $\mathscr{I}^+$. The dashed region represents the rest of the spacetime, which may take a number of different forms in different examples.

aren't actually concerned with what happens infinitely far away; we use infinity as a proxy for "well outside the black hole," and imagine that spacetime sufficiently far away from the hole can be approximated by Minkowski space.

As mentioned in the Chapter 5, a spacetime that looks Minkowskian at infinity is referred to as asymptotically flat. The meaning of this concept is made clear in a conformal diagram such as in Figure 6.1. From our discussion in Appendix H of the conformal diagram for Minkowski, we know that conformal infinity comes in five pieces: future and past timelike infinity $i^\pm$, future and past null infinity $\mathscr{I}^\pm$, and spatial infinity i^0. An asymptotically flat spacetime (or region of spacetime) is one for which $\mathscr{I}^\pm$ and i^0 have the same structure as for Minkowski; timelike infinity is not necessary. Such spacetimes will have the general form shown in Figure 6.1.

With this picture, it is clear how we should think of the future event horizon: it is the surface beyond which timelike curves cannot escape to infinity. Recalling that the causal past J^- of a region is the set of all points we can reach from that region by moving along past-directed timelike paths, the event horizon can be equivalently defined as the boundary of $J^-(\mathscr{I}^+)$, the causal past of future null infinity. (The event horizon is really the boundary of the *closure* of this set, but we're not being rigorous.) Analogous definitions hold for the past horizon. As we have seen in the case of maximally extended Schwarzschild, there may be more than one asymptotically flat region in a spacetime, and correspondingly more than one event horizon.

From the definition, it is clear that the event horizon is a null hypersurface. Properties of null hypersurfaces are discussed in Appendix D; here we can recall the major features. A hypersurface Σ can be defined by $f(x) =$ constant for some function $f(x)$. The gradient $\partial_\mu f$ is normal to Σ; if the normal vector is null, the hypersurface is said to be null, and the normal vector is also tangent to Σ. Null hypersurfaces can be thought of as a collection of null geodesics $x^\mu(\lambda)$, called the generators of the hypersurface. The tangent vectors ξ^μ to these geodesics are proportional to the normal vectors,

$$\xi^{\mu} = \frac{dx^{\mu}}{d\lambda} = h(x) g^{\mu\nu} \partial_{\nu} f, \tag{6.1}$$

and therefore also serve as normal vectors to the hypersurface. We may choose the function $h(x)$ so that the geodesics are affinely parameterized, so the tangent vectors will obey

$$\xi_{\mu}\xi^{\mu} = 0, \quad \xi^{\mu}\nabla_{\mu}\xi^{\nu} = 0. \tag{6.2}$$

For future event horizons, the generators may end in the past (for example, when a black hole is formed by stellar collapse) but will always continue indefinitely into the future (and similarly with future and past interchanged).

Because the event horizon is a global concept, it might be difficult to actually locate one when you are handed a metric in an arbitrary set of coordinates. Fortunately, in this chapter we will be concerned with quite special metrics—stationary, asymptotically flat, and containing event horizons with spherical topology. In such spacetimes, there are convenient coordinate systems in which there is a simple way to identify the event horizon. For the Schwarzschild solution, the event horizon is a place where the light cones "tilt over" so that $r = 2GM$ is a null surface rather than a timelike surface, as $r =$ constant would be for large r. Light-cone tilting is clearly a coordinate-dependent notion (it doesn't happen, for example, in Kruskal coordinates), but the metrics of concern to us will allow for analogous constructions. A stationary metric has a Killing vector ∂_t that is asymptotically timelike, and we can adapt the metric components to be time-independent ($\partial_t g_{\mu\nu} = 0$). On hypersurfaces $t =$ constant, we can choose coordinates (r, θ, ϕ) in which the metric at infinity looks like Minkowski space in spherical polar coordinates. Hypersurfaces $r =$ constant will be timelike cylinders with topology $S^2 \times \mathbf{R}$ at $r \to \infty$. Now imagine we have chosen our coordinates cleverly, so that as we decrease r from infinity the $r =$ constant hypersurfaces remain timelike until some fixed $r = r_{\mathrm{H}}$, for which the surface is everywhere null. (In nonclever coordinates, $r =$ constant hypersurfaces will become null or spacelike for some values of θ and ϕ but remain timelike for others.) This will clearly represent an event horizon, since timelike paths crossing into the region $r < r_{\mathrm{H}}$ will never be able to escape back to infinity. Determining the point at which $r =$ constant hypersurfaces become null is easy; $\partial_{\mu} r$ is a one-form normal to such hypersurfaces, with norm

$$g^{\mu\nu}(\partial_{\mu} r)(\partial_{\nu} r) = g^{rr}. \tag{6.3}$$

We are looking for the place where the norm of our one-form vanishes; hence, in the coordinates we have described, the event horizon $r = r_{\mathrm{H}}$ will simply be the hypersurface at which g^{rr} switches from being positive to negative,

$$g^{rr}(r_{\mathrm{H}}) = 0. \tag{6.4}$$

This criterion clearly works for Schwarzschild, for which $g^{rr} = 1 - 2GM/r$. We will present the Reissner–Nordström and Kerr solutions in coordinates that are similarly adapted to the horizons.

The reason why we make such a big deal about event horizons is that they are nearly inevitable in general relativity. This conclusion is reached by concatenating two interesting results: Singularities are nearly inevitable, and singularities are hidden behind event horizons. Of course both results hold under appropriate sets of assumptions; it is not that hard to come up with spacetimes that have no singularities (Minkowski would be an example), nor is it even that hard to find singularities without horizons (as we will see below in our discussion of charged black holes). But we believe that "generic" solutions will have singularities hidden behind horizons.

The ubiquity of singularities is guaranteed by the **singularity theorems** of Hawking and Penrose. Before these theorems were proven, it was possible to hope that collapse to a Schwarzschild singularity was an artifact of spherical symmetry, and typical geometries would remain nonsingular (as happens, for example, in Newtonian gravity). But the Hawking–Penrose theorems demonstrate that once collapse reaches a certain point, evolution to a singularity is inevitable. The way we know there is a singularity is through geodesic incompleteness—there exists some geodesic that cannot be extended within the manifold, but nevertheless ends at a finite value of the affine parameter. The way we know collapse has reached a point of no return is the appearance of a **trapped surface**. To understand what a trapped surface is, first picture a two-sphere in Minkowski space, taken as a set of points some fixed radial distance from the origin, embedded in a constant-time slice. If we follow null rays emanating into spacetime from this spatial sphere, one set (pointed inward) will describe a shrinking set of spheres, while the other (pointed outward) will describe a growing set of spheres. But this would not be the case for a sphere of fixed radius $r < 2GM$ in the Schwarzschild geometry; inside the event horizon, both sets of null rays emanating from such a sphere would evolve to smaller values of r (since r is a timelike coordinate), and thus to smaller areas $4\pi r^2$. This is what is meant by a trapped surface: a compact, spacelike, two-dimensional submanifold with the property that outgoing future-directed light rays *converge* in both directions everywhere on the submanifold. (The formal definition of "converge" is that the expansion θ, as described in the discussion of geodesic congruences in Appendix F, is negative.)

With these definitions in hand, we can present an example of a singularity theorem.

> Let M be a manifold with a generic metric $g_{\mu\nu}$, satisfying Einstein's equation with the strong energy condition imposed. If there is a trapped surface in M, there must be either a closed timelike curve or a singularity (as manifested by an incomplete timelike or null geodesic).

In this case, by "a generic metric" we mean that the **generic condition** is satisfied for both timelike and null geodesics. For timelike geodesics, the generic condition states that every geodesic with tangent vector U^μ must have at least one point on which $R_{\alpha\beta\gamma\delta}U^\alpha U^\delta \neq 0$; for null geodesics, the generic condition states that every geodesic with tangent vector k^μ must have at least one point on which $k_{[\alpha}R_{\beta]\gamma\delta[\epsilon}k_{\zeta]}k^\gamma k^\delta \neq 0$. These fancy conditions simply serve to exclude very special metrics for which the curvature consistently vanishes in some directions.

Singularity theorems exist in many forms, proceeding from various different sets of assumptions. The moral of the story seems to be that typical time-dependent solutions in general relativity usually end in singularities. (Or begin in them; some theorems imply the existence of cosmological singularities, such as the Big Bang.) This represents something of a problem for GR, in the sense that the theory doesn't really apply to the singularities themselves, whose existence therefore represents an incompleteness of description. The traditional attitude toward this issue is to hope that a sought-after quantum theory of gravity will somehow resolve the singularities of classical GR.

In the meantime, we can take solace in the idea that singularities are hidden behind event horizons. This belief is encompassed in the **cosmic censorship conjecture**:

> Naked singularities cannot form in gravitational collapse from generic, initially nonsingular states in an asymptotically flat spacetime obeying the dominant energy condition.

A **naked singularity** is one from which signals can reach $\mathscr{I}^+$; that is, one that is not hidden behind an event horizon. Notice that the conjecture refers to the formation of naked singularities, not their existence; there are certainly solutions in which spacelike naked singularities exist in the past (such as the Schwarzschild white hole) or timelike singularities exist for all times (such as in super-extremal charged black holes, discussed below). The cosmic censorship conjecture has not been proven, although a great deal of effort has gone into finding convincing counterexamples, without success. The requirement that the initial data be in some sense "generic" is important, as numerical experiments have shown that finely-tuned initial conditions are able to give rise to naked singularities. A precise proof of some form of the cosmic censorship conjecture remains one of the outstanding problems of classical general relativity.[3]

A consequence of cosmic censorship (or of certain equivalent assumptions) is that classical black holes never shrink, they only grow bigger. The size of a black hole is measured by the area of the event horizon, by which we mean the spatial area of the intersection of the event horizon with a spacelike slice. We then have Hawking's **area theorem**:

> Assuming the weak energy condition and cosmic censorship, the area of a future event horizon in an asymptotically flat spacetime is non-decreasing.

[3]For a review of cosmic censorship see R.M. Wald, "Gravitational Collapse and Cosmic Censorship," `http://arxiv.org/abs/gr-qc/9710068`.

For a Schwarzschild black hole, the area depends monotonically on the mass, so this theorem implies that Schwarzschild black holes can only increase in mass. But for spinning black holes this is no longer the case; the area depends on a combination of mass and angular momentum, and we will see below that we can actually extract energy from a black hole by decreasing its spin. We can also decrease the mass of a black hole through quantum-mechanical Hawking radiation; this can be traced to the fact that quantum field theory in curved spacetime can violate the weak energy condition.

6.3 ■ KILLING HORIZONS

In the Schwarzschild metric, the Killing vector $K = \partial_t$ goes from being timelike to spacelike at the event horizon. In general, if a Killing vector field χ^μ is null along some null hypersurface Σ, we say that Σ is a **Killing horizon** of χ^μ. Note that the vector field χ^μ will be normal to Σ, since a null surface cannot have two linearly independent null tangent vectors.

The notion of a Killing horizon is logically independent from that of an event horizon, but in spacetimes with time-translation symmetry the two are closely related. Under certain reasonable conditions (made explicit below), we have the following classification:

> Every event horizon Σ in a stationary, asymptotically flat spacetime is a Killing horizon for some Killing vector field χ^μ.
>
> If the spacetime is static, χ^μ will be the Killing vector field $K^\mu = (\partial_t)^\mu$ representing time translations at infinity.
>
> If the spacetime is stationary but not static, it will be axisymmetric with a rotational Killing vector field $R^\mu = (\partial_\phi)^\mu$, and χ^μ will be a linear combination $K^\mu + \Omega_{\rm H} R^\mu$ for some constant $\Omega_{\rm H}$.

For example, below we will examine the Kerr metric for spinning black holes, in which the event horizon is a Killing horizon for a linear combination of the Killing vectors for rotations and time translations. In Kerr, the hypersurface on which ∂_t becomes null is actually timelike, so is not a Killing horizon.

Let's be precise about the conditions under which this classification scheme actually holds.[4] Carter has shown that, for static black holes, the event horizon is a Killing horizon for K^μ; this is a purely geometric fact, which holds even without invoking Einstein's equation. In the stationary case, if we assume the existence of a rotational Killing field R^μ with the property that 2-planes spanned by K^μ and R^μ are orthogonal to a family of two-dimensional surfaces, then the event horizon will be a Killing horizon for a linear combination of the two Killing fields, again from purely geometric considerations. If on the other hand we only assume that the black hole is stationary, we cannot prove in general that the event horizon

[4]For a discussion see R. M. Wald, "The thermodynamics of black holes," *Living Rev. Rel.* **4**, 6 (2001), `http://arxiv.org/gr-qc/9912119`.

is axisymmetric. Given Einstein's equation and some conditions on the matter fields, Hawking was able to show that the event horizon of any stationary black hole must be a Killing horizon for some vector field, and furthermore that such horizons must either be stationary or axisymmetric. For the rest of this chapter we will speak as if the above classification holds; however, making assumptions about matter fields is notoriously tricky, and we should keep in mind the possibility in principle of finding black holes that are not static or axisymmetric, for which the event horizon might not be a Killing horizon.

It's important to point out that, while event horizons for stationary asymptotically flat spacetimes will typically be Killing horizons, it's easy to have Killing horizons that have nothing to do with event horizons. Consider Minkowski space in inertial coördinates, $ds^2 = -dt^2 + dx^2 + dy^2 + dz^2$; clearly there are no event horizons in this spacetime. The Killing vector that generates boosts in the x-direction is

$$\chi = x\partial_t + t\partial_x, \tag{6.5}$$

with norm

$$\chi_\mu \chi^\mu = -x^2 + t^2. \tag{6.6}$$

This goes null at the null surfaces

$$x = \pm t, \tag{6.7}$$

which are therefore Killing horizons. By combining the boost Killing vector with translational and rotational Killing vectors, we can move these horizons through the manifold; there are Killing horizons all over. In more interesting spacetimes, of course, there will be fewer Killing vector fields, and the associated horizons (if any) will have greater physical significance.

To every Killing horizon we can associate a quantity called the **surface gravity**. Consider a Killing vector χ^μ with Killing horizon Σ. Because χ^μ is a normal vector to Σ, along the Killing horizon it obeys the geodesic equation,

$$\chi^\mu \nabla_\mu \chi^\nu = -\kappa \chi^\nu, \tag{6.8}$$

where the right-hand side arises because the integral curves of χ^μ may not be affinely parameterized. The parameter κ is the surface gravity; it will be constant over the horizon, except for a "bifurcation two-sphere" where the Killing vector vanishes and κ can change sign. (This happens, for example, at the center of the Kruskal diagram in the Schwarzschild solution.) Using Killing's equation $\nabla_{(\mu}\chi_{\nu)} = 0$ and the fact that $\chi_{[\mu}\nabla_\nu \chi_{\sigma]} = 0$ (since χ^μ is normal to Σ), it is straightforward to derive a nice formula for the surface gravity:

$$\kappa^2 = -\tfrac{1}{2}(\nabla_\mu \chi_\nu)(\nabla^\mu \chi^\nu). \tag{6.9}$$

The expression on the right-hand side is to be evaluated at the horizon Σ. You are encouraged to check this formula yourself.

The surface gravity associated with a Killing horizon is in principle arbitrary, since we can always scale a Killing field by a real constant and obtain another Killing field. In a static, asymptotically flat spacetime, the time-translation Killing vector $K = \partial_t$ can be normalized by setting

$$K_\mu K^\mu(r \to \infty) = -1. \tag{6.10}$$

This in turn fixes the surface gravity of any associated Killing horizon. If we are in a stationary spacetime, where the Killing horizon is associated with a linear combination of time translations and rotations, fixing the normalization of $K = \partial_t$ also fixes this linear combination, so the surface gravity remains unique.

The reason why κ is called the "surface gravity" becomes clear only when the spacetime is static. In that case we have the following interpretation:

> In a static, asymptotically flat spacetime, the surface gravity is the acceleration of a static observer near the horizon, as measured by a static observer at infinity.

To make sense of such a statement, let's first consider static observers. By a static observer we mean one whose four-velocity U^μ is proportional to the time-translation Killing field K^μ:

$$K^\mu = V(x)U^\mu. \tag{6.11}$$

Since the four-velocity is normalized to $U_\mu U^\mu = -1$, the function V is simply the magnitude of the Killing field,

$$V = \sqrt{-K_\mu K^\mu}, \tag{6.12}$$

and hence ranges from zero at the Killing horizon to unity at infinity. V is sometimes called the "redshift factor," since it relates the emitted and observed frequencies of a photon as measured by static observers. Recall that the conserved energy of a photon with four-momentum p^μ is $E = -p_\mu K^\mu$, while the frequency measured by an observer with four-velocity U^μ will be $\omega = -p_\mu U^\mu$. Therefore

$$\omega = \frac{E}{V}, \tag{6.13}$$

and a photon emitted by static observer 1 will be observed by static observer 2 to have wavelength $\lambda = 2\pi/\omega$ given by

$$\lambda_2 = \frac{V_2}{V_1}\lambda_1. \tag{6.14}$$

In particular, at infinity where $V = 1$, we will observe a wavelength $\lambda_\infty = \lambda_1/V_1$.

Now we turn to the idea of "acceleration as viewed from infinity." A static observer will not typically be moving on a geodesic; for example, particles tend to fall into black holes rather than hovering next to them at fixed spatial coordinates.

We can express the four-acceleration $a^\mu = U^\sigma \nabla_\sigma U^\mu$ in terms of the redshift factor as

$$a_\mu = \nabla_\mu \ln V, \tag{6.15}$$

as you can easily check. The magnitude of the acceleration,

$$a = \sqrt{a_\mu a^\mu} = V^{-1}\sqrt{\nabla_\mu V \nabla^\mu V}, \tag{6.16}$$

will go to infinity at the Killing horizon—it will take an infinite acceleration to keep an object on a static trajectory. But an observer at infinity will detect the acceleration to be "redshifted" by a factor V; this turns out to be the surface gravity. Thus, we claim that

$$\kappa = Va = \sqrt{\nabla_\mu V \nabla^\mu V}, \tag{6.17}$$

evaluated at the horizon Σ. You can check that this expression agrees with (6.9). The surface gravity is the product of zero (V) and infinity (a), but will typically be finite. When we say that the observed acceleration is redshifted, we have in mind stretching a test string from a static object at the horizon to an observer at infinity, and measuring the acceleration on the end of the string at infinity. (It is worth taking the time to see if you can promote this hand-waving argument to something more rigorous.)

What goes wrong with the above considerations if the spacetime is stationary but not static? We still have an asymptotically time-translation Killing vector $K = \partial_t$, and we can define stationary observers as ones whose four-velocities are parallel to K^μ, as in (6.11); the redshift will continue to be given by (6.14). The problem is that K^μ won't become null at a Killing horizon, but generally at some timelike surface outside the horizon. This place where $K^\mu K_\mu = 0$ is called the **stationary limit surface** (or sometimes "ergosurface"), since inside this surface K^μ is spacelike, and consequently no observer can remain stationary, even if it is still outside the event horizon. Such an observer has to move with respect to the Killing field, but need not move in the direction of the black hole. From (6.12) and (6.14), the redshift of a stationary observer diverges as we approach the stationary limit surface, which is therefore also called the **infinite redshift surface**. As we will see in our discussion of the Kerr metric, the region between the stationary limit surface and the event horizon, the ergosphere, is a place where timelike paths are inevitably dragged along with the rotation of the black hole. We will continue to use "surface gravity" as a label in stationary spacetimes, which we will calculate using the Killing vector χ^μ, which actually does go null on the event horizon, even if the resulting quantity cannot be interpreted as the gravitational acceleration of a stationary observer as seen at infinity.

Let's apply these notions to Schwarzschild to see how they work. For the metric

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\Omega^2, \tag{6.18}$$

the Killing vector and static four-velocity are

$$K^\mu = (1,\ 0,\ 0,\ 0), \quad U^\mu = \left[\left(1 - \frac{2GM}{r}\right)^{-1/2},\ 0,\ 0,\ 0\right], \tag{6.19}$$

so the redshift factor is

$$V = \sqrt{1 - \frac{2GM}{r}}. \tag{6.20}$$

(Note the agreement with our calculation of the redshift in the previous chapter.) From (6.15), the acceleration is

$$a_\mu = \frac{GM}{r^2\left(1 - \frac{2GM}{r}\right)}\nabla_\mu r, \tag{6.21}$$

where of course $\nabla_\mu r = \delta^r_\mu$. The magnitude of the acceleration is thus

$$a = \frac{GM}{r^2\left(1 - \frac{2GM}{r}\right)^{1/2}}. \tag{6.22}$$

The surface gravity is $\kappa = Va$ evaluated at the event horizon $r = 2GM$, and

$$Va = \frac{GM}{r^2}, \tag{6.23}$$

so the surface gravity of a Schwarzschild black hole is

$$\kappa = \frac{1}{4GM}. \tag{6.24}$$

It might seem surprising that the surface gravity decreases as the mass increases, but a glance at (6.23) reveals what is going on; at fixed radius increasing M acts to increase the combination Va, but increasing the mass also increases the Schwarzschild radius, and that effect wins out. Thus, the surface gravity of a big black hole is actually weaker than that of a small black hole; this is consistent with an examination of the tidal forces, which are also smaller for bigger black holes.

6.4 ■ MASS, CHARGE, AND SPIN

Since we have claimed above that the most general stationary black-hole solution to general relativity is characterized by mass, charge, and spin, we should consider how these quantities might be defined in GR. Charge is the easiest to consider, so we start there; more details are found in our discussion of Stokes's theorem in Appendix E. We'll look specifically at electric charge, although magnetic charge could be examined in the same way.

Maxwell's equations relate the electromagnetic field strength tensor $F_{\mu\nu}$ to the electric current four-vector J_e^μ,

$$\nabla_\nu F^{\mu\nu} = J_e^\mu. \tag{6.25}$$

The charge passing through a spacelike hypersurface Σ is given by an integral over coordinates x^i on the hypersurface,

$$\begin{aligned} Q &= -\int_\Sigma d^3x\sqrt{\gamma}\, n_\mu J_e^\mu \\ &= -\int_\Sigma d^3x\sqrt{\gamma}\, n_\mu \nabla_\nu F^{\mu\nu}, \end{aligned} \tag{6.26}$$

where γ_{ij} is the induced metric, and n^μ is the unit normal vector, associated with Σ. The minus sign ensures that a positive charge density and a future-pointing normal vector will give a positive total charge. Stokes's theorem can then be used to express the charge as a boundary integral,

$$Q = -\int_{\partial\Sigma} d^2x\sqrt{\gamma^{(2)}}\, n_\mu \sigma_\nu F^{\mu\nu}, \tag{6.27}$$

where the boundary $\partial\Sigma$, typically a two-sphere at spatial infinity, has metric $\gamma_{ij}^{(2)}$ and outward-pointing normal vector σ^μ. The magnetic charge could be determined by replacing $F^{\mu\nu}$ with the dual tensor $*F^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$. Thus, to calculate the total charge, we need know only the behavior of the electromagnetic field at spatial infinity. In Appendix E we do an explicit calculation for a point charge in Minkowski space, which yields a predictable result but serves as a good check that our conventions work out correctly.

We turn now to the concept of the total energy (or mass) of an asymptotically flat spacetime. This is a much trickier notion than that of the charge; for one thing, energy-momentum is a tensor rather than a vector in general relativity, and for another, the energy-momentum tensor $T_{\mu\nu}$ only describes the properties of matter, not of the gravitational field. But recall that in Chapter 3 we discussed how we could nevertheless define a conserved total energy if spacetime were stationary, with a timelike Killing vector field K^μ. We first construct a current

$$J_T^\mu = K_\nu T^{\mu\nu}, \tag{6.28}$$

where $T^{\mu\nu}$ is the energy-momentum tensor. Because this current is divergenceless (from Killing's equation and conservation of $T^{\mu\nu}$), we can find a conserved energy by integrating over a spacelike surface Σ,

$$E_T = \int_\Sigma d^3x\sqrt{\gamma}\, n_\mu J_T^\mu, \tag{6.29}$$

just as for the charge. As interesting as this expression is, there are clearly some inadequacies with it. For example, consider the Schwarzschild metric. It has a

Killing vector, but $T^{\mu\nu}$ vanishes everywhere. Is the energy of a Schwarzschild black hole therefore zero? On both physical and mathematical grounds, there is reason to suspect not; there is a singularity, after all, which renders the integral difficult to evaluate. Furthermore, a Schwarzschild black hole can evolve from a massive star with a definite nonzero energy, and we might like that energy to be conserved. It is worth searching for an alternative definition of energy that better captures our intuitive picture for black hole spacetimes.

Sticking for the moment to spacetimes with a timelike Killing vector K^μ, consider a new current

$$J_R^\mu = K_\nu R^{\mu\nu}. \tag{6.30}$$

Using Einstein's equation, we can equivalently write this as

$$J_R^\mu = 8\pi G K_\nu \left(T^{\mu\nu} - \tfrac{1}{2} T g^{\mu\nu} \right). \tag{6.31}$$

The Ricci tensor is not divergenceless; instead we have the contracted Bianchi identity,

$$\nabla_\mu R^{\mu\nu} = \tfrac{1}{2} \nabla^\nu R. \tag{6.32}$$

But this and Killing's equation suffice to guarantee that our new current is conserved. To see this, we simply compute

$$\nabla_\mu J_R^\mu = (\nabla_\mu K_\nu) R^{\mu\nu} + K_\nu (\nabla_\mu R^{\mu\nu}). \tag{6.33}$$

The first term vanishes because $R^{\mu\nu}$ is symmetric and $\nabla_\mu K_\nu$ is antisymmetric (from Killing's equation). Using (6.32) we therefore have

$$\nabla_\mu J_R^\mu = \tfrac{1}{2} K_\nu \nabla^\nu R = 0, \tag{6.34}$$

which we know vanishes because the directional derivative of R vanishes along a Killing vector, (3.178).

As before, we can define a conserved energy associated with this current,

$$E_R = \frac{1}{4\pi G} \int_\Sigma d^3x \sqrt{\gamma}\, n_\mu J_R^\mu, \tag{6.35}$$

where the normalization is chosen for future convenience. The energy E_R will be independent of the spacelike hypersurface Σ, and hence conserved. This notion of energy has a significant advantage over E_T, arising from the fact that E_R can be rewritten as a surface integral over a two-sphere at spatial infinity. To see this, recall from (3.177) that any Killing vector satisfies $\nabla_\mu \nabla_\nu K^\mu = K^\mu R_{\mu\nu}$; the current itself can thus be written as a total derivative,

$$J_R^\mu = \nabla_\nu (\nabla^\mu K^\nu), \tag{6.36}$$

so that

$$E_R = \frac{1}{4\pi G}\int_\Sigma d^3x\sqrt{\gamma}\,n_\mu\nabla_\nu(\nabla^\mu K^\nu). \tag{6.37}$$

Note that, from raising indices on Killing's equation, $\nabla^\mu K^\nu = -\nabla^\nu K^\mu$. We can therefore again use Stokes's theorem just as we did for electric charge, to write E_R as an integral at spatial infinity,

$$\boxed{E_R = \frac{1}{4\pi G}\int_{\partial\Sigma} d^2x\sqrt{\gamma^{(2)}}\,n_\mu\sigma_\nu\nabla^\mu K^\nu.} \tag{6.38}$$

This expression is the **Komar integral** associated with the timelike Killing vector K^μ; it can be interpreted as the total energy of a stationary spacetime.

To convince ourselves that we're on the right track, let's calculate the Komar integral for Schwarzschild, with metric (6.18). The normal vectors, normalized to $n_\mu n^\mu = -1$ and $\sigma_\mu\sigma^\mu = +1$, have nonzero components

$$n_0 = -\left(1-\frac{2GM}{r}\right)^{1/2}, \quad \sigma_1 = \left(1-\frac{2GM}{r}\right)^{-1/2}, \tag{6.39}$$

with other components vanishing. We therefore have

$$n_\mu\sigma_\nu\nabla^\mu K^\nu = -\nabla^0 K^1. \tag{6.40}$$

The Killing vector is $K^\mu = (1, 0, 0, 0)$, so we can readily calculate

$$\begin{aligned}
\nabla^0 K^1 &= g^{00}\nabla_0 K^1 \\
&= g^{00}\left(\partial_0 K^1 + \Gamma^1_{0\lambda}K^\lambda\right) \\
&= g^{00}\Gamma^1_{00}K^0 \\
&= -\left(1-\frac{2GM}{r}\right)^{-1}\frac{GM}{r^2}\left(1-\frac{2GM}{r}\right) \\
&= -\frac{GM}{r^2}.
\end{aligned} \tag{6.41}$$

The metric on the two-sphere at infinity is

$$\gamma^{(2)}_{ij}\,dx^i dx^j = r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{6.42}$$

so that

$$\sqrt{\gamma^{(2)}} = r^2\sin\theta. \tag{6.43}$$

Putting it all together, the energy of a Schwarzschild black hole is

$$\begin{aligned} E_R &= \frac{1}{4\pi G}\int d\theta d\phi\, r^2 \sin\theta \left(\frac{GM}{r^2}\right) \\ &= M. \end{aligned} \tag{6.44}$$

This is of course the desired result, explaining the normalization chosen in (6.35).

Despite getting the right answer, we should think about what just happened. In particular, we obtained this energy by integrating the current $J^\mu_R = K_\nu R^{\mu\nu}$ over a spacelike slice, finding that the result could be written as an integral at spatial infinity. But for Schwarzschild, the metric solves the vacuum Einstein equation, $R_{\mu\nu} = 0$; it therefore seems difficult to get a nonzero answer from integrating J^μ_R, just as it did for (6.29). If we think about the structure of the maximally extended Schwarzschild solution, we realize that we could draw two kinds of spacelike slices: those that extend through the wormhole to the second asymptotic region, and those that end on the singularity. If the slice extends through the wormhole, the other asymptotic region provides another component to $\partial\Sigma$, and thus another contribution to (6.38); this contribution would exactly cancel, so the total energy would indeed be zero. If the slice intersected the singularity, we wouldn't know quite how to deal with it. Nevertheless, in either case it is sensible to treat our result (6.44) as the correct answer. The point is that, since (6.38) involves contributions only at spatial infinity, it should be a valid expression for the energy no matter what happens in the interior. We could even imagine time-dependent behavior in the interior; so long as K^μ was *asymptotically* a timelike Killing vector, the Komar energy will be well-defined. We could, for example, consider spherically symmetric gravitational collapse from an initially static star. Evaluating the integral (6.35) directly over Σ would give a sensible answer for the total mass, which should not change as the star collapsed to a black hole (we are imagining spherical symmetry, so that gravitational radiation cannot carry away energy to infinity). So the Komar integral (6.38), which would be valid before the collapse, may be safely interpreted as the energy even after collapse to a black hole. Of course for some purposes we might want to allow for energy loss through gravitational radiation, in which case we need to be careful about how we extend our slice to infinity; one can define a "Bondi mass" at future null infinity which allows us to keep track of energy loss through radiation.

Another worry about the Komar formula is whether it is really what we should think of as the "energy," which is typically the conserved quantity associated with time translation invariance. The best argument in favor of this interpretation is simply that E_R is certainly a conserved quantity of some sort, and it agrees with what we think should be the energy of Schwarzschild (and of a collection of masses in the Newtonian limit, as you could check), so what else could it be? Alternatively, one could think about a Hamiltonian formulation of general relativity, and carefully define the generator of time translations in an asymptotically flat spacetime, and then identify that with the total energy. This was first done by Arnowitt, Deser, and Misner, and their result is known as the **ADM energy**. In an asymptotically flat spacetime, we can write the metric just as we do in perturbation theory,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{6.45}$$

except that here we only ask that the components $h_{\mu\nu}$ be small at spatial infinity, not necessarily everywhere. The ADM energy can then be written as an integral over a two-sphere at spatial infinity, as

$$E_{\rm ADM} = \frac{1}{16\pi G}\int_{\partial\Sigma} d^2x\sqrt{\gamma^{(2)}}\,\sigma^i\left(\partial_j h^j{}_i - \partial_i h^j{}_j\right), \tag{6.46}$$

where spatial indices are raised with δ^{ij} (the spatial metric at infinity). This formula looks coordinate-dependent, but is actually well-defined under our assumptions. If $h_{\mu\nu}$ is time-independent at infinity, it can be verified that the ADM energy and the Komar energy actually agree. This gives us even more confidence that the Komar integral really represents the energy. However, there is a sense in which the ADM energy is more respectable; for example, the Komar integral can run into trouble if we have long-range scalar fields nonminimally coupled to gravity. But for our immediate purposes the Komar energy is quite acceptable.

One quality that we would like something called "energy" to have is that it be positive for any physical configuration; otherwise a zero-energy state could decay into pieces of positive energy and negative energy. The energy conditions discussed in Chapter 4 give a notion of positive energy for matter fields, but we might worry about a negative gravitational contribution leading to problems. Happily, in GR we have the **positive energy theorem**, first proven by Shoen and Yau:

> The ADM energy of a nonsingular, asymptotically flat spacetime obeying Einstein's equation and the dominant energy condition is nonnegative. Furthermore, Minkowski is the only such spacetime with vanishing ADM energy.

If we allow for singularities, there are clearly counterexamples, such as Schwarzschild with $M < 0$. However, if a spacetime with a singularity (such as Schwarzschild with $M > 0$) is reached as the evolution of nonsingular initial data, the theorem will apply. Thus we seem to be safe from negative-energy isolated systems in general relativity.

Finally, we may turn to spin (angular momentum), which is perfectly straightforward after our discussion of energy. Imagine that we have a rotational Killing vector $R = \partial_\phi$. In exact analogy with the time-translation case, we can define a conserved current

$$J^\mu_\phi = R_\nu R^{\mu\nu}, \tag{6.47}$$

which will lead to an expression for the conserved angular momentum J as an integral over spatial infinity,

$$\boxed{J = -\frac{1}{8\pi G}\int_{\partial\Sigma} d^2x\sqrt{\gamma^{(2)}}\, n_\mu \sigma_\nu \nabla^\mu R^\nu.} \tag{6.48}$$

(It is too bad that "J" is used for both the current and the angular momentum, just as it is too bad that "R" is both the rotational Killing vector and the Ricci tensor. But there are only so many letters to go around.) Just as with the energy, this expression will still be valid even if R^μ is only asymptotically a Killing vector. Note that the normalization is different than in the energy integral; it could be justified, for example, by evaluating the expression for slowly-moving masses with weak gravitational fields.

6.5 ■ CHARGED (REISSNER–NORDSTRÖM) BLACK HOLES

We turn now to the exact solutions representing electrically charged black holes. Such solutions are not extremely relevant to realistic astrophysical situations; in the real world, a highly-charged black hole would be quickly neutralized by interactions with matter in the vicinity of the hole. But charged holes nevertheless illustrate a number of important features of more general situations. In this case the full spherical symmetry of the problem is still present; we know therefore that we can write the metric as

$$ds^2 = -e^{2\alpha(r,t)}dt^2 + e^{2\beta(r,t)}dr^2 + r^2 d\Omega^2. \tag{6.49}$$

Now, however, we are no longer in vacuum, since the hole will have a nonzero electromagnetic field, which in turn acts as a source of energy-momentum. The energy-momentum tensor for electromagnetism is given by

$$T_{\mu\nu} = F_{\mu\rho}F_\nu{}^\rho - \tfrac{1}{4}g_{\mu\nu}F_{\rho\sigma}F^{\rho\sigma}, \tag{6.50}$$

where $F_{\mu\nu}$ is the electromagnetic field strength tensor. Since we have spherical symmetry, the most general field strength tensor will have components

$$\begin{aligned} F_{tr} &= f(r,t) = -F_{rt} \\ F_{\theta\phi} &= g(r,t)\sin\theta = -F_{\phi\theta}, \end{aligned} \tag{6.51}$$

where $f(r,t)$ and $g(r,t)$ are some functions to be determined by the field equations, and components not written are zero. F_{tr} corresponds to a radial electric field, while $F_{\theta\phi}$ corresponds to a radial magnetic field. For those of you wondering about the $\sin\theta$, recall that the thing that should be independent of θ and ϕ is the radial component of the magnetic field, $B^r = \epsilon^{01\mu\nu}F_{\mu\nu}$. For a spherically symmetric metric,

$$\epsilon^{\rho\sigma\mu\nu} = \frac{1}{\sqrt{-g}}\tilde{\epsilon}^{\rho\sigma\mu\nu}$$

is proportional to $(\sin\theta)^{-1}$, so we want a factor of $\sin\theta$ in $F_{\theta\phi}$.

The field equations in this case are both Einstein's equation and Maxwell's equations:

$$g^{\mu\nu}\nabla_\mu F_{\nu\sigma} = 0$$
$$\nabla_{[\mu}F_{\nu\rho]} = 0. \tag{6.52}$$

The two sets are coupled together, since the electromagnetic field strength tensor enters Einstein's equation through the energy-momentum tensor, while the metric enters explicitly into Maxwell's equations.

The difficulties are not insurmountable, however, and a procedure similar to the one we followed for the vacuum case leads to a solution for the charged case as well. We will not go through the steps explicitly, but merely quote the final answer. The solution is known as the **Reissner–Nordström (RN) metric**, and is given by

$$\boxed{ds^2 = -\Delta dt^2 + \Delta^{-1}dr^2 + r^2 d\Omega^2,} \tag{6.53}$$

where

$$\Delta = 1 - \frac{2GM}{r} + \frac{G(Q^2 + P^2)}{r^2}. \tag{6.54}$$

In this expression, M is once again interpreted as the mass of the hole; Q is the total electric charge, and P is the total magnetic charge. Isolated magnetic charges (monopoles) have never been observed in nature, but that doesn't stop us from writing down the metric that they would produce if they did exist. There are good theoretical reasons to think that monopoles may exist if forces are "grand unified" at very high energies, but they must be very heavy and extremely rare. Of course, a black hole could possibly have magnetic charge even if there aren't any monopoles. In fact, the electric and magnetic charges enter the metric in the same way, so we are not introducing any additional complications by keeping P in our expressions. Conservatives are welcome to set $P = 0$ if they like. The electromagnetic fields associated with this solution are given by

$$E_r = F_{rt} = \frac{Q}{r^2}$$
$$B_r = \frac{F_{\theta\phi}}{r^2 \sin\theta} = \frac{P}{r^2}. \tag{6.55}$$

The $1/r^2$ dependence of these fields is just what we are used to in flat space; of course, here we know that this depends on our precise choice of radial coordinate.

The RN metric has a true curvature singularity at $r = 0$, as could be checked by computing the curvature invariant scalar $R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$. The horizon structure, however, is more complicated than in Schwarzschild. In the discussion of event horizons above, we suggested that $g^{rr} = 0$ would be a useful diagnostic for locat-

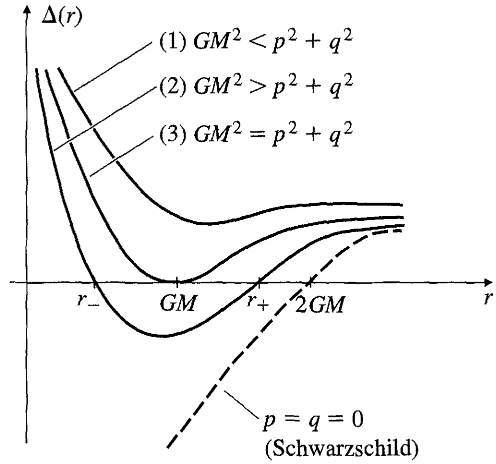

FIGURE 6.2 The function $\Delta(r) = 1 - 2GM/r + G(Q^2 + P^2)/r^2$ for the Reissner–Nordström solutions; zeroes indicate the location of an event horizon.

ing event horizons, if we had cleverly chosen coordinates so that this condition is satisfied at some fixed value of r. Fortunately the coordinates of (6.53) have this property, and the event horizon will be located at

$$g^{rr}(r) = \Delta(r) = 1 - \frac{2GM}{r} + \frac{G(Q^2 + P^2)}{r^2} = 0. \tag{6.56}$$

This will occur at

$$r_{\pm} = GM \pm \sqrt{G^2M^2 - G(Q^2 + P^2)}. \tag{6.57}$$

As shown in Figure 6.2, this might constitute two, one, or zero solutions, depending on the relative values of GM^2 and $Q^2 + P^2$. We therefore consider each case separately.

Case One: $GM^2 < Q^2 + P^2$

In this case the coefficient Δ is always positive (never zero), and the metric is completely regular in the (t, r, θ, ϕ) coordinates all the way down to $r = 0$. The coordinate t is always timelike, and r is always spacelike. But still there is the singularity at $r = 0$, which is now a timelike line. Since there is no event horizon, there is no obstruction to an observer traveling to the singularity and returning to report on what was observed. This is a naked singularity, as discussed earlier. A careful analysis of the geodesics reveals that the singularity is repulsive—timelike geodesics never intersect $r = 0$; instead they approach and then reverse course and move away. (Null geodesics can reach the singularity, as can nongeodesic timelike curves.)

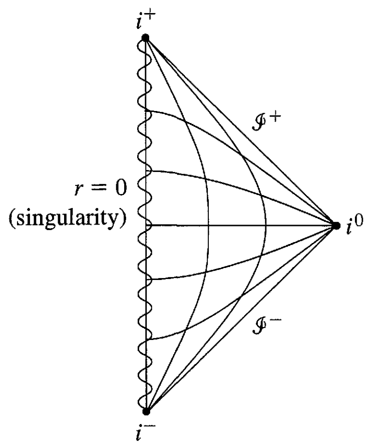

FIGURE 6.3 Conformal diagram for Reissner–Nordström solution with $Q^2 + P^2 > GM^2$. There is a naked singularity at the origin.

As $r \to \infty$ the solution approaches flat spacetime, and as we have just seen the causal structure seems normal everywhere. The conformal diagram will therefore be just like that of Minkowski space, except that now $r = 0$ is a singularity, as shown in Figure 6.3.

The nakedness of the singularity offends our sense of decency, as well as the cosmic censorship conjecture. In fact, we should never expect to find a black hole with $GM^2 < Q^2 + P^2$ as the result of gravitational collapse. Roughly speaking, this condition states that the total energy of the hole is less than the contribution to the energy from the electromagnetic fields alone—that is, the mass of the matter that carried the charge would have had to be negative. This solution is therefore generally considered to be unphysical. Notice also that there are no Cauchy surfaces in this spacetime, since timelike lines can begin and end at the singularity.

Case Two: $GM^2 > Q^2 + P^2$

We expect this situation to apply in realistic gravitational collapse; the energy in the electromagnetic field is less than the total energy. In this case the metric coefficient $\Delta(r)$ is positive at large r and small r, and negative inside the two vanishing points $r_\pm = GM \pm \sqrt{G^2M^2 - G(Q^2 + P^2)}$. The metric has coordinate singularities at both r_+ and r_-; in both cases these could be removed by a change of coordinates as we did with Schwarzschild.

The surfaces defined by $r = r_\pm$ are both null, and they are both event horizons. The singularity at $r = 0$ is a timelike line, not a spacelike surface as in Schwarzschild. If you are an observer falling into the black hole from far away, r_+ is just like $2GM$ in the Schwarzschild metric; at this radius r switches from being a spacelike coordinate to a timelike coordinate, and you necessarily move in the direction of decreasing r. Witnesses outside the black hole also see the same phenomena that they would outside an uncharged hole—the infalling observer is seen to move more and more slowly, and is increasingly redshifted.

But the inevitable fall from r_+ to ever-decreasing radii only lasts until you reach the null surface $r = r_-$, where r switches back to being a spacelike coordinate and the motion in the direction of decreasing r can be arrested. Therefore you do not have to hit the singularity at $r = 0$; this is to be expected, since $r = 0$ is a timelike line (and therefore not necessarily in your future). In fact you can choose either to continue on to $r = 0$, or begin to move in the direction of increasing r back through the null surface at $r = r_-$. Then r will once again be a timelike coordinate, but with reversed orientation; you are forced to move in the direction of *increasing* r. You will eventually be spit out past $r = r_+$ once more, which is like emerging from a white hole into the rest of the universe. From here you can choose to go back into the black hole—this time, a different hole than the one you entered in the first place—and repeat the voyage as many times as you like. This little story corresponds to the conformal diagram in Figure 6.4, which of course can be derived more rigorously by choosing appropriate coordinates and analytically extending the Reissner–Nordström metric as far as it will go.

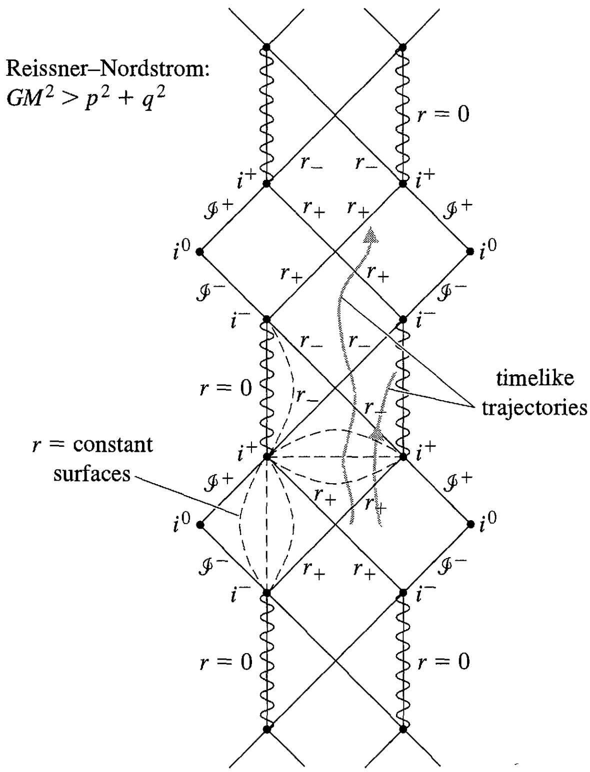

FIGURE 6.4 Conformal diagram for Reissner–Nordström solution with $GM^2 > Q^2 + P^2$. There are an infinite number of copies of the region outside the black hole.

How much of this is science, as opposed to science fiction? Probably not much. If you think about the world as seen from an observer inside the black hole who is about to cross the event horizon at r_-, you notice that the observer can look back in time to see the entire history of the external (asymptotically flat) universe, at least as seen from the black hole. But they see this (infinitely long) history in a finite amount of their proper time—thus, any signal that gets to them as they approach r_- is infinitely blueshifted. Therefore it is likely that any nonspherically-symmetric perturbation that comes into an RN black hole will violently disturb the geometry we have described. It's hard to say what the actual geometry will look like, but there is no very good reason to believe that it must contain an infinite number of asymptotically flat regions connecting to each other via various wormholes.[5]

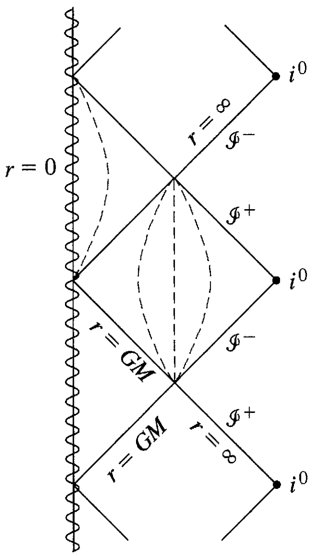

FIGURE 6.5 Conformal diagram for the extremal Reissner–Nordström solution, $GM^2 = Q^2 + P^2$. There is a naked singularity at the origin, and an infinite number of external regions.

Case Three: $GM^2 = Q^2 + P^2$

This case is known as the **extreme** Reissner–Nordström solution. On the one hand the extremal hole is an amusing theoretical toy; this solution is often examined in studies of the role of black holes in quantum gravity. In supersymmetric theories, extremal black holes can leave certain symmetries unbroken, which is a considerable aid in calculations. On the other hand it appears unstable, since adding just a little bit of matter will bring it to Case Two.

The extremal black holes have $\Delta(r) = 0$ at a single radius, $r = GM$. This represents an event horizon, but the r coordinate is never timelike; it becomes null at $r = GM$, but is spacelike on either side. The singularity at $r = 0$ is a timelike line, as in the other cases. So for this black hole you can again avoid the singularity and continue to move to the future to extra copies of the asymptotically flat region, but the singularity is always "to the left." The conformal diagram is shown in Figure 6.5.

A fascinating property of extremal black holes is that the mass is in some sense balanced by the charge. More specifically, two extremal holes with same-sign charges will attract each other gravitationally, but repel each other electromagnetically, and it turns out that these effects precisely cancel. Indeed, we can find *exact* solutions to the coupled Einstein–Maxwell equations representing any number of such black holes in a stationary configuration. To see this, turn first to the Reissner–Nordström metric itself, and let's stick with electric charges rather than magnetic charges, just for simplicity. At extremality, $GM^2 = Q^2$, and the metric takes the form

$$ds^2 = -\left(1 - \frac{GM}{r}\right)^2 dt^2 + \left(1 - \frac{GM}{r}\right)^{-2} dr^2 + r^2 d\Omega^2. \tag{6.58}$$

By defining a shifted radial coordinate

$$\rho = r - GM, \tag{6.59}$$

[5]For some work on this issue, see E. Poisson and W. Israel, *Phys. Rev. D* **41**, 1796 (1990).

the metric takes the isotropic form

$$ds^2 = -H^{-2}(\rho)\mathrm{d}t^2 + H^2(\rho)[\mathrm{d}\rho^2 + \rho^2 d\Omega^2], \tag{6.60}$$

where

$$H(\rho) = 1 + \frac{GM}{\rho}. \tag{6.61}$$

Because $\mathrm{d}\rho^2 + \rho^2 d\Omega^2$ is just the flat metric in three spatial dimensions, we can write (6.60) equally well as

$$ds^2 = -H^{-2}(\vec{x})\mathrm{d}t^2 + H^2(\vec{x})[\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2], \tag{6.62}$$

where H can be written

$$H = 1 + \frac{GM}{|\vec{x}|}. \tag{6.63}$$

In the original r coordinate, the electric field of the extremal solution can be expressed in terms of a vector potential A_μ as

$$E_r = F_{rt} = \frac{Q}{r^2} = \partial_r A_0, \tag{6.64}$$

where the timelike component of the vector potential is

$$A_0 = -\frac{Q}{r}, \tag{6.65}$$

and we imagine the spatial components vanish (having set the magnetic field to zero). In our new ρ coordinate, and with the extremality condition $Q^2 = GM^2$, this becomes

$$A_0 = -\frac{\sqrt{G}M}{\rho + GM}, \tag{6.66}$$

or equivalently

$$\sqrt{G}A_0 = H^{-1} - 1. \tag{6.67}$$

But now let's forget that we know that H obeys (6.61), and simply plug the metric (6.62) and the electrostatic potential (6.67) into Einstein's equation and Maxwell's equations, imagining that H is time-independent ($\partial_0 H = 0$) but otherwise unconstrained. We can straightforwardly show (see the Exercises) that they can be simultaneously satisfied by any time-independent function $H(\vec{x})$ that obeys

$$\nabla^2 H = 0, \tag{6.68}$$

where $\nabla^2 = \partial_x^2 + \partial_y^2 + \partial_z^2$. This is simply Laplace's equation, and it is straightforward to write down all of the solutions that are well-behaved at infinity; they take the form

$$H = 1 + \sum_{a=1}^{N} \frac{GM_a}{|\vec{x} - \vec{x}_a|}, \tag{6.69}$$

for some set of N spatial points defined by $\vec{x}_a$. These points describe the locations of N extremal RN black holes with masses M_a and charges $Q_a = \sqrt{G}M_a$. This multi-extremal-black hole metric is undoubtedly one of the most remarkable exact solutions to Einstein's equation.

6.6 ■ ROTATING (KERR) BLACK HOLES

We could go into a good deal more detail about the charged solutions, but let's instead move on to rotating black holes. To find the exact solution for the metric in this case is much more difficult, since we have given up on spherical symmetry. Instead we look for solutions with axial symmetry around the axis of rotation that are also stationary (a timelike Killing vector). Although the Schwarzschild and Reissner–Nordström solutions were discovered soon after general relativity was invented, the solution for a rotating black hole was found by Kerr only in 1963. His result, the **Kerr metric**, is given by the following mess:

$$\boxed{\begin{aligned} ds^2 = &-\left(1 - \frac{2GMr}{\rho^2}\right) dt^2 - \frac{2GMar\sin^2\theta}{\rho^2}(dt\, d\phi + d\phi\, dt) \\ &+ \frac{\rho^2}{\Delta}dr^2 + \rho^2 d\theta^2 + \frac{\sin^2\theta}{\rho^2}\left[(r^2 + a^2)^2 - a^2\Delta\sin^2\theta\right] d\phi^2, \end{aligned}} \tag{6.70}$$

where

$$\boxed{\Delta(r) = r^2 - 2GMr + a^2} \tag{6.71}$$

and

$$\boxed{\rho^2(r, \theta) = r^2 + a^2\cos^2\theta.} \tag{6.72}$$

The two constants M and a parameterize the possible solutions. To verify that the mass M is equal to the Komar energy (6.38) is straightforward but tedious, while a is the angular momentum per unit mass,

$$a = J/M, \tag{6.73}$$

where J is the Komar angular momentum (6.48). It is easy to include electric and magnetic charges Q and P, simply by replacing $2GMr$ with $2GMr - G(Q^2 +$

P^2); the result is the **Kerr–Newman metric**. The associated one-form potential has nonvanishing components

$$A_t = \frac{Qr - Pa\cos\theta}{\rho^2}, \quad A_\phi = \frac{-Qar\sin^2\theta + P(r^2 + a^2)\cos\theta}{\rho^2}. \tag{6.74}$$

All of the essential phenomena persist in the absence of charges, so we will set $Q = P = 0$ from now on.

The coordinates (t, r, θ, ϕ) are known as **Boyer–Lindquist coordinates**. It is straightforward to check that as $a \to 0$ they reduce to Schwarzschild coordinates. If we keep a fixed and let $M \to 0$, however, we recover flat spacetime but not in ordinary polar coordinates. The metric becomes

$$ds^2 = -dt^2 + \frac{(r^2 + a^2\cos^2\theta)}{(r^2 + a^2)}dr^2 + (r^2 + a^2\cos^2\theta)^2 d\theta^2 + (r^2 + a^2)\sin^2\theta\, d\phi^2, \tag{6.75}$$

and we recognize the spatial part of this as flat space in ellipsoidal coordinates, as shown in Figure 6.6. They are related to Cartesian coordinates in Euclidean 3-space by

$$\begin{aligned} x &= (r^2 + a^2)^{1/2}\sin\theta\, \cos\phi \\ y &= (r^2 + a^2)^{1/2}\sin\theta\, \sin\phi \\ z &= r\cos\theta. \end{aligned} \tag{6.76}$$

There are two Killing vectors of the metric (6.70), both of which are manifest; since the metric coefficients are independent of t and ϕ, both $K = \partial_t$ and $R = \partial_\phi$ are Killing vectors. Of course R^μ expresses the axial symmetry of the solution. The vector K^μ is not orthogonal to $t =$ constant hypersurfaces, and in fact is not orthogonal to any hypersurfaces at all; hence this metric is stationary, but not

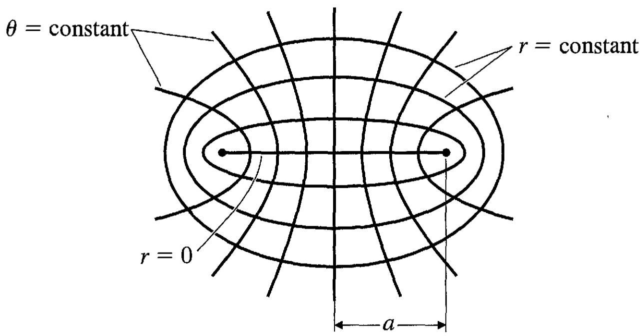

FIGURE 6.6 Ellipsoidal coordinates (r, θ), used in the Kerr metric. $r = 0$ is a two-dimensional disk; the intersection of $r = 0$ with $\theta = \pi/2$ is the ring at the boundary of this disk.

static. This makes sense; the black hole is spinning, so it's not static, but it is spinning in exactly the same way at all times, so it's stationary. Alternatively, the metric can't be static because it's not time-reversal invariant, since that would reverse the angular momentum of the hole.

The Kerr metric also possesses a Killing tensor. These were defined in Chapter 3 as any symmetric $(0, n)$ tensor $\sigma_{\mu_1 \cdots \mu_n}$ satisfying

$$\nabla_{(\lambda}\sigma_{\mu_1 \cdots \mu_n)} = 0. \tag{6.77}$$

In the Kerr geometry we can define the $(0, 2)$ tensor

$$\sigma_{\mu\nu} = 2\rho^2 l_{(\mu}n_{\nu)} + r^2 g_{\mu\nu}. \tag{6.78}$$

In this expression the two vectors l and n are given (with indices raised) by

$$\begin{aligned} l^\mu &= \frac{1}{\Delta}\left(r^2 + a^2, \Delta, 0, a\right) \\ n^\mu &= \frac{1}{2\rho^2}\left(r^2 + a^2, -\Delta, 0, a\right). \end{aligned} \tag{6.79}$$

Both vectors are null and satisfy

$$l^\mu l_\mu = 0, \quad n^\mu n_\mu = 0, \quad l^\mu n_\mu = -1. \tag{6.80}$$

With these definitions, you can check for yourself that $\sigma_{\mu\nu}$ is a Killing tensor.

We have chosen coordinates for Kerr such that the event horizons occur at those fixed values of r for which $g^{rr} = 0$. Since $g^{rr} = \Delta/\rho^2$, and $\rho^2 \geq 0$, this occurs when

$$\Delta(r) = r^2 - 2GMr + a^2 = 0. \tag{6.81}$$

As in the Reissner–Nordström solution, there are three possibilities: $GM > a$, $GM = a$, and $GM < a$. The last case features a naked singularity, and the extremal case $GM = a$ is unstable, just as in Reissner–Nordström. Since these cases are of less physical interest, we will concentrate on $GM > a$. Then there are two radii at which Δ vanishes, given by

$$r_\pm = GM \pm \sqrt{G^2M^2 - a^2}. \tag{6.82}$$

Both radii are null surfaces that will turn out to be event horizons; a side view of a Kerr black hole is portrayed in Figure 6.7. The analysis of these surfaces proceeds in close analogy with the Reissner–Nordström case; it is straightforward to find coordinates that extend through the horizons.

Because Kerr is stationary but not static, the event horizons at $r_\pm$ are not Killing horizons for the asymptotic time-translation Killing vector $K = \partial_t$. The norm of K^μ is given by

$$K^\mu K_\mu = -\frac{1}{\rho^2}(\Delta - a^2 \sin^2\theta). \tag{6.83}$$

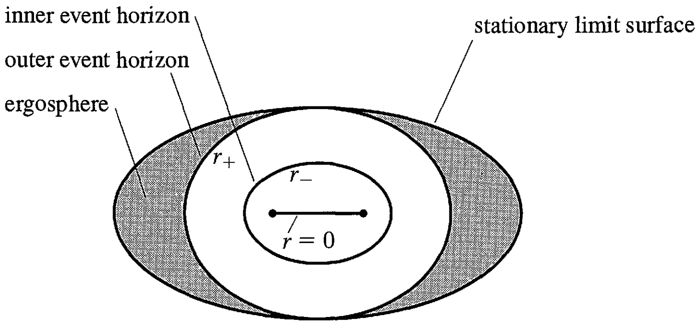

FIGURE 6.7 Horizon structure around the Kerr solution (side view). The event horizons are null surfaces that demarcate points past which it becomes impossible to return to a certain region of space. The stationary limit surface, in contrast, is timelike except where it is tangent to the event horizon (at the poles); it represents the place past which it is impossible to be a stationary observer. The ergosphere between the stationary limit surface and the outer event horizon is a region in which it is possible to enter and leave again, but not to remain stationary.

This does not vanish at the outer event horizon; in fact, at $r = r_+$ (where $\Delta = 0$), we have

$$K^\mu K_\mu = \frac{a^2}{\rho^2}\sin^2\theta \geq 0. \tag{6.84}$$

So the Killing vector is already spacelike at the outer horizon, except at the north and south poles ($\theta = 0, \pi$) where it is null. The locus of points where $K^\mu K_\mu = 0$ is of course the stationary limit surface, and is given by

$$(r - GM)^2 = G^2M^2 - a^2\cos^2\theta, \tag{6.85}$$

while the outer event horizon is given by

$$(r_+ - GM)^2 = G^2M^2 - a^2. \tag{6.86}$$

There is thus a region between these two surfaces, known as the **ergosphere**. Inside the ergosphere, you must move in the direction of the rotation of the black hole (the ϕ direction); however, you can still move toward or away from the event horizon (and have no trouble exiting the ergosphere). The ergosphere is evidently a place where interesting things can happen even before you cross the horizon; more details on this later.

Before rushing to draw conformal diagrams, we need to understand the nature of the true curvature singularity; this does not occur at $r = 0$ in this spacetime, but rather at $\rho = 0$ (where the curvature invariant $R_{\rho\sigma\mu\nu}R^{\rho\sigma\mu\nu}$ diverges). Since $\rho^2 = r^2 + a^2\cos^2\theta$ is the sum of two manifestly nonnegative quantities, it can only vanish when both quantities are zero, or

$$r = 0, \quad \theta = \frac{\pi}{2}. \tag{6.87}$$

This seems like a funny result, but remember that $r = 0$ is not a point in space, but a disk; the set of points $r = 0$, $\theta = \pi/2$ is actually the *ring* at the edge of this disk. The rotation has "softened" the Schwarzschild singularity, spreading it out over a ring.

What happens if you go inside the ring? A careful analytic continuation (which we will not perform) would reveal that you exit to another asymptotically flat spacetime, but not an identical copy of the one you came from. The new spacetime is described by the Kerr metric with $r < 0$. As a result, Δ never vanishes and there are no horizons. The conformal diagram, Figure 6.8, is much like that for Reissner–Nordström, except now you can pass through the singularity. Because the Kerr metric is not spherically symmetric, the conformal diagram is not quite

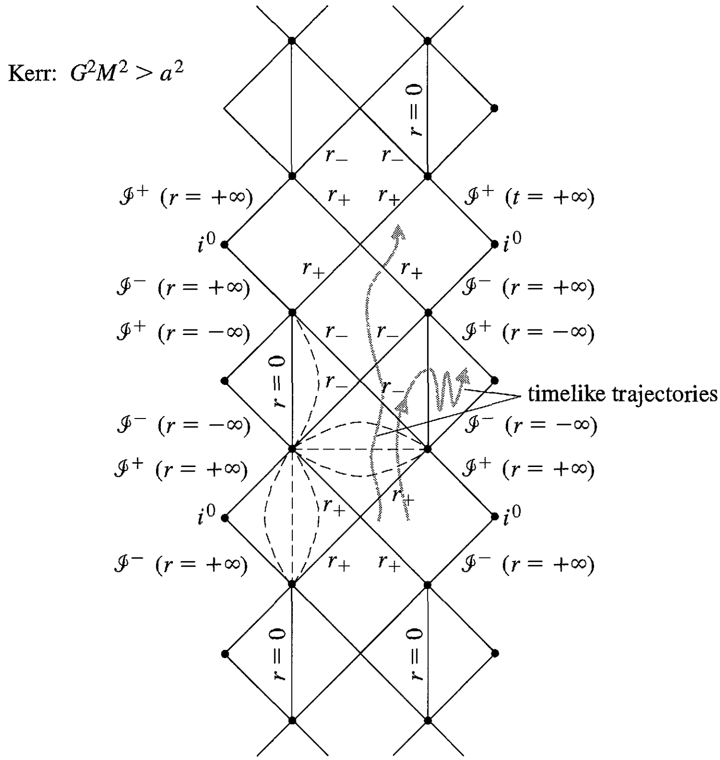

FIGURE 6.8 Conformal diagram for the Kerr solution with $G^2M^2 > a^2$. As with the analogous charged solution, there are an infinite number of copies of the region outside the black hole.

as faithful as in the previous cases; a single point on the diagram represents fixed values of t and r, and will have a different geometry for different values of θ.

Not only do we have the usual strangeness of these distinct asymptotically flat regions connected to ours through the black hole, but the region near the ring singularity has additional pathologies: closed timelike curves. If you consider trajectories that wind around in ϕ while keeping θ and t constant and r a small negative value, the line element along such a path is

$$ds^2 \approx a^2 \left(1 + \frac{2GM}{r}\right) d\phi^2, \tag{6.88}$$

which is negative for small negative r. Since these paths are closed, they are obviously CTCs. You can therefore meet yourself in the past, with all that entails.

Of course, everything we say about the analytic extension of Kerr is subject to the same caveats we mentioned for Schwarzschild and Reissner–Nordström; it is unlikely that realistic gravitational collapse leads to these bizarre spacetimes. It is nevertheless always useful to have exact solutions. Furthermore, for the Kerr metric strange things are happening even if we stay outside the event horizon, to which we now turn.

We begin by considering more carefully the angular velocity of the hole. Obviously the conventional definition of angular velocity will have to be modified somewhat before we can apply it to something as abstract as the metric of spacetime. Let us consider the fate of a photon that is emitted in the ϕ direction at some radius r in the equatorial plane ($\theta = \pi/2$) of a Kerr black hole. The instant it is emitted its momentum has no components in the r or θ direction, and therefore the condition that the trajectory be null is

$$ds^2 = 0 = g_{tt}dt^2 + g_{t\phi}(dt d\phi + d\phi dt) + g_{\phi\phi}d\phi^2. \tag{6.89}$$

This can be immediately solved to obtain

$$\frac{d\phi}{dt} = -\frac{g_{t\phi}}{g_{\phi\phi}} \pm \sqrt{\left(\frac{g_{t\phi}}{g_{\phi\phi}}\right)^2 - \frac{g_{tt}}{g_{\phi\phi}}}. \tag{6.90}$$

If we evaluate this quantity on the stationary limit surface of the Kerr metric, we have $g_{tt} = 0$, and the two solutions are

$$\frac{d\phi}{dt} = 0, \quad \frac{d\phi}{dt} = \frac{a}{2G^2M^2 + a^2}. \tag{6.91}$$

The nonzero solution has the same sign as a; we interpret this as the photon moving around the hole in the same direction as the hole's rotation. The zero solution means that the photon directed against the hole's rotation doesn't move at all in this coordinate system. Note that we haven't given a full solution to the photon's trajectory, only shown that its instantaneous velocity is zero. This is an example of a phenomenon known as the "dragging of inertial frames"; it is explored more in one of the exercises to Chapter 7. Massive particles, which must move more slowly than photons, are necessarily dragged along with the hole's rotation once they are inside the stationary limit surface. This dragging continues as we approach the outer event horizon at r_+; we can define the angular velocity of the event horizon itself, $\Omega_{\rm H}$, to be the minimum angular velocity of a particle at the horizon. Directly from (6.90) we find that

$$\Omega_{\rm H} = \left(\frac{d\phi}{dt}\right)_{-}(r_+) = \frac{a}{r_+^2 + a^2}. \tag{6.92}$$

6.7 ■ THE PENROSE PROCESS AND BLACK-HOLE THERMODYNAMICS

Black hole thermodynamics is one of the most fascinating and mysterious subjects in general relativity. To get there, however, let us begin with something apparently very straightforward: motion along geodesics in the Kerr metric. We know that such a discussion will be simplified by considering the conserved quantities associated with the Killing vectors $K = \partial_t$ and $R = \partial_\phi$. For the purposes at hand we can restrict our attention to massive particles, for which we can work with the four-momentum

$$p^\mu = m\frac{dx^\mu}{d\tau}, \tag{6.93}$$

where m is the rest mass of the particle. Then we can take as our two conserved quantities the actual energy and angular momentum of the particle,

$$E = -K_\mu p^\mu = m\left(1 - \frac{2GMr}{\rho^2}\right)\frac{dt}{d\tau} + \frac{2mGMar}{\rho^2}\sin^2\theta\,\frac{d\phi}{d\tau} \tag{6.94}$$

and

$$L = R_\mu p^\mu = -\frac{2mGMar}{\rho^2}\sin^2\theta\,\frac{dt}{d\tau} + \frac{m(r^2+a^2)^2 - m\Delta a^2\sin^2\theta}{\rho^2}\sin^2\theta\,\frac{d\phi}{d\tau}. \tag{6.95}$$

These differ from the definitions for the conserved quantities used in the last chapter, where E and L were taken to be the energy and angular momentum *per unit mass*. They are conserved either way, of course.

The minus sign in the definition of E is there because at infinity both K^μ and p^μ are timelike, so their inner product is negative, but we want the energy to be positive. Inside the ergosphere, however, K^μ becomes spacelike; we can therefore imagine particles for which

$$E = -K_\mu p^\mu < 0. \tag{6.96}$$

The extent to which this bothers us is ameliorated somewhat by the realization that *all* particles must have positive energies if they are outside the stationary limit surface; therefore a particle inside the ergosphere with negative energy must either remain in the ergosphere, or be accelerated until its energy is positive if it is to escape.

Still, this realization leads to a way to extract energy from a rotating black hole; the method is known as the **Penrose process**. The idea is simple; starting from outside the ergosphere, you arm yourself with a large rock and leap toward the black hole. If we call the four-momentum of the (you + rock) system $p^{(0)\mu}$, then the energy $E^{(0)} = -K_\mu p^{(0)\mu}$ is certainly positive, and conserved as you move along your geodesic. Once you enter the ergosphere, you hurl the rock with all your might, in a very specific way. If we call your momentum $p^{(1)\mu}$ and that of the rock $p^{(2)\mu}$, then at the instant you throw it we have conservation of momentum just as in special relativity:

$$p^{(0)\mu} = p^{(1)\mu} + p^{(2)\mu}. \tag{6.97}$$

Contracting with the Killing vector K_μ gives

$$E^{(0)} = E^{(1)} + E^{(2)}. \tag{6.98}$$

But, if we imagine that you are arbitrarily strong (and accurate), you can arrange your throw such that $E^{(2)} < 0$, as per (6.96). Furthermore, Penrose was able to show that you can arrange the initial trajectory and the throw as shown in Figure 6.9, such that afterward you follow a geodesic trajectory back outside the stationary limit surface into the external universe. Since your energy is conserved

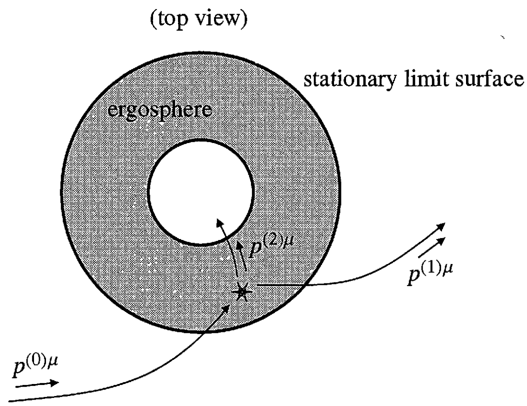

FIGURE 6.9 The Penrose process (top view). An object falls toward a Kerr black hole and splits in two while in the ergosphere (within the stationary limit surface, but outside the outer event horizon). One piece falls into the horizon with a negative energy $E^{(2)}$, while the other escapes to infinity with a larger energy than that of the original infalling object.

along the way, at the end we will have

$$E^{(1)} > E^{(0)}. \tag{6.99}$$

Thus, you have emerged with *more* energy than you entered with.

There is no such thing as a free lunch; the energy you gained came from somewhere, and that somewhere is the black hole. In fact, the Penrose process extracts energy from the rotating black hole by decreasing its angular momentum; you have to throw the rock against the hole's rotation to get the trick to work. To see this more precisely, recall that we claimed earlier in this chapter that any event horizon in a stationary spacetime would be a Killing horizon for some Killing vector. For Kerr this is a linear combination of the time-translation and rotational Killing vectors,

$$\chi^\mu = K^\mu + \Omega_{\rm H} R^\mu, \tag{6.100}$$

where $\Omega_{\rm H}$ is precisely the angular velocity of the horizon as defined in (6.92). Using $K = \partial_t$ and $R = \partial_\phi$, it is straightforward to verify that χ^μ becomes null at the outer event horizon. The statement that the particle with momentum $p^{(2)\mu}$ crosses the event horizon "moving forward in time" is simply

$$p^{(2)\mu}\chi_\mu < 0. \tag{6.101}$$

Plugging in the definitions of E and L, we see that this condition is equivalent to

$$L^{(2)} < \frac{E^{(2)}}{\Omega_{\rm H}}. \tag{6.102}$$

Since we have arranged $E^{(2)}$ to be negative, and $\Omega_{\rm H}$ positive, we see that the particle must have a negative angular momentum—it is moving against the hole's rotation. Once you have escaped the ergosphere and the rock has fallen inside the event horizon, the mass and angular momentum of the hole are what they used to be plus the negative contributions of the rock:

$$\begin{aligned} \delta M &= E^{(2)} \\ \delta J &= L^{(2)}, \end{aligned} \tag{6.103}$$

where $J = Ma$ is the angular momentum of the black hole. Then (6.102) becomes a limit on how much you can decrease the angular momentum:

$$\delta J < \frac{\delta M}{\Omega_{\rm H}}. \tag{6.104}$$

If we exactly reach this limit, as the rock we throw in becomes more and more null, we have the "ideal" process, in which $\delta J = \delta M/\Omega_{\rm H}$.

We will now use these ideas to verify that, although you can use the Penrose process to extract energy from the black hole (thereby decreasing M), you cannot violate the area theorem: The area of the event horizon is nondecreasing. Although the mass decreases, the angular momentum must also decrease, in a combination which only allows the area to increase. To see this, let's calculate the area of the outer event horizon, which is located at

$$r_+ = GM + \sqrt{G^2M^2 - a^2}. \tag{6.105}$$

The induced metric γ_{ij} on the horizon (where i and j run over $\{\theta, \phi\}$) can be found straightforwardly by setting $r = r_+$ (so $\Delta = 0$), $\mathrm{d}t = 0$ and $\mathrm{d}r = 0$ in (6.70):

$$\begin{aligned} \gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j &= ds^2(\mathrm{d}t = 0, \mathrm{d}r = 0, r = r_+) \\ &= (r_+^2 + a^2\cos^2\theta)\mathrm{d}\theta^2 + \left[\frac{(r_+^2 + a^2)^2\sin^2\theta}{r_+^2 + a^2\cos^2\theta}\right]\mathrm{d}\phi^2. \end{aligned} \tag{6.106}$$

The horizon area is then the integral of the induced volume element,

$$A = \int \sqrt{|\gamma|}\, d\theta d\phi. \tag{6.107}$$

The determinant is

$$|\gamma| = (r_+^2 + a^2)^2 \sin^2\theta, \tag{6.108}$$

so the horizon area is simply

$$A = 4\pi(r_+^2 + a^2). \tag{6.109}$$

To show that the area doesn't decrease, it is convenient to work instead in terms of the **irreducible mass** of the black hole, defined by

$$\begin{aligned} M_{\mathrm{irr}}^2 &= \frac{A}{16\pi G^2} \\ &= \frac{1}{4G^2}(r_+^2 + a^2) \\ &= \frac{1}{2}\left(M^2 + \sqrt{M^4 - (Ma/G)^2}\right) \\ &= \frac{1}{2}\left(M^2 + \sqrt{M^4 - (J/G)^2}\right). \end{aligned} \tag{6.110}$$

We can differentiate to obtain, after a bit of work, how M_{irr} is affected by changes in the mass or angular momentum,

$$\delta M_{\mathrm{irr}} = \frac{a}{4GM_{\mathrm{irr}}\sqrt{G^2M^2 - a^2}}(\Omega_{\mathrm{H}}^{-1}\delta M - \delta J). \tag{6.111}$$

Then our limit (6.104) becomes

$$\delta M_{\text{irr}} > 0. \tag{6.112}$$

The irreducible mass can never be reduced; hence the name. It follows that the maximum amount of energy we can extract from a black hole before we slow its rotation to zero is

$$M - M_{\text{irr}} = M - \frac{1}{\sqrt{2}}\left(M^2 + \sqrt{M^4 - (J/G)^2}\right)^{1/2}. \tag{6.113}$$

The result of this complete extraction is a Schwarzschild black hole of mass M_{irr}. It turns out that the best we can do is to start with an extreme Kerr black hole; then we can get out approximately 29% of its total energy.

The irreducibility of M_{irr} leads immediately to the fact that the area A can never decrease. From (6.110) and (6.111) we have

$$\delta A = 8\pi G \frac{a}{\Omega_{\text{H}}\sqrt{G^2M^2 - a^2}}(\delta M - \Omega_{\text{H}}\delta J), \tag{6.114}$$

which can be recast as

$$\delta M = \frac{\kappa}{8\pi G}\delta A + \Omega_{\text{H}}\delta J, \tag{6.115}$$

where we have introduced

$$\kappa = \frac{\sqrt{G^2M^2 - a^2}}{2GM(GM + \sqrt{G^2M^2 - a^2})}. \tag{6.116}$$

The quantity κ is of course just the surface gravity of the Kerr solution, as you could verify by plugging (6.100) into (6.9).

Equations like (6.115) first started people thinking about a correspondence between black holes and thermodynamics. Consider the first law of thermodynamics,

$$dE = T\,dS - p\,dV, \tag{6.117}$$

where T is the temperature, S is the entropy, p is the pressure, and V is the volume, so the pdV term represents work we do to the system. It is natural to think of the term $\Omega_{\text{H}}\delta J$ in (6.115) as work that we do on the black hole by throwing rocks into it. Then the correspondence begins to take shape if we think of identifying the thermodynamic quantities energy, entropy, and temperature with the black-hole mass, area, and surface gravity:

$$\begin{aligned} E &\leftrightarrow M \\ S &\leftrightarrow A/4G \\ T &\leftrightarrow \kappa/2\pi. \end{aligned} \tag{6.118}$$

(Remember we are using units in which $\hbar = c = k = 1$.) In the context of classical general relativity the analogy is essentially perfect, with each law of thermodynamics corresponding to a law of black hole mechanics. A system in thermal equilibrium will have settled to a stationary state, corresponding to a stationary black hole. The zeroth law of thermodynamics states that in thermal equilibrium the temperature is constant throughout the system; the analogous statement for black holes is that stationary black holes have constant surface gravity on the entire horizon. This will be true, at least under the same reasonable assumptions under which the event horizon is a Killing horizon. As we have seen, the first law (6.117) is equivalent to (6.115). The second law, that entropy never decreases, is simply the statement that the area of the horizon never decreases. Finally, the usual statement of the third law is that it is impossible to achieve $T = 0$ in any physical process, or that the entropy must go to zero as the temperature goes to zero. For black holes this doesn't quite work; it turns out that $\kappa = 0$ corresponds to extremal black holes, which don't necessarily have a vanishing area. But the thermodynamic third law doesn't really work either, in the sense that there are ordinary physical systems that violate it; the third law applies to some situations but is not truly fundamental.

We have cheated a little in proposing the correspondence (6.118); you will notice that by equating TdS with $\kappa dA/8\pi G$ we do not know how to separately normalize S/A or T/κ, only their combination. As we will discuss in Chapter 9, however, Hawking showed that quantum fields in a black-hole background allow the hole to radiate at a temperature $T = \kappa/2\pi$. Once this is known, we can interpret $A/4G$ as an actual entropy of the black hole. Bekenstein has proposed a **generalized second law**, that the combined entropy of matter and black holes never decreases:

$$\delta\left(S + \frac{A}{4G}\right) \geq 0. \tag{6.119}$$

The generalized second law can actually be proven under a variety of assumptions. Usually, however, we like to associate the entropy of a system with the logarithm of the number of accessible quantum states. There is therefore some tension between this concept and the no-hair theorem, which indicates that there are very few possible states for a black hole of fixed charge, mass, and spin (only one, in fact). It seems likely that this behavior is an indication of a profound feature of the interaction between quantum mechanics and gravitation.

6.8 ■ EXERCISES

1. Show that the coupled Einstein–Maxwell equations can be simultaneously solved by the metric (6.62) and the electrostatic potential (6.67) if $H(\vec{x})$ obeys Laplace's equation,

$$\nabla^2 H = 0. \tag{6.120}$$

2. Consider the orbits of massless particles, with affine parameter λ, in the equatorial plane of a Kerr black hole.

(a) Show that

$$\left(\frac{dr}{d\lambda}\right)^2 = \frac{\Sigma^2}{\rho^4}(E - LW_+(r))(E - LW_-(r)), \tag{6.121}$$

where $\Sigma^2 = (r^2 + a^2)^2 - a^2\Delta(r)\sin^2\theta$, E and L are the conserved energy and angular momentum, and you have to find expressions for $W_\pm(r)$.

(b) Using this result, and assuming that $\Sigma^2 > 0$ everywhere, show that the orbit of a photon in the equatorial plane cannot have a turning point inside the outer event horizon r_+. This means that ingoing light rays cannot escape once they cross r_+, so it really is an event horizon.

3. In the presence of an electromagnetic field, a particle of charge e and mass m obeys

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\tau}\frac{dx^\sigma}{d\tau} = \frac{e}{m}F^\mu{}_\nu\frac{dx^\nu}{d\tau}. \tag{6.122}$$

Imagine that such a particle is moving in the field of a Reissner–Nordström black hole with charge Q and mass M.

(a) Show that the energy

$$E = m\left(1 - \frac{2GM}{r} + \frac{GQ^2}{r^2}\right)\frac{dt}{d\tau} + \frac{eQ}{r} \tag{6.123}$$

is conserved.

(b) Will a Penrose-type process work for a charged black hole? What is the change in the black hole mass, δM, for the maximum physical process?

4. Consider de Sitter space in static coordinates:

$$ds^2 = -\left(1 - \frac{\Lambda}{3}r^2\right)dt^2 + \frac{dr^2}{1 - \frac{\Lambda}{3}r^2} + r^2 d\Omega^2.$$

This space has a Killing vector ∂_t that is timelike near $r = 0$ and null on a Killing horizon. Locate the radial position of the Killing horizon, r_K. What is the surface gravity, κ, of the horizon? Consider the Euclidean signature version of de Sitter space obtained by making the replacement $t \to i\tau$. Show that a coordinate transformation can be made to make the Euclidean metric regular at the horizon, so long as τ is made periodic.

5. What is the magnetic field seen by an observer orbiting a Riessner–Nordström black hole of electric charge Q and mass M in a circular orbit with circumference $2\pi R$?

6. Consider a Kerr black hole with an accretion disk of negligible mass in the equatorial plane. Assume that particles in the disk follow geodesics (that is, ignore any pressure support). Now suppose the disk contains some iron atoms that are being excited by a source of radiation. When the iron atoms de-excite they emit radiation with a known frequency ν_0, as measured in their rest frame. Suppose we detect this radiation far from the black hole (we also lie in the equatorial plane). What is the observed frequency of photons emitted from either edge of the disk, and from the center of the disk? Consider cases where the disk and the black hole are rotating in the same and opposite directions. Can we use these measurements to determine the mass and angular momentum of the black hole?

CHAPTER

7 Perturbation Theory and Gravitational Radiation

7.1 ■ LINEARIZED GRAVITY AND GAUGE TRANSFORMATIONS

When we first derived Einstein's equation, we checked that we were on the right track by considering the Newtonian limit. We took this to mean not only that the gravitational field was weak, but also that it was static (no time derivatives), and that test particles were moving slowly. The weak-field limit described in this chapter is less restrictive, assuming that the field is still weak but it can vary with time, and without any restrictions on the motion of test particles. This will allow us to discuss phenomena that are absent or ambiguous in the Newtonian theory, such as gravitational radiation (where the field varies with time) and the deflection of light (which involves fast-moving particles).

The weakness of the gravitational field is once again expressed as our ability to decompose the metric into the flat Minkowski metric plus a small perturbation,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1. \tag{7.1}$$

We will restrict ourselves to coordinates in which $\eta_{\mu\nu}$ takes its canonical form, $\eta_{\mu\nu} = \mathrm{diag}(-1, +1, +1, +1)$. The assumption that $h_{\mu\nu}$ is small allows us to ignore anything that is higher than first order in this quantity, from which we immediately obtain

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}, \tag{7.2}$$

where $h^{\mu\nu} = \eta^{\mu\rho}\eta^{\nu\sigma}h_{\rho\sigma}$. As before, we can raise and lower indices using $\eta^{\mu\nu}$ and $\eta_{\mu\nu}$, since the corrections would be of higher order in the perturbation. In fact, we can think of the linearized version of general relativity (where effects of higher than first order in $h_{\mu\nu}$ are neglected) as describing a theory of a symmetric tensor field $h_{\mu\nu}$ propagating on a flat background spacetime. This theory is Lorentz invariant in the sense of special relativity; under a Lorentz transformation $x^{\mu'} = \Lambda^{\mu'}{}_{\mu}x^{\mu}$, the flat metric $\eta_{\mu\nu}$ is invariant, while the perturbation transforms as

$$h_{\mu'\nu'} = \Lambda^{\mu'}{}_{\mu}\Lambda^{\nu'}{}_{\nu}h_{\mu\nu}. \tag{7.3}$$

Note that we could have considered small perturbations about some other background spacetime besides Minkowski space. In that case the metric would have been written $g_{\mu\nu} = g^{(0)}_{\mu\nu} + h_{\mu\nu}$, and we would have derived a theory of a symmetric tensor propagating on the curved space with metric $g^{(0)}_{\mu\nu}$. Such an approach is necessary, for example, in cosmology.

We want to find the equations of motion obeyed by the perturbations $h_{\mu\nu}$, which come by examining Einstein's equation to first order. We begin with the Christoffel symbols, which are given by

$$\begin{aligned}\Gamma^{\rho}_{\mu\nu} &= \tfrac{1}{2}g^{\rho\lambda}(\partial_\mu g_{\nu\lambda} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu}) \\ &= \tfrac{1}{2}\eta^{\rho\lambda}(\partial_\mu h_{\nu\lambda} + \partial_\nu h_{\lambda\mu} - \partial_\lambda h_{\mu\nu}).\end{aligned} \tag{7.4}$$

Since the connection coefficients are first-order quantities, the only contribution to the Riemann tensor will come from the derivatives of the Γ's, not the Γ^2 terms. Lowering an index for convenience, we obtain

$$\begin{aligned}R_{\mu\nu\rho\sigma} &= \eta_{\mu\lambda}\partial_\rho\Gamma^{\lambda}_{\nu\sigma} - \eta_{\mu\lambda}\partial_\sigma\Gamma^{\lambda}_{\nu\rho} \\ &= \tfrac{1}{2}(\partial_\rho\partial_\nu h_{\mu\sigma} + \partial_\sigma\partial_\mu h_{\nu\rho} - \partial_\sigma\partial_\nu h_{\mu\rho} - \partial_\rho\partial_\mu h_{\nu\sigma}).\end{aligned} \tag{7.5}$$

The Ricci tensor comes from contracting over μ and ρ, giving

$$R_{\mu\nu} = \tfrac{1}{2}(\partial_\sigma\partial_\nu h^{\sigma}{}_{\mu} + \partial_\sigma\partial_\mu h^{\sigma}{}_{\nu} - \partial_\mu\partial_\nu h - \Box h_{\mu\nu}), \tag{7.6}$$

which is manifestly symmetric in μ and ν. In this expression we have defined the trace of the perturbation as $h = \eta^{\mu\nu}h_{\mu\nu} = h^{\mu}{}_{\mu}$, and the d'Alembertian is simply the one from flat space, $\Box = -\partial_t^2 + \partial_x^2 + \partial_y^2 + \partial_z^2$. Contracting again to obtain the Ricci scalar yields

$$R = \partial_\mu\partial_\nu h^{\mu\nu} - \Box h. \tag{7.7}$$

Putting it all together we obtain the Einstein tensor:

$$\begin{aligned}G_{\mu\nu} &= R_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}R \\ &= \tfrac{1}{2}(\partial_\sigma\partial_\nu h^{\sigma}{}_{\mu} + \partial_\sigma\partial_\mu h^{\sigma}{}_{\nu} - \partial_\mu\partial_\nu h - \Box h_{\mu\nu} - \eta_{\mu\nu}\partial_\rho\partial_\lambda h^{\rho\lambda} + \eta_{\mu\nu}\Box h).\end{aligned} \tag{7.8}$$

Consistent with our interpretation of the linearized theory as one describing a symmetric tensor on a flat background, the linearized Einstein tensor (7.8) can be derived by varying the following Lagrangian with respect to $h_{\mu\nu}$:

$$\begin{aligned}\mathcal{L} = \tfrac{1}{2}\big[&(\partial_\mu h^{\mu\nu})(\partial_\nu h) - (\partial_\mu h^{\rho\sigma})(\partial_\rho h^{\mu}{}_{\sigma}) + \tfrac{1}{2}\eta^{\mu\nu}(\partial_\mu h^{\rho\sigma})(\partial_\nu h_{\rho\sigma}) \\ &- \tfrac{1}{2}\eta^{\mu\nu}(\partial_\mu h)(\partial_\nu h)\big].\end{aligned} \tag{7.9}$$

You are asked to verify the appropriateness of the Lagrangian in the exercises.

The linearized field equation is of course $G_{\mu\nu} = 8\pi G T_{\mu\nu}$, where $G_{\mu\nu}$ is given by (7.8) and $T_{\mu\nu}$ is the energy-momentum tensor, calculated to zeroth order in $h_{\mu\nu}$. We do not include higher-order corrections to the energy-momentum tensor because the amount of energy and momentum must itself be small for the weak-field limit to apply. In other words, the lowest nonvanishing order in $T_{\mu\nu}$ is automatically of the same order of magnitude as the perturbation. Notice that the conservation law to lowest order is simply $\partial_\mu T^{\mu\nu} = 0$. We will often be concerned with the vacuum equation, which as usual is just $R_{\mu\nu} = 0$, where $R_{\mu\nu}$ is given by (7.6).

With the linearized field equation in hand, we are almost prepared to set about solving it. First, however, we should deal with the thorny issue of gauge invariance. This issue arises because the demand that $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ does not completely specify the coordinate system on spacetime; there may be other coordinate systems in which the metric can still be written as the Minkowski metric plus a small perturbation, but the perturbation will be different. Thus, the decomposition of the metric into a flat background plus a perturbation is not unique. To examine this issue, we will draw upon ideas about diffeomorphisms discussed in Appendices A and B; readers who have not yet read those sections can skip to equation (7.14) and the two paragraphs following, which contain the essential ideas.

Let's think about gauge invariance from a highbrow point of view. The notion that the linearized theory can be thought of as one governing the behavior of tensor fields on a flat background can be formalized in terms of a *background spacetime* M_b, a *physical spacetime* M_p, and a diffeomorphism $\phi : M_b \rightarrow M_p$. As manifolds M_b and M_p are the same (since they are diffeomorphic), but we imagine that they possess some different tensor fields; on M_b we have defined the flat Minkowski metric $\eta_{\mu\nu}$, while on M_p we have some metric $g_{\alpha\beta}$ that obeys Einstein's equation. (We imagine that M_b is equipped with coordinates x^μ and M_p is equipped with coordinates y^α, although these will not play a prominent role.) The diffeomorphism ϕ allows us to move tensors back and forth between the background and physical spacetimes, as in Figure 7.1. Since we would like to construct our linearized theory as one taking place on the flat background spacetime, we are interested in the pullback $(\phi^* g)_{\mu\nu}$ of the physical metric. We can define the perturbation as the difference between the pulled-back physical metric

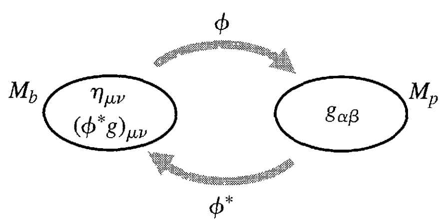

FIGURE 7.1 A diffeomorphism relating the background spacetime M_b (with flat metric $\eta_{\mu\nu}$) to the physical spacetime M_p.

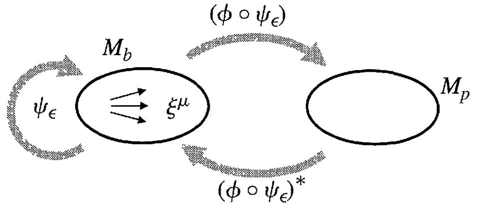

FIGURE 7.2 A one-parameter family of diffeomorphisms ψ_ϵ, generated by the vector field ξ^μ on the background spacetime M_b.

and the flat one:

$$h_{\mu\nu} = (\phi^* g)_{\mu\nu} - \eta_{\mu\nu}. \tag{7.10}$$

From this definition, there is no reason for the components of $h_{\mu\nu}$ to be small; however, if the gravitational fields on M_p are weak, then for *some* diffeomorphisms ϕ we will have $|h_{\mu\nu}| \ll 1$. We therefore limit our attention only to those diffeomorphisms for which this is true. Then the fact that $g_{\alpha\beta}$ obeys Einstein's equation on the physical spacetime means that $h_{\mu\nu}$ will obey the linearized equation on the background spacetime (since ϕ, as a diffeomorphism, can be used to pull back Einstein's equation themselves).

In this language, the issue of gauge invariance is simply that there are a large number of permissible diffeomorphisms between M_b and M_p (where "permissible" means that the perturbation is small). Consider a vector field $\xi^\mu(x)$ on the background spacetime. This vector field generates a one-parameter family of diffeomorphisms $\psi_\epsilon : M_b \to M_b$, as shown in Figure 7.2. For ϵ sufficiently small, if ϕ is a diffeomorphism for which the perturbation defined by (7.10) is small, then so will $(\phi \circ \psi_\epsilon)$ be, although the perturbation will have a different value. Specifically, we can define a family of perturbations parameterized by ϵ:

$$\begin{aligned} h^{(\epsilon)}_{\mu\nu} &= [(\phi \circ \psi_\epsilon)^* g]_{\mu\nu} - \eta_{\mu\nu} \\ &= [\psi^*_\epsilon(\phi^* g)]_{\mu\nu} - \eta_{\mu\nu}. \end{aligned} \tag{7.11}$$

The second equality is based on the fact that the pullback under a composition is given by the composition of the pullbacks in the opposite order, which follows from the fact that the pullback itself moves things in the opposite direction from the original map. Plugging in the relation (7.10), we find

$$\begin{aligned} h^{(\epsilon)}_{\mu\nu} &= \psi^*_\epsilon(h + \eta)_{\mu\nu} - \eta_{\mu\nu} \\ &= \psi^*_\epsilon(h_{\mu\nu}) + \psi^*_\epsilon(\eta_{\mu\nu}) - \eta_{\mu\nu}, \end{aligned} \tag{7.12}$$

since the pullback of the sum of two tensors is the sum of the pullbacks. Now we use our assumption that ϵ is small; in this case $\psi^*_\epsilon(h_{\mu\nu})$ will be equal to $h_{\mu\nu}$ to lowest order, while the other two terms give us a Lie derivative:

$$\begin{aligned} h^{(\epsilon)}_{\mu\nu} &= \psi^*_\epsilon(h_{\mu\nu}) + \epsilon\left[\frac{\psi^*_\epsilon(\eta_{\mu\nu}) - \eta_{\mu\nu}}{\epsilon}\right] \\ &= h_{\mu\nu} + \epsilon\mathcal{L}_\xi\eta_{\mu\nu}. \end{aligned} \tag{7.13}$$

In Appendix B we show that the Lie derivative of the metric along a vector field ξ_μ is $\mathcal{L}_\xi g_{\mu\nu} = 2\nabla_{(\mu}\xi_{\nu)}$. In the current context the background metric is flat, and covariant derivatives become partial derivatives; we therefore have

$$\boxed{h^{(\epsilon)}_{\mu\nu} = h_{\mu\nu} + 2\epsilon\partial_{(\mu}\xi_{\nu)}.} \tag{7.14}$$

This formula represents the change of the metric perturbation under an infinitesimal diffeomorphism along the vector field $\epsilon\xi^\mu$: we will call this a **gauge transformation** in linearized theory.

The diffeomorphisms ψ_ϵ provide a different representation of the same physical situation, while maintaining our requirement that the perturbation be small. Therefore, the result (7.12) tells us what kind of metric perturbations denote physically equivalent spacetimes—those related to each other by $2\epsilon\partial_{(\mu}\xi_{\nu)}$, for some vector ξ^μ. The invariance of our theory under such transformations is analogous to traditional gauge invariance of electromagnetism under $A_\mu \to A_\mu + \partial_\mu\lambda$. (The analogy is different from another analogy we draw with electromagnetism in Appendix J, relating local Lorentz transformations in the orthonormal-frame formalism to changes of basis in an internal vector bundle.) In electromagnetism the invariance comes about because the field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is left unchanged by gauge transformations; similarly, we find that the transformation (7.14) changes the linearized Riemann tensor by

$$\begin{aligned} \delta R_{\mu\nu\rho\sigma} &= \tfrac{1}{2}(\partial_\rho\partial_\nu\partial_\mu\xi_\sigma + \partial_\rho\partial_\nu\partial_\sigma\xi_\mu + \partial_\sigma\partial_\mu\partial_\nu\xi_\rho + \partial_\sigma\partial_\mu\partial_\rho\xi_\nu \\ &\qquad - \partial_\sigma\partial_\nu\partial_\mu\xi_\rho - \partial_\sigma\partial_\nu\partial_\rho\xi_\mu - \partial_\rho\partial_\mu\partial_\nu\xi_\sigma - \partial_\rho\partial_\mu\partial_\sigma\xi_\nu) \\ &= 0. \end{aligned} \tag{7.15}$$

Our abstract derivation of the appropriate gauge transformation for the metric perturbation is verified by the fact that it leaves the curvature (and hence the physical spacetime) unchanged.

Gauge invariance can also be understood from the slightly more lowbrow but considerably more direct route of infinitesimal coordinate transformations. Our diffeomorphism ψ_ϵ can be thought of as changing coordinates from x^μ to $x^\mu - \epsilon\xi^\mu$. (The minus sign, which is unconventional, comes from the fact that the "new" metric is pulled back from a small distance forward along the integral curves, which is equivalent to replacing the coordinates by those a small distance backward along the curves.) Following through the usual rules for transforming tensors under coordinate transformations, you can derive precisely (7.14)—although you have to cheat somewhat by equating components of tensors in two different coordinate systems.

7.2 ■ DEGREES OF FREEDOM

With the expression (7.8) for the linearized Einstein tensor, and the expression (7.14) for the effect of gauge transformations, we could immediately set about choosing a gauge and solving Einstein's equation. However, we can accumulate some additional physical insight by first choosing a fixed inertial coordinate system in the Minkowski background spacetime, and decomposing the components of the metric perturbation according to their transformation properties under spatial rotations. You might worry that such a decomposition is contrary to the coordinate-independent spirit of general relativity, but it is really no different than decomposing the electromagnetic field strength tensor into electric and magnetic fields. Even though both **E** and **B** are components of a $(0, 2)$ tensor, it is nevertheless sometimes convenient to assume the role of some fixed observer and think of them as three-vectors.[1]

The metric perturbation is a $(0, 2)$ tensor, but symmetric rather than antisymmetric. Under spatial rotations, the 00 component is a scalar, the $0i$ components (equal to the $i0$ components) form a three-vector, and the ij components form a two-index symmetric spatial tensor. This spatial tensor can be further decomposed into a trace and a trace-free part. (In group theory language, we are looking for "irreducible representations" of the rotation group. In other words, we decompose the tensor into individual pieces, which transform only into themselves under spatial rotations.) We therefore write $h_{\mu\nu}$ as

$$\begin{aligned} h_{00} &= -2\Phi \\ h_{0i} &= w_i \\ h_{ij} &= 2s_{ij} - 2\Psi\delta_{ij}, \end{aligned} \tag{7.16}$$

where Ψ encodes the trace of h_{ij}, and s_{ij} is traceless:

$$\begin{aligned} \Psi &= -\tfrac{1}{6}\delta^{ij}h_{ij} \\ s_{ij} &= \tfrac{1}{2}\left(h_{ij} - \tfrac{1}{3}\delta^{kl}h_{kl}\delta_{ij}\right). \end{aligned} \tag{7.17}$$

The entire metric is thus written as

$$\boxed{ds^2 = -(1+2\Phi)dt^2 + w_i(dt\,dx^i + dx^i dt) + [(1-2\Psi)\delta_{ij} + 2s_{ij}]dx^i dx^j.} \tag{7.18}$$

[1] The discussion here follows that in E. Bertschinger, "Cosmological Dynamics," a talk given at Summer School on Cosmology and Large Scale Structure (Session 60), Les Houches, France, 1–28 Aug 1993; `http://arXiv.org/abs/astro-ph/9503125`. Bertschinger focuses on cosmological perturbation theory, in which spacelike hypersurfaces are expanding with time, but it is simple enough to specialize to the case of a nonexpanding universe.

We have not chosen a gauge or solved any equations, just defined some convenient notation. The traceless tensor s_{ij} is known as the *strain*, and will turn out to contain gravitational radiation. Sometimes the decomposition of the spatial components into trace and trace-free parts is not helpful, and we can just stick with h_{ij}; we will use whichever notation is appropriate in individual cases. Note that, just as in Chapter 1, the spatial metric is now simply δ_{ij}, and we can freely raise and lower spatial indices without changing the components.

To get a feeling for the physical interpretation of the different fields in the metric perturbation, we consider the motion of test particles as described by the geodesic equation. The Christoffel symbols for (7.18) are

$$\begin{aligned}
\Gamma^0_{00} &= \partial_0\Phi \\
\Gamma^i_{00} &= \partial_i\Phi + \partial_0 w_i \\
\Gamma^0_{j0} &= \partial_j\Phi \\
\Gamma^i_{j0} &= \partial_{[j}w_{i]} + \tfrac{1}{2}\partial_0 h_{ij} \\
\Gamma^0_{jk} &= -\partial_{(j}w_{k)} + \tfrac{1}{2}\partial_0 h_{jk} \\
\Gamma^i_{jk} &= \partial_{(j}h_{k)i} - \tfrac{1}{2}\partial_i h_{jk}.
\end{aligned} \tag{7.19}$$

In these expressions we have stuck with h_{ij} rather than s_{ij} and Ψ, since they enter only in the combination $h_{ij} = 2s_{ij} - 2\Psi\delta_{ij}$. The distinction will become important once we start taking traces to get to the Ricci tensor and Einstein's equation. Since we have fixed an inertial frame, it is convenient to express the four-momentum $p^\mu = dx^\mu/d\lambda$ (where $\lambda = \tau/m$ if the particle is massive) in terms of the energy E and three-velocity $v^i = dx^i/dt$, as

$$p^0 = \frac{dt}{d\lambda} = E, \quad p^i = Ev^i. \tag{7.20}$$

Then we can take the geodesic equation

$$\frac{dp^\mu}{d\lambda} + \Gamma^\mu_{\rho\sigma}p^\rho p^\sigma = 0, \tag{7.21}$$

move the second term to the right-hand side so that it takes on the appearance of a force term, and divide both sides by E to obtain

$$\frac{dp^\mu}{dt} = -\Gamma^\mu_{\rho\sigma}\frac{p^\rho p^\sigma}{E}. \tag{7.22}$$

The $\mu = 0$ component describes the evolution of the energy,

$$\frac{dE}{dt} = -E\left[\partial_0\Phi + 2(\partial_k\Phi)v^k - \left(\partial_{(j}w_{k)} - \frac{1}{2}\partial_0 h_{jk}\right)v^j v^k\right]. \tag{7.23}$$

You might think that the energy should be conserved, but $E = p^0 = m\gamma$ only includes the "inertial" energy of the particle—in the slowly-moving limit, the rest energy and the kinetic energy—and not the energy from interactions with the gravitational field.

The spatial components $\mu = i$ of the geodesic equation become

$$\frac{dp^i}{dt} = -E\left[\partial_i\Phi + \partial_0 w_i + 2(\partial_{[i}w_{j]} + \partial_0 h_{ij})v^j + \left(\partial_{(j}h_{k)i} - \frac{1}{2}\partial_i h_{jk}\right)v^j v^k\right]. \tag{7.24}$$

To interpret this physically, it is convenient to define the "gravito-electric" and "gravito-magnetic" three-vector fields,

$$\begin{aligned} G^i &\equiv -\partial_i\Phi - \partial_0 w_i \\ H^i &\equiv (\nabla \times \vec{w})^i = \epsilon^{ijk}\partial_j w_k, \end{aligned} \tag{7.25}$$

which bear an obvious resemblance to the definitions of the ordinary electric and magnetic field in terms of a scalar and vector potential. Then (7.24) becomes

$$\frac{dp^i}{dt} = E\left[G^i + (\vec{v} \times H)^i - 2(\partial_0 h_{ij})v^j - \left(\partial_{(j}h_{k)i} - \frac{1}{2}\partial_i h_{jk}\right)v^j v^k\right]. \tag{7.26}$$

The first two terms on the right-hand side describe how the test particle, moving along a geodesic, responds to the scalar and vector perturbations Φ and w_i in a way reminiscent of the Lorentz force law in electromagnetism. We also find couplings to the spatial perturbations h_{ij}, of linear and quadratic order in the three-velocity. The relative importance of the different perturbations will of course depend on the physical situation under consideration, as we will soon demonstrate.

In addition to the motion of test particles, we should examine the field equations for the metric perturbations, which are of course the linearized Einstein equations. The Riemann tensor in our variables is

$$\begin{aligned} R_{0j0l} &= \partial_j\partial_l\Phi + \partial_0\partial_{(j}w_{l)} - \tfrac{1}{2}\partial_0\partial_0 h_{jl} \\ R_{0jkl} &= \partial_j\partial_{[k}w_{l]} - \partial_0\partial_{[k}h_{l]j} \\ R_{ijkl} &= \partial_j\partial_{[k}h_{l]i} - \partial_i\partial_{[k}h_{l]j}, \end{aligned} \tag{7.27}$$

with other components related by symmetries. We contract using $\eta^{\mu\nu}$ to obtain the Ricci tensor,

$$\begin{aligned} R_{00} &= \nabla^2\Phi + \partial_0\partial_k w^k + 3\partial_0^2\Psi \\ R_{0j} &= -\tfrac{1}{2}\nabla^2 w_j + \tfrac{1}{2}\partial_j\partial_k w^k + 2\partial_0\partial_j\Psi + \partial_0\partial_k s_j{}^k \\ R_{ij} &= -\partial_i\partial_j(\Phi - \Psi) - \partial_0\partial_{(i}w_{j)} + \Box\Psi\delta_{ij} - \Box s_{ij} + 2\partial_k\partial_{(i}s_{j)}{}^k, \end{aligned} \tag{7.28}$$

where $\nabla^2 = \delta^{ij}\partial_i\partial_j$ is the three-dimensional flat Laplacian. Since the Ricci tensor involves contractions, the trace-free and trace parts of the spatial perturbations now enter in different ways. Finally, we can calculate the Einstein tensor,

$$\begin{aligned} G_{00} &= 2\nabla^2\Psi + \partial_k\partial_l s^{kl} \\ G_{0j} &= -\tfrac{1}{2}\nabla^2 w_j + \tfrac{1}{2}\partial_j\partial_k w^k + 2\partial_0\partial_j\Psi + \partial_0\partial_k s_j{}^k \\ G_{ij} &= (\delta_{ij}\nabla^2 - \partial_i\partial_j)(\Phi - \Psi) + \delta_{ij}\partial_0\partial_k w^k - \partial_0\partial_{(i}w_{j)} \\ &\quad + 2\delta_{ij}\partial_0^2\Psi - \Box s_{ij} + 2\partial_k\partial_{(i}s_{j)}{}^k - \delta_{ij}\partial_k\partial_l s^{kl}. \end{aligned} \tag{7.29}$$

Using this expression in Einstein's equation $G_{\mu\nu} = 8\pi G T_{\mu\nu}$ reveals that only a small fraction of the metric components are true degrees of freedom of the gravitational field; the rest obey constraints that determine them in terms of the other fields. To see this, start with $G_{00} = 8\pi G T_{00}$, which we write using (7.29) as

$$\nabla^2\Psi = 4\pi G T_{00} - \tfrac{1}{2}\partial_k\partial_l s^{kl}. \tag{7.30}$$

This is an equation for Ψ with no time derivatives; if we know what T_{00} and s_{ij} are doing at any time, we can determine what Ψ must be (up to boundary conditions at spatial infinity). Thus, Ψ is not by itself a propagating degree of freedom; it is determined by the energy-momentum tensor and the gravitational strain s_{ij}. Next turn to the $0j$ equation, which we write as

$$(\delta_{jk}\nabla^2 - \partial_j\partial_k)w^k = -16\pi G T_{0j} + 4\partial_0\partial_j\Psi + 2\partial_0\partial_k s_j{}^k. \tag{7.31}$$

This is an equation for w^i with no time derivatives; once again, if we know the energy-momentum tensor and the strain (from which we can find Ψ), the vector w^i will be determined. Finally, the ij equation is

$$\begin{aligned} (\delta_{ij}\nabla^2 - \partial_i\partial_j)\Phi &= 8\pi G T_{ij} + (\delta_{ij}\nabla^2 - \partial_i\partial_j - 2\delta_{ij}\partial_0^2)\Psi \\ &\quad - \delta_{ij}\partial_0\partial_k w^k + \partial_0\partial_{(i}w_{j)} + \Box s_{ij} - 2\partial_k\partial_{(i}s_{j)}{}^k - \delta_{ij}\partial_k\partial_l s^{jl}. \end{aligned} \tag{7.32}$$

Once again, we see that there are no time derivatives acting on Φ, which is therefore determined as a function of the other fields.

Thus, the only propagating degrees of freedom in Einstein's equations are those in the strain tensor s_{ij}; as we will see, these are used to describe gravitational waves. The other components of $h_{\mu\nu}$ are determined in terms of s_{ij} and the matter fields—they do not require separate initial data. In alternative theories, such as those discussed in Section 4.8 with either additional fields or higher-order terms in the action, the other components of the metric may become dynamical variables. As we discuss briefly at the end of Section 7.4, propagating tensor fields give rise upon quantization to particles of different spins, depending on the behavior of the field under spatial rotations. Thus, the scalars Φ and Ψ would be spin-0, the vector w_i would be spin-1, and the tensor s_{ij} is spin-2. Only the spin-2 piece is a true particle excitation in ordinary GR.

In the previous section we showed how gauge transformations $h_{\mu\nu} \to h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu$ are generated by a vector field ξ^μ. Henceforth we set the parameter ϵ of (7.14) equal to unity, and think of the vector field ξ^μ itself as being small. Under such a transformation, the different metric perturbation fields change by

$$\begin{aligned}
\Phi &\to \Phi + \partial_0 \xi^0 \\
w_i &\to w_i + \partial_0 \xi^i - \partial_i \xi^0 \\
\Psi &\to \Psi - \tfrac{1}{3}\partial_i \xi^i \\
s_{ij} &\to s_{ij} + \partial_{(i}\xi_{j)} - \tfrac{1}{3}\partial_k \xi^k \delta_{ij},
\end{aligned} \tag{7.33}$$

as you can easily check. Just as in electromagnetism and other gauge theories, different gauges can be appropriate to different circumstances; here we list some popular choices.

Consider first the **transverse gauge** (a generalization of the conformal Newtonian or Poisson gauge sometimes used in cosmology.) The transverse gauge is closely related to the Coulomb gauge of electromagnetism, $\partial_i A^i = 0$. We begin by fixing the strain to be spatially transverse,

$$\partial_i s^{ij} = 0, \tag{7.34}$$

by choosing ξ^j to satisfy

$$\nabla^2 \xi^j + \tfrac{1}{3}\partial_j \partial_i \xi^i = -2\partial_i s^{ij}. \tag{7.35}$$

The value of ξ^0 is still undetermined, so we can use this remaining freedom to render the vector perturbation transverse,

$$\partial_i w^i = 0, \tag{7.36}$$

by choosing ξ^0 to satisfy

$$\nabla^2 \xi^0 = \partial_i w^i + \partial_0 \partial_i \xi^i. \tag{7.37}$$

The meaning of transverse becomes clear upon taking the Fourier transform, after which a vanishing divergence implies that a tensor is orthogonal to the wave vector. Neither (7.35) nor (7.37) completely fixes the value of ξ^μ; they are both second-order differential equations in spatial derivatives, which require boundary conditions to specify a solution. For our present purposes, it suffices that solutions will always exist. The conditions (7.34) and (7.36) together define the transverse gauge. In this gauge, Einstein's equation becomes

$$\boxed{G_{00} = 2\nabla^2 \Psi = 8\pi G T_{00},} \tag{7.38}$$

$$\boxed{G_{0j} = -\tfrac{1}{2}\nabla^2 w_j + 2\partial_0\partial_j\Psi = 8\pi G T_{0j},} \tag{7.39}$$

and

$$\boxed{G_{ij} = (\delta_{ij}\nabla^2 - \partial_i\partial_j)(\Phi - \Psi) - \partial_0\partial_{(i}w_{j)} + 2\delta_{ij}\partial_0^2\Psi - \Box s_{ij} = 8\pi G T_{ij}.} \tag{7.40}$$

In the remainder of this chapter, we will use these equations to find weak-field solutions in different situations.

Another popular gauge is known as the **synchronous gauge**. It is equivalent to the choice of Gaussian normal coordinates, discussed in Appendix D. It may be thought of as the gravitational analogue of the temporal gauge of electromagnetism, $A^0 = 0$, since it kills off the nonspatial components of the perturbation. We begin by setting the scalar potential Φ to vanish,

$$\Phi = 0, \tag{7.41}$$

by choosing ξ^0 to satisfy

$$\partial_0\xi^0 = -\Phi. \tag{7.42}$$

This leaves us the ability to choose ξ^i. We can set the vector components to zero,

$$w^i = 0, \tag{7.43}$$

by choosing ξ^i to satisfy

$$\partial_0\xi^i = -w^i + \partial_i\xi^0. \tag{7.44}$$

The metric in synchronous gauge therefore takes on the attractive form

$$ds^2 = -dt^2 + (\delta_{ij} + h_{ij})dx^i dx^j. \tag{7.45}$$

This is just a matter of gauge choice, and is applicable to any spacetime slightly perturbed away from Minkowski. It is straightforward to write down Einstein's equation in synchronous gauge, but we won't bother as we won't actually be using it in the rest of this chapter.

In addition to transverse and synchronous gauges, in calculating the production of gravitational waves it is convenient to use yet a third choice, the Lorenz/harmonic gauge. As we will discuss below, it is equivalent to setting

$$\partial_\mu h^\mu{}_\nu - \tfrac{1}{2}\partial_\nu h = 0, \tag{7.46}$$

where $h = \eta^{\mu\nu}h_{\mu\nu}$. This gauge does not have any especially simple expression in terms of our decomposed perturbation fields, but it does make the linearized Einstein equation take on a particularly simple form.

Before moving on to applications of the weak-field limit, we conclude our discussion of degrees of freedom by drawing attention to the distinction between our *algebraic* decomposition of the metric perturbation components in (7.16), and an additional decomposition that becomes possible if we consider tensor *fields* rather than tensors defined at a point. This additional decomposition helps to bring out the physical degrees of freedom more directly, and is crucial in cosmological perturbation theory. Its basis is the standard observation that a vector field can be decomposed into a transverse part $w^i_\perp$ and a longitudinal part $w^i_\parallel$:

$$w^i = w^i_\perp + w^i_\parallel, \tag{7.47}$$

where a transverse vector is divergenceless and a longitudinal vector is curl-free,

$$\partial_i w^i_\perp = 0, \qquad \epsilon^{ijk}\partial_j w_{\parallel k} = 0. \tag{7.48}$$

Notice that these are differential equations, so clearly they only make sense when applied to tensor fields. A transverse vector can be represented as the curl of some other vector ξ^i, although the choice of ξ^i is not unique unless we impose a subsidiary condition such as $\partial_i\xi^i = 0$. A longitudinal vector is the divergence of a scalar λ,

$$w^i_\perp = \epsilon^{ijk}\partial_j\xi_k, \qquad w_{\parallel i} = \partial_i\lambda. \tag{7.49}$$

Just like our original decomposition of the metric perturbation into scalar, vector and tensor pieces, this decomposition of a vector field into parts depending on a scalar and a transverse vector is invariant under spatial rotations. The scalar λ clearly represents one degree of freedom; the vector ξ^i looks like three degrees of freedom, but one of these is illusory due to the nonuniqueness of the choice of ξ^i (which you will notice is equivalent to the freedom to make gauge transformations $\xi_i \to \xi_i + \partial_i\omega$). There are thus three degrees of freedom in total, as there should be to describe the original vector field w^i.

A similar procedure applies to the traceless symmetric tensor s^{ij}, which can be decomposed into a transverse part $s^{ij}_\perp$, a solenoidal part $s^{ij}_{\rm S}$, and a longitudinal part $s^{ij}_\parallel$,

$$s^{ij} = s^{ij}_\perp + s^{ij}_{\rm S} + s^{ij}_\parallel. \tag{7.50}$$

The transverse part is divergenceless, while the divergence of the solenoidal part is a transverse (divergenceless) vector, and the divergence of the longitudinal part is a longitudinal (curl-free) vector:

$$\partial_i s^{ij}_\perp = 0$$

$$\partial_i \partial_j s_{\mathrm{S}}^{ij} = 0$$
$$\epsilon^{jkl} \partial_k \partial_i s_{\parallel}{}^{i}{}_{j} = 0. \tag{7.51}$$

This means that the longitudinal part can be derived from a scalar field θ, and the solenoidal part can be derived from a transverse vector ζ^i,

$$s_{\parallel ij} = \left(\partial_i \partial_j - \tfrac{1}{3}\delta_{ij}\nabla^2\right)\theta$$
$$s_{\mathrm{S}ij} = \partial_{(i}\zeta_{j)}, \tag{7.52}$$

where

$$\partial_i \zeta^i = 0. \tag{7.53}$$

Thus, the longitudinal part describes a single degree of freedom, while the solenoidal part describes two degrees of freedom. The transverse part cannot be further decomposed; it describes the remaining two degrees of freedom of the symmetric traceless 3×3 tensor s_{ij}. Later in this chapter we will introduce the transverse-traceless gauge for describing gravitational waves propagating in vacuum; in this gauge, the only nonvanishing metric perturbation is the transverse tensor perturbation $s_{\perp}^{ij}$.

With this decomposition of tensor fields, we have succeeded in writing the original ten-component metric perturbation $h_{\mu\nu}$ in terms of four scalars (Φ, Ψ, λ, and θ) with one degree of freedom each, two transverse vectors (ξ^i and ζ^i) with two degrees of freedom each, and one transverse-traceless tensor ($s_{\perp}^{ij}$) with two degrees of freedom. People refer to this set of fields when they speak of "scalar," "vector," and "tensor" modes. We can then decompose the energy-momentum tensor in a similar way, write Einstein's equation in terms of these variables, and isolate the physical (gauge-invariant) degrees of freedom. We won't use this decomposition in this book, but you should be aware of its existence when referring to the literature.

7.3 ■ NEWTONIAN FIELDS AND PHOTON TRAJECTORIES

We previously defined the "Newtonian limit" as describing weak fields for which sources were static and test particles were slowly moving. In this section we will extend this definition somewhat, still restricting ourselves to static sources but allowing the test particles to move at any velocity. There is clearly an important difference, as we previously only needed to consider effects of the g_{00} component of the metric, but we will find that relativistic particles respond to spatial components of the metric as well.

We can model our static gravitating sources by dust, a perfect fluid for which the pressure vanishes. (Most of the matter in the universe is well approximated by dust, including stars, planets, galaxies, and even dark matter.) We work in the rest frame of the dust, where the energy-momentum tensor takes the form

$$T_{\mu\nu} = \rho U_\mu U_\nu = \begin{pmatrix} \rho & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{pmatrix}. \tag{7.54}$$

Since our background is flat Minkowski space, it is straightforward to accommodate moving sources by simply Lorentz-transforming into their rest frame; what we are unable to deal with in this limit is multiple sources with large relative velocities.

Turn to Einstein's equation in the transverse gauge, (7.38)–(7.40). For static sources we drop all time-derivative terms, and simultaneously plug in the energy-momentum tensor (7.54), to obtain

$$\begin{aligned} \nabla^2\Psi &= 4\pi G\rho \\ \nabla^2 w_j &= 0 \\ (\delta_{ij}\nabla^2 - \partial_i\partial_j)(\Phi - \Psi) - \nabla^2 s_{ij} &= 0. \end{aligned} \tag{7.55}$$

We will look for solutions that are both nonsingular and well-behaved at infinity; consequently, only those fields that are sourced by the right-hand side will be nonvanishing. For example, the second equation in (7.55) immediately implies $w^i = 0$. We next take the trace of the third equation (summing over δ^{ij}):

$$2\nabla^2(\Phi - \Psi) = 0. \tag{7.56}$$

This enforces equality of the two scalar potentials,

$$\Phi = \Psi. \tag{7.57}$$

Recall that in our initial discussion of the Newtonian limit in Chapter 4, we argued that the 00 component Φ of the perturbation (which is responsible for the motion of nonrelativistic particles) obeyed the Poisson equation; from (7.55) it appears as if it is actually the scalar perturbation Ψ to the spatial components that obeys this equation. The implicit connection is provided by (7.56), which sets the two potentials equal when the trace of T_{ij} (the sum of the three principle pressures) vanishes. Finally we can plug $\Phi = \Psi$ into the last equation of (7.55) to get

$$\nabla^2 s_{ij} = 0, \tag{7.58}$$

which implies $s_{ij} = 0$ for a well-behaved solution.

The perturbed metric for static Newtonian sources is therefore

$$\boxed{ds^2 = -(1 + 2\Phi)dt^2 + (1 - 2\Phi)(dx^2 + dy^2 + dz^2),} \tag{7.59}$$

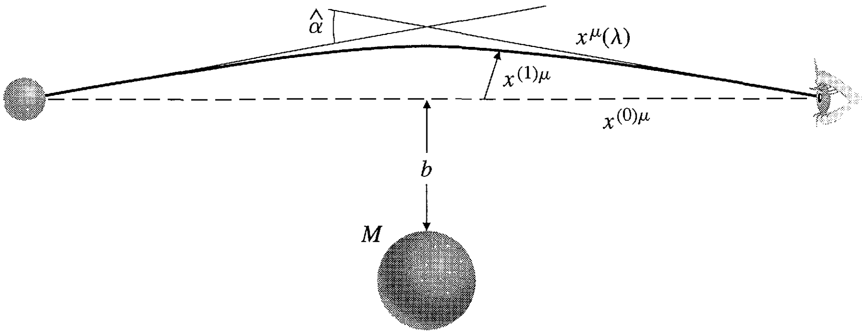

FIGURE 7.3 A deflected geodesic $x^\mu(\lambda)$, decomposed into a background geodesic $x^{(0)\mu}$ and a perturbation $x^{(1)\mu}$. The deflection angle $\widehat{\alpha}$ represents (minus) the amount by which the wave vector rotates along the path. A single mass M with impact parameter b is depicted, although the setup is more general.

or equivalently

$$h_{\mu\nu} = \begin{pmatrix} -2\Phi & & & \\ & -2\Phi & & \\ & & -2\Phi & \\ & & & -2\Phi \end{pmatrix}, \tag{7.60}$$

where the potential obeys the conventional Poisson equation,

$$\nabla^2\Phi = 4\pi G\rho. \tag{7.61}$$

This is an important extension of our result from Chapter 4, since we now know the perturbation of the spatial metric as well as h_{00}.

Now let us consider the path of a photon (or other massless particle) through this geometry; in other words, solve the perturbed geodesic equation for a null trajectory $x^\mu(\lambda)$.[2] The geometry we consider is portrayed in Figure 7.3. Recall that our philosophy is to consider the metric perturbation as a field defined on a flat background spacetime. Similarly, we can decompose the geodesic into a background path plus a perturbation,

$$x^\mu(\lambda) = x^{(0)\mu}(\lambda) + x^{(1)\mu}(\lambda), \tag{7.62}$$

where $x^{(0)\mu}$ solves the geodesic equation in the background (in other words, is just a straight null path). We then evaluate all quantities along the background path, to solve for $x^{(1)\mu}(\lambda)$. For this procedure to make sense, we need to assume that the potential Φ is not appreciably different along the background and true geodesics; this condition amounts to requiring that $x^{(1)i}\partial_i\Phi \ll \Phi$. If this condition is not true, however, all is not lost. If we consider only very short paths, the deviation

[2]The approach we use is outlined in T. Pyne and M. Birkinshaw, *Astrophys. Journ.* **458**, 46 (1996), `http://arxiv.org/abs/astro-ph/9504060`.

$x^{(1)\mu}$ will necessarily be small, and our approximation will be valid. But then we can assemble larger paths out of such short segments. As a result, we will derive true equations, but the paths over which we integrate will be the *actual* path $x^\mu(\lambda)$, rather than the background path $x^{(0)\mu}(\lambda)$. As long as this is understood, our results will be valid for any trajectories in the perturbed spacetime.

For convenience we denote the wave vector of the background path as k^μ, and the derivative of the deviation vector as ℓ^μ:

$$k^\mu \equiv \frac{dx^{(0)\mu}}{d\lambda}, \quad \ell^\mu \equiv \frac{dx^{(1)\mu}}{d\lambda}. \tag{7.63}$$

The condition that a path be null is of course

$$g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = 0, \tag{7.64}$$

which we must solve order-by-order. At zeroth order we simply have $\eta_{\mu\nu}k^\mu k^\nu = 0$, or

$$(k^0)^2 = (\vec{k})^2 \equiv k^2, \tag{7.65}$$

where $\vec{k}$ is the three-vector with components k^i. This equation serves as the definition of the constant k. Then at first order we obtain

$$2\eta_{\mu\nu}k^\mu \ell^\nu + h_{\mu\nu}k^\mu k^\nu = 0, \tag{7.66}$$

or

$$-k\ell^0 + \vec{\ell}\cdot\vec{k} = 2k^2\Phi. \tag{7.67}$$

We now turn to the perturbed geodesic equation,

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda} = 0. \tag{7.68}$$

The Christoffel symbols can be found by setting $w^i = 0$ and $h_{ij} = -2\Phi\delta_{ij}$ in (7.19):

$$\begin{aligned} \Gamma^0_{0i} &= \Gamma^i_{00} = \partial_i\Phi, \\ \Gamma^i_{jk} &= \delta_{jk}\partial_i\Phi - \delta_{ik}\partial_j\Phi - \delta_{ij}\partial_k\Phi. \end{aligned} \tag{7.69}$$

The zeroth-order geodesic equation simply tells us that $x^{(0)\mu}$ is a straight trajectory, while at first order we have

$$\frac{d\ell^\mu}{d\lambda} = -\Gamma^\mu_{\rho\sigma}k^\rho k^\sigma. \tag{7.70}$$

There are no factors of ℓ^μ on the right-hand side, since the Christoffel symbols are already first-order in the perturbation. The $\mu = 0$ component of (7.70) is

$$\frac{d\ell^0}{d\lambda} = -2k(\vec{k} \cdot \vec{\nabla}\Phi), \tag{7.71}$$

while the spatial components are

$$\frac{d\vec{\ell}}{d\lambda} = -2k^2\vec{\nabla}_\perp \Phi. \tag{7.72}$$

Here we have introduced the gradient transverse to the path, defined as the total gradient minus the gradient along the path,

$$\begin{aligned} \vec{\nabla}_\perp \Phi &\equiv \vec{\nabla}\Phi - \vec{\nabla}_\parallel \Phi \\ &= \vec{\nabla}\Phi - k^{-2}(\vec{k} \cdot \vec{\nabla}\Phi)\vec{k}. \end{aligned} \tag{7.73}$$

In all of these expressions, the path means the background path.

Note that, to first order in Φ, the spatial wave vector perturbation $\vec{\ell}$ is orthogonal to the original spatial wave vector $\vec{k}$. To see this, we can get an expression for ℓ^0 by integrating (7.71) to get

$$\begin{aligned} \ell^0 &= \int \frac{d\ell^0}{d\lambda} d\lambda \\ &= -2k \int (\vec{k} \cdot \vec{\nabla}\Phi) d\lambda \\ &= -2k \int \left(\frac{d\vec{x}}{d\lambda} \cdot \vec{\nabla}\Phi\right) d\lambda \\ &= -2k \int \vec{\nabla}\Phi \cdot d\vec{x} \\ &= -2k\Phi. \end{aligned} \tag{7.74}$$

The constant of integration is fixed by demanding that $\ell^0 = 0$ when $\Phi = 0$. Plugging this into (7.67) reveals

$$\vec{\ell} \cdot \vec{k} = k\ell^0 + 2k^2\Phi = 0, \tag{7.75}$$

verifying that $\vec{\ell}$ and $\vec{k}$ are orthogonal to first order.

The **deflection angle** $\hat{\alpha}$ is the amount by which the original spatial wave vector is deflected as it travels from a source to the observer; it is a two-dimensional vector in the plane perpendicular to $\vec{k}$. (We use the notation $\hat{\alpha}$ rather than $\vec{\alpha}$, as the latter is used for the reduced deflection angle introduced in Chapter 8.) From the geometry portrayed in Figure 7.3, the deflection angle can be expressed as

$$\hat{\vec{\alpha}} = -\frac{\Delta\vec{\ell}}{k}, \tag{7.76}$$

where the minus sign simply accounts for the fact that the deflection angle is measured by an observer looking backward along the photon path. The rotation of the wave vector can be calculated from (7.72) as

$$\begin{aligned}\Delta\vec{\ell} &= \int \frac{d\vec{\ell}}{d\lambda} d\lambda \\ &= -2k^2 \int \vec{\nabla}_\perp \Phi \, d\lambda. \end{aligned} \tag{7.77}$$

The deflection angle can therefore be expressed as an integral over the physical spatial distance traversed, $s = k\lambda$, as

$$\boxed{\hat{\vec{\alpha}} = 2 \int \vec{\nabla}_\perp \Phi \, ds.} \tag{7.78}$$

We can evaluate the deflection angle in the case of a point mass, where we imagine the background path to be along the x-direction with an impact parameter defined by a transverse vector $\vec{b}$ pointing from the path to the mass at the point of closest approach. Setting $b = |\vec{b}|$, the potential is

$$\Phi = -\frac{GM}{r} = -\frac{GM}{(b^2 + x^2)^{1/2}}, \tag{7.79}$$

and its transverse gradient is therefore

$$\vec{\nabla}_\perp \Phi = \frac{GM}{(b^2 + x^2)^{3/2}} \vec{b}. \tag{7.80}$$

The deflection angle is thus

$$\begin{aligned}\hat{\alpha} &= 2GMb \int \frac{dx}{(b^2 + x^2)^{3/2}} \\ &= \frac{4GM}{b}, \end{aligned} \tag{7.81}$$

where the integral has been taken from $-\infty$ to ∞, presuming that both source and observer are very far from the deflecting mass. Note that $c = 1$ in our units; a factor of c^2 should be inserted in the denominator of (7.81) in other systems.

Deflection of light by the Sun was historically a crucial test of general relativity. Einstein proposed three such tests: precession of the perihelion of Mercury, gravitational redshift, and deflection of light. The precession of Mercury's perihelion was successfully explained by GR, but this explained a discrepancy that had already been observed; gravitational redshift was not observed until much later, so deflection of light was the first time that Einstein's theory correctly predicted a phenomenon that had not yet been detected. A famous expedition led by Eddington observed the positions of stars near the Sun during a 1919 total eclipse; the observations were in agreement with the GR prediction, leading to front-page stories in newspapers around the world. The predicted effect is quite small: for the Sun we have $GM_\odot/c^2 = 1.48 \times 10^5$ cm, and the solar radius is $R_\odot = 6.96 \times 10^{10}$ cm, leading to a maximum deflection angle of $\widehat{\alpha} = 1.75$ arcsecs. Later re-evaluation of Eddington's results has cast doubt upon whether he actually obtained the precision that was originally claimed; contemporary measurements use high-precision interferometric observations of quasars passing behind the Sun to obtain very accurate tests of GR (which it has so far passed). Meanwhile, observation of light deflection by astrophysical sources such as galaxies and stars has become a vibrant area of research, under the name of "gravitational lensing." Of course in these circumstances we rarely know the mass of the lens well enough to provide precision tests of GR; instead, it is more common to use the observed deflection angle as a way to measure the mass. We will discuss lensing more in Chapter 8.

In addition to the deflection of light, in 1964 Shapiro pointed out another observable consequence of weak-field general relativity on photon trajectories: gravitational time delay. The total coordinate time elapsed along a null path is

$$t = \int \frac{dx^0}{d\lambda}\, d\lambda. \tag{7.82}$$

We are putting ourselves in the position of an observer far from any sources, at rest in the background inertial frame, so coordinate time is our proper time. In the presence of a Newtonian potential, the photons appear to "slow down" with respect to the background light cones, leading to an additional time delay of

$$\begin{aligned}\Delta t &= \int \frac{dx^{(1)0}}{d\lambda}\, d\lambda \\ &= \int \ell^0\, d\lambda \\ &= -2k \int \Phi\, d\lambda,\end{aligned} \tag{7.83}$$

or

$$\Delta t = -2 \int \Phi\, ds. \tag{7.84}$$

According to our rules, the integral is performed over the background path. In addition to this Shapiro delay, there can be an additional "geometric" time delay because the spatial distance traversed by the real path is longer than that of the background path. For deflection of light by the Sun the geometric delay effect is negligible, but in cosmological applications it can be comparable to the Shapiro effect. The time delay has been observed, most precisely by making use of spacecraft rather than naturally-occurring objects; for details see Will (1981).

The motion of photons through a Newtonian potential, leading to both the deflection of light and the gravitational time delay, could equivalently be derived by imagining that the photons are propagating in a medium with refractive index

$$n = 1 - 2\Phi, \tag{7.85}$$

to first order. Indeed, we could have found the equations of motion for the photon by using Fermat's principle of least time; you are asked to demonstrate this in the exercises.

7.4 ■ GRAVITATIONAL WAVE SOLUTIONS

An even more exciting application of the weak-field limit is to gravitational radiation. Here we are studying the freely-propagating degrees of freedom of the gravitational field, requiring no local sources for their existence (although they can of course be generated by such sources). We therefore turn once again to the weak-field equations in transverse gauge, (7.38)–(7.40), this time keeping time derivatives but completely turning off the energy-momentum tensor, $T_{\mu\nu} = 0$. The 00 equation is then

$$\nabla^2\Psi = 0, \tag{7.86}$$

which with well-behaved boundary conditions implies $\Psi = 0$. Then the $0j$ equation is

$$\nabla^2 w_j = 0, \tag{7.87}$$

which again implies $w_j = 0$.

We turn next to the trace of the ij equation, which (plugging in the above results) yields

$$\nabla^2\Phi = 0, \tag{7.88}$$

which implies $\Phi = 0$.

We are therefore left with the trace-free part of the ij equation, which becomes a wave equation for the traceless strain tensor:

$$\Box s_{ij} = 0. \tag{7.89}$$

Although it has been convenient thus far to work with s_{ij}, it is far more common in the literature to find expressions written in terms of the entire metric perturbation $h_{\mu\nu}$, but in an ansatz where all of the other degrees of freedom (Φ, Ψ, w_i) are set to zero (and s_{ij} is transverse). This is commonly known as the **transverse**

traceless gauge, in which we have

$$h_{\mu\nu}^{\mathrm{TT}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & 2s_{ij} & \\ 0 & & & \end{pmatrix}. \tag{7.90}$$

The equation of motion is then

$$\Box h_{\mu\nu}^{\mathrm{TT}} = 0. \tag{7.91}$$

To make it easier to compare with other resources, in our discussion of gravitational waves we will use $h_{\mu\nu}^{\mathrm{TT}}$ rather than s_{ij}, keeping in mind that $h_{\mu\nu}^{\mathrm{TT}}$ is purely spatial, traceless and transverse:

$$\begin{aligned} h_{0\nu}^{\mathrm{TT}} &= 0 \\ \eta^{\mu\nu} h_{\mu\nu}^{\mathrm{TT}} &= 0 \\ \partial_\mu h_{\mathrm{TT}}^{\mu\nu} &= 0. \end{aligned} \tag{7.92}$$

From the wave equation (7.91) we begin finding solutions. Those familiar with the analogous problem in electromagnetism will notice that the procedure is almost precisely the same. A particularly useful set of solutions to this wave equation are the plane waves, given by

$$h_{\mu\nu}^{\mathrm{TT}} = C_{\mu\nu} e^{ik_\sigma x^\sigma}, \tag{7.93}$$

where $C_{\mu\nu}$ is a constant, symmetric, $(0, 2)$ tensor, which is obviously traceless and purely spatial:

$$\begin{aligned} C_{0\nu} &= 0 \\ \eta^{\mu\nu} C_{\mu\nu} &= 0. \end{aligned} \tag{7.94}$$

Of course $e^{ik_\sigma x^\sigma}$ is complex, while $h_{\mu\nu}^{\mathrm{TT}}$ is real; we carry both real and imaginary parts through the calculation, and take the real part at the end. The constant vector k^σ is the wave vector. To check that we have a solution, we plug in:

$$\begin{aligned} 0 &= \Box h_{\mu\nu}^{\mathrm{TT}} \\ &= \eta^{\rho\sigma} \partial_\rho \partial_\sigma h_{\mu\nu}^{\mathrm{TT}} \\ &= \eta^{\rho\sigma} \partial_\rho (ik_\sigma h_{\mu\nu}^{\mathrm{TT}}) \\ &= -\eta^{\rho\sigma} k_\rho k_\sigma h_{\mu\nu}^{\mathrm{TT}} \\ &= -k_\sigma k^\sigma h_{\mu\nu}^{\mathrm{TT}}. \end{aligned} \tag{7.95}$$

Since, for an interesting solution, not all of the components of $h^{\mathrm{TT}}_{\mu\nu}$ will be zero everywhere, we must have

$$k_\sigma k^\sigma = 0. \tag{7.96}$$

The plane wave (7.93) is therefore a solution to the linearized equation if the wave vector is null; this is loosely translated into the statement that gravitational waves propagate at the speed of light. The timelike component of the wave vector is the frequency of the wave, and we write $k^\sigma = (\omega, k^1, k^2, k^3)$. (More generally, an observer moving with four-velocity U^μ would observe the wave to have a frequency $\omega = -k_\mu U^\mu$.) Then the condition that the wave vector be null becomes

$$\omega^2 = \delta_{ij} k^i k^j. \tag{7.97}$$

Of course our wave is far from the most general solution; any (possibly infinite) number of distinct plane waves can be added together and will still solve the linear equation (7.91). Indeed, any solution can be written as such a superposition.

We still need to ensure that the perturbation is transverse. This means that

$$\begin{aligned} 0 &= \partial_\mu h^{\mu\nu}_{\mathrm{TT}} \\ &= i C^{\mu\nu} k_\mu e^{i k_\sigma x^\sigma}, \end{aligned} \tag{7.98}$$

which is only true if

$$k_\mu C^{\mu\nu} = 0. \tag{7.99}$$

We say that the wave vector is orthogonal to $C^{\mu\nu}$.

Our solution can be made more explicit by choosing spatial coordinates such that the wave is traveling in the x^3 direction; that is,

$$k^\mu = (\omega, 0, 0, k^3) = (\omega, 0, 0, \omega), \tag{7.100}$$

where we know that $k^3 = \omega$ because the wave vector is null. In this case, $k^\mu C_{\mu\nu} = 0$ and $C_{0\nu} = 0$ together imply

$$C_{3\nu} = 0. \tag{7.101}$$

The only nonzero components of $C_{\mu\nu}$ are therefore C_{11}, C_{12}, C_{21}, and C_{22}. But $C_{\mu\nu}$ is traceless and symmetric, so in general we can write

$$C_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & C_{11} & C_{12} & 0 \\ 0 & C_{12} & -C_{11} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{7.102}$$

Thus, for a plane wave in this gauge traveling in the x^3 direction, the two components C_{11} and C_{12} (along with the frequency ω) completely characterize the wave.

To get a feeling for the physical effect of a passing gravitational wave, consider the motion of test particles in the presence of a wave. It is certainly insufficient to solve for the trajectory of a single particle, since that would only tell us about the values of the coordinates along the world line. In fact, for any single particle we can find transverse traceless coordinates in which the particle appears stationary to first order in $h^{\rm TT}_{\mu\nu}$. To obtain a coordinate-independent measure of the wave's effects, we consider the relative motion of nearby particles, as described by the geodesic deviation equation. If we consider some nearby particles with four-velocities described by a single vector field $U^\mu(x)$ and separation vector S^μ, we have

$$\frac{D^2}{d\tau^2}S^\mu = R^\mu{}_{\nu\rho\sigma}U^\nu U^\rho S^\sigma . \tag{7.103}$$

We would like to compute the right-hand side to first order in $h^{\rm TT}_{\mu\nu}$. If we take our test particles to be moving slowly, we can express the four-velocity as a unit vector in the time direction plus corrections of order $h^{\rm TT}_{\mu\nu}$ and higher; but we know that the Riemann tensor is already first order, so the corrections to U^ν may be ignored, and we write

$$U^\nu = (1, 0, 0, 0). \tag{7.104}$$

Therefore we only need to compute $R^\mu{}_{00\sigma}$, or equivalently $R_{\mu 00\sigma}$. From (7.5) we have

$$R_{\mu 00\sigma} = \tfrac{1}{2}(\partial_0\partial_0 h^{\rm TT}_{\mu\sigma} + \partial_\sigma\partial_\mu h^{\rm TT}_{00} - \partial_\sigma\partial_0 h^{\rm TT}_{\mu 0} - \partial_\mu\partial_0 h^{\rm TT}_{\sigma 0}). \tag{7.105}$$

But $h^{\rm TT}_{\mu 0} = 0$, so

$$R_{\mu 00\sigma} = \tfrac{1}{2}\partial_0\partial_0 h^{\rm TT}_{\mu\sigma}. \tag{7.106}$$

Meanwhile, for our slowly-moving particles we have $\tau = x^0 = t$ to lowest order, so the geodesic deviation equation becomes

$$\frac{\partial^2}{\partial t^2}S^\mu = \frac{1}{2}S^\sigma \frac{\partial^2}{\partial t^2}h^{{\rm TT}\mu}{}_\sigma . \tag{7.107}$$

For our wave traveling in the x^3 direction, this implies that only S^1 and S^2 will be affected—the test particles are only disturbed in directions perpendicular to the wave vector. This is of course familiar from electromagnetism, where the electric and magnetic fields in a plane wave are perpendicular to the wave vector.

Our wave is characterized by the two numbers, which for future convenience we will rename as follows:

$$\begin{aligned} h_+ &= C_{11} \\ h_\times &= C_{12}, \end{aligned} \tag{7.108}$$

so that

$$C_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{7.109}$$

Let's consider their effects separately, beginning with the case $h_\times = 0$. Then we have

$$\frac{\partial^2}{\partial t^2} S^1 = \frac{1}{2} S^1 \frac{\partial^2}{\partial t^2} (h_+ e^{ik_\sigma x^\sigma}) \tag{7.110}$$

and

$$\frac{\partial^2}{\partial t^2} S^2 = -\frac{1}{2} S^2 \frac{\partial^2}{\partial t^2} (h_+ e^{ik_\sigma x^\sigma}). \tag{7.111}$$

These can be immediately solved to yield, to lowest order,

$$S^1 = \left(1 + \tfrac{1}{2} h_+ e^{ik_\sigma x^\sigma}\right) S^1(0) \tag{7.112}$$

and

$$S^2 = \left(1 - \tfrac{1}{2} h_+ e^{ik_\sigma x^\sigma}\right) S^2(0). \tag{7.113}$$

Thus, particles initially separated in the x^1 direction will oscillate in the x^1 direction, and likewise for those with an initial x^2 separation. That is, if we start with a ring of stationary particles in the x-y plane, as the wave passes they will bounce back and forth in the shape of a "+," as shown in Figure 7.4. On the other hand, the equivalent analysis for the case where $h_+ = 0$ but $h_\times \neq 0$ would yield the solution

$$S^1 = S^1(0) + \tfrac{1}{2} h_\times e^{ik_\sigma x^\sigma} S^2(0) \tag{7.114}$$

and

$$S^2 = S^2(0) + \tfrac{1}{2} h_\times e^{ik_\sigma x^\sigma} S^1(0). \tag{7.115}$$

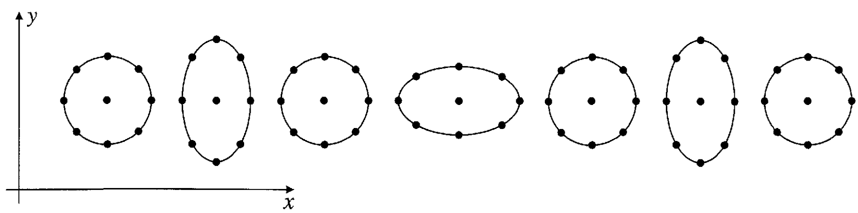

FIGURE 7.4 The effect of a gravitational wave with + polarization is to distort a circle of test particles into ellipses oscillating in a + pattern.

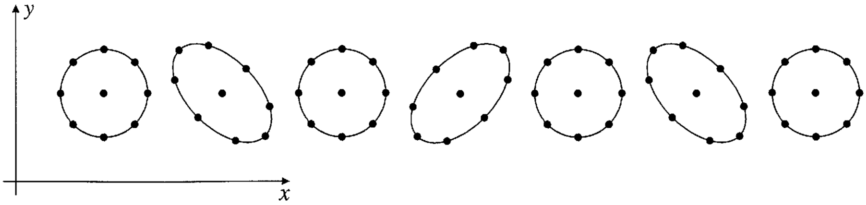

FIGURE 7.5 The effect of a gravitational wave with $\times$ polarization is to distort a circle of test particles into ellipses oscillating in a $\times$ pattern.

In this case the circle of particles would bounce back and forth in the shape of a "$\times$," as shown in Figure 7.5. The notation h_+ and $h_\times$ should therefore be clear. These two quantities measure the two independent modes of linear polarization of the gravitational wave, known as the "plus" and "cross" polarizations. If we liked we could consider right- and left-handed circularly polarized modes by defining

$$h_R = \frac{1}{\sqrt{2}}(h_+ + ih_\times),$$
$$h_L = \frac{1}{\sqrt{2}}(h_+ - ih_\times). \tag{7.116}$$

The effect of a pure h_R wave would be to rotate the particles in a right-handed sense, as shown in Figure 7.6, and similarly for the left-handed mode h_L. Note that the individual particles do not travel around the ring; they just move in little epicycles.

We can relate the polarization states of classical gravitational waves to the kinds of particles we would expect to find upon quantization. The spin of a quantized field is directly related to the transformation properties of that field under spatial rotations. The electromagnetic field has two independent polarization states described by vectors in the x-y plane; equivalently, a single polarization mode is invariant under a rotation by 360° in this plane. Upon quantization this theory yields the photon, a massless spin-1 particle. The neutrino, on the other hand, is also a massless particle, described by a field that picks up a minus sign under rotations by 360°; it is invariant under rotations of 720°, and we say it has

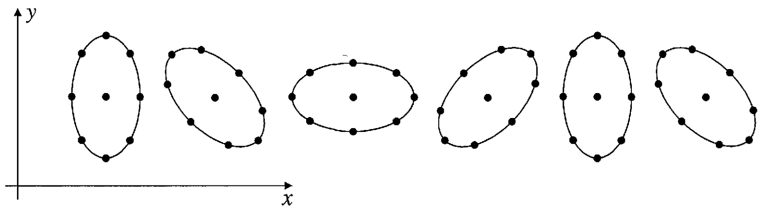

FIGURE 7.6 The effect of a gravitational wave with R polarization is to distort a circle of test particles into an ellipse that rotates in a right-handed sense.

spin-$\frac{1}{2}$. The general rule is that the spin S is related to the angle θ under which the polarization modes are invariant by $S = 360°/\theta$. The gravitational field, whose waves propagate at the speed of light, should lead to massless particles in the quantum theory. Noticing that the polarization modes we have described are invariant under rotations of 180° in the x-y plane, we expect the associated particles—gravitons—to be spin-2. We are a long way from detecting such particles (and it would not be a surprise if we never detected them directly), but any respectable quantum theory of gravity should predict their existence.

In fact, starting with a theory of spin-2 gravitons and requiring some simple properties provides a nice way to *derive* the full Einstein's equation of general relativity. Imagine starting with the Lagrangian (7.9) for the symmetric tensor $h_{\mu\nu}$, but now imagining that this "really is" a physical field propagating in Minkowski spacetime rather than a perturbation to a dynamical metric. (This Lagrangian doesn't include couplings to matter, but it is straightforward to do so.) Now make the additional demand that $h_{\mu\nu}$ couple to its own energy-momentum tensor (discussed below), as well as to the matter energy-momentum tensor. This induces higher-order nonlinear terms in the action, and consequently induces additional "energy-momentum" terms of even higher order. By repeating this process, an infinite series of terms is introduced, but the series can be summed to a simple expression, perhaps because you already know the answer—the Einstein–Hilbert action (possibly with some higher-order terms). In the process, we find that matter couples to the unique combination $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. In other words, by asking for a theory of a spin-2 field coupling to the energy-momentum tensor, we end up with the fully nonlinear glory of general relativity. The background metric $\eta_{\mu\nu}$ becomes completely unobservable. Of course, some of the global geometric aspects of GR are obscured by this procedure, which ultimately is just another way of justifying Einstein's equation.

While we are noting amusing things, let's point out that the behavior of gravitational waves yields a clue as to why string theory gives rise to a quantum theory of gravity. Consider the fundamental vibrational modes of a loop of string, as shown in Figure 7.7. There are three lowest-energy modes for a loop of string: an over-

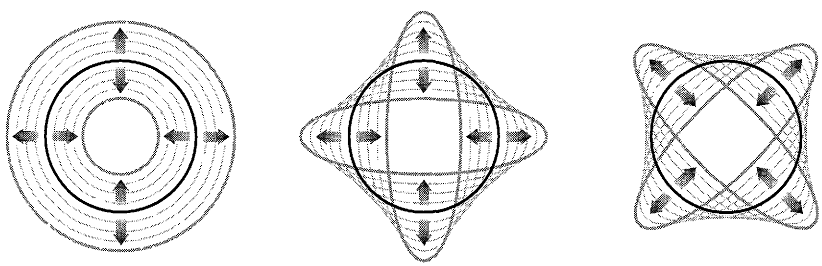

FIGURE 7.7 The three fundamental vibrational modes of a loop of string. The overall "breathing" mode (far left) is invariant under rotations, and gives rise to a spin-0 particle. The other two modes match the two polarizations of a gravitational wave, and represent the two states of a massless spin-2 particle.

all oscillation of its size, plus two independent ways it can oscillate into ellipses. These give rise to three massless degrees of freedom: a spin-0 particle (the dilaton) and a massless spin-2 particle (the graviton). Notice the obvious similarity between the string oscillations and the motion of test particles under the influence of a gravitational wave; this is no accident, and is the reason why quantized strings inevitably give rise to gravity. (String theory was originally investigated as a theory of the strong interactions, but different models would inevitably predict an unnecessary massless spin-2 particle; eventually it was realized that this flaw could be a virtue, if the theory came to be thought of as a quantum theory of gravity.) The extra unwanted spin-0 (scalar) mode reflects the fact that string theory actually predicts a scalar-tensor theory of gravity (as discussed in Section 4.8) rather than ordinary GR. Since a massless scalar of this sort is not observed in nature, some mechanism must work to give a mass to the scalar at low energies.

7.5 ■ PRODUCTION OF GRAVITATIONAL WAVES

With plane-wave solutions to the linearized vacuum equation in our possession, it remains to discuss the generation of gravitational radiation by sources. For this purpose it is necessary to consider Einstein's equation coupled to matter, $G_{\mu\nu} = 8\pi G T_{\mu\nu}$. Because $T_{\mu\nu}$ doesn't vanish, the metric perturbation will include nonzero scalar and vector components as well as the strain tensor representing gravitational waves; we cannot assume that our solution takes the transverse-traceless form (7.90). Instead, we will keep the entire perturbation $h_{\mu\nu}$ and solve for the produced gravitational wave far from the source, where we can then impose transverse-traceless gauge.

There are still some convenient simplifications we can introduce, even in the presence of sources. We first define the trace-reversed perturbation,

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \tfrac{1}{2} h \eta_{\mu\nu}. \tag{7.117}$$

The name of the trace-reversed perturbation makes sense, since

$$\bar{h} = \eta^{\mu\nu} \bar{h}_{\mu\nu} = -h. \tag{7.118}$$

Obviously we can reconstruct the original perturbation from the trace-reversed form, so no information has been lost. Note also that, if we are in vacuum far away from any sources and can go to transverse-traceless gauge, the trace-reversed perturbation will be equal to the original perturbation,

$$\bar{h}^{\mathrm{TT}}_{\mu\nu} = h^{\mathrm{TT}}_{\mu\nu}. \tag{7.119}$$

Meanwhile, we are still free to choose some sort of gauge. Under a gauge transformation (7.14), the trace-reversed perturbation transforms as

$$\bar{h}_{\mu\nu} \to \bar{h}_{\mu\nu} + 2\partial_{(\mu}\xi_{\nu)} - \partial_\lambda \xi^\lambda \eta_{\mu\nu}. \tag{7.120}$$

By choosing a gauge parameter ξ_μ satisfying

$$\Box\xi_\mu = -\partial_\lambda \bar{h}^\lambda{}_\mu, \tag{7.121}$$

we can therefore set

$$\partial_\mu \bar{h}^{\mu\nu} = 0. \tag{7.122}$$

This condition is known as the **Lorenz gauge**, analogous with the similar condition $\partial_\mu A^\mu = 0$ often used in electromagnetism.[3] Note that the original perturbation is not transverse in this gauge; rather, we have

$$\partial_\mu h^{\mu\nu} = \tfrac{1}{2}\partial^\nu h. \tag{7.123}$$

Plugging the definition of the trace-reversed perturbation into our expression for the Einstein tensor (7.8), and using the Lorenz gauge condition, yields the very concise expression

$$G_{\mu\nu} = -\tfrac{1}{2}\Box\bar{h}_{\mu\nu}. \tag{7.124}$$

The analogous expression in terms of the original perturbation $h_{\mu\nu}$ is slightly messier; this is the reason for introducing $\bar{h}_{\mu\nu}$. The linearized Einstein equation in this gauge is therefore simply a wave equation for each component,

$$\Box\bar{h}_{\mu\nu} = -16\pi G T_{\mu\nu}. \tag{7.125}$$

The solution to such an equation can be obtained using a Green function, in precisely the same way as the analogous problem in electromagnetism. Here we will review the outline of the method, following Wald (1984).

The Green function $G(x^\sigma - y^\sigma)$ for the d'Alembertian operator $\Box$ is the solution of the wave equation in the presence of a delta-function source:

$$\Box_x G(x^\sigma - y^\sigma) = \delta^{(4)}(x^\sigma - y^\sigma), \tag{7.126}$$

where $\Box_x$ denotes the d'Alembertian with respect to the coordinates x^σ. The usefulness of such a function resides in the fact that the general solution to an equation such as (7.125) can be written

$$\bar{h}_{\mu\nu}(x^\sigma) = -16\pi G \int G(x^\sigma - y^\sigma) T_{\mu\nu}(y^\sigma)\, d^4y, \tag{7.127}$$

as can be verified immediately. (Notice that no factors of $\sqrt{-g}$ are necessary, since our background is simply flat spacetime.) The solutions to (7.126) have of course been worked out long ago, and they can be thought of as either "retarded" or "advanced," depending on whether they represent waves traveling forward or

[3]Note the spelling. The "gauge" was originated by Ludwig Lorenz (1829–1891), while the more famous "transformation" was invented by Hendrick Antoon Lorentz (1853–1928). See J.D. Jackson and L.B. Okun, *Rev. Mod. Phys.* **73**, 663 (2001).

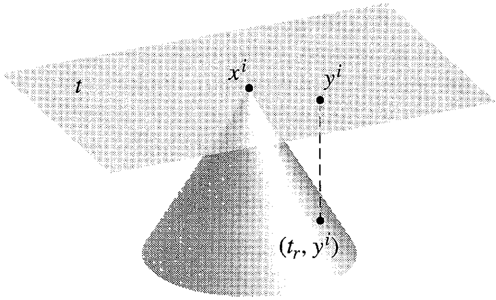

FIGURE 7.8 Disturbances in the gravitational field at (t, x^i) are calculated in terms of events on the past light cone.

backward in time. Our interest is in the retarded Green function, which represents the accumulated effect of signals to the past of the points under consideration. It is given by

$$G(x^\sigma - y^\sigma) = -\frac{1}{4\pi|\mathbf{x}-\mathbf{y}|}\delta[|\mathbf{x}-\mathbf{y}| - (x^0 - y^0)]\,\theta(x^0 - y^0). \qquad (7.128)$$

Here we have used boldface to denote the spatial vectors $\mathbf{x} = (x^1, x^2, x^3)$ and $\mathbf{y} = (y^1, y^2, y^3)$, with norm $|\mathbf{x}-\mathbf{y}| = [\delta_{ij}(x^i - y^i)(x^j - y^j)]^{1/2}$. The theta function $\theta(x^0 - y^0)$ equals 1 when $x^0 > y^0$, and zero otherwise. The derivation of (7.128) would take us too far afield, but it can be found in any standard text on electrodynamics or partial differential equations in physics.

Upon plugging (7.128) into (7.127), we can use the delta function to perform the integral over y^0, leaving us with

$$\bar{h}_{\mu\nu}(t, \mathbf{x}) = 4G\int \frac{1}{|\mathbf{x}-\mathbf{y}|}T_{\mu\nu}(t - |\mathbf{x}-\mathbf{y}|, \mathbf{y})\, d^3y, \qquad (7.129)$$

where $t = x^0$. The term "retarded time" is used to refer to the quantity

$$t_r = t - |\mathbf{x}-\mathbf{y}|. \qquad (7.130)$$

The interpretation of (7.129) should be clear: the disturbance in the gravitational field at $(t, \mathbf{x})$ is a sum of the influences from the energy and momentum sources at the point $(t_r, \mathbf{x}-\mathbf{y})$ on the past light cone, as depicted in Figure 7.8.

Let us take this general solution and consider the case where the gravitational radiation is emitted by an isolated source, fairly far away, comprised of nonrelativistic matter; these approximations will be made more precise as we go on. First we need to set up some conventions for Fourier transforms, which always make life easier when dealing with oscillatory phenomena. Given a function of spacetime $\phi(t, \mathbf{x})$, we are interested in its Fourier transform (and inverse) with respect to time alone,

$$\widetilde{\phi}(\omega, \mathbf{x}) = \frac{1}{\sqrt{2\pi}} \int dt\, e^{i\omega t} \phi(t, \mathbf{x}),$$
$$\phi(t, \mathbf{x}) = \frac{1}{\sqrt{2\pi}} \int d\omega\, e^{-i\omega t} \widetilde{\phi}(\omega, \mathbf{x}). \tag{7.131}$$

Taking the transform of the metric perturbation, we obtain

$$\begin{aligned}
\widetilde{\bar{h}}_{\mu\nu}(\omega, \mathbf{x}) &= \frac{1}{\sqrt{2\pi}} \int dt\, e^{i\omega t} \bar{h}_{\mu\nu}(t, \mathbf{x}) \\
&= \frac{4G}{\sqrt{2\pi}} \int dt\, d^3y\, e^{i\omega t} \frac{T_{\mu\nu}(t - |\mathbf{x}-\mathbf{y}|, \mathbf{y})}{|\mathbf{x}-\mathbf{y}|} \\
&= \frac{4G}{\sqrt{2\pi}} \int dt_r\, d^3y\, e^{i\omega t_r} e^{i\omega|\mathbf{x}-\mathbf{y}|} \frac{T_{\mu\nu}(t_r, \mathbf{y})}{|\mathbf{x}-\mathbf{y}|} \\
&= 4G \int d^3y\, e^{i\omega|\mathbf{x}-\mathbf{y}|} \frac{\widetilde{T}_{\mu\nu}(\omega, \mathbf{y})}{|\mathbf{x}-\mathbf{y}|}.
\end{aligned} \tag{7.132}$$

In this sequence, the first equation is simply the definition of the Fourier transform, the second line comes from the solution (7.129), the third line is a change of variables from t to t_r, and the fourth line is once again the definition of the Fourier transform.

We now make the approximations that our source is isolated, far away, and slowly moving. This means that we can consider the source to be centered at a (spatial) distance r, with the different parts of the source at distances $r + \delta r$ such that $\delta r \ll r$, as shown in Figure 7.9. Since it is slowly moving, most of the radiation emitted will be at frequencies ω sufficiently low that $\delta r \ll \omega^{-1}$. (Essentially, light traverses the source much faster than the components of the source itself do.) Under these approximations, the term $e^{i\omega|\mathbf{x}-\mathbf{y}|}/|\mathbf{x}-\mathbf{y}|$ can be replaced by $e^{i\omega r}/r$ and brought outside the integral. This leaves us with

$$\widetilde{\bar{h}}_{\mu\nu}(\omega, \mathbf{x}) = 4G \frac{e^{i\omega r}}{r} \int d^3y\, \widetilde{T}_{\mu\nu}(\omega, \mathbf{y}). \tag{7.133}$$

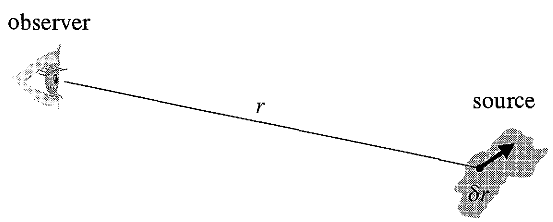

FIGURE 7.9 A source of size δr, at a distance r from the observer.

In fact there is no need to compute all of the components of $\widetilde{\bar{h}}_{\mu\nu}(\omega, \mathbf{x})$, since the Lorenz gauge condition $\partial_\mu \bar{h}^{\mu\nu}(t, \mathbf{x}) = 0$ in Fourier space implies

$$\widetilde{\bar{h}}^{0\nu} = -\frac{i}{\omega}\partial_i \widetilde{\bar{h}}^{i\nu}. \tag{7.134}$$

We therefore only need to concern ourselves with the spacelike components of $\widetilde{\bar{h}}_{\mu\nu}(\omega, \mathbf{x})$, and recover $\widetilde{\bar{h}}^{0\nu}$ from (7.134). The first thing to do is to set $\nu = j$ to find $\widetilde{\bar{h}}^{0j}$ from $\widetilde{\bar{h}}^{ij}$, which we would then use to find $\widetilde{\bar{h}}^{00}$ from $\widetilde{\bar{h}}^{i0}$. From (7.133) we therefore want to take the integral of the spacelike components of $\widetilde{T}_{\mu\nu}(\omega, \mathbf{y})$. We begin by integrating by parts in reverse:

$$\int d^3y\, \widetilde{T}^{ij}(\omega, \mathbf{y}) = \int \partial_k(y^i \widetilde{T}^{kj})\, d^3y - \int y^i(\partial_k \widetilde{T}^{kj})\, d^3y. \tag{7.135}$$

The first term is a surface integral which will vanish since the source is isolated, while the second can be related to $\widetilde{T}^{0j}$ by the Fourier-space version of $\partial_\mu T^{\mu\nu} = 0$:

$$-\partial_k \widetilde{T}^{k\mu} = i\omega \widetilde{T}^{0\mu}. \tag{7.136}$$

Thus,

$$\begin{aligned}
\int d^3y\, \widetilde{T}^{ij}(\omega, \mathbf{y}) &= i\omega \int y^i \widetilde{T}^{0j}\, d^3y \\
&= \frac{i\omega}{2}\int (y^i \widetilde{T}^{0j} + y^j \widetilde{T}^{0i})\, d^3y \\
&= \frac{i\omega}{2}\int \left[\partial_l(y^i y^j \widetilde{T}^{0l}) - y^i y^j(\partial_l \widetilde{T}^{0l})\right] d^3y \\
&= -\frac{\omega^2}{2}\int y^i y^j \widetilde{T}^{00}\, d^3y.
\end{aligned} \tag{7.137}$$

The second line is justified since we know that the left-hand side is symmetric in i and j, while the third and fourth lines are simply repetitions of reverse integration by parts and conservation of $T^{\mu\nu}$. It is conventional to define the **quadrupole moment tensor** of the energy density of the source,

$$I_{ij}(t) = \int y^i y^j T^{00}(t, \mathbf{y})\, d^3y, \tag{7.138}$$

a constant tensor on each surface of constant time. The overall normalization of the quadrupole tensor is a matter of convention, and by no means universal, so be careful in comparing different references. In terms of the Fourier transform of the quadrupole moment, our solution takes on the compact form

$$\widetilde{\bar{h}}_{ij}(\omega, \mathbf{x}) = -2G\omega^2 \frac{e^{i\omega r}}{r}\tilde{I}_{ij}(\omega). \tag{7.139}$$

We can transform this back to t to obtain the **quadrupole formula**,

$$\boxed{\bar{h}_{ij}(t, \mathbf{x}) = \frac{2G}{r} \frac{d^2 I_{ij}}{dt^2}(t_r),} \tag{7.140}$$

where as before $t_r = t - r$.

The gravitational wave produced by an isolated nonrelativistic object is therefore proportional to the second derivative of the quadrupole moment of the energy density at the point where the past light cone of the observer intersects the source. In contrast, the leading contribution to electromagnetic radiation comes from the changing *dipole* moment of the charge density. The difference can be traced back to the universal nature of gravitation. A changing dipole moment corresponds to motion of the center of density—charge density in the case of electromagnetism, energy density in the case of gravitation. While there is nothing to stop the center of charge of an object from oscillating, oscillation of the center of mass of an isolated system violates conservation of momentum. (You can shake a body up and down, but you and the earth shake ever so slightly in the opposite direction to compensate.) The quadrupole moment, which measures the shape of the system, is generally smaller than the dipole moment, and for this reason, as well as the weak coupling of matter to gravity, gravitational radiation is typically much weaker than electromagnetic radiation.

One case of special interest is the gravitational radiation emitted by a binary star (two stars in orbit around each other). For simplicity let us consider two stars of mass M in a circular orbit in the x^1-x^2 plane, at distance R from their common center of mass, as shown in Figure 7.10. We will treat the motion of the stars in the Newtonian approximation, where we can discuss their orbit just as Kepler would have. Circular orbits are most easily characterized by equating the force due to gravity to the outward "centrifugal" force:

$$\frac{GM^2}{(2R)^2} = \frac{Mv^2}{R}, \tag{7.141}$$

which gives us

$$v = \left(\frac{GM}{4R}\right)^{1/2}. \tag{7.142}$$

The time it takes to complete a single orbit is simply

$$T = \frac{2\pi R}{v}, \tag{7.143}$$

but more useful to us is the angular frequency of the orbit,

$$\Omega = \frac{2\pi}{T} = \left(\frac{GM}{4R^3}\right)^{1/2}. \tag{7.144}$$

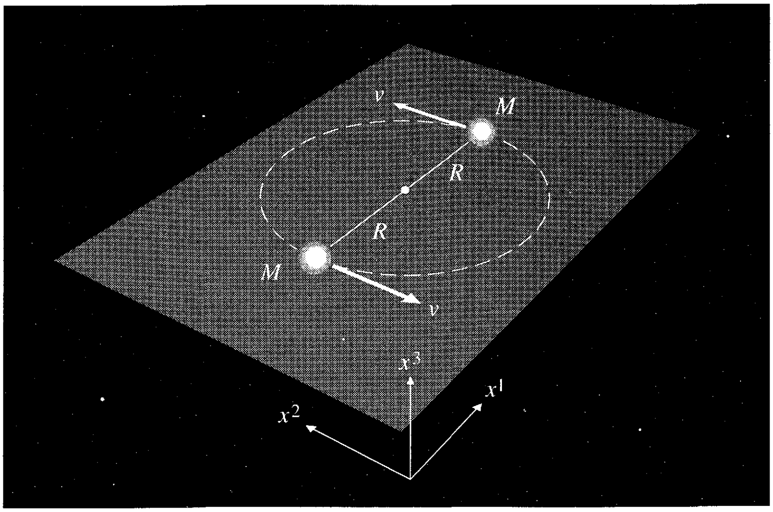

FIGURE 7.10 A binary star system. Two stars of mass M orbit in the x^1-x^2 plane with an orbital radius R.

In terms of Ω we can write down the explicit path of star a,

$$x_a^1 = R\cos\Omega t, \qquad x_a^2 = R\sin\Omega t, \tag{7.145}$$

and star b,

$$x_b^1 = -R\cos\Omega t, \qquad x_b^2 = -R\sin\Omega t. \tag{7.146}$$

The corresponding energy density is

$$\begin{aligned} T^{00}(t,\mathbf{x}) = M\delta(x^3)\big[&\delta(x^1 - R\cos\Omega t)\delta(x^2 - R\sin\Omega t) \\ &+ \delta(x^1 + R\cos\Omega t)\delta(x^2 + R\sin\Omega t)\big]. \end{aligned} \tag{7.147}$$

The profusion of delta functions allows us to integrate this straightforwardly to obtain the quadrupole moment from (7.138):

$$\begin{aligned} I_{11} &= 2MR^2\cos^2\Omega t = MR^2(1+\cos 2\Omega t) \\ I_{22} &= 2MR^2\sin^2\Omega t = MR^2(1-\cos 2\Omega t) \\ I_{12} = I_{21} &= 2MR^2(\cos\Omega t)(\sin\Omega t) = MR^2\sin 2\Omega t \\ I_{i3} &= 0. \end{aligned} \tag{7.148}$$

From this in turn it is easy to get the components of the metric perturbation from (7.140):

$$\bar{h}_{ij}(t, \mathbf{x}) = \frac{8GM}{r}\Omega^2 R^2 \begin{pmatrix} -\cos 2\Omega t_r & -\sin 2\Omega t_r & 0 \\ -\sin 2\Omega t_r & \cos 2\Omega t_r & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{7.149}$$

The remaining components of $\bar{h}_{\mu\nu}$ could be derived from demanding that the Lorenz gauge condition be satisfied.

7.6 ■ ENERGY LOSS DUE TO GRAVITATIONAL RADIATION

It is natural at this point to talk about the energy emitted via gravitational radiation. Such a discussion, however, is immediately beset by problems, both technical and philosophical. As we have mentioned before, there is no true local measure of the energy in the gravitational field. Of course, in the weak field limit, where we think of gravitation as being described by a symmetric tensor propagating on a fixed background metric, we might hope to derive an energy-momentum tensor for the fluctuations $h_{\mu\nu}$, just as we would for electromagnetism or any other field theory. To some extent this is possible, but still difficult. As a result of these difficulties there are a number of different proposals in the literature for what we should use as the energy-momentum tensor for gravitation in the weak field limit; all of them are different, but for the most part they give the same answers for physically well-posed questions such as the rate of energy emitted by a binary system.

At a technical level, the difficulties begin to arise when we consider what form the energy-momentum tensor should take. We have previously mentioned the energy-momentum tensors for electromagnetism and scalar field theory, both of which share an important feature—they are quadratic in the relevant fields. By hypothesis, our approach to the weak field limit has been to keep only terms that are linear in the metric perturbation. Hence, in order to keep track of the energy carried by the gravitational waves, we will have to extend our calculations to at least second order in $h_{\mu\nu}$. In fact we have been cheating slightly all along. In discussing the effects of gravitational waves on test particles, and the generation of waves by a binary system, we have been using the fact that test particles move along geodesics. But as we know, this is derived from the covariant conservation of energy-momentum, $\nabla_\mu T^{\mu\nu} = 0$. In the order to which we have been working, however, we actually have $\partial_\mu T^{\mu\nu} = 0$, which would imply that test particles move on straight lines in the flat background metric. This is a symptom of the inability of the weak field limit to describe self-gravitating systems. In practice, the best that can be done is to solve the weak field equation to some appropriate order, and then justify after the fact the validity of the solution. We will follow the procedure outlined in Chapters 35 and 36 of Misner, Thorne, and Wheeler (1973), where additional discussion of subtleties may be found. See also Wald (1984) and Schutz (1985).

Let us now examine Einstein's vacuum equation $R_{\mu\nu} = 0$ to second order, and see how the result can be interpreted in terms of an energy-momentum tensor for the gravitational field. We expand both the metric and the Ricci tensor,

$$g_{\mu\nu} = \eta_{\mu\nu} + h^{(1)}_{\mu\nu} + h^{(2)}_{\mu\nu}$$
$$R_{\mu\nu} = R^{(0)}_{\mu\nu} + R^{(1)}_{\mu\nu} + R^{(2)}_{\mu\nu}, \tag{7.150}$$

where $R^{(1)}_{\mu\nu}$ is taken to be of the same order as $h^{(1)}_{\mu\nu}$, while $R^{(2)}_{\mu\nu}$ and $h^{(2)}_{\mu\nu}$ are of order $(h^{(1)}_{\mu\nu})^2$. Because we work in a flat background, the zeroth-order equation $R^{(0)}_{\mu\nu} = 0$ is automatically solved. The first-order vacuum equation is simply

$$R^{(1)}_{\mu\nu}[h^{(1)}] = 0, \tag{7.151}$$

which determines the first-order perturbation $h^{(1)}_{\mu\nu}$ (up to gauge transformations). The second-order perturbation $h^{(2)}_{\mu\nu}$ will be determined by the second-order equation

$$R^{(1)}_{\mu\nu}[h^{(2)}] + R^{(2)}_{\mu\nu}[h^{(1)}] = 0. \tag{7.152}$$

The notation $R^{(1)}_{\mu\nu}[h^{(2)}]$ indicates the parts of the expanded Ricci tensor that are linear in the metric perturbation, as given by (7.6), applied to the second-order perturbation $h^{(2)}_{\mu\nu}$; meanwhile, $R^{(2)}_{\mu\nu}[h^{(1)}]$ stands for the quadratic part of the expanded Ricci tensor,

$$\begin{aligned} R^{(2)}_{\mu\nu} = {} & \tfrac{1}{2}h^{\rho\sigma}\partial_\mu\partial_\nu h_{\rho\sigma} + \tfrac{1}{4}(\partial_\mu h_{\rho\sigma})\partial_\nu h^{\rho\sigma} + (\partial^\sigma h^\rho{}_\nu)\partial_{[\sigma}h_{\rho]\mu} - h^{\rho\sigma}\partial_\rho\partial_{(\mu}h_{\nu)\sigma} \\ & + \tfrac{1}{2}\partial_\sigma(h^{\rho\sigma}\partial_\rho h_{\mu\nu}) - \tfrac{1}{4}(\partial_\rho h_{\mu\nu})\partial^\rho h - (\partial_\sigma h^{\rho\sigma} - \tfrac{1}{2}\partial^\rho h)\partial_{(\mu}h_{\nu)\rho}, \end{aligned} \tag{7.153}$$

applied to the first-order perturbation $h^{(1)}_{\mu\nu}$. There are no cross terms, as they would necessarily be higher order.

Now let's write the vacuum equation as $G_{\mu\nu} = 0$; this is of course equivalent to $R_{\mu\nu} = 0$, but will enable us to express the result in a suggestive form. At second order we have

$$R^{(1)}_{\mu\nu}[h^{(2)}] - \tfrac{1}{2}\eta^{\rho\sigma}R^{(1)}_{\rho\sigma}[h^{(2)}]\eta_{\mu\nu} = 8\pi G t_{\mu\nu}, \tag{7.154}$$

where we have defined

$$t_{\mu\nu} \equiv -\frac{1}{8\pi G}\left\{R^{(2)}_{\mu\nu}[h^{(1)}] - \frac{1}{2}\eta^{\rho\sigma}R^{(2)}_{\rho\sigma}[h^{(1)}]\eta_{\mu\nu}\right\}. \tag{7.155}$$

Notice a couple of things about this expression. First, we have not included terms of the form $h^{(1)\rho\sigma}R^{(1)}_{\rho\sigma}[h^{(1)}]$, since $R^{(1)}_{\mu\nu}[h^{(1)}] = 0$. Second, the left-hand side of (7.154) is not the full second-order Einstein tensor, as we have moved terms involving $R^{(2)}_{\mu\nu}[h^{(1)}]$ to the right-hand side and provocatively relabeled them as an energy-momentum tensor for the first-order perturbations, $t_{\mu\nu}$. Such an identification seems eminently reasonable; $t_{\mu\nu}$ is a symmetric tensor, quadratic in $h_{\mu\nu}$, which represents how the perturbations affect the spacetime metric in just the way that a matter energy-momentum tensor would. (Linear terms in $h_{\mu\nu}$ have no effect, since $G^{(1)}_{\mu\nu}[h^{(1)}]$ is simply set to zero by the first-order equation.) Notice that $t_{\mu\nu}$ is also conserved, in the background flat-space sense,

$$\partial_\mu t^{\mu\nu} = 0, \tag{7.156}$$

which we know from the Bianchi identity $\partial_\mu G^{\mu\nu} = 0$.

Unfortunately there are some limitations on our interpretation of $t_{\mu\nu}$ as an energy-momentum tensor. Of course it is not a tensor at all in the full theory, but we are leaving that aside by hypothesis. More importantly, it is not invariant under gauge transformations (infinitesimal diffeomorphisms), as you could check by direct calculation. One way of circumventing this difficulty is to average the energy-momentum tensor over several wavelengths, an operation we denote by angle brackets $\langle\cdots\rangle$. This procedure has both philosophical and practical advantages. From a philosophical viewpoint, we know that our ability to choose Riemann normal coordinates at any one point makes it impossible to define a reliable measure of the gravitational energy-momentum that is purely local (defined at each point in terms of the metric and its first derivatives at precisely that point). If we average over several wavelengths, however, we may hope to capture enough of the physical curvature in a small region to describe a gauge-invariant measure. From a practical standpoint, any terms that are derivatives (as opposed to products of derivatives) will average to zero,

$$\langle \partial_\mu (X)\rangle = 0. \tag{7.157}$$

We are therefore empowered to integrate by parts under the averaging brackets,

$$\langle A(\partial_\mu B)\rangle = -\langle(\partial_\mu A)B\rangle, \tag{7.158}$$

which will greatly simplify our expressions.

With this in mind, let us calculate $t_{\mu\nu}$ as defined in (7.155), using the expression (7.153) for the second-order Ricci tensor. (Henceforth we will no longer use superscripts on the metric perturbation, as we will only be interested in the first-order perturbation.) Although part of the motivation for averaging is to obtain a gauge-invariant answer, the actual calculation is a mess, so for illustrative purposes we will carry it out in transverse-traceless gauge,

$$\partial^\mu h^{\mathrm{TT}}_{\mu\nu} = 0, \quad h^{\mathrm{TT}} = 0. \tag{7.159}$$

Don't forget that we are only allowed to choose this gauge in vacuum. The non-vanishing parts of $R^{(2)\mathrm{TT}}_{\mu\nu}$ in this gauge can be written as

$$\begin{aligned} R^{(2)\mathrm{TT}}_{\mu\nu} = {} & \tfrac{1}{2} h^{\rho\sigma}_{\mathrm{TT}} \partial_\mu \partial_\nu h^{\mathrm{TT}}_{\rho\sigma} + \tfrac{1}{4}(\partial_\mu h^{\mathrm{TT}}_{\rho\sigma})\partial_\nu h^{\rho\sigma}_{\mathrm{TT}} + \tfrac{1}{2}\eta^{\rho\lambda}(\partial^\sigma h^{\mathrm{TT}}_{\rho\nu})\partial_\sigma h^{\mathrm{TT}}_{\lambda\mu} \\ & - \tfrac{1}{2}(\partial^\sigma h^{\mathrm{TT}}_{\rho\nu})\partial^\rho h^{\mathrm{TT}}_{\sigma\mu} - h^{\rho\sigma}_{\mathrm{TT}} \partial_\rho \partial_{(\mu} h^{\mathrm{TT}}_{\nu)\sigma} + \tfrac{1}{2} h^{\rho\sigma}_{\mathrm{TT}} \partial_\sigma \partial_\rho h^{\mathrm{TT}}_{\mu\nu}. \end{aligned} \tag{7.160}$$

Now let's apply the averaging brackets, and integrate by parts where convenient. The last three terms in (7.160) all go away, as integration by parts leads to divergences that vanish. We are left with

$$\left\langle R^{(2)\mathrm{TT}}_{\mu\nu}\right\rangle = -\tfrac{1}{4}\left\langle(\partial_\mu h^{\mathrm{TT}}_{\rho\sigma})(\partial_\nu h^{\rho\sigma}_{\mathrm{TT}}) + 2\eta^{\rho\lambda}(\Box h^{\mathrm{TT}}_{\rho\nu})h^{\mathrm{TT}}_{\lambda\mu}\right\rangle. \tag{7.161}$$

But the perturbation obeys the first-order equation of motion, which sets $\Box h^{\mathrm{TT}}_{\mu\nu} = 0$. So we are finally left with

$$\left\langle R^{(2)\mathrm{TT}}_{\mu\nu}\right\rangle = -\tfrac{1}{4}\left\langle(\partial_\mu h^{\mathrm{TT}}_{\rho\sigma})(\partial_\nu h^{\rho\sigma}_{\mathrm{TT}})\right\rangle. \tag{7.162}$$

We can take the trace to get the curvature scalar; after integration by parts we again find a $\Box h^{\mathrm{TT}}_{\mu\nu}$ term which we set to zero, so

$$\left\langle \eta^{\mu\nu} R^{(2)\mathrm{TT}}_{\mu\nu}\right\rangle = 0. \tag{7.163}$$

These expressions can be inserted into (7.155) to obtain a simple expression for the gravitational-wave energy-momentum tensor in transverse-traceless gauge:

$$\boxed{t_{\mu\nu} = \frac{1}{32\pi G}\left\langle(\partial_\mu h^{\mathrm{TT}}_{\rho\sigma})(\partial_\nu h^{\rho\sigma}_{\mathrm{TT}})\right\rangle.} \tag{7.164}$$

Remember that, in this gauge, nonspatial components vanish, $h^{\mathrm{TT}}_{0\nu} = 0$. You will therefore sometimes see the above expression written with spatial indices ij instead of spacetime indices $\rho\sigma$; the two versions are clearly equivalent. If we had been strong enough to do the corresponding calculation without first choosing a gauge, we would have found

$$\begin{aligned} t_{\mu\nu} = \frac{1}{32\pi G}\Big\langle &(\partial_\mu h_{\rho\sigma})(\partial_\nu h^{\rho\sigma}) - \frac{1}{2}(\partial_\mu h)(\partial_\nu h) \\ &- (\partial_\rho h^{\rho\sigma})(\partial_\mu h_{\nu\sigma}) - (\partial_\rho h^{\rho\sigma})(\partial_\nu h_{\mu\sigma})\Big\rangle. \end{aligned} \tag{7.165}$$

A bit of straightforward manipulation suffices to check that this expression is actually gauge invariant, as you are asked to show in the exercises.

Let's calculate the transverse-traceless expression (7.164) for a single plane wave,

$$h^{\mathrm{TT}}_{\mu\nu} = C_{\mu\nu}\sin(k_\lambda x^\lambda). \tag{7.166}$$

We have taken the real part and set the phase arbitrarily so that the wave is a sine rather than cosine. The energy-momentum tensor is then

$$t_{\mu\nu} = \frac{1}{32\pi G} k_\mu k_\nu C_{\rho\sigma} C^{\rho\sigma}\left\langle\cos^2(k_\lambda x^\lambda)\right\rangle. \tag{7.167}$$

Averaging the $\cos^2$ term over several wavelengths yields

$$\left\langle \cos^2(k_\lambda x^\lambda) \right\rangle = \tfrac{1}{2}. \tag{7.168}$$

For simplicity we can take the wave to be moving along the z-axis, so that

$$k_\lambda = (-\omega, 0, 0, \omega) \tag{7.169}$$

the minus sign coming from lowering an index on k^λ, and from (7.109),

$$C_{\rho\sigma}C^{\rho\sigma} = 2(h_+^2 + h_\times^2). \tag{7.170}$$

It is more common in the gravitational-wave literature to express observables in terms of the ordinary frequency $f = \omega/2\pi$, rather than the angular frequency ω. Putting it all together reveals

$$t_{\mu\nu} = \frac{\pi}{8G} f^2(h_+^2 + h_\times^2) \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}. \tag{7.171}$$

As we will discuss in the next section, typical gravitational-wave sources we might expect to observe at Earth will have frequencies between 10^{-4} and 10^4 Hz, and amplitudes $h \sim 10^{-22}$. It is therefore useful to express the energy flux in the z direction, $-T_{0z}$, at an order-of-magnitude level as

$$-T_{0z} \sim 10^{-4} \left(\frac{f}{\mathrm{Hz}}\right)^2 \frac{(h_+^2 + h_\times^2)}{(10^{-21})^2} \frac{\mathrm{erg}}{\mathrm{cm}^2 \cdot \mathrm{s}}. \tag{7.172}$$

This is the amount of energy that could in principle be deposited in each square centimeter of a detector every second. As pointed out by Thorne,[4] this is actually a substantial energy flux, especially at the upper end of the frequency range. For comparison purposes, a supernova at cosmological distances is characterized by a peak electromagnetic flux of approximately 10^{-9} erg/cm^2/s; the gravitational-wave signal, however, only lasts for milliseconds, while the visible electromagnetic signal extends for months.

Now let's use our formula for the gravitational-wave energy-momentum tensor to calculate the rate of energy loss from a system emitting gravitational radiation according to the quadrupole formula (7.140). The total energy contained in gravitational radiation on a surface Σ of constant time is defined as

$$E = \int_\Sigma t_{00}\, d^3x, \tag{7.173}$$

while the total energy radiated through to infinity may be expressed as

[4]K.S. Thorne, in *Three Hundred Years of Gravitation*, Cambridge: Cambridge University Press, 1987.

$$\Delta E = \int P\, dt, \tag{7.174}$$

where the power P is

$$P = \int_{S^2_\infty} t_{0\mu} n^\mu r^2 d\Omega. \tag{7.175}$$

Here, the integral is taken over a two-sphere at spatial infinity S^2_∞, and n^μ is a unit spacelike vector normal to S^2_∞. In polar coordinates $\{t, r, \theta, \phi\}$, the components of the normal vector are

$$n^\mu = (0, 1, 0, 0). \tag{7.176}$$

We would like to calculate the power P using our expression for $t_{\mu\nu}$, (7.164). The first issue we face is that this expression is written in terms of the transverse-traceless perturbation, while the quadrupole formula (7.140) is written in terms of the spatial components $\bar{h}_{ij}$ of the Lorenz-gauge trace-reversed perturbation. The simplest procedure (although it's not that simple) is to first convert $\bar{h}_{ij}$ into transverse-traceless gauge, which is permissible because we are interested in the behavior of the waves in vacuum, far from the source from which they are emitted, plug into the formula for $t_{\mu\nu}$, then convert back into nontransverse-traceless form. Let's see how this works.

We begin by introducing the (spatial) projection tensor

$$P_{ij} = \delta_{ij} - n_i n_j, \tag{7.177}$$

which projects tensor components into a surface orthogonal to the unit vector n^i. (See Appendix D for more discussion.) In our case, we choose n^i to point along the direction of propagation of the wave, so that P_{ij} will project onto the two-sphere at spatial infinity. We can use the projection tensor to construct the transverse-traceless version of a symmetric spatial tensor X_{ij} via

$$X^{\mathrm{TT}}_{ij} = \left(P_i{}^k P_j{}^l - \tfrac{1}{2} P_{ij} P^{kl}\right) X_{kl}. \tag{7.178}$$

You can check for yourself that X^{TT}_{ij} is indeed transverse and traceless. Because it is traceless, $\bar{h}^{\mathrm{TT}}_{ij}$ is equal to the original perturbation h^{TT}_{ij}; plugging into the quadrupole formula (7.140), we get

$$h^{\mathrm{TT}}_{ij} = \bar{h}^{\mathrm{TT}}_{ij} = \frac{2G}{r} \frac{d^2 I^{\mathrm{TT}}_{ij}}{dt^2}(t - r), \tag{7.179}$$

where the transverse-traceless part of the quadrupole moment is also constructed via (7.178). In fact the quadrupole moment defined by (7.138) is not the most convenient quantity to use in expressing the generated wave, as it involves an integral over the energy density that might be difficult to determine. Instead we can use the **reduced quadrupole moment**,

$$J_{ij} = I_{ij} - \tfrac{1}{3}\delta_{ij}\delta^{kl} I_{kl}, \tag{7.180}$$

which is just the traceless part of I_{ij}. The reduced quadrupole moment has the nice property of being the coefficient of the r^{-3} term in the multipole expansion of the Newtonian potential,

$$\Phi = -\frac{GM}{r} - \frac{G}{r^3} D_i x^i - \frac{3G}{2r^5} J_{ij} x^i x^j + \cdots, \tag{7.181}$$

and is therefore more readily approximated for realistic sources. (Here D_i is the dipole moment, $D_i = \int T^{00} x^i \, d^3x$.) Of course, the transverse-traceless part of the quadrupole moment is the same as the transverse-traceless part of the reduced (that is, traceless) quadrupole moment, so (7.179) becomes

$$h_{ij}^{\mathrm{TT}} = \frac{2G}{r} \frac{d^2 J_{ij}^{\mathrm{TT}}}{dt^2}(t - r). \tag{7.182}$$

To calculate the power, we are interested in $t_{0\mu} n^\mu = t_{0r}$. Because the quadrupole moment depends only on the retarded time $t_r = t - r$, we have

$$\begin{aligned} \partial_0 h_{ij}^{\mathrm{TT}} &= \frac{2G}{r} \frac{d^3 J_{ij}^{\mathrm{TT}}}{dt^3}, \\ \partial_r h_{ij}^{\mathrm{TT}} &= -\frac{2G}{r} \frac{d^3 J_{ij}^{\mathrm{TT}}}{dt^3} - \frac{2G}{r^2} \frac{d^2 J_{ij}^{\mathrm{TT}}}{dt^2} \\ &\approx -\frac{2G}{r} \frac{d^3 J_{ij}^{\mathrm{TT}}}{dt^3}, \end{aligned} \tag{7.183}$$

where we have dropped the r^{-2} term because we are interested in the $r \to \infty$ limit. The important component of the energy-momentum tensor is therefore

$$t_{0r} = -\frac{G}{8\pi r^2} \left\langle \left(\frac{d^3 J_{ij}^{\mathrm{TT}}}{dt^3} \right) \left(\frac{d^3 J_{\mathrm{TT}}^{ij}}{dt^3} \right) \right\rangle. \tag{7.184}$$

The next step is to convert back to J_{ij} from the transverse-traceless part. Applying (7.178) and some messy algebra, it is straightforward to show that

$$X_{ij}^{\mathrm{TT}} X_{\mathrm{TT}}^{ij} = X_{ij} X^{ij} - 2X_i{}^j X^{ik} n_j n_k + \tfrac{1}{2} X^{ij} X^{kl} n_i n_j n_k n_l - \tfrac{1}{2} X^2 + X X^{ij} n_i n_j, \tag{7.185}$$

where $X = \delta^{ij} X_{ij}$. Because J_{ij} is traceless, we have

$$J_{ij}^{\mathrm{TT}} J_{\mathrm{TT}}^{ij} = J_{ij} J^{ij} - 2J_i{}^j J^{ik} n_j n_k + \tfrac{1}{2} J^{ij} J^{kl} n_i n_j n_k n_l, \tag{7.186}$$

and the power is

$$P = -\frac{G}{8\pi}\int_{S^2_\infty}\left\langle \frac{d^3J_{ij}}{dt^3}\frac{d^3J^{ij}}{dt^3} - 2\frac{d^3J_i{}^j}{dt^3}\frac{d^3J^{ik}}{dt^3}n_jn_k + \frac{1}{2}\frac{d^3J^{ij}}{dt^3}\frac{d^3J^{kl}}{dt^3}n_in_jn_kn_l\right\rangle d\Omega. \tag{7.187}$$

To evaluate this expression, it is best to switch back to Cartesian coordinates in space, where $n^i = x^i/r$. The quadrupole tensors are independent of the angular coordinates, since they are defined by integrals over all of space. We may therefore pull them outside the integral, and use the identities

$$\begin{aligned}\int d\Omega &= 4\pi\\ \int n_in_j\,d\Omega &= \frac{4\pi}{3}\delta_{ij}\\ \int n_in_jn_kn_l\,d\Omega &= \frac{4\pi}{15}(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}).\end{aligned} \tag{7.188}$$

When all is said and done, the expression for the power collapses to

$$P = -\frac{G}{5}\left\langle \frac{d^3J_{ij}}{dt^3}\frac{d^3J^{ij}}{dt^3}\right\rangle, \tag{7.189}$$

where we should remember that the quadrupole moment is evaluated at the retarded time $t_r = t - r$. Our formula has a minus sign because it represents the rate at which the energy is changing, and radiating sources will be losing energy.

For the binary system represented by (7.148), the reduced quadrupole moment is

$$J_{ij} = \frac{MR^2}{3}\begin{pmatrix}(1+3\cos 2\Omega t) & 3\sin 2\Omega t & 0\\ 3\sin 2\Omega t & (1-3\cos 2\Omega t) & 0\\ 0 & 0 & -2\end{pmatrix}, \tag{7.190}$$

and its third time derivative is therefore

$$\frac{d^3J_{ij}}{dt^3} = 8MR^2\Omega^3\begin{pmatrix}\sin 2\Omega t & -\cos 2\Omega t & 0\\ -\cos 2\Omega t & -\sin 2\Omega t & 0\\ 0 & 0 & 0\end{pmatrix}. \tag{7.191}$$

The power radiated by the binary is thus

$$P = -\frac{128}{5}GM^2R^4\Omega^6, \tag{7.192}$$

or, using expression (7.144) for the frequency,

$$P = -\frac{2}{5}\frac{G^4M^5}{R^5}. \tag{7.193}$$

Of course, energy loss through the emission of gravitational radiation has been observed. In 1974 Hulse and Taylor discovered a binary system, PSR1913+16, in which both stars are very small, so classical effects are negligible, or at least under control, and one is a pulsar. The period of the orbit is eight hours, extremely small by astrophysical standards. The fact that one of the stars is a pulsar provides a very accurate clock, with respect to which the change in the period as the system loses energy can be measured. The result is consistent with the prediction of general relativity for energy loss through gravitational radiation.

7.7 ■ DETECTION OF GRAVITATIONAL WAVES

One of the highest-priority goals of contemporary gravitational physics and astrophysics is to detect gravitational radiation directly. (By direct we mean "by observing the influence of the gravitational wave on test bodies," in contrast to observing the indirect effect of energy loss, as in the binary pulsar.) There is every reason to believe that such a detection will happen soon, either in already-existing gravitational-wave observatories or those being planned for the near future. Once we detect gravitational radiation, of course, the goal will immediately become to extract useful astrophysical information from the observations. Our current understanding of the universe outside the Solar System comes almost exclusively from observations of electromagnetic radiation, with some additional input from neutrinos and cosmic rays; the advent of gravitational-wave astrophysics will open an entirely new window onto energetic phenomena in the distant universe.[5]

Before discussing how we might go about detecting astrophysical gravitational waves, we should think about what sources are likely to be most readily observable. The first important realization is that the necessary conditions for the generation of appreciable gravitational radiation are very different from those for electromagnetic radiation. The difference can be traced to the fact that gravitational waves are produced by the bulk motion of large masses, while electromagnetic waves are produced (typically) by incoherent excitations of individual particles. Electromagnetic radiation can therefore be produced by a source that is static in bulk, such as a star, which is a substantial advantage to the astronomer. However, gravitational waves are produced coherently by large moving masses (every particle in the mass contributes in the same sense to the wave), which can partially compensate for the impossibility of emission from static sources.

We therefore need massive sources with substantial bulk motions. As a simple example, consider the binary system of Section 7.5, in which both stars have mass M and the orbital radius is R. We will cheat somewhat by applying the Newtonian formulae for the orbital parameters in a regime where GR has begun to become important, but this will suffice for an order-of-magnitude estimate. The relevant parameters can be distilled down to the Schwarzschild radius $R_S = 2GM/c^2$,

[5]For an overview of gravitational-wave astrophysics, see S.A. Hughes, S. Márka, P.L. Bender, and C.J. Hogan, "New physics and astronomy with the new gravitational-wave observatories," `http://arxiv.org/astro-ph/0110349`.

the orbital radius R, and the distance r between us and the binary. (We will now restore explicit factors of c, to facilitate comparison with experiment.) In terms of these, the frequency of the orbit and thus of the produced gravitational waves is approximately

$$f = \frac{\Omega}{2\pi} \sim \frac{cR_S^{1/2}}{10R^{3/2}}. \tag{7.194}$$

From the formula (7.149) for the resulting perturbation, we can estimate the gravitational-wave amplitude received as

$$h \sim \frac{R_S^2}{rR}. \tag{7.195}$$

Let's see what this implies for the kind of source we might hope to observe. A paradigmatic example is the coalescence of a black-hole/black-hole binary. For typical parameters we can take both black holes to be 10 solar masses, the binary to be at cosmological distances $\sim$ 100 Mpc, and the components to be separated by ten times their Schwarzschild radii:

$$\begin{aligned} R_S &\sim 10^6 \text{ cm} \\ R &\sim 10^7 \text{ cm} \\ r &\sim 10^{26} \text{ cm}. \end{aligned} \tag{7.196}$$

Such a source is thus characterized by

$$f \sim 10^2 \text{ s}^{-1}, \qquad h \sim 10^{-21}. \tag{7.197}$$

If we are to have any hope of detecting the coalescence of a binary with these parameters, we need to be sensitive to frequencies near 100 Hz and strains of order 10^{-21} or less.

Fortunately, these parameters are within the reach of our experimental capabilities (with the heroic efforts of many scientists). The most promising technique for gravitational-wave detection currently under consideration is interferometry, and here we will stick exclusively to a discussion of interferometers, although it is certainly conceivable that a new technology could be invented that would have better sensitivity.

Recall that the physical effect of a passing gravitational wave is to slightly perturb the relative positions of freely-falling masses. If two test masses are separated by a distance L, the change in their distance will be roughly

$$\frac{\delta L}{L} \sim h. \tag{7.198}$$

Imagine that we contemplate building an observatory with test bodies separated by some distance of order kilometers. Then to detect a wave with amplitude of order $h \sim 10^{-21}$ would require a sensitivity to changes of

$$\delta L \sim 10^{-16} \left(\frac{h}{10^{-21}}\right) \left(\frac{L}{\text{km}}\right) \text{ cm.} \tag{7.199}$$

Compare this to the size of a typical atom, set by the Bohr radius,

$$a_0 \sim 5 \times 10^{-9} \text{ cm,} \tag{7.200}$$

or for that matter the size of a typical nucleus, of approximately a Fermi,

$$1 \text{ fm} = 10^{-13} \text{ cm.} \tag{7.201}$$

The point we are belaboring here is that a feasible terrestrial gravitational-wave observatory will have to be sensitive to changes in distance much smaller than the size of the constituent atoms out of which any conceivable test masses would have to be made.

Laser interferometers provide a way to overcome the difficulty of measuring such miniscule perturbations. Consider the schematic set-up portrayed in Figure 7.11. A laser (typically with characteristic wavelength $\lambda \sim 10^{-4}$ cm) is directed at a beamsplitter, which sends the photons down two evacuated tubes of length L. At the ends of the cavities are test masses, represented by mirrors suspended from pendulums. The light actually bounces off partially-reflective mirrors near the beamsplitter, so that a typical photon travels up and down the cavity

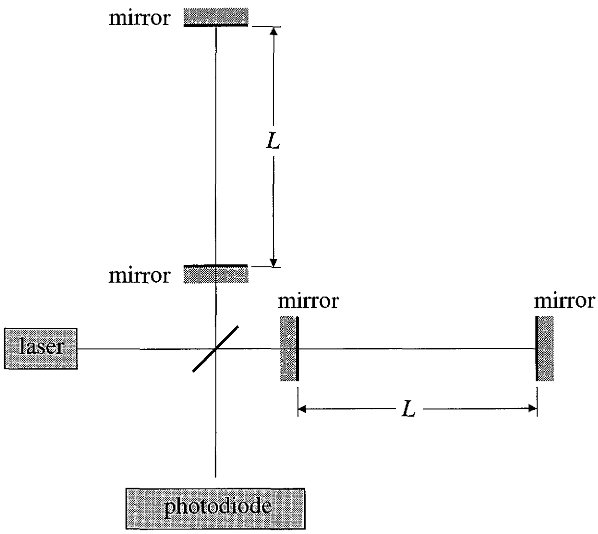

FIGURE 7.11 A schematic design for a gravitational-wave interferometer.

of order 100 times before returning to the beamsplitter and being directed into a photodiode. The system is arranged such that, if the test masses are perfectly stationary, the returning beams destructively interfere, sending no signal to the photodiode. As we have seen, the effect of a passing gravitational wave will be to perturb orthogonal lengths in opposite senses, leading to a phase shift in the laser pulse that will disturb the destructive interference. During 100 round trips through the cavity arms, the accumulated phase shift will be

$$\delta\phi \sim 200\left(\frac{2\pi}{\lambda}\right)\delta L \sim 10^{-9}, \qquad (7.202)$$

where 200 rather than 100 represents the fact that the shifts in the two arms add together. Such a tiny shift can be measured if the number of photons N is sufficiently large to overcome the "shot noise"; in particular, if $\sqrt{N} > \delta\phi$.

The technological challenges associated with building sufficiently quiet and sensitive gravitational-wave observatories are being tackled in a number of different locations, including the United States (LIGO), Italy (Virgo), Germany (GEO), Japan (TAMA), and Australia (ACIGA). LIGO (Laser Interferometric Gravitational-Wave Observatory) is presently the most advanced detector; it consists of two facilities (one in Washington state and one in Louisiana), each with four-kilometer arms. A single gravitational-wave observatory will be unable to localize a source's position on the sky; multiple detectors will be crucial for this task (as well as for verifying that an apparent signal is actually real).

Fundamental noise sources limit the ability of terrestrial observatories to detect low-frequency gravitational waves. Figure 7.12 shows the sensitivity regions, as a function of frequency, for two dramatically different designs: a terrestrial observatory such as LIGO, and a space-based mission such as LISA (Laser Interferometer Space Antenna). The general principle behind LISA is the same as any other interferometer, but the implementation is (or will be, if it is actually built) dramatically different. Current designs envision three spacecraft orbiting the Sun at approximately 30 million kilometers behind the Earth, separated from each other by 5 million kilometers. Due to the much larger separations, LISA is sensitive to frequencies in the vicinity of 10^{-2} Hz. The sensitivities portrayed in this plot should be taken as suggestive, as they depend on integration times and other factors.

Many potential noise sources confront the gravitational-wave astronomer. For ground-based observatories, the dominant effect at low frequencies is typically seismic noise, while at high frequencies it comes from photon shot noise and at intermediate frequencies from thermal noise. Advanced versions of ground-based detectors may be able to compensate for seismic noise at low frequencies, but will encounter irreducible noise from gravity gradients due to atmospheric phenomena or objects (such as cars) passing nearby. Satellite observatories, of course, are immune from such effects. Instead, the fundamental limitations are expected to come from errors in measuring changes in the distances between the spacecraft (or more properly, between the shielded proof masses within the spacecraft) and from nongravitational accelerations of the spacecraft.

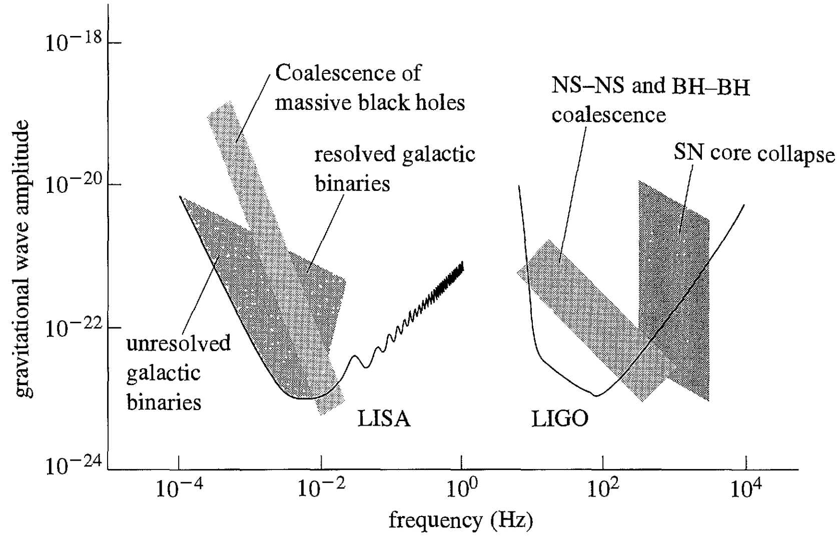

FIGURE 7.12 Sensitivities as a function of frequency for representative ground-based (LIGO) and space-based (LISA) gravitational-wave observatories, along with the expected signals from possible sources. Figure from the LISA collaboration home page (`http://lisa.jpl.nasa.gov/`).

We can conclude with a very brief overview of possible sources for gravitational-wave observatories. We have already mentioned the possibility of compact binaries of various sorts. For ground-based observatories, such sources will not become visible until they are very close to coalescence, and then only if the components are sufficiently massive (neutron stars or black holes). Extrapolating from what we know about such systems suggests that there may be several coalescences per year within a distance of a few hundred Mpc. Another promising possibility is core collapse in massive stars, giving rise to supernovae. Although a perfectly spherically-symmetric collapse would not generate any gravitational waves, realistic events are expected to be subject to instabilities that would break this symmetry. An exciting prospect is the coordinated observation of supernovae by ordinary telescopes and gravitational-wave observatories. Lastly, among possible sources for ground-based observatories are periodic sources such as (not-completely-axially-symmetric) rotating neutron stars. The amplitudes from such sources are expected to be small, but not necessarily completely out of reach of advanced detectors.

The interesting sources for space-based detectors are somewhat different. Most importantly, the known population of binaries in our galaxy will certainly provide a gravitational-wave signal of detectable magnitude. Indeed, unresolved binaries represent a source of confusion noise for the detector, as it will be impossible to pick out individual low-intensity sources from the background. Nevertheless, numerous higher-intensity sources should be easily observable. In addition, various processes in the evolution of supermassive black holes (greater than 1000 $M_\odot$, such as those found in the centers of galaxies) lead to interesting sources: the formation of such objects, their subsequent growth via accretion of smaller objects, and possible coalescence of multiple supermassive holes. Tracking the evolution of the gravitational-wave signal from a solar-mass black hole orbiting and eventually falling into a supermassive hole will allow for precision mapping of the spacetime metric, providing a novel test of GR.

In addition to waves produced by localized sources, we also face the possibility of stochastic gravitational-wave backgrounds. By this we mean an isotropic set of gravitational waves, perhaps generated in the early universe, characterized by a smoothly-varying power spectrum as a function of frequency. One possibility is a nearly scale-free spectrum of gravitational waves produced by inflation, as discussed in Chapter 8. Such waves will be essentially impossible to detect directly on the ground (falling perhaps five orders of magnitude below the capabilities of advanced detectors), or even by LISA, but could conceivably be observable by a next-generation space-based mission. More likely, any such waves will first become manifest in the polarization of the cosmic microwave background. Another possibility, however, is generation of primordial gravitational waves from a violent (first-order) phase transition. Such waves will have a spectrum with a well-defined peak frequency, related to the temperature T of the phase transition by

$$f_{\text{peak}} \sim 10^{-3} \left(\frac{T}{1000 \text{ GeV}} \right) \text{ Hz}. \tag{7.203}$$

Thus, a first-order electroweak phase transition ($T \sim 200$ GeV) falls within the band potentially observable by LISA. This is especially intriguing, as some models of baryogenesis require a strong phase transition at this scale; it is provocative to think that we could learn something significant about electroweak physics through a gravitational experiment.

7.8 ■ EXERCISES

1. Show that the Lagrangian (7.9) gives rise to the linearized version of Einstein's equation.

2. Consider a thin spherical shell of matter, with mass M and radius R, slowly rotating with an angular velocity Ω.
   (a) Show that the gravito-electric field $\vec{G}$ vanishes, and calculate the gravito-magnetic field $\vec{H}$ in terms of M, R, and Ω.
   (b) The nonzero gravito-magnetic field caused by the shell leads to dragging of inertial frames, known as the **Lense–Thirring effect**. Calculate the rotation (relative to the inertial frame defined by the background Minkowski metric) of a freely-falling observer sitting at the center of the shell. In other words, calculate the precession of the spatial components of a parallel-transported vector located at the center.

3. Fermat's principle states that a light ray moves along a path of least time. For a medium with refractive index $n(\mathbf{x})$, this is equivalent to extremizing the time

$$t = \int n(\mathbf{x})[\delta_{ij}dx^i dx^j]^{1/2} \tag{7.204}$$

along the path. Show that Fermat's principle, with the refractive index given by $n = 1 - 2\Phi$, leads to the correct equation of motion for a photon in a spacetime perturbed by a Newtonian potential.

4. Show that the Lorenz gauge condition $\partial_\mu \bar{h}^{\mu\nu} = 0$ is equivalent to the **harmonic gauge** condition. This gauge is defined by

$$\Box x^\mu = 0, \tag{7.205}$$

where each coordinate x^μ is thought of as a scalar function on spacetime. (Any function satisfying $\Box f = 0$ is known as an "harmonic function.")

5. In the exercises for Chapter 3, we introduced the metric

$$ds^2 = -(\mathrm{d}u\, \mathrm{d}v + \mathrm{d}v\, \mathrm{d}u) + a^2(u)\mathrm{d}x^2 + b^2(u)\mathrm{d}y^2, \tag{7.206}$$

where a and b are unspecified functions of u. For appropriate functions a and b, this represents an *exact* gravitational plane wave.

(a) Calculate the Christoffel symbols and Riemann tensor for this metric.

(b) Use Einstein's equation in vacuum to derive equations obeyed by $a(u)$ and $b(u)$.

(c) Show that an exact solution can be found, in which both a and b are determined in terms of an *arbitrary* function $f(u)$.

6. Two objects of mass M have a head-on collision at event $(0, 0, 0, 0)$. In the distant past, $t \to -\infty$, the masses started at $x \to \pm\infty$ with zero velocity.

(a) Using Newtonian theory, show that $x(t) = \pm(9Mt^2/8)^{1/3}$.

(b) For what separations is the Newtonian approximation reasonable?

(c) Calculate $h^{\mathrm{TT}}_{xx}(t)$ at $(x, y, z) = (0, R, 0)$.

7. Gravitational waves can be detected by monitoring the distance between two free flying masses. If one of the masses is equipped with a laser and an accurate clock, and the other with a good mirror, the distance between the masses can be measured by timing how long it takes for a pulse of laser light to make the round-trip journey. How would you want your detector oriented to register the largest response from a plane wave of the form

$$ds^2 = -\mathrm{d}t^2 + [1 + A\cos(\omega(t - z))]\,\mathrm{d}x^2 + [1 - A\cos(\omega(t - z))]\,\mathrm{d}y^2 + \mathrm{d}z^2?$$

If the masses have a mean separation L, what is the largest change in the arrival time of the pulses caused by the wave? What frequencies ω would go undetected?

8. The gravitational analog of *bremsstrahlung* radiation is produced when two masses scatter off each other. Consider what happens when a small mass m scatters off a large mass M with impact parameter b and total energy $E = 0$. Take $M \gg m$ and $M/b \ll 1$. The motion of the small mass can be described by Newtonian physics, since $M/b \ll 1$. If the orbit lies in the (x, y) plane and if the large mass sits at

$(x, y, z) = (0, 0, 0)$, calculate the gravitational wave amplitude for both polarizations at $(x, y, z) = (0, 0, r)$. Since the motion is not periodic, the gravitational waves will be burst-like and composed of many different frequencies. On physical grounds, what do you expect the dominant frequency to be? Estimate the total energy radiated by the system. How does this compare to the peak kinetic energy of the small mass?

Hint: The solution for the orbit can be found in Goldstein (2002). The solution is:

$$r = \frac{2b}{1 + \cos\theta},$$

$$t = \sqrt{\frac{2b^3}{M}} \left(\tan\frac{\theta}{2} + \frac{1}{3}\tan^3\frac{\theta}{2} \right).$$

Time runs from $t = (-\infty, \infty)$. Rather than using the above implicit solution for $\theta(t)$ you might want to use

$$\dot{\theta} = \sqrt{\frac{M}{8b^3}} (1 + \cos\theta)^2 .$$

9. Verify that the expression (7.165) for the gravitational-wave energy-momentum tensor is invariant under gauge transformations $h_{\mu\nu} \to h_{\mu\nu} + 2\partial_{(\mu}\xi_{\nu)}$.

10. Show that the integral expression (7.173) for the total energy in gravitational perturbations is independent of the spatial hypersurface Σ.