Special and General Relativity

Todd H. Rider thor@riderinstitute.org 13 November 2019

Any suggestions for improvements would be greatly appreciated.

There was a young lady named Bright,
Whose speed was far faster than light;
She set out one day
In a relative way,
And returned home the previous night.

–Arthur Henry Reginald Buller, in *Punch* (December 19, 1923)

...when the Rabbit actually *took a watch out of its waistcoat-pocket*, and looked at it, and then hurried on, Alice started to her feet... and was just in time to see it pop down a large rabbit-hole under the hedge.

In another moment down went Alice after it, never once considering how in the world she was to get out again...

Either the well was very deep, or she fell very slowly, for she had plenty of time as she went down to look about her, and to wonder what was going to happen next. First, she tried to look down and make out what she was coming to, but it was too dark to see anything...

–Lewis Carroll, *Alice's Adventures in Wonderland* (1865)

Overview

Special relativity describes the weird things that happen when you travel close to the speed of light, including time slowing down, lengths contracting, and not being able to accelerate past light speed no matter how hard you try. These weird effects are manifestations of the facts that space and time are interrelated, and mass and energy are interrelated.

General relativity describes the weird things that happen when you get close to very strong gravitational fields, including time slowing down, planetary orbits and even light being bent, and stranger things like black holes from which even light cannot escape. These effects are due to the fact that mass and energy are each capable of warping space and time. The degree and character of the warpage are determined by very nasty four-dimensional tensors (4x4 or even 4x4x4x4 matrices), but in most cases, the effects may be described and even calculated using a much more physically intuitively approach. Important problems which can be addressed either the easy way or the hard way include classic tests of general relativity (gravitational redshifting of light, gravitational bending of light, precession of Mercury's perihelion, and gravitational waves), the gravitational fields of stars, black holes, and wormholes, and models for the overall structure and history of the universe. Certain quantum mechanical effects in general relativity can also be addressed even without a full-fledged quantum theory of gravity.

Contents

1 Special Relativity

Special relativity describes the weird things that happen when you travel close to the speed of light, and general relativity describes the weird things that happen when you get close to very strong gravitational fields. A thorough understanding of special relativity, which describes the interrelationship between space and time, is a prerequisite for studying general relativity, in which both space and time are distorted by gravity. Therefore we will cover special relativity thoroughly before moving on to general relativity. Readers may choose the helpful hogwash version or the scary math version. And if you think the helpful hogwash version has scary math, just wait until you see the scary math version. For more information, see [1]-[2].

1.1 Helpful Hogwash Version

The sort of weird effects that occur due to special relativity are not the sort of thing you see in everyday life, so it can be somewhat difficult to wrap your brain around why they happen. To assist your brain wrapping, we are going to give three completely different, brief, and hopefully helpful hogwash explanations for special relativity. If you don't like one of them, try one of the other two.

Keep calm and turn the page...

Helpful Hogwash Explanation Number 1

Two foundational concepts lead to special relativity, as illustrated graphically in Fig. 1.

- There is no fixed stationary reference frame in the universe. You may have experienced being in a car with another car moving along side of it, or inside a train with another train moving along side of it, and momentarily being unsure if your vehicle was moving one way or the other vehicle was moving the other way. As illustrated in Fig. 1(a), if objects are moving with constant velocities relative to each other (not accelerating), one cannot say which one if any is really stationary. Velocity is relative. (In space, no one can hear you scream.)
- Maxwell's equations describe light and other electromagnetic waves moving at the speed of light $c \approx 2.9979 \times 10^8$ m/sec, without regard to the speed of any observer (*Electromagnetism* ?.?). Therefore, observers always see light traveling at light speed, no matter how fast the observers are going. Imagine a spaceship chasing a light wave through space [Fig. 1(b)]. If the ship is going at 90% of light speed, it will see the light moving at only 10% of light speed unless either (1) time slows down on the ship (making the light appear to move faster), or (2) distances contract on the ship (making the light appear to cover more ground). In fact, both effects occur.

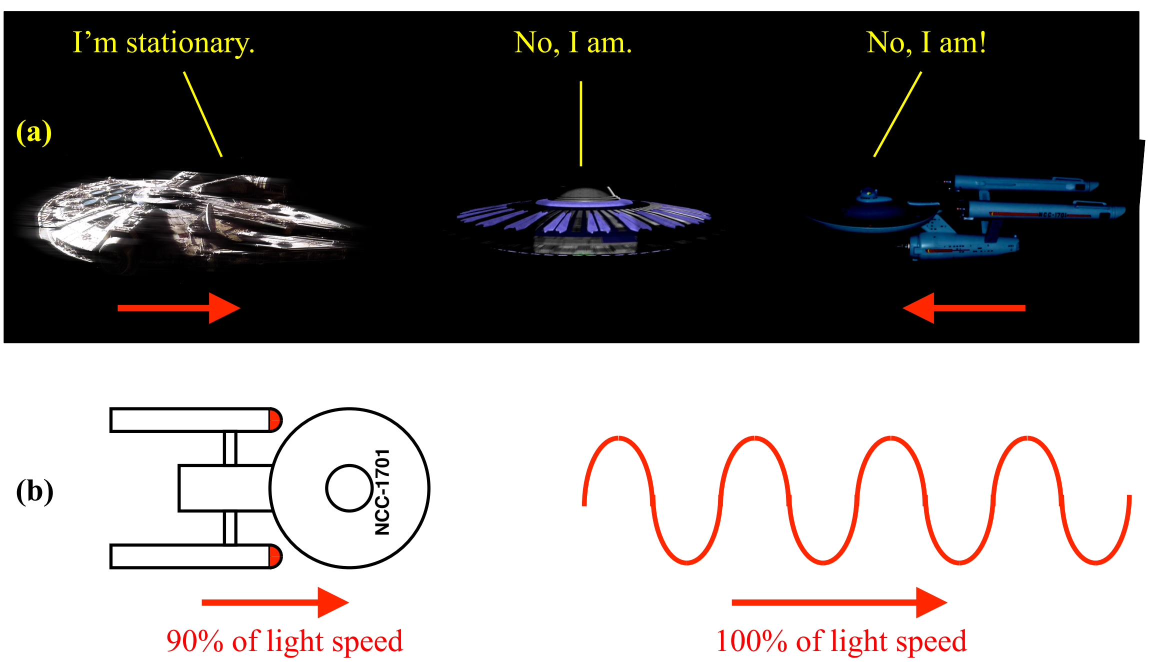

Figure 1. The two foundational concepts for special relativity. (a) There is no fixed stationary reference frame. If objects are moving with constant velocities relative to each other (not accelerating), one cannot say which one if any is really stationary. **(b)** An observer always sees light traveling at light speed, no matter how fast the observer is going.

Helpful Hogwash Explanation Number 2

Alternatively, consider the view of space and time illustrated in Fig. 2. Distance is a real number and time is an imaginary number (meaning it uses $i = \sqrt{-1}$). The horizontal axis represents position in space. The vertical axis is the time on Earth, assumed to be a "stationary" reference frame to simplify the example; the time is multiplied by the speed of light c so that both axes have units of length. Light travels on a diagonal line where position/time = c.

If a spaceship travels away from Earth at a constant velocity v, an observer on Earth would say that the ship had traveled a distance $v\,t_{\text{Earth}}$ after an elapsed Earth time t_{Earth}. Since the ship is moving at a constant velocity, observers on the ship would feel that they are in a "stationary" reference frame and are moving straight forward through time, so the elapsed ship time is given by the hypotenuse of the red triangle, the length along the ship's trajectory. The ship measures its time axis in units of ict, just as the Earth does.

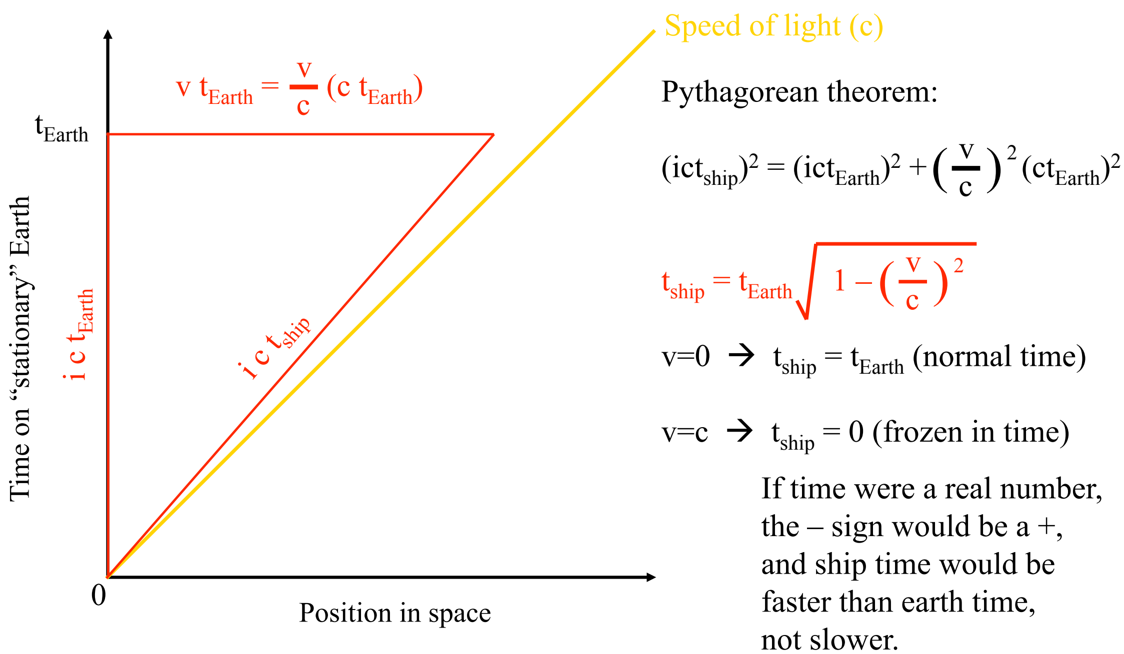

Figure 2. The interrelationship between space and time in special relativity. The main axes show the perspective of an observer on "stationary" Earth watching a spaceship move away at velocity v. If $v = 0$, the ship moves vertically through time at velocity c without traveling in space. If $v = c$, the ship travels at the maximum speed (c) through space but does not age (the elapsed ship time is 0).

Using the Pythagorean theorem to relate the lengths of the three sides yields an equation for converting between Earth time and ship time:

$$t_{\text{ship}} = t_{\text{Earth}} \sqrt{1 - \left(\frac{v}{c}\right)^2} = \frac{t_{\text{Earth}}}{\gamma}, \quad \text{where} \tag{1}$$

$$\gamma \equiv \frac{1}{\sqrt{1-(v/c)^2}} = \frac{1}{\sqrt{1-\beta^2}}. \tag{2}$$

$$\beta \equiv \frac{v}{c} \tag{3}$$

The - sign in the square root of Eqs. (1) and (2) arises because time is imaginary, and it means that ship time is slower than Earth time. This strange mixture of real spatial axes and an imaginary time axis is called **Minkowski** space-time. If time were a real number, the - sign inside the square root sign would be a +, and ship time would be faster than Earth time. That all-real coordinate system is called **Euclidean** space-time, but it does not describe our universe.

Time onboard the spaceship moves more slowly than time on the earth by a factor of γ. If there were people on the ship, they would move, think, and age more slowly by this factor. Figure 3 graphs values of γ for velocities ranging from $v = 0$ to $v = c$.

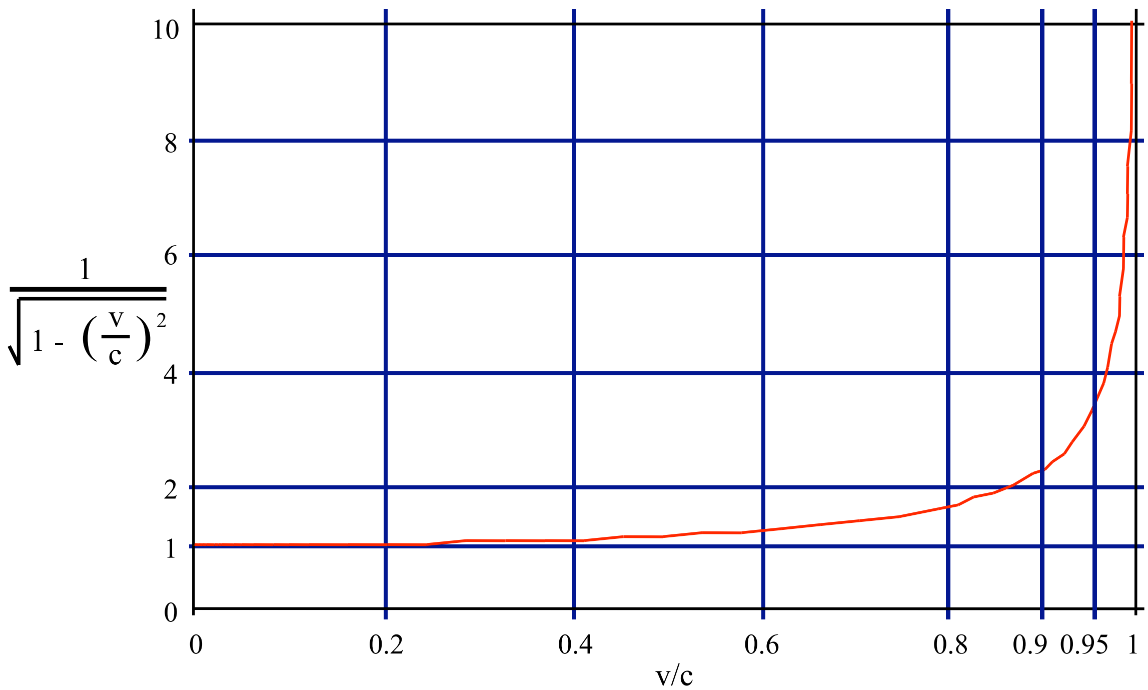

Figure 3. Graph of $\gamma = [1 - (v/c)^2]^{-1/2}$ multiplier versus velocity. This is the factor by which time slows down, length in the direction of travel contracts, and the effective mass increases for an object traveling at velocity v relative to a "stationary" observer.

For these same reasons, lengths in the direction of travel also appear to be shorter by this same factor γ, as viewed by observers on the ship:

$$\Delta x_{\text{ship}} \quad = \quad \Delta x_{\text{Earth}} \sqrt{1 - \left(\frac{v}{c}\right)^2} \quad = \quad \frac{\Delta x_{\text{Earth}}}{\gamma} \tag{4}$$

Thus for observers on the ship, time slows down and distances contract, making light still appear to be moving at light speed, as predicted in the second bullet point on p. 3.

Note from Figs. 2 and 3 that for $v = 0$, all of these weird effects disappear, and the spaceship experiences space and time the same way that observers on earth do. At the opposite extreme, for $v = c$, γ becomes infinite; the ship hurtles through space at light speed relative to the earth, but any astronauts onboard are frozen in time and do not age. In practice, only light and other massless particles can reach this speed. Spaceships and other objects with mass can only approach light speed arbitarily closely before γ and the equations of motion blow up.

Helpful Hogwash Explanation Number 3

If Figs. 2 and 3 seems too scary, another way to view the time effects is from the perspective of time onboard the spaceship, regardless of how fast the ship is going. Put this way, you are always moving at light speed, either through time or space. The vertical axis in Fig. 4 represents your speed through time, while the horizontal axis represents your speed through space. If you aren't moving through space, you are moving through time at the speed of light (getting older at 1 second per second). As your speed through space increases, your speed through time decreases (you age at less than 1 second per second). If you go at the speed of light through space, you stop moving through time entirely (you appear frozen in time). You can't get past the frozen-in-time point to start going backward in time. This description is equivalent to that in Fig. 2, just remaining in the point of view of the ship regardless of its velocity relative to the external reference frame.

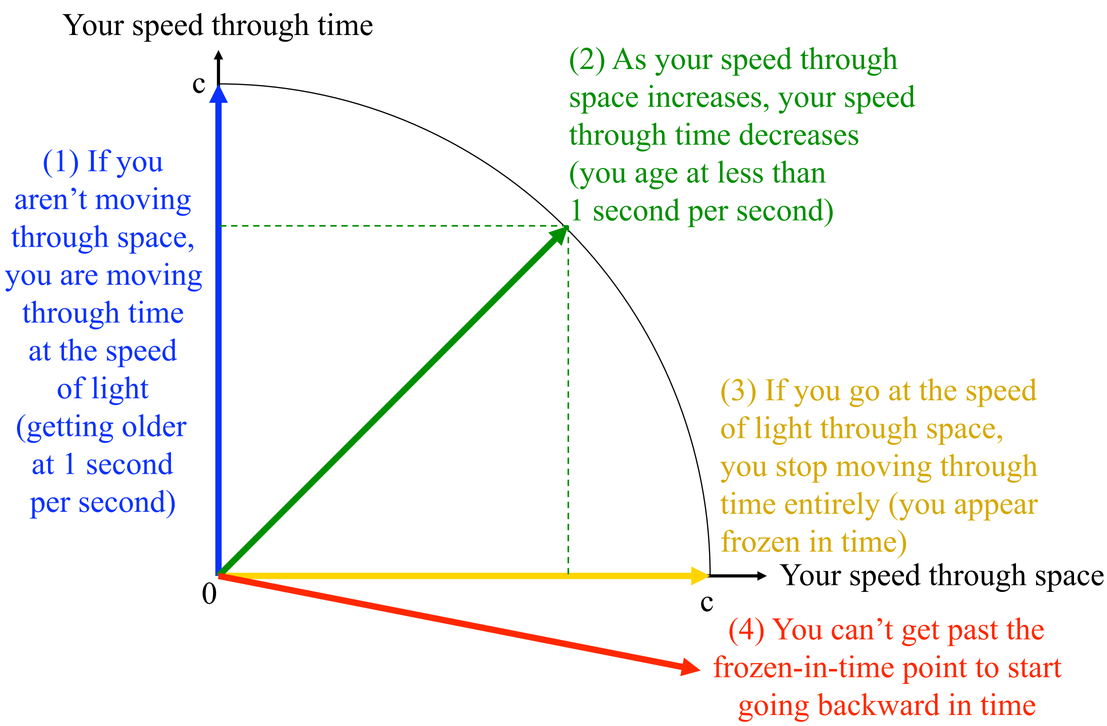

Figure 4. Effects of special relativity from the point of view of a moving observer. You are always moving at light speed, either through time or space. The vertical axis represents your speed through time, while the horizontal axis represents your speed through space. (1) If you aren't moving through space, you are moving through time at the speed of light (getting older at 1 second per second). (2) As your speed through space increases, your speed through time decreases (you age at less than 1 second per second). (3) If you go at the speed of light through space, you stop moving through time entirely (you appear frozen in time). (4) You can't get past the frozen-in-time point to start going backward in time.

Post-Hogwash Stress Syndrome

If any of those three hogwash explanations made sense to you, we can move along. If they didn't, you can write your own book on relativity.

Since objects must travel more slowly than light, their trajectories are constrained to remain within a **light cone** in space and time, as shown in Fig. 5. The vertical direction represents time traveling upward from the past to the future. Two spatial dimensions are shown as the horizontal axis and the axis perpendicular to the page. For simplicity of illustration, the third spatial dimension is not shown. In this coordinate system, if the current position and time of an object are known, the object must have been within the lower cone at all times in the past, and it must remain within the upper cone at all times in the future. The slope of the cone (dividing space by time) is the speed of light c. The object cannot visit places and times outside the cone without traveling faster than light, which is generally frowned upon.

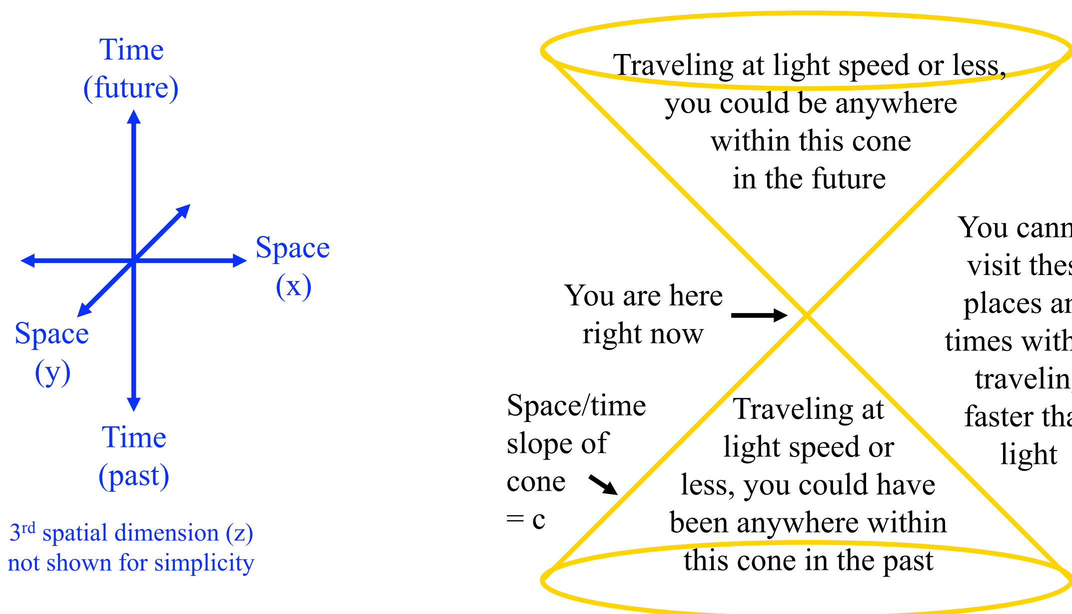

Figure 5. Light cone. The vertical direction represents time traveling upward from the past to the future. Two of the three spatial dimensions are shown as the horizontal axis and the axis perpendicular to the page. If the current position and time of an object are known, the object must have been within the lower cone at all times in the past, and it must remain within the upper cone at all times in the future. The slope of the cone (dividing space by time) is the speed of light c.

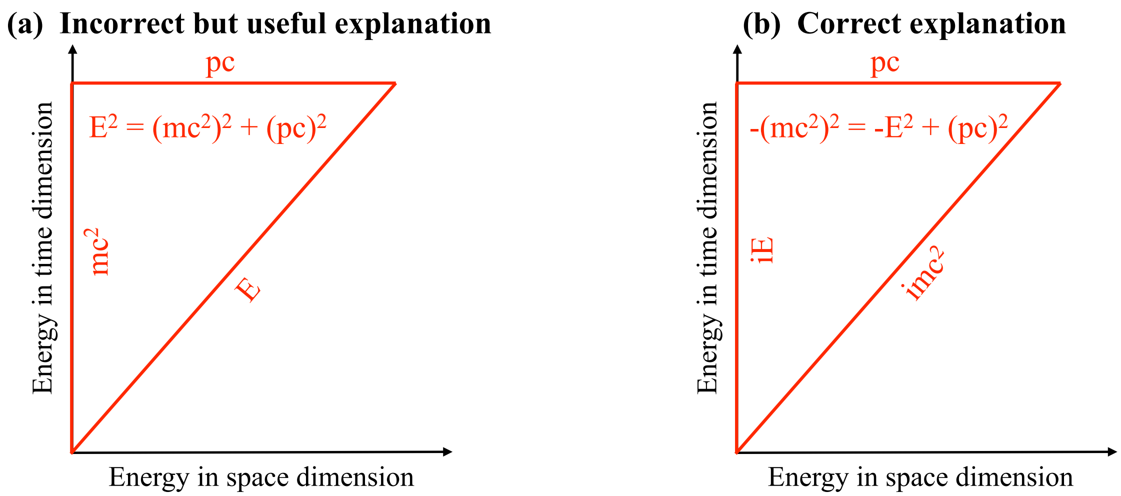

Figure 6. The relationship between mass, momentum, and energy in special relativity. **(a)** An incorrect but mnemonically useful explanation. **(b)** The correct explanation.

Figure 6 shows how mass, energy, and momentum can be viewed in special relativity. An incorrect but mnemonically useful explanation [Fig 6(a)] is that a moving object has two orthogonal types of energy, pure rest energy mc^2 due to the object's motion through time, and pure kinetic energy pc due to the object's momentum p through space. Matter that is not moving only has rest energy mc^2, whereas light only has kinetic energy pc. If the total energy E is the vector sum of these two components, doing the Pythagorean theorem thing yields the correct relativistic equation for E:

$$E = \sqrt{(mc^2)^2 + (pc)^2} = \gamma mc^2 \tag{5}$$

$$\approx mc^2 + \frac{1}{2}mv^2 + \text{smaller terms for } v \ll c \tag{6}$$

Likewise the momentum is

$$p = \gamma mv \tag{7}$$

$$\approx mv \text{ for } v \ll c \tag{8}$$

Equations (5) and (7) show that these effects are equivalent to increasing the rest mass by a factor of γ from Fig. 3; as an object approaches the speed of light, its effective mass becomes so large that one can never exert enough force to accelerate the object all the way to light speed.

For $v/c \ll 1$, the momentum p in Eq. (7) reduces to the usual nonrelativistic expression $p = mv$ as shown in Eq. (8), and E in Eq. (5) reduces to a simple sum of the rest energy mc^2 and the nonrelativistic kinetic energy $mv^2/2$, as shown in Eq. (6).

The correct explanation [Fig. 6(b)] is that the total energy E is what extends into the time dimension, and it is imaginary just as time coordinates are imaginary in Fig. 2. The kinetic energy due to an object's momentum through space is indeed pc. An observer moving through both space and time at a constant velocity could consider themselves at rest, with only a rest energy imc^2 (imaginary since that is the time dimension for the moving observer). Using the Pythagorean theorem with these side lengths again gives the correct answer for the energy.

Lest you think we just made up all this weirdness, particle accelerators such as the Large Hadron Collider (LHC) shown in Fig. 7 routinely accelerate electrons, protons, and other particles to nearly the speed of light and confirm that their time slows down, that their effective mass increases, that they can never reach the speed of light, and so forth.

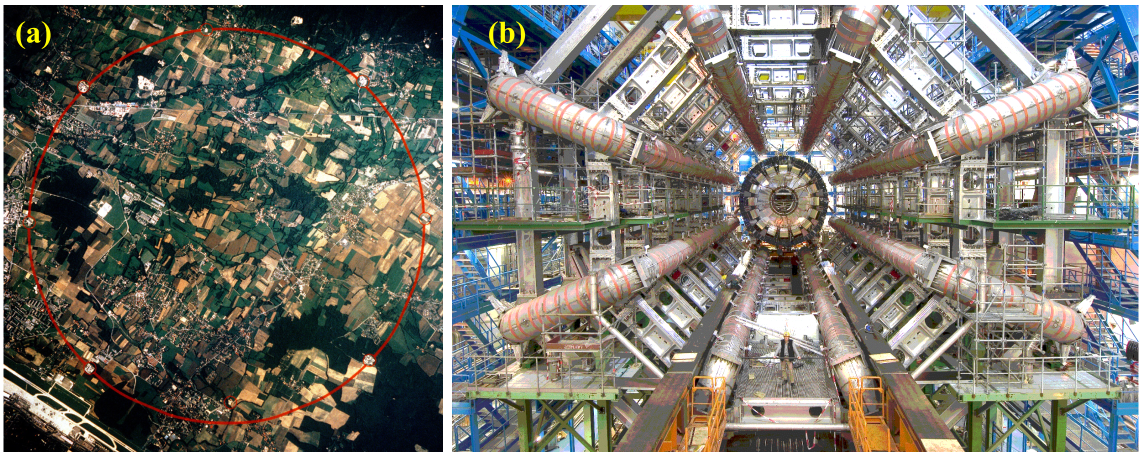

Figure 7. Particle accelerators are the ultimate demonstration of special relativity. (a) An overhead view of the CERN Large Hadron Collider (LHC) in Switzerland. The underground accelerator follows the red circle and is 17 miles long. **(b)** An inside view of the LHC. Note the person standing near the bottom for scale. The LHC imparts energies up to 7 trillion electron-volts per particle. Particle accelerators like this can push electrons, protons, and other particles up to 99.999999999% of light speed, but never faster than light. At such high velocities, time slows down as expected, as demonstrated by the fact that particle decays take much longer than they would for particles at rest. Similarly, the observed mass increases and length decreases correspond with the predictions of special relativity.

1.2 Scary Math Version

If the explanations of Section 1.1 made too much sense to you, fear not, we can cover the same ground using much more confusing math. Specifically, special relativity may be written much more formally using matrices called tensors. This tensor notation is an especially important tool for general relativity, so you might as well learn it now for special relativity.

The position vector of something in four-dimensional space-time is

$$x^\mu = (x^0, x^1, x^2, x^3) \tag{9}$$

$$= (ct, x, y, z) \qquad \text{in rectangular coordinates} \tag{10}$$

$$= (ct, r, \theta, \phi) \qquad \text{in spherical coordinates} \tag{11}$$

The superscripts in Eq. (9) denote the four coordinates of space and time and should not be confused with exponents. As in the hogwash explanation for special relativity, time is multiplied by c, $x^0 \equiv ct$, so that the time dimension can be measured in units of length just like the three spatial dimensions. Unlike in the hogwash version, however, time is not multiplied by i; that same function will be performed by the scary tensors in just a moment.

In converting from rectangular to spherical coordinates as in Eqs. (10)-(11), θ is measured from the z-axis and ϕ is measured in the x-y plane from the x-axis, following the conventions of *Applied Mathematics* ?.?.

Greek indices like μ and ν denote all four dimensions (time and space, dimensions 0 through 3), whereas Latin indices like j and k denote spatial dimensions only (dimensions 1 through 3).

The **Einstein convention for implicit summation** saves cluttering up equations with a lot of $\sum$ summation signs: repetition of the same index label on the same side of an equation denotes summation over all values of that index. Use of implicit summation will be illustrated very shortly.

$\eta_{\mu\nu}$ is the reference metric of space-time that is not warped by gravitational fields. Unwarped space-time is also called flat or Minkowski space-time and is what we are used to in special relativity. Depending on whether coordinates are rectangular or spherical, the metric is:

$$\eta_{\mu\nu} \equiv \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}_{\text{rectangular}} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta \end{pmatrix}_{\text{spherical}} \tag{12}$$

Hey, we warned you this would be scary.

The distance between two infinitesimally close points in space-time is defined to be ds. In flat space-time, it is determined by $\eta_{\mu\nu}$:

$$ds^2 = \eta_{\mu\nu}\, dx^\mu\, dx^\nu \tag{13}$$

$$= \sum_\mu \sum_\nu \eta_{\mu\nu}\, dx^\mu\, dx^\nu \tag{14}$$

$$= -(c\,dt)^2 + dx^2 + dy^2 + dz^2 \tag{15}$$

Equation (13) uses the Einstein implicit summation convention, as shown more explicitly in Eq. (14). Be warned that implicit summation will be used with wild abandon from now on.

The -1 in the upper left (time-time) position of the metric serves to introduce a - sign in the first term of Eq. (15) Thus the metric tensor serves the same role here as defining time to be imaginary did in Fig. 2, and it leads to the same result. The matrix inverse of $\eta_{\mu\nu}$ is $\eta^{\mu\nu}$. The corresponding individual elements of these two matrices are actually the same.

In special relativity (or in general relativity when space-time is essentially flat), one raises and lowers indices using $\eta_{\mu\nu}$ and $\eta^{\mu\nu}$. For example, $x_\mu = \eta_{\mu\nu} x^\nu$ and $x^\mu = \eta^{\mu\nu} x_\nu$. Using this notation, Eq. (13) can be rewritten in an even more compact form:

$$ds^2 = dx_\mu\, dx^\mu \tag{16}$$

Likewise, designating E/c as the zeroth or time component p^0 of a momentum four-vector $p^\mu = (E/c, p^1, p^2, p^3)$, the relation between energy, momentum, and mass can be written as

$$-(mc)^2 = p_\mu\, p^\nu \tag{17}$$

$$= -\left(\frac{E}{c}\right)^2 + |\mathbf{p}|^2 \tag{18}$$

in agreement with Eq. (5).

Relative to some reference frame, a motionless spaceship moves straight through time, but a moving ship moves at an angle ϕ through time and space. Thus when the ship changes from motionless to moving at this velocity, it is equivalent to rotating in space-time by an angle ϕ. This is called a **Lorentz transformation** or **Lorentz rotation** and is illustrated in Fig. 8(a). As shown in Fig. 8(b), boosting a ship's velocity by an angle ϕ_1 in space-time, and then boosting it again by an angle ϕ_2, increases the ship's total velocity, but the total velocity can never reach the speed of light.

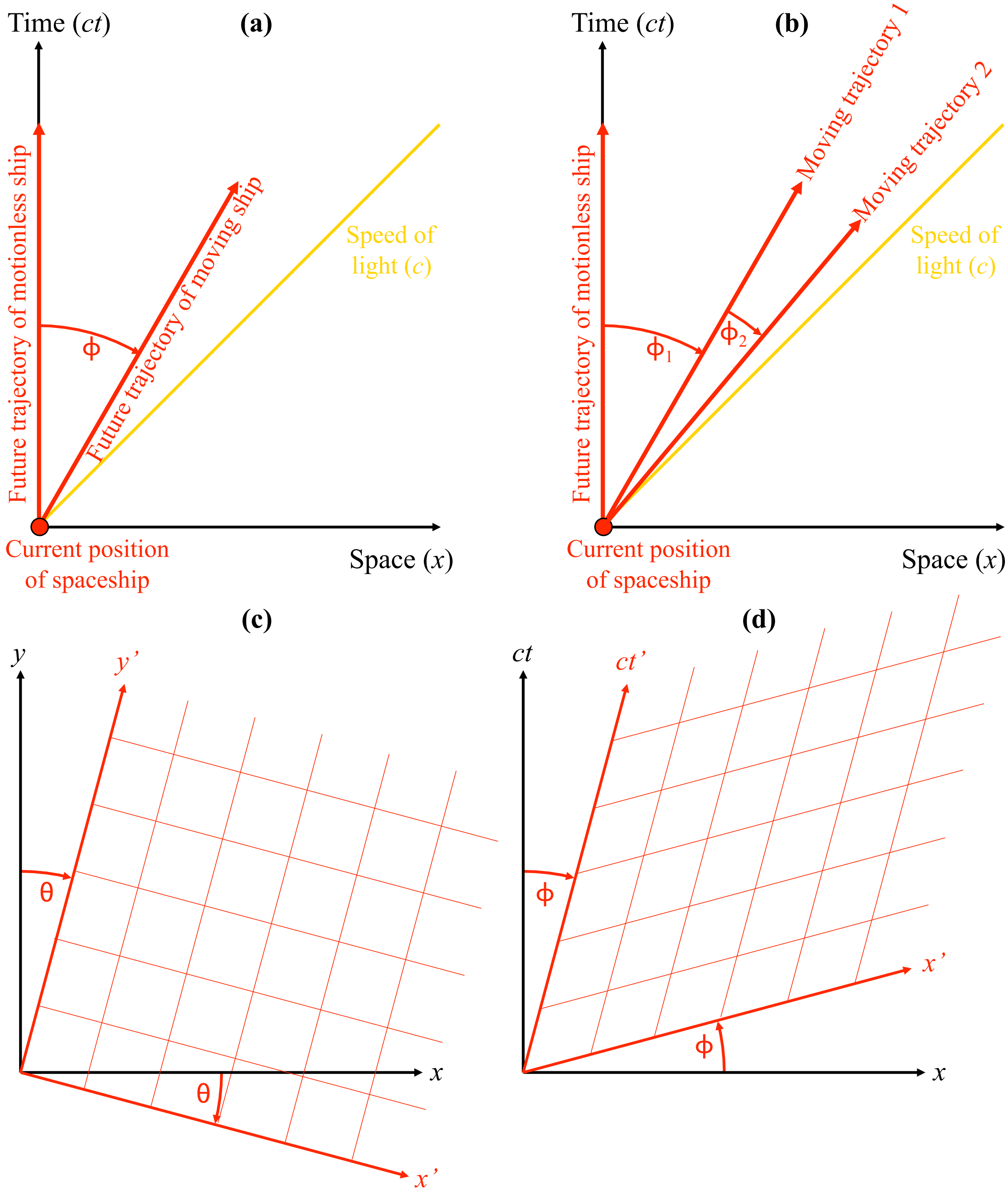

Figure 8. Lorentz rotation. (a) Relative to some reference frame, a motionless spaceship moves straight through time, but a moving ship travels at an angle ϕ through time and space. Thus when the ship changes from motionless to moving at this velocity, it is equivalent to rotating in space-time by an angle ϕ. **(b)** Boosting a ship's velocity by an angle ϕ_1 in space-time, and then boosting it again by an angle ϕ_2, increases the ship's total velocity, but the total velocity can never reach the speed of light. **(c)** A spatial rotation by angle θ changes the coordinate axes x and y to new axes x' and y'. **(d)** A space-time rotation by angle ϕ changes the axes from ct and x to ct' and x'. Since time coordinates are imaginary numbers and spatial coordinates are real numbers, the new axes appear distorted to a "motionless" observer using ct and x coordinates, but ct' and x' appear orthogonal to each other and perfectly normal to a moving observer using those new coordinates.

As illustrated in Fig. 8(c), a spatial rotation by angle θ changes the coordinate axes x and y to new axes x' and y':

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\theta \; - \; y\sin\theta \\ x\sin\theta \; + \; y\cos\theta \end{pmatrix} \tag{19}$$

As depicted in Fig. 8(d), a space-time rotation by angle ϕ changes the axes from ct and x to ct' and x'. The transformations from Eqs. (1)-(4) can be formalized in terms of tensors and then expressed in terms analogous to Eq. (19) for purely spatial rotation:

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ct\gamma \; - \; x\gamma\beta \\ -ct\gamma\beta \; + \; x\gamma \\ y \\ z \end{pmatrix} \tag{20}$$

$$= \begin{pmatrix} \cosh\phi & -\sinh\phi & 0 & 0 \\ -\sinh\phi & \cosh\phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ct\cosh\phi \; - \; x\sinh\phi \\ -ct\sinh\phi \; + \; x\cosh\phi \\ y \\ z \end{pmatrix} \tag{21}$$

Since time coordinates are imaginary numbers and spatial coordinates are real numbers, the angle ϕ between them is actually an imaginary number. Writing ϕ as a real number causes Eq. (21) to be written in terms of cosh and sinh instead of cos and sin as in Eq. (19). This imaginary business also makes the new axes in Fig. 8(d) appear distorted to a "motionless" observer using ct and x coordinates, but ct' and x' would appear orthogonal to each other and perfectly normal to a moving observer using those new coordinates.

In this nifty notation, velocity is simply $v/c = \tanh\phi$. Adding two velocities $v_1/c = \tanh\phi_1$ and $v_2/c = \tanh\phi_2$ yields a total velocity v_{total}/c:

$$\frac{v_{\text{total}}}{c} = \tanh(\phi_1 + \phi_2) = \frac{\tanh\phi_1 + \tanh\phi_2}{1 + \tanh\phi_1 \tanh\phi_2} = \frac{(v_1/c) + (v_2/c)}{1 + (v_1 v_2/c^2)} \tag{22}$$

Thus two space-time angles ϕ_1 and ϕ_2 can just be added together, even though their corresponding velocities cannot simply be added together ($v_1 + v_2 \neq v_{\text{total}}$), since the total velocity can never reach the speed of light. [Note that in the nonrelativistic limit of v_1 and v_2 much less than the speed of light, Eq. (22) does reduce to the simple addition of velocities.]

2 Principles of General Relativity–Helpful Hogwash Version

This section will explain the key ideas of general relativity in a very physically intuitive fashion, as opposed to the highly abstract mathematical presentations that are generally given in all the textbooks [3]-[11] and summarized in Section 3. This section will explain the starting point for general relativity, then discuss how gravitational fields affect time and space.

2.1 Starting Point for General Relativity

The starting point for general relativity is that gravitational attraction is indistinguishable from acceleration. As shown in Fig. 9(a), someone standing in an elevator cannot tell if it is resting in the earth's gravitational field versus way out in space but accelerating. Similarly, someone floating in an elevator cannot tell if it is falling in the earth's gravitational field versus floating way out in space, as illustrated in Fig. 9(b). This is the **equivalence principle**: a region with a uniform gravitational field (e.g. the gravitational acceleration at the earth's surface, $g \approx 9.807\ \mathrm{m/sec^2}$) and a uniformly accelerating reference frame are equivalent.

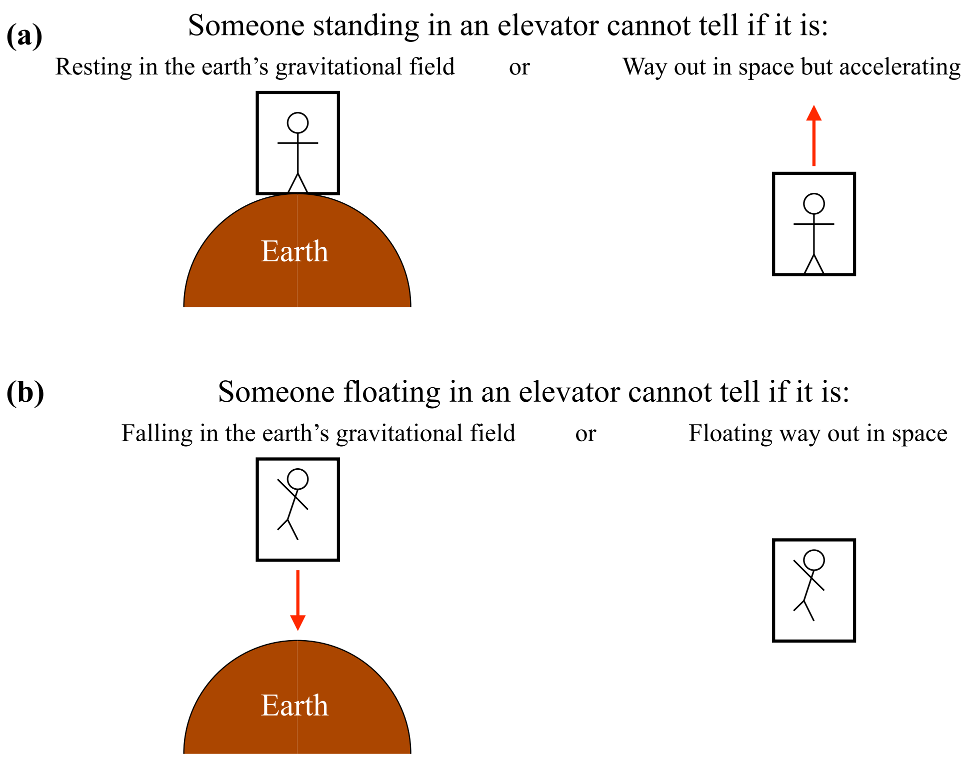

Figure 9. Starting point for general relativity. (a) Someone standing in an elevator cannot tell if it is resting in the earth's gravity or way out in space but accelerating. **(b)** Similarly, someone floating in an elevator cannot tell if it is falling in the earth's gravity or floating way out in space.

Special relativity was called special because it only dealt with a special case: different frames of reference that were moving with constant velocities relative to each other. General relativity is called general because it can be used to analyze frames of reference that do not have constant velocities, i.e. frames of reference that are accelerating or decelerating. The notion that gravitational attraction and acceleration are equivalent means that general relativity describes gravitational fields. Just as traveling near light speed can "shorten" both time and space, as described in Section 1, strong gravitational fields also shorten time and space, as will be explained in the next two subsections.

2.2 Gravity Shortens Time

As shown in Fig. 10, electromagnetic waves gain energy (are blue-shifted to shorter wavelengths and higher frequencies) falling into a gravitational field and lose energy (are red-shifted to longer wavelengths and lower frequencies) escaping from a gravitational field.

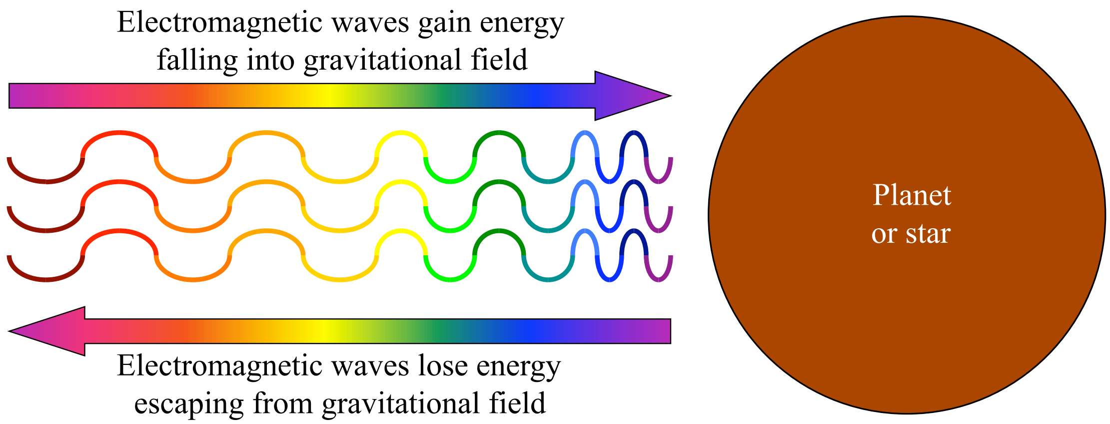

Figure 10. Gravity shortens time. Electromagnetic waves are blue-shifted to higher energies and frequencies when they fall into a gravitational field, and they are red-shifted to lower energies and frequencies when they escape from a gravitational field. Light has no mass so classically its energy should not be affected by a gravitational field, yet it is. The explanation is that time is actually moving more slowly inside a gravitational field than outside, making frequencies from inside the field look lower to an observer outside the field, and frequencies from outside the field look higher to an observer inside the field. The effect is not merely an artifact of signal transmission between the inside and outside of the gravitational field.

The blue shifting of light falling into a gravitational field corresponds exactly to what would be predicted by considering the energy gained by an object falling into a classical (Newtonian) gravitational well. Consider a motionless object of mass m at an observer's position far from the well; the object has rest energy mc^2 and no kinetic energy, making its total rest + kinetic energy just $E = mc^2$. If this object were to "fall" down to a distance r from the center of the gravitational well, the gravitational field would give it a kinetic energy equal to GMm/r, where $G \approx 6.67 \cdot 10^{-11}\mathrm{m}^3\mathrm{kg}^{-1}\mathrm{sec}^{-2}$ is Newton's gravitational constant and M is the mass creating the gravitational field. The total rest + kinetic energy of the object would then be

$$E' \;=\; mc^2 + \frac{GMm}{r} \;=\; mc^2\left(1 + \frac{GM}{c^2 r}\right) . \tag{23}$$

Therefore the change in energy relative to the energy outside of the gravitational well is:

$$\frac{\Delta E}{E} \;=\; \frac{GM}{c^2 r} . \tag{24}$$

Because the energy E and frequency f of light waves are directly proportional to each other ($E = hf$, in which h is Planck's constant), this simple argument predicts a frequency shift of

$$\frac{\Delta f}{f} = \frac{GM}{c^2 r} . \qquad (25)$$

for light waves falling into the gravitational field.

The time period of wave oscillations is inversely proportional to the frequency of the oscillations, so the time of the wave oscillations decreases by the same fraction,

$$\frac{\Delta t}{t} = \frac{GM}{c^2 r} . \qquad (26)$$

Light has no mass so classically its energy should not be affected by a gravitational field, yet it is. The explanation is that relative to the time outside a gravitational field, time inside a field moves more slowly by the fraction given in Eq. (26), making frequencies from inside the field look lower to an observer outside the field, and frequencies from outside the field look higher to an observer inside the field. The effect is not merely an artifact of signal transmission between the inside and outside of the gravitational field. Someone inside a gravitational field would age more slowly than someone outside, just as someone traveling near light speed would age more slowly than the frame of reference of a "stationary" observer in special relativity. Thus gravity "shortens" time by the fraction in Eq. (26).

(As an aside, it is good that the energy shift applies to photons even though they have no mass. If photons were unaffected by gravity, stuff in the form of a mass could be dropped down a gravitational well, then at least in theory completely converted to photons and transmitted back out of the well. As a mass falling on the way down the stuff would gain energy, but it would not lose this energy on the way back up when it is in the form of photons. One could thus generate energy from nothing. This unphysical outcome is prevented by the fact that photons climbing out of gravitational wells actually lose energy.)

Because gravitational fields cause time to run more slowly than it does in gravity-free regions, they can also "tilt" the forward direction of time toward a particular direction in space. In Fig. 11, the vertical axis is the elapsed time and the horizontal axis is the position in space. If a motionless (in this frame of reference) massive star or planet with a strong gravitational field is placed in the middle of the spatial axis, the star or planet will move forward through time, traveling vertically from the past to the future.

Each blue line in Fig. 11 represents a "simultaneous" moment throughout space. The gravitational field of the star or planet slows down time in its vicinity, so the blue lines near the massive object do not travel into the future as quickly.

A "motionless" test object located nearby goes forward through time by moving perpendicular to each simultaneous moment as shown in Fig. 11. Due to time moving more slowly inside a gravitational field, the object's path through time (the red arrow) veers toward the center of the gravitational field. Time for the test object becomes "tilted" toward the star or planet. And tilting in space-time is just a Lorentz rotation (Fig. 8), or a change of velocity–an acceleration. Since the trajectory doesn't just tilt but continues to curve, the test object's velocity will continue to change, as it accelerates toward the star or planet and (if it has enough time before impact) asymptotically approaches the 45^o angle that describes the speed of light on such a space-time graph.

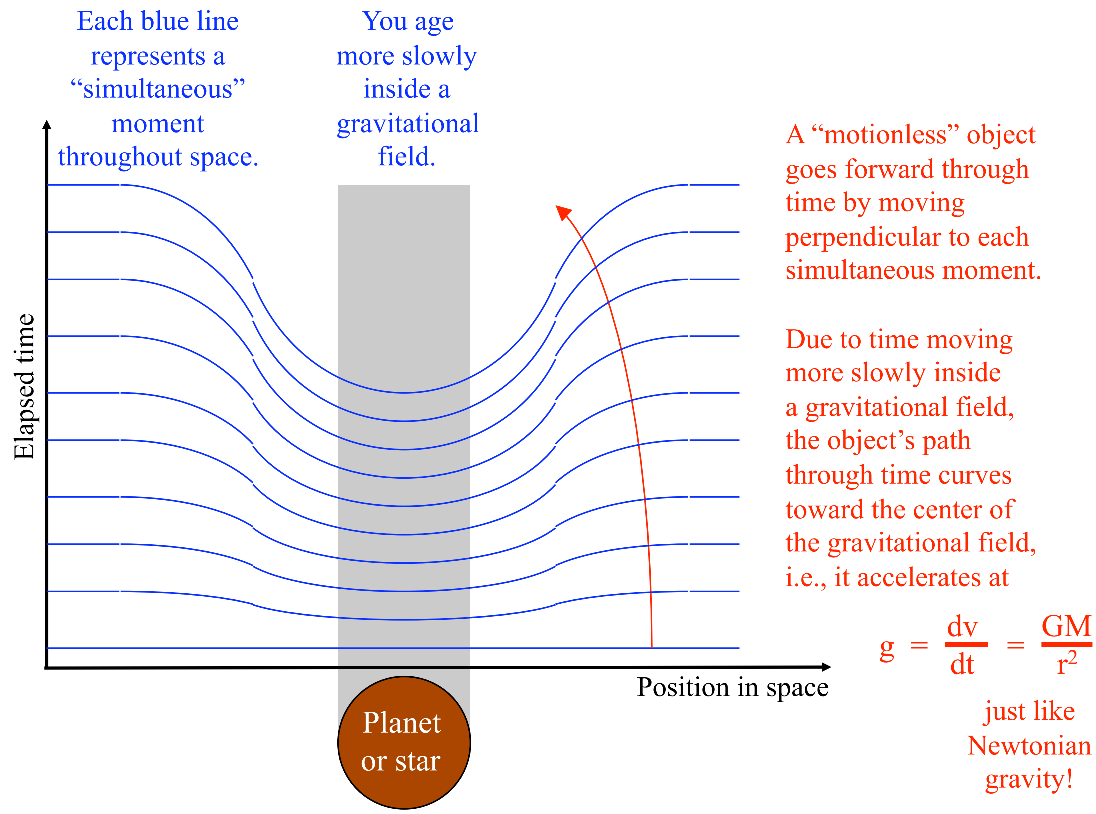

Figure 11. Gravity shortens time (continued). The vertical axis is the elapsed time and the horizontal axis is the position in space. A planet or star with a strong gravitational field is at rest in this frame of reference and travels vertically from the past to the future. Each blue line represents a "simultaneous" moment throughout space. You age more slowly inside a gravitational field, so the blue lines near the planet or star do not travel into the future as quickly. A "motionless" test object goes forward through time by moving perpendicular to each simultaneous moment. Due to time moving more slowly inside a gravitational field, the object's path through time (the red arrow) curves toward the center of the gravitational field, i.e., it accelerates at $g = dv/dt = GM/r^2$.

This behavior is just what we would expect from Newtonian physics without all of the fuss over general relativity–the test object falls toward the star and continues to accelerate as it falls. Specifically, the incremental change in a test object's velocity $d(v/c)$ is directly related to the radial variation in the flow of time, $(c\,dt)\,\frac{d}{dr}[1 - GM/(c^2 r)]$:

$$d\left(\frac{v}{c}\right) \quad = \quad (c\,dt)\ \frac{d}{dr}\left(1 - \frac{GM}{c^2 r}\right) \quad = \quad (c\,dt)\ \frac{GM}{c^2 r^2} \tag{27}$$

Using Eq. (27), the acceleration of the test object toward the star or planet in Fig. 11 is

$$g \quad = \quad \frac{dv}{dt} \quad = \quad \frac{GM}{r^2} \tag{28}$$

just like Newtonian gravity! This is the basis of the equivalence principle between gravity and acceleration as described in Section 2.1.

2.3 Gravity Shortens Space

As shown in Fig. 12(a), if you cut a wedge of angle $\Delta\phi$ from a flat (two-dimensional) paper circle and then tape the cut edges together, the circle becomes a cone that projects into an extra (third) dimension. Similarly, a gravitational field effectively removes an angle $\Delta\phi$ from the usual 2π radians of a circle enclosing the field [Fig. 12(b)]. The circumference of circular orbits around massive objects will be disproportionately small in comparison with the radius of the orbit.

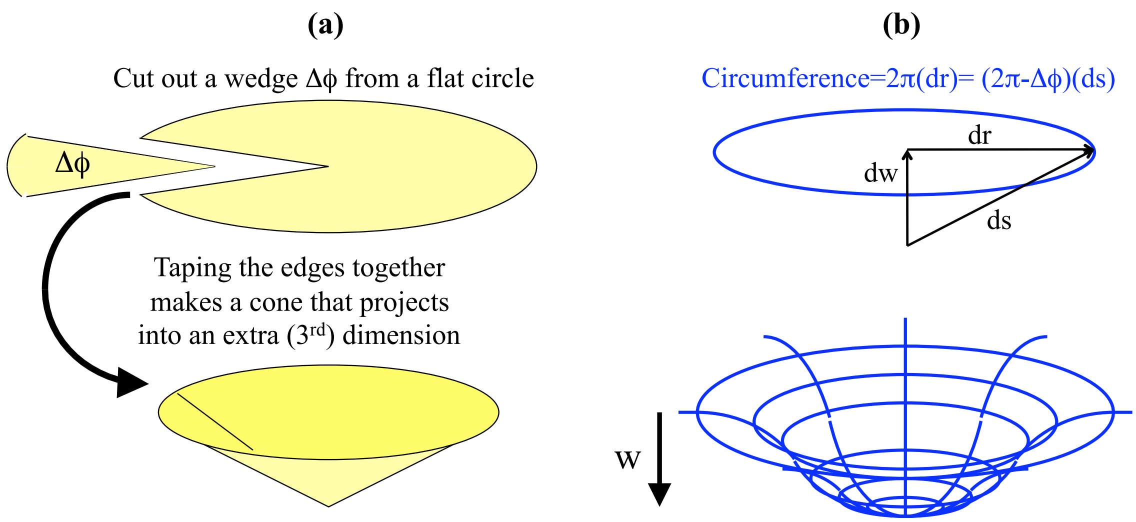

Figure 12. Gravity shortens space. (a) If you cut a wedge of angle $\Delta\phi$ from a flat (two-dimensional) paper circle and then tape the cut edges together, the circle becomes a cone that projects into an extra (third) dimension. **(b)** A gravitational field effectively removes an angle $\Delta\phi$ from the usual 2π radians of a circle enclosing the field. This is equivalent to space being bent into an extra (fourth) spatial dimension w. This warping gets steeper as one approaches a massive object, then gradually levels off inside the object since less and less of the object's mass is enclosed by smaller and smaller circumferences.

Since space and time are interchangeable, gravity shortens space (the circumference of the circle) by the same fraction as it shortens time in Eq. (26),

$$\frac{\Delta\phi}{2\pi} = \frac{\Delta t}{t} \tag{29}$$

$$= \frac{GM}{c^2 r} \, . \tag{30}$$

As shown in Fig. 12(b), this spatial distortion is equivalent to space being bent into an extra (fourth) spatial dimension w that is perpendicular to the usual x, y, and z spatial dimensions, just as the paper in Fig. 12(a) is forced to bend into an extra (third) dimension. If the spatial warping removes an angle of $\Delta\phi$ from the expected 2π radians of circumference surrounding the gravitational well, the change in the circumference is proportional to the change in the effective radius of a circle around the well:

$$\frac{2\pi - \Delta\phi}{2\pi} = \frac{dr}{ds} = \sqrt{\frac{dr^2}{dr^2 + dw^2}} = \frac{1}{\sqrt{1 + (dw/dr)^2}} \tag{31}$$

$$\approx 1 - \frac{1}{2}\left(\frac{dw}{dr}\right)^2 , \tag{32}$$

in which Eq. (32) used a Taylor expansion assuming that $dw/dr \ll 1$.

By combining Eqs. (30) and (32), one can calculate the extra spatial dimension w:

$$\frac{dw}{dr} \approx \sqrt{\frac{2GM}{c^2 r}}$$

$$\Rightarrow \quad w \approx \sqrt{\frac{8GMr}{c^2}} . \tag{33}$$

This warping gets steeper as one approaches a massive object, then gradually levels off inside the object since less and less of the object's mass is enclosed by smaller and smaller circumferences. (This is true for objects like stars and planets with distributed mass, but not for black holes with lots of mass in the center.)

Thus space is like a "rubber sheet" (representing only two dimensions of space for simplicity) that can be warped into a higher "**embedding**" dimension as shown in Fig. 12(b); a massive object (e.g. a star) will create a pothole in the rubber sheet. The embedding dimension is a fifth dimension perpendicular to the usual four (3 space + 1 time); it may or may not have any real physical significance as an extra unseen dimension, but at the very least it is useful for visualizing the spatial curvature. (Spatial curvature does not strictly necessitate the physical existence of this embedding dimension, since one could view curvature in terms of a variable amount or "density" of "real distance" per "unseen grid distance" in different directions.)

In the same way that straight lines on the rubber sheet become curved when the gravitational depression is made, light and inertial "straight-line" motion are curved by the spatial warpage of a massive object. The warped-space versions of what used to be straight lines are called geodesic lines, and they will be described mathematically by the nasty Christoffel symbols in Section 3.1.

3 Principles of General Relativity–Scary Math Version

Having covered the principles of general relativity in a physically intuitive fashion in Section 2, we will now present the official mathematical version that has for so long made students quake in their sneakers when they think of general relativity. The terms defined in this section are staples of general relativity, but they consist almost entirely of gory algebra with very little simple physical insight into what they mean precisely or why they are defined as they are. The modest physical insights from the better textbooks [3]-[11] are given below. Fortunately, the simpler, more intuitive approach of Section 2 is sufficient for analyzing the most frequently encountered problems in general relativity. In other words, if Section 3 becomes too mathematically violent, you can skip it and move on to the remaining sections without missing much except post-traumatic stress disorder.

The scary math to be covered falls into four different areas: space-time curvature, the stress-energy tensor, Einstein's nonlinear field equation that connects those two, and a linearized version of Einstein's equation that is easier to solve.

3.1 Space-Time Curvature

A large part of what makes the math of general relativity so frightening is that is uses tensors–really big tensors. A subscript or superscript index denotes the four dimensions of time plus space. A rank-0 tensor without indices, like G, is a scalar, or just a number. A rank-1 tensor like x^μ is a 4×1 matrix, or a vector. A rank-2 tensor like $\eta_{\mu\nu}$ is a 4×4 matrix, which occupies two dimensions of a page when written out as in Eq. (12). So far no problem, right? We've done physics calculations since we were in preschool with scalars, vectors, and two-dimensional matrices.

The problem is that general relativity involves rank-3 tensors like $\Gamma^\alpha{}_{\beta\gamma}$ that are 4×4×4 matrices. In other words, if the complete matrix is written all together, it occupies three dimensions, the two in the page and the one coming out of the page, and it looks like a cube built out of 64 individual numbers. But even worse, general relativity has rank-4 tensors like $R^\alpha{}_{\beta\gamma\delta}$ that are 4×4×4×4 matrices! Try writing that easily unless your notebook pages are four-dimensional tesseracts.

Whereas the metric of flat space-time in special relativity is $\eta_{\mu\nu}$ from Eq. (12), the metric of general (possibly warped) space-time in general relativity is $g_{\mu\nu}$, so that the distance ds between two infinitesimally close points in space-time is

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \quad \text{for curved space-time} \tag{34}$$

In other words, in general relativity, one raises and lowers indices using $g_{\mu\nu}$ and $g^{\mu\nu}$ (instead of $\eta_{\mu\nu}$ and $\eta^{\mu\nu}$ as in special relativity), where $g^{\mu\nu}$ is defined to be the matrix inverse of $g_{\mu\nu}$. The metric tensor is symmetric under the interchange of indices: $g_{\mu\nu} = g_{\nu\mu}$.

Another definition sometimes encountered is $g \equiv \det(g_{\mu\nu})$.

Thankfully, at least G still means the old-fashioned Newton's gravitational constant $G \approx 6.67 \times 10^{-11}$ m^3 kg^{-1} sec^{-2}.

As a very simple, compact notation, commas in subscripts or superscripts denote differentiation with respect to the index or indices after the comma. For example:

$$\begin{aligned} X_{\mu,\nu} &\equiv \partial_\nu X_\mu \equiv \partial X_\mu / \partial x^\nu \\ X_{\mu,}{}^{\nu} &\equiv \partial^\nu X_\mu \equiv \partial X_\mu / \partial x_\nu \end{aligned}$$

The Christoffel symbol or affine connection describes how straight lines in flat space-time get bent in warped space-time:

$$\Gamma^{\alpha}{}_{\beta\gamma} = \Gamma^{\alpha}{}_{\gamma\beta} \equiv \frac{1}{2} g^{\alpha\sigma}(g_{\sigma\beta,\gamma} + g_{\sigma\gamma,\beta} - g_{\beta\gamma,\sigma}) \qquad \textbf{Christoffel symbol} \text{ or } \textbf{affine connection} \qquad (35)$$

Freely moving particles in curved (Riemann) space follow geodesic lines, the same way freely moving particles would follow straight lines in flat space:

$$\frac{d^2x^{\alpha}}{ds^2} + \Gamma^{\alpha}{}_{\beta\gamma}\frac{dx^{\beta}}{ds}\frac{dx^{\gamma}}{ds} = 0 \qquad \text{geodesic line in Riemann space} \qquad (36)$$

In flat space $\Gamma^{\alpha}{}_{\beta\gamma} = 0$ for all components. Using this fact and rewriting things in terms of the proper time τ, where $(d\tau)^2 = -(ds)^2$, the equation of motion reduces to

$$\frac{d^2x^{\alpha}}{d\tau^2} = 0 \qquad \text{classical law of inertia.} \qquad (37)$$

The curvature of what used to be straight lines must be considered when taking derivatives of things that vary with time or spatial position. This prompts the introduction of a fancy new derivative (called a **covariant derivative**) that uses $\Gamma^{\alpha}{}_{\beta\gamma}$ to account for the effects of curvature:

$$A^{\mu}{}_{;\nu} \equiv A^{\mu}{}_{,\nu} + \Gamma^{\mu}{}_{\nu\sigma}A^{\sigma} \qquad \text{covariant derivative of vector } A^{\mu} \text{ with respect to } x^{\nu} \qquad (38)$$

Note that a semicolon is used to distinguish this type of derivative from its less sophisticated cousin, which only uses a measly comma.

The Riemann curvature tensor describes space-time warping:

$$R^{\alpha}{}_{\beta\gamma\delta} \equiv \Gamma^{\alpha}{}_{\beta\delta,\gamma} - \Gamma^{\alpha}{}_{\beta\gamma,\delta} + \Gamma^{\alpha}{}_{\sigma\gamma}\Gamma^{\sigma}{}_{\beta\delta} - \Gamma^{\alpha}{}_{\sigma\delta}\Gamma^{\sigma}{}_{\beta\gamma} \qquad \textbf{Riemann curvature tensor} \qquad (39)$$

The Riemann curvature tensor's components are all zero if and only if the space is flat. In a region free of matter and energy, $R^{\alpha}{}_{\beta\gamma\delta}$ may still have nonzero components, indicating that the space in that region is warped by matter or energy in nearby regions.

The Ricci tensor is a partially contracted version of the Riemann tensor and describes *changes* in the space-time warping due to the presence of mass or energy:

$$R_{\mu\nu} = R_{\nu\mu} \equiv R^{\sigma}{}_{\mu\sigma\nu} \qquad \textbf{Ricci curvature tensor} \qquad (40)$$

The Ricci tensor's components are all zero for a certain region if there is no matter or energy in that region; even if $R_{\mu\nu} = 0$ for all components in a given region of space-time, that region may still be warped.

Contracting the Ricci tensor all the way gives the Ricci scalar curvature:

$$R \equiv R^{\nu}{}_{\nu} = g^{\mu\nu}R_{\mu\nu} \qquad \textbf{Ricci scalar curvature} \qquad (41)$$

The Einstein tensor is very similar to the Ricci tensor:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \qquad \textbf{Einstein curvature tensor} \qquad (42)$$

Like the Ricci tensor, the Einstein tensor describes changes in the space-time warping specifically due to the local presence of matter-energy. Its components are all zero for a certain region if there is no matter-energy in that region, although a region of space-time may still be warped even if $G_{\mu\nu} = 0$ for all components there. The similarities between $G_{\mu\nu}$ and $R_{\mu\nu}$ imply that:

$$G_{\mu\nu} = 0 \iff R_{\mu\nu} = 0\,. \qquad (43)$$

3.2 Stress-Energy Tensor

$T^{\mu\nu}$ is the **stress-energy tensor** due to matter or energy. "Momentum-energy tensor" would probably be a better name, since the physical meaning of the stress-energy tensor's components is:

$$T^{\mu 0} = p^\mu/\text{volume} \quad \begin{cases} T^{00} = \text{density of mass-energy} \\ T^{j0} = 1/c \text{ times the energy flux in the } j \text{ direction} \\ \quad\;\; = 1/c \text{ times the density of the } j \text{ component of momentum} \end{cases}$$

$$\begin{aligned} T^{jk} &= j \text{ component of force per area acting on } k \text{ side of an infinitesimal cube} \\ &= k \text{ component of flux of } j \text{ component of momentum} \end{aligned}$$

The three-dimensional matrix T^{jk} may also be identified with the three-dimensional matrix in *Fluid Mechanics and Aerodynamics* ?.? which describes the pressure and viscous forces within a fluid.

The stress-energy tensor is symmetric: $T^{\mu\nu} = T^{\nu\mu}$.

For a gas with mass density ρ, pressure p, and velocity v_μ, the stress-energy tensor is:

$$T^{\mu\nu} \equiv \left(\rho + \frac{p}{c^2}\right) v_\mu v_\nu - g_{\mu\nu} p \;=\; \begin{pmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}_{\text{rectangular, flat space, rest frame}} \tag{44}$$

The stress-energy tensor for electromagnetic fields is:

$$T^{\mu\nu} = \frac{1}{4\pi}\left(F^{\mu\alpha}F^{\nu}_{\ \alpha} - \frac{1}{4} g^{\mu\nu} F_{\alpha\beta}F^{\alpha\beta}\right) \qquad \begin{cases} T^{00} = \frac{1}{8\pi}\left(E^2 + B^2\right) \\ T^{0j} = T^{j0} = \frac{1}{4\pi}\left(\mathbf{E}\times\mathbf{B}\right)^j \\ T^{jk} = \frac{1}{4\pi}[-(E^jE^k + B^jB^k) + \frac{1}{2}(E^2+B^2)\delta^{jk}] \end{cases} \tag{45}$$

On the extreme right-hand-side of Eq. (45), the components of the stress-energy tensor have been written for flat space-time and Cartesian spatial coordinates.

The conservation laws for energy and momentum may be written in terms of the stress-energy tensor as

$$T^{\mu\nu}{}_{;\nu} \equiv T^{\mu\nu}{}_{,\nu} + \Gamma^{\mu}{}_{\sigma\mu} T^{\sigma\nu} + \Gamma^{\nu}{}_{\sigma\nu} T^{\mu\sigma} = 0 \qquad \textbf{conservation of energy and momentum} \tag{46}$$

Note that whereas the covariant derivative of a rank-1 tensor requires only one term involving the affine connection, the covariant derivative of a rank-2 tensor like $T^{\mu\nu}$ contains two terms involving the affine connection. Of course, in flat space-time the semicolon could be replaced by a comma and one could forget that one had ever had the misfortune to hear anything about affine connections.

By assuming space-time to be flat and separating Eq. (46) into its energy and momentum components, one finds

$$T^{0\nu}{}_{,\nu} \equiv -\frac{1}{c}\frac{\partial T^{00}}{\partial t} + \frac{\partial T^{01}}{\partial x^1} + \frac{\partial T^{02}}{\partial x^2} + \frac{\partial T^{03}}{\partial x^3} = 0 \quad \text{conservation of energy} \tag{47}$$

$$T^{i\nu}{}_{,\nu} \equiv -\frac{1}{c}\frac{\partial T^{i0}}{\partial t} + \frac{\partial T^{i1}}{\partial x^1} + \frac{\partial T^{i2}}{\partial x^2} + \frac{\partial T^{i3}}{\partial x^3} = 0 \quad \text{conservation of } i \text{ component of momentum} \tag{48}$$

The detailed physical meaning of Eqs. (47) and (48) may be seen by considering the definitions of each component of $T^{\mu\nu}$ which were given above.

3.3 Einstein's Gravitational Field Equation

Just as $\nabla^2\Phi = -4\pi\rho$ for a Newtonian gravitational potential Φ in the presence of mass density ρ, in general relativity the relationship between gravitational strength and mass (or energy, since $E = mc^2$ for a mass at rest) is:

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \qquad \textbf{Einstein's equation} \tag{49}$$

It follows that just as $\nabla^2\Phi = 0$ for Newtonian gravity in regions of zero mass density, so also the absence of local matter-energy ($T^{\mu\nu} = 0$) in general relativity implies that

$$G_{\mu\nu} = 0 \ \ \text{(or equivalently } R_{\mu\nu} = 0\text{)} \qquad \text{in regions free of matter-energy.} \tag{50}$$

Note that this does not mean that there is no space-time curvature (zero Riemann curvature tensor) in such regions, only that the Einstein curvature tensor is zero there.

Einstein's field equation (49) may be rewritten in a simplified form as the pressure needed to impart a certain curvature to space:

$$\begin{aligned} \text{Pressure } (T_{\mu\nu}) \text{ [in N/m}^2\text{]} &= \frac{c^4}{8\pi G} \times \text{ spatial curvature } (G_{\mu\nu}) \text{ [in 1/m}^2\text{]} \\ &= 4.83 \times 10^{42} \times \text{ spatial curvature } (G_{\mu\nu}) \text{ [in 1/m}^2\text{]} \end{aligned} \tag{51}$$

Because c is so large and G is so small, it requires a pressure of 5×10^{42} N/m^2 to create a spatial curvature of 1 m^{-2}. That means that it takes a heck of a lot of pressure or matter-energy to significantly warp space-time. In other words, space-time is a deformable substance but an incredibly stiff one.

In the author's opinion there is not really a good derivation for Einstein's equation, so it is simplest to take the equation as an assumption that appears to model reality fairly well, just as Newton's earlier law of gravity was an assumption aimed at modeling reality. In one common derivation of Einstein's equation, one first *assumes* a certain form for the gravitational action and then minimizes this action to arrive at Einstein's equation; unfortunately, this method merely exchanges the initial assumption about the field equation for an initial assumption about the gravitational action. Other common derivations consider which candidate curvature tensors one might define which would be as simple as possible and yet would obey all the appropriate mathematical and physical constraints; this approach is also just a form of shifting the initial assumptions, and it entails a lot of frightening mathematical abstraction as well. Readers interested in these derivations may consult [7].

3.4 Linearized Gravitational Field Equations

For more insight into what the equations of general relativity mean, it is helpful to linearize the equations. The linearized equations are also useful in their own right for calculating various physical effects which do not involve strong gravitational fields or extreme curvature of space-time. Analogies between linearized gravity and electromagnetism (EM) will be pointed out and exploited along the way.

If there are only small perturbations $h_{\mu\nu} \ll 1$ to space-time, the metric may be expressed as a linear combination of the perturbations and the metric for unperturbed, flat Minkowski space-time:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \ . \tag{52}$$

It is useful to define the D'Alembertian operator, which is the 4-dimensional space-time equivalent of the 3-dimensional Laplacian operator ∇^2:

$$^2 \equiv \left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \qquad \textbf{D'Alembertian operator} \ . \tag{53}$$

The linearized versions of the Christoffel symbols and curvature tensors for the above metric are:

$$\begin{aligned}
\Gamma^{\alpha}{}_{\beta\gamma} &\approx \frac{1}{2}\eta^{\alpha\sigma}(h_{\sigma\beta,\gamma} + h_{\sigma\gamma,\beta} - h_{\beta\gamma,\sigma}) \\
R_{\alpha\beta\gamma\delta} &\approx -\frac{1}{2}(h_{\alpha\gamma,\beta\delta} + h_{\beta\delta,\alpha\gamma} - h_{\alpha\delta,\beta\gamma} - h_{\beta\gamma,\alpha\delta} \\
R_{\mu\nu} &\approx -\frac{1}{2}\left[h_{,\mu\nu} +^2 h_{\mu\nu} - \eta^{\sigma\rho}(h_{\mu\sigma,\rho\nu} + h_{\nu\sigma,\rho\mu}) \right]
\end{aligned}$$

Two useful quantities are defined as:

$$h \equiv \eta^{\mu\nu} h_{\mu\nu} \qquad \text{scalar perturbation} \tag{54}$$

$$\overline{h_{\mu\nu}} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h \qquad \textbf{gravitational potential} \text{ (like vector potential } A^{\mu} \text{ in EM)} \tag{55}$$

Similarly, one may write:

$$\overline{h} \equiv \eta^{\mu\nu}\overline{h_{\mu\nu}} = -h \ ; \qquad \text{scalar gravitational potential} \tag{56}$$

$$h_{\mu\nu} = \overline{h_{\mu\nu}} - \frac{1}{2}\eta_{\mu\nu}\overline{h} \ . \tag{57}$$

Continuing the analogy with electromagnetism, one finds:

$$\partial_{\nu}\overline{h^{\mu\nu}} = 0 \qquad \textbf{Lorentz (or Hilbert) gauge condition} \text{ (like } \partial_{\nu}A^{\nu} = 0 \text{ in EM)} \tag{58}$$

In the Lorentz gauge the Ricci tensor and related parameters may be simplified:

$$R_{\mu\nu} \approx -\frac{1}{2}{}^2 h_{\mu\nu}$$

$$R \approx -\frac{1}{2}{}^2 h$$

$$G_{\mu\nu} \approx -\frac{1}{2}{}^2 \left(h_{\mu\nu} - \frac{1}{2}h\eta_{\mu\nu}\right) = -\frac{1}{2}{}^2\overline{h_{\mu\nu}}\,.$$

With this last expression for the Einstein tensor, Einstein's gravitational equation becomes:

$${}^2\overline{h_{\mu\nu}} \approx -\frac{16\pi G}{c^4}T_{\mu\nu} \quad \textbf{linearized Einstein's equation} \text{ (like } {}^2A^\nu = -4\pi J^\nu \text{ in EM)} \tag{59}$$

By analogy with the solution of the electrostatic potential in electromagnetic theory,

$$\nabla^2\Phi = -4\pi\rho \;\rightarrow\; \Phi(\mathbf{x}, t) = \int \frac{d^3\mathbf{x}'\;\rho(\mathbf{x}, t - |\mathbf{x} - \mathbf{x}'|)}{|\mathbf{x} - \mathbf{x}'|}\,.$$

the linearized gravitational field equation may be solved:

$${}^2\overline{h_{\mu\nu}} = -\frac{16\pi G}{c^4}T_{\mu\nu} \;\rightarrow\; \overline{h_{\mu\nu}}(\mathbf{x}, t) = \frac{4G}{c^4}\int \frac{d^3\mathbf{x}'\;T_{\mu\nu}(\mathbf{x}, t - |\mathbf{x} - \mathbf{x}'|)}{|\mathbf{x} - \mathbf{x}'|}\,. \tag{60}$$

Incidentally, in the weak-field limit, the Newtonian gravitational potential Φ may be recovered from the metric via the relation:

$$g_{00} \approx 1 + \frac{2\Phi}{c^2} + O\left(\frac{v}{c}\right)\,. \tag{61}$$

4 Classic Tests of General Relativity

While general relativity can cause a number of interesting phenomena, this section will focus on four well-known tests of the theory: gravitational time dilation, gravitational deflection of light, the precession of Mercury's orbital perihelion around the sun, and gravitational waves. All of these effects may be explained accurately and simply using physical insight into how gravitational fields shorten space and time.

4.1 Gravitational Time Dilation

The first test of general relativity is that time moves more slowly in a gravitational field. It is straightforward to measure this effect by comparing atomic clocks on the earth with atomic clocks in global positioning system (GPS) satellites orbiting the earth (Fig. 13).

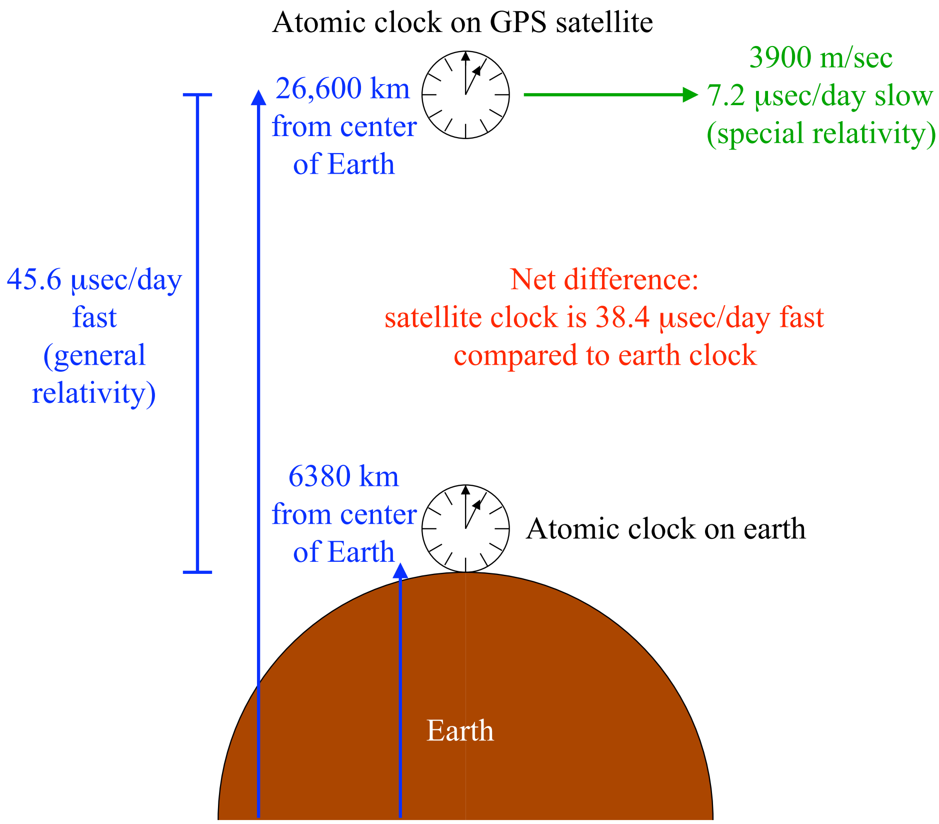

Figure 13. First test of general relativity: time moves more slowly in a gravitational field. An atomic clock on a global positioning system (GPS) satellite may be compared with an atomic clock on Earth. Because the GPS clock is orbiting at 3900 m/sec, special relativity dictates that it will be 7 μsec/day slow relative to the Earth clock. However, because the GPS clock is 20,200 km further away from the Earth's gravitational field than the Earth clock, general relativity predicts it will be 45 μsec/day slow relative to the Earth clock. Thus the net prediction is that the GPS clock will be 38 μsec/day slow relative to the Earth clock, and this is indeed the observed amount. Smaller but measurable effects have also been found using atomic clocks on airplanes.

An accurate calculation of the difference between earth and GPS clocks involves both special relativity (since the orbiting clock is moving much faster than the earth clock) and general relativity (since the orbiting clock is further out of the earth's gravitational field). Both effects will be calculated below.

In the rotating frame of reference of a satellite in a circular orbit, the inward gravitational acceleration GM/r^2 is exactly balanced by the outward centifugal acceleration v^2/r, so the satellite's orbital velocity is

$$v = \sqrt{\frac{GM}{r}} \tag{62}$$

$$= 3870 \ \text{m/sec} \qquad \text{for GPS satellite} \tag{63}$$

using the earth's mass $M = 5.98 \times 10^{24}$ kg and GPS satellite orbital radius $r = 2.66 \times 10^7$ m. An atomic clock on the surface of the earth at the equator has a velocity of 46.4 m/sec from the rotation of the earth; clocks closer to the poles would have less velocity. The velocity of the earth clock will be neglected in our special relativity calculations, since it is an order of magnitude smaller than the velocity of the orbiting clock.

For small velocities, the special relativistic time dilation from Eq. (1) may be approximated as:

$$\begin{aligned} \frac{\Delta t}{t} &= \gamma - 1 = \frac{1}{\sqrt{1-(v/c)^2}} - 1 \\ &\approx \frac{1}{1-(v/c)^2/2} - 1 \approx 1 + \frac{1}{2}\left(\frac{v}{c}\right)^2 - 1 = \frac{1}{2}\left(\frac{v}{c}\right)^2 \end{aligned} \tag{64}$$

$$\approx 8.32 \times 10^{-11} \approx 7.19 \ \mu\text{sec/day} \qquad \text{for GPS satellite} \tag{65}$$

where we have used the satellite velocity from Eq. (63). Since time passes more slowly for moving objects than objects at rest, the GPS clock will be slower than the earth clock by this 7.2 μsec, not including general relativistic effects that will now be calculated.

Using Eq. (26), the general relativistic time difference between the earth and satellite clocks is:

$$\frac{\Delta t}{t} = \frac{GM}{c^2}\left(\frac{1}{r_{\text{orbit}}} - \frac{1}{r_{\text{earth}}}\right) \tag{66}$$

$$\approx -5.28 \times 10^{-10} \approx -45.6 \ \mu\text{sec/day} \qquad \text{for GPS satellite} \tag{67}$$

Because time moves more rapidly in regions of lower gravity than in regions of higher gravity, the GPS clock will be faster than the earth clock by this 45.6 μsec/day, not including the special relativistic effect calculated above. Combining the two effects, the GPS clock should be 38.4 μsec/day faster than the earth clock, which agrees well with actual measurements.

4.2 Deflection of Light by a Gravitational Field

The second test of general relativity is that massive objects such as the sun can bend the trajectories of light beams, even though photons of light do not have any mass. Thus for a star whose light passes near the sun [Fig. 14(a)] on its way to the earth, observers on earth would see an image of the star that is shifted from the star's actual position. Gravitational space shortening and gravitational time shortening each deflect passing light toward the sun; these two effects may be computed separately and then added together to get the total deflection.

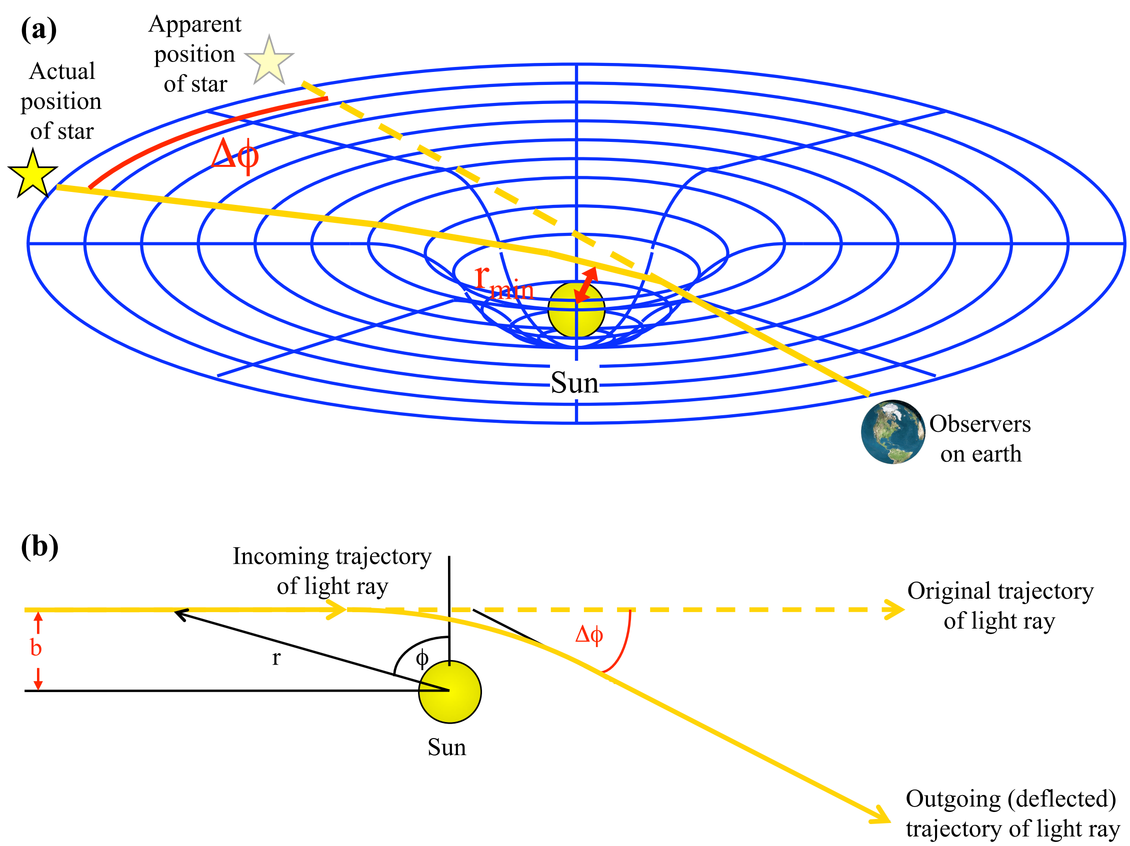

Figure 14. Second test of general relativity: bending of light by a gravitational field. **(a)** Starlight traveling a minimum distance r_{min} from the sun on its way to Earth is deflected by the sun's gravitational field by an angle $\Delta\phi$. Thus observers on earth would see an image of the star that is shifted from the star's actual position. In order to experimentally determine the deflection, the apparent position of the star relative to other stars can be compared when the starlight passes near the sun versus other times when the starlight passes much further from the sun. **(b)** The deflection angle may be calculated using general relativity. The impact parameter is b, distance at any point from the center of the sun is r, and angle relative to closest approach is ϕ.

As shown in Fig. 14(b), it will be assumed that light passes the sun with an impact parameter b. If ϕ is the angle around the sun relative to the point of closest approach, the radial distance of the light from the star for a given angle is $r = b/\cos\phi$, neglecting deflection. This approximation works very well for small deflection angles.

Gravitational shortening of space deflects the light by an angle:

$$\begin{aligned}(\Delta\phi)_{\mathrm{space}} &\approx \int d\phi \frac{GM}{c^2 r} \\ &\approx \int d\phi \cos\phi \frac{GM}{c^2 b} \\ &\approx \frac{2GM}{c^2 b}\ . \end{aligned} \tag{68}$$

Gravitational shortening of time tilts the time axis of the light toward the star. Along the direction of travel, the forward tilt which the light picks up on its way toward the star is cancelled by the backward tilt induced as the light leaves the star. The only net time tilting is perpendicular to the direction of travel, in the direction toward the star. Thus the deflection due to time shortening is:

$$\begin{aligned}(\Delta\phi)_{\mathrm{time}} &\approx \int d\beta \cos\phi \;=\; \int \left(dt\frac{d\beta}{dt}\right)\left(\frac{b}{r}\right) \\ &\approx \frac{GMb}{c}\int_{-\infty}^{\infty}\frac{dt}{r^3} \\ &\approx \frac{GMb}{c}\int_{-\infty}^{\infty}\frac{dt}{(b^2+c^2t^2)^{3/2}} \\ &\approx \frac{2GM}{c^2 b}\ . \end{aligned} \tag{69}$$

For small deflections, the impact parameter in Eqs. (68) and (69) can be approximated by the minimum distance from the center of the sun, $b \approx r_{\mathrm{min}}$. As expected from Eq. (29), the gravitational space-shortening and time-shortening effects are equal and additive (they are both toward the sun), so the total deflection is:

$$(\Delta\phi)_{\mathrm{total}} \approx \frac{4GM}{c^2 r_{\mathrm{min}}} \qquad \textbf{light trajectory deflection due to gravity} \tag{70}$$

$$= 1.75\ \ \mathrm{arcsec} \qquad \text{for starlight just grazing the sun's surface} \tag{71}$$

using observed values for the sun's mass $M_{\odot} = 1.989 \times 10^{30}$ kg and radius $R_{\odot} = 6.96 \times 10^{8}$ m. Solar eclipses are the best opportunity to image stars whose light passes near the sun without being overwhelming the imaging equipment with light from the sun itself. Measured values agree well with the prediction of general relativity.

Because gravitational fields can bend light, massive objects can act as gravitational lenses. Light from an object behind a gravitational lens can be bent around the sides of the gravitational lens, creating a multiple image of the object that emitted the light. Many gravitational lenses have been detected by looking for such multiple images.

4.3 Shifting of Mercury's Perihelion

The third test of general relativity is the precession of orbits in a gravitational field. A specific example is the planet Mercury, which follows a slightly elliptical orbit close to the sun. In contrast to the predictions of nonrelativistic classical mechanics, the angular locations of the perihelion (point closest to the sun) and aphelion (point furthest from the sun) of the orbit slowly precess around the sun, as shown in Fig. 15.

The total precession angle $\Delta\phi$ per orbit is actually a combination of: (a) special relativistic effects on space, since Mercury's velocity is a small but nonzero fraction of light speed, (b) special relativistic effects on time, (c) general relativistic effects on space, since the sun's gravity distorts space slightly, and (d) general relativistic effects on time. All of these effects will be calculated separately and then added together to find the total $\Delta\phi$.

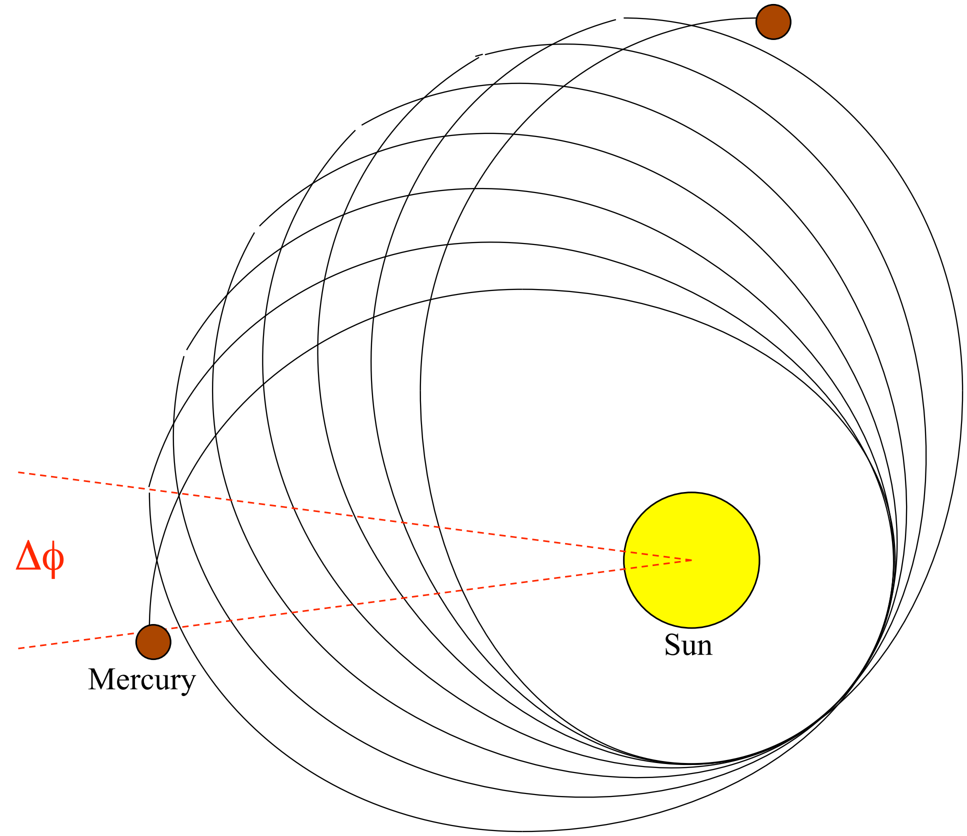

Figure 15. Third test of general relativity: precession of orbits in a gravitational field. In nonrelativistic classical mechanics, orbits in a central gravitational field return to their precise starting point after making a 360^o revolution, barring interfering effects from drag, other orbiting bodies, etc. In contrast, relativity predicts that the orbit will precess by an angle $\Delta\phi$ per revolution. This effect is pronounced for orbits close to a strong gravitational field, such as Mercury's orbit around the sun. The precession of Mercury's orbit has been measured and agrees well with the relativistic prediction.

Mercury's orbit is only slightly elliptical, so to simplify calculations it may be treated as approximately circular with radius r and velocity v from Eq. (62). Because this velocity is a small but nonzero fraction of light speed, special relativistic effects will alter space and time by a fraction:

$$\begin{aligned} \frac{\Delta x}{x} &= \frac{\Delta t}{t} = \gamma - 1 \approx \frac{1}{2}\left(\frac{v}{c}\right)^2 \\ &\approx \frac{1}{2}\frac{GM}{c^2 r} \end{aligned} \tag{72}$$

where the first line of Eq. (72) used Eq. (64) and the second line used Eq. (62).

In Mercury's moving frame of reference, special relativistic effects contract the space around its orbit by the fraction in Eq. (72). Since the orbit normally covers 2π radians, the spatial contraction alters each orbit by 2π times the fractional change of Eq. (72):

$$(\Delta\phi)_{\text{special relativity, space}} = \pi \frac{GM}{c^2 r} . \tag{73}$$

Likewise, Mercury's time will run more slowly than that of a "motionless" external observer by the fractional amount in Eq. (72). Thus when an external observer thinks that Mercury has made a full revolution of period T, the planet only thinks that it has orbited for a time of $[1-(GM)/(2c^2 r)]T$. Therefore, after the planet has made a "complete" revolution as judged by an external observer, the planet will continue to make another fraction $GM/(2c^2 r)$ of an orbit before it is satisfied that it has completed one full orbit. Since an orbit has 2π radians (neglecting spatial distortion, which would cause second-order effects in this case), the special relativistic time dilation makes each orbit precess by an angle:

$$(\Delta\phi)_{\text{special relativity, time}} = \pi \frac{GM}{c^2 r} . \tag{74}$$

From Eq. (30), general relativistic effects shorten space by removing from a circular orbit around the gravitational field an angle:

$$(\Delta\phi)_{\text{general relativity, space}} = 2\pi \frac{GM}{c^2 r} . \tag{75}$$

From Eq. (26), an external observer would see time on Mercury (deep inside the sun's gravitational field) run more slowly by a fractional amount $GM/c^2 r$. This causes orbital precession for the same reason that the special relativistic time dilation did. The orbital period corresponds to 2π radians, so the precession due to general relativistic time dilation is:

$$(\Delta\phi)_{\text{general relativity, time}} = 2\pi \frac{GM}{c^2 r} . \tag{76}$$

As expected, space and time are equivalent, so the special relativistic space and time effects are equal to each other, and the general relativistic space and time effects are equal to each other. All of these effects alter the orbit in the same direction. Adding together all the effects of Eqs. (73) through (76), the total precession is

$$(\Delta\phi)_{\text{total}} = 6\pi \frac{GM}{c^2 r} \tag{77}$$

$$\approx 0.10 \ \frac{\text{arcsec}}{\text{orbit}} = 0.43 \ \frac{\text{arcsec}}{\text{year}} \qquad \text{for Mercury's orbit} \tag{78}$$

using the sun's mass $M_\odot = 1.989 \times 10^{30}$ kg and Mercury's average orbital radius $r = 5.8 \times 10^{10}$ m. The precession of Mercury's orbit has been measured and agrees well with this prediction.

4.4 Gravitational Waves

The fourth test of general relativity involves gravitational waves. The linearized gravitational field equation from Eq. (60) may be viewed as a wave equation, as is especially evident in the absence of sources ($T_{\mu\nu} = 0$):

$$\begin{aligned} {}^2\overline{h_{\mu\nu}} = \left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\overline{h_{\mu\nu}} &= 0 \\ \rightarrow \quad \overline{h_{\mu\nu}}(\mathbf{x}, t) &= \mathrm{Re}\left\{C_{\mu\nu}\exp\left(ik_\alpha x^\alpha\right)\right\} \\ &= \mathrm{Re}\left\{C_{\mu\nu}\exp\left[i(\mathbf{k}\cdot\mathbf{x} - \omega t)\right]\right\} \ , \end{aligned} \tag{79}$$

where $C_{\mu\nu}$ is a glorified (tensor) amplitude, the three-dimensional wavevector includes the spatial components of k^α, $\mathbf{k} \equiv (k^1, k^2, k^3)$, and the angular frequency of the wave is $\omega \equiv ck^0$.

The wave equation implies that $k_\alpha k^\alpha = 0$. This condition may be rewritten as $\omega^2 - c^2|\mathbf{k}|^2 = 0$, which is the usual relation for dispersionless waves traveling at the speed of light. Similarly, the Lorentz gauge condition of Eq. (58) means that the wave must satisfy the relation $C^{\mu\alpha}k_\alpha = 0$.

Thus there can be gravitational waves–ripples in the "rubber sheet" of space-time which propagate at the speed of light (Fig. 16). Whereas electromagnetic waves are oscillating electromagnetic fields, gravity waves are actually oscillations of the distances between points–the empty space between two points alternately stretches and contracts. As described in Section 3.3, space-time is very stiff, so it takes an enormous amount of energy to create even small gravity waves.

Gravity waves can theoretically be directly detected by using a laser interferometer to sense fluctuations in the distance between two or more spatial points, such as orbiting satellites (Fig. 16). Unfortunately, gravity waves are weak enough and potential sources of them are scarce enough that gravity waves have not yet been directly detected in this fashion.

However, gravity waves have been detected by an indirect method. Orbiting massive bodies should act as a quadrupole source of gravitational radiation, and because gravity waves carry energy with them, such orbiting bodies should lose energy. Like most general relativistic phenomena, this effect is small, but it has been detected in the specific case of a binary star system (PSR 1913+16) in which one of the stellar objects is a pulsar.

Gravitational waves only behave in a linear fashion when their amplitude is small, as has been assumed above. Strong gravity waves can carry a large amount of energy, and the presence of a large amount of energy can create gravitational fields, hence the nonlinearity. This is equivalent to what would happen if electromagnetic waves carried electric charge, since electric charge is capable of starting electromagnetic waves. For strong gravitational waves one must abandon the linearized theory above and return to the full, nonlinear Einstein equation for gravitational fields.

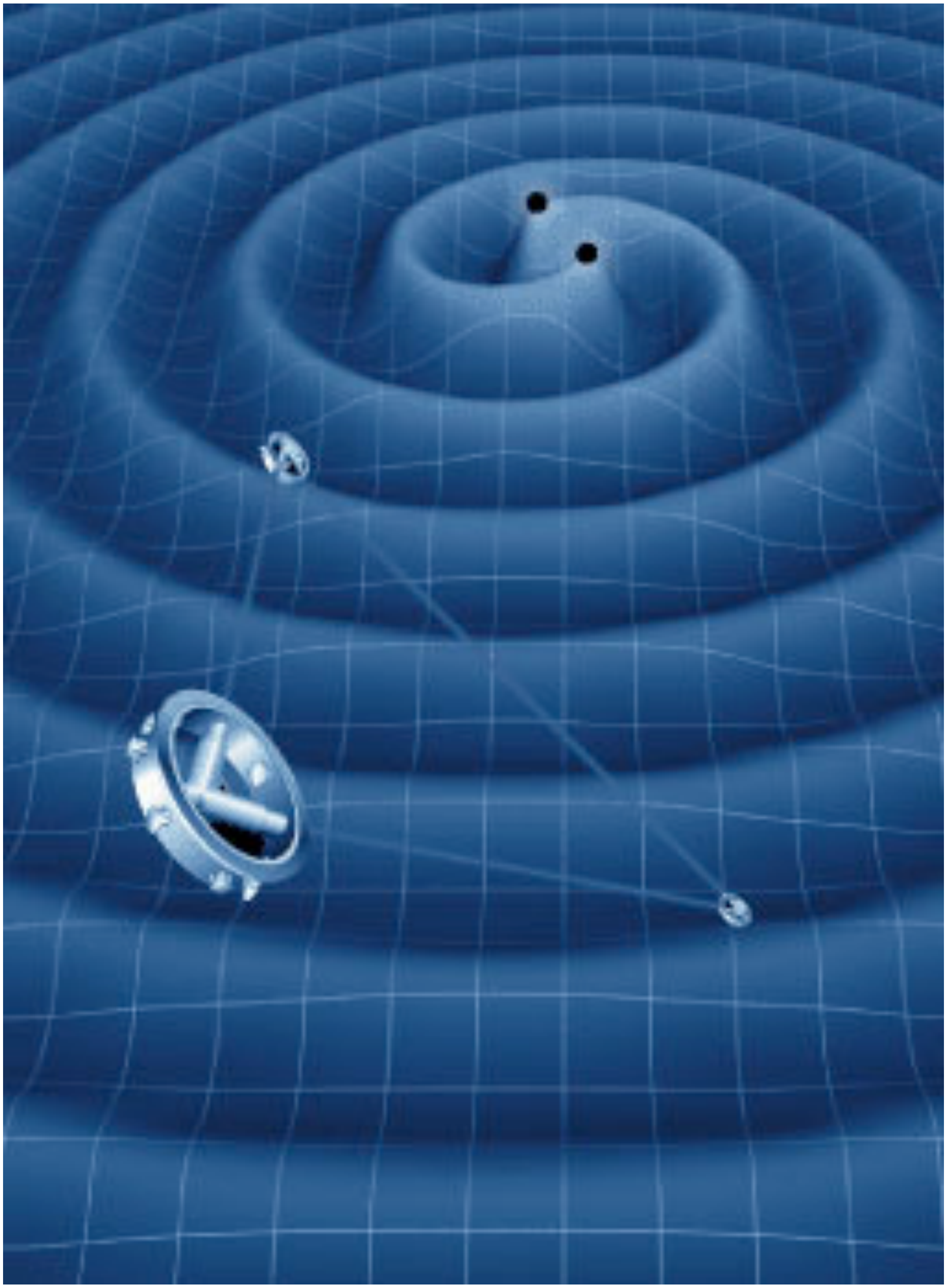

Figure 16. Fourth test of general relativity: two massive bodies orbiting each other produce gravitational waves. Such waves were first detected indirectly via the energy loss of the orbiting bodies as the gravitational waves radiate away some of the system's energy. In principle, gravitational waves could also be detected directly by measuring the very small variations they induce in the distances among three satellites.

5 Stars, Black Holes, and Wormholes

One of the few exact solutions to Einstein's equation is the Schwarzschild solution, which describes the space-time around a massive, motionless, spherically symmetric object. This solution will be derived and applied to stars and simple black holes. Then we will consider modified forms of the solution for more complex black holes and wormholes. For more detailed information on all of these topics, see [7]-[13].

5.1 Schwarzschild Solution

We will present two different greatly simplified derivations of the Schwarzschild solution.

Derivation 1

In the absence of any gravitational warping, the metric for Eq. (34) would be $g_{\mu\nu} = \eta_{\mu\nu}$ from Eq. (12). To account for gravitational effects, we can begin by considering two motionless observers, one at a distance r from a mass M, and the other infinitely far from the mass. From Eq. (26), increments of time dt_r for the first observer will move more slowly than increments of time dt_∞ for the second observer by a factor of

$$dt_r \approx \left(1 - \frac{GM}{c^2 r}\right) dt_\infty \, , \tag{80}$$

in which $GM/(c^2 r)$ is typically far smaller than 1. Squaring both sides of Eq. (80) and neglecting the $[GM/(c^2 r)]^2$ term since it is so small, one finds

$$dt_r^2 \approx \left(1 - \frac{2GM}{c^2 r}\right) dt_\infty^2 \, . \tag{81}$$

The times of the two observers are related by the time-time component g_{00} of the metric tensor:

$$dt_r^2 = -g_{00}\, dt_\infty^2 \tag{82}$$

$$g_{00} \approx -1 + \frac{2GM}{c^2 r} \, . \tag{83}$$

As discussed in Section 2.3, the effect of gravitational warping on space is very similar to its effect on time, so the same term $2GM/(c^2 r)$ is added to the spatial components of the metric tensor (in rectangular cartesian coordinates):

$$g_{11} = g_{22} = g_{33} \approx 1 + \frac{2GM}{c^2 r} \, . \tag{84}$$

Thus the complete Schwarzshild metric is

$$g_{\mu\nu} \approx \begin{pmatrix} -1 + \frac{2GM}{c^2 r} & 0 & 0 & 0 \\ 0 & 1 + \frac{2GM}{c^2 r} & 0 & 0 \\ 0 & 0 & 1 + \frac{2GM}{c^2 r} & 0 \\ 0 & 0 & 0 & 1 + \frac{2GM}{c^2 r} \end{pmatrix}_{\text{rectangular}} \tag{85}$$

$$= \begin{pmatrix} -\left(1 - \frac{2GM}{c^2 r}\right) & 0 & 0 & 0 \\ 0 & \left(1 + \frac{2GM}{c^2 r}\right) & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta \end{pmatrix}_{\text{spherical}} \, . \tag{86}$$

Because of the spherical symmetry about the center of the mass, the spherical form of the metric is the more useful form.

The last step in the derivation is to ensure that the metric is an exact and not merely an approximate solution of the Einstein field equation (49). This requires lots of gory algebra which we will spare the reader so we can keep our PG rating. Suffice it to say that all of the components of the spherical metric in Eq. (85) are exactly correct, except for the radial component, $g_{rr} \approx (1 + 2GM/c^2r)$, which turns out to be a linearized Taylor expansion of the actual nonlinear radial component, $g_{rr} = \left(1 - 2GM/c^2r\right)^{-1}$.

With this change, the full-fledged nonlinear metric for the region around a motionless, spherically symmetric mass is:

$$g_{\mu\nu} \equiv \begin{pmatrix} -\left(1-\frac{2Gm}{c^2r}\right) & 0 & 0 & 0 \\ 0 & \left(1-\frac{2Gm}{c^2r}\right)^{-1} & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2\sin^2\theta \end{pmatrix}_{\text{spherical}} \qquad \textbf{Schwarzschild metric} \qquad (87)$$

Derivation 2

A somewhat more rigorous derivation of the Schwarzschild metric uses the linearized gravitational equations of Section 3.4 to calculate the gravitational field surrounding a motionless, spherically symmetric mass M which has density ρ and occupies a region of space denoted by V.

Since the mass is not moving, the only nonzero component of the stress-energy tensor within the interior of the mass will be $T_{00} = \rho c^2$. Outside of the mass, all of the components of the stress-energy tensor are zero. Because of the relation between the gravitational potential $\overline{h_{\mu\nu}}$ and the stress-energy tensor, the only nonzero component of the gravitational potential is

$$\begin{aligned} \overline{h_{00}}(\mathbf{x}, t) &= \frac{4G}{c^4}\int_V \frac{d^3\mathbf{x}'\ \rho c^2}{|\mathbf{x}-\mathbf{x}'|} \\ &= \frac{4GM}{c^2r}\ , \end{aligned} \qquad (88)$$

where the volume integral of the density simply yielded the total mass M and where r is the distance from the center of the mass.

The contracted (scalar) form of the gravitational potential is then

$$\begin{aligned} \overline{h} &= \eta^{00}\overline{h_{00}} = (-1)\left(\frac{4GM}{c^2r}\right) \\ &= -\frac{4GM}{c^2r}\ . \end{aligned} \qquad (89)$$

Now the linear perturbation to the metric may be calculated:

$$\begin{aligned} h_{00} &= \overline{h_{00}} - \frac{1}{2}\eta_{00}\overline{h} \\ &= \frac{4GM}{c^2r} - \frac{1}{2}(-1)\left(-\frac{4GM}{c^2r}\right) = \frac{2GM}{c^2r} \end{aligned} \qquad (90)$$

$$\begin{aligned} h_{jj} &= \overline{h_{jj}} - \frac{1}{2}\eta_{jj}\overline{h} \\ &= 0 - \frac{1}{2}(+1)\left(-\frac{4GM}{c^2 r}\right) = \frac{2GM}{c^2 r}\,. \end{aligned} \tag{91}$$

Therefore the complete linearized metric is

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} = \begin{pmatrix} -1+\frac{2GM}{c^2r} & 0 & 0 & 0 \\ 0 & 1+\frac{2GM}{c^2r} & 0 & 0 \\ 0 & 0 & 1+\frac{2GM}{c^2r} & 0 \\ 0 & 0 & 0 & 1+\frac{2GM}{c^2r} \end{pmatrix}_{\text{rectangular}}$$

This is the same intermediate result as we got from derivation 1 above. As in derivation 1, this metric can be converted from rectangular to spherical coordinates and corrected for nonlinear effects from the Einstein field equation to obtain the same exact solution–Eq. (87).

The quantity $2GM/c^2$ which appears prominently in the Schwarzschild metric is given a special name:

$$R_s \equiv \frac{2GM}{c^2} \approx \frac{M}{M_\odot} 2.95 \text{ km} \qquad \textbf{Schwarzschild radius} \tag{92}$$

in which the mass M has been compared to the mass $M_\odot \approx 1.99 \times 10^{30}$ kg of the sun.

Using Eq. (92), the Schwarzschild metric from Eq. (87) may be rewritten as:

$$g_{\mu\nu} \equiv \begin{pmatrix} -\left(1-\frac{R_s}{r}\right) & 0 & 0 & 0 \\ 0 & \left(1-\frac{R_s}{r}\right)^{-1} & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2\sin^2\theta \end{pmatrix}_{\text{spherical}} \tag{93}$$

The significance of the Schwarzschild radius will be explained in Section 5.2. For now, note that extremely weird things happen to both time (g_{00}) and space (g_{11}) when $r = R_s$ in Eq. (93).

It is important to mention that the Schwarzschild solution is valid only at radii larger than the radius of the object which creates the gravitational field. As the radial position of the observer passes the surface of the object and moves closer to the center, less and less of the object's matter-energy will contribute to the local gravitational field acting on the observer. (That's basically just the gravitational equivalent of Coulomb's law–you only feel the effects of the stuff that is at smaller radial distances from the center than you are.) Therefore, inside the object the gravitational well will forego the precipitous plunge that appears in the Schwarzschild solution near $r = R_s$ and instead will generally level off at some value at the object's center, as illustrated in Fig. 14. But the Schwarzschild solution is still applicable to the space external to these objects. Because $r \gg R_s$ in these cases, Taylor expansions of the Schwarzschild solution may be used to simplify the analysis.

5.2 Schwarzschild Black Holes

The basic idea of black holes can be derived even without special or general relativity. As shown in Fig. 17, a certain escape velocity is required to completely escape from the gravitational field of an object.

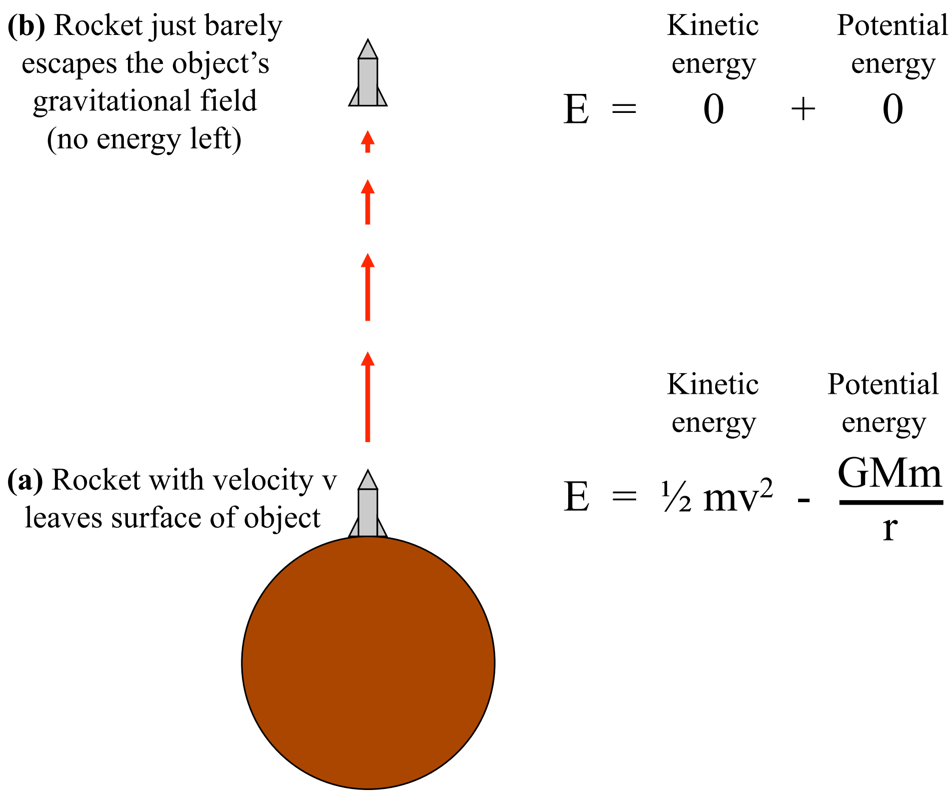

Figure 17. Escape velocity from the gravitational field of an object. (a) A rocket with velocity v leaves the surface of the object; the rocket has an energy $E = mv^2/2 - GMm/r$. **(b)** If the rocket just barely escapes the object's gravitational field, it has no energy left $E = 0$. Equating the energies from (a) and (b) yields the escape velocity $v = \sqrt{2GM/r}$, the minimum initial velocity that will allow the rocket to leave the gravitational field without falling back into it.

If a rocket with velocity v leaves the surface of the object at point (**a**), the rocket has a positive kinetic energy $mv^2/2$ and a negative gravitational potential energy $-GMm/r$, or a total energy

$$E = \frac{1}{2}mv^2 - \frac{GMm}{r} . \tag{94}$$

If the rocket just barely escapes the object's gravitational field, it has no energy left by the time it arrives at point (**b**) far from the object, so the rocket's kinetic plus potential energy is

$$E = 0 . \tag{95}$$

Assuming the rocket simply coasts from point (**a**) to point (**b**) without firing its engines, its energy is conserved–the rocket's total kinetic plus gravitational potential energy is the same at the two points. Equating the energies from Eqs. (94) and (95) yields the escape velocity

$$v_{\text{escape}} = \sqrt{\frac{2GM}{r}} . \tag{96}$$

The escape velocity is the minimum initial velocity that will allow the rocket to leave the object's gravitational field (starting from an initial distance r from the center of the object) without falling back into it.

For reference, escape velocities from the surfaces of several objects in the solar system are given in Table 1.

Object	M **(kg)**	r **(m)**	v_{escape} **(m/sec)**
Asteroid	$< 10^{19}$	$< 10^5$	< 100
Moon	7.35×10^{22}	1.74×10^6	2380
Earth	5.97×10^{24}	6.38×10^6	11,200
Sun	1.99×10^{30}	6.96×10^8	618,000

Table 1. Escape velocities from the surface of an asteroid, the moon, the earth, and the sun.

Note that the escape velocity is higher (more initial kinetic energy is required) to escape from objects with a higher mass M, or to escape from a closer distance r to an object where the gravitational attraction is stronger.

Setting the escape velocity from Eq. (96) equal to the maximum possible velocity, the speed of light c, one finds that even light cannot escape if it gets closer than the Schwarzschild radius $r = R_s$ to an object of mass M:

$$R_s \equiv \frac{2GM}{c^2} \approx \frac{M}{M_\odot} 2.95 \text{ km} \qquad \textbf{Schwarzschild radius} \tag{97}$$

in which $M_\odot$ is the mass of the sun. Even though this derivation did not use special or general relativity, it found the same value for the Schwarzschild radius as the relativistic derivations that led to Eq. (92). An object that has all of its mass confined to a radius smaller than R_s is called a black hole, since its gravitational field is strong enough to trap even light. In order to become a black hole, the sun would have to be compressed to a radius of approximately 3 km, and the earth would have to be compressed to a radius of $R_s \approx 1$ cm. The boundary $r = R_s$ is called the **event horizon**, since any events that occur inside it can never be observed by or reported to observers on the outside. Black holes are like Las Vegas–what happens in them stays in them, and they can both eat a hole in your paycheck too.

Most astrophysical objects do not have enough mass concentrated into a small enough volume for the Schwarzschild radius to matter. The radius of these objects is much larger than R_s and the gravitational well levels off inside the objects without ever reaching the bizarre conditions described on the following pages.

Both simple calculations (e.g., Newtonian gravitational force $-GMm/r^2$) and full-fledged general relativistic calculations predict that the gravitational field of a black hole becomes infinitely strong at $r = 0$, a point called the singularity. In practice, quantum gravitational effects that are poorly understood at this time may intervene to prevent the formation of an infinite singularity. In any event, it is clear that even if someone miraculously survived the trip to the center of a black hole, they would have a very bad day once they arrived there.

As was described for Fig. 5, light cones point from the past to the future and depict the possible locations that one could be in the future or that one could have been in the past, given one's specific location in the present and the constraint that one cannot travel faster than light. Furthermore, as was described for Fig. 11, the forward time direction tilts toward the center of a gravitational field. Therefore, the physics of black holes may be viewed in terms of how light cones tilt over as one approaches and enters a black hole, as shown in Fig. 18.

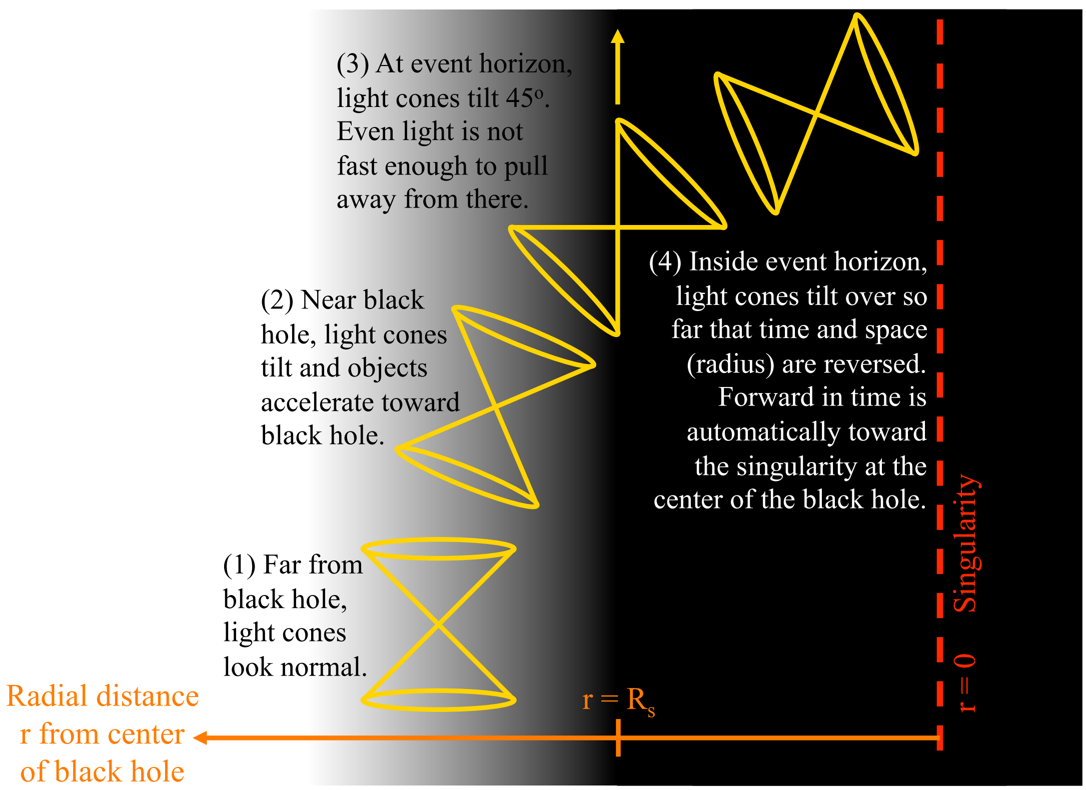

Figure 18. Tilting of light cones near a black hole. The horizontal axis is the radial distance r from the center of the black hole, and the vertical axis is time. (1) Far from the black hole, light cones are not tilted. (2) Near the black hole, light cones tilt toward $r = 0$–objects accelerate toward the hole. (3) At $r = R_S$, light cones tilt at a 45^o angle in space-time, and even light cannot escape. (4) For $r < R_S$, light cones actually tilt so far that time and space (radial position) are reversed; forward in time is automatically toward the singularity at $r = 0$.

As illustrated in Fig. 18, the orientation of a light cone changes as one travels from a point far away from a black hole to the event horizon and then inside the event horizon:

1. Far from the black hole, light cones look normal and point vertically from the past to the future.

2. Near the black hole, light cones tilt toward the black hole and thus objects accelerate toward the hole; this would be true even in the gravitational field of an object that is not a black hole.

3. As we have already found, what is unique about a black hole is that even light cannot escape at the Schwarzshild radius. An equivalent way of viewing this is that at $r = R_s$, light cones tilt at a 45^o angle in space time. Even light is not fast enough to pull away from there, and at best it can only follow the yellow arrow in Fig. 18 straight forward in time without climbing to larger radial distances from the black hole.

4. Inside the event horizon, light cones actually tilt over so far that time and space (radial position) are reversed. In other words, forward in time is automatically toward the singularity at the center of the black hole, $r = 0$, so there is no way to avoid hitting the singularity once one crosses the event horizon. [Another way to see the interchange of space and time inside a black hole is to note that in going from $r > r_S$ to $r < r_S$, the time and radial components of the Schwarzschild metric in Eq. (87) both change sign; the time component goes from negative to positive and the radial component changes from positive to negative. This means that inside a black hole, space (at least in the radial direction) and time effectively become interchanged.] For the same reason that space and time are interchanged inside a black hole, spacelike quantities (e.g. the radial component of momentum) and timelike ones (e.g. energy) are also interchanged; this fact will play a role in the discussion of Hawking radiation in Section 6.1.

Objects crossing the event horizon tilt 45^o in space-time, just as if they were moving at light speed. Thus to external observers, the objects appear frozen in time, taking infinitely long to cross the event horizon. Astronauts entering a black hole would watch the entire future of the universe, then cross the event horizon and hit the singularity in just microseconds of slowed-down ship time.

If matter from a collapsing star asymptotically approaches the Schwarzschild radius as viewed by an observer far away, can you really say that a black hole ever forms before the entire history of the universe is over? Or if a black hole has somehow already managed to come into existence, can you say that it ever manages to finish acquiring new mass that is falling toward the event horizon? For these reasons, black holes were sometimes called "frozen stars" in the older physics literature, especially in the Soviet Union. In practice, within a finite time in the history of the universe, infalling matter can get close enough to the Schwarzschild radius that radiation from it is so red-shifted as to be essentially undetectable to an observer far away. Moreover, the infalling matter is so close to the Schwarzschild radius that an external observer attempting to detect that matter by coming to sample it would end up joining it on its plunge to the event horizon, which would occur within a finite time in the frame of reference of an infalling observer.

Although to an observer far from a black hole, a spaceship falling into the black hole would take infinitely long to reach the event horizon, the results are quite different as perceived by the crew of the ill-fated ship. Using increments of shipboard time dt_{ship}, people on board the ship would measure their radial velocity toward or away from the black hole as dr/dt_{ship}. Assuming for simplicity that the ship falls straight toward the black hole (no angular velocity) and that its velocity is simply due to the gravitational acceleration (no firing of rocket engines, no kinetic energy when the ship was far from the black hole), Eq. (94) may be written as:

$$\begin{aligned} 0 &= \frac{1}{2}m\left(\frac{dr}{dt_{\mathrm{ship}}}\right)^2 - \frac{GMm}{r}, \quad \text{or} \\ \frac{dr}{dt_{\mathrm{ship}}} &= \pm\sqrt{\frac{2GM}{r}} = \pm c\sqrt{\frac{R_s}{r}} \end{aligned} \tag{98}$$

Choosing the - sign in Eq. (98) for inward travel (decreasing radial distance from the center of the black hole) and integrating, one finds the onboard ship time required to fall from an initial radial distance r to the singularity at $r = 0$:

$$\Delta t_{\mathrm{ship},\, r\to 0} = \frac{2}{3}\frac{r^{3/2}}{c\sqrt{R_s}} = r_{\mathrm{km}}^{3/2}\sqrt{\frac{M_\odot}{M}}1.29\ \mu\mathrm{sec} \tag{99}$$

Note that t_{ship} changes smoothly as r crosses the event horizon at R_s. Thus to people on the spaceship, time continues to advance properly as they approach the event horizon, cross it, and pass into the interior of the black hole. However, if the people were monitoring signals from the rest of the universe, they would witness the universe around them rapidly age and pass through its entire future history before they passed through the event horizon.

Using Eq. (99), the onboard ship time required to fall from the event horizon to the singularity is

$$\Delta t_{\mathrm{ship},\, R_s\to 0} = \frac{2}{3}\frac{R_s}{c} = \frac{M}{M_\odot}\,6.56\ \mu\mathrm{sec} \tag{100}$$

Therefore once the ship crosses the event horizon, it reaches the singularity within a matter of microseconds, as measured by clocks on board the ship. As far as is currently known, gravitational fields would become infinitely strong at the singularity. Black holes are not a good vacation destination.

Subtracting Eq. (100) from Eq. (99), the onboard ship time required to fall from radial distance r to the event horizon is

$$\Delta t_{\mathrm{ship},\, r\to R_s} = \frac{2}{3}\frac{r^{3/2} - R_s^{3/2}}{c\sqrt{R_s}} = r_{\mathrm{km}}^{3/2}\sqrt{\frac{M_\odot}{M}}1.29 - \frac{M}{M_\odot}\,6.56\ \mu\mathrm{sec} \tag{101}$$

From the Schwarzschild metric of Eq. (87), the relation between time t_∞ for a distant observer and time for a ship at radial distance r from a black hole is

$$dt_\infty \approx \frac{dt_{\mathrm{ship}}}{\sqrt{1 - R_s/r}} \tag{102}$$

Thus when $r = R_s$, Eq. (102) blows up and an infinitesimally small increment of ship time dt_{ship} corresponds to an infinitely long amount of time for the observer far from the black hole.

5.3 Stars and Stellar Collapse

Toward the end of its life, after a star has consumed most of the hydrogen in its core, it collapses and heats up until it is hot enough to fuse the hydrogen that remains in its mantle; this is called shell burning. The fusion reactions in the mantle cause the surface layers of the star to expand drastically and cool, resulting in an enormous **red giant** star. Further contraction and heating of the core can lead to the fusion of helium and progressively higher elements. In these final stages, a star often blows off its outer layers.

The remaining stellar core collapses to form one of three objects, depending on its mass M:

1. White dwarf. If $M < M_{\text{white dwarf max}}$, the star contracts until its degenerate electrons' Fermi pressure supports it, forming a white dwarf that slowly radiates away its residual energy.

2. Neutron star. If $M_{\text{white dwarf max}} < M < M_{\text{neutron star max}}$, gravity overcomes the Fermi pressure at the white dwarf stage. The star continues to contract until it squeezes its protons and electrons together to form neutrons by inverse beta decay. Such a neutron star acts like a giant nucleus of neutrons and is supported against further collapse by the neutrons' Fermi pressure.

3. Black hole. If $M > M_{\text{neutron star max}}$, gravity overcomes even the Fermi pressure of the degenerate neutrons and the star continues to collapse. The star becomes a black hole once its radius becomes less than the **Schwarzschild radius** R_s.

The limiting masses $M_{\text{white dwarf max}}$ and $M_{\text{neutron star max}}$ may be estimated in a simple fashion. Consider a collapsed star with a density n of degenerate fermions. The degenerate fermions are electrons in a white dwarf and neutrons in a neutron star. For a very dense star that approaches the appropriate limiting mass, the degenerate particles will have such a large Fermi energy that they will be relativistic. From the *Statistical Physics* ?.?, the Fermi energy per degenerate relativistic fermion is

$$E_F \sim \hbar c n^{1/3} \sim \frac{\hbar c N^{1/3}}{R} , \tag{103}$$

where n has been estimated from the total number N of the fermions and the radius R of the star.

The gravitational potential energy per degenerate fermion is roughly

$$E_G \sim -\frac{G M m_{\text{per fermion}}}{R} = -\frac{G m_{\text{per fermion}}^2 N}{R} , \tag{104}$$

in which the total mass of the collapsed star has been written as $M = N m_{\text{per fermion}}$, where $m_{\text{per fermion}}$ is the amount of stellar mass per degenerate fermion. In a white dwarf composed of typical fusion products, on average one proton and one neutron accompany each electron, so $m_{\text{per fermion}} = m_e + m_p + m_n \approx 2m_n$. In a neutron star, most of the mass is in the form of the degenerate neutrons, so $m_{\text{per fermion}} = m_n$.

Thus the total energy per degenerate fermion consists of two terms, both of which vary like $1/R$:

$$E \sim E_F + E_G \sim \frac{\hbar c N^{1/3}}{R} - \frac{G m_{\text{per fermion}}^2 N}{R} , \tag{105}$$

The collapsed star will seek a stable equilibrium that minimizes E. If $E > 0$ in Eq. (105), the energy will be decreased by increasing R. Part of the fermions will then become nonrelativistic, causing the radial dependence of E_F to change to $\sim 1/R^2$. Eventually E_G will begin to dominate over E_F and the system will reach equilibrium at some finite radius R.

On the other hand, if $E < 0$ in Eq. (105), the energy can be decreased without bound by decreasing R, and the star will be unstable to further gravitational collapse. Setting $E = 0$ in Eq. (105), the threshold for stability occurs at a maximum number of degenerate fermions,

$$N_{\text{max}} \;=\; 3\left(\frac{\hbar c}{G}\right)^{3/2} \frac{1}{m^3_{\text{per fermion}}}\,, \tag{106}$$

where the coefficient 3 comes from calculating the Fermi and gravitational energies more precisely by numerically integrating to determine the density profile [12].

The maximum fermion number is equivalent to a maximum stellar mass of

$$M_{\text{max}} \;\equiv\; N_{\text{max}}\, m_{\text{per fermion}} \;\approx\; 3\left(\frac{\hbar c}{G}\right)^{3/2} \frac{1}{m^2_{\text{per fermion}}}\,. \tag{107}$$

The radius of a stable collapsed star near the limiting mass may be found by inserting Eq. (106) into Eq. (103) and assuming that the fermions are somewhat relativistic, $E_F \sim m_{\text{of fermion}} c^2$ (here $m_{\text{of fermion}}$ is the actual mass of each degenerate fermion):

$$R \;\approx\; 3\sqrt{\frac{\hbar^3}{Gc}}\;\frac{1}{m_{\text{of fermion}}\, m_{\text{per fermion}}}\,. \tag{108}$$

The factor of 3 in Eq. (108) again comes from more detailed calculations [12].

Applying Eqs. (107) and (108) to a white dwarf ($m_{\text{of fermion}} = m_e$, $m_{\text{per fermion}} \approx 2m_n$) produces

$$M_{\text{white dwarf max}} \;\approx\; 1.5 M_\odot \qquad R \;\approx\; 10^7\ \text{m} \qquad \overline{\rho} \;\sim\; 10^6\ \text{g/cm}^3 \qquad \textbf{White dwarf} \tag{109}$$

Equations (107) and (108) may also be applied to a neutron star ($m_{\text{of fermion}} = m_{\text{per fermion}} = m_n$), although general relativity and other effects lower the maximum mass by about a factor of 2 [12]:

$$M_{\text{neutron star max}} \;\approx\; 3 M_\odot \qquad R \;\approx\; 10^4\ \text{m} \qquad \overline{\rho} \;\sim\; 10^{15}\ \text{g/cm}^3 \qquad \textbf{Neutron star} \tag{110}$$

Thus white dwarfs have masses up to $1.5M_\odot$, neutron stars are in the range $1.5 - 3M_\odot$, and black holes formed from stars have masses greater than $3M_\odot$.

Pulsars are particularly interesting types of neutron stars that emit periodic bursts of radiation. When a star collapses, its magnetic field is compressed and greatly strengthened. Trapped electrons above the magnetic poles emit narrow beams of cyclotron radiation. A pulsar rotates about an axis that is different from its magnetic axis, causing beams of cyclotron radiation to sweep the sky during each rotation. Depending on the pulsar, the cyclotron radiation may be anywhere from radio to gamma wavelengths, and the rotation period may range from milliseconds to seconds [12].

How do we know that something as weird as a black hole actually exists? Although we cannot see inside a black hole, we can see the radiation from matter that is compressed, accelerated, and heated as it approaches a black hole. The heated infalling matter forms a relatively thin **accretion disk** (due to its angular momentum) as it spirals around the black hole. Although a black hole itself can have no magnetic field, the spiraling infalling matter creates a strong magnetic field, and accelerated charged particles escaping along the field lines lead to large, intense, **polar relativistic jets**. Figure 19 shows a Hubble Space Telescope image of a giant black hole at the center of elliptical galaxy NGC 4261. Numerous other black holes have been identified and studied.

For more information on stellar structure and evolution, see *Plasma Physics and Fusion* 5.

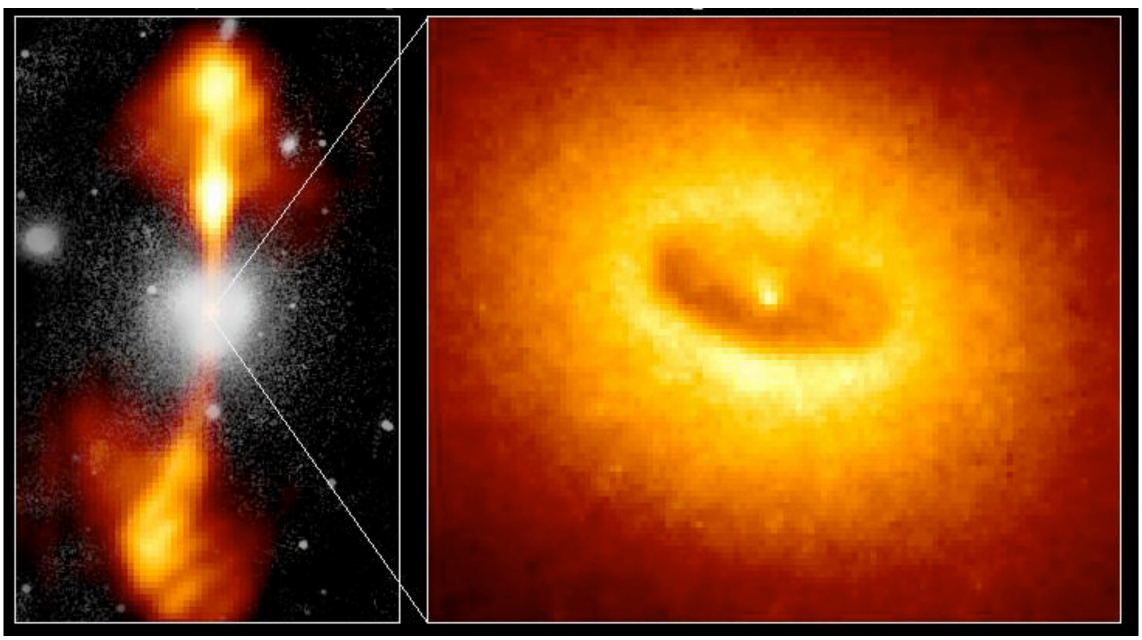

Figure 19. Giant black hole at the center of elliptical galaxy NGC 4261. A black hole's strong gravitational field heats approaching matter until it glows. We can see that glowing matter around the black hole, even though we cannot see inside the black hole. Note the upper and lower polar jets in the left photo, and the accretion disk in the higher-magnification right photo. Numerous other black holes have been identified and studied. (Public domain image from http://upload.wikimedia.org/wikipedia/commons/3/3a/Ngc4261.jpg.)

5.4 Rotating Black Holes

Most stars rotate, so by conservation of angular momentum during the collapse of a star to become a black hole, presumably most black holes have significant rotation. If a black hole is rotating, it creates four major effects that are not produced by non-rotating black holes.

A. The first effect is called **frame dragging** and is depicted in Fig. 20. A rapidly rotating object with a strong gravitational field (which does not necessarily have to be black hole) can essentially drag the nearby surrounding spatial region with it in its rotation. Getting the surrounding space to "move" in the angular direction is equivalent to Lorentz-boosting the space in the angular direction. The forward direction of time becomes tilted in the angular direction of the object's rotation, or in other words light cones are tilted both in the inward radial direction due to the object's gravity and in the angular direction of rotation. This introduces off-diagonal $t\phi$ components (assuming ϕ is the direction of rotation) into the metric describing space-time around rotating objects.

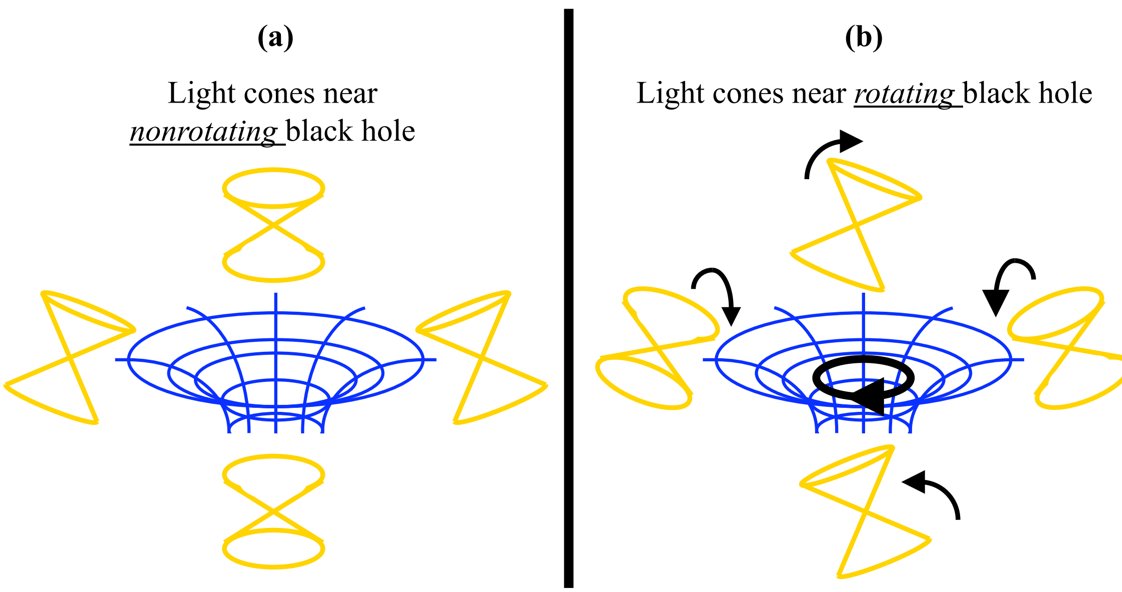

Figure 20. First effect of rotating black holes: frame dragging. (a) Light cones near a nonrotating black hole tilt toward the center of the black hole. **(b)** A rotating black hole twists the surrounding space-time in the direction of the hole's rotation, so light cones near the black hole tilt toward the center of the hole and in the direction of the hole's rotation. The tilting of the light cones forces objects near the black hole to orbit in the direction of the hole's rotation. As the hole's speed of rotation increases, the light cones eventually tip over so far that one can travel through the future of one light cone into the past of the next light cone. In other words, you could travel into the past by looping the right way around a rapidly rotating black hole. While this time travel mechanism becomes possible for black hole rotation speeds less than the speed of light, there may be as-yet-undiscovered physical constraints that prevent this time travel approach.

B. The second effect caused by black hole rotation is that the singularity is a ring instead of a point, as illustrated in Fig. 21. The correct radius of this **ring singularity** can be calculated even without special or general relativity. As a star collapses to become a black hole, its angular momentum $L = \langle Mvr \rangle$ remains constant, where M is the mass of the star, and the angular position r and rotation velocity v of the mass have been averaged over the volume of the star. As the star's radius shrinks the rotation velocity v increases, just as spinning ice skaters rotate more rapidly when they pull their arms in. Of course, matter cannot quite reach the speed of light c, so the minimum radius to which all the mass of the star can contract is

$$R_{\text{ring}} \quad = \quad \frac{L}{Mc} \tag{111}$$

Thus all the matter from the star collapses to form a ring with this radius. All the scary math of special and general relativity gives the same result for those who prefer to do it the hard way.

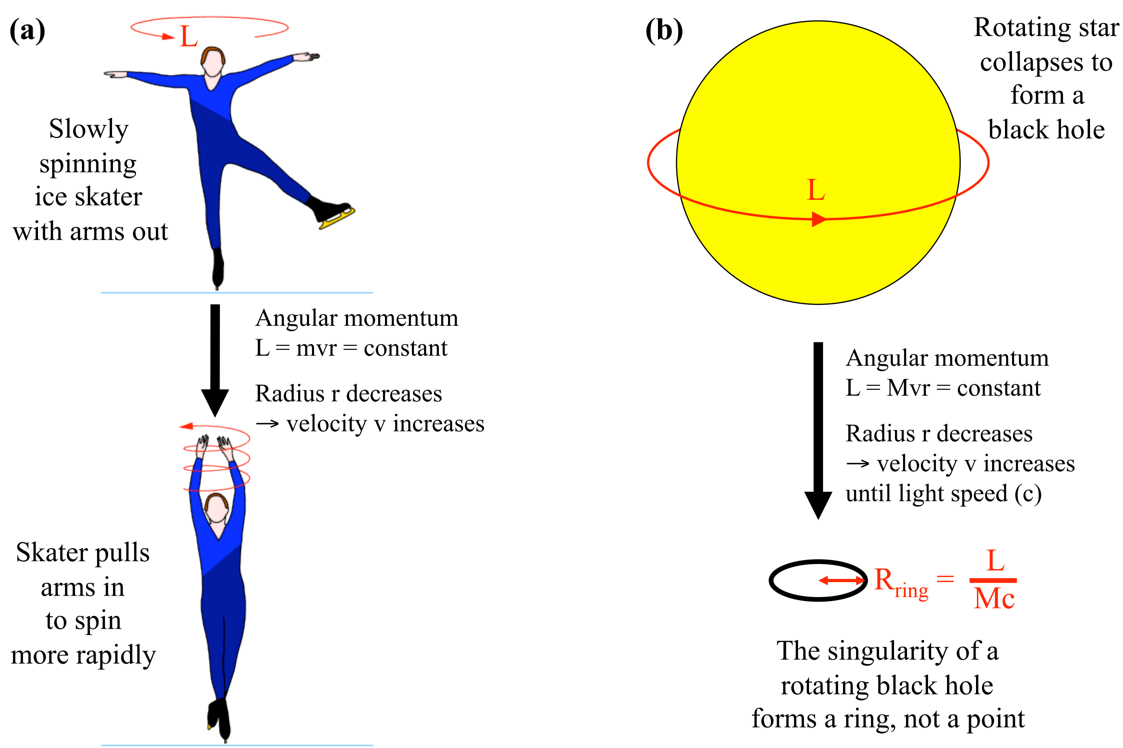

Figure 21. Second effect of rotating black holes: ring singularity. (a) Conservation of angular momentum $L = mvr$ means that if slowly spinning ice skaters pull in their arms (radius r decreases), their velocity v will increase and they will spin more rapidly. **(b)** Likewise, if a rotating star with angular momentum $L = Mvr$ collapses to form a black hole (radius r decreases), its velocity v will increase. However, since v cannot exceed the speed of light c, r cannot decrease to form a point singularity as occurs in a non-spinning black hole. The minimum value for the radius of the singularity occurs when $v = c$, $R_{\text{ring}} = L/Mc$. Thus the singularity forms a ring with this radius.

C. The third effect of rotation is that the one event horizon of a nonrotating black hole becomes two separate inner and outer event horizons for a rotating black hole (Fig. 22). Lots of scary math [7, 8] yields the radii $R_{\text{event},+}$ and $R_{\text{event},-}$ of the outer and inner event horizons:

$$R_{\text{event},\pm} = \frac{R_s}{2} \pm \sqrt{\left(\frac{R_s}{2}\right)^2 - R_{\text{ring}}^2} \tag{112}$$

The **ergosurface** indicates the distance from the center where the frame dragging becomes so strong that even light can only orbit in the direction of the black hole's rotation. Just as there are two event horizons, there are two ergosurfaces for rotating black holes, and both are oblate spheroids. More scary math [7, 8] yields their radii:

$$R_{\text{ergo},\pm} = \frac{R_s}{2} \pm \sqrt{\left(\frac{R_s}{2}\right)^2 - R_{\text{ring}}^2 \cos^2\theta} \tag{113}$$

The **Kerr metric** is the solution for the space-time around a rotating black hole. For no rotation, $R_{\text{ring}} = 0$, the outer radii in Eqs. (112) and (113) become the Schwarzschild radius R_s, the inner radii disappear into the singularity at $r = 0$, and the Kerr metric becomes the Schwarzschild metric. If a black hole has enough angular momentum, $L/mc = R_{\text{ring}} > R_s/2$, the event horizon radii in Eq. (112) do not have real solutions and the event horizons vanish, permitting outside observers to freely see and explore the **naked singularity**, unless currently poorly understood physical principles impose **cosmic censorship** that prohibits the formation of naked singularities.

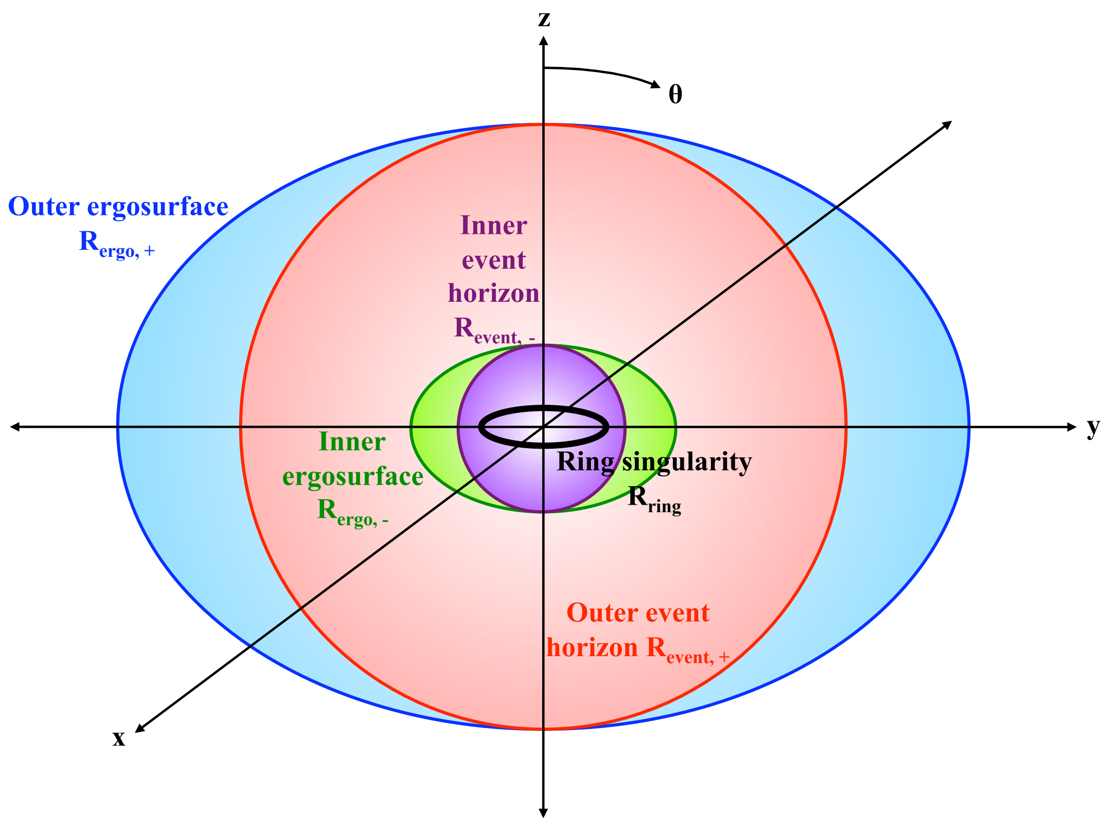

Figure 22. Third effect of rotating black holes: inner and outer event horizons and ergosurfaces. Whereas a nonrotating black hole only has one event horizon, a rotating black hole has two event horizons as well as two ergosurfaces, within which the frame dragging becomes so strong that even light can only orbit in the direction of the black hole's rotation.

D. The fourth effect is that large amounts of energy can be extracted from rotating black holes. Figure 23 shows a simple example that was first proposed by Roger Penrose. A spacecraft could dive into the outer ergosphere of a rotating black hole, jettison its trash into the black hole, and escape out of the ergosphere. Because a rotating black hole tilts light cones in the ergosphere in the direction of rotation and energy continues to decrease the deeper one goes into a gravitational well, if the trash enters the event horizon of the black hole with an angular momentum opposite that of the black hole, the trash can actually have a negative relativistic energy. Adding a negative energy to the black hole is equivalent to extracting energy from the black hole. By conservation of energy and angular momentum, the net energy (mass) and angular momentum of the black hole decrease, and the energy and angular momentum of the spacecraft increase by the same amounts. In theory, this process could be repeated to extract all of the rotational energy from a Kerr black hole until it became a nonrotating Schwarzschild black hole.

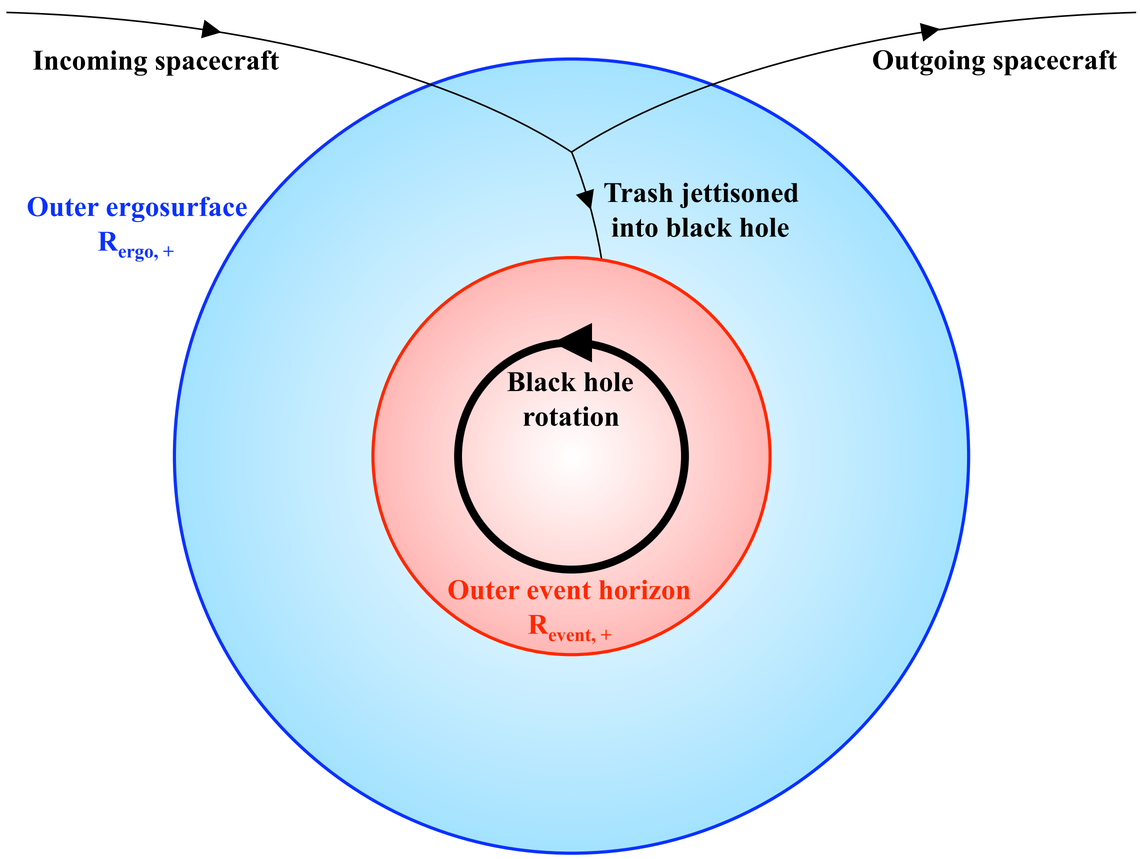

Figure 23. Fourth effect of rotating black holes: energy can be extracted from a rotating black hole, as shown in the top view of a Penrose process. A spacecraft can enter the outer ergosphere of a rotating black hole, jettison its trash into the event horizon opposite the direction of the black hole's rotation, and leave the ergosphere with more energy than it originally had, thereby extracting some of the rotational energy from the black hole.

In addition to extracting energy already stored in a rotating black hole, one can extract a significant fraction of the rest energy of an object approaching a black hole, as may be seen by a simple classical calculation even without special and general relativistic corrections. The total energy E of an object of mass m in a circular orbit at velocity $v = \sqrt{GM/r}$ from Eq. (62) and distance r from a black hole is:

$$\begin{aligned} E &= mc^2 + \frac{1}{2}mv^2 - \frac{GMm}{r} \\ &= mc^2\left(1 - \frac{1}{4}\frac{R_s}{r}\right) \end{aligned} \qquad (114)$$

Equation (114) predicts that an object can come from far away and lose up to 8% of its rest energy (for example by radiation or collisions) before settling into the closest stable orbit around a nonrotating black hole, $r = 3R_s$ (see [7]-[11] for the details of orbits around black holes). More detailed relativistic calculations slightly lower that value to 6%, but that is still much larger than the $\sim 0.7\%$ of rest mass that is converted to energy by the fusion of hydrogen to helium-4 in stars.

However, for a black hole rotating so fast that it is on the verge of having a naked singularity, an object can stably orbit in the direction of the black hole's rotation just outside the outer event horizon at $r = R_s/2$ from Eq. (112). Equation (114) predicts that an object can come from far away and lose up to 50% of its rest energy before settling into this closest stable orbit. Relativistic corrections slightly lower that value to 42%, or seven times more energy than can be derived from an object approaching a nonrotating black hole. Most black holes are probably rotating more slowly than this maximal case, so your actual mileage may vary, but the bottom line is that matter falling into black holes can release ∼9-60 times more energy than the fusion reactions that power stars.

This enormous energy source is believed to be the secret of **active galactic nuclei**, giant rotating black holes that lie at the centers of galaxies and distant quasars and radiate vast amounts of power, especially in polar jets perpendicular to the plane of rotation (Fig. 19). Matter spiraling inward to the black hole forms an accretion disk and is accelerated, compressed, and heated until it emits X-rays. Although the black hole itself has no magnetic field, it is surrounded by intense magnetic fields, and the rapid rotation of the black hole converts a significant fraction of the infalling particles' rest energy to energy radiated in the polar jets, in somewhat the same way that rotating magnetic coils can produce electrical power in a generator.

Stephen Hawking showed that black holes can only have three distinguishing features visible to the outside world: their mass M, angular momentum L, and electric charge Q. Since black holes have no other visible features, this finding is often whimsically summarized as "black holes have no hair." We have already covered the Schwarzschild and Kerr metrics for uncharged nonrotating and rotating black holes. The space-time solution for an electrically charged nonrotating black hole is called the **Reissner-Nordstrom metric**, and the space-time solution for a charged rotating black hole is called the **Kerr-Newman metric.** Solutions for charged black holes are not as physically relevant as those for uncharged black holes, since a charged black hole would rapidly suck in particles of the opposite charge until it was approximately neutral. For more information on the solutions for rotating and/or charged black holes, see [7]-[13].

5.5 Wormholes

If you fell into a black hole, you would die in half a dozen different ways. However, with sufficiently advanced technology it might be possible "tame" a black hole to form a wormhole through which spacecraft could safely pass. This section will explain each phenomenon that makes a black hole fatal, then describe theoretical solutions that could eliminate each problem to make a safe wormhole. For much more detailed information, see [13] and references cited therein.

Problem 1. A black hole does not lead anywhere.

Solution 1. Two black holes could be connected to form a tunnel or wormhole, as shown in Fig. 24. The wormhole may connect one part of our universe to a different part of our universe, and the distance through the wormhole may be much shorter than the distance through normal space. Alternatively, the wormhole may connect two different times together, or it may connect our universe to another universe.

John Wheeler has suggested that microscopic wormholes may appear and disappear in the "quantum foam" of space-time, so in theory one might nurture one of those to create a larger, long-lived wormhole. Another possible solution is that quantum gravitational processes may permit two previously unconnected black holes to link together to create a wormhole.

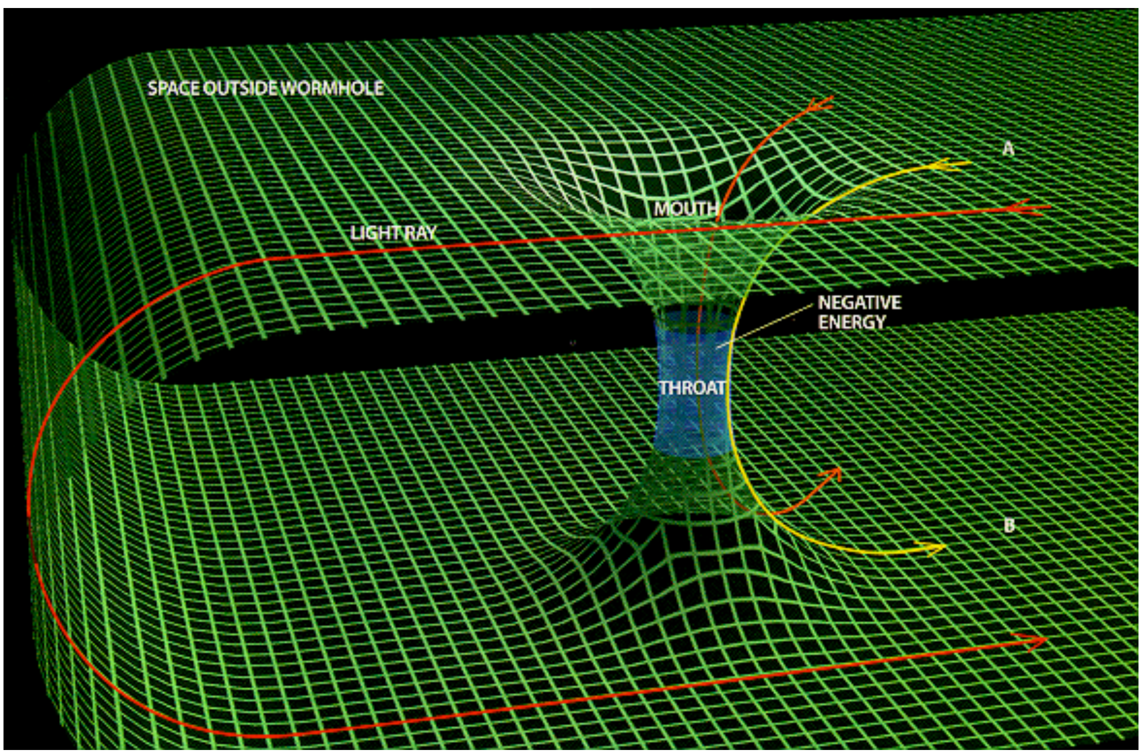

Figure 24. Wormhole. A wormhole is basically two black holes that connect together to form a "subway" tube. The wormhole may connect one part of our universe to a different part of our universe, and the distance through the wormhole may be much shorter than the distance through normal space. Alternatively, the wormhole may connect two different times together, or it may connect our universe to another universe. Due to gravitational attraction, the throat of the wormhole will pinch off, dividing the wormhole into two unconnected black holes, unless the throat is actively propped open by negative energy, which would exert gravitational repulsion on the walls of the wormhole.

Problem 2. Your time slows down as you approach the event horizon, and the universe ends before you finally fall in. Even if you could survive the wormhole, there would be nothing to which you could return.

Solution 2. The event horizon could be eliminated by using the proper distribution of mass/energy to create the wormhole, instead of having all the mass at the center.

Problem 3. As shown in Fig. 25, electromagnetic waves falling into a black hole gets blue-shifted to become very high-energy radiation. Starlight gets blue-shifted to higher and higher energies as it falls into the gravitational field. Even weak radio waves or infrared light becomes intense gamma rays as it approaches the event horizon. Anyone near or inside the black hole would be fried by the intense incoming radiation.

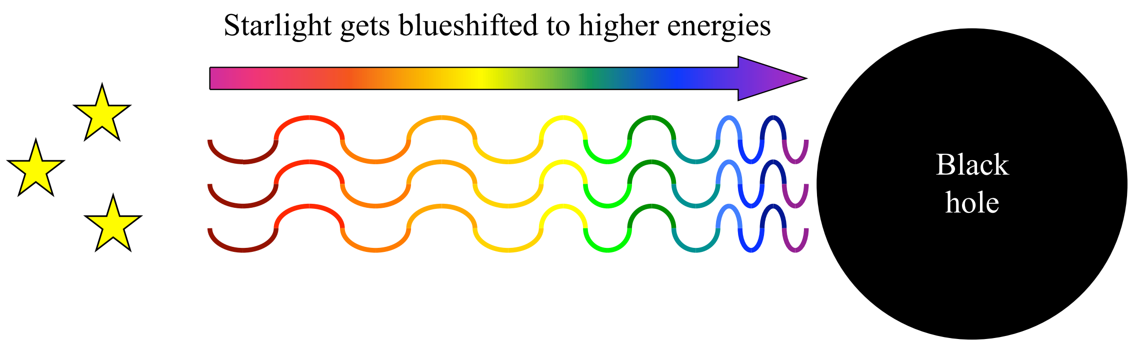

Figure 25. Electromagnetic waves falling into a black hole gets blue-shifted to become very high-energy radiation. Starlight gets blue-shifted to higher and higher energies as it falls into the gravitational field. Even weak radio waves or infrared light becomes intense gamma rays as it approaches the event horizon. Anyone near or inside the black hole would be fried by the intense incoming radiation.

Solution 3. The event horizon and the accompanying blue-shifting of infalling radiation could be eliminated by using the proper distribution of mass/energy to create the wormhole, as in Solution 2.

Problem 4. Tidal forces rip you apart long before you reach the center, as shown in Fig. 26. The gravitational field increases (there is a gravitational gradient) as one approaches a black hole. Sufficiently close to the black hole, the gravitational field can be much stronger in one place even than the field a meter or two further away from the black hole. These tidal forces stretch any object falling into the black hole until they tear it apart. Smaller black holes pack their gravitational fields into a smaller space than larger black holes, so the tidal forces are much worse near a small black hole than at the same distance from a large black hole.

Using the Newtonian approximation $a = GM/r^2$ for simplicity, the difference in gravitational acceleration across an object of width Δr_m in meters is

$$\Delta a \;=\; \Delta r_m \left|\frac{da}{dr}\right| \;=\; 2\Delta r_m \,\frac{GM}{r^3} \tag{115}$$

From Eqs. (97) and (115), the tidal acceleration difference at the Schwarzschild radius is

$$\Delta a \;=\; 1.03\times 10^{10}\left(\frac{M_\odot}{M}\right)^2 \Delta r_m \,\frac{\text{m}}{\text{sec}^2} \;=\; 1.05\times 10^{9}\left(\frac{M_\odot}{M}\right)^2 \Delta r_m \text{ g}\,, \tag{116}$$

where g ≈ 9.807 m/sec^2 is the gravitational acceleration at the earth's surface. Thus the difference in acceleration between the head and feet of a person ($\Delta r_m \approx 2$ m) passing the Schwarzschild radius of a black hole of one solar mass is 2×10^9 g. A person would be turned into spaghetti long before they reached the Schwarzschild radius.

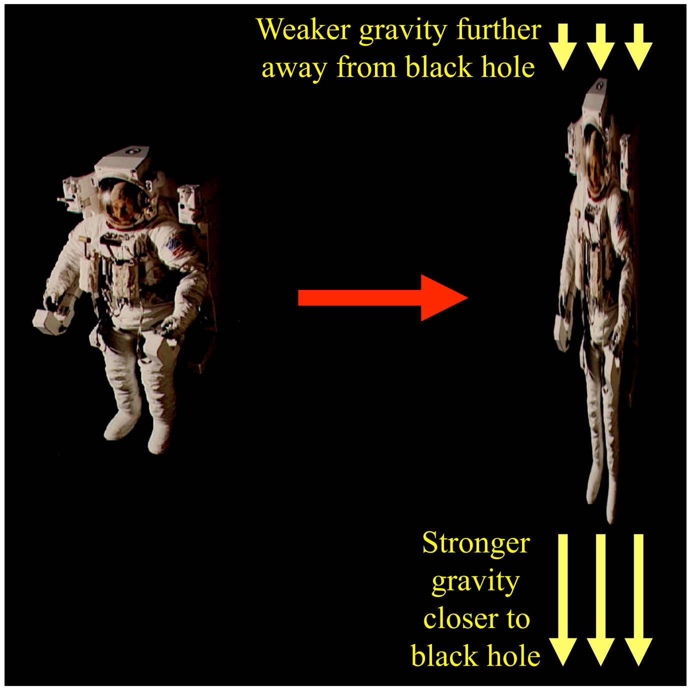

Figure 26. Tidal forces near a black hole. The gravitational field increases (there is a gravitational gradient) as one approaches a black hole. Sufficiently close to the black hole, the gravitational field can be much stronger in one place even than the field a meter or two further away from the black hole. These tidal forces stretch any object falling into the black hole until they tear it apart. Smaller black holes pack their gravitational fields into a smaller space than larger black holes, so the tidal forces are much worse near a small black hole than at the same distance from a large black hole.

Solution 4. The larger the hole is, the smaller the gravity gradients and tidal forces will be. For example, in Eq. (116), if the hole has 46,000 solar masses, the acceleration difference on a person will be less than 1 g at the Schwarzschild radius. Distributing the mass/energy creating the wormhole instead of concentrating it all at the center could also greatly help, just as it might help to solve problems 2 and 3.

Problem 5. You cannot avoid hitting the singularity at the center.

Solution 5. While section 5.2 showed that it is impossible to avoid hitting the singularity of a nonrotating black hole, section 5.4 showed that the singularity of a rotating black hole is a ring. Thus you can travel from the outside world to the center of the hole without hitting the ring singularity, especially if your trajectory is not in the plane of the ring. Alternatively, wormholes might have the correct distribution of mass/energy to eliminate the singularity entirely. Another possibility is that the singularity might exist only as the final state of the wormhole in the far future, so you can avoid it as long as you do not remain inside the wormhole too long.

Problem 6. Due to the gravitational attraction of itself, or of anything passing through it, the throat of the wormhole will pinch off, dividing the wormhole into two unconnected black holes before anything can make it through.

Solution 6. Once could use the gravitational repulsion (antigravity) of negative energy to repel the walls of the wormhole and hold the wormhole's throat open, as shown in Fig. 24. Fig. 27 illustrates one way to create such negative energy. In quantum vacuum fluctuations, photons, electrons, positrons, or other particles can appear from nowhere, then disappear again. The Casimir effect is a way to create negative energy, or less energy than the vacuum. Whereas any electromagnetic wave can appear in the vacuum, only waves meeting the boundary conditions can appear between two parallel metal plates. In other words, if the vacuum contains the usual amount of nothing, then there is less than nothing between the plates. The pressure of the usual amount of nothing in the vacuum will overcome the pressure of the less than nothing between the plates and push the plates together. Other ways to create negative energy include squeezed light, moving mirrors, etc. [13].

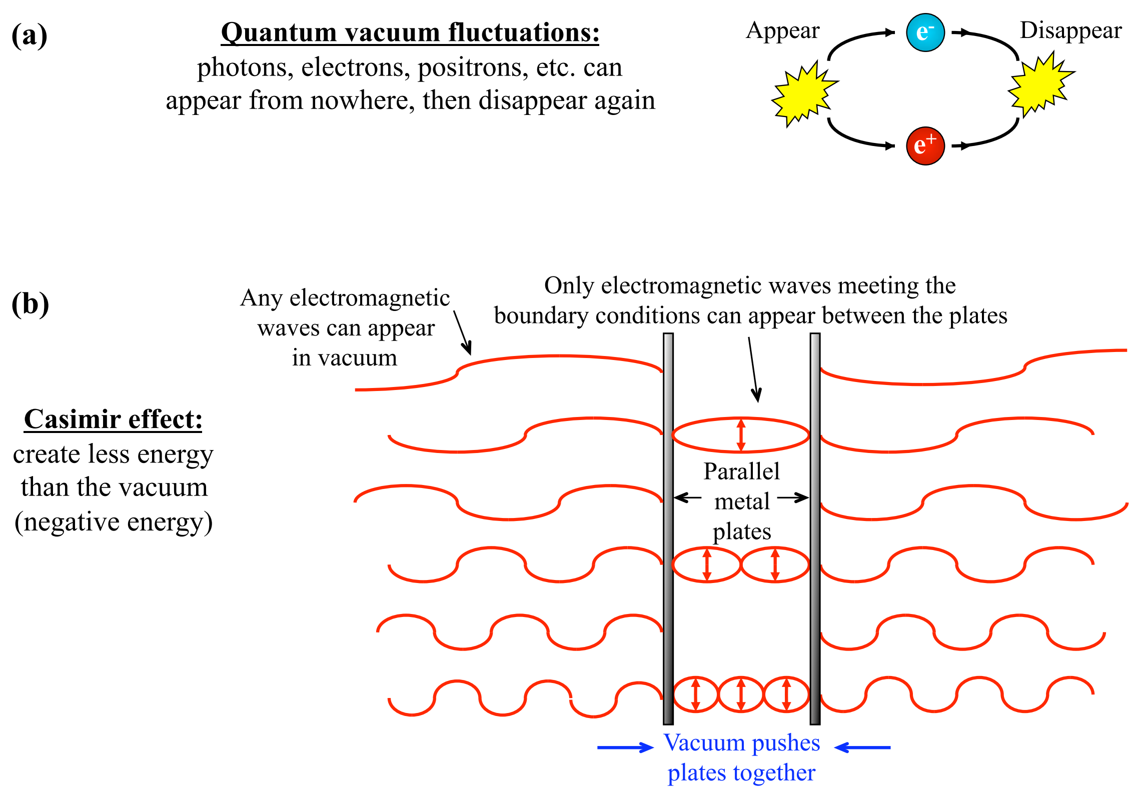

Figure 27. An example source of negative energy to hold wormholes open. (a) In quantum vacuum fluctuations, photons, electrons, positrons, or other particles can appear from nowhere, then disappear again. **(b)** The Casimir effect is one way to create negative energy, or less energy than the vacuum. Whereas any electromagnetic wave can appear in the vacuum, only waves meeting the boundary conditions can appear between two parallel metal plates. In other words, if the vacuum contains the usual amount of nothing, then there is less than nothing between the plates. The pressure of the usual amount of nothing in the vacuum will overcome the pressure of the less than nothing between the plates and push the plates together.

6 Quantum Effects in General Relativity

Quantum effects can cause interesting modifications to general relativity, some of which can be predicted even without a complete theory of quantum gravity. In this section we will consider three effects: Hawking radiation from black holes, the Planck scale, and the cosmological constant or vacuum energy. For more information on quantum gravity and its theoretical difficulties, see *Relativistic Quantum Field Theory* 5.

6.1 Hawking Radiation from Black Holes

Hawking radiation is a quantum mechanism by which matter or energy can actually escape from black holes. According to quantum theory, pairs of virtual particles and antiparticles continually appear from the vacuum and then promptly annihilate with each other and disappear. As shown in Fig. 28, Hawking radiation occurs when such a particle-antiparticle pair appear just outside a black hole's event horizon and one member of the pair crosses the event horizon and becomes trapped in the black hole; without a companion with which to annihilate, the other particle (or antiparticle) freely continues to exist and appears to have been "emitted" by the black hole.

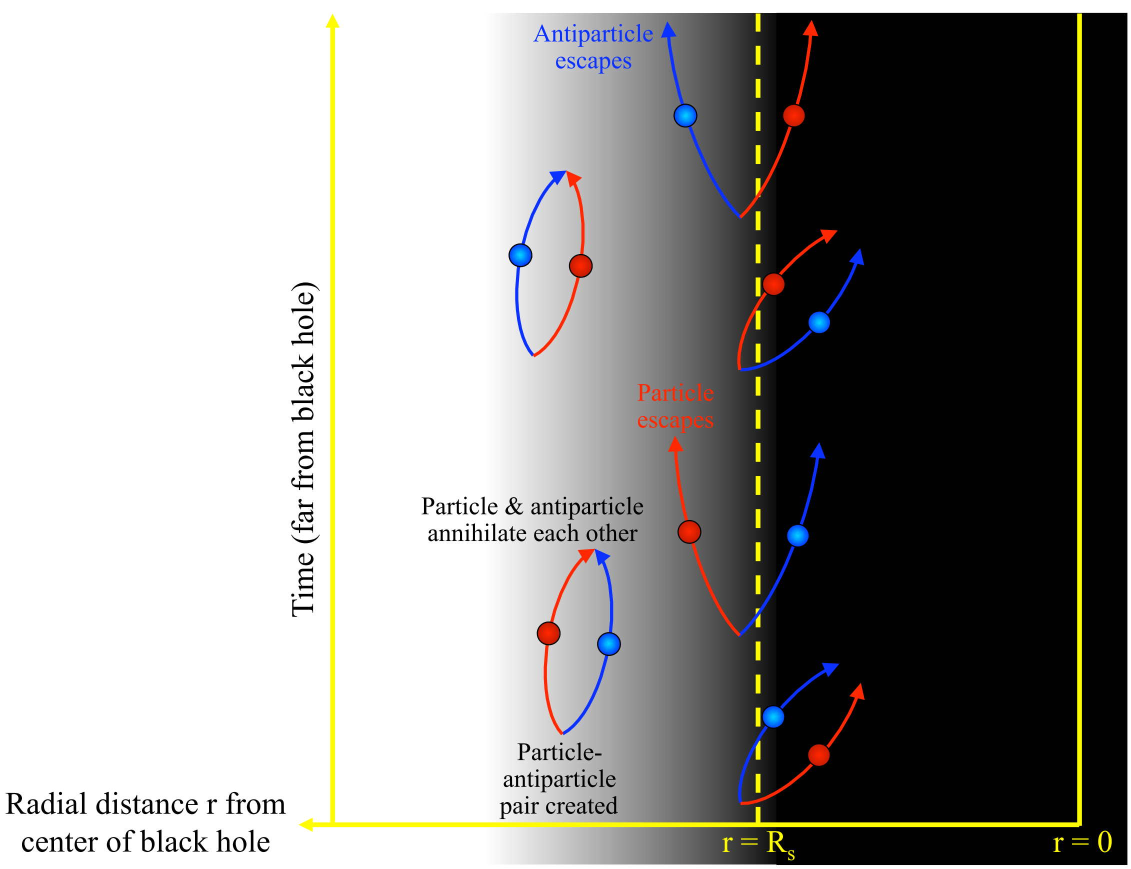

Figure 28. Hawking radiation. In empty space, particle-antiparticle pairs continually appear and then annihilate each other. If a pair appears near the event horizon of a black hole, one member may fall into the black hole, allowing its partner to escape annihilation. To outside observers, the escaping particle or antiparticle appears to have been emitted by the black hole, and the mass of the black hole is reduced accordingly.

Specifically, consider the pair creation of two photons, one with four-momentum $(+pc, -\mathbf{p})$ and the other with $(-pc, +\mathbf{p})$; the net four-momentum of the two photons is thus zero. The first photon has the correct positive energy appropriate for its momentum and could therefore continue to exist indefinitely if it did not annihilate with the other photon. However, the second photon's negative energy prevents it from being a permanent "real" particle, and so it can only exist for a time $\Delta t \sim \hbar/pc$. If during that brief time it crosses the black hole's event horizon, it will enter a region where spacelike and timelike vectors are interchanged; this converts the photon's previously unacceptable negative energy $-pc$ into an acceptable momentum, and its previous momentum into an acceptable positive energy. Therefore the photon becomes a real particle inside the black hole. When the photon was outside the event horizon, it had a negative energy; by crossing the event horizon and adding this negative energy to the black hole, energy has effectively been removed from the black hole. This energy compensates for the energy needed to create the other (positive-energy) photon and ensures that energy is conserved in the long run.

The emission of photons makes the black hole appear to be radiating like a black body with a certain temperature, which may be calculated. Because photons may be emitted from any part of the event horizon, whose dimensions are described by the Schwarzschild radius R_s, the momentum of emitted photons is uncertain by an amount Δp,

$$\Delta p \;\sim\; \frac{\hbar}{R_s} \;=\; \frac{\hbar}{2GM/c^2}\ . \tag{117}$$

This uncertainty in momentum (or photon energy Δpc) can be characterized in terms of thermal fluctuations at a temperature T (with k_B defined as Boltzmann's constant):

$$\Delta p \;\approx\; \frac{k_B T}{c}\ . \tag{118}$$

Equating these two expressions for Δp and solving for T yields

$$T \;\sim\; \frac{\hbar c^3}{2k_B GM}\ . \tag{119}$$

This result is within a factor of 4π of the result of Hawking's much more rigorous derivation,

$$T \;=\; \frac{\hbar c^3}{8\pi k_B GM} \qquad \textbf{Hawking temperature of a black hole's event horizon} \tag{120}$$

Due to the black-body-like Hawking radiation, a black hole loses energy at a rate

$$\frac{d(Mc^2)}{dt} \;=\; \left(4\pi R_s^2\right)\left(\sigma_{SB} T^4\right) \;\propto\; M^{-2}\ , \tag{121}$$

where $\sigma_{SB} \equiv (2\pi^5 k_B^4)/(15h^3c^2) \approx 5.67\times10^{-8}$ W/(m^2 ^{o}K^4) is the Stefan-Boltzmann constant. Thus smaller black holes have higher temperatures and lose energy more rapidly than larger ones. Very small black holes should completely evaporate in one final explosion of Hawking radiation, unless other presently unknown quantum gravity effects intervene.

Integrating Eq. (121) and rearranging, the time for a black hole of initial mass M_0 to evaporate via Hawking radiation is:

$$\tau_{\text{evaporation}} \;=\; \frac{5120\pi G^2}{\hbar c^4}\, M_0^3 \;\approx\; 10\left(\frac{M_0}{10^{11}\ \text{kg}}\right)^3 \text{ billion years} \tag{122}$$

Thus a small black hole of $\sim 10^{11}$ kg or less created early in the 13.7-billion-year history of the universe would evaporate by now, but more typical black holes of stellar mass would require far longer than the life of the universe to evaporate.

6.2 Planck Scale

The point at which general relativity and quantum physics collide head-on is defined by the Planck scale. Estimated values for the Planck scale can be found via the following hogwash argument.

A particle-like object of mass M creates gravitational effects which come with an associated length scale, the Schwarzschild radius: $R_s = 2GM/c^2$. Similarly, such an object exhibits quantum effects on a length scale of the Compton length, $r_C = \hbar/Mc$.

Quantum effects dominate over gravitational effects when $r_C \gg R_s$, as it is for elementary particles. Gravitational effects dominate over quantum effects when $R_s \gg r_C$, as is the case for macroscopic astronomical objects. However, both gravitational and quantum effects must be taken into consideration when $R_s \sim r_C$, or equivalently when

$$M \sim M_P \equiv \sqrt{\hbar c/G} \approx 2.18 \cdot 10^{-8}\ \text{kg} \quad \textbf{Planck mass} \tag{123}$$

Related quantities may also be defined:

$$E_P \equiv M_P c^2 \approx 1.22 \cdot 10^{19}\ \text{GeV} \quad \textbf{Planck energy} \tag{124}$$

$$L_P \equiv \hbar/M_P c \approx 1.62 \cdot 10^{-35}\ \text{m} \quad \textbf{Planck length} \tag{125}$$

$$T_P \equiv L_P/c \approx 5.39 \cdot 10^{-44}\ \text{sec} \quad \textbf{Planck time} \tag{126}$$

Note that the Planck scale is far beyond anything which humans can presently achieve. The Planck mass is far greater than the masses of any known particles, the Planck energy is far higher than conceivable particle accelerators can attain, and the Planck length and time scales are far smaller than can currently be resolved. But if you could ever reach the Planck scale, you would need a full-fledged theory of quantum gravity to describe things. Unfortunately that doesn't seem to be coming anytime soon. See see *Relativistic Quantum Field Theory* 5 for more details on the headaches of quantum gravity theories.

At present, the best guess of what happens at the Planck scale is that there are chaotic quantum fluctuations in space-time, such that the very notion of measurable space and time coordinates breaks down. When viewed at this very fine scale, the smooth "rubber sheet" of space-time may actually be a quantum foam of space-time bubbles and microscopic wormholes continually appearing, rearranging, and disappearing.

6.3 Vacuum Energy and Cosmological Constant

The most general form of Einstein's equation includes a cosmological constant Λ:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \; . \tag{127}$$

In theory a nonzero cosmological constant could be caused by quantum vacuum energy effects. The cosmological constant appears to be zero now, but in the very early universe it may have been nonzero and caused an exponential inflation of the size of the universe, as will be discussed later.

The cosmological constant can be absorbed into the stress-energy tensor by introducing terms for the quantum vacuum energy density and pressure [13]:

$$T_{\mu\nu}^{\text{total}} = T_{\mu\nu}^{\text{standard}} + T_{\mu\nu}^{\text{vacuum}} \tag{128}$$

in which

$$T_{\mu\nu}^{\text{vacuum}} \equiv -\frac{c^2\Lambda}{8\pi G} g_{\mu\nu} = \begin{pmatrix} \frac{c^4\Lambda}{8\pi G} & 0 & 0 & 0 \\ 0 & -\frac{c^4\Lambda}{8\pi G} & 0 & 0 \\ 0 & 0 & -\frac{c^4\Lambda}{8\pi G} & 0 \\ 0 & 0 & 0 & -\frac{c^4\Lambda}{8\pi G} \end{pmatrix}_{\text{rectangular, flat space}} \tag{129}$$

Thus the quantum vacuum effects modeled by the cosmological constant act like a gas with a positive energy density,

$$\rho_{\Lambda} c^2 = c^4 \Lambda / 8\pi G \; , \tag{130}$$

(assuming $\Lambda > 0$), but a *negative* pressure,

$$p_{\Lambda} = -c^4 \Lambda / 8\pi G \; . \tag{131}$$

The vacuum energy density is determined by fundamental quantum physics considerations and cannot change if the vacuum energy's volume is changed. The negative pressure, while sounding strange, ensures that if the volume occupied by the vacuum energy increases by an amount dV, the associated work $dW = -p\, dV$ will be sufficient to provide the energy needed to *create* more vacuum energy to fill the new volume.

7 Cosmology

One of the major applications of general relativity lies in the field of cosmology [14]-[17]. General relativity may be used to derive equations modeling the structure and history of the universe.

7.1 Observational Data about the Universe

There are several observed characteristics that must be part of any realistic model of the universe, either as fundamental assumptions on which the model is founded or as theoretical predictions of the model that support its validity.

The composition of the universe at present appears to be ∼5% matter, ∼23% dark matter, ∼0.006% radiative energy, and ∼72% dark energy, although the relative amounts and importance of these components has changed throughout the history of the universe:

- **Matter** is ordinary particles like protons, neutrons, and electrons, a few oddball particle types like neutrinos (see *Relativistic Quantum Field Theory* for more information), and black holes into which matter has collapsed. Most of the matter is in the form of stars, which are generally surrounded by much smaller amounts of matter in the form of planets and such. Stars are grouped into galaxies. An average galaxy has $\sim 10^{11}$ stars, and there are at least $\sim 10^{11}$ galaxies in the observable universe. Not counting heavier elements that have subsequently been produced in stars, matter dating from the origin of the universe is ∼ 75% hydrogen and ∼25% helium-4 by mass, as will be explained by **Big Bang nucleosynthesis** in Section 7.5.

- **Dark matter** exerts a gravitational attraction like normal matter yet appears to be composed of something other than the known particles, is very difficult to detect directly, and thus is currently poorly understood. Evidence for dark matter includes: (1) The rotation velocities of many galaxies indicate a gravitational pull from ∼5x more matter in those galaxies than all their visible stars plus the expected number of planets, white dwarfs, neutron stars, and black holes. (2) Gravitational bending of light by some galaxy clusters is also ∼5x greater than should be caused by all the visible stars plus the expected number of planets and burned out stars.

- **Radiative energy** is generally electromagnetic radiation or photons, spanning the spectrum from radio waves to visible light to X-rays, and played a much more important role in the early history of the universe, when the universe was much smaller and hotter. Most of the radiative energy is in the form of **cosmic microwave background radiation**, which began as high-temperature thermal radiation in the earliest phases of the universe and has subsequently cooled to the microwave regime, as shown in Fig. 29.

- **Dark energy**, also known as vacuum energy or the cosmological constant (Section 6.3), is a form of energy that permeates all space in the universe and exerts an antigravitational effect, trying to push the universe apart instead of pull it together. Dark energy may arise from various quantum effects, none of which are well understood, and in fact multiple forms of dark energy may have played different roles at different times in the history of the universe.

The cosmological principle is the assumption that the universe is homogenous and isotropic – it basically looks the same everywhere and in all directions. It has the virtue of simplicity, and it is supported by astronomical observations such as the isotropy of the cosmic microwave background radiation.

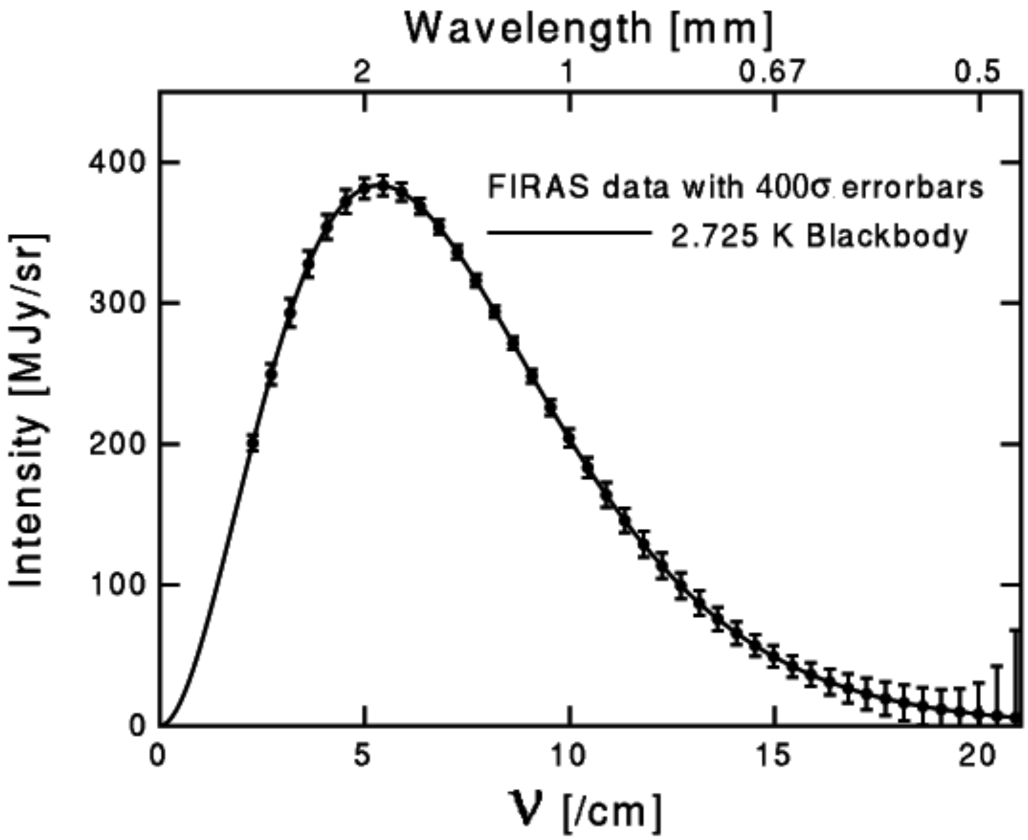

Figure 29. The measured spectrum of cosmic microwave background radiation corresponds almost perfectly to black-body radiation with a temperature of 2.7^{o}K. The actual error bars are so small that they have been multiplied by a factor of 400 just to make them visible on the graph. [Data from COsmic Background Explorer (COBE) satellite's Far InfraRed Absolute Spectrophotometer (FIRAS) instrument.]

Distances to nearby stars can be measured via parallax; the angle from earth to a star changes as the earth moves from one side of the sun to the other, just as our two eyes separated by a certain distance can judge distances to objects. To measure distances so great that parallax is negligible, astronomers use the apparent brightness of stars of known intensity, since the apparent brightness is proportional to (distance)$^{-2}$. As examples, the pulsation rate of Cepheid variable stars depends on their intensity in a known fashion, type Ia supernovae all have the same intensity because they occur when white dwarfs acquire extra matter and exceed a specific critical mass, and Hertzsprung-Russell diagrams (*Plasma Physics and Fusion* 5.5) plotting the intensities and peak emission wavelengths of all the stars in each galaxy fall along the same known curve. The nearest stars to our sun are several light-years away, our galaxy (the Milky Way) is $\sim 100,000$ light-years in diameter, other galaxies range from a few million to several billion light-years away from ours, and the furthest detectable objects are ~ 13 billion light-years away.

Ages of stars can be estimated by their luminosities, surface temperatures, and composition (via spectroscopy), which indicate their consumption rate and how much of their hydrogen they have consumed to form helium and heavier elements via nuclear fusion (see *Plasma Physics and Fusion* 5 for more information). The ages of stellar remnants such as white dwarfs can be estimated from their surface temperature and cooling rate. Our sun formed approximately 4.5 billion years ago and is roughly halfway through its life cycle. The oldest detectable stars and stellar remnants in the universe are $\sim$ 13 billion years old. This agrees with the furthest visible objects in the universe being $\sim$ 13 billion light-years away, since light would have left them $\sim$ 13 billion years ago to reach us on earth now. It also agrees well with the age of the universe from Hubble's law (below).

Velocities of stars relative to the earth can be measured by Doppler shifts of known spectral lines for hydrogen, helium, and other elements in stars. Stars moving toward us compress their emitted electromagnetic waves and thus appear blue-shifted, whereas stars moving away from us stretch out their emitted electromagnetic waves and appear red-shifted. From *Electromagnetism* ?.?, the Doppler shift of a star moving away from us at velocity $v \ll c$ will lengthen its emitted wavelengths λ by an amount $\Delta\lambda$, often denoted in cosmology by the ratio z:

$$z \equiv \frac{\Delta\lambda}{\lambda} = \frac{v}{c} \,. \tag{132}$$

Special relativity would introduce corrections to Eq. (132) for velocities approaching the speed of light, but those corrections can be safely ignored for most velocities measured in cosmology. Additional corrections are also introduced if the speed of a star relative to our location has changed over the history of the universe, but those will also be ignored here for simplicity.

Hubble's law. All galaxies (or clusters of galaxies) in the universe appear to be moving away from each other with a velocity v proportional to the distance d between the galaxies, $v = Hd$, where H is Hubble's constant, as shown in Fig. 30. (Hubble's constant is approximately constant right now, but in general it varies with time during the history of the universe, so we use H_0 to denote its current value.) This observation is known as Hubble's law and may be attributed to stretching of the cosmic scale factor R (a measure of the size of the universe):

$$H(t) = \frac{dR/dt}{R(t)} \,. \tag{133}$$

The best current estimate for the Hubble constant is

$$H_0 \approx 71 \, \frac{\text{km/sec}}{\text{Megaparsec}} \approx 2.3 \times 10^{-18} \text{ sec}^{-1} \,, \tag{134}$$

in which 1 Megaparsec $\approx 3.262 \times 10^6$ light-years $\approx 3.086 \times 10^{22}$ m.

Extrapolating backward in time from Hubble's law, the universe appears to have started in a **Big Bang**. Without knowing the details of how the universe has accelerated or decelerated during its past history, one can use the current expansion rate to make a rough estimate of the universe's age. Assuming constant expansion at the current rate, the current age t_0 of the universe is

$$t_0 = \frac{1}{H_0} \approx 4.35 \times 10^{17} \text{ sec} \approx 13.8 \text{ billion years} \tag{135}$$

As we will see in Section 7.5, this very simple estimate is almost exactly the correct answer. It also agrees well with the oldest stellar ages and furthest distances we can measure.

As shown in Fig. 30(b), the universe seems to be stretching uniformly everywhere, rather than expanding outward from one central point in three-dimensional space. General relativity is required to accurately model the uniform expansion of space-time, as will be done in Section 7.4. Yet even simple Newtonian calculations of three-dimensional expansion can yield correct answers if handled correctly, as will be shown in Section 7.3. Note that the red shift of a distant galaxy can be interpreted either as a Doppler shift due to that galaxy's relative velocity away from us, or as stretching of the wavelengths since empty space itself has stretched while the electromagnetic waves were in transit from that galaxy to us. The two views are equivalent, since the relative velocities of galaxies are caused by the stretching and expansion of the universe.

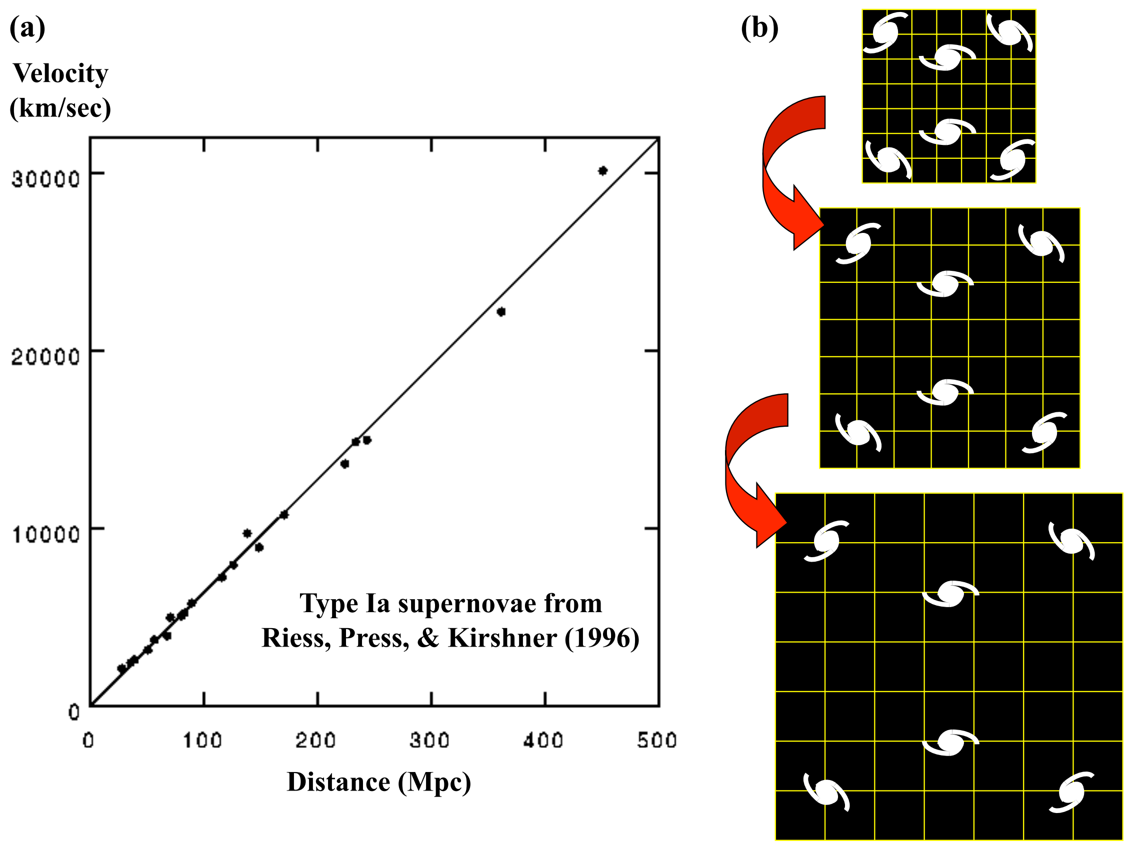

Figure 30. Hubble's law. (a) Numerous astronomical measurements indicate that galaxies (or clusters of galaxies) in all directions are moving away from our galaxy with velocities directly proportional to their distance from us–the further away they are, the faster they are receding. **(b)** Assuming that we are not by some wild coincidence located at the center of the universe, all parts of the universe should see this same effect. This means that the whole universe is "stretching" uniformly everywhere, rather than objects rushing outward from a central point in space.

7.2 Structure of the Universe

In deriving the structure of the universe from scratch, we will use the cosmological principle that the universe is homogenous and isotropic–it basically looks the same everywhere. In accordance with the cosmological principle, the large-scale curvature of space-time will be assumed to be the same at all spatial locations in the universe.

To further simplify matters, we will leave time out of things for now and just consider a snapshot of how the space-time structure of the universe looks at some fixed time t. As explained in Section 2.3, gravity can warp our three-dimensional space into an extra (fourth) spatial dimension w, sometimes called the embedding dimension since our three dimensions are embedded in that extra dimension that we cannot directly see. Due to the assumptions that the universe is isotropic, homogeneous, and has the same space-time curvature everywhere, we are led to consider the universe as a three-dimensional surface of revolution in four dimensions:

$$\frac{R^2}{k} \;=\; w^2 + x^2 + y^2 + z^2 \;=\; w^2 + r^2 \ . \tag{136}$$

in which $R^2(t)/k$ is the square of the radius of curvature of the universe. $R(t)$ is a scale factor which can vary during the history of the universe, and k, the sign of the squared radius of curvature, has three possible values, as illustrated in Fig. 31:

> $\mathbf{k = +1}$. For this choice of k, the curvature has a real value and Eq. (136) describes a three-dimensional spherical surface embedded in a four-dimensional Euclidean (all dimensions real, not imaginary) space. This is called a **closed universe**, since you can head off in what appears to be a straight line and eventually come back to the same position, circumnavigating the universe like a cosmic Magellan.
>
> $\mathbf{k = -1}$. For this choice, the radius of curvature has an imaginary value–our three real-valued spatial dimensions curve into an imaginary fourth embedding dimension w. Equation (136) becomes $r^2 + w^2 = -R^2$, or, written in a more revealing form, $-r^2 - w^2 = R^2$. If only r^2 were negative, this would be the equation for a three-dimensional hyperboloid surface embedded in a four-dimensional Euclidean space. But w^2 is also negative, which confirms that the embedding dimension is imaginary with respect to the three spatial dimensions. Thus $k = -1$ corresponds to a universe which is shaped like a three-dimensional hyperboloid surface embedded in four-dimensional Minkowski space. Note that there are now two imaginary dimensions running around (time and the spatial embedding dimension) and that each one can be combined with the normal three spatial dimensions to make a Minkowski space. Don't confuse the two. Anyway, although this type of universe has the same nonzero curvature everywhere, it curves in a funny way such that a "straight-line" trajectory never comes back to where it started, unlike the situation with $k = +1$. Therefore the case $k = -1$ is called an **open universe**.

k = 0 describes a **flat universe** whose space is purely Euclidean and not warped into an embedding dimension. Don't worry about the fact that $k = 0$ means that we're actually dividing by zero in some of the equations here. $k = 0$ just means that R^2/k is infinite, and the surface of a sphere of infinite radius is effectively a flat plane.

As will be discussed in Section 7.5, the flat $k = 0$ case appears to describe our universe. For generality, we will continue to consider cases in which k might assume the other values, but in practice we probably don't have to worry about our universe being curved in weird ways.

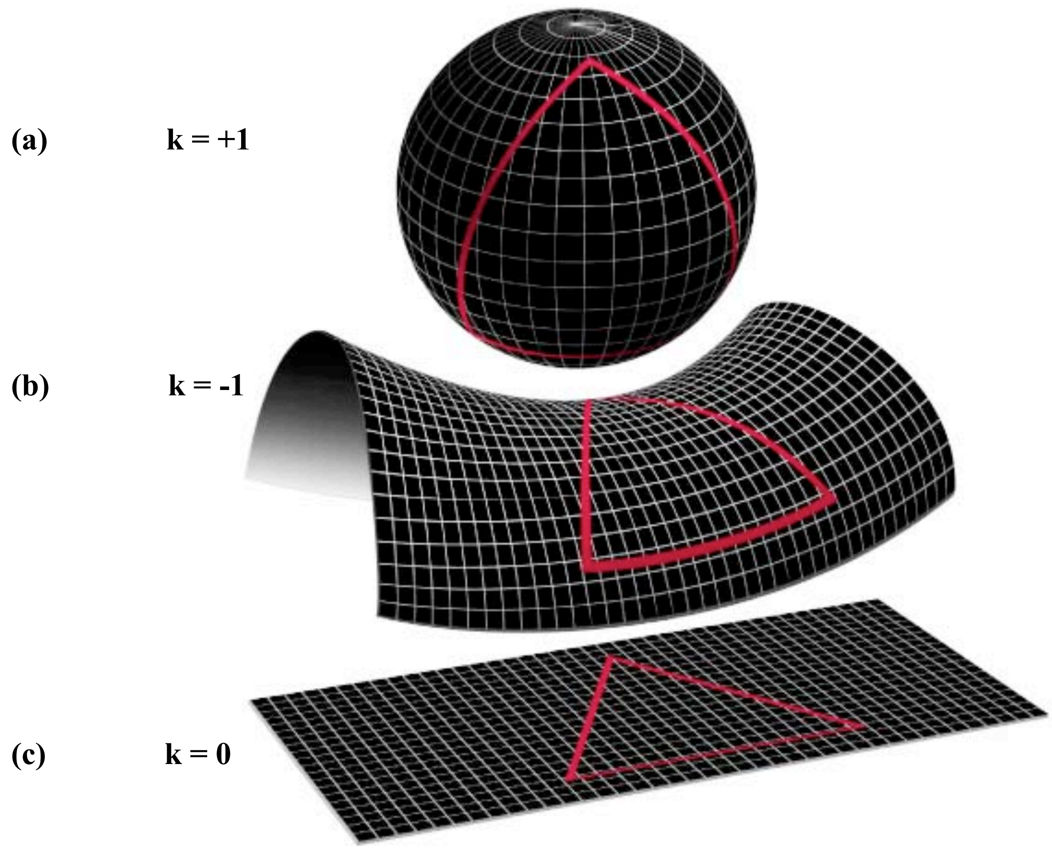

Figure 31. Four-dimensional shape of the universe for various values of k. (a) If $k = +1$, the universe is curved into a real fourth spatial dimension to form a closed sphere. Traveling in what seems like a straight line in three dimensions will eventually return you to your starting point. Due to the curvature, the sum of three angles in a triangle will actually be more than 180^o. **(b)** If $k = -1$, the universe is curved into an imaginary fourth spatial dimension in such a way that it never closes back on itself. This universe is uniformly curved everywhere, but because the mix of imaginary and real dimensions cannot be accurately depicted in real dimensions here, the $k = -1$ universe appears distorted into a saddle shape here. Due to the imaginary curvature, the sum of three angles in a triangle will be less than 180^o. **(c)** If $k = 0$, the universe is "flat" with no curvature into a fourth spatial dimension. The sum of three angles in a triangle is 180^o. (Public domain figure from http://upload.wikimedia.org/wikipedia/commons/9/98/End_of_universe.jpg.)

Eq. (136) can be used to calculate the space-time metric of the universe. Differentiating Eq. (136) while holding time (and thus $R(t)$) constant yields $w\ dw = -r\ dr$, since $dR = 0$. Squaring this result produces $w^2 dw^2 = r^2 dr^2$, which may be combined with Eq. (136) to obtain an expression for dw^2:

$$dw^2 = \frac{r^2 dr^2}{w^2} = \frac{r^2 dr^2}{R^2 - r^2} . \tag{137}$$

This expression for dw^2 may be substituted into the metric:

$$\begin{aligned} ds^2 &= -c^2 dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) + dw^2 \\ &= -c^2 dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) + \frac{r^2 dr^2}{\frac{R^2}{k} - r^2} \\ &= -c^2 dt^2 + \frac{R^2 dr^2}{R^2 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) . \end{aligned}$$

Changing to a rescaled variable $r' \equiv r/R$, this becomes the **Robertson-Walker metric** for a homogeneous, isotropic universe:

$$ds^2 = -c^2 dt^2 + [R(t)]^2 \left[\frac{dr'^2}{1 - kr'^2} + r'^2(d\theta^2 + \sin^2\theta d\phi^2) \right] \quad \textbf{Robertson-Walker metric} \tag{138}$$

Equivalently, the Robertson-Walker metric written in tensor form (using r' as the radial coordinate) is:

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & \left(\frac{R^2}{1-kr'^2}\right) & 0 & 0 \\ 0 & 0 & R^2 r'^2 & 0 \\ 0 & 0 & 0 & R^2 r'^2 \sin^2\theta \end{pmatrix}_{\text{spherical}} \tag{139}$$

$k = 0$ appears to describe our universe the best, in which case Eq. (139) reduces to the metric for flat space-time, except the three spatial dimensions can be expanded by a factor R without affecting the time dimension.

7.3 Time Development of the Universe–Helpful Hogwash Version

Before deriving the equations for the time development of the universe via rigorous (and mathematically complex) general relativistic methods, the same equations will first be derived by simple Newtonian arguments in order to gain some physical insight.

First energy conservation equation

As shown in Fig. 30(b), the universe seems to be stretching uniformly everywhere, rather than expanding outward from one central point in three-dimensional space. General relativity is required to accurately model the uniform expansion of space-time, as will be done in Section 7.4. Yet even simple Newtonian calculations of three-dimensional expansion can yield correct answers if handled properly. At a sufficiently macroscopic scale, the individual galaxies and other objects that make up the universe may for simplicity be treated as part of a continuum the same way that the particles in a fluid may be modeled as a continuum. This simplification is known as **Weyl's postulate**. Mathematically, this means that the universe's contents may be modeled as a gas with average density ρ and pressure p.

To begin the Newtonian calculations, we will consider a test particle of mass m at the "edge" of the universe, a distance R from the "center" of the universe. The particle's total energy E is the sum of its positive kinetic energy and negative gravitational potential energy:

$$E = \frac{1}{2}mv^2 - \frac{GMm}{R} , \tag{140}$$

The particle's velocity relative to the "center" of the universe is the same as the expansion velocity of the "edge" of the universe,

$$v = \frac{dR}{dt} \equiv \dot{R} . \tag{141}$$

To avoid clutter in equations, we will use overhead dots to denote time derivatives–one dot denotes the first derivative with respect to time, two dots denote the second derivative, etc.

The mass M attracting the particle is all of the matter closer to the "center" than the particle is,

$$M = \frac{4}{3}\pi R^3 \rho . \tag{142}$$

The energy E may be incorporated into a new quantity k:

$$k \equiv -\frac{2E}{mc^2} \tag{143}$$

The quantity k is a constant because all of the components of its definition are constants. However, since the mass m of the test particle is arbitrary, only the sign of k really matters. For this reason, one may limit discussion to only the values $k = -1,\ 0,\ +1$. As will be shown in Section 7.3 by arguments from general relativity, this quantity is the same k that specifies the curvature of the universe, as was discussed in Section 7.1:

$\mathbf{k = +1}$ in Eq. (143) means that the test particle's energy is negative, so it (and the other things in the universe) will not have enough energy to escape the gravitational pull of the universe; in this case the universe will eventually stop expanding and collapse back on itself.

$\mathbf{k = -1}$ in Eq. (143) implies the test particle's energy is positive, so it (and everything else in the universe) can continue moving outward indefinitely.

k = 0 marks the boundary between these two outcomes; for $k = 0$ the universe will just barely be able to continue to expand indefinitely and avoid gravitational collapse.

Substituting Eqs. (141)-(143) into (140) and rearranging, one finds

$$\frac{\dot{R}^2}{R^2} + \frac{c^2}{R^2}\,k = \frac{8\pi G}{3}\,\rho \qquad \textbf{First energy conservation equation} \tag{144}$$

Despite the scary appearance of Eq. (144), it still means simply that the total energy of particles in the universe is conserved, just as Eq. (140) does. The unusual arrangement of Eq. (144) facilitates comparison with the general relativistic result that will be derived in Section 7.3.

Critical density

By using $v = \dot{R} = HR$, Eq. (140) may be written as

$$E = \frac{4}{3}\pi G m R^2 \left(\frac{3H^2}{8\pi G} - \rho\right) = \frac{4}{3}\pi G m R^2 \left(\rho_{\text{crit}} - \rho\right) . \tag{145}$$

in which the critical density is

$$\rho_{\text{crit}} \equiv \frac{3H^2}{8\pi G} \tag{146}$$

The value of the density relative to ρ_{crit} is closely related to E and k:

$$\begin{array}{lll} \rho < \rho_{\text{crit}} \rightarrow E > 0 \text{ and } k = -1 & \text{open universe (expands indefinitely)} \\ \rho = \rho_{\text{crit}} \rightarrow E = 0 \text{ and } k = 0 & \text{flat universe (just barely expands indefinitely)} \\ \rho > \rho_{\text{crit}} \rightarrow E < 0 \text{ and } k = +1 & \text{closed universe (eventually collapses)} \end{array} \tag{147}$$

Second energy conservation equation

As the universe expands, its volume V increases by an amount dV, and it does an amount of work $p\,dV$. The energy of the universe must decrease by a corresponding amount $dE = -p\,dV$. Assuming that the universe is expanding much more slowly than the speed of light (as the measured Hubble constant indeed indicates), the energy of the universe is just the rest energy of its matter, so $dE = d(\rho V)c^2$. These concepts lead to a relationship between ρ and p for the universe:

$$\frac{d}{dt}E = -p\frac{d}{dt}V \tag{148}$$

$$\Longrightarrow c^2\frac{d}{dt}(\rho R^3) = -p\frac{d}{dt}(R^3) \tag{149}$$

$$\Longrightarrow \dot{\rho} = -3\,\frac{\dot{R}}{R}\left(\rho + \frac{p}{c^2}\right) \tag{150}$$

Combining Eqs. (144) and (150), one finds

$$2\,\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{c^2}{R^2}\,k = -\frac{8\pi G}{c^2}\,p \qquad \textbf{Second energy conservation equation} \tag{151}$$

Again, despite its scary appearance, Eq. (151) still just means that the total energy in the universe is conserved, just as Eqs. (144) and (148) do. The strange form in which Eq. (151) is written will aid comparison with the result to be derived from general relativity in Section 7.3. Equation (151) is directly related to Eq. (144) by the connection between density and pressure in Eq. (149).

7.4 Time Development of the Universe–Scary Math Version

Now the general relativistic derivation for the time development of the universe will be presented. The strategy of the derivation is first to calculate the stress-energy tensor $T_{\mu\nu}$ for a universe-filling "gas" using the Robertson-Walker metric $g_{\mu\nu}$ from Eq. (139), then to work out the Einstein curvature tensor $G_{\mu\nu}$ corresponding to the Robertson-Walker metric. When Einstein's equation (127) is used to relate these expressions for $T_{\mu\nu}$ and $G_{\mu\nu}$, the result is equations for the time dependence of R, the scale factor of the universe.

The stress-energy tensor $T_{\mu\nu}$ for a gas with density ρ and pressure p from Eq. (44) (using $v_\mu = v_\nu = (c, 0, 0, 0)$ assuming the gas particles are at rest in their local reference frames) may be adapted for the Robertson-Walker metric from Eq. (139):

$$T_{\mu\nu} = \left(\rho + \frac{p}{c^2}\right) v_\mu v_\nu + p g_{\mu\nu} = \begin{pmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & \left(\frac{pR^2}{1-kr'^2}\right) & 0 & 0 \\ 0 & 0 & pR^2 r'^2 & 0 \\ 0 & 0 & 0 & pR^2 r'^2 \sin^2\theta \end{pmatrix}_{\text{spherical}} \tag{152}$$

Starting with the Robertson-Walker metric from Eq. (139) and the definitions of Section 3.1, one can muddle through lots of scary algebra to sequentially work out the corresponding Christoffel symbol ${\Gamma^\alpha}_{\beta\gamma}$, Riemann curvature tensor ${R^\alpha}_{\beta\gamma\delta}$, Ricci curvature tensor $R_{\mu\nu}$, Ricci curvature scalar $R_{\text{Ricci scalar}}$ (using the name in the subscript here to avoid confusion with the scale factor of the universe R), and ultimately the Einstein curvature tensor $G_{\mu\nu}$:

$${\Gamma^1}_{01} = {\Gamma^1}_{10} = \frac{1}{2} g^{11} g_{11,0} = \frac{\dot{R}}{Rc}\,; \qquad \text{etc.} \tag{153}$$

$${R^1}_{010} = {R^2}_{020} = {R^3}_{030} = -\frac{\ddot{R}}{Rc^2}\,; \quad {R^0}_{000} = 0\,; \qquad \text{etc.} \tag{154}$$

$$R_{\mu\nu} = \begin{pmatrix} -3\frac{\ddot{R}}{Rc^2} & 0 & 0 & 0 \\ 0 & \left(\frac{2k + R\ddot{R}/c^2 + 2\dot{R}^2/c^2}{1-kr'^2}\right) & 0 & 0 \\ 0 & 0 & r'^2\left(2k + \frac{R\ddot{R}}{c^2} + 2\frac{\dot{R}^2}{c^2}\right) & 0 \\ 0 & 0 & 0 & r'^2 \sin^2\theta\left(2k + \frac{R\ddot{R}}{c^2} + 2\frac{\dot{R}^2}{c^2}\right) \end{pmatrix} \tag{155}$$

$$R_{\text{Ricci scalar}} = g^{\mu\nu} R_{\mu\nu} = -\frac{6}{R^2}\left(k + \frac{R\ddot{R}}{c^2} + \frac{\dot{R}^2}{c^2}\right) \tag{156}$$

$$G_{00} = 3\,\frac{\dot{R}^2}{R^2 c^2} + 3\frac{k}{R^2} \qquad \text{and} \qquad G_{11} = -\frac{k + 2R\ddot{R}/c^2 + \dot{R}^2/c^2}{1 - kr'^2}\,. \tag{157}$$

Applying Einstein's equation Eq. (127) (with the cosmological constant Λ included) to the time (00) and radial (11) components of $T_{\mu\nu}$ and $G_{\mu\nu}$ yields two equations:

$$\frac{\dot{R}^2}{R^2} + \frac{c^2}{R^2}\,k - \frac{c^2}{3}\Lambda = \frac{8\pi G}{3}\,\rho \qquad \textbf{First Friedmann-Lemaître equation} \quad (158)$$

$$2\,\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{c^2}{R^2}\,k - c^2\Lambda = -\frac{8\pi G}{c^2}\,p \qquad \textbf{Second Friedmann-Lemaître equation} \quad (159)$$

These equations confirm the results of the Newtonian derivation from Eqs. (144) and (151), except that they also account for the possibility of a cosmological constant Λ. They verify that the net effect of the cosmological constant is to contribute an effective positive energy density $\rho_{\Lambda}c^2 = c^2\Lambda/8\pi G$ to the total energy density and an effective negative pressure $p_{\Lambda} = -c^2\Lambda/8\pi G$ to the total pressure, as was predicted in Eqs. (130) and (131). As with Eqs. (144) and (151), Eqs. (158) and (159) are glorified energy conservation equations that are directly interrelated via Eq. (149) connecting the density ρ to the pressure p.

Equations (158) and (159) may be combined to find the deceleration of the cosmic scale factor:

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2} - \frac{c^2\Lambda}{4\pi G}\right) = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2} - 2\rho_{\Lambda}\right) . \qquad (160)$$

Substituting $\dot{R} = HR$ into Eq. (158) and solving for k confirms the critical density found earlier in Eqs. (145) and (146):

$$k = \frac{8\pi G R^2}{3c^2}\left(\rho + \frac{c^2\Lambda}{8\pi G} - \frac{3H^2}{8\pi G}\right) = \frac{8\pi G R^2}{3c^2}\left(\rho + \rho_{\Lambda} - \rho_{\rm crit}\right) , \qquad (161)$$

in which $\rho_{\rm crit} = 3H^2/8\pi G$ as in Eq. (146) and ρ_{Λ} is the contribution from the cosmological constant/vacuum energy. The density ρ could be subdivided into the density contributions $\rho_{\rm matter}$ from matter (both regular matter and dark matter) and $\rho_{\rm rad}$ from electromagnetic radiation. What matters in Eq. (161) is the sum $\Sigma\rho = \rho_{\rm matter} + \rho_{\rm rad} + \rho_{\Lambda}$. As before, the universe is closed with $k = +1$ for $\Sigma\rho > \rho_{\rm crit}$, flat with $k = 0$ for $\Sigma\rho = \rho_{\rm crit}$, and open with $k = -1$ for $\Sigma\rho < \rho_{\rm crit}$.

Hardcore cosmologists like to define the ratios of the total and component densities to the critical density:

$$\Omega \equiv \frac{\rho}{\rho_{\rm crit}} \qquad \Omega_{\rm matter} \equiv \frac{\rho_{\rm matter}}{\rho_{\rm crit}} \qquad \Omega_{\rm rad} \equiv \frac{\rho_{\rm rad}}{\rho_{\rm crit}} \qquad \Omega_{\Lambda} \equiv \frac{\rho_{\Lambda}}{\rho_{\rm crit}} \qquad (162)$$

$$\Omega = \Omega_{\rm matter} + \Omega_{\rm rad} + \Omega_{\Lambda} \qquad (163)$$

Using these definitions, Eq. (161) may be rewritten as

$$\Omega(t) - 1 = \frac{kc^2}{[H(t)R(t)]^2} = \frac{kc^2}{(dR/dt)^2} \qquad (164)$$

7.5 History of the Universe

The history and age of the universe depend on how much stuff there is in the universe, and whether that stuff is primarily matter (either regular matter or dark matter), electromagnetic radiation, vacuum energy, or some combination of these. Our universe appears to have experienced different eras when different components dominated its expansion, as shown in Fig. 32. We will consider these different eras separately in the following pages.

Empty universe

In an expanding but empty universe with negligible density, pressure, and vacuum energy ($\rho = 0$, $p = 0$, and $\Lambda = 0$), there is no gravitational attraction to alter the speed of expansion over time, and Eq. (160) yields

$$\ddot{R} = 0 \qquad \Longrightarrow \qquad R(t) = H_0 t \,. \tag{165}$$

Thus if the universe is empty, its current age is

$$t_0 = \frac{1}{H_0} \approx 13.8 \text{ billion years} \tag{166}$$

just as in Eq. (135). This appears to be almost exactly the right answer for our universe, but for the wrong reason. Observational data indicates that our universe is far from empty, and in fact is filled with a density right at the critical value ρ_{crit}. Thus we will now turn our attention to cases in which the universe is filled with matter, radiation, vacuum energy, or combinations thereof.

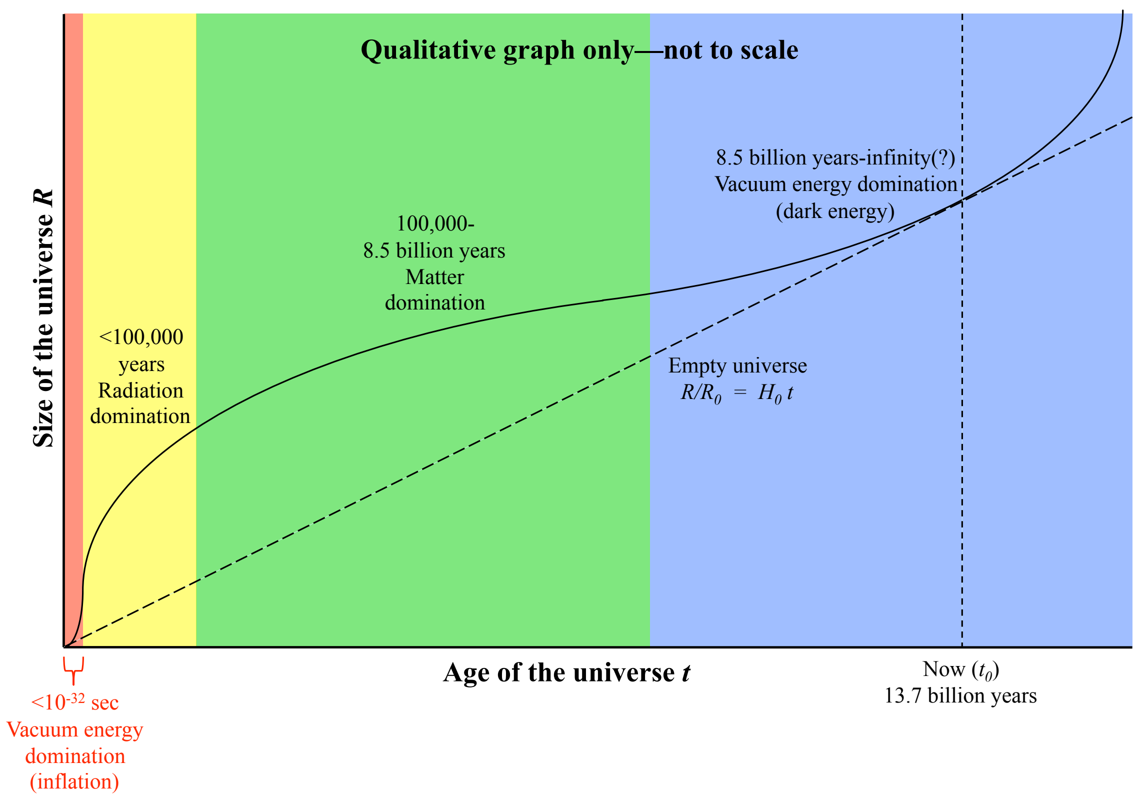

Figure 32. The expansion of the size of the universe R with time t has been dominated first by vacuum energy (inflation), then radiation, then matter, and now a different form of vacuum energy (dark energy). For comparison, the dashed diagonal line shows an empty universe that expands at a constant rate.

Matter dominance

For most of the 20th century it was believed that our universe was dominated by matter. If radiation may be neglected, the total mass of the universe is conserved,

$$\frac{4}{3}\pi R^3 \rho_{\rm matter} \;\equiv\; \frac{4}{3}\pi R_0^3 \rho_{\rm matter\ 0}\ , \tag{167}$$

where R_0 and $\rho_{\rm matter\ 0}$ are the current scale factor and matter density (ordinary matter plus dark matter, whatever the heck that is).

Using Eq. (167), Eq. (158) may be rewritten as:

$$\dot{R}^2 \;=\; \frac{8\pi G R_0^3 \rho_{\rm matter\ 0}}{3R} \;+\; \frac{1}{3}\Lambda c^2 R^2 \;-\; c^2 k \tag{168}$$

Equation (168) may be integrated to find the time dependence of R for various values of k and Λ. The simplest case to solve is that of a flat universe ($k = 0$) and with no vacuum energy ($\Lambda = 0$), which is called the Einstein-de Sitter model for the universe. For this case the solution is

$$R(t) \;=\; (6\pi G R_0^3 \rho_{\rm matter\ 0})^{1/3}\, t^{2/3}\ . \tag{169}$$

As Eq. (168) shows, even universe models with $k = \pm 1$ initially expand like the Einstein-de Sitter universe when R is relatively small, especially when ρ is very close to $\rho_{\rm crit}$.

For the Einstein-de Sitter model, the universe's age may be determined from the Hubble constant:

$$t_0 \;=\; \frac{2}{3}\frac{R}{\dot{R}} \;=\; \frac{2}{3}\frac{1}{H_0} \;\approx\; 9.2 \text{ billion years} \tag{170}$$

From Eqs. (169) and (170), a matter-dominated universe expands as:

$$\frac{R}{R_0} \;\approx\; \left(\frac{t}{9.2 \text{ billion years}}\right)^{2/3} \tag{171}$$

It was long believed that the Einstein-de Sitter model best described our universe, but more precise measurements of H_0 revealed that the age given by Eq. (170) is too short and the age of the oldest stars in the universe is too long ($\sim$ 13 billion years). Thus our universe has not been matter-dominated for at least part of its history to date.

Radiation dominance

The very early universe was radiation-dominated; this caused a different time dependence of R, which will now be calculated. First note that a photon's wavelength gets stretched as R increases, so the photon's energy E or frequency f is $E \propto f \propto 1/R$. The number of photons per volume also decreases like $\sim 1/R^3$ as the volume of the universe increases. Therefore the radiation energy density is $\rho_{\rm rad} \propto E/R^3 \propto 1/R^4$, or in terms of the current radiation density $\rho_{\rm rad\ 0}$ in the universe:

$$\rho_{\rm rad} \;=\; \rho_{\rm rad\ 0}\left(\frac{R_0}{R}\right)^4\ . \tag{172}$$

For radiation, the relationship between pressure and density is $3p = \rho c^2$. By using this information in Eq. (160) and setting $\Lambda = 0$, one finds

$$\frac{\ddot{R}}{R} \;=\; \Omega_{\rm rad\ 0}\, H_0^2 \left(\frac{R_0}{R}\right)^4\ . \tag{173}$$

Integrating Eq. (173) reveals that $R(t) \propto \sqrt{t}$ for a radiation-dominated universe:

$$R(t) \;=\; \left(\frac{\Omega_{\rm rad\,0}}{2}\right)^{1/4} R_0\sqrt{H_0 t}\;. \qquad (174)$$

Radiation was much more important in the early universe because $\rho_{\rm rad} \propto 1/R^4$, whereas $\rho_{\rm matter} \propto 1/R^3$, so the ratio at any time t is:

$$\frac{\rho_{\rm rad}(t)}{\rho_{\rm matter}(t)} \;=\; \frac{\Omega_{\rm rad\,0}}{\Omega_{\rm matter\,0}}\,\frac{R_0}{R(t)}\;. \qquad (175)$$

The universe transitioned from an era of radiation dominance to an era of matter dominance, $\rho_{\rm rad}(t_{\rm transition}) = \rho_{\rm matter}(t_{\rm transition})$, when the size of the universe was $R_{\rm transition}$:

$$\frac{R_{\rm transition}}{R_0} \;=\; \frac{\Omega_{\rm rad\,0}}{\Omega_{\rm matter\,0}} \;\approx\; 2\times 10^{-4}\;. \qquad (176)$$

Using Eq. (176) with $\Omega_{\rm rad\,0} \approx 6\times 10^{-5}$ and $\Omega_{\rm matter\,0} \approx 0.28$ and solving Eq. (174) for the time of transition between the radiation-dominated and matter-dominated eras of the universe, one finds

$$t_{\rm transition} \;=\; \sqrt{\frac{2}{\Omega_{\rm rad\,0}}}\left(\frac{R_{\rm transition}}{R_0}\right)^2\frac{1}{H_0} \;=\; \sqrt{\frac{2}{\Omega_{\rm rad\,0}}}\left(\frac{\Omega_{\rm rad\,0}}{\Omega_{\rm matter\,0}}\right)^2\frac{1}{H_0} \;\approx\; 100{,}000 \;\text{ years} \qquad (177)$$

The universe is filled with thermal radiation (the cosmic microwave background radiation) with a current temperature of 2.725^oK, or 2.35×10^{-4} electron-Volts (eV). Since the temperature of thermal radiation varies as $T \propto 1/R$ just as the photons' energy does, the temperature of the radiation at other times in the history of the universe (not just the radiation dominated era) was

$$T_{\rm rad}(t) \;\approx\; 2.725\,\frac{R_0}{R(t)}\;{}^o{\rm K} \;\approx\; 2.35\times 10^{-4}\,\frac{R_0}{R(t)}\;\;{\rm eV} \qquad (178)$$

Combining Eqs. (174) and (178) and using $H_0 \approx 2.3\times 10^{-18}\ {\rm sec}^{-1}$, the temperature of the universe at any time t (in seconds) during the radiation-dominated era was

$$T(t) \;\approx\; \frac{2.09\times 10^6\ {\rm eV}}{\sqrt{t}\ ({\rm sec})} \qquad (179)$$

Big Bang nucleosynthesis fused some hydrogen into helium-4 and trace amounts of other elements during the first few minutes of the universe. Weak nuclear force interactions freely converted protons into neutrons and vice versa until the universe cooled below $T = 0.8$ MeV [$t \approx 7$ sec from Eq. (179)], and neutrons fused with protons to become deuterium and ultimately helium-4 once the universe cooled to $T = 0.1$ MeV [$t \approx 437$ sec from Eq. (179)]. Free neutrons decay exponentially with a mean lifetime of ~886 sec and had to survive ~430 sec before they were stabilized by becoming part of nuclei, so ~ exp(-430 sec/886 sec) of the initial neutrons survived. Protons have a rest energy of 938.3 MeV and neutrons have a rest energy of 939.6 MeV, so in thermal equilibrium at 0.8 MeV when they were last created, the number of neutrons n_n was much smaller than the number of protons n_p due to the 1.3 MeV of extra energy required by the neutrons and the Boltzmann exponential factor (*Statistical Physics* ?.?), $n_n/n_p \approx$ exp(-1.3 MeV/0.8 MeV). Combining these factors, the relative amount of neutrons that were created and survived to form nuclei is

$$\frac{n_n}{n_p} \;\approx\; \exp\left(-\frac{1.3\ {\rm MeV}}{0.8\ {\rm MeV}}\right)\exp\left(-\frac{430\ {\rm sec}}{886\ {\rm sec}}\right) \;\approx\; \frac{1}{8} \qquad (180)$$

This is very close to the answer $n_n/n_p \approx 1/7$ from more detailed calculations–the universe has about one neutron for every seven protons. With that ratio, two neutrons and 14 protons can form one helium-4 and 12 hydrogen atoms, which agrees well with the observed amount. Not counting subsequent nucleosynthesis in stars, the universe has $\sim 12/13$ hydrogen and $\sim 1/13$ helium atoms by number, or ~75% hydrogen and ~25% helium-4 by mass. There is also ~0.01% deuterium and much smaller amounts of helium-3, lithium-6, lithium-7, and beryllium-7, as predicted by detailed calculations and confirmed by spectroscopic observations of the universe [14]-[17].

Decoupling of radiation from matter occurred when the universe cooled to $\sim 3000^o$K ~ 0.3 eV. Prior to this time, all or much of the matter in the universe was ionized (with some electrons not bound by nuclei) and thus strongly interactive with electromagnetic radiation and in thermal equilibrium with it. After this time, almost all of the electrons were bound in atoms and the matter became transparent to photons and no longer interacted with them. The present cosmic microwave background radiation dates to this time, when the size of the universe was $R/R_0 \approx 9 \times 10^{-4}$ using Eq. (178). Equations (176) and (177) found that the radiation-dominated era lasted $\sim 100,000$ years until $R/R_0 \approx 2 \times 10^{-4}$. Assuming subsequent complete matter dominance for simplicity, Eq. (171) shows that the time required for the universe to grow from $R/R_0 \approx 2 \times 10^{-4}$ to 9×10^{-4} was $\sim 200,000$ years. Thus the cosmic microwave background attained its final form as ~3000 ^{o}K blackbody radiation $\sim 300,000$ years after the Big Bang and has been redshifted to 2.725 ^{o}K blackbody radiation as the universe has expanded in the ~13.7 billion years since that time.

Vacuum energy dominance

The expansion of the universe proceeds very differently if the universe is dominated by vacuum energy Λ instead of matter or radiation. In this case, Eq. (160) becomes

$$\frac{\ddot{R}}{R} = \frac{c^2\Lambda}{3} \,. \tag{181}$$

The solution of Eq. (181) is exponential **inflation** of the size of the universe on a timescale τ:

$$R(t) \propto e^{t/\tau} \qquad \text{where} \qquad \tau \equiv \sqrt{\frac{3}{c^2\Lambda}} \tag{182}$$

Section 6.3 provided the physical basis for understanding this exponential inflation. From Eqs. (130) and (131), vacuum energy fills all of space with a constant energy density (which is determined by fundamental quantum physics considerations and cannot change if the vacuum energy's volume is changed) but a negative pressure. The negative pressure ensures that if the volume occupied by the vacuum energy increases by an amount dV, the associated work $dW = -p\,dV$ will be sufficient to provide the energy needed to *create* more vacuum energy to fill the new volume. Thus the universe gains energy by expanding, just as the mythical substance flubber gained energy by bouncing in the movies. The universe expands a bit, which gives it some energy and make it want to expand more, which gives it more energy and makes it want to expand even more, and thus the size of the universe grows exponentially. An alternate but equivalent view is that the vacuum energy exerts an antigravitational force, pushing the universe apart instead of trying to pull it together like the gravitational attraction of normal energy and matter.

Current astronomical observations indicate that there is a small but nonzero Λ, called dark energy, which may have always been present but is only now becoming dominant as the expansion of the universe causes normal energy and matter to become more dilute and thus to exert less influence. If this is correct, the coming billions of years will see our universe begin to expand exponentially and completely isolate each galaxy from the other galaxies. Using the measured dark energy density $\Omega_\lambda \equiv c^2\Lambda/(3H_0^2) \approx 0.72$, the time constant for this exponential fate of the universe is:

$$\tau = \frac{1}{H_0\sqrt{\Omega_\Lambda}} \approx 16.3 \text{ billion years} \tag{183}$$

Of course, we currently have no idea what exactly dark energy is, so it is possible that it could dissipate, increase, or do something else in the far future of the universe. The recent history of the universe, in which there has been a mixture of dark energy and matter, will be discussed shortly.

Cosmologists believe that a different form of vacuum energy Λ existed for a brief time after the Big Bang. The universe may have come into existence with empty space itself in a metastable excited quantum energy state (called a false vacuum) instead of an actual ground energy state (the true vacuum). The universe would then have significant vacuum energy and its size would grow exponentially, until space decayed to the true vacuum, at which point that initial vacuum energy would be converted into radiation and matter . There are at least three reasons to believe that the universe went through a brief initial period of exponential growth, known as **inflation** [14]-[17]:

- **Flatness ($\Omega = 1$) of the universe.** The flatness of the universe (how close the total density of radiation, matter, and vacuum energy is to the critical density) evolves differently over the history of the universe, depending on whether the universe is truly flat ($k = 0$), open ($k = -1$), or closed ($k = 1$) and dominated by matter, radiation, or vacuum energy. Using Eq. (164) with Eqs. (169), (174), and (182), one finds:

$$|\Omega(t) - 1| = \frac{|k|\,c^2}{(dR/dt)^2} \begin{cases} = 0 & \text{for } k = 0 \\ \propto t^{2/3} & \text{for } k = \pm 1 \text{ and matter-dominance} \\ \propto t & \text{for } k = \pm 1 \text{ and radiation-dominance} \\ \propto \exp(-2t/\tau) & \text{for } k = \pm 1 \text{ and vacuum-energy-dominance} \end{cases} \tag{184}$$

  Thus if the universe begins perfectly flat with $k = 0$ and $\Omega = 1$, it will remain that way for all time. However, if the universe is radiation dominated (as it was in its early history) or matter dominated (as it has been for most of its history) and begins with even a little curvature ($k = \pm 1$ and $|\Omega - 1|$ small but nonzero), that curvature will grow rapidly as time passes, as shown in Eq. (184). [These results are based on Eqs. (169) and (174), whose assumption of negligible curvature will break down once the curvature becomes too large, so $|\Omega(t) - 1|$ will eventually level off.] For our universe to be as flat as it is now ($|\Omega(t_0) - 1| < 10^{-2}$ for $t_0 \approx 4 \times 10^{17}$ sec), it would have had to be incredibly, utterly close to flat ($|\Omega(t) - 1| < 10^{-14}$ for $t \sim 1$ sec) in the earliest stages of the universe, which is an exceedingly tight constraint. On the other hand, if inflation occurred during the early universe, $|\Omega - 1|$ would have approached 0 exponentially, as shown by Eq. (184). Physically, this just means that even if the universe has a spherical or saddle shape as in Fig. 31, if inflation blew that shape up to a very large size and we could only see a very small amount of it (only those regions whose light was able to reach earth within 13.7 billion years), that very small piece of a homongous spherical or saddle-shaped universe would look essentially flat.

- **Dilution of strange relic particles from the Big Bang.** Most grand unified theory models of how forces and particles were unified at the time of the Big Bang predict the creation of magnetic monopoles and other weird particles during or immediately after the Big Bang. Rapid inflation of the size of the universe would greatly dilute the concentration of these particles and explain why we don't see any signs of them now.

- **Identical conditions between causally disconnected regions of the universe (horizon problem).** The cosmic microwave background radiation temperature, the density of matter, physical laws, and other properties appear to be the same as far as we can see in one direction in the universe (> 13 billion light-years away) and also as far as we can see in the opposite direction, even though those regions are too far apart (> 26 billion light-years) to have had any influence on each other during the ∼ 13.7-billion-year history of the universe. However, if inflation made the early universe expand much faster than light (that's legal–it's just empty space between things expanding faster than light, not objects themselves moving through surrounding space faster than light), all the regions we can see could have originally been part of the same causally connected region before inflation. In fact, the early universe may have had different regions with different densities, temperatures, and even fundamental physical constants and laws, and all the universe we can currently see expanded from one of those small homogeneous regions. Far beyond the visible universe, far beyond where we could travel without exceeding the speed of light, may lie other regions of the universe with very different properties.

Current measurements suggest that exponential inflation may have proceeded for at least ∼ 60 time constants all within the first $\sim 10^{-32}$ seconds after the Big Bang [14]-[17].

Dominance by both vacuum energy and matter

Our universe appears to be flat ($k = 0$ and $\Omega = 1$) with significant amounts of matter and vacuum energy but negligible radiation density for most of its history ($\Omega_{\text{matter}} + \Omega_\Lambda = 1$). Using these conditions, Eq. (168) may be rewritten as:

$$\dot{R} \;=\; H_0 R_0 \left[(1-\Omega_\Lambda)\frac{R_0}{R} + \Omega_\Lambda \frac{R^2}{R_0^2} \right]^{1/2} \tag{185}$$

Using Eq. (185), one can calculate the age of our universe:

$$t_0 \;=\; \int_0^{R_0} \frac{dR}{\dot{R}} \;=\; \frac{1}{H_0}\int_0^{R_0} \frac{dR}{R_0}\left[(1-\Omega_\Lambda)\frac{R_0}{R} + \Omega_\Lambda \frac{R^2}{R_0^2} \right]^{-1/2} \tag{186}$$

$$\;=\; \frac{2}{3}\,\frac{1}{H_0}\,\frac{1}{\sqrt{\Omega_\Lambda}}\,\ln\left(\frac{1+\sqrt{\Omega_\Lambda}}{\sqrt{1-\Omega_\Lambda}} \right) \tag{187}$$

in which the integral is done by cheating and looking it up in a book of scary integrals.

The age of a matter- and vacuum-energy-filled universe comes out to be about the same as the age of an empty universe, $t_0 \approx H_0^{-1}$, when $\Omega_\Lambda \approx 0.74$ and $\Omega_{\text{matter}} \approx 0.26$, which is almost exactly what we have measured for the actual universe. The physical explanation for this result is that for most of the history of our universe to this point, the "antigravity" repulsion of the vacuum energy has nearly balanced the gravitational attraction of the ordinary and dark matter, so our universe has been expanding at a nearly constant rate.

In the future, the influence of vacuum energy's antigravity will increase and matter's gravity will decrease, and the universe will expand more and more rapidly, ultimately returning to the exponential inflation of a vacuum-energy-dominated universe as calculated previously. (The vacuum energy that produced inflation in the very early universe probably had a different origin than the vacuum energy that is mildly accelerating the universe's expansion now and that will ultimately drive future inflation.)

References

[1] Edwin F. Taylor and John A. Wheeler, *Space-time Physics* (Freeman, San Francisco, 1966). A very readable introduction to special relativity.

[2] Lewis C. Epstein, *Relativity Visualized* (Insight Press, 1985). A nonmathematical yet very physically insightful introduction to special and general relativity.

[3] I.R. Kenyon, *General Relativity* (Oxford University Press, Oxford, 1990). This is probably the best introductory general relativity textbook–very concise without sacrificing much detail or scope, with shorter yet better explained derivations than most other general relativity books.

[4] Edwin F. Taylor and John A. Wheeler, *Exploring Black Holes: Introduction to General Relativity* (Addison Wesley Longman, New York, 2000). A very readable if incomplete and idiosyncratic introduction to general relativity.

[5] John A. Wheeler, *A Journey into Gravity and space-time* (Scientific American, New York, 1990). A very simple introduction to general relativity with beautiful illustrations and virtually no equations.

[6] Bernard Schutz, *Gravity from the Ground Up: An Introductory Guide to Gravity and General Relativity* (Cambridge University Press, 2003). This is another introductory book that avoids extensive math and gives fairly good explanations.

[7] Charles W. Misner, Kip S. Thorne, and John A. Wheeler, *Gravitation* (Freeman, New York, 1973). Often called the "Big Black Phone Book of Gravity," this textbook looks truly imposing, but it gives clearer and more physically intuitive pictures of general relativity theory than can be found in most other textbooks (which is not saying much). Other nice features are its presentations of multiple ways of looking at certain things (e.g. multiple, different derivations of Einstein's gravitational field equation) and its comprehensive nature (at least for its time; there have been a few new and possibly even useful ideas in general relativity theory since 1973).

[8] Ray d'Inverno, *Introducing Einstein's Relativity* (Oxford University Press, Oxford, 1992).

[9] Hans C. Ohanian and Remo Ruffini, *Gravitation and space-time* (2nd edition, W.W. Norton, 1994).

[10] James B. Hartle, *Gravity: An Introduction to Einstein's General Relativity* (Benjamin Cummings, 2003).

[11] Sean Carroll, *Space-time and Geometry: An Introduction to General Relativity* (Benjamin Cummings, 2003).
These four books are longer than Kenyon, shorter than Misner, Thorne, and Wheeler, and more up-to-date than the latter. They are roughly comparable to each other, although sometimes one will have better coverage on a particular topic than the others.

[12] Stuart L. Shapiro and Saul A. Teukolsky, *Black Holes, White Dwarfs, and Neutron Stars* (Wiley, New York, 1983). Explains how general relativity affects stellar structure and stellar evolution, and discusses stellar collapse to white dwarfs, neutron stars, and black holes.

[13] Matt Visser, *Lorentzian Wormholes: From Einstein to Hawking* (American Institute of Physics, Woodbury, NY, 1995). Begins with a nice little summary of some of the main points of general relativity, then gives a good treatment of some modern issues in general relativity theory: wormholes, time travel, and quantum gravity.

[14] Jeremy Bernstein, *An Introduction to Cosmology* (Prentice Hall, Englewood Cliffs, NJ, 1995).

[15] Andrew Liddle, *An Introduction to Modern Cosmology* (2nd edition, Wiley, New York, 2003). Both of these books are very readable introductions to cosmology that provide much physical insight and try to keep the math from getting very hairy.

[16] Edward W. Kolb and Michael S. Turner, *The Early Universe* (Addison Wesley, Reading, MA, 1994).

[17] Scott Dodelson, *Modern Cosmology* (Academic Press, 2003).
These are both modern, thorough treatments of cosmology that incorporate both general relativity and quantum field theory. Dodelson is more up-to-date, but Kolb and Turner contains some material not in Dodelson.