4 Equivalence of Gravitation and Inertia

- After a review of the Newtonian theory of gravitation in terms of its potential function, we take the first step in the study of general relativity (GR) with the introduction of the equivalence principle (EP).
- The weak EP, the equality of the gravitational and inertial masses, was extended by Einstein to the strong EP, the equivalence between inertia and gravitation for all interactions. This implies the existence of a local inertial frame (the frame in free fall) at every spacetime point. In a sufficiently small region, a local inertial observer will sense no gravitational effects.
- The equivalence of acceleration and gravity means that GR (physics laws valid in all coordinate systems, including accelerating frames) must be a theory of gravitation.
- The strong EP has physical consequences on time: gravitational redshift, time dilation, and the gravitational retardation of light speed leading to bending of light rays.
- Motivated by the EP, Einstein proposed a curved spacetime description of the gravitational field.

Soon after completing his formulation of special relativity (SR) in 1905, Einstein started working on a relativistic theory of gravitation. In this chapter, we cover mostly the period 1907–1911, when Einstein relied heavily on the equivalence principle (EP) to extract some general relativity (GR) results. Not until the end of 1915 did he work out fully the ideas of GR. By studying the consequences of the EP, he concluded that the proper language of GR is Riemannian geometry. In this, Einstein was helped by his ETH classmate and later colleague Marcel Grossmann (1878–1936). The mathematics of curved space will be introduced in Chapter 5. The curved spacetime representation of gravitational fields immediately suggests the geodesic equation as the GR equation of motion. In Chapter 6, we then begin the more difficult subject of the GR field equation and its solution.

A College Course on Relativity and Cosmology. First Edition. Ta-Pei Cheng.

4.1 Seeking a relativistic theory of gravitation

Before discussing GR, Einstein's field theory of gravitation, we review Newton's theory, whose field equation and equation of motion can be expressed in terms of the gravitational potential.

4.1.1 Newtonian potential: a summary

Newton formulated his theory of gravitation using a force that acts instantaneously between distant objects (action-at-a-distance force):

$$\mathbf{F}(\mathbf{r}) = -G_{\mathrm{N}}\frac{mM}{r^2}\hat{\mathbf{r}}, \tag{4.1}$$

where G_{N} is Newton's constant, M the point-source mass, m the test mass, and $\mathbf{r}$ the displacement of m from M.

Just as in electrostatics, where electric field is force per unit charge, $\mathbf{F}(\mathbf{r}) = q\mathbf{E}(\mathbf{r})$, we can cast Newton's gravitational force in the form

$$\mathbf{F} = m\mathbf{g}. \tag{4.2}$$

This defines the gravitational field $\mathbf{g}(\mathbf{r})$ as the gravitational force per unit mass. In terms of this field, Newton's law of gravitation for a point source of mass M reads

$$\mathbf{g}(\mathbf{r}) = -G_{\mathrm{N}}\frac{M}{r^2}\hat{\mathbf{r}}. \tag{4.3}$$

Just as Coulomb's law is equivalent to Gauss's law for the electric field, this field (4.3) can be expressed for an arbitrary mass distribution as Gauss's law for the gravitational field:

$$\oint_S \mathbf{g}\cdot d\mathbf{A} = -4\pi G_{\mathrm{N}} M. \tag{4.4}$$

The area integral on the left-hand side is the gravitational field flux out of a closed surface S, and M on the right-hand side is the total mass enclosed inside S. This integral representation of Gauss's law (4.4) can be converted into a differential equation. We first turn both sides into volume integrals. On the left-hand side, we use the divergence theorem (the area integral of the field flux equals the volume integral of the divergence of the field), while on the right-hand side, we express the mass in terms of the mass density function ρ:

$$\int \nabla\cdot\mathbf{g}\, dV = -4\pi G_{\mathrm{N}}\int \rho\, dV. \tag{4.5}$$

Since this relation holds for any volume, the integrands on both sides must be equal:

$$\nabla\cdot\mathbf{g} = -4\pi G_{\mathrm{N}}\rho. \tag{4.6}$$

This is Newton's field equation in differential form. We define the gravitational potential[1] $\Phi(\mathbf{r})$ through the field with $\mathbf{g} \equiv -\boldsymbol{\nabla}\Phi$, so the field equation (4.6) becomes

$$\nabla^2\Phi = 4\pi G_{\mathrm{N}}\rho. \tag{4.7}$$

[1] Recall the familiar example of the potential $\Phi = -G_{\mathrm{N}}M/r$ for a spherically symmetric source with total mass M.

To obtain the gravitational equation of motion, we insert (4.2) into Newton's second law, $\mathbf{F} = m\mathbf{a}$, canceling mass to get

$$\frac{d^2\mathbf{r}}{dt^2} = \mathbf{g}. \tag{4.8}$$

Thus the gravitational motion of a particle is totally independent of any of its properties (mass, charge, etc.). The acceleration can be expressed in terms of the gravitational potential as

$$\frac{d^2\mathbf{r}}{dt^2} = -\boldsymbol{\nabla}\Phi. \tag{4.9}$$

We note that the Newtonian field theory of gravitation, as embodied in (4.7) and (4.9), is not compatible with SR, because space and time coordinates are not treated on equal footings. Newtonian theory is a static field theory. Stated in another way, these equations are comparable to Coulomb's law in electrostatics. They do not account for the effects of motion (i.e., magnetism). This incompleteness just reflects the underlying assumption of an action-at-a-distance force, which implies an infinite speed of signal transmission, incompatible with the basic postulate of relativity.

4.1.2 Einstein's motivation for general relativity

Einstein's theory of gravitation has a unique history. It was not prompted by any empirical failure of Newton's theory, but resulted from pure thought, the theoretical speculation of one person drawing the consequences of fundamental principles. It sprang fully formed from Einstein's mind. As an old physicists' saying goes, "Einstein just stared at his own navel and came up with GR!"

From Einstein's published papers, one can infer a few interconnected motivations:

1. **Seeking a relativistic theory of gravitation** The Newtonian theory is not compatible with special relativity, as it invokes an action-at-a-distance force, implying infinitely fast signal transmission. Furthermore, inertial frames of reference, which are fundamental to SR, lose their privileged status in the presence of gravity.
2. "**Space is not a thing**" This is how Einstein phrased his conviction that physics laws should not depend on reference frames, which express the relationships among physical processes in the world but do not have independent existence.
3. **Why are inertial and gravitational masses equal?** Einstein strove for a deeper understanding of this empirical fact.

Einstein generalized Newton's field equation (4.7) and the equation of motion (4.9) to develop a relativistic theory valid in all coordinate systems.

4.2 The equivalence principle: from Galileo to Einstein

This section presents several properties of gravitation. They all follow from what Einstein called the principle of the equivalence of gravity and inertia. The final formulation of Einstein's theory of gravitation, the general theory of relativity, automatically and precisely includes this EP. Historically, it motivated a series of discoveries that ultimately led Einstein to the geometric theory of gravity, which models the gravitational field as warped spacetime.

4.2.1 Inertial mass vs. gravitational mass

One of the distinctive features of gravity is that its equation of motion (4.9) is totally independent of the test particle's properties. This comes about because of the cancellation of the mass factors in $m\mathbf{g}$ and $m\mathbf{a}$. Actually, these two masses correspond to very different concepts:

- The inertial mass m_{I} in Newton's second law,

$$\mathbf{F} = m_{\mathrm{I}}\mathbf{a}, \tag{4.10}$$

  enters into the description of the response of a particle to all forces.
- The gravitational mass m_{G} in Newton's law of gravitation,

$$\mathbf{F} = m_{\mathrm{G}}\mathbf{g}, \tag{4.11}$$

  reflects the response of a particle to a particular force: gravity. The gravitational mass m_{G} may be viewed (in analogy to electromagnetic theory) as the gravitational charge placed in a given gravitational field $\mathbf{g}$.

Now consider two objects, A and B, composed of different materials: one of copper and the other of wood. When they are let go in a given gravitational field $\mathbf{g}$ (e.g., dropped from the Leaning Tower of Pisa), they will, according to (4.10) and (4.11), obey the equations of motion:

$$(\mathbf{a})_{\mathrm{A}} = \left(\frac{m_{\mathrm{G}}}{m_{\mathrm{I}}}\right)_{\mathrm{A}} \mathbf{g}, \quad (\mathbf{a})_{\mathrm{B}} = \left(\frac{m_{\mathrm{G}}}{m_{\mathrm{I}}}\right)_{\mathrm{B}} \mathbf{g}. \tag{4.12}$$

Part of Galileo's great legacy is his experimental observation that all bodies fall with the same acceleration—that free fall is universal. The equality $(\mathbf{a})_{\mathrm{A}} = (\mathbf{a})_{\mathrm{B}}$ then leads to

$$\left(\frac{m_{\mathrm{G}}}{m_{\mathrm{I}}}\right)_{\mathrm{A}} = \left(\frac{m_{\mathrm{G}}}{m_{\mathrm{I}}}\right)_{\mathrm{B}}. \tag{4.13}$$

Because this mass ratio is universal for all substances, it can be set, by appropriate choice of units, equal to unity. We can simply say

$$m_{\rm I} = m_{\rm G}. \tag{4.14}$$

Even at the level of atomic physics, matter is made up of protons, neutrons, and electrons (all having different interactions) bound together with different binding energies. It is difficult to find an a priori reason to expect such a relation as (4.13). As we shall see, this is the empirical foundation underlying the geometric formulation of GR, the relativistic theory of gravity.

There is no record that Galileo ever measured the acceleration of freely falling objects (dropped from the Tower of Pisa or elsewhere)[2]. But he did measure the acceleration of objects sliding down an inclined plane; this slowed-down motion made measurements feasible. Newton achieved the same end by using a pendulum to measure the ratio $m_{\rm I}/m_{\rm G}$ for different objects.

[2] Interestingly Galileo provided a theoretical argument, a thought experiment, in the first chapter of his *Discourse and Mathematical Demonstration of Two New Sciences*, in support of the idea that all substances should fall with the same acceleration. Consider any falling object: without this universality of free fall, the tendency of different components of the object to fall differently would give rise to internal stress and could cause certain objects to spontaneously disintegrate. The nonobservation of this phenomenon could then be taken as evidence for equal accelerations.

Exercise 4.1 Physical examples of $m_{\rm I}/m_{\rm G}$ dependence

(a) For the frictionless inclined plane (with angle θ) in Fig. 4.1(a), find the acceleration's dependence on the ratio $m_{\rm I}/m_{\rm G}$. Thus a violation of the equivalence principle would show up as a material dependence in the time required for a material block to slide down the plane. (b) For the simple pendulum (with string length L) in Fig. 4.1(b), find the oscillation period's dependence on the ratio $m_{\rm I}/m_{\rm G}$.

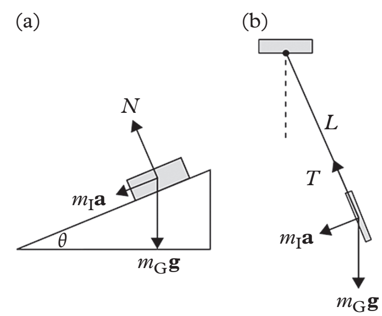

Figure 4.1 *Both the gravitational mass and inertial mass enter in these phenomena: (a) a sliding object on an inclined plane, where **N** is the normal force, and (b) oscillations of a pendulum, where **T** is the tension force in the string.*

4.2.2 Einstein: "my happiest thought"

In the course of writing a review article on SR in 1907, Einstein came upon what he later termed, "my happiest thought." He recalled the fundamental experimental result of Galileo that all objects fall with the same acceleration. Since all bodies accelerate the same way, an observer in a freely falling laboratory will not be able to detect any gravitational effect (on a point particle) in this frame. That is to say, gravity is transformed away in reference frames in free fall.

Principle of equivalence stated Imagine an astronaut in a freely falling spaceship. Because all objects fall with the same acceleration, a released object will not fall with respect to the spaceship. Thus, from the viewpoint of the astronaut, gravity is absent; everything becomes weightless. To Einstein, this vanishing of the gravitational effect was so significant that he elevated it (in order to focus on it) to a physical principle, the equivalence principle:

$$\begin{pmatrix} \text{Physics in a frame freely falling in a gravitational field} \\ \text{is equivalent to} \\ \text{physics in an inertial frame without gravity} \end{pmatrix}.$$

Namely, within a freely falling frame, where the acceleration exactly cancels the uniform gravitational field, no sign of either acceleration or gravitation can be found by any physical means. Correspondingly,

$$\begin{pmatrix} \text{Physics in a nonaccelerating frame with a gravitational field } \mathbf{g} \\ \text{is equivalent to} \\ \text{physics in a frame without gravity but accelerating with } \mathbf{a} = -\mathbf{g} \end{pmatrix}.$$

Absence of gravity in an inertial frame According to the EP, accelerating frames of reference can be treated in exactly the same way as frames with gravity. This suggests a new definition of an inertial frame, without reference to any external environment such as fixed stars, as a frame in which there is no gravity. Our freely falling spaceship can thereby be deemed an inertial frame of reference; our astronaut is now an inertial observer. Einstein realized the unique position of gravitation in the theory of relativity. Namely, he understood that the question was not how to incorporate gravity into SR, but rather how to use gravitation as a means to broaden the principle of relativity from inertial frames to all coordinate systems, including accelerating frames.

From EP to gravity as the structure of spacetime If we confine ourselves to the physics of mechanics, the EP is just a restatement of $m_{\mathrm{I}} = m_{\mathrm{G}}$. But once promoted to a principle, it allowed Einstein to extend this equivalence between inertia and gravitation to all physics—not just to mechanics, but also to electromagnetism, etc. This generalized version is sometimes called the strong equivalence principle. Thus the weak EP is just the statement of $m_{\mathrm{I}} = m_{\mathrm{G}}$, while the strong EP is the principle of equivalence applied to all physics.[3] Henceforth, we shall still call the strong equivalence principle EP for short. Because the motion of a test body in a gravitational field is independent of the properties of the body, Einstein came up with the idea that the effect of gravity on the body can be attributed directly to some spacetime feature, and that gravity is nothing but the structure of a warped spacetime.

[3] Or, the EP says that one can form an inertial frame at any point in spacetime in which matter satisfies the laws of SR. The strong EP implies the validity of this equivalence principle for all laws of nature, not just mechanics.

Exercise 4.2 Two EP brain-teasers

Even in mechanics, in some instances, the (weak) EP can be very useful in helping us to obtain physics results with simple analysis. Here are two notable examples: (a) Use the EP to explain the observation (see Fig. 4.2a) that a helium balloon leans forward *in a (forward-) accelerating vehicle. (b) On his 76th birthday, Einstein received a gift from his Princeton neighbor Eric Rogers.[4] It was a toy composed of a ball attached by a spring to the inside of a bowl (a toilet plunger), which was just the right size to hold the ball. The upright bowl was fastened to a broomstick as in Fig 4.2(b). What is the* surefire way, *suggested by the EP, to pop the ball back into the bowl each time?*

[4] Referenced in, e.g., (French 1979, p 131).

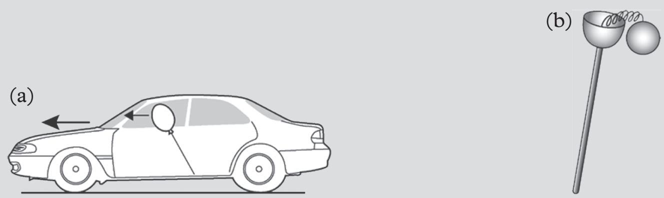

Figure 4.2 *Illustrations for the two EP brain-teasers in Exercise 4.2.*

4.3 EP leads to gravitational time dilation and light deflection

The strong EP implies that gravity can bend a light ray, shift the frequency of an electromagnetic wave, and cause clocks to run slow. Ultimately, these results suggested to Einstein that the proper framework to describe the relativistic effects of gravity is a curved spacetime. These gravitational time dilation phenomena will be interpreted as reflecting the warping of spacetime in the time direction.

4.3.1 Gravitational redshift and time dilation

To deduce the SR effects of relative motion, we often compare observations made in different coordinate frames. Similarly, one can obtain the effects of gravity on certain physical phenomena using the following general procedure:

1. One first considers the description by an observer inside a spaceship in free fall. According to the EP, there is no gravitational effect in this inertial frame, and SR applies.
2. One then considers the same situation from the viewpoint of an observer watching the spaceship from outside: there is a gravitational field in which the first (freely falling) observer is seen to be accelerating.
3. The combined effects of acceleration and gravity, as seen by the second observer, must then reproduce the SR description as recorded by the inertial observer in free fall. Physics should be independent of coordinate choices.

Bending of a light ray—a qualitative account

Let us first study the effect of gravity on a light ray traveling (horizontally) across a spaceship that is falling in a constant (vertical) gravitational field $\mathbf{g}$. The EP informs us that from the viewpoint of the astronaut in the spaceship, there is no

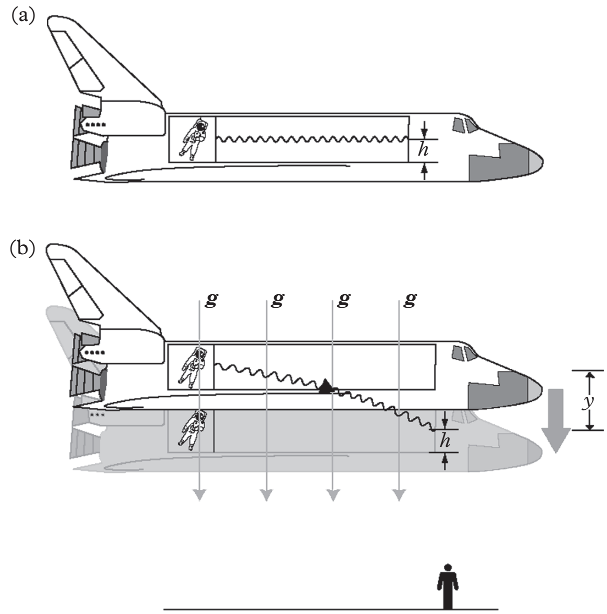

Figure 4.3 *According to the EP, a light ray will fall in a gravitational field. (a) To the astronaut in the freely falling spaceship (an inertial observer), the light trajectory is straight. (b) To an observer outside the spaceship, the astronaut is accelerating (falling) in a gravitational field. The light ray will be bent so that it reaches the opposite side at a height $y = gt^2/2$ below the initial point; it falls with the spaceship.*

detectable effect associated either with gravity or with acceleration: the light travels straight across the spaceship from one side to the other; it is emitted at a height h and received at the same height h on the opposite side of the spaceship as in Fig. 4.3(a). But to an observer outside, the spaceship is accelerating (falling) in a gravitational field $\mathbf{g}$. To this outside observer, the light ray bends as it traverses the falling spaceship as in Fig. 4.3(b). Thus the EP predicts that gravity bends light.

We do not ordinarily see light curving; for the gravitational field and distance scale with which we are familiar, this bending effect is unobservably small. Consider a lab with a width of 300 m. A light ray travels across the lab in 1 μs. During this interval, the light drops (is bent by) an extremely small amount: $y = gt^2/2 \simeq 5 \times 10^{-12}\,\text{m} = 0.05\,\text{Å}$. This suggests that one needs to seek such effects where large masses and great distances are involved—as in astronomical settings.

Gravitational redshift

Above, we discussed the effect of a gravitational field on a light ray whose trajectory is transverse to the field direction. Now let us consider the situation when the field is parallel (or antiparallel) to the ray's path as in Fig. 4.4.

Here we have a receiver placed at a distance h directly above the emitter in a downward-pointing gravitational field $\mathbf{g}$. Just as in the transverse case considered above, we first describe the situation from the viewpoint of the astronaut (in free fall), Fig. 4.4(a). The EP informs us that the spaceship in free fall is an inertial

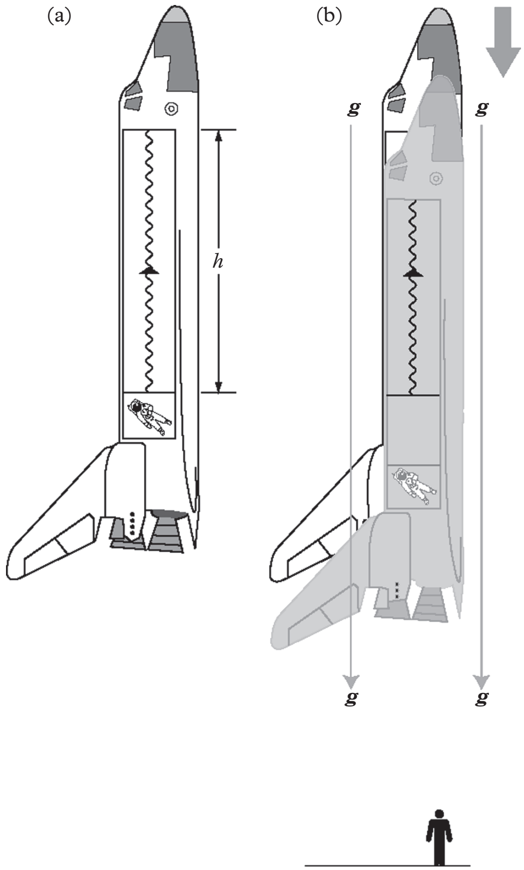

Figure 4.4 *The EP implies a gravitational redshift. Light is sent upward and received at a height h. (a) To an inertial observer in the freely falling spaceship, there is no frequency shift. (b) To an observer outside the spaceship, this astronaut is accelerating in a gravitational field. In this frame, a gravitational redshift cancels the Doppler blueshift, so both observers agree (as they must) that the frequency does not change.*

frame, with no physical effects associated with gravity or acceleration. The astronaut should not detect any frequency shift; the received light frequency ω_{rec} is the same as the emitted frequency ω_{em}:

$$(\Delta\omega)_{\text{ff}} = (\omega_{\text{rec}} - \omega_{\text{em}})_{\text{ff}} = 0, \tag{4.15}$$

where the subscript ff reminds us that these are the values as seen by an observer in free fall.

From the viewpoint of the observer outside, the spaceship is accelerating (falling) in a gravitational field as in Fig. 4.4(b). Assume that this spaceship starts to fall at the moment of light emission. It takes a finite amount of time $\Delta t = h/c$ for the light signal to reach the receiver on the ceiling (to lowest order in gh/c^2, we can neglect the change in the receiver's position during the light's travel). During that time, the falling receiver accelerates to a downward velocity $\Delta u = g\Delta t$. The familiar Doppler formula in the low-velocity approximation (3.54) would lead us to expect a frequency shift of

$$\left(\frac{\Delta\omega}{\omega}\right)_{\text{Doppler}} = \frac{\Delta u}{c}. \tag{4.16}$$

Since the receiver above is moving toward the emitter below, the light waves must be compressed; this shift must be toward the blue:

$$(\Delta\omega)_{\text{Doppler}} = (\omega_{\text{rec}} - \omega_{\text{em}})_{\text{Doppler}} > 0. \tag{4.17}$$

But, as stated in (4.15), the inertial observer (in free fall) sees no such shift; the received frequency does not deviate from the emitted frequency. This physical result must hold for both observers, so the blueshift in (4.17) must somehow be canceled. To the observer outside the spaceship, gravity is also present. We can recover the nullshift result if the light is redshifted by gravity by just the right amount to cancel the Doppler blueshift of (4.16):

$$\left(\frac{\Delta\omega}{\omega}\right)_{\text{gravity}} = -\frac{\Delta u}{c}. \tag{4.18}$$

We now express the relative velocity on the right-hand side in terms of the gravitational potential difference $\Delta\Phi$ between the two locations:

$$\Delta u = g\Delta t = \frac{gh}{c} = \frac{\Delta\Phi}{c}. \tag{4.19}$$

By combining (4.18) and (4.19), we obtain the gravitational frequency shift:

$$\frac{\Delta\omega}{\omega} = -\frac{\Delta\Phi}{c^2}; \tag{4.20}$$

namely,[5]

$$\frac{\omega_1 - \omega_2}{\omega_2} = -\frac{\Phi_1 - \Phi_2}{c^2}. \tag{4.21}$$

Light emitted at a lower gravitational potential (Φ_2) will be received at a higher gravitational potential ($\Phi_1 > \Phi_2$) with a lower frequency ($\omega_1 < \omega_2$); that is, it is redshifted, even though the emitter and the receiver are not in relative motion. Likewise, light emitted at a higher potential appears blueshifted to a receiver at a lower potential.

The Pound–Rebka experiment In principle, this gravitational redshift can be tested by a careful examination of the spectral emission lines from an astronomical object (hence from a deep gravitational potential well). However, such an effect can easily be masked by the standard Doppler shifts due to the thermal motion

[5] Whether the denominator is ω_1 or ω_2, the difference is of higher order and can be ignored in these leading-order formulae.

of the emitting atoms. Consequently, conclusive data did not exist in the first few decades after Einstein's paper. It was not until 1960 that the gravitational redshift was first convincingly verified in a series of terrestrial experiments in which Robert Pound (1919–2010) and his collaborators measured the tiny frequency shift of radiation traveling up $h = 22.5$ m, the height of an elevator shaft in the building housing the Harvard Physics Department:

$$\left|\frac{\Delta\omega}{\omega}\right| = \left|\frac{gh}{c^2}\right| = O(10^{-15}). \tag{4.22}$$

Normally, it is not possible to fix the frequency of an emitter or absorber to a very high accuracy, because of the energy shift due to thermal recoils of the atoms. However, owing to the Mössbauer effect,[6] the emission line width in a rigid crystal is as narrow as possible—limited only by the quantum mechanical uncertainty principle, $\Delta t \Delta E \geq \hbar$, where Δt is the lifetime of the unstable (excited) state. Thus a long-lived state would have a particularly small energy/frequency spread. The Harvard experimenters (Pound and Rebka 1960) worked with an excited isotope of iron, $^{57}\mathrm{Fe}^*$, produced through the nuclear beta decay of cobalt-57. It transitions to the ground state by emitting a gamma ray: $^{57}\mathrm{Fe}^* \to {}^{57}\mathrm{Fe} + \gamma$. The gamma ray emitted this way at the bottom of the elevator shaft, after climbing the 22.5 m, could no longer be resonantly absorbed by a sheet of iron in the ground state placed at the top of the shaft. To prove that the radiation had been redshifted by just the right $O(10^{-15})$ amount, Pound and Rebka introduced an (ordinary) Doppler blueshift by moving the detector slowly toward the emitter at just the right speed to compensate for the gravitational redshift. Thus, the radiation could again be absorbed. What was the speed with which they had to move the receiver? From (4.20) and (4.16), we have

$$\frac{gh}{c^2} \underset{\text{gravity}}{=} \left|\frac{\Delta\omega}{\omega}\right| \underset{\text{Doppler}}{=} \frac{u}{c}, \tag{4.23}$$

with

$$u = \frac{gh}{c} = \frac{9.8\ \mathrm{m/s^2} \times 22.5\ \mathrm{m}}{3 \times 10^8\ \mathrm{m/s}} = 7.35 \times 10^{-7}\ \mathrm{m/s}. \tag{4.24}$$

This is such a small speed that it would take $h/u = c/g = O(3 \times 10^7\ \mathrm{s}) \simeq 1$ year to cover the same elevator shaft height. Of course this velocity is just the one attained by an object freely falling for the time interval $O(10^{-7}\ \mathrm{s})$ that it takes the light to climb the elevator shaft. This is the compensating effect we invoked in our derivation of the gravitational redshift at the beginning of this section.

Gravitational time dilation

From our understanding of the Doppler effect, this gravitational frequency shift looks absurd. How can an observer, stationary with respect to the emitter, receive

[6] The Mössbauer effect: Atomic recoil can reduce the energy of an emitted photon. Since the emitting atom is surrounded by other atoms in thermal motion, this recoil is uncontrollable. (We can picture the atom as part of a vibrating lattice.) As a result, the photon energies in different emission events can vary considerably, resulting in a significant spread of their frequencies. This rules out high-precision measurements of the atomic frequency for purposes such as testing the gravitational redshift. But, in 1958, Rudolf Mössbauer (1929–2011) made a breakthrough when he pointed out, and verified by observation, that crystals with high Debye–Einstein temperature (i.e., having a rigid crystalline structure) could pick up the recoil by the entire crystal. Namely, in such a situation, the emitting atom has a huge effective mass. Consequently, the atom carries away no recoil energy; the photon can pick up all the energy lost by the emitting atom, and the frequency of the emitted radiation is as precise as it can be.

a different number of wave crests per unit time than the emitted rate? Here is Einstein's radical yet simple answer: while the number of wave crests does not change, the time unit itself changes in the presence of gravity. Clocks at different gravitational potentials run at different rates; there is a gravitational time dilation effect.

Frequency is proportional to the inverse of the local proper time rate $\omega \sim 1/d\tau$; the gravitational frequency shift formula (4.21) can be converted to a time dilation formula:

$$\frac{d\tau_1 - d\tau_2}{d\tau_2} = \frac{\Phi_1 - \Phi_2}{c^2}. \tag{4.25}$$

Namely, a clock at higher gravitational potential ($\Phi_1 > \Phi_2$) will run faster ($d\tau_1 > d\tau_2$); a lower clock runs slow. The fast/slow descriptions reflect the larger/smaller elapsed time intervals. All observers agree on this, since $d\tau_1$ and $d\tau_2$ are scalars. Another derivation of (4.25) will be presented in Box 4.1.

Contrast this gravitational time dilation with the distinct SR effect of (2.22): $dt = \gamma\, d\tau$. If two observers are in relative motion, each perceives the other's clock to run slow by a factor of γ. Obviously, one must be careful to understand which clocks rate is being described and by whom. Recall the twin paradox discussed in Box 2.3. The time dilation of Al's clock in motion means its time $d\tau_A$ is measured by a stationary Bill to run slow, $dt_B > d\tau_A$ as $dt_B = \gamma_A\, d\tau_A$; when the twins meet, Bill has aged 50 years compared with Al's 30. Consider a comparable case in which Bill lives on the top floor of a high-rise building and Al at the bottom ($\Phi_B > \Phi_A$). The twins will again age differently, now because of gravitational time dilation. Bill's clock runs faster; as the years pass by, he will be older than the low-level-dwelling Al. In Section 4.3.2, orbiting satellites provide yet another concrete example illustrating these different types of time dilation.

Box 4.1 Gravitational time dilation—another derivation

As discussed in our SR chapters, the standard method to compare clock readings is through the exchange of light signals. Our way of arriving at the gravitational time dilation result (4.25) followed this standard method—an exchange of light signals and a comparison of light frequencies—leading to the gravitational redshift and then time dilation results.

Here we present another derivation of (4.25) that will display its relation to (and its compatibility with) the familiar SR effect of time dilation due to relative motion. Two clocks are located at different gravitational potential points, Clock-1 at Φ_1 and Clock-2 at Φ_2. Let Clock-3 fall freely in this gravitational field (see Fig. 4.5); when it passes Clock-1, it has the speed u_1 and when it passes Clock-2, the speed u_2. At the instant when Clock-3 passes Clock-1, both clocks are at the same gravitation potential Φ_1. Thus a comparison of the clocks' rates involves only the SR effect due to their relative motion of speed u_1. A similar comparison can be carried out when Clock-3 passes Clock-2 at Φ_2 with

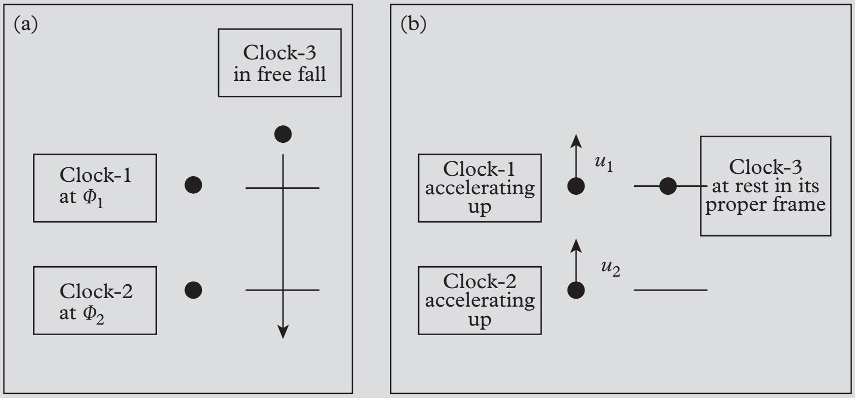

Figure 4.5 *Comparing clock rates at different gravitational potential points: (a) Clock-3 in free fall passing Clock-1 at Φ_1 with speed u_1 and Clock-2 at Φ_2 with speed u_2. (b) Clock-1 is seen moving with speed $-u_1$ in the inertial (free-fall and gravity-free) frame of Clock-3 when the latter passes Clock-1. A picture similar to (b) can be drawn when Clock-3 passes Clock-2.*

speed u_2. One can relate the proper clock rates $d\tau_1$ and $d\tau_2$ to the coordinate time rate in Clock-3's freely falling frame of reference. By the SR time dilation formula at the respective Φ_1 and Φ_2 points,

$$dt_3^{(1)} = \gamma_1\, d\tau_1, \quad \text{and} \quad dt_3^{(2)} = \gamma_2\, d\tau_2, \tag{4.26}$$

with $\gamma_{1,2} = (1 - u_{1,2}^2/c^2)^{-1/2}$. Because Clock-3 is in free fall, gravity is absent in its inertial reference frame; time dt_3 passes at a constant rate as the other clocks accelerate by: $dt_3^{(1)} = dt_3^{(2)}$. The time dilation result (4.25) can then be derived by connecting the two equations in (4.26):

$$\begin{aligned}\frac{d\tau_1}{d\tau_2} &= \frac{\gamma_2}{\gamma_1} = \left(\frac{1 - u_1^2/c^2}{1 - u_2^2/c^2}\right)^{1/2} \\ &\simeq 1 - \frac{1}{2}\frac{u_1^2 - u_2^2}{c^2} = 1 + \frac{\Phi_1 - \Phi_2}{c^2},\end{aligned} \tag{4.27}$$

where, to reach the second line, we have dropped terms $O(u^4/c^4)$ in the power series expansions of the denominator and of the square root. At the last equality, we have used the low-velocity version (consistent with our presentation) of the energy conservation relation for the freely falling Clock-3. The change in kinetic energy must equal minus the change in potential energy: $\frac{1}{2}m\Delta u^2 = -m\Delta\Phi$.

This derivation of (4.25) shows that gravitational time dilation is entirely compatible with the previously known SR time dilation effect—just as we showed its compatibility with the Doppler frequency shift (4.16) in our first derivation of gravitational redshift (4.20).

4.3.2 Relativity and the operation of GPS

The Global Positioning System (GPS) is capable of fixing locations on earth by exchange of electromagnetic signals between the ground and a network of orbiting satellites. More than 24 satellites are distributed more-or-less uniformly among orbits around the globe (e.g., four satellites on each of six equally spaced orbital planes), so that for any point on earth, $\mathbf{r}$, there are at least four satellites at $\mathbf{r}_i$ $(i = 1, 2, 3, 4)$ above its horizon (Fig. 4.6). Radar signals encode each satellite's position and time at transmission $(\mathbf{r}_i, t_i)$. The GPS receiver on the ground then calculates the four unknowns of its position and time of reception $(\mathbf{r}, t)$ by solving the four simultaneous equations $|\mathbf{r}_i - \mathbf{r}| = c|t_i - t|$. Since distance is obtained from timing measurements, $\Delta r = c\Delta t$, one needs an extremely accurate knowledge of the transmission time, as a difference of one nanosecond (10^{-9} s) translates into a distance of about one foot, $10^{-9}\,\text{s} \times 3 \times 10^8\,\text{m/s} = 0.3\,\text{m}$.

To achieve such accuracy, each satellite carries an atomic clock. Still, all the time measurements must account for the fact that the clocks on the satellites are moving at high speeds and are at different gravitational potentials with respect to the ground location. How large are the resulting SR and gravitational time dilation effects? In the next paragraph, we carry out an order-of-magnitude calculation to show that both effects are significant, and furthermore that the gravitational effect is opposite to, and about six times bigger than, the effect due to relative motion.

Satellites high up in space and moving with high speed To determine the relativistic effects, we need to know the orbital radius of the satellite (r_S) and how fast it is moving (v_S). Each of the satellites is set to have a period of 12 hours, so that a satellite passes overhead of a given observer on earth twice each day. We will find (r_S, v_S) by solving the coupled equations representing the orbital period (12 hours) and the gravitational equation of motion with the centripetal acceleration $a = v^2/r$:

$$\begin{aligned} \frac{2\pi r_S}{v_S} &= 12\,\text{h} = 12 \times 3600\,\text{s}, \\ \frac{G_N M_\oplus}{r_S^2} &= \frac{F}{m} = a = \frac{v_S^2}{r_S}, \end{aligned} \tag{4.28}$$

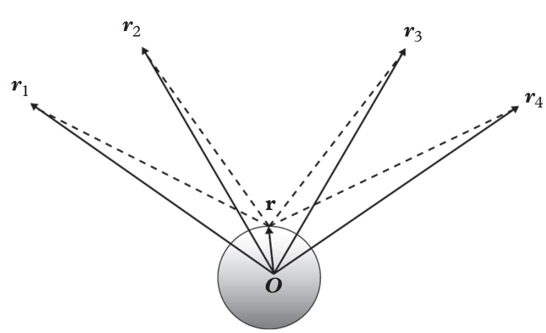

Figure 4.6 *Light signals are exchanged between a location r on earth and four GPS satellites located at r_i $(i = 1,2,3,4)$.*

where $M_\oplus$ is the mass of the earth. Solving these two equations, we obtain the satellite's orbital radius in terms of the earth's radius $r_\oplus$,

$$r_S = 2.7 \times 10^7 \text{m} = 4.2\, r_\oplus, \tag{4.29}$$

and the tangential speed $v_S = 3.87$ km/s. Such a velocity gives a relativistic beta factor of $\beta_S = v_S/c = 1.3 \times 10^{-5}$ and a Lorentz gamma factor of

$$\gamma_S = (1-\beta_S^2)^{-1/2} = 1 + 0.83 \times 10^{-10}. \tag{4.30}$$

Time dilation due to relative motion From this, we can calculate the SR time dilation correction, as both the satellite clock and the earthbound clock are in motion. Let t_0 be the time recorded by a clock at rest with respect to the coordinate origin (the center of the earth). The time dilation of the moving satellite clock is $dt_0 = \gamma_S\, d\tau_S$ and that of the clock on earth, $dt_0 = \gamma_\oplus\, d\tau_\oplus$. Thus the relative dilation effect due to the motion of the satellite and the ground clocks around the earth is given by

$$\left(\frac{d\tau_\oplus}{d\tau_S}\right)_{\text{mo}} = \frac{\gamma_S}{\gamma_\oplus} = 1 + \frac{v_S^2 - v_\oplus^2}{2c^2}. \tag{4.31}$$

Since the ground clock speed $v_\oplus$ depends on its location on earth and is small,[7] we will drop the $v_\oplus^2$ term. The fractional time difference due to relative motion is, according to (4.30),

$$\mathcal{F}_{\text{mo}} \equiv \left(\frac{d\tau_\oplus - d\tau_S}{d\tau_S}\right)_{\text{mo}} = \left(\frac{d\tau_\oplus}{d\tau_S}\right)_{\text{mo}} - 1$$
$$= \frac{1}{2}\beta_S^2 = 0.83 \times 10^{-10}. \tag{4.32}$$

With $d\tau_S < d\tau_\oplus$, we say that the satellite clock runs slower than the clock on the ground. This is so because the satellite time is dilated more than the ground clock, reflecting their relative speed $\gamma_S > \gamma_\oplus$.

Time dilation due to relative heights The gravitational time dilation formula (4.25) yields a fractional time difference of

$$\mathcal{F}_{\text{grav}} \equiv \left(\frac{d\tau_\oplus - d\tau_S}{d\tau_S}\right)_{\text{grav}} = \frac{\Phi_\oplus - \Phi_S}{c^2}, \tag{4.33}$$

where $\Phi_S > \Phi_\oplus$ as the satellite is located at a potential that is less negative,

$$\Phi_\oplus = -\frac{G_N M_\oplus}{r_\oplus} \quad \text{and} \quad \Phi_S = -\frac{G_N M_\oplus}{r_S} = \frac{\Phi_\oplus}{4.2}. \tag{4.34}$$

[7] This is the case even for a receiver located on the equator with $v_\oplus = 2\pi R_\oplus/24\,\text{h}$. Comparing it with $v_S = 2\pi r_S/12\,\text{h}$, we have $v_S/v_\oplus = 2r_S/R_\oplus = 8.4$; thus the correction $(v_\oplus/v_S)^2$ is only about 1%.

In the last term, we have inserted the result from (4.29). Given that $\Phi_\oplus = -gr_\oplus$, where $g = 9.8\,\text{m/s}^2$ is the gravitational acceleration at the earth's surface, one immediately obtains

$$\mathcal{F}_{\text{grav}} = \left(\frac{d\tau_\oplus}{d\tau_S}\right)_{\text{grav}} - 1 \tag{4.35}$$

$$= -\frac{gr_\oplus}{c^2}\left(1 - \frac{1}{4.2}\right) = -5.3 \times 10^{-10}.$$

With the elapsed satellite time interval being larger, $d\tau_S > d\tau_\oplus$, we say that the satellite clock runs faster than the clock on the ground, in contrast to the effect due to relative motion.

Full relativistic correction We see that this gravitational correction is about six times as great as that due to SR time dilation (4.32) and in the opposite direction. When we combine the leading-order corrections due to the gravitational potential difference (4.33) and to the relative motion (4.31), the total relativistic correction may be written as

$$\left(\frac{d\tau_\oplus}{d\tau_S}\right)_{\text{rel}} = \left(\frac{d\tau_\oplus}{d\tau_S}\right)_{\text{grav}} \times \left(\frac{d\tau_\oplus}{d\tau_S}\right)_{\text{mo}} \tag{4.36}$$

$$= 1 + \frac{\Phi_\oplus - \Phi_S}{c^2} - \frac{v_\oplus^2 - v_S^2}{2c^2}.$$

As we shall see, this result is contained in the full GR solution to be discussed in Chapter 6 (cf. Example 6.1). Without taking the relativistic correction into consideration, errors will rapidly accumulate in the GPS time measurements. For example, in a period of only one minute, by ignoring time dilation due to relative motion, one would incur an error of $\mathcal{F}_{\text{mo}} \times 60\,\text{s} \simeq 5\,\text{ns}$, while the error from ignoring the gravitational time dilation would be $\mathcal{F}_{\text{grav}} \times 60\,\text{s} \simeq -30\,\text{ns}$; taken together, one gets an error of about 25 ns. As discussed in the introductory paragraph, one nanosecond corresponds to a distance of about a foot. Without taking relativistic effects into account, there is no way that the GPS system could pinpoint locations within a few feet or so in accuracy.

Exercise 4.3 A GPS calculation

We expect that the gravitational time dilation could be reduced if the satellite traveled in a lower orbit. Figure out how low the satellite's orbital radius r_S must be so that the time dilation effects due to gravity and due to relative motion cancel each other? What will be its period? Just as we did above, you may neglect $v_\oplus$. (a) From the requirement that $\mathcal{F}_{\text{mo}} + \mathcal{F}_{\text{grav}} = 0$, you should be able to deduce the relation of velocity v_S to orbital radius r_S:

$$v_S^2 = 2gr_\oplus\left(1 - \frac{r_\oplus}{r_S}\right). \tag{4.37}$$

Note that for this problem, it is entirely adequate to approximate the gamma factor by $\gamma \equiv (1-\beta^2)^{-1/2} \simeq 1 + \frac{1}{2}\beta^2$. *(b) From the centripetal acceleration equation, together with (4.37), one can then find the solution:*

$$v_S^2 = \frac{2}{3} g r_\oplus \qquad \text{and} \qquad r_S = \frac{3}{2} r_\oplus. \tag{4.38}$$

(c) What will be the resultant orbital period (in contrast to the real GPS system's 12-hour period)? From these numbers, you should be able to conclude that such a system of low-flying satellites may have some practical problems.

4.3.3 The EP calculation of light deflection

The EP implies that clocks run at different rates at locations where the gravitational potentials are different. Such effects will lead to different speed measurements—even the speed of light can be measured to have different values! We are familiar with light speeds in different media being characterized by varying indices of refraction. Gravitational time dilation implies that even in vacuum there is an effective index of refraction when a gravitational potential is present. Since potential usually varies in space (i.e., its gradient, the gravitational field, is usually nonzero), this index is generally a position-dependent function.

Gravity-induced index of refraction in free space At a given position r with gravitational potential $\Phi(r)$, a determination of light speed with respect to the local proper time $d\tau$ and local proper length $d\rho$ gives

$$\frac{d\rho}{d\tau} = c. \tag{4.39}$$

This speed c is a universal constant. On the other hand, the light speed will deviate from c according to the elapsed time and distance displacement measured by the clock and ruler at a different position. In fact, a common choice of coordinate time and distance is that given by the clock and ruler located far away from the gravitational source. Equation (4.25) then gives us the relation between the local time ($d\tau = d\tau_1$) measured at r and the coordinate time ($dt = d\tau_2$) measured where $\Phi_2 = 0$:

$$d\tau = \left(1 + \frac{\Phi(r)}{c^2}\right) dt. \tag{4.40}$$

What about the gravitational effect on length measurement? The deflection of light by a gravitational source was first predicted by Einstein in 1911. The calculation was based on EP-implied gravitational time dilation alone. This was before Einstein had developed the idea of gravity as structure of spacetime. Since most of the first strong-EP effects to be discussed were those of gravity on time measurement, Einstein did not discuss in his paper the influence of gravity on length measurement. Thus, in effect, he set $d\rho = dr$, and this is what we do here as well. In this way, to a remote observer, light speed is reduced by gravity[8] (Φ being negative):

[8] As we shall see in Section 6.4.2, this deviation of measured light speed from c is half as much as the full GR solution.

$$[c(r)]_{\mathrm{EP}} \equiv \frac{dr}{dt} = \frac{1+\Phi(r)/c^2}{d\tau}\,d\rho = \left(1+\frac{\Phi(r)}{c^2}\right)c, \tag{4.41}$$

which varies from position to position as the gravitational potential varies. For the faraway observer, the effect of the gravitational field can be viewed as introducing an index of refraction in space:

$$\frac{1}{[n(r)]_{\mathrm{EP}}} \equiv \frac{[c(r)]_{\mathrm{EP}}}{c} = \left(1+\frac{\Phi(r)}{c^2}\right). \tag{4.42}$$

Let us reemphasize some key concepts behind this position-dependent speed of light. We are not suggesting that the deviation of $c(r)$ from the constant c means that the speed of light measured by a local observer has changed, or that the velocity of light is no longer a universal constant in the presence of gravitational fields. Rather, it reflects the physics that clocks at different gravitational potentials run at different rates. For an observer located far from the gravitational source (whose proper time is conveniently taken to be the coordinate time), the velocity of light appears to slow down. A dramatic example is offered by black holes (to be discussed in Chapter 7). Because of infinite gravitational time dilation, it would take an infinite amount of coordinate time for a light signal to leave a black hole (thus, to the remote observer, no light can escape from a black hole), even though to a local observer, his proper time seems to flow normally.

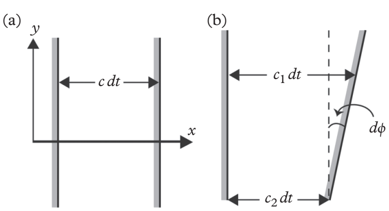

Figure 4.7 *Wavefronts of light trajectories: (a) Wavefronts of a straight-moving trajectory in the absence of gravity. (b) Tilting of wavefronts in a medium with an index of refraction varying in the vertical y direction so that $c_1 > c_2$. The bending of the resultant trajectory is signified by the small angular deflection $d\phi$.*

Bending of light ray calculated using Huygens' construction We can use this position-dependent index of refraction to calculate the bending of a light ray by a transverse gravitational field via the Huygens' construction. Consider a plane light wave propagating in the $+x$ direction. At each time interval Δt, the wavefront advances a distance $c\Delta t$; see Fig. 4.7(a). The existence of a transverse gravitational field (in the y direction) means a nonvanishing derivative of the gravitational potential, $d\Phi/dy \neq 0$. A change of the gravitational potential implies a change in $c(r)$, which leads to tilting of the wavefronts. We can calculate the angle of the bending of the light ray by using the diagram in Fig. 4.7(b):

$$d\phi \simeq \frac{(c_1 - c_2)\,dt}{dy} \simeq \frac{d[c(r)](dx/c)}{dy}. \tag{4.43}$$

Working in the limit of weak gravity with small $\Phi(r)/c^2$ (or equivalently $n \simeq 1$), we can relate $d[c(r)]$ to a change in the index of refraction via (4.42):

$$d\left[\frac{c(r)}{c}\right]_{\mathrm{EP}} = d\left[\frac{1}{n(r)}\right]_{\mathrm{EP}} = \frac{d\Phi(r)}{c^2}. \tag{4.44}$$

Thus, integrating (4.43) over the entire path, we obtain the total deflection angle:

$$[\delta\phi]_{\mathrm{EP}} = \int [d\phi]_{\mathrm{EP}} = \frac{1}{c^2}\int_{-\infty}^{\infty} \frac{\partial \Phi}{\partial y}\, dx = \frac{1}{c^2}\int_{-\infty}^{\infty} (\nabla\Phi \cdot \hat{\mathbf{y}})\, dx. \tag{4.45}$$

The integrand is the gravitational acceleration perpendicular to the light path. We shall apply this formula to the case of a spherical source with $\Phi = -G_{\mathrm{N}}M/r$ and $\nabla\Phi = G_{\mathrm{N}}M\hat{\mathbf{r}}/r^2$. Although the gravitational field will no longer be a simple uniform field in the $\hat{\mathbf{y}}$ direction, our approximate result can still be used, because the bending takes place mostly in the small region of $r \simeq r_{\min}$. (See Fig. 4.8.) We have

$$[\delta\phi]_{\mathrm{EP}} = \frac{G_{\mathrm{N}}M}{c^2}\int_{-\infty}^{\infty} \frac{\hat{\mathbf{r}}\cdot\hat{\mathbf{y}}}{r^2}\, dx = \frac{G_{\mathrm{N}}M}{c^2}\int_{-\infty}^{\infty} \frac{y}{r^3}\, dx, \tag{4.46}$$

where we have used $\hat{\mathbf{r}}\cdot\hat{\mathbf{y}} = \cos\theta = y/r$. An inspection of Fig. 4.8 also shows that, for small deflection, we can approximate $y \simeq r_{\min}$; hence

$$r = (x^2 + y^2)^{1/2} \simeq (x^2 + r_{\min}^2)^{1/2}, \tag{4.47}$$

leading to

$$[\delta\phi]_{\mathrm{EP}} = \frac{G_{\mathrm{N}}M}{c^2}\int_{-\infty}^{\infty} \frac{r_{\min}}{\left(x^2 + r_{\min}^2\right)^{3/2}}\, dx = \frac{2G_{\mathrm{N}}M}{c^2 r_{\min}}. \tag{4.48}$$

It should be noted again: this deflection result follows from an EP calculation which is based on gravitational time dilation alone. As we shall discuss in Section 6.4.2, this is half of the GR result $[\delta\varphi]_{\mathrm{GR}} = 2[\delta\varphi]_{\mathrm{EP}}$, because the curved spacetime description of GR implies in addition a gravitational length contraction.

4.3.4 Energetics of light transmission in a gravitational field

Because light gravitates (i.e., it bends and redshifts in a gravitational field), it is tempting to imagine that a photon has a (gravitational) mass. This may well lead to some erroneous conclusions regarding the energetics of light traveling in a gravitational field. Box 4.2 addresses this problem.

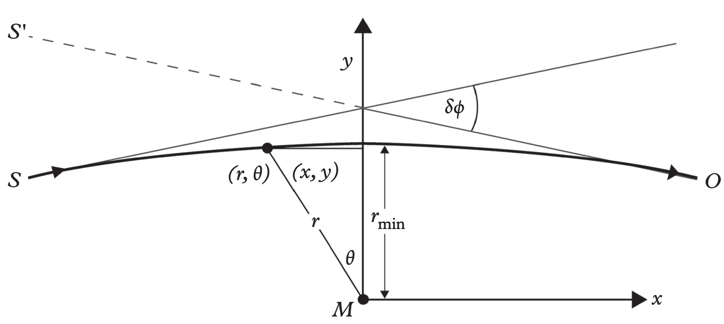

Figure 4.8 *Angle of deflection $\delta\phi$ of light by a mass M. A point on the light trajectory (solid curve) can be labeled either as (x, y) or as (r, θ). The source at S would appear to the observer at O to be located at a shifted position S'.*

Box 4.2 Energy considerations for gravitating light

Erroneous energy considerations

When considering the propagation of a light ray in a gravitational field, one might argue as follows: from the viewpoint of relativity, there is no fundamental difference between mass and energy, $E = m_I c^2$. The equivalence $m_I = m_G$ means that any energy also has a nonzero gravitational mass $m_G = E/c^2$, and hence will gravitate. The gravitational redshift formula (4.20) can be derived by regarding such a light pulse as losing kinetic energy when climbing out of a gravitational potential well.[9] Applying similar reasoning to the problem of gravitational light deflection, one can derive the result (4.48) by using the Newtonian mechanics formula[9a] for a moving mass with velocity u being gravitationally deflected by a spherically symmetric mass M (just as in Fig. 4.8),

$$[\delta\phi]_{\text{Newton}} = \frac{2G_N M}{u^2 r_{\min}}. \tag{4.49}$$

For the case of a photon with $u = c$, this would just reproduce the EP result shown in (4.48). Thus it so happens that the EP and Newtonian results coincide. As stated earlier, the predicted deflection is half of the correct GR prediction. Nevertheless, such an approach to understanding the effect of gravity on a light ray is conceptually incorrect, because

- A photon has no mass, so it cannot be described as a nonrelativistic massive object having a gravitational potential energy.
- This approach makes no connection to the underlying physics of gravitational time dilation.

The correct energy consideration

The energetics of gravitational redshift should be properly considered as follows. Light is emitted and received through atomic transitions between two atomic energy levels of a given atom:[10] $E_1 - E_2 = \hbar\omega$. We can treat the emitting and receiving atoms as nonrelativistic massive objects. Thus, when sitting at a higher gravitational potential, the receiver atom has more energy than the emitter atom:

$$E_{\text{rec}} = E_{\text{em}} + mgh. \tag{4.50}$$

We can replace the mass by E/c^2, so that, to leading order, $E_{\text{rec}} = (1 + gh/c^2)E_{\text{em}}$. There is a multiplicative energy shift of the atomic levels. This implies that all the energy levels (and their differences) of the receiving atom are blueshifted (increased energy, increased frequency) with respect to those of the emitter atom by

[9] For instance, the (Pound and Rebka 1960) paper has the title, "Apparent weight of photons."

[9a] Equation (4.49) is quoted in small-angle approximation of a general result that can be found in textbooks on mechanics. See, e.g., Eq. (4.37) in (Kibble 1985).

[10] We have used the fact that the energy of a light ray is proportional to its frequency. For most of us, the quantum relation $E = \hbar\omega$ comes immediately to mind, but this proportionality also holds in classical electromagnetism, where the field can be pictured as a collection of harmonic oscillators; see, e.g., (Cheng 2013, Section 3.1).

$$(E_1 - E_2)_{\text{rec}} = \left(1 + \frac{gh}{c^2}\right)(E_1 - E_2)_{\text{em}}; \tag{4.51}$$

hence there is a fractional shift of atomic energy

$$\left(\frac{\Delta E}{E}\right)_{\text{atom}} = \frac{gh}{c^2} = \frac{\Delta\Phi}{c^2}. \tag{4.52}$$

On the other hand, the traveling light pulse, neither gaining nor losing energy along its trajectory, has the same energy as the emitting atom. But it will be seen by the blueshifted receiver atom as redshifted:

$$\left(\frac{\Delta E}{E}\right)_{\gamma} = -\frac{\Delta\Phi}{c^2} = \frac{\Delta\omega}{\omega}, \tag{4.53}$$

which is the previously obtained result (4.20). This approach is conceptually correct, as

- Atoms can be treated as nonrelativistic objects having gravitational potential energy *mgh*.
- This derivation is entirely consistent with gravitational time dilation. The gravitational frequency shift does not result from any change in photon properties. It comes about because the standards of frequency (i.e., time) are different at different locations. In fact, this present approach gives us a physical picture of why clocks must run at different rates at different gravitational potentials. An atom is the most basic form of a clock, whose time rates are determined by transition frequencies. The fact that atoms have different gravitational potential energies (hence different energy levels) naturally give rise to different transitional frequencies, and hence different clock rates.

The various results called "Newtonian"

We should also clarify the often-encountered practice of calling results such as (4.49) Newtonian. By this it is meant that the result can be derived in the pre-Einsteinian-relativity framework, in which particles can take on any speed we wish them to have. Consequently, it is entirely correct to use the Newtonian mechanics formula for a light particle that happens to propagate with speed c. However, one should be aware of the difference between this Newtonian (pre-relativistic) framework and the proper Newtonian limit of relativistic physics (which we shall specify in Section 5.3.3) of nonrelativistic velocity and a static weak gravitational field. In this contemporary sense, (4.48) is not a result valid in the Newtonian limit (cf. Section 6.4.2).

Einstein's inference of a curved spacetime

Aside from the principle of relativity, the EP is the most important physical principle underlying Einstein's formulation of a geometric theory of gravity. Not only did it allow accelerating frames to be treated on equal footing with inertial frames, thus giving early glimpses of GR phenomenology, but also the study of the EP physics of time change led Einstein to propose that gravity represents the structure of a curved spacetime.[11] We shall explain this connection after learning in the following chapters some mathematics of curved spaces.

[11] In particular, that the gravitational equation of motion (4.9) is totally independent of any property of the test particle suggested to Einstein that the gravitational field, unlike other force fields, is related to some fundamental feature of spacetime.

Review questions

1. Write out, in terms of the gravitational potential $\Phi(x)$, the field equation and the equation of motion for Newton's theory of gravitation. What is the distinctive feature of this equation of motion (as opposed to that for other forces)?
2. What is inertial mass? What is gravitational mass? Give the simplest experimental evidence that their ratio is a universal constant (i.e., independent of the material composition of the object).
3. What is the equivalence principle? What is the weak EP? The strong EP?
4. Give a qualitative argument that the EP implies gravitational bending of a light ray.
5. Provide two derivations of the formula for gravitational frequency shift: $\Delta\omega/\omega = -\Delta\Phi/c^2$. (a) Use the idea that gravity can be transformed away by taking a reference frame in free fall. (b) Use the idea that atomic energy levels will be shifted in a gravitational field.
6. Derive the gravitational time dilation formula, $\Delta\tau/\tau = \Delta\Phi/c^2$, in two ways: (a) from the gravitational frequency shift formula; (b) from the consideration of three identically constructed clocks—two stationary at potential points Φ_1 and Φ_2, and the third in free fall passing by the first two.
7. GPS requires very precise time reading of clocks on the satellites that send electromagnetic signals to fix locations within a few meters on earth. What relativistic effects must be taken into account in order for this arrangement to work?
8. The presence of a gravitational field implies the presence of an effective index of refraction in free space. Does this mean that the speed of light is not absolute? What is the physical consequence of this index of refraction. When viewed from later developments, why was Einstein's 1911 calculation not complete?
9. Find the deviation from c when the light speed measured by an observer far from the gravitational source, when only gravitational time dilation is taken into account.