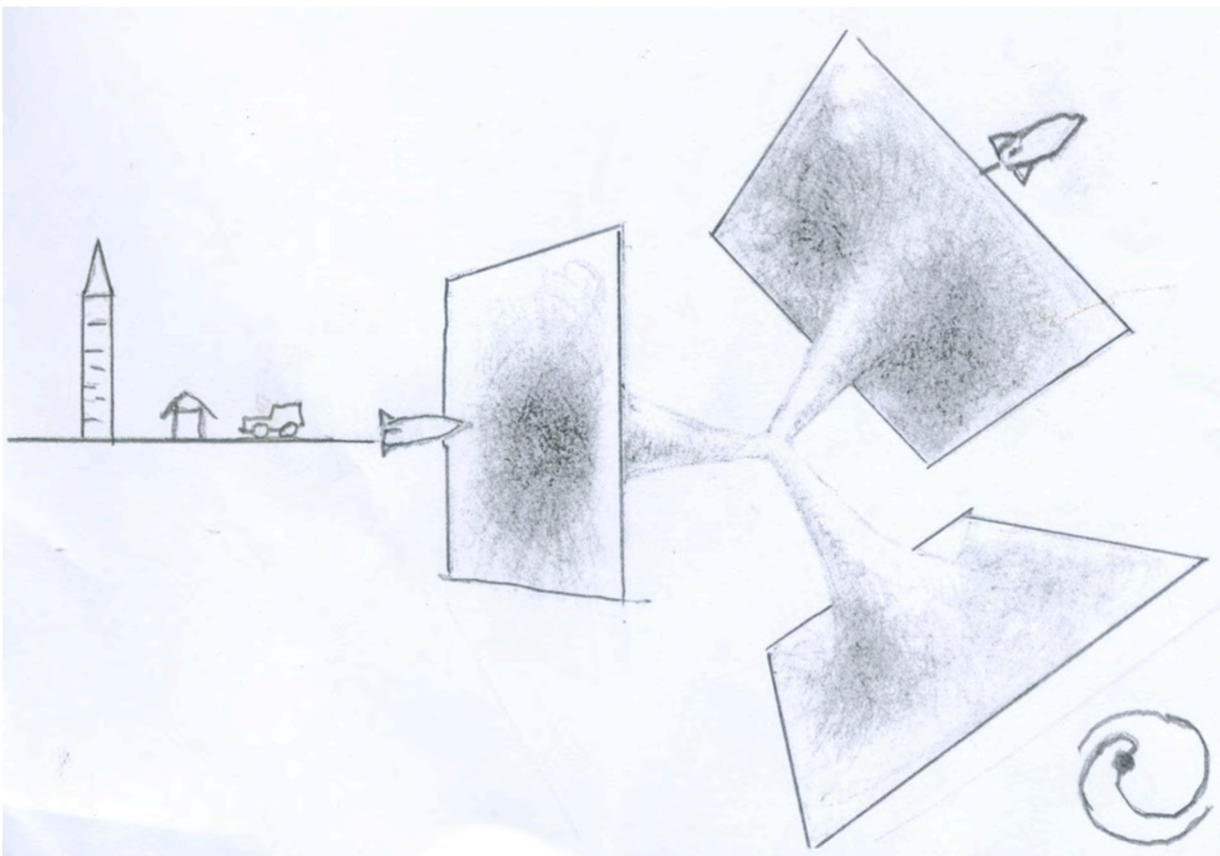
EXPLORING BLACK HOLES

Introduction to General Relativity

Second Edition



funnel image by Thomas Dahill

Edwin F. Taylor

John Archibald Wheeler

Edmund Bertschinger

Selected Physical and Astronomical Constants

[Inside front cover, left]

Gravitational constant	$G = 6.67 \times 10^{-11}$ meter^3/(kilogram-second^2)

EARTH

Mass of Earth	$M_{\mathrm{E}} = 5.97 \times 10^{24}$ kilogram $= 4.44 \times 10^{-3}$ meter
Equatorial radius of Earth	$r_{\mathrm{E}} = 6.38 \times 10^{6}$ meter
Gravitational acceleration on Earth	$g_{\mathrm{E}} = 9.81$ meter/second^2 $= 1.09 \times 10^{-16}$ meter^{-1}
Mean distance of Earth from Sun $\equiv$ 1 astronomical unit (AU)	1 AU $= 1.50 \times 10^{11}$ meter
Mean speed of Earth in its orbit around Sun	$v_{\mathrm{E}} = 2.98 \times 10^{4}$ meter/second

OUR SUN

Mass of the Sun	$M_{\mathrm{Sun}} = 1.99 \times 10^{30}$ kilogram $= 1.48 \times 10^{3}$ meter
Radius of the Sun	$r_{\mathrm{Sun}} = 6.96 \times 10^{8}$ meter

OUR UNIVERSE

Age of the Universe	$= (13.7 \pm 0.1) \times 10^{9}$ year $= 1.30 \times 10^{26}$ meter
Hubble constant now	$H_0 = 72 \pm 2$ (kilometer/second)/Megaparsec $= 7.37 \times 10^{-11}$ (light year) $^{-1}$ $= 7.78 \times 10^{-27}$ meter^{-1}

Conversion Factors

Speed of light in a vacuum (by definition): $c \equiv 2.997\,924\,58 \times 10^{8}$ meter/second

1 second $\approx 3.00 \times 10^{8}$ meter of light-travel time

1 meter of light-travel time $= 3.34 \times 10^{-9}$ second

1 year $= 3.16 \times 10^{7}$ second $= 9.46 \times 10^{15}$ meter of light-travel time

1 light-year$= 3.16 \times 10^{7}$ second $\times c$ $= 9.46 \times 10^{15}$ meters

1 kilometer $= 0.621$ mile

1 kilogram $= 7.42 \times 10^{-28}$ meter of mass $= 5.61 \times 10^{32}$ electron-volt

1 parsec $= 3.26$ light year

Approximation

$(1+\epsilon)^n \approx 1 + n\epsilon + O\left(\epsilon^2\right)$ $\qquad$ $(|\epsilon| \ll 1$ and $|n\epsilon| \ll 1)$

For time in unit of meter, see Chapter 1, Speeding
For mass in unit of meter, see Chapter 3, Curving

[Inside front cover, right]

Can I see a black hole at all? If I can see it, what does a black hole look like? Does it *look* black? Where do black holes exist in the Universe? Does the black hole look different when I fall toward it? What does it *feel* like to fall into a black hole? Am I comfortable? Do I see the stars overhead as I fall into a black hole? If so, do these stars change position or color as I fall? How fast do I fall? Does my speed reach or exceed the speed of light? Once inside, can I receive messages and packages from my friends on the outside? Is it true that, once inside, I cannot send anything to my friends on the outside, not even a light signal? *Why* can't I send them messages? How long do I live once I fall into a black hole? Will I reach the center alive? Can I see the crunch-point ahead of me? What is the last thing I see? Is my death quick and painless? What happens to the mass of a black hole when it swallows me or swallows a stone? How does the orbit of a stone around a black hole differ from the orbit of a planet around our Sun? Newton says a planet stays in orbit because the Sun exerts a gravitational force on it. How does Einstein explain this orbit? If Newton and Einstein disagree, how do we decide between them? How close to a black hole can I move in a circular orbit? Can I use a black hole to travel rapidly forward in time? backward in time? What are the upper and lower limits on the mass of a star, a white dwarf, a neutron star, a black hole? Which of these bodies require general relativity for its correct description? In what sense are space and time unified? Why do things fall in my everyday life on Earth? Does the term *relativity* mean that everything is relative? What does *curvature* mean? How can I observe curvature? How many different observed effects does curvature account for? How does the Global Positioning System fail if we ignore general relativity? How much does light change direction as it passes the Sun or a black hole? Does the amount of change in direction depend on the color of the light? How does an astronomical object focus light from a distant galaxy and what does the image of that distant galaxy look like? Can light go into a permanent orbit around a black hole? How fast can a black hole spin? Does a spinning black hole drag space around with it? What does "drag space" *mean*; how can I observe it? Can I extract energy from a spinning black hole? What is a quasar? Do spinning black holes power quasars; if so, how? What are gravitational waves? What can gravitational waves tell us about the Universe that light cannot? How far away is the most distant galaxy that we see? Is the Universe just a big black hole? How did the Universe begin? What is the Universe made of? What does the *Big Bang* mean and how do we know it happened? Did time and space exist before the Big Bang? Why does the Universe expand with time? What is the Universe expanding into? Why is the expansion of the Universe speeding up? Where is the center of the Universe, anyway? How will the Universe end—or will it go on forever? Will the Earth exist twenty billion years from now? Whether or not the Earth exists then, what will the heavens look like from the position of our solar system?

Curiosity, like coffee, is an acquired need. Just a titillation
at the beginning, it becomes with training a raging passion.
—Nicholas S. Thompson

I have no special talents. I am only passionately curious.
—Albert Einstein

[Title page, right-hand side]

EXPLORING BLACK HOLES

INTRODUCTION TO GENERAL RELATIVITY

Second Edition

Edwin F. Taylor

John Archibald Wheeler

Edmund Bertschinger

[Right-hand page]

.

It is not my purpose in this discussion to represent the general theory of relativity as a system that is as simple and as logical as possible, and with the minimum number of axioms; but my main object here is to develop this theory in such a way that the reader will feel that the path we have entered upon is psychologically the natural one, and that the underlying assumptions will seem to have the highest possible degree of security.

—Albert Einstein

Everything important is, at bottom, utterly simple.

—John Archibald Wheeler

[REFERENCES:

First quote: Albert Einstein, "The Foundation of the General Theory of Relativity" in *The Principle of Relativity*, translated by W. Perrett and G. B. Jeffery, 1952 Dover Publications, ISBN 0486600815, page 118.

Second quote in Robert Oerter in *The Theory of Almost Everything: The Standard Model, The Unsung Triumph of Modern Physics*, 2006.]

Download file name: AAFrontMatter170421v6.pdf

CONTENTS, Exploring Black Holes, Second Edition

Chapter 1. Speeding

- *What is the key idea of relativity?*
- *Everything is relative, right?*
- *"Space and time form a unity called spacetime." Huh?*
- *Do people in relative motion* age *differently? Do they* feel *the slowing down/speeding up of their aging?*
- *What is the farthest galaxy I can possibly visit in person?*
- *Can an advanced civilization create a rocket to carry "humanity" anywhere in our galaxy? How soon can we on Earth learn of their discoveries?*
- *How do relativistic expressions for energy and momentum differ from those of Newton?*
- *When and why does special relativity break down, and what warns us that this is about to happen?*

DOWNLOAD FILE NAME: **Ch01Speeding170508v1.pdf**

CHAPTER 1

Speeding

Edmund Bertschinger & Edwin F. Taylor *

I've completely solved the problem. My solution was to analyze the concept of time. Time cannot be absolutely defined, and there is an inseparable relation between time and signal velocity.

—Albert Einstein, May 1905, to his friend Michele Besso

1.1 ■ SPECIAL RELATIVITY

Special relativity and general relativity

Special relativity distinguished from **General relativity**

Special relativity describes the very fast and reveals the unities of both space-time and mass-energy. **General relativity**, a **Theory of Gravitation**, describes spacetime and motion near a massive object, for example a star, a galaxy, or a black hole. The present chapter reviews a few key concepts of special relativity as an introduction to general relativity.

Begin relativity with a stone wearing a wristwatch.

What is at the root of relativity? Is there a single, simple idea that launches us along the road to understanding? At the beginning of *Alice in Wonderland* a rabbit rushes past carrying a pocket watch. At the beginning of our relativity adventure a small stone wearing a wristwatch flies past us.

Observe two events in laboratory frame.

The wristwatch ticks once at Event 1, then ticks again at Event 2. At each event the stone emits a flash of light. The top panel of Figure 1 shows these events as observed in the laboratory frame. We assume that the laboratory is an **inertial reference frame**.

Definition: **inertial frame**

DEFINITION 1. Inertial frame
An **inertial reference frame**, which we usually call an **inertial frame**, is a region of spacetime in which Newton's first law of motion holds: *A free stone at rest remains at rest; a free stone in motion continues that motion at constant speed in a straight line.*

We are interested in the records of these two events made by someone in the laboratory. We call this someone, the **observer**:

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

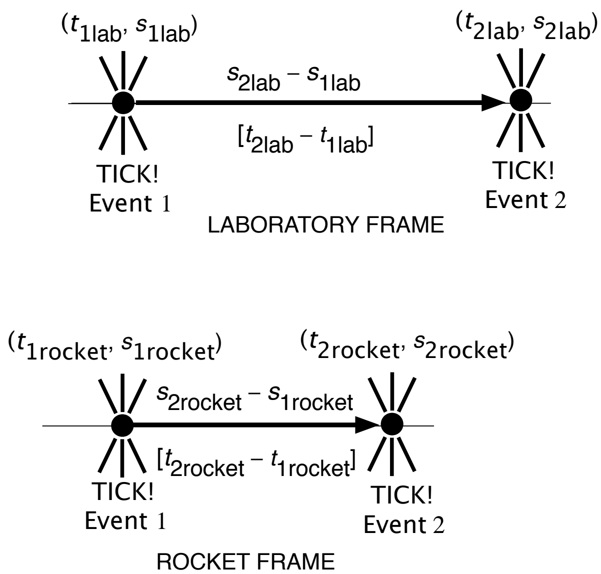


FIGURE 1 A free stone moves through a laboratory at constant speed. The stone wears a wristwatch that ticks as it emits a first flash at Event 1 and a second flash at Event 2. *Top panel:* The laboratory observer records Event 1 at coordinates $(t_{1\text{lab}}, s_{1\text{lab}})$ and Event 2 at coordinates $(t_{2\text{lab}}, s_{2\text{lab}})$. *Bottom panel:* An unpowered rocket ship streaks through the laboratory; the observer riding in the rocket ship records Event 1 at rocket coordinates $(t_{1\text{rocket}}, s_{1\text{rocket}})$ and Event 2 at $(t_{2\text{rocket}}, s_{2\text{rocket}})$. Each observer calculates the distance and time lapse between the two events, displayed on the line between them.

Definition: **inertial observer**

DEFINITION 2. Observer $\equiv$ inertial observer

An **inertial observer** is an observer who makes measurements using the space and time coordinates of any given inertial frame. In this book we *choose* to report *every* measurement and observation using an inertial frame. Therefore in this book **observer $\equiv$ inertial observer**.

The top panel of Figure 1 summarizes the records of the laboratory observer, who uses the standard notation $(t_{1\text{lab}}, s_{1\text{lab}})$ for the lab-measured time and space coordinates of Event 1 and $(t_{2\text{lab}}, s_{2\text{lab}})$ for the coordinates of Event 2.

The laboratory observer calculates the *difference* between the time coordinates of the two events and the *difference* between the space coordinates of the two events that she measures in her frame. The top panel of Figure 1 labels these results.

Next an unpowered rocket moves through the laboratory along the line connecting Event 1 and Event 2. An observer who rides in the rocket measures the coordinates of the two events and constructs the bottom panel in Figure 1.

Now the key result of special relativity: There is a surprising relation between the coordinate differences measured in laboratory and rocket frames, both of which are inertial frames. Here is that expression:

Surprise: Both observers calculate the same wristwatch time between two events.

$$\tau^2 = (t_{2\text{lab}} - t_{1\text{lab}})^2 - (s_{2\text{lab}} - s_{1\text{lab}})^2 = (t_{2\text{rocket}} - t_{1\text{rocket}})^2 - (s_{2\text{rocket}} - s_{1\text{rocket}})^2 \quad (1)$$

The expression on the left side of (1) is the square of the so-called **wristwatch time** τ, which we define explicitly in the following section. Special relativity says that the wristwatch time lapse of the stone that moves directly between events can be predicted (calculated) by both laboratory and rocket observers, each using his or her own time and space coordinates. The middle expression in (1) contains only laboratory coordinates of the two events. The right-hand expression contains only rocket coordinates of the same two events. Each observer predicts (calculates) the same value of the stone's wristwatch time lapse as it travels between these two events.

Fuller Explanation: *Spacetime Physics,* Chapter 1. Chapter 2, Section 2.6, shows how to synchronize the clocks in each frame with one another. Or look up **Einstein-Poincaré synchronization**.

1.2 ■ WRISTWATCH TIME

Every observer agrees on the advance of wristwatch time.

Einstein said to Besso (initial quote): "Time cannot be absolutely defined . . ." Equation (1) exhibits this ambiguity: the laboratory time lapse, rocket time lapse, and wristwatch time lapse between two ticks of the stone's wristwatch *can all be different from one another.* But equation (1) tells us much more: It shows how any inertial observer whatsoever can use the space and time coordinate separations between ticks measured in her frame to calculate the unique **wristwatch time** τ, the time lapse between ticks recorded on the stone's wristwatch as it moves from Event 1 to Event 2.

Example of **wristwatch time** or **aging**

DEFINITION 3. Wristwatch time = aging

Equation (1) and Figure 1 show an example of the **wristwatch time** τ between two events, in this case the time lapse recorded on a wristwatch that is present at both events and travels uniformly between them. Wristwatch time is sometimes called **aging**, because it is the amount by which the wearer of the wristwatch gets older as she travels directly between this pair of events. Another common name for wristwatch time is **proper time**, which we do not use in this book.

We, the authors of this book, rate (1) as one of the greatest equations in physics, perhaps in all of science. Even the famous equation $E = mc^2$ is a child of equation (1), as Section 1.7 shows.

Truth be told, equation (1) is not limited to events along the path of a stone; it also applies to any pair of events in flat spacetime, no matter how large their coordinate separations in any one frame. In the general case, equation (1) is called the spacetime **interval** between these two events.

Definition: **interval**

DEFINITION 4. Interval

The spacetime **interval** is an expression whose inputs are the distance separation and time separation between a pair of events measured in an inertial frame. The term “interval” refers to the whole equation (1). There are three different possible outputs, three types of interval:

Case 1: Timelike interval, $\tau^2 > 0$ this section
Case 2: Spacelike interval, $\tau^2 < 0$ Section 1.3
Case 3: Lightlike interval, $\tau^2 = 0$ Section 1.4

These three categories span all possible relations between a pair of events in special relativity. When $(t_{2\text{lab}} - t_{1\text{lab}})^2$ is greater than $(s_{2\text{lab}} - s_{1\text{lab}})^2$, then we have the case we analyzed for two events that may lie along the path of a stone. We call this a **timelike interval** because the magnitude of the time part of the interval is greater than that of its space part.

What happens when $(s_{2\text{lab}} - s_{1\text{lab}})^2$ is greater than $(t_{2\text{lab}} - t_{1\text{lab}})^2$ in (1), so the interval is negative? We call this a **spacelike interval** because the magnitude of the space part of the interval is greater than that of its time part. In this case we interchange $(t_{2\text{lab}} - t_{1\text{lab}})^2$ and $(s_{2\text{lab}} - s_{1\text{lab}})^2$ to yield a positive quantity we call σ^2, whose different physical interpretation we explore in Section 1.3.

What happens when $(s_{2\text{lab}} - s_{1\text{lab}})^2$ is equal to $(t_{2\text{lab}} - t_{1\text{lab}})^2$ in (1), so the interval has the value zero? We call this a **null interval** or **lightlike interval**, as explained in Section 1.4.

Measure space and time separations in the same unit, which *you* choose.

Note: All separations in (1) must be measured in the same unit; otherwise they cannot appear as separate terms in the same equation. But we are free to choose the common unit: it can be **years** (of time) and **light-years** (of distance). A light-year is the distance light travels in a vacuum in one year. Or we can use **meters** (of distance) along with **light-meters** (of time). A light-meter of time is the time it takes light to travel one meter in a vacuum—about 3.34×10^{-9} second. Alternative expressions for light-meter are **meter of light-travel time** or simply **meter of time**.

Speed of light equals unity.

Distance and time expressed in the same unit? Then the *speed of light* has the value unity, with *no* units:

$$c = \frac{1 \text{ light-year of distance}}{1 \text{ year of time}} = \frac{1 \text{ meter of distance}}{1 \text{ light-meter of time}} = 1 \qquad (2)$$

Why the letter c? The Latin word *celeritas* means “swiftness” or “speed.”

Stone’s speed: a fraction of light speed

So much for the speed of *light*. How do we measure the speed of a *stone* using space and time separations between ticks of its wristwatch? Typically the value of the stone’s speed depends on the reference frame with respect to which we measure these separations. In the top panel of Figure 1, its speed in

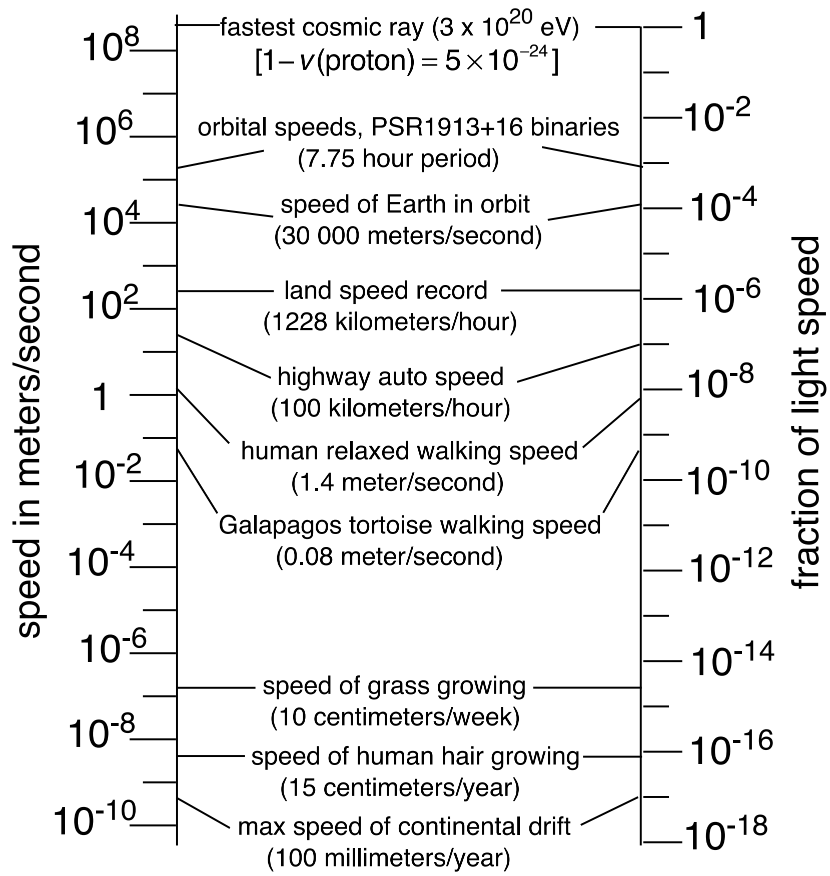


FIGURE 2 The speed ladder. Some typical speeds encountered in Nature.

the laboratory frame is $v_{\text{lab}} = (s_{2\text{lab}} - s_{1\text{lab}})/(t_{2\text{lab}} - t_{1\text{lab}})$. In the bottom panel, its speed in the rocket frame is $v_{\text{rocket}} = (s_{2\text{rocket}} - s_{1\text{rocket}})/(t_{2\text{rocket}} - t_{1\text{rocket}})$. Typically the values of these two speeds differ from one another. However, both values are less than one. Figure 2 samples the range of speeds encountered in Nature.

Equation (1) is so important that we use it to define **flat spacetime.**

Definition:
flat spacetime

DEFINITION 5. Flat spacetime

Flat spacetime is a spacetime region in which equation (1) is valid for every pair of events.

The interval in equation (1) has an important property that will follow us through special and general relativity: it has the same value when calculated using either laboratory or rocket coordinates. We say that wristwatch time is an **invariant quantity**.

Sample Problems 1. Wristwatch Times

PROBLEM 1A
An unpowered rocket ship moves at constant speed to travel 3 light-years in 5 years, this time and distance measured in the rest frame of our Sun. What is the time lapse for this trip recorded on a clock carried with the spaceship?

SOLUTION 1A
The two events that start and end the spaceship's journey are separated in the Sun frame by $s_{2\mathrm{Sun}} - s_{1\mathrm{Sun}} = 3$ light-years and $t_{2\mathrm{Sun}} - t_{1\mathrm{Sun}} = 5$ years. Equation (1) gives the resulting wristwatch time:

$$\tau^2 = 5^2 - 3^2 = 25 - 9 = 16 \text{ years}^2 \qquad (3)$$
$$\tau = 4 \text{ years}$$

which is *less* than the time lapse measured in the Sun frame.

PROBLEM 1B
An elementary particle created in the target of a particle accelerator arrives 5 meters of time later at a detector 4 meters from the target, as measured in the laboratory. The wristwatch of the elementary particle records what time between creation and detection?

SOLUTION 1B
The events of creation and detection are separated in the laboratory frame by $s_{2\mathrm{lab}} - s_{1\mathrm{lab}} = 4$ meters and $t_{2\mathrm{lab}} - t_{1\mathrm{lab}} = 5$ meters of time. Equation (1) tells us that

$$\tau^2 = 5^2 - 4^2 = 25 - 16 = 9 \text{ meters}^2 \qquad (4)$$
$$\tau = 3 \text{ meters}$$

Again, the wristwatch time for the particle is less than the time recorded in the laboratory frame.

PROBLEM 1C
In Problem 1B the two events are separated by a distance of 4 meters, which means that it takes light 4 meters of light-travel time to move between them. But Solution 1B says that the particle's wristwatch records only 3 meters of time as the particle moves from the first to the second event. Does this mean that the particle travels faster than light?

SOLUTION 1C
This difficulty is common in relativity. The phrase "time between two events" has no unique value (initial quote of this chapter). The *time* depends on *which clock* measures the time, in this case either the laboratory clocks, which measure laboratory time separation $t_{2\mathrm{lab}} - t_{1\mathrm{lab}}$, or the particle's wristwatch, which measures lapsed wristwatch time τ. Equation (1) already warns us that these two measures of time may not have the same value. Indeed a particle that moves faster and faster, covering a greater and greater distance $s_{2\mathrm{lab}} - s_{1\mathrm{lab}}$ in the same laboratory time lapse $t_{2\mathrm{lab}} - t_{1\mathrm{lab}}$, records a wristwatch time τ that gets smaller and smaller (Sample Problems 2), finally approaching—as a limit—the value zero, in which case a light flash has replaced the particle (Section 1.4). But for a particle with mass, the distance $s_{2\mathrm{lab}} - s_{1\mathrm{lab}}$ it travels in the laboratory frame is always less than the laboratory time $t_{2\mathrm{lab}} - t_{1\mathrm{lab}}$ that it takes the particle to move that distance. In other words, its laboratory speed will always be less than one, the speed of light. No particle can move faster than light moves in a vacuum. (Convince the scientific community that this statement is false, and your name will go down in history!)

DEFINITION 6. Invariant

Definition: **invariant**

Formally, a quantity is an **invariant** when it keeps the same value under some transformation. Equation (1) shows the interval between any pair of events along the path of a free stone to have the same value when calculated using coordinate separations in any inertial frame. Transformations of coordinate separations between inertial frames are called **Lorentz transformations** (Section 1.10), so we say that the interval is a **Lorentz invariant**. However, the interval must also be an invariant under even more general transformations, not just Lorentz transformations, because all observers—not just those in inertial frames—will agree on the stone's wristwatch time lapse between any two given events. As a consequence, we most often drop the adjective *Lorentz* and use just the term **invariant**.

Fuller Explanation: *Spacetime Physics,* Chapter 1, Spacetime: Overview

Sample Problems 2. Speeding to Andromeda

At approximately what constant speed v_{Sun} with respect to our Sun must a spaceship travel so that its occupants age only 1 year during a trip from Earth to the Andromeda galaxy? Andromeda lies 2 million light-years distant from Earth in the Sun's rest frame.

SOLUTION The word *approximately* in the statement of the problem tells us that we may make some assumptions. We assume that a single inertial frame can stretch all the way from Sun to Andromeda, so special relativity applies. Equation (1) leads us to predict that the speed v_{Sun} of the spaceship measured in the Sun frame is very close to unity, the speed of light. That allows us to set $(1 + v_{\text{Sun}}) \approx 2$ in the last of the following steps:

$$\begin{aligned}
\tau^2 &= (t_{2\text{Sun}} - t_{1\text{Sun}})^2 - (s_{2\text{Sun}} - s_{1\text{Sun}})^2 \qquad (5)\\
&= (t_{2\text{Sun}} - t_{1\text{Sun}})^2 \left[1 - \left(\frac{s_{2\text{Sun}} - s_{1\text{Sun}}}{t_{2\text{Sun}} - t_{1\text{Sun}}}\right)^2\right]\\
&= (t_{2\text{Sun}} - t_{1\text{Sun}})^2 \left(1 - v_{\text{Sun}}^2\right)\\
&= (t_{2\text{Sun}} - t_{1\text{Sun}})^2 (1 - v_{\text{Sun}})(1 + v_{\text{Sun}})\\
&\approx 2\,(t_{2\text{Sun}} - t_{1\text{Sun}})^2 (1 - v_{\text{Sun}})
\end{aligned}$$

Equate the first and last expressions in (5) to obtain

$$1 - v_{\text{Sun}} \approx \frac{\tau^2}{2\,(t_{2\text{Sun}} - t_{1\text{Sun}})^2} \qquad (6)$$

IF the spaceship speed v_{Sun} is very close to the speed of light, THEN the Sun-frame time for the trip to Andromena is very close to the time that light takes to make the trip: 2 million years. Substitute this value for $t_{2\text{Sun}} - t_{1\text{Sun}}$ and also demand that the wristwatch time on the spaceship (the aging of the occupants during their trip) be $\tau = 1$ year. The result is

$$\begin{aligned}
1 - v_{\text{Sun}} &\approx \frac{1\,\text{year}^2}{2 \times 4 \times 10^{12}\,\text{year}^2} \qquad (7)\\
&= \frac{10^{-12}}{8} = 1.25 \times 10^{-13}
\end{aligned}$$

Equation (7) expresses the result in sensible scientific notation. However, your friends may be more impressed if you report the speed as a fraction of the speed of light: $v_{\text{Sun}} = 0.999\,999\,999\,999\,875$. This result justifies our assumption that v_{Sun} is close to unity. *Additional question*: What is the *distance* $(s_{2\text{rocket}} - s_{1\text{rocket}})$ between Earth and Andromeda measured in the rocket frame?

1.3 ■ RULER DISTANCE

Everyone agrees on the ruler distance between two events.

Two firecrackers explode one meter apart and *at the same time,* as measured in a given inertial frame: in *this* frame the explosions are **simultaneous.** No stone—not even a light flash—can travel the distance between these two explosions in the zero time available in this frame. Therefore equation (1) cannot give us a value of the wristwatch time between these two events.

Use simultaneous explosions to measure length of a rod.

Simultaneous explosions are thus useless for measuring time. But they are perfect for measuring length. *Question:* How do you measure the length of a rod, whether it is moving or at rest in, say, the laboratory frame? *Answer:* Set off two firecrackers at opposite ends of the rod and *at the same time* $(t_{2\text{lab}} - t_{1\text{lab}} = 0)$ in that frame. Then *define* the rod's length in the laboratory frame as the *distance* $(s_{2\text{lab}} - s_{1\text{lab}})$ between this pair of explosions simultaneous in that frame.

Relativity of simultaneity

Special relativity warns us that another observer who flies through the laboratory typically does *not* agree that the two firecrackers exploded at the same time as recorded on her rocket clocks. This effect is called the **relativity of simultaneity**. The relativity of simultaneity is the bad news (and for many people the most difficult idea in special relativity). But here's the good news: All inertial observers, whatever their state of relative motion, can calculate the distance σ between explosions as recorded in the frame in which they do occur simultaneously. This calculation uses Case 2 of the interval (Definition 4):

Spacelike interval σ

$$\sigma^2 \equiv -\tau^2 = (s_{2\text{lab}} - s_{1\text{lab}})^2 - (t_{2\text{lab}} - t_{1\text{lab}})^2 \qquad \text{(spacelike interval)} \quad (8)$$
$$= (s_{2\text{rocket}} - s_{1\text{rocket}})^2 - (t_{2\text{rocket}} - t_{1\text{rocket}})^2$$

The Greek letter *sigma*, σ, in (8)—equivalent to the Roman letter s—is the length of the rod defined as the distance between explosions at its two ends measured in a frame in which these explosions are simultaneous.

Equation (8) does not define a different kind of interval; it is merely shorthand for the equation for Case 2 in Definition 4 in which $\tau^2 < 0$.

Actually, we do not need a rod or ruler to make use of this equation (though we keep *ruler* as a label). Take any two events for which $\tau^2 < 0$. Then there exists an inertial frame in which these two events occur at the same time; we use this frame to define the **ruler distance** σ between these two events:

DEFINITION 7. Ruler distance

The **ruler distance** σ between two events is the distance between these events measured by an inertial observer in whose frame the two events occur at the same time. Another common name for ruler distance is **proper distance**, which we do not use in this book.

Definition: **ruler distance**

Equation (8) tells us that every inertial observer can calculate the ruler distance between two events using the space and time separations between these events measured in his or her own frame.

Fuller Explanation: *Spacetime Physics,* Chapter 6, Regions of Spacetime

1.4 ■ LIGHTLIKE (NULL) INTERVAL

Everyone agrees on the null value of the interval between two events connected by a direct light flash that moves in a vacuum.

Now think of the case in which the lab-frame space separation $(s_{2\text{lab}} - s_{1\text{lab}})$ between two events is equal to the time separation $(t_{2\text{lab}} - t_{1\text{lab}})$ between them. In this case anything that moves uniformly between them must travel at the speed of light $v_{\text{lab}} = (s_{2\text{lab}} - s_{1\text{lab}})/(t_{2\text{lab}} - t_{1\text{lab}}) = 1$. Physically, only a direct light flash can move between this pair of events. We call the result a **lightlike interval**:

$$\tau^2 = -\sigma^2 = 0 = (s_{2\text{lab}} - s_{1\text{lab}})^2 - (t_{2\text{lab}} - t_{1\text{lab}})^2 \quad \text{(lightlike interval)} \quad (9)$$
$$= (s_{2\text{rocket}} - s_{1\text{rocket}})^2 - (t_{2\text{rocket}} - t_{1\text{rocket}})^2$$

Because of its zero value, the lightlike interval is also called the **null interval**.

DEFINITION 8. Lightlike (null) interval

A **lightlike interval** is the interval between two events whose space

Definition: **lightlike interval** or **null interval**

Sample Problems 3. Causation

Three events have the following space and time coordinates as measured in the laboratory frame in meters of distance and meters of time. All three events lie along the x-axis in the laboratory frame. (Temporarily suppress the subscript "lab" in this Sample Problem.)

Event A: $(t_{\mathrm{A}}, x_{\mathrm{A}}) = (2, 1)$
Event B: $(t_{\mathrm{B}}, x_{\mathrm{B}}) = (7, 4)$
Event C: $(t_{\mathrm{C}}, x_{\mathrm{C}}) = (5, 6)$

Classify the intervals between each pair of these events as timelike, lightlike, or spacelike:

(a) between events A and B
(b) between events A and C
(c) between events B and C

In each case say whether or not it is possible for one of the events in the pair (which one?) to cause the other event of the pair, and if so, by what possible means.

SOLUTION
The interval between events A and B is:

$$\tau^2 = (7-2)^2 - (4-1)^2 = 5^2 - 3^2 \qquad (10)$$
$$= 25 - 9 = +16$$

The time part is greater than the space part, so the interval between the events is *timelike*. Event A could have caused Event B, for example by sending a stone moving directly between them at a speed $v_{\mathrm{lab}} = 3/5$. (There are other possible ways for Event A to cause Event B, for example by sending a light flash that sets off an explosion between the two locations, with a fragment of the explosion reaching Event B at the scheduled time, and so forth. Our analysis says only that Event A *can* cause Event B, but it does not *force* Event A to cause Event B. Someone standing next to an object located at the x-coordinate of Event B could simply kick that object at the scheduled time of Event B.)

The interval between events A and C is:

$$\tau^2 = (5-2)^2 - (6-1)^2 = 3^2 - 5^2 \qquad (11)$$
$$= 9 - 25 = -16$$

The space part is greater than the time part, so the interval between the events is *spacelike*. Neither event can cause the other, because to do so an object would have to travel between them at a speed greater than that of light.

The interval between events B and C is:

$$\tau^2 = (7-5)^2 - (4-6)^2 = 2^2 - 2^2 \qquad (12)$$
$$= 4 - 4 = 0$$

The space part is equal to the time part, so the interval between the events is *lightlike*. Event C can cause Event B, but only by sending a direct light signal to it.

Challenge: How can we rule out the possibility that event B causes event A, or that event B causes event C? Would your answers to these questions be different if the same events are observed in some other frame in rapid motion with respect to the laboratory? (Answer in Exercise 1.)

separation and time separation are equal in every inertial frame. Only a direct light flash can connect these two events. Because these space and time separations are equal, the interval has the value zero, so is also called the **null interval**.

Comment 1. Einstein's derivation of special relativity
Divide both sides of (9) by $(t_{2,\mathrm{frame}} - t_{1,\mathrm{frame}})^2$, where "frame" is either "lab" or "rocket." The result tells us that the speed in any inertial frame is one, $v_{\mathrm{lab}} = v_{\mathrm{rocket}} = 1$. Einstein derived (9) starting with the *assumption* that the speed of light is the same in all inertial frames.

Fuller Explanation: *Spacetime Physics,* Chapter 6, Regions of Spacetime.

1.5 ■ WORLDLINE OF A WANDERING STONE; THE LIGHT CONE

A single curve tells all about the motion of our stone.

Grasp a stone in your hand and move it alternately in one direction, then in the opposite direction along the straight edge of your desk. Choose the $x_{\rm lab}$ axis along this line. Then the stone's motion is completely described by the function $x_{\rm lab}(t_{\rm lab})$. No matter how complicated this back-and-forth motion is, we can view it at a glance when we plot $x_{\rm lab}$ along the horizontal axis of a graph whose vertical axis represents the time $t_{\rm lab}$. Figure 3 shows such a curve, which we call a **worldline**.

Definition: **worldline**

DEFINITION 9. Worldline

A **worldline** is the path through spacetime taken by a stone or light flash. By Definition 3, the total wristwatch time (aging) along the worldline is the sum of wristwatch times between sequential events along the worldline from a chosen initial event to a chosen final event. The wristwatch time is an invariant; it has the same value when calculated using either laboratory or rocket coordinates. Therefore specification of a worldline requires neither coordinates nor the metric.

Comment 2. Plotting the worldline

Figure 3 shows a worldline plotted in laboratory coordinates. Typically a given worldline will look different when plotted in rocket coordinates. We plot a worldline in whatever coordinates we are using. Worldlines can be plotted in spacetime diagrams for both flat and curved spacetime.

In the worldline of Figure 3 the stone starts at initial event O. As time passes—as time advances upward in the diagram—the stone moves first to the right. Then the stone slows down, that is it covers less distance to the right per unit time, and comes to rest momentarily at event Z. (The vertical tangent to the worldline at Z tells us that the stone covers zero laboratory distance there: it is instantaneously at rest at Z.) Thereafter the stone accelerates to the left in space until it arrives at event P.

Limits on worldline slope

What possible future worldlines are available to the stone that arrives at event P? Any material particle must move at less than the speed of light. In other words, it travels less than one meter of distance in one meter of light-travel time. Therefore its future worldline must make an "angle with the vertical" somewhere between minus 45 degrees and plus 45 degrees in Figure 3, in which space and time are measured in the same units and plotted to the same scale. These limits on the slope of the stone's worldline—which apply to every event on every worldline—emerge as dashed lines from event P in Figure 3. These dashed lines are worldlines of light rays that move in opposite $x_{\rm lab}$-directions and cross at the event P. We call these crossed light rays a **light cone.** Figure 4 displays the cone shape.

Definition: **light cone**

DEFINITION 10. Light cone

The **light cone** of an event is composed of the set of all possible worldlines of light that intersect at that event and define its past and

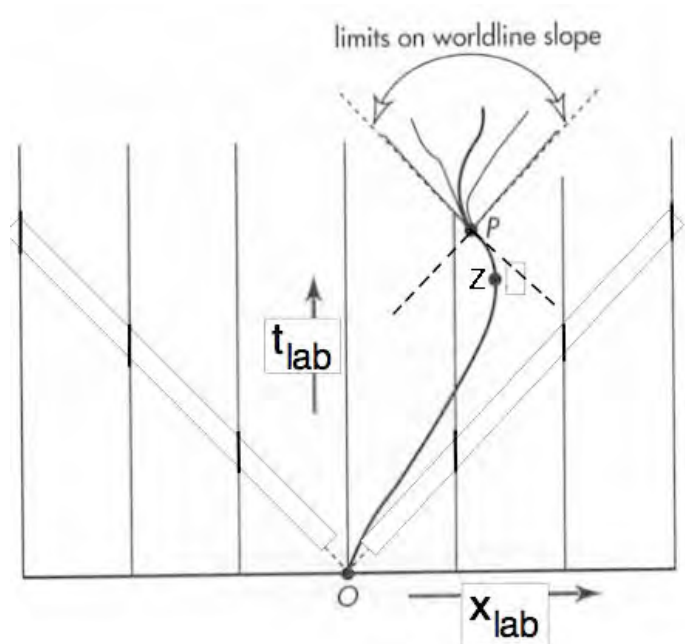


FIGURE 3 Curved **worldline** of a stone moving back and forth along a single straight spatial line in the laboratory. A point on this diagram, such as Z or P, combines x_{lab}-location (horizontal direction) with t_{lab}-location (vertical direction); in other words a point represents a spacetime *event*. The dashed lines through P are worldlines of light rays that pass through P. We call these crossed lines *the light cone of P*. For the cone shape, see Figure 4.

future (Figure 4). We also call it a light *cone* when it is plotted using one space dimension plus time, as in Figure 3, and when plotted using three space dimensions plus time—even though we cannot visualize the resulting four-dimensional spacetime plot.

THE LIGHT CONE AND CAUSALITY

. . the light cone provides a mathematical tool for the analysis of [general relativity] additional to the usual tools of metric geometry. We believe that this tool still remains to be put to full use, and that causality is the physical principle which will guide this future development.

—Robert W. Fuller and John Archibald Wheeler

More complete explanation: *Spacetime Physics,* Chapter 5, Trekking Through Spacetime

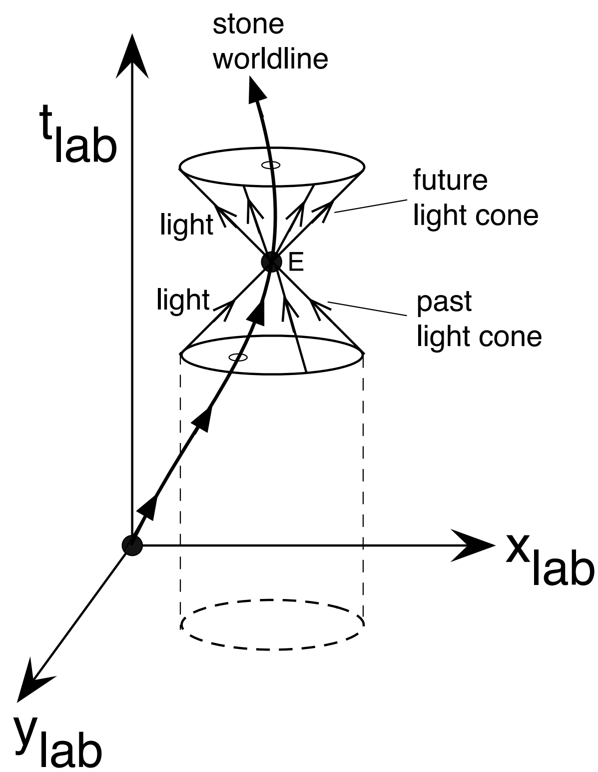


FIGURE 4 *Light cone* of Event E that lies on the worldline of a stone, plotted for two space dimensions plus time. The light cone consists of the upward-opening future light cone traced out by the expanding circular light flash that the stone emits at Event E, plus the downward-opening past light cone traced out by a contracting circular light flash that converges on Event E.

1.6 ■ THE TWIN “PARADOX” AND THE PRINCIPLE OF MAXIMAL AGING

The Twin Paradox leads to a definition of natural motion.

To get ready for curved spacetime (whatever that means), look more closely at the motion of a free stone in *flat spacetime* (Definition 5), where special relativity correctly describes motion.

Twin Paradox predicts motion of a stone.

A deep description of motion arises from the famous **Twin Paradox.** One twin—say a boy—relaxes on Earth while his fraternal twin sister frantically travels to a distant star and returns. When the two meet again, the stay-at-home brother has aged more than his traveling sister. (To predict this outcome, extend Sample Problem 1A to include return of the traveler to the point of origin.) Upon being reunited, the “twins” no longer look similar: the traveling sister is *younger*: she has aged less than her stay-at-home brother. Very strange! But (almost) no one who has studied relativity doubts the difference in age, and every minute of every day somewhere on Earth a measurement with a fast-moving particle verifies it.

Being at rest is one *natural* motion.

Which twin has the motion we can call *natural*? Isaac Newton has a definition of natural motion. He would say, "A twin at rest tends to remain at rest." So it is the stay-at-home twin who moves in the natural way. In contrast, the out-and-back twin suffers the acceleration required to change her state of motion, from outgoing motion to incoming motion, so the twins can meet again in person. At least at her turnaround, the motion of the traveling twin is forced, *not natural.*

Moving uniformly is another *natural* motion.

Viewed from the second, relatively moving, inertial frame of the twin sister, the stay-at-home boy initially moves away from her with constant speed in a straight line. Again, his motion is *natural.* Newton would say, "A twin in uniform motion tends to continue this motion at constant speed in a straight line." So the motion of the stay-on-Earth twin is also natural from the viewpoint of his sister's frame in uniform relative motion—or from the viewpoint of any frame moving uniformly with respect to the original frame. In *any* such frame, the time lapse on the wristwatch of the stay-at-home twin can be calculated from the interval (1).

But there *is* a difference between the stay-at-home brother on Earth and the sister: She moves outward to a star, *then turns around* and returns to her Earthbound brother. So when her trip is over, everyone must agree: It is the brother who follows "natural" motion from parting event to reunion event. And it is the stay-at-home brother—whose wristwatch records the greater elapsed time—who **ages** the most.

Natural motion: Maximal wristwatch time.

The lesson we draw from the Twin Paradox in flat spacetime is that *natural* motion is the motion that maximizes the wristwatch time between *any* pair of events along its path. Now we can state the **Principle of Maximal Aging in flat spacetime**.

Definition: **Principle of Maximal Aging**

DEFINITION 11. The Principle of Maximal Aging (flat spacetime)

The **Principle of Maximal Aging** states that the worldline a free stone follows between a pair of events in flat spacetime is the worldline for which the wristwatch time is a maximum compared with every possible alternative worldline between these events. The free stone follows the worldline of *maximal aging* between these two events.

?

Objection 1. *Why should I believe the Principle of Maximal Aging? Newton never talks about this weird idea! What does this so-called "Principle" mean, anyway?*

!

Response: For now the Principle of Maximal Aging is simply a restatement of the observation that in flat spacetime a free stone follows a straight worldline. It repeats Newton's First Law of Motion: A free stone at rest or in motion maintains that condition. Why bother? Because general relativity revises and extends the Principle of Maximal Aging to predict the motion of a free stone in curved spacetime.

Objection 2. *Wait! Have you really resolved the Twin Paradox? Both the twin sister and the twin brother sees his or her twin moving away, then moving back. Motion is relative, remember? The view of each twin is symmetrical, not only during the outward trip but also during the return trip. There is no difference between them. The experience of the two twins is identical; you cannot wriggle out of this essential symmetry! You have failed to explain why their wristwatches have different readings when they reunite.*

Nice point. But you forget that the experience of the two twins is *not* identical. Fill in details of the story: When the twin sister arrives at the distant star and reverses her starship's direction of motion, that reversal throws her against the forward bulkhead. Ouch! She starts home with a painful lump on the right side of her forehead. Then when her ship slows down so she can stand next to her stay-at-home brother, she forgets her seat belt again. *Result:* a second painful lump, this time on the left side of her forehead. In contrast, her brother remains relaxed and uninjured during their entire separation. When the twins stand side by side, can *each* of them tell *which twin* has gone to the distant star? Of course! *More: Every passing observer*—whatever his or her speed or direction of motion—sees and reports the difference between the twins: "injured sister; smiling brother." *Everyone* agrees on this difference. No contradiction and no confusion. "Paradox" resolved.

Comment 3. The Quintuplet "Paradox"

In the last sentence of Definition 11, The Principle of Maximal Aging, notice the word "every" in the phrase "is a maximum compared with *every* alternative path...between the given initial and final events." We are not just talking twins here, but triplets, quadruplets, quintuplets—indeed endless multiple births. *Example,* Figure 5: One quintuplet—**Quint** #1—follows the worldline of maximal aging between the two anchoring events by moving uniformly between them. Each of the other quints also starts from the same Initial Event A and ends at the same Final Event B, but follows a different alternative worldline—changes velocity—between initial and final events. When all the quints meet at the final event, all four traveling quints are younger than their uniformly-moving sibling, but typically by different amounts. *Every traveler, #2 through #5, who varies velocity between the two end-events is younger than its uniformly-moving sibling, Quint #1.* The Principle of Maximal Aging singles out one worldline among the limitless number of alternative worldlines between two end-events and demands that the free stone follow **this** worldline—and no other.

An infinite number of alternative worldlines: the free stone chooses one.

QUERY 1. Analyze the Quintuplet Paradox

Answer the following questions about the Quintuplet Paradox illustrated in Figure 5.

A. Which of the five quints ages the *most* between end-events A and B? (Trick question!)
B. Which of the five quints ages the *least* between end-events A and B?
C. List the numbered worldlines in order, starting with the worldline along which the aging is the *least* and ending with the worldline along which the aging is the *most*.

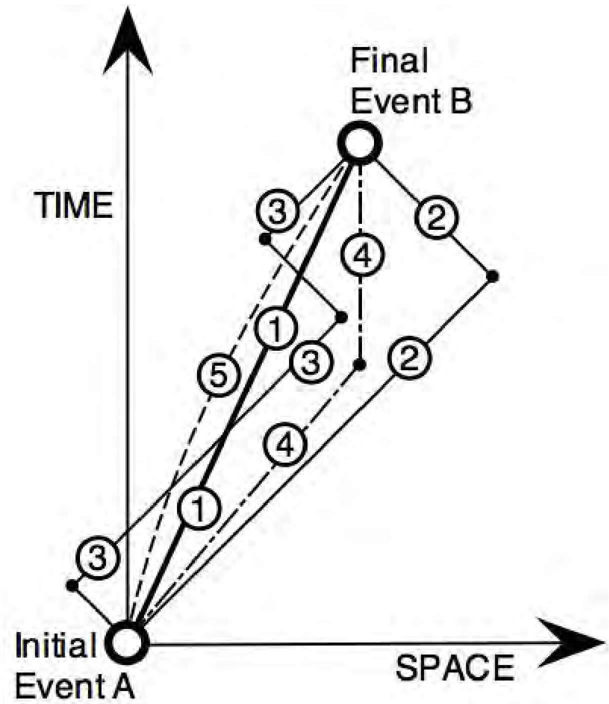


FIGURE 5 The Quintuplet Paradox: Five alternative worldlines track the motion of five different quintuplets (**quints**) between Initial Event A and Final Event B along a spatial straight line. Quint #1 follows the (thick) worldline of maximal aging between A and B. Quint #2 moves along the (thin) worldline at 0.999 of the speed of light outward and then back again. Quint #3 follows a worldline (also a thin line) at the same *speed* as #2, but with three reversals of direction. Quint #4 shuffles (dot-dash line) to the spatial position of Final Event B, then relaxes there until her siblings join her at Event B. The (dashed) worldline of Quint #5 hugs worldline #1—the worldline of Maximal Aging—but does not quite follow it.

D. True or false? If the dashed worldline of Quint #5 skims close enough to that of Quint #1—while still being separate from it—then Quint #5 will age the same as Quint #1 between end-events A and B.

E. *Optional:* Suppose we view the worldlines of Figure 5 with respect to a frame in which Event A and Event B occur at the same spatial location. Whose *inertial* rest frame does this correspond to? Will your answers to Items A through D be different in this case?

Fuller Explanation: Twin “paradox:” *Spacetime Physics,* Chapter 4, Section 4.6.

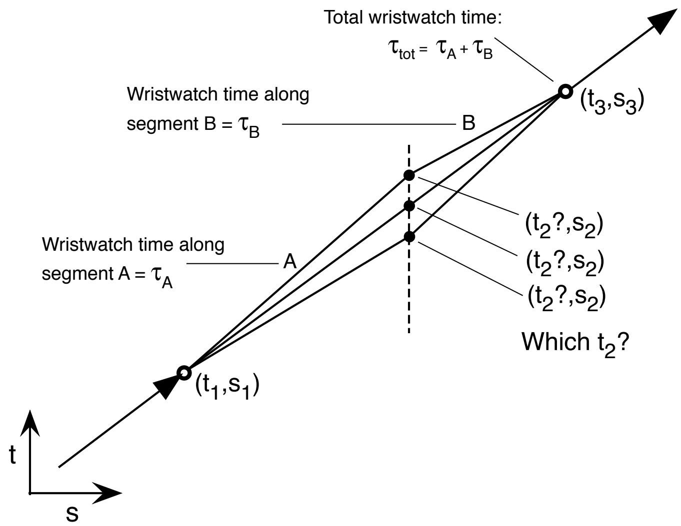


FIGURE 6 Figure for the derivation of the energy of a stone. Examine two adjacent segments, A and B, along an extended worldline plotted in, say, the laboratory frame. Choose three events at the endpoints of these two segments with coordinates (t_1, s_1), (t_2, s_2), and (t_3, s_3). All coordinates are fixed except t_2. Vary t_2 to find the maximum value of the total aging $\tau_{\rm tot}$ (Principle of Maximal Aging). *Result:* an expression for the stone's energy E.

1.7 ■ ENERGY IN SPECIAL RELATIVITY

The Principle of Maximal Aging tells us the energy of a stone.

Here is a modern translation (from Latin) of Isaac Newton's famous First Law of Motion:

Newton's First Law of motion

> **Newton's first law of motion:** Every body perseveres in its state of being at rest or of moving uniformly straight forward except insofar as it is compelled to change its state by forces impressed.

Validity of Newton's First Law in special relativity . . .

. . . leads to relativistic expression for energy.

In modern terminology, Newton's First Law says that, as measured in an inertial frame in flat spacetime, a free stone moves along a *straight worldline*, that is with constant speed along a straight path in space. We assumed the validity of Newton's First Law in defining the inertial frame (Definition 1, Section 1.1). In the present section the Principle of Maximal Aging again verifies this validity of the First Law. *Extra surprise!* This process will help us to derive the relativistic expression for the stone's energy E.

Figure 6 illustrates the method: Consider two adjacent segments, A and B, of the stone's worldline with fixed events at the endpoints. Vary t_2 of the middle event to find the value that gives a maximum for the total wristwatch time $\tau_{\rm tot}$ along the adjacent segments. Now the step-by-step derivation:

1. The wristwatch time between the first and second events along the worldline is the square root of the interval between them:

$$\tau_{\mathrm{A}} = \left[(t_2 - t_1)^2 - (s_2 - s_1)^2\right]^{1/2} \qquad (13)$$

To prepare for the derivative that leads to maximal aging, differentiate this expression with respect to t_2. (All other coordinates of the three events are fixed.)

$$\frac{d\tau_{\mathrm{A}}}{dt_2} = \frac{t_2 - t_1}{\left[(t_2 - t_1)^2 - (s_2 - s_1)^2\right]^{1/2}} = \frac{t_2 - t_1}{\tau_{\mathrm{A}}} \qquad (14)$$

2. The wristwatch time between the second and third events along the worldline is the square root of the interval between them:

$$\tau_{\mathrm{B}} = \left[(t_3 - t_2)^2 - (s_3 - s_2)^2\right]^{1/2} \qquad (15)$$

Again, to prepare for the derivative that leads to extremal aging, differentiate this expression with respect to t_2:

$$\frac{d\tau_{\mathrm{B}}}{dt_2} = -\frac{t_3 - t_2}{\left[(t_3 - t_2)^2 - (s_3 - s_2)^2\right]^{1/2}} = -\frac{t_3 - t_2}{\tau_{\mathrm{B}}} \qquad (16)$$

3. The total wristwatch time τ_{tot} from event #1 to event #3—the total aging between these two events—is the sum of the wristwatch time τ_{A} between the first two events plus the wristwatch time τ_{B} between the last two events:

$$\tau_{\mathrm{tot}} = \tau_{\mathrm{A}} + \tau_{\mathrm{B}} \qquad (17)$$

Principle of Maximal Aging finds time t_2 for middle event.

4. Now ask: At what intermediate t_2 will a free stone pass the intermediate point in space s_2 and emit the second flash #2? Answer by using the Principle of Maximal Aging: The time t_2 will be such that the total aging τ_{tot} in (17) is a maximum. To find this maximum take the derivative of τ with respect to t_2 and set the result equal to zero. Add the final expressions (14) and (16) to obtain:

$$\frac{d\tau_{\mathrm{tot}}}{dt_2} = \frac{t_2 - t_1}{\tau_{\mathrm{A}}} - \frac{t_3 - t_2}{\tau_{\mathrm{B}}} = 0 \qquad (18)$$

Quantity whose value is the same for adjoining segments

6. In equation (18) the time $(t_2 - t_1)$ is the lapse of laboratory time for the stone to traverse segment A. Call this time t_{A}. The time $(t_3 - t_2)$ is the lapse of laboratory time for the stone to traverse segment B. Call this time t_{B}. Then rewrite (18) in the simple form

$$\frac{t_{\mathrm{A}}}{\tau_{\mathrm{A}}} = \frac{t_{\mathrm{B}}}{\tau_{\mathrm{B}}} \qquad (19)$$

This result yields a maximum τ_{tot}, *not* a minimum; see Exercise 4.

7. We did not say *which* pair of adjoining segments along the worline we were talking about, so equation (19) must apply to *every* pair of adjoining segments *anywhere* along the path. Suppose that there are three such adjacent segments. If the value of the expression is the same for, say, the first and second segments and also the same for the second and third segments, then it must be the same for the first and third segments. Continue in this way to envision a whole series of adjoining segments, labeled A, B, C, D,..., for each of which equation (19) applies, leading to the set of equations

$$\frac{t_{\mathrm{A}}}{\tau_{\mathrm{A}}} = \frac{t_{\mathrm{B}}}{\tau_{\mathrm{B}}} = \frac{t_{\mathrm{C}}}{\tau_{\mathrm{C}}} = \frac{t_{\mathrm{D}}}{\tau_{\mathrm{D}}} \to \frac{dt_{\mathrm{lab}}}{d\tau} \tag{20}$$

where all coordinate values are given in the laboratory frame.

Differences shrink to differentials

Comment 4. Differences to differentials
The last step, with the arrow, in (20) is a momentous one. We take the calculus limit by shrinking to differentials—infinitesimals—all the differences in physical quantities. In Figure 6, for example, segments A and B shrink to infinitesimals. Why is this step important? Because in general relativity, curvature of spacetime means that relations between adjacent events are described accurately only when *adjacent* events are differentially close to one another. If they are far apart, the two events may be in regions of different spacetime curvature.

What does the result (20) mean? We now show that $dt_{\mathrm{lab}}/d\tau$ in (20) is the expression for energy per unit mass of a free stone in the laboratory frame. The differential form of (1) yields:

$$d\tau^2 = dt_{\mathrm{lab}}^2 - ds_{\mathrm{lab}}^2 = dt_{\mathrm{lab}}^2\left(1 - ds_{\mathrm{lab}}^2/dt_{\mathrm{lab}}^2\right) = dt_{\mathrm{lab}}^2\left(1 - v_{\mathrm{lab}}^2\right) \tag{21}$$

Combine (20) with (21):

$$\frac{dt_{\mathrm{lab}}}{d\tau} = \frac{1}{\left(1 - v_{\mathrm{lab}}^2\right)^{1/2}} \tag{22}$$

Working in a single inertial frame, we have just found that $dt/d\tau$ is unchanging along the worldline of a free stone, which by Definition 11 is the worldline of maximal aging. It follows that v_{lab} is constant. Hence the Principle of Maximal Aging leads to the result that in flat spacetime the free stone moves at constant speed. (The derivation of relativistic momentum in Section 1.8 shows that the free stone's *velocity* is also constant, so that it moves along a straight worldline in every inertial frame.)

We show below that at low speeds (22) reduces to Newton's expression for kinetic energy plus rest energy, all divided by the stone's mass m. This supports our decision to call the expression in (22) the energy per unit mass of the stone:

$$\frac{E_{\text{lab}}}{m} = \frac{dt_{\text{lab}}}{d\tau} = \frac{1}{(1 - v_{\text{lab}}^2)^{1/2}} = \gamma_{\text{lab}} \qquad (23)$$

The last expression in (23) introduces a symbol—Greek lower case gamma—that we use to simplify later equations.

$$\gamma_{\text{lab}} \equiv \frac{1}{\left(1 - v_{\text{lab}}^2\right)^{1/2}} \qquad (24)$$

We call E_{lab}/m a **constant of motion** because the free stone's energy does not change as it moves in the laboratory frame. This may seem trivial for a stone that moves with constant speed in a straight line. In general relativity, however, we will find an "energy" that is a constant of motion for a free stone in orbit around a center of gravitational attraction.

We applied the Principle of Maximal Aging to motion in the laboratory frame. An almost identical derivation applies in the rocket frame. Coordinates of the initial and final events will differ from those in Figure 6, but the result will still be that $dt_{\text{rocket}}/d\tau$ is constant along the free stone's worldline:

$$\frac{E_{\text{rocket}}}{m} = \frac{dt_{\text{rocket}}}{d\tau} = \frac{1}{(1 - v_{\text{rocket}}^2)^{1/2}} = \gamma_{\text{rocket}} \qquad (25)$$

Typically the value of the energy will be different in different inertial frames. We expect this, because the speed of a stone is not necessarily the same in different frames.

Equations (23) and (25) tell us that the energy of a stone in a given inertial frame increases without limit when the stone's speed approaches the value one, the speed of light, in that frame. Therefore the speed of light is the limit of the speed of a stone—or of any particle with mass—measured in any inertial frame. The other limit of (23) is a stone at rest in the laboratory. In this case, equation (23) reduces to

$$E_{\text{lab}} = m \qquad \text{(when speed of stone } v_{\text{lab}} = 0\text{)} \qquad (26)$$

We express m, the mass of the stone, in units of energy. If you insist on using conventional units, such as joules for energy and kilograms for mass, then a conversion factor c^2 intrudes into our simple expression. The result is the most famous equation in all of physics:

$$E_{\text{lab,conv}} = m_{\text{conv}} c^2 \qquad \text{(when speed of stone } v_{\text{lab}} = 0\text{)} \qquad (27)$$

Here the intentionally-awkward subscript "conv" means "conventional units."

Rest energy

Equations (26) and (27) both quantify the *rest energy* of a stone; both tell us

Sample Problems 4. Energy Magnitudes

PROBLEM 4A

The "speed ladder" in Figure 2 shows that the fastest wheeled vehicle moves on land at a speed approximately $v \approx 10^{-6}$. The kinetic energy of this vehicle is what fraction of its rest energy?

SOLUTION 4A

For such an "everyday" speed, the approximation on the right side of equation (28) should be sufficiently accurate. Then $v^2 \approx 10^{-12}$ and approximate equation (28) tells us that:

$$\frac{\text{kinetic energy}}{\text{rest energy}} = \frac{mv^2}{2m} = \frac{v^2}{2} \approx 5 \times 10^{-13} \quad (29)$$

PROBLEM 4B

With what speed v must a stone move so that its kinetic energy equals its rest energy?

SOLUTION 4B

This problem requires relativistic analysis. Equation (23) gives total energy and (26) gives rest energy. Kinetic energy is the difference between the two:

$$\frac{E_{\text{lab}} - m}{m} = \frac{1}{(1-v^2)^{1/2}} - 1 = 1 \quad (30)$$

from which

$$1 - v^2 = \frac{1}{2^2} = \frac{1}{4} \quad (31)$$

so that

$$v = \left(\frac{3}{4}\right)^{1/2} = 0.866 \quad (32)$$

This speed is a fraction of the speed of light, which means that $v_{\text{conv}} = 0.866 \times 3.00 \times 10^8$ meters/second $= 2.60 \times 10^8$ meters/second.

PROBLEM 4C

Our Sun radiates 3.86×10^{26} watts of light. How much mass does it convert to radiation every second?

SOLUTION 4C

This problem provides exercise in converting units. One watt is one joule/second. The units of energy are the units of (force $\times$ distance) or (mass $\times$ acceleration $\times$ distance). Therefore the units of joule are kilogram-meter^2/second^2. From (27):

$$\begin{aligned} m &= \frac{E_{\text{conv}}}{c^2} \quad (33) \\ &= \frac{3.86 \times 10^{26}\ \text{kilogram-meters}^2/\text{second}^2}{(3.00 \times 10^8\ \text{meters/second})^2} \\ &\approx 4.3 \times 10^9\ \text{kilograms} \\ &\approx 4.3 \times 10^6\ \text{metric tons} \end{aligned}$$

This is the mass—a few million metric tons—that our Sun, a typical star, converts into radiation every second.

that mass itself is a treasure trove of energy. On Earth, nuclear reactions release less than one percent of this available energy. In contrast, a particle-antiparticle annihilation can release *all* of the mass of the combining particles in the form of radiant energy (gamma rays).

At everyday speeds, the expression for E_{lab} in (23) reduces to an expression that contains Newton's kinetic energy. How do we get to Newton's case? Simply ask: How fast do things move around us in our everyday lives? At this writing, the fastest speed achieved by a wheeled vehicle on land is 1228 kilometers per hour (Figure 2), which is 763 miles per hour or 280 meters per second. As a fraction of light speed, this vehicle moves at $v = 9.3 \times 10^{-7}$ (no units). For such a small fraction, we can use a familiar approximation (inside the front cover):

$$\begin{aligned} E_{\text{lab}} &= \frac{m}{(1 - v_{\text{lab}}^2)^{1/2}} = m\left(1 - v_{\text{lab}}^2\right)^{-1/2} \approx m\left(1 + \frac{v_{\text{lab}}^2}{2}\right) \quad (28) \\ &\approx m + \frac{1}{2}mv_{\text{lab}}^2 = m + (KE)_{\text{Newton}} \qquad (v_{\text{lab}} \ll 1) \end{aligned}$$

You can verify that the approximation is highly accurate when v_{lab} has the value of the land speed record—and is an even better approximation for the everyday speeds of a bicycle or football. The final term in (28) is Newton's (low speed) expression for the kinetic energy of the stone. The first term is the rest energy of the stone, equation (26).

We can also separate the relativistic expression for energy into rest energy and kinetic energy. Define the relativistic kinetic energy of a stone in any frame with the equation

$$KE \equiv E - m = m(\gamma - 1) \qquad \text{(any frame, any speed)} \qquad (34)$$

Comment 5. Deeper than Newton?
Newton's First Law of Motion, quoted at the beginning of this section, was his brilliant assumption. In the present section we have derived this result using the Principle of Maximal Aging. Is our result deeper than Newton's? We think so, because the Principle of Maximal Aging has wider application than special relativity. It informs our predictions for the motion of a stone around both the non-spinning and the spinning black hole. Deep indeed!

Fuller Explanation: Energy in flat spacetime: *Spacetime Physics,* Chapter 7, Momenergy.

1.8 ■ MOMENTUM IN SPECIAL RELATIVITY

The interval plus the Principle of Maximal Aging give us an expression for the linear momentum of a stone.

To derive the relativistic expression for the momentum of a stone, we use a method similar to that for the derivation of energy in Section 1.7. Figure 7 corresponds to Figure 6, which we used to derive the stone's energy. Momentum has components in all three space directions; first we derive its x_{lab} component, which we write as $p_{x,\text{lab}}$. In the momentum case the time t_2 for the intermediate flash emission is *fixed,* while we vary the space coordinate s_2 of this intermediate event to find the location that yields maximum wristwatch time between initial and final events. We ask you to carry out this derivation in the exercises. The result is a second expression whose value is constant for a free stone in either the laboratory frame or the rocket frame:

$$\frac{p_{x,\text{lab}}}{m} = \frac{dx_{\text{lab}}}{d\tau} = \frac{v_{x,\text{lab}}}{(1 - v_{\text{lab}}^2)^{1/2}} = \gamma_{\text{lab}}\, v_{x,\text{lab}} \qquad (35)$$

$$\frac{p_{x,\text{rocket}}}{m} = \frac{dx_{\text{rocket}}}{d\tau} = \frac{v_{x,\text{rocket}}}{(1 - v_{\text{rocket}}^2)^{1/2}} = \gamma_{\text{rocket}}\, v_{x,\text{rocket}} \qquad (36)$$

where v_{lab} and v_{rocket} are each constant in the respective frame, and γ was defined in (24). Expressions for the y_{lab} and z_{lab} components of momentum

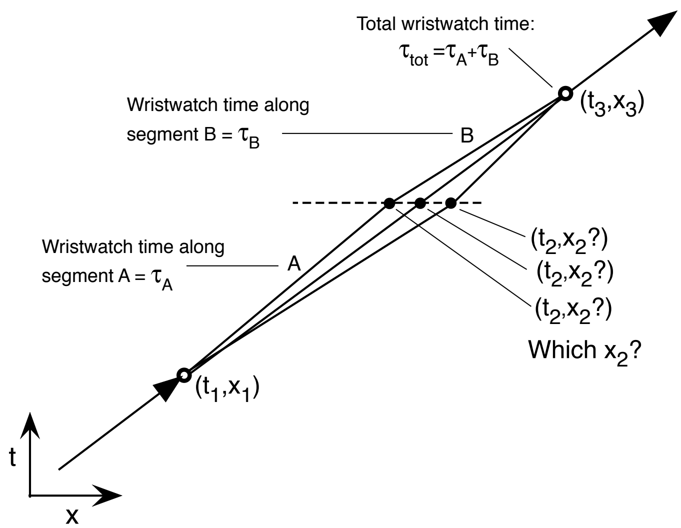


FIGURE 7 Figure for the derivation of the x-component of momentum of a stone. You will carry out this derivation in the exercises.

$p_{x,\mathrm{lab}}/m = dx_{\mathrm{lab}}/d\tau$ is a constant of motion.

are similar to (35) and (36). The result for each component of momentum reminds us that the free stone moves with constant speed in a straight line in every inertial frame.

Each component of the free stone's momentum in the laboratory frame is a *constant of motion*, like its energy E_{lab}/m in the laboratory frame, because each component of momentum does not change as the free stone moves in the laboratory frame. Momentum components of the stone in the rocket frame are also constants of motion, though equations (35) and (36) show that corresponding components in the two frames are not equal, because the stone's velocity is not the same in the two frames.

At slow speed, $v \ll 1$, we recover Newton's components of momentum in both frames. This justifies our calling components in (35) and (36) *momentum.*

Fuller Explanation: Momentum in flat spacetime: *Spacetime Physics,* Chapter 7, Momenergy.

1.9 ■ MASS IN RELATIVITY

The mass m of a stone is an invariant!

Find mass from energy and momentum.

An important relation among mass, energy, and momentum follows from the timelike interval and our relativistic expressions for energy and momentum. Suppose a moving stone emits two flashes differentially close together in distance ds_{lab} and in time dt_{lab}, with similar differentials in the rocket frame. Then (1) gives the lapse of wristwatch time $d\tau$:

$$d\tau^2 = dt_{\mathrm{lab}}^2 - ds_{\mathrm{lab}}^2 = dt_{\mathrm{rocket}}^2 - ds_{\mathrm{rocket}}^2 \qquad (37)$$

Box 1. No Mass Change with Speed!

The fact that no stone moves faster than the speed of light is sometimes "explained" by saying that "the mass of a stone increases with speed," leading to what is called "relativistic mass" whose increase prevents acceleration to a speed greater than that of light. This interpretation can be applied consistently, but what could it mean in practice? Someone riding along with the faster-moving stone detects no change in the number of atoms in the stone, nor any change whatever in the individual atoms, nor in the binding energy between atoms. Where's the "change" in what is claimed to be a "changing mass"? We observe no change in the stone that can possibly account for the varying value of its "relativistic mass."

Our viewpoint in this book is that mass is a *Lorentz invariant*, something whose value is the same for all inertial observers when they use (39) or (40) to reckon the mass. In relativity, every invariant is a diamond. Do not throw away a diamond!

To preserve the diamond of invariant mass, we will never—outside the confines of this box—use the phrase "rest mass." (Horrors!). Why not? Because "rest mass" (Ouch!) implies that there is such a thing as "non-rest mass"—mass that changes with speed. Oops, there goes your precious diamond down the drain.

In contrast, the phrase *rest energy* is fine; it *is true* that energy changes with speed; the energy of a stone *does* have different values as measured by inertial observers in uniform relative motion. In the special case of a stone at rest in any inertial frame, however, the value of its rest energy *in that frame* is equal to the value of its mass—equation (26)—provided you use the same units for mass as for energy.

"Rest mass"? NO!
Rest energy? YES!

For more on this subject see *Spacetime Physics,* **Dialog: Use and Abuse of the Concept of Mass,** pages 246–251.

Divide equation (37) through by the invariant $d\tau^2$ and multiply through by the invariant m^2 to obtain

$$m^2 = \left(m\frac{dt_{\text{lab}}}{d\tau}\right)^2 - \left(m\frac{ds_{\text{lab}}}{d\tau}\right)^2 = \left(m\frac{dt_{\text{rocket}}}{d\tau}\right)^2 - \left(m\frac{ds_{\text{rocket}}}{d\tau}\right)^2 \qquad (38)$$

Substitute expressions (23) and (35) for energy and momentum to obtain:

$$m^2 = E_{\text{lab}}^2 - p_{\text{lab}}^2 = E_{\text{rocket}}^2 - p_{\text{rocket}}^2 \qquad (39)$$

In (39) mass, energy, and momentum are all expressed in the same units, such as kilograms or electron-volts. In conventional units (subscript "conv"), the equation has a more complicated form. In either frame:

$$(m_{\text{conv}}c^2)^2 = E_{\text{conv}}^2 - p_{\text{conv}}^2 c^2 \qquad (40)$$

Stone's energy (also momentum) may be different for different observers. . .

. . . but its mass has the same (invariant!) value in all frames.

Equations (39) and (40) are central to special relativity. There is nothing like them in Newton's mechanics. The stone's energy E typically has different values when measured in different inertial frames that are in uniform relative motion. Also the stone's momentum p typically has different values when measured in different frames. However, the values of these two quantities in *any* given inertial frame can be used to determine the value of the stone's mass m, which is independent of the inertial frame. *The stone's mass m is a Lorentz invariant* (Definition 6 and Box 1).

Fuller Explanation: Mass and momentum-energy in flat spacetime: *Spacetime Physics,* Chapter 7, Momenergy.

1.10 ■ THE LORENTZ TRANSFORMATION

Relative motion; relative observations

Lorentz transform from lab to rocket

To develop special relativity, Einstein assumed that the laws of physics are the same in every inertial frame, an assertion called **The Principle of Relativity**. Let two different inertial frames, such as those of a laboratory and an unpowered rocket ship, be in uniform relative motion with respect to one another. Special relativity is valid in each of these frames. *More:* Special relativity links the coordinates of an event in one frame with the coordinates of the same event in the other frame; it also relates the energy and momentum components of a stone measured in one frame to the corresponding quantities measured in the other frame. Let an inertial (unpowered) rocket frame pass with relative velocity $v_{\rm rel}$ in the x-direction through an overlapping laboratory frame. Call the laboratory coordinate separations between two events $(\Delta t_{\rm lab}, \Delta x_{\rm lab}, \Delta y_{\rm lab}, \Delta z_{\rm lab})$ and the rocket coordinate separations between the same events $(\Delta t_{\rm rocket}, \Delta x_{\rm rocket}, \Delta y_{\rm rocket}, \Delta z_{\rm rocket})$. From now on we use the Greek letter capital delta, Δ, as a shorthand for separation, to avoid lengthy expressions, for example $\Delta t_{\rm lab} = t_{2,\rm lab} - t_{1,\rm lab}$. These separations are related by the **Lorentz transformation equations**:

$$\begin{aligned}
\Delta t_{\rm rocket} &= \gamma_{\rm rel}\left(\Delta t_{\rm lab} - v_{\rm rel}\Delta x_{\rm lab}\right) \\
\Delta x_{\rm rocket} &= \gamma_{\rm rel}\left(\Delta x_{\rm lab} - v_{\rm rel}\Delta t_{\rm lab}\right) \\
\Delta y_{\rm rocket} &= \Delta y_{\rm lab} \quad \text{and} \quad \Delta z_{\rm rocket} = \Delta z_{\rm lab}
\end{aligned} \tag{41}$$

Lorentz transform from rocket to lab

where equation (24) defines $\gamma_{\rm rel}$. We do not derive these equations here; see Fuller Explanation at the end of this section. The reverse transformation, from rocket to laboratory coordinates, follows from symmetry: replace $v_{\rm rel}$ by $-v_{\rm rel}$ and interchange rocket and lab labels in (41) to obtain

$$\begin{aligned}
\Delta t_{\rm lab} &= \gamma_{\rm rel}\left(\Delta t_{\rm rocket} + v_{\rm rel}\Delta x_{\rm rocket}\right) \\
\Delta x_{\rm lab} &= \gamma_{\rm rel}\left(\Delta x_{\rm rocket} + v_{\rm rel}\Delta t_{\rm rocket}\right) \\
\Delta y_{\rm lab} &= \Delta y_{\rm rocket} \quad \text{and} \quad \Delta z_{\rm lab} = \Delta z_{\rm rocket}
\end{aligned} \tag{42}$$

Transform energy and momentum from lab to rocket

For a pair of events infinitesimally close to one another, we can reduce differences in (42) and (41) to coordinate differentials. Further: It is also valid to divide the resulting equations through by the Lorentz invariant differential $d\tau$ and multiply through by the invariant mass m. Then substitute from equations (23) and (35). *Result:* Two sets of equations that transform the energy E and the components $(p_{\rm x}, p_{\rm y}, p_{\rm z})$ of the momentum of a stone between these two frames:

$$\begin{aligned} E_{\text{rocket}} &= \gamma_{\text{rel}}\left(E_{\text{lab}} - v_{\text{rel}} p_{\text{x,lab}}\right) \\ p_{\text{x,rocket}} &= \gamma_{\text{rel}}\left(p_{\text{x,lab}} - v_{\text{rel}} E_{\text{lab}}\right) \\ p_{\text{y,rocket}} &= p_{\text{y,lab}} \quad \text{and} \quad p_{\text{z,rocket}} = p_{\text{z,lab}} \end{aligned} \tag{43}$$

Transform energy and momentum from rocket to lab

Here $p_{\text{x,rocket}}$ is the x-component of momentum in the rocket frame, and so forth. The reverse transformation, again by symmetry:

$$\begin{aligned} E_{\text{lab}} &= \gamma_{\text{rel}}\left(E_{\text{rocket}} + v_{\text{rel}} p_{\text{x,rocket}}\right) \\ p_{\text{x,lab}} &= \gamma_{\text{rel}}\left(p_{\text{x,rocket}} + v_{\text{rel}} E_{\text{rocket}}\right) \\ p_{\text{y,lab}} &= p_{\text{y,rocket}} \quad \text{and} \quad p_{\text{z,lab}} = p_{\text{z,rocket}} \end{aligned} \tag{44}$$

We can now predict and compare measurements in inertial frames in relative motion. And remember, special relativity assumes that every inertial frame extends without limit in every direction and for all time.

Lorentz boost

Comment 6. Nomenclature: Lorentz boost

Often a Lorentz transformation is called a **Lorentz boost**. The word *boost* does not mean sudden change, but rather a change in the frame from which we make measurements and observations.

Comment 7. Constant of motion *vs.* invariant

An *invariant* is not the same as a *constant of motion.* Here is the difference:

An invariant is a quantity that has the same value *in all inertial frames.* Two sample invariants: (a) the wristwatch time between any two events, (b) the mass of a stone. The term *invariant* must always tell or imply what the change is that leads to the same result. Carefully stated, we would say: “The wristwatch time between two events and the mass of a stone are each invariant with respect to a Lorentz transformation between the laboratory and the rocket frame.”

By contrast, a *constant of motion* is a quantity that stays unchanged along the worldline of a free stone *as calculated in a given inertial frame.* Two sample constants of motion: (a) the energy and (b) the momentum of a free stone as observed or measured in, say, the laboratory frame. In other inertial frames moving relatively to the lab frame, the energy and momentum of the stone are also constants of motion; however, these quantities typically have *different values in different inertial frames.*

Conclusion: Invariants (diamonds) and constants of motion (rubies) are both truly precious.

Fuller Explanation: *Spacetime Physics,* Special Topic: Lorentz Transformation.

1.11 ■ LIMITS ON LOCAL INERTIAL FRAMES

Limits on the extent of an inertial frame in curved spacetime

Flat spacetime is the arena in which special relativity describes Nature. The power of special relativity applies strictly only in an inertial frame—or in each one of a collection of overlapping inertial frames in uniform relative motion. In every inertial frame, by definition, a free stone released from rest remains at rest and a free stone launched with a given velocity maintains the magnitude and direction of that velocity.

Limits on size of local inertial frames? We need general relativity.

If it were possible to embrace the Universe with a single inertial frame, then special relativity would describe our Universe, and we would not need general relativity. But we *do* need general relativity, precisely because typically an inertial frame is inertial in only a limited region of space and time. Near a center of attraction, every inertial frame must be **local.** An inertial frame can be set up, for example, inside a sufficiently small "container," such as (a) an unpowered rocket ship in orbit around Earth or Sun, or (b) an elevator on Earth whose cables have been cut, or (c) an unpowered rocket ship in interstellar space. In each such inertial frame, for a limited extent of space and time, we find no evidence of gravity.

Inertial frame cannot be too large, because . . .

Well, *almost* no evidence. Every inertial enclosure in which we ride near Earth cannot be too large or fall for too long a frame time without some unavoidable change in relative motion between a pair of free stones in the enclosure. Why? Because each one of a pair of widely separated stones within a large enclosed space is affected differently by the nonuniform gravitational field of Earth—as Newton would say. For example, two stones released from rest side by side are both attracted toward the center of Earth, so they move closer together as measured inside a falling long narrow horizontal railway coach (Figure 8, left panel). Their motion toward one another has nothing to do with gravitational attraction between these stones, which is entirely negligible.

As another example, think of two stones released from rest far apart vertically, one directly above the other in a long narrow vertical falling railway coach (Figure 8, right panel). For vertical separation, their gravitational accelerations toward Earth are both in the same direction. However, the stone nearer Earth is more strongly attracted to Earth, so gradually leaves the other stone behind, according to Newton's analysis. As a result, viewed from inside the coach the two stones move farther apart. *Conclusion:* The large enclosure is not an inertial frame.

. . . tidal accelerations occur in large frames.

A rider in either railway car such as those shown in Figure 8 sees the pair of horizontally-separated stones accelerate *toward* one another and a pair of vertically-separated stones accelerate *away* from one another. These relative motions earn the name **tidal accelerations,** because they arise from the same kind of nonuniform gravitational field that accounts for ocean tides on Earth—tides due to the field of the Moon, which is stronger on the side of Earth nearer the Moon.

Unavoidable tidal accelerations? Then unavoidable spacetime curvature!

As we fall toward the center of attraction, there is no way to avoid the relative—*tidal*—accelerations at different locations in the long railway car. We

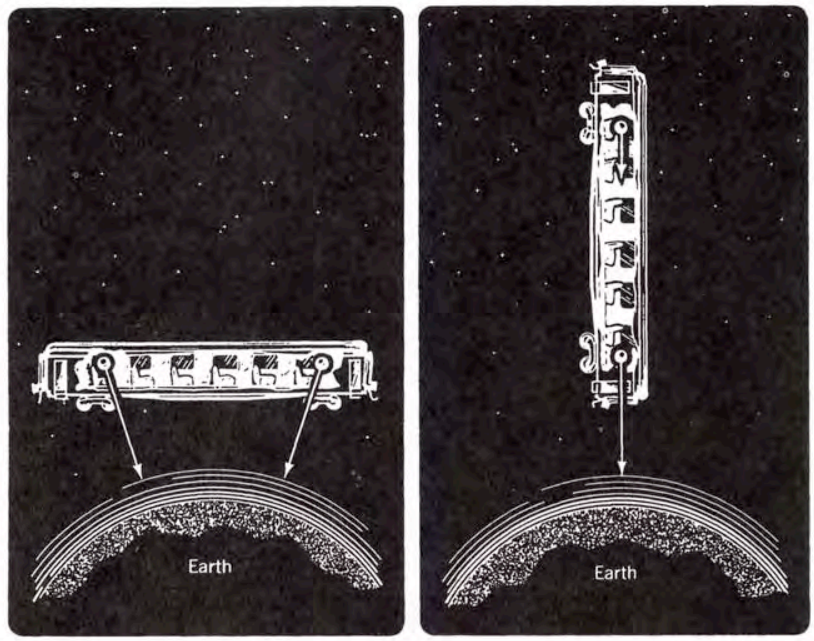


FIGURE 8 Einstein's old-fashioned railway coach in free fall, showing relative accelerations of a pair of free stones, as described by Newton (not to scale). *Left panel:* Two horizontally separated free stones are both attracted toward the center of Earth, so as viewed by someone who rides in the falling horizontal railway car, this pair of stones accelerate *toward* one another. *Right panel:* A free stone nearer Earth has a greater acceleration than that of a free stone farther from Earth. As viewed by someone who rides in the falling vertical railway car, this pair of free stones accelerate *away* from one another. We call these relative accelerations **tidal accelerations**.

can do nothing to eliminate tidal accelerations completely. These relative accelerations are central indicators of the **curvature of spacetime.**

Even though we cannot completely eliminate tidal accelerations near a center of gravitational attraction, we can often reduce them sufficiently so that they do not affect the results of a local measurement that takes place entirely in that frame.

Make every measurement in a local inertial frame.

Conclusion: Almost everywhere in the Universe we can set up a *local* inertial frame in which to carry out a measurement. Throughout this book we *choose* to make every observation and measurement and carry out every experiment in a local inertial frame. This leads to one of the key ideas in this book (see back cover):

We *choose* to report every measurement and observation using an inertial frame—a *local* inertial frame in curved spacetime.

But the local inertial frame tells only part of the story. How can we analyze a pair of events widely separated near the Earth, near the Sun, or near

General relativity: patchwork quilt of inertial frames.

a neutron star—events too far apart to be enclosed in a single inertial frame? For example, how do we describe the motion of a comet whose orbit completely encircles the Sun, with an orbital period of many years? The comet passes through a whole series of local inertial frames, but cannot be tracked using a single global inertial frame—which does not exist. Special relativity has reached its limit! To describe motion that oversteps a single local inertial frame, we must turn to a theory of curved spacetime such as Einstein's general relativity—his **Theory of Gravitation**—that we start in Chapter 3, Curving.

Comment 8. Which way does wristwatch time flow?
In your everyday life, time flows out of what you call your past, into what you call your future. We label this direction **the arrow of time**. But equation (37) contains only squared differentials, which allows wristwatch time lapse to be negative—to run backward—instead of forward along your worldline. So why does your life flow in only one direction—from past to future on your wristwatch? A subtle question! We do not answer it here. In this book we simply assume one-way flow of wristwatch time along any worldline. This assumption will lead us on an exciting journey!

Fuller Explanation: *Spacetime Physics,* Chapter 2, Falling Free, and Chapter 9, Gravity: Curved Spacetime in Action.

1.12 ■ GENERAL RELATIVITY: OUR CURRENT TOOLKIT

Ready for a theory of curved spacetime.

General relativity: amazing predictive power

The remainder of this book introduces Einstein's general theory of relativity, currently our most powerful toolkit for understanding gravitational effects. You will be astonished at the range of observations that general relativity describes and correctly predicts, among them gravitational waves, space dragging, the power of quasars, deflection and time delay of light passing a center of attraction, the tiny precession of the orbit of planet Mercury, the focusing of light by astronomical objects, and the existence of gravitational waves. It even makes some predictions about the fate of the Universe.

General relativity faces extension or revision.

In spite of its immense power, Einstein's general relativity has some inadequacies. General relativity is incompatible with quantum mechanics that describes the structure of atoms. Sooner or later a more fundamental theory is sure to replace general relativity and surmount its limits.

What makes up 96% of the Universe?

We now have strong evidence that so-called "baryonic matter"—everything we can see and touch on Earth (including ourselves) and everything we currently see in the heavens—constitutes only about four percent of the *stuff* that affects the expansion of the Universe. What makes up the remaining 96 percent? Current theories of **cosmology**—the study of the history and evolution of the Universe (Chapter 15)—examine this question using general relativity. But an alternative possibility is that general relativity itself requires modification at these huge scales of distance and time.

Theoretical research into quantum gravity is active; so are experimental tests looking for violations of general relativity, experiments whose outcomes

In the meantime, general relativity is a powerful toolkit.

might guide a new synthesis. Meanwhile, Einstein's general relativity is highly successful and increasingly important as an everyday toolkit. The conceptual issues it raises (and often satisfies) are profound and are likely to be part of any future modification. Welcome to this deep, powerful, and intellectually delicious subject!

Comment 9. Truth in labeling: "Newton" and "Einstein"
Throughout this book we talk about Newton and Einstein as if each were responsible for the current form of his ideas. This is false: Newton published nothing about kinetic energy; Einstein did not believe in the existence of black holes. Hundreds of people have contributed—and continue to contribute—to the ongoing evolution and refinement of ideas created by these giants. We do not intend to slight past or living workers in the field. Rather, we use "Newton" and "Einstein" as labels to indicate which of their worlds we are discussing at any point in the text.

?

Objection 3. *You have told me a lot of weird stuff in this chapter, but I am interested in truth and reality. Do moving clocks* ***really*** *run slow? Are clocks synchronized in one frame* ***really*** *unsynchronized in a relatively-moving frame? Give me the truth about* ***reality****!*

!

Truth and *reality* are mighty words indeed, but in both special and general relativity they are distractions; we strongly suggest that you avoid them as you study these subjects. Why? Because they direct your attention away from the key question that relativity is designed to answer: *What does this inertial observer measure and report?* Ask THAT question and you are ready for general relativity!

Fuller Explanation: *Spacetime Physics,* Chapter 9, Gravity: Curved Spacetime in Action

> *Now Besso has departed from this strange world a little ahead of me. That means nothing. We who believe in physics, know that the distinction between past, present and future is only a stubbornly persistent illusion.*
>
> —Albert Einstein, 21 March 1955, in a letter to Michele Besso's family; Einstein died 18 April 1955.

Comment 10. Chapter preview and summary
This book does not provide formal chapter previews or summaries. To preview the material, read the section titles and questions on the left hand initial page of each chapter, then skim through the marginal comments. Do the same to summarize material and to recall it at a later date.

1.13 ■ EXERCISES

1. Answer to challenge problem in Sample Problem 3:

Event B cannot cause either Event A or Event C because it occurs *after* those events in the given frame. The temporal order of events with a timelike relation will not change, no matter from what frame they are observed: See Section 2.6, entitled "The Difference between Space and Spacetime."

2. Spatial Separation I

Two firecrackers explode at the same place in the laboratory and are separated by a time of 3 seconds as measured on a laboratory clock.

A. What is the spatial distance between these two events in a rocket in which the events are separated in time by 5 seconds as measured on rocket clocks?

B. What is the relative speed v_{rel} between rocket and laboratory frames?

3. Spatial Separation II

Two firecrackers explode in a laboratory with a time difference of 4 seconds and a space separation of 5 light-seconds, both space and time measured with equipment at rest in the laboratory. What is the distance between these two events in a rocket in which they occur at the same time?

4. Maximum wristwatch time

Show that equation (18) corresponds to a maximum, not a minimum, of total wristwatch time of the stone, equation (17), as it travels across two adjacent segments of its worldline.

5. Space Travel

An astronaut wants to travel to a star 33 light-years away. He wants the trip to last 33 years. (He wants to *age* 33 years during the trip.) How fast should he travel? (The answer is NOT $v = 1$.)

6. Traveling Clock Loses Synchronization

An airplane flies from Budapest to Boston, about 6700 kilometers, at a speed of 350 meters/second. It carries a clock that was initially synchronized with a clock in Budapest and another one in Boston. When the clock arrives in Boston, will the clock aboard the plane be fast or slow compared to the one in Boston, and by how much? Neglect the curvature and rotation of the Earth, as well as the short phases of acceleration and deceleration of the plane at takeoff and landing.

7. Successive Lorentz Boosts

Consider two successive Lorentz transformations: the first transformation from lab frame L to runner frame R, and a second transformation from runner frame R to super-runner frame S. The runner frame moves with speed v_1 relative to the lab frame. And the super-runner frame moves with speed v_2 relative to the runner frame; this, along the same line of motion that R moves relative to L.

Write the two transformations, from L to R, and from R to S, and combine them to obtain events coordinates in the S frame in terms of the events coordinates in the L frame. Show that the result is equivalent to a single Lorentz transformation from L to S, with speed $v_{\rm rel}$ given by:

$$v_{\rm rel} = \frac{v_1 + v_2}{1 + v_1 v_2} \tag{45}$$

Use equation (45) to verify the slogan, *For light, one plus one equals one.*

8. Tilted Meter Stick

A spaceship moves directly toward Earth, say along the x-axis at constant speed $v_{\rm rel}$ with respect to Earth. A meter stick is stationary in the spaceship but oriented at an angle $\alpha_{\rm S}$ with respect to the forward line of relative motion. As they pass one another: (a) What angle does the Earth observer measure the meter stick to make with his x-axis? (b) What is the length of the stick measured by the earth observer? (c) Answer parts (a) and (b) for the cases $\alpha_{\rm S} = 90^{\rm o}$ and $\alpha_{\rm S} = 0^{\rm o}$. (d) For the case $v_{\rm rel} = 0.75$ and $\alpha_{\rm S} = 60^{\rm o}$, what are the numerical results of parts (a) and (b)?

9. Super Cosmic Rays

The Pierre Auger Observatory is an array of cosmic ray detectors lying on the vast plain *Pampa Amarilla* (yellow prairie) in western Argentina, just east of the Andes Mountains. The purpose of the observatory is to study cosmic rays of the highest energies. The highest energy cosmic ray detected had an energy of 3×10^{20} electron-volts.

A. A regulation tennis ball has a mass of 57 grams. If this tennis ball is given a kinetic energy of 3×10^{20} electron volts, how fast will it move, in meters per second? (*Hint:* Try Newton's mechanics.)

B. Suppose a proton has the energy 3×10^{20} electron-volts. How long would it take this proton to cross our galaxy (take the galaxy diameter to be 10^5 light-years) as measured on the proton's wristwatch? Give your answer in seconds.

C. What is the diameter of the galaxy measured in the rest frame of the proton?

10. Mass-Energy Conversion

A. How much mass does a 100-watt bulb dissipate (in heat and light) in one year?

B. Pedaling a bicycle at full throttle, you generate approximately one-half horsepower of *useful* power. (1 horsepower = 746 watts). The human body is about 25 percent efficient; that is, 25 percent of the food burned can be converted to useful work. How long a time will you have to ride your bicycle in order to lose 1 kilogram by direct conversion of mass to energy? Express your answer in years. (One year = 3.16×10^7 seconds.) How can weight-reducing gymnasiums stay in business? What is misleading about the way this exercise is phrased?

C. One kilogram of hydrogen combines chemically with 8 kilograms of oxygen to form water; about 10^8 joules of energy is released. A very good chemical balance is able to detect a fractional change in mass of 1 part in 10^8. By what factor is this sensitivity more than enough—or insufficient—to detect the fractional change of mass in this reaction?

11. Departure from Newton

Use equations (33) and (34) to check the Newtonian limit of the expression for kinetic energy:

A. An asteroid that falls from rest at a great distance reaches Earth's surface with a speed of 10 kilometers/second (if we neglect atmospheric resistance). By what percent is Newton's prediction for kinetic energy in error for this asteroid?

B. At what speed does the all-speed expression for kinetic energy (34) yield a kinetic energy that differs from Newton's prediction—embodied in equation (33)—by one percent? ten percent? fifty percent? seventy-five percent? one hundred percent? Use the percentage expression $100 \times [KE - (KE)_{\text{Newton}}]/KE$, where KE is the relativistic expression for kinetic energy.

12. Units and Conversions

A. Show that the speed of a stone in an inertial frame (as a fraction of the speed of light) is given by the expression

$$v_{\text{inertial}} = \left(\frac{ds}{dt}\right)_{\text{inertial}} = \left(\frac{p}{E}\right)_{\text{inertial}} \tag{46}$$

B. What speed v does (46) predict when the mass of the particle is zero, as is the case for a flash of light? Is this result the one you expect?

C. The mass and energy of particles in beams from accelerators is often expressed in GeV, that is billions of electron-volts. Journal articles describing these measurements refer to particle momentum in units of GeV/c. Explain.

13. The Pressure of Light

A flash of light has zero mass. Use equation (40), in conventional units, to answer the following questions.

A. You can feel on your hand an object with the weight of 1 gram mass. Shine a laser beam downward on a black block of wood that you hold in your hand. You detect an increased force as if the block of wood had increased its mass by one gram. What power does the laser beam deliver, in watts?

B. The block of wood described in part A absorbs the energy of the laser beam. Will the block burst into flame?

14. Derivation of the Expression for Momentum

A. Carry out the derivation of the relativistic expression for momentum described in Section 1.8. Lay out this derivation in a series of numbered steps that parallel those for the derivation of the energy in Section 1.7.

B. Write an expression for p in conventional units.

15. Verifying energy-momentum transformation equations

Derive transformation equations (43) and (44) using the procedure outlined just before these equations.

16. Newtonian transformation

Show that for Newton, where all velocities are small compared to the speed of light, the Lorentz transformation equations (41) reduce to the familiar Galilean transformation equations and lead to the universality of time.

17. The Photon

NOTE: Exercises 13 through 18 are related to one another.

A. A photon is a quantum of light, a particle with zero mass. Apply equation (39) for a photon moving only in the $\pm x$-direction. Show that in this conversion to light, $p_{\text{x}} \rightarrow \pm E$.

B. Write down the Lorentz transformation equations (43) and (44) for a photon moving in the positive x-direction.

C. Write down the Lorentz transformation equations (43) and (44) for a photon moving in the negative x-direction.

D. Show that *it does not matter* what units you use for E in your photon Lorentz transformation equations, as long as the units for each occurrence of E are the same.

18. One-Dimensional Doppler Equations

A mongrel equation (neither classical nor quantum-mechanical) connects the quantum energy E of a single photon with the frequency f of a classical electromagnetic wave. In conventional units, this equation is:

$$E_{\mathrm{conv}} = hf_{\mathrm{conv}} \qquad \text{(photon, conventional units)} \tag{47}$$

where f_{conv} is the frequency in oscillations per second and h is **Planck's constant**. In SI units, E_{conv} has the unit joules, and h has the value $h = 6.63 \times 10^{-34}$ joule-second.

A. Substitute (47) into your transformation equations for the photon, and replace γ_{rel} in those equations with its definition $(1 - v_{\mathrm{rel}}^2)^{-1/2}$. Planck's constant disappears from the resulting equations between frequency f_{lab} in the laboratory frame and frequency f_{rocket} in the rocket frame:

$$f_{\mathrm{lab}} = \left[\frac{1 \pm v_{\mathrm{rel}}}{1 \mp v_{\mathrm{rel}}}\right]^{1/2} f_{\mathrm{rocket}} \qquad (\pm x,\ \text{light}) \tag{48}$$

$$f_{\mathrm{rocket}} = \left[\frac{1 \mp v_{\mathrm{rel}}}{1 \pm v_{\mathrm{rel}}}\right]^{1/2} f_{\mathrm{lab}} \qquad (\pm x,\ \text{light}) \tag{49}$$

These are the **one-dimensional Doppler equations** for light moving in either direction along the x-axis.

B. The relation between frequency f_{conv} and wavelength λ_{conv} for a classical plane wave in an inertial frame, in conventional units

$$f_{\mathrm{conv}}\lambda_{\mathrm{conv}} = c \qquad \text{(classical plane wave)} \tag{50}$$

Rewrite equations (48) and (49) for the relation between laboratory wavelength λ_{lab} and rocket wavelength $\lambda_{\mathrm{rocket}}$.

19. Speed-Control Beacon

An advanced civilization sets up a beacon on a planet near the crowded center of our galaxy and asks travelers approaching directly or receding directly from the beacon to use the Doppler shift to measure their speed relative to the beacon, with a speed limit at $v = 0.2$ relative to that beacon. The beacon emits light of a single *proper* wavelength λ_0, that is, the wavelength measured in the rest frame of the beacon. Four index colors are:

$$\begin{aligned}\lambda_{\rm red} &= 680 \times 10^{-9}\text{meter} = 680 \text{ nanometers} \\ \lambda_{\rm yellow} &= 580 \times 10^{-9}\text{meter} = 580 \text{ nanometers} \\ \lambda_{\rm green} &= 525 \times 10^{-9}\text{meter} = 525 \text{ nanometers} \\ \lambda_{\rm blue} &= 475 \times 10^{-9}\text{meter} = 475 \text{ nanometers}\end{aligned} \tag{51}$$

A. Choose the beacon proper wavelength λ_0 so that a ship approaching at half the speed limit, $v = 0.1$, sees green light. What is the proper wavelength λ_0 of the beacon beam? What color do you see when you stand next to the beacon?

B. As your spaceship moves directly toward the beacon described in Part A, you see the beacon light to be blue. What is your speed relative to the beacon? Is this below the speed limit?

C. In which direction, toward or away from the beacon, are you traveling when you see the beacon to be red? What is your speed relative to the beacon? Is this below the speed limit?

20. Radar

An advanced civilization uses radar to help enforce the speed limit in the crowded center of our galaxy. Radar relies on the fact that with respect to its rest frame a spaceship reflects a signal back with a frequency equal to the incoming frequency measured in its frame.

A. Show that a radar signal of frequency f_0 at the source is received back from a directly approaching ship with the reflected frequency $f_{\rm reflect}$ given by the expression:

$$f_{\rm reflect} = \frac{1+v}{1-v} f_0 \qquad \text{(radar)} \tag{52}$$

where v is the speed of the spaceship with respect to the signal source.

B. What is the wavelength $\lambda_{\rm reflect}$ of the signal reflected back from a spaceship approaching at the speed limit of $v = 0.2$?

C. The highway speed of a car is very much less than the speed of light. Use the approximation formula inside the front cover to find the following approximate expression for $f_{\rm reflect} - f_0$:

$$f_{\rm reflect} - f_0 \approx 2 v f_0 \qquad \text{(highway radar)} \tag{53}$$

The Massachusetts State Highway Patrol uses radar with microwave frequency $f_0 = 10.525 \times 10^9$ cycles/second. By how many cycles/second is the reflected beam shifted in frequency when reflected from a car approaching at 100 kilometers/hour (or 27.8 meters/second)?

21. Two-dimensional Velocity Transformations

An electron moves in the laboratory frame with components of velocity $(v_{\mathrm{x,lab}}, v_{\mathrm{y,lab}})$ and in the rocket frame with components of velocity $(v_{\mathrm{x,rocket}}, v_{\mathrm{y,rocket}})$.

A. Use the differential form of the Lorentz transformation equations (42) to relate the velocity components of the electron in laboratory and rocket frames:

$$v_{\mathrm{x,lab}} = \frac{v_{\mathrm{x,rocket}} + v_{\mathrm{rel}}}{1 + v_{\mathrm{rel}} v_{\mathrm{x,rocket}}} \qquad v_{\mathrm{y,lab}} = \frac{v_{\mathrm{y,rocket}}}{\gamma_{\mathrm{rel}} \left(1 + v_{\mathrm{rel}} v_{\mathrm{x,rocket}}\right)} \tag{54}$$

This is called the **Law of Transformation of Velocities**.

B. With a glance at the Lorentz transformation (42) and its inverse (41), make an argument that to derive the inverse of (54), one simply replaces v_{rel} with $-v_{\mathrm{rel}}$ and interchanges lab and rocket labels, leading to:

$$v_{\mathrm{x,rocket}} = \frac{v_{\mathrm{x,lab}} - v_{\mathrm{rel}}}{1 - v_{\mathrm{rel}} v_{\mathrm{x,lab}}} \qquad v_{\mathrm{y,rocket}} = \frac{v_{\mathrm{y,lab}}}{\gamma_{\mathrm{rel}} \left(1 - v_{\mathrm{rel}} v_{\mathrm{x,lab}}\right)} \tag{55}$$

C. Does the law of transformation of velocities allow the electron to move faster than light when observed in the laboratory frame? For example, suppose that in the rocket frame the electron moves in the positive x_{rocket}-direction with velocity $v_{\mathrm{x,rocket}} = 0.75$ and the rocket frame also moves in the same direction with the same relative speed $v_{\mathrm{rel}} = 0.75$. What is the value of the velocity $v_{\mathrm{x,lab}}$ of the electron in the laboratory frame?

D. Suppose two light flashes move with opposite velocities $v_{\mathrm{x,rocket}} = \pm 1$ in the rocket frame. What are the corresponding velocities $v_{\mathrm{x,lab}}$ of the two light flashes in the laboratory frame?

E. Light moves with velocity components $(v_{\mathrm{x,rocket}}, v_{\mathrm{y,rocket}}, v_{\mathrm{z,rocket}}) = (0, -1, 0)$ in the rocket frame. *Predict* the magnitude $|v_{\mathrm{lab}}|$ of its velocity measured in the laboratory frame. Does a calculation verify your prediction?

22. Aberration of light

Light that travels in one direction in the laboratory travels in another direction in the rocket frame unless the light moves along the line of relative motion of the two frames. This difference in light travel direction is called **aberration**.

A. Transform the angle of light propagation in two spatial dimensions. Recall that laboratory and rocket x-coordinates lie along the same line, and in each frame measure the angle ψ of light motion with respect to this common forward x-direction. Make the following argument: Light travels with the speed one, which is the hypotenuse of the velocity component triangle. Therefore for light $v_{\text{x,inertial}} \equiv v_{\text{x,inertial}}/1 = \cos\psi$. Show that this argument converts the first of equations (54) to:

$$\cos\psi_{\text{lab}} = \frac{\cos\psi_{\text{rocket}} + v_{\text{rel}}}{1 + v_{\text{rel}}\cos\psi_{\text{rocket}}} \qquad \text{(light)} \qquad (56)$$

B. From equation (39) show that for light tracked in any inertial frame $|p_{\text{inertial}}| = E_{\text{inertial}}$. Hence $p_{\text{x,inertial}}/E_{\text{inertial}} = \cos\psi$ and the first of equations (44) becomes, for light

$$E_{\text{lab}} = E_{\text{rocket}}\gamma_{\text{rel}}\left(1 + v_{\text{rel}}\cos\psi_{\text{rocket}}\right) \qquad \text{(light)} \qquad (57)$$

C. Make an argument that to derive the inverses of (56) and (57), you simply replace v_{rel} with $-v_{\text{rel}}$ and interchange laboratory and rocket labels, to obtain the aberration equations:

$$\cos\psi_{\text{rocket}} = \frac{\cos\psi_{\text{lab}} - v_{\text{rel}}}{1 - v_{\text{rel}}\cos\psi_{\text{lab}}} \qquad \text{(light)} \qquad (58)$$

$$E_{\text{rocket}} = E_{\text{lab}}\gamma_{\text{rel}}\left(1 - v_{\text{rel}}\cos\psi_{\text{lab}}\right) \qquad \text{(light)} \qquad (59)$$

D. A source at rest in the rocket frame emits light uniformly in all directions in that frame. Consider the 50 percent of this light that goes into the forward hemisphere in the rocket frame. Show that in the laboratory frame this light is concentrated in a narrow forward cone of half-angle $\psi_{\text{headlight,lab}}$ given by the following equation:

$$\cos\psi_{\text{headlight,lab}} = v_{\text{rel}} \qquad \text{(headlight effect)} \qquad (60)$$

The transformation that leads to concentration of light in the forward direction is called the **headlight effect**.

23. Cherenkov Radiation

Can an electron move faster than light? No and yes. No, an electron cannot move faster than light *in a vacuum*; yes, it can move faster than light in a medium in which light moves more slowly than its standard speed in a vacuum. P. A. Cherenkov shared the 1958 Nobel Prize for this discovery that an electron emits coherent radiation when it moves faster than light moves in any medium.

What is the minimum kinetic energy that an electron must have to emit Cherenkov radiation while traveling through water, where the speed of light is $v_{\text{light}} \approx 0.75$? Express this kinetic energy as both the fraction (kinetic energy)$/m$ of its mass m and in electron-volts (eV). Type “Cherenkov radiation" into a computer search engine to see images of the blue light due to Cherenkov radiation emitted by a radioactive source in water.

24. Live Forever?

Luc Longtin shouts, "I can live forever! Here is a variation of equation (1): $\Delta\tau^2 = \Delta t^2_{\text{Earth}} - \Delta s^2_{\text{Earth}}$. Relativity allows the possibility that $\Delta\tau \ll \Delta t_{\text{Earth}}$. In the limit, $\Delta\tau \to 0$, so the hour hand on my wristwatch does not move. Eternal life!

"I have decided to ride a 100 kilometer/hour train back and forth my whole life. THEN I will age much more slowly." Comment on Luc's ecstatic claim without criticizing him.

A. When he carries out his travel program, how much younger will 100-year-old Luc be than his stay-at-home twin brother Guy?

B. Suppose Luc rides a spacecraft in orbit around Earth (speed given in Figure 2). In this case, how much younger will 100-year-old Luc be than brother Guy?

C. Suppose Luc manages to extend his life measured in Earth-time by riding on a fast cosmic ray (speed given in Figure 2). When Luc returns to Earth in his old age, it is clear that his brother Guy will no longer be among the living. However, would Luc *experience* his life as much longer than he would have experienced it if he remained on Earth? That is, would he "enjoy a longer life" in some significant sense, for example counting many times the total number of heartbeats experienced by Guy?

1.14 ■ REFERENCES

Initial quote: *The Ultimate Quotable Einstein*, Collected and Edited by Alice Calaprice, 2011, Princeton University Press, page 354.

Final quote: same reference, page 113.

Spacetime Physics, Introduction to Special Relativity, Second Edition, Edwin F. Taylor and John Archibald Wheeler, W. H. Freeman and Co., New York, 1992, ISBN 0-7167-2327-1. This is our own special relativity text, to which we make reference at the end of several sections in Chapter 1 and elsewhere in the present book.

Quote at the end of Section 1.5: Robert W. Fuller and John A. Wheeler, "Causality and Multiply Connected Space-Time," Physical Review, Volume 128, Number 2, October 15, 1962, pages 919-929.

Newton's first law of motion (Section 1.7) quoted from *The Principia*, A new translation by I. Bernard Cohen and Anne Whitman, University of California Press, 1999, page 416.

Comment 8: Kip S. Thorne's *Black Holes and Time Warps: Einstein's Outrageous Legacy*, W. W. Norton Co., New York, 1994. An introduction that includes the many people who forged the physics of the black hole.

Download file name: Ch01Speeding170508v1.pdf

Chapter 2. The Bridge: Special Relativity to General Relativity

- *How can I get rid of gravity? (Do not try this at home!)*
- *Kansas is on a curved Earth; why can we use a flat road map of Kansas?*
- *How can I find the shortest path between two points on a curved surface?*
- *How does a stone move in curved spacetime?*
- *What is the fundamental difference between space and spacetime?*

CHAPTER 2 The Bridge: Special Relativity to General Relativity

Edmund Bertschinger & Edwin F. Taylor *

Law 1. Every body perseveres in its state of being at rest or of moving uniformly straight forward except insofar as it is compelled to change its state by forces impressed.

—Isaac Newton

At that moment there came to me the happiest thought of my life . . . for an observer falling freely from the roof of a house no gravitational field exists during his fall—at least not in his immediate vicinity. That is, if the observer releases any objects, they remain in a state of rest or uniform motion relative to him, respectively, independent of their unique chemical and physical nature. Therefore the observer is entitled to interpret his state as that of "rest."

—Albert Einstein

2.1 ■ LOCAL INERTIAL FRAME

We can always and (almost!) anywhere "let go" and drop into a local inertial frame.

No force of gravity in inertial frame

Law 1 above, Newton's First Law of Motion, is the same as our definition of an inertial frame (Definition 1, Section 1.1). For Newton, gravity is just one of many forces that can be "impressed" on a body. Einstein, in what he called the happiest thought of his life, realized that on Earth, indeed as far as we know *anywhere in the Universe*—except on the singularity inside the black hole—we can find a local "free-fall" frame in which an observer does not feel gravity. We understand instinctively that *always* and *anywhere* we can remove

Local inertial frame available anywhere

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

FIGURE 1 Vito Ciaravino, a University of Michigan student, experiences weightlessness as he rides the Vomit Comet. NASA photo.

the floor or cut the cable that holds us up and immediately drop into a **local inertial frame**. *There is no force of gravity* in Einstein's inertial frame—"at least not in his immediate vicinity."

In curved spacetime inertial frame is local.

Einstein's phrase "in his [the observer's] immediate vicinity" brings a warning: Generally, an inertial frame is *local*. Section 1.11 showed that tidal effects can limit the extent of distances and times measured in a frame in which special relativity is valid and correctly describes motions and other observations.

Inertial frame ≡ free-fall frame

We call a local inertial frame a *free-fall frame*, even though from some viewpoints the frame may not be falling. A rising rocket immediately after burnout above Earth's atmosphere provides a free-fall frame, even while it continues temporarily to climb away from the surface. So does an unpowered spaceship in interstellar space, which is not "falling" toward anything.

Laws of physics identical in every inertial frame.

Vito Ciaravino (Figure 1) floats freely inside the Vomit Comet, a NASA model C9 cargo plane guided to follow, for 25 to 30 seconds, the same trajectory above Earth's surface that a free projectile would follow in the absence of air resistance (Figure 2). As Vito looks around inside the cabin, he cannot tell whether his local container is seen by people outside to be rising or falling—or tracing out some other free-fall orbit. Indeed, he might forgetfully think for a moment that his capsule is floating freely in interstellar space. The Principle of Relativity tells us that the laws of physics are the same in *every* free-fall frame.

Newton claims that tidal accelerations are merely the result of the variation in gravity's force from place to place. But Einstein asserts: *There is no such thing as the force of gravity.* Rather, gravitational effects (including tides) are evidence of spacetime curvature. In Chapter 3 we find that tides are

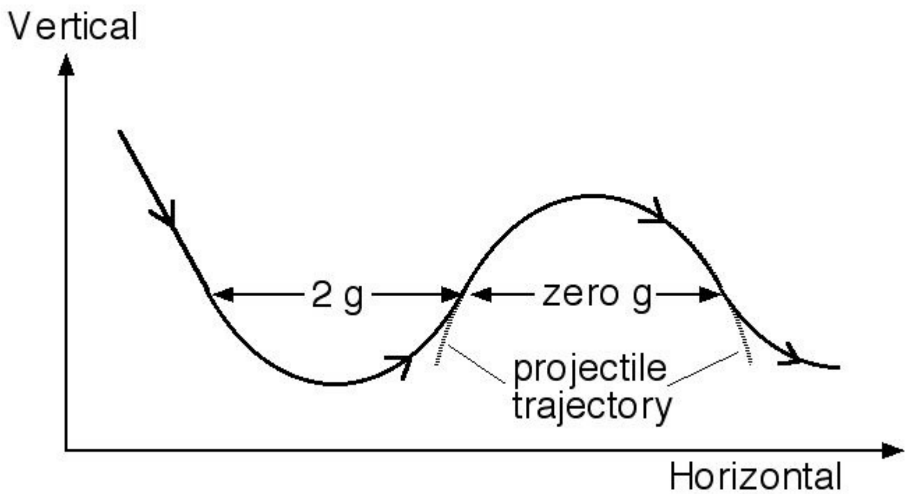


FIGURE 2 Trajectory followed by the Vomit Comet airplane above Earth's surface. Portions of the trajectory marked "2 g" and "zero g" are parabolas. During the zero-g segment, which lasts up to 30 seconds, the plane is guided to follow the trajectory of a free projectile in the absence of air resistance. By guiding the plane through different parabolic trajectories, the pilot can (temporarily!) duplicate the gravity on Mars (one-third of g on Earth) or the Moon (one-sixth of g on Earth).

Spacetime curvature has many effects.

but one consequence of spacetime curvature. Many effects of curvature cannot be explained or even described using Newton's single universal frame in which gravity is a force like any other. General relativity is not just an alternative to Newton's laws; it bursts the bonds of Newton's vision and moves far beyond it.

Curved surface compared to curved spacetime

Flat and curved surfaces in *space* can illuminate, by analogy, features of flat and curved *spacetime.* In the present chapter we use this analogy between a flat or curved surface, on the one hand, and flat or curved spacetime, on the other hand, to bridge the transition between special relativity (SR) and general relativity (GR).

2.2 ■ FLAT MAPS: LOCAL PATCHES ON CURVED SURFACES

Planning short and long trips on Earth's spherical surface

General relativity sews together local inertial frames.

Spacetime curvature makes it impossible to use a single inertial frame to relate events that are widely separated in spacetime. General relativity makes the connection by allowing us to choose a *global coordinate system* that effectively sews together local inertial frames. General relativity's task is similar to yours when you lay out a series of adjacent small flat maps to represent a long path between two widely separated points on Earth. We now examine this analogy in detail.

Flat Kansas map "good enough" for local traveler.

Figure 3 is a flat road map of the state of Kansas, USA. Someone who plans a trip within Kansas can use the **map scale** at the bottom of this map to convert centimeters of length on the map between two cities to kilometers

FIGURE 3 Road map of the state of Kansas, USA. Kansas is small enough, relative to the entire surface of Earth, so that projecting Earth's features onto this flat map does not significantly distort separations or relative directions. (Copyright geology.com)

that he drives between these cities. The map reader has confidence that using the same map scale at different locations in Kansas will not lead to significant errors in predicting separations between cities—because "flat Kansas" conforms pretty well to the curved surface of Earth. Figure 4 shows a flat patch bigger than Kansas on which map distortions will still be negligible for most everyday purposes. In contrast, at the edge of Earth's profile in Figure 4 is an edge-on view of a much larger flat surface. A projection from the rounded Earth surface onto this larger flat surface inevitably leads to some small distortions of separations compared to those actually measured along the curved surface of Earth. We define a **space patch** as a flat surface on which a projected map is sufficiently distortion-free for whatever purpose we are using the map.

Definition: **space patch**

DEFINITION 1. Space patch

A **space patch** is a flat surface purposely limited in size so that a map projected onto it from a curved surface does not result in significant distortions of separations between locations for the purpose of a given measurement or journey.

Single flat map not accurate for a long trip.

Let's plan an overland trip along a path that we choose between the city of Amsterdam in the Netherlands and the city of Vladivostok in Siberia. We recognize that on a single flat map the path of our long trip will be distorted. How then do we reckon the trip length from Amsterdam to Vladivostok? This total length for a long trip across much of the globe can be estimated using a series of local flat maps on slightly overlapping space patches (Figure 5). We sum the short separations across these small flat maps to reckon the total length of the long, winding path from Amsterdam to Vladivostok.

On each local flat map we are free to fix positions using a square array of perpendicular coordinates ("Cartesian coordinates") in north-south

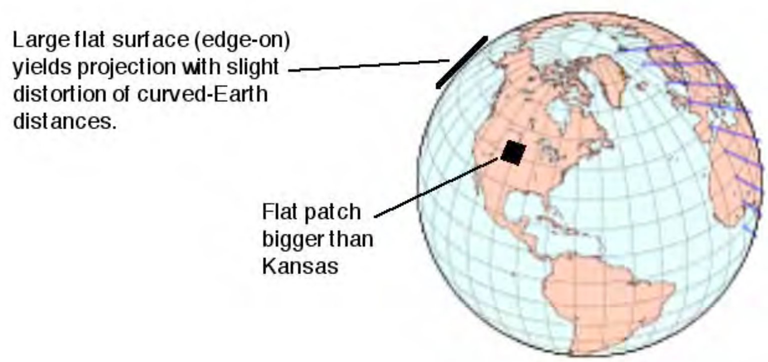


FIGURE 4 Small space patch and large flat plane tangent to Earth's surface. Projecting Earth's features onto the large flat plane can lead to distortion of those features on the resulting flat map. For precise mapmaking, the larger surface does not satisfy the requirements of a *space patch*.

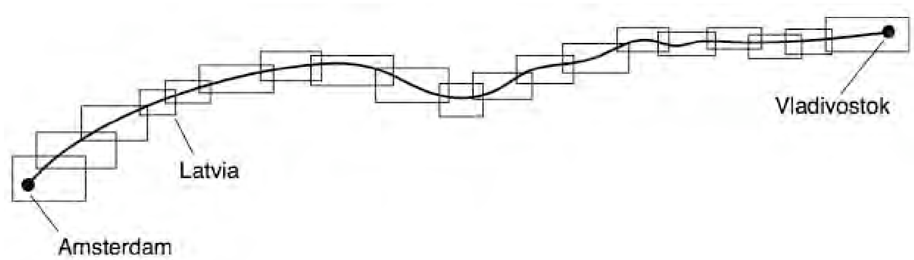


FIGURE 5 To reckon the total length of the path between Amsterdam and Vladivostok, sum the short separations across a series of small, overlapping, flat maps lined up along our chosen path. One of these small, flat maps covers all of Latvia. The smaller each map is—and the greater the total number of flat maps along the path—the more accurately will the sum of measured distances across the series of local maps represent the actually-measured total length of the entire path between the two cities.

On each small flat map, use the Pythagorean Theorem.

(y-coordinate) and east-west (x-coordinate) directions applied to that particular patch, for example on our regional map of Latvia. The distance or space separation between two points, $\Delta s_{\mathrm{Latvia}}$, that we calculate using the Pythagorean Theorem applied to the flat Latvian map is *almost equal* to the separation that we would measure using a tape measure that conforms to Earth's curved surface. Use the name **local space metric** to label the local, approximate Pythagorean theorem:

$$\Delta s^2_{\mathrm{Latvia}} \approx \Delta x^2_{\mathrm{Latvia}} + \Delta y^2_{\mathrm{Latvia}} \quad \text{(local space metric on Latvian patch)} \quad (1)$$

Δ means **increment,** a finite but small separation.

Comment 1. Notation for Approximate Metrics

Equation (1) displays the notation that we use throughout this book for an approximate metric on a flat patch. First, the symbol capital delta, Δ, stands for **increment,** a measurable but still small separation that gives us "elbow room" to make measurements. This replaces the unmeasurably small quantity indicated by the zero-limit calculus differential d. Second, the approximately equal sign, $\approx$, acknowledges that, even though our flat surface is small, projection onto it from the curved surface inevitably leads to some small distortion. Finally, the subscript label, such as "Latvia," on each incremental variable names the local patch.

We order flat maps from each nation through which we travel from Amsterdam to Vladivostok and measure little separations on each map (Figure 5). In equation (1), from our choice of axes, Δy_{Latvia} aligns itself with a great circle that passes through the north geographic pole, while Δx_{Latvia} lies in the perpendicular east-west direction.

Geographic north and magnetic north yield same Δs.

On a more ancient local flat map, the coordinate separation $\Delta y_{\text{Latvia,rot}}$ may lie in the direction of magnetic north, a direction directly determined with a compass. Choose $\Delta x_{\text{Latvia,rot}}$ to be perpendicular to $\Delta y_{\text{Latvia,rot}}$. Then in rotated coordinates using magnetic north the same incremental separation between points along our path is given by the alternative local space metric

$$\Delta s^2_{\text{Latvia}} \approx \Delta x^2_{\text{Latvia,rot}} + \Delta y^2_{\text{Latvia,rot}} = \Delta x^2_{\text{Latvia}} + \Delta y^2_{\text{Latvia}} \qquad (2)$$

Pythagorean Theorem valid on rotated flat maps.

These two local maps are rotated relative to one another. But the value of the left side is the same. Why? First, because the value of the left side is *measured directly*; it does not depend on any coordinate system. Second, the values of the two right-hand expressions in (2) are equal because the Pythagorean theorem applies to all flat maps. *Conclusion:* Relative rotation does not change the predicted value of the incremental separation Δs_{Latvia} between nearby points along our path. So when we sum individual separations to find the total length of the trip, we make no error when we use a variety of maps if their only difference is relative orientation toward north.

2.3 ■ GLOBAL COORDINATE SYSTEM ON EARTH

Global space metric using latitude and longitude

Use latitude and longitude.

A professional mapmaker (cartographer) gently laughs at us for laying side by side all those tiny flat maps obtained from different and possibly undependable sources. She urges us instead to use the standard global coordinate system of latitude and longitude on Earth's surface (Figure 6). She points out that a hand-held Global Positioning System (GPS) receiver (Chapter 4) verifies to high accuracy our latitude and longitude at any location along our path. Combine these readings with a global map—perhaps already installed in the GPS receiver—to make easy the calculation of differential displacements ds on each local map, which we then sum (integrate) to predict the total length of our path.

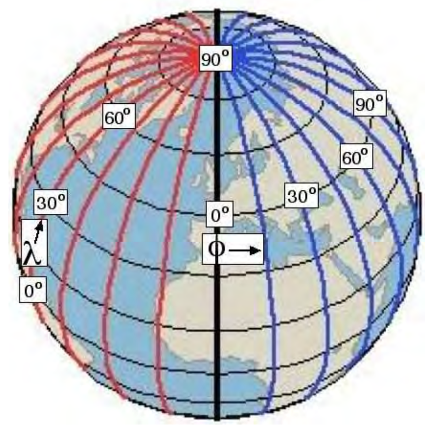


FIGURE 6 Conventional global coordinate system for Earth using angles of latitude λ and longitude ϕ.

Space metric in global coordinates

What price do we pay for the simplicity and accuracy of latitude and longitude coordinates? Merely our time spent receiving a short tutorial on the surface geometry of a sphere. Our cartographer lays out Figure 6 that shows angles of latitude λ and longitude ϕ, then gives us a third version of the space metric—call it a **global space metric**—that uses global coordinates to provide the same incremental separation ds between nearby locations as does a local flat map:

$$ds^2 = R^2 \cos^2 \lambda \, d\phi^2 + R^2 d\lambda^2 \qquad (0 \leq \phi < 2\pi \text{ and } -\pi/2 \leq \lambda \leq +\pi/2) \quad (3)$$

Here R is the radius of Earth. For a quick derivation of (3), see Figure 7.

Global space metric contains coordinates as well as differentials.

Why does the function $\cos \lambda$ appear in (3) in the term with coordinate differential $d\phi$? Because north and south of the equator, curves of longitude converge toward one another, meeting at the north and south poles. When we move 15° of longitude near the equator we travel a much longer east-west path than when we move 15° of longitude near the north pole or south pole. Indeed, very close to either pole the traveler covers 15° of longitude when he strolls along a very short east-west path.

RIDDLE: A bear walks one kilometer south, then one kilometer east, then one kilometer north and arrives back at the same point from which she started. Three questions:

1. What color is the bear?
2. Through how many degrees of longitude does the bear walk eastward?
3. How many kilometers must the bear travel to cover the same number of degrees of longitude when she walks eastward on Earth's equator?

The global space metric (3) is powerful because it describes the differential separation ds between adjacent locations *anywhere* on Earth's surface.

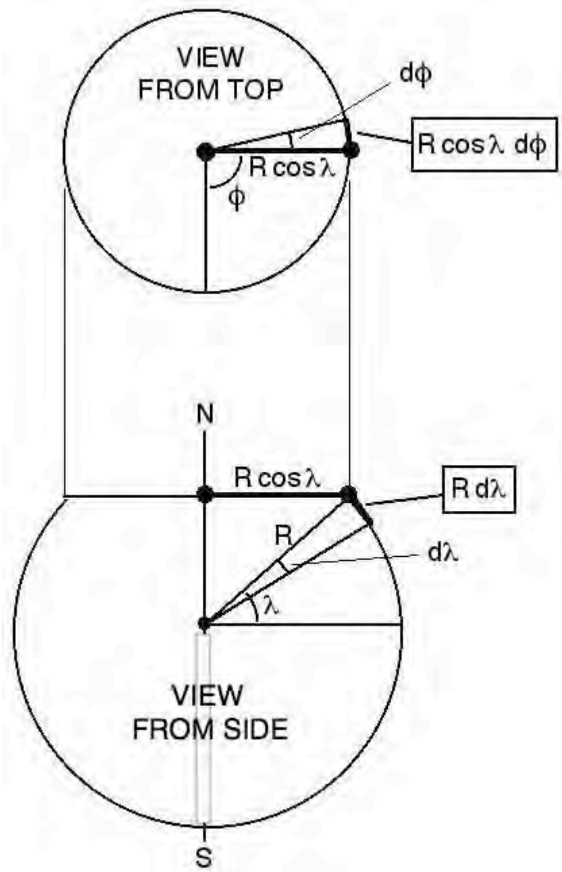


FIGURE 7 Derive the global space metric (3), as the sum of the squares of the north-south and east-west sides of a little box on Earth's surface. The north-south side of the little box is $Rd\lambda$, where R is the radius of Earth and $d\lambda$ is the differential change in latitude. The east-west side is $R\cos\lambda\, d\phi$. The global space metric (3) adds the squares of these sides (Pythagorean Theorem!) to find the square of the differential separation ds^2 across the diagonal of the little box.

Adapt global metric on a small patch . . .

However, we still want to relate global coordinates to a local measurement that we make anywhere on Earth. To achieve this goal, recall that on every space patch Earth's surface is effectively flat. On this patch we apply our comfortable local Cartesian coordinates, which allow us to use our super-comfortable Pythagorean Theorem—but only locally!

For example the latitude λ does not vary much across Latvia, so we can use a constant (average) $\bar{\lambda}$. Then we write:

$$\begin{aligned}\Delta s^2_{\rm Latvia} &\approx R^2\cos^2\bar{\lambda}\,\Delta\phi^2 + R^2\Delta\lambda^2 \qquad \text{(in or near Latvia)} \\ &\approx \Delta x^2_{\rm Latvia} + \Delta y^2_{\rm Latvia}\end{aligned} \tag{4}$$

. . . to make a local metric with Cartesian coordinates.

In the first line of (4) the coefficient R^2 is a constant. (We idealize the Earth as a sphere with the same radius to every point on its surface.) Then the coefficient $R^2\cos^2\bar{\lambda}$ is also constant, but in this case only across the local patch with average latitude $\bar{\lambda}$. Oh, joy! Constant coefficients allow us to define local Cartesian frame coordinates that lead to the second line in equation (4):

$$\Delta x_{\rm Latvia} \equiv R\cos\bar{\lambda}\Delta\phi \quad \text{and} \quad \Delta y_{\rm Latvia} \equiv R\Delta\lambda \qquad \text{(in or near Latvia)} \tag{5}$$

Over and over again in this book we go from a global metric to a local metric, following steps similar to those of equations (4) and (5).

Comment 2. No reverse transformation

Important note: This global-to-local conversion cannot be carried out in reverse. A local metric tells us nothing at all about the global metric from which it was derived. The reason is simple and fundamental: A space patch is, by definition, *flat*: it carries no information whatsoever about the curvature of the surface from which it was projected.

Integrate differential separation ds to calculate exact length of long path.

Global space metric (3) provides only the differential separation ds between two adjacent points that have the "vanishingly small" separation demanded by calculus. To predict the measured length of a path from Amsterdam to Vladivostok, use integral calculus to integrate ("sum") this differential ds along the entire path. *Calculus advantage:* Because all increments are vanishingly small (for which each differential patch of Earth has, in this limit, no curvature at all), their integrated sum—the total length—is completely accurate. Similarly, when we use local space metrics (1) or (2) to approximate the total length, we sum the small separations across local maps, each of which is confined to a single patch. *Multiple-patch advantage:* We can use Cartesian coordinates to make direct local measurements, then simply sum our results to obtain an approximate total distance.

Find shortest path

Suppose that our goal is to find a path of shortest length between these two cities. Along our original path, we move some of the intermediate points perpendicular to the path and recalculate its total length, repeating the calculus integration or summation until any alteration of intermediate segments no longer decreases the total path length between our fixed end locations, Amsterdam and Vladivostok. We say that the path that results from this process has the shortest length of all neighboring paths between these two cities on Earth. Everyone, using any global coordinate system or set of local frame coordinates whatsoever, agrees that we have found the path of shortest length near our original path.

Use any global coordinate system whatsoever.

Does Earth care what global coordinate system we use to indicate positions on it? Not at all! An accident of history (and international politics) fixed the zero of longitude at Greenwich Observatory near London, England. If Earth did not rotate, there would be no preferred axis capped by the north pole; we could place this pole of global coordinates anywhere on the surface.

Squiggly global coordinates lead to same predictions.

No one can stop us from abandoning latitude and longitude entirely and constructing a global coordinate system that uses a set of squiggly lines on Earth's surface as coordinate curves (subject only to some simple requirements of uniqueness and smoothness). That squiggly coordinate system leads to a global space metric more complicated than (3), but one equally capable of providing the invariant differential separation ds on Earth's surface—a differential separation whose value is identical for *every* global coordinate system. *We can use the global space metric to translate differences in (arbitrary!) global coordinates into measurable separations on a space patch.*

Many global metrics for the surface of a given potato

Generalize further: Think of a potato—or a similarly odd-shaped asteroid. Cover the potato with an inscribed global coordinate system and derive from that coordinate system a space metric that tells us the differential separation ds between any two adjacent points on the potato. Typically this space metric will be a function of coordinates as well as of coordinate differentials, because the surface of the potato curves more at some places and curves less at other places. Then change the coordinate system and find another space metric. And again. *Every* global space metric gives the *same* value of ds, the *invariant* (measureable) separation between the *same* two adjacent points on the potato.

Everyone agrees on the total length of a given path.

Next draw an arbitrary continuous curve connecting two points far apart on the potato. Use any of the metrics again to compute the total length along this curve by summing the short separations between each successive pair of points. *Result:* Since every global space metric yields the same incremental separation between each pair of nearby points on that curve, it will yield the same total length for a given curve connecting two distant points on that surface. *The length of the curve is* ***invariant****; it has the same value whatever global coordinate system we use.*

Everyone agrees that a given path is shortest.

Finally, find a curve with a shortest total length along the surface of Earth between two fixed endpoints. Since every global space metric gives the same length for a curve connecting two points on the surface, therefore every global space metric leads us to this same path of minimum length near to our original path.

One can draw a powerful analogy between the properties of a curved surface and those of curved spacetime. We now turn to this analogy.

2.4 ■ MOTION OF A STONE IN CURVED SPACETIME

A free stone moves so that its wristwatch time along each segment of its worldline is a maximum.

Relativity describes not just the separation between two nearby *points* along a traveler's *path*, but the space*time* separations between two nearby *events* that lie along the *worldline* of a moving stone. Time and space are inexorably tied together in the observation of motion.

Stone follows a straight worldline in local inertial frame.

How does a free stone move? We know the special relativity answer: With respect to an inertial frame, a free stone moves along a straight worldline, that is with constant speed on a straight trajectory in space. The Twin Paradox (Section 1.6) gives us an alternative description of free motion in an inertial frame, namely the *Principle of Maximal Aging for flat spacetime:* A free stone moves with respect to an inertial frame so that its wristwatch time between initial and final events is a maximum.

How to generalize to GR Principle of Maximal Aging?

How do we generalize the special-relativity Principle of Maximal Aging in order to predict the motion of a stone in curved spacetime? At the outset we don't know the answer to this question, so we adopt a method similar to the one we used for our trip from Amsterdam to Vladivostok: There we laid a series of adjacent flat maps along the path (Figure 5) to create a map book or atlas that displays all the maps intermediate between the two distant cities. Then we determined the incremental separation along the straight segments of path on each flat map; finally we summed these incremental separations to reckon the total length of our journey.

Start the spacetime analog with the spacetime metric in flat spacetime—equation (1.35):

$$d\tau^2 = dt^2_{\text{lab}} - ds^2_{\text{lab}} = dt^2_{\text{rocket}} - ds^2_{\text{rocket}} \qquad \text{(flat spacetime)} \qquad (6)$$

where dt_{lab} and ds_{lab} are the differential local frame time and space separations respectively between an adjacent pair of events in a particular frame, and $d\tau$ is the invariant (frame-independent) differential wristwatch time between them.

Use adjacent inertial frames.

Next we recall Einstein's "happiest thought" (initial quote) and decide to cover the stone's long worldline with a series of adjacent local inertial frames. We need to stretch differentials in (6) to give us advances in wristwatch time *that we can measure* between event-pairs along the worldline. (By definition, nobody can measure directly the "vanishingly small" differentials of calculus.) Around each pair of nearby events along a worldline we install a local inertial frame. Write the metric for each local inertial frame to reflect the fact that local spacetime is only approximately flat:

$$\Delta\tau^2 \approx \Delta t^2_{\text{inertial}} - \Delta s^2_{\text{inertial}} \qquad \text{("locally flat" spacetime)} \qquad (7)$$

This approximation for the spacetime interval is analogous to the approximate equations (1) and (4) for Latvia. Equation (7) extends rigorous spacetime metric (6) to measurable quantities beyond the reach of differentials but keeps each pair of events within a sufficiently small spacetime region so that distortions due to spacetime curvature can be ignored as we carry out a particular measurement or observation. We call such a finite region of spacetime a **spacetime patch**. The effectively flat spacetime patch allows us to extend metric (6) to a finite region in curved spacetime large enough to accommodate local coordinate increments and local measurements. Equation (7) employs these local increments, indicated by the symbol capital delta, Δ, to label a small but finite difference.

Spacetime patch

DEFINITION 2. Spacetime patch

A **spacetime patch** is a region of spacetime large enough not to be limited to differentials but small enough so that curvature does not noticeably affect the outcome of a given measurement or observation on that patch.

Comment 3. What do "large enough" and "small enough" mean?

Our definition of a patch describes its size using the phrases "large enough" and "small enough". What do these phrases mean? Can we make them exact? Sure, but only when we apply them to a particular experiment. For every experiment, we can learn how to estimate a maximum local spatial size and a maximum local time lapse of the spacetime patch so that we will not detect effects of curvature on the results of our experiment. Until we choose a specific experiment, we cannot decide whether or not it takes place in a sufficiently small spacetime patch to escape effects of spacetime curvature.

Apply special relativity in local inertial frame.

Equation (7) implies that we have applied local inertial coordinates to the patch. We call the result a **local inertial frame**, and use special relativity to describe motion in it. In particular the expression for a stone's energy—equation (28) in Section 1.7—is valid for this local frame:

$$\frac{E_{\text{inertial}}}{m} = \lim_{\Delta\tau\to 0}\frac{\Delta t_{\text{inertial}}}{\Delta\tau} = \frac{1}{\left(1 - v^2_{\text{inertial}}\right)^{1/2}} \tag{8}$$

Here v_{inertial} and E_{inertial} are the speed and energy of the stone, respectively, measured in the local inertial frame using the tools of special relativity. The *maximum* size of a local inertial frame will depend on the sensitivity of our current measurement to local curvature. However, the *minimum* size of this frame is entirely under our control. In equation (8) we go to the differential limit to describe the instantaneous speed of a stone.

We assert but do not prove that we can set up a local inertial frame—Einstein's happiest thought—almost everywhere in the Universe. For more details on the spacetime patch and its coordinates, see Section 5.7.

Use adjoining (flat) spacetime patches.

Now we generalize the special relativistic Principle of Maximal Aging to the motion of a stone in curved spacetime. Applying the Principle of Maximal Aging to a single local inertial frame tells us nothing new; it just leads to the original prediction: motion along a straight worldline in an inertial frame—this time a local one. How do we determine the effect of spacetime *curvature*? Generalize as little as possible by using *two* adjoining flat patches.

DEFINITION 3. Principle of Maximal Aging (Special and General Relativity)

The Principle of Maximal Aging says that a free stone follows a worldline through spacetime (flat or curved) such that its wristwatch time (aging) is a maximum across every pair of adjoining spacetime patches.

In Sections 1.7 and 1.8 we used the Principle of Maximal Aging to find expressions for the energy and the linear momentum, constants of motion of a free stone in flat spacetime. In Section 6.2, the Principle of Maximal Aging is

Two GR tools:
1. spacetime metric
2. Principle of Maximal Aging

central to finding an expression for the so-called *global energy,* a global constant of motion for the free stone near a black hole. Section 8.2 extends the use of the Principle of Maximal Aging to derive an expression for the so-called *global angular momentum,* a second constant of motion for a free stone near a black hole. (Near a center of attraction, linear momentum is not a constant of motion for a free stone, but angular momentum is.) Chapter 11 adapts the Principle to describe the global motion of the fastest particle in the Universe: the photon. **The spacetime metrics (global and local) and the Principle of Maximal Aging are the major tools we use to study general relativity.**

2.5 ■ GLOBAL SPACETIME METRIC IN CURVED SPACETIME

Wristwatch time between a pair of nearby events anywhere in a large spacetime region

Search for metric in global coordinates.

The cartographer laughed at us for fooling around with flat maps valid only over tiny portions of a curved surface in space. She displayed a metric (3) in global latitude and longitude coordinates, a *global space metric* that delivers the differential separation ds between two nearby stakes driven into the ground differentially close to one other anywhere on Earth's curved surface. Is there a corresponding *global spacetime metric* that delivers the differential wristwatch time $d\tau$ between adjacent events expressed in global spacetime coordinates for the curved spacetime region around, say, a black hole?

GR global metric delivers $d\tau$.

Yes! The global spacetime metric is the primary tool of general relativity. Instead of tracing a path from Amsterdam to Vladivostok across the curved surface of Earth, we want to trace the worldline of a stone through spacetime in the vicinity of a (non-spinning or spinning) Earth, neutron star, or black hole. To do this, we set up a convenient (for us) global spacetime coordinate system. We submit these coordinates plus the distribution of mass-energy (plus pressure, it turns out) to Einstein's general relativity equations. Einstein's equations return to us a global spacetime metric for our submitted coordinate system and distribution of mass-energy-pressure. This metric is the key tool that describes curved spacetime, just as the space metric in (3) was our key tool to describe a curved surface in space.

How do we use the global spacetime metric? Its inputs consist of global coordinate expressions and differential global coordinate separations—such as dt, dr, $d\phi$—between an adjacent pair of events. The output of the spacetime metric is the differential wristwatch time $d\tau$ between these events. We then convert the global metric to a local one by stretching the differentials d to increments Δ, for example in (7), that track the wristwatch time of the stone as it moves across a local inertial frame. If the stone is free—that is, if its motion follows only the command of the local spacetime structure—then the Principle of Maximal Aging tells us that the stone moves so that its summed wristwatch time is maximum across every pair of adjoining spacetime patches along its worldline.

Use any global coordinate system whatsoever.

Does the black hole care what global coordinate system we use in deriving our global spacetime metric? Not at all! General relativity allows us to use *any global coordinate system whatsoever*, subject only to some requirements of smoothness and uniqueness (Section 5.8). The metric for every alternative global coordinate system predicts the same value for the wristwatch time summed along the stone's worldline. We have (almost) complete freedom to choose our global coordinate system.

Contents of GR global metric

What does one of these global spacetime metrics around a black hole look like? On the left will be the squared differential of the wristwatch time $d\tau^2$. On the right is an expression that depends on the mass-energy-pressure of the center of attraction, on its spin if it is rotating, and on differentials of the global coordinates between adjacent events. Moreover, by analogy to equation (3) and Figure 7, the spacetime separation between adjacent events can also depend on their location, so we expect global coordinates to appear on the right side of the global spacetime metric as well. For a black hole, the result is a global spacetime metric with the general form:

$$d\tau^2 = \text{Function of} \left\{ \begin{array}{l} 1.\,\text{central mass/energy/pressure,} \\ 2.\,\text{spin, if any,} \\ 3.\,\text{global coordinate location,} \\ 4.\,\text{differentials of} \\ \ \text{global coordinates} \end{array} \right\} \quad \text{(black hole metric)} \qquad (9)$$

Curvature requires use of differentials in the metric.

Why do differentials appear in equation (9)? Think of the analogy to a spatial surface. On a (flat) Euclidean plane we are not limited to differentials, but can use total separations: the Pythagorean theorem is usually written $a^2 + b^2 = c^2$. However, on a curved surface such as that of a potato, this formula is not valid globally. The Pythagorean theorem, when applied to Earth's surface, is true only locally, in its approximate incremental form (1) and (2). Metrics in curved spacetime are similarly limited to differentials. However, we will repeatedly use transformations from global coordinates to local coordinates—similar to the global-to-flat-map transformation of equations (5)—to provide a comfortable local inertial frame metric (7) in which to make measurements and observations and to analyze results with special relativity.

Different metrics for the same and different spacetimes

Chapter 3, Curving, introduces one global spacetime metric, the Schwarzschild metric of the form (9) in the vicinity of the simplest black hole, a black hole with mass but no spin. Study of the Schwarzschild metric reveals many central concepts of general relativity, such as stretching of space and warping of time. Chapter 7, Inside the Black Hole, displays a *different* global metric for the *same* nonrotating black hole. Chapters 17 through 21 use a metric of the form (9) for a spinning black hole. Metrics with forms different from (9) describe gravitational waves (Chapter 16), and the expanding Universe (Chapters 14 and 15). In each case we apply the Principle of Maximal Aging to predict the motion of a stone or photon—and for the expanding universe the motion of a galaxy—in the region of curved spacetime under study.

Complete description of spacetime

The global coordinate system plus the global metric, taken together, provide a *complete* description of the spacetime region to which they apply, such as around a black hole. (Strictly speaking, the global coordinate system must include information about the range of each coordinate, a range that describes its "connectedness"—technical name, its *topology*.)

2.6 ■ THE DIFFERENCE BETWEEN SPACE AND SPACETIME

Cause and effect are central to science.

Minus sign in metric implies cause and effect.

The formal difference between *space* metrics such as (1) and (3) and *spacetime* metrics such as (6) and (7) is the negative sign in the spacetime metric between the space part and the time part. This negative sign establishes a fundamental relation between events in spacetime geometry: that of a possible *cause and effect*. Cause and effect are meaningless in space geometry; geometric structures are timeless (a feature that delighted the ancient Greeks). No one says, "The northern hemisphere of Earth caused its southern hemisphere." In spacetime, however, one event can *cause* some other event. (We already know from Chapter 1 that for some event-pairs, one event *cannot* cause the other.)

How is causation (or its impossibility) implied by the minus sign in the spacetime metric? See this most simply in the interval equation for flat spacetime with one space dimension:

$$\tau^2 = t_{\text{lab}}^2 - x_{\text{lab}}^2 = t_{\text{rocket}}^2 - x_{\text{rocket}}^2 \qquad \text{(flat spacetime)} \tag{10}$$

Light cones partition spacetime.

Figure 8 shows the consequences of this minus sign for events in the past and future of selected Event A. The relations between coordinates of the same event on the two diagrams are calculated using the Lorentz transformation (Section 1.10). The left panel in Figure 8 shows the laboratory spacetime diagram. Light flashes that converge on or are emitted from Event A trace out past and future *light cones*. These light cones provide boundaries for events in the past that can influence A and events in the future that A can influence. For example, thin lines that converge on Event A from events B, C, and D in its past could be worldlines of stones projected from these earlier events, any one of which could cause Event A. Similarly, thin lines diverging from A and passing through events E, F, and G in its future could be worldlines of stones projected from Event A that cause these later events.

Cause and effect are preserved.

The right panel of Figure 8 shows the rocket spacetime diagram, which displays the same events plotted in the left side laboratory diagram. The key idea illustrated in Figure 8 is that the worldline of a stone projected, for example, from Event A to event G in the laboratory spacetime diagram is transcribed as the worldline of the same stone projected from A to G (although with a different speed) in the rocket diagram. If this stone projected from A *causes* event G in one frame, then it will cause event G in both frames—and indeed in all possible inertial frames that surround Event A. *More:* As the laboratory observer clocks a stone to move with a speed less than that of light in the laboratory frame, the rocket observer also clocks the

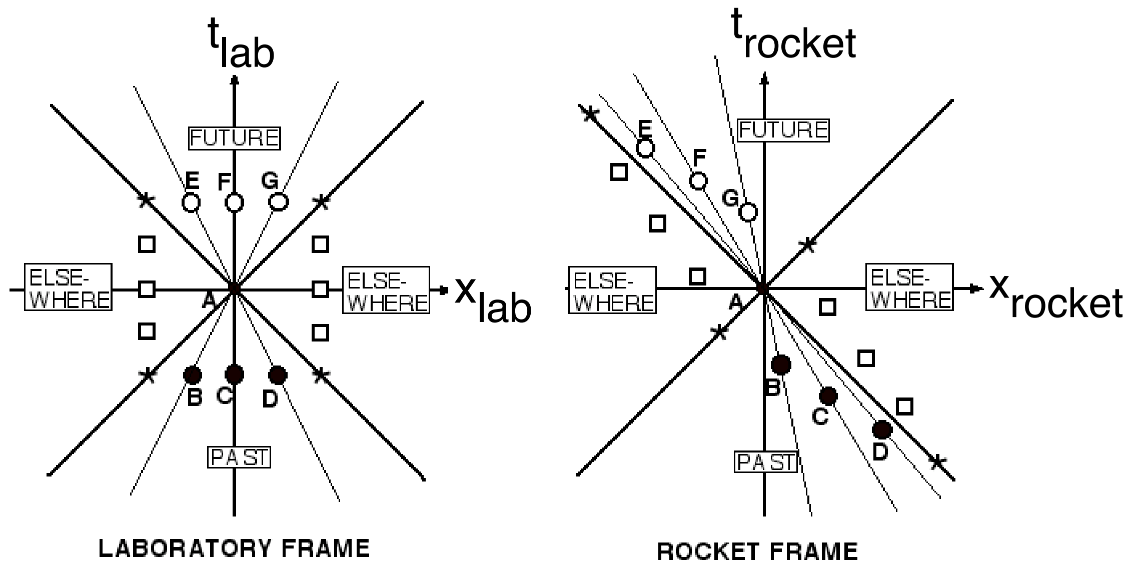


FIGURE 8 Preservation of cause and effect in special relativity. The laboratory spacetime diagram is on the left, an unpowered rocket spacetime diagram is on the right. Both diagrams plot a central Event A, and other events that may or may not cause A or be caused by A. Heavy diagonal lines are worldlines for light flashes that pass through Event A and form light cones that partition spacetime into PAST, FUTURE, and ELSEWHERE with respect to Event A. Little black-filled circles in the past of A plot events that can cause Event A in both frames. Little open circles in the future of A plot events that Event A can cause in both frames or in any other overlapping inertial frame. Little open squares plot events that cannot cause Event A and that cannot be caused by Event A in these frames or in any other inertial frame. Every ELSEWHERE event has a *spacelike* relation to Event A (Section 1.3).

stone to move with a speed less than that of light in the rocket frame. *Still more:* Events B, C, and D in the past of Event A in the laboratory frame remain in the past of Event A in the rocket frame; cause and effect can never be reversed! The spacetime interval (10) guarantees all these results and preserves cause-and-effect relationships in every physical process.

Impossibility of cause and effect is also preserved.

In contrast, events shown as little open boxes in the regions labeled ELSEWHERE in laboratory and rocket spacetime diagrams can neither cause Event A nor be caused by Event A. Why not? Because a worldline between any little box and Event A in the laboratory frame would have a slope of magnitude less than one, so a speed (the inverse of slope) greater than that of light, a speed forbidden to stone or light flash. *More:* These worldlines represent faster-than-light speed in every rocket frame as well.

No event in the regions marked ELSEWHERE can have a cause-and-effect relation with selected Event A when observed in any overlapping free-fall frame whatsoever. In this case the *impossibility* of cause and effect is guaranteed by the spacetime interval, which becomes spacelike between these two events: equation (10) becomes $\sigma^2 = s^2_{\text{frame}} - t^2_{\text{frame}}$ for any overlapping frame.

Comment 4. Before or after?
Note that some events in the ELSEWHERE region that occur *before* Event A in the laboratory frame occur *after* Event A in the rocket frame and *vice versa*. Does this destroy cause and effect? No, because none of these events can either cause Event A or be caused by Event A. Nature squeezes out of every contradiction!

Invariant wristwatch time

Figure 8 shows that time separation between event A and any event in its past or future light cone is typically different when measured in the two inertial frames, $\Delta t_{\text{rocket}} \neq \Delta t_{\text{lab}}$, as is their space separation, $\Delta x_{\text{rocket}} \neq \Delta x_{\text{lab}}$. But equation (10) assures us that the stone’s wristwatch time $\Delta\tau$ along the straight worldline between any of these events and A has the same value for the observers in any overlapping inertial frame.

Spacetime metric: the guardian of cause and effect

TWO-SENTENCE SUMMARY

The space metric—with its plus sign—is guardian of the invariant separation in space.

The spacetime metric—with its minus sign—is guardian of the invariant interval (cause and effect) in spacetime.

It gets even better: Figure 5 in Section 1.6 and the text that goes with it already tell us that the minus sign in the spacetime metric is the source of the Principle of Maximal Aging: in an inertial frame the straight worldline (which a free stone follows) is the one with *maximal* wristwatch time.

2.7 ■ DIALOG: GOODBYE “DISTANCE.” GOODBYE “TIME.”

Throw distance alone and time alone out of general relativity!

Reader: You make a big deal about using events to describe everything and using your mighty metric to connect these events. So what does the metric tell us about the *distance* between two events in curved spacetime?

Authors: The metric, by itself, tells us nothing whatsoever about the distance between two events.

Are you kidding? If general relativity cannot tell me the distance between two events, what use is it?

The word “distance” by itself does not belong in a book on general relativity.

You must be mad! Your later chapters include Expanding Universe and Cosmology, which surely describe distances. Now and then the news tells us about a more precise measurement of the time back to the Big Bang.

The word “time” by itself does not belong in a book on general relativity.

How can you possibly exclude "distance" and "time" from general relativity?

Herman Minkowski predicted this exclusion in 1908, as Einstein started his seven-year trudge from special to general relativity. Minkowski declared, "Henceforth space by itself and time by itself are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."

So Minkowski saw this coming.

Yes. We replace Minkowski's word "space" with the more precise word "distance." And get rid of his "doomed to fade" prediction, which has already taken place. Then Minkowski's up-dated statement reads, "DISTANCE BY ITSELF AND TIME BY ITSELF ARE DEAD! LONG LIVE SPACETIME!"

Spare me your dramatics. Do you mean to say that nowhere in describing general relativity do you write "the distance between these two events is 16 meters" or "the time between these two events is six years"?

Not unless we make a mistake.

So if I catch you using either one of these words—"distance" or "time"—I can shout, "Gottcha!"

Sure, if either word stands alone. Our book does talk about different kinds of distance and different kinds of time, but we try never to use either word by itself. Instead, we must always put a label on either word, even in the metric description of event separation.

Okay Dude, what are the labels for a pair of events described by the metric itself?

Differential or adjacent.

Aha, now we're getting somewhere. What do "differential" and "adjacent" mean?

"Differential" refers to the zero-limit calculus separation between events used in a metric, such as metric (6) for flat spacetime or metric (9) for curved spacetime. "Adjacent" means the same, but we also use it more loosely to label the separation between events described by a local approximate metric, such as (7).

Please give examples of "differential" separations between events in a metric.

Only three possible kinds of separation: (1) Differential ***spacelike*** *separation* $d\sigma$. *(2) Differential* ***timelike*** *separation* $d\tau$. *And of course (3) differential* ***lightlike****—"null"—separation* $d\sigma = d\tau = 0$.

But each of those is on the left side of the metric. What about coordinates on the right side of the metric?

You get to choose those coordinates yourself, so they have no direct connection to any physical measurement or observation.

You mean I can choose any coordinate system I want for the right side of the metric?

Almost. When you submit your set of global coordinates to Einstein's equations—for example when Schwarzschild submitted his black-hole global coordinates—Einstein's equations send back the metric. There are also a couple of simple requirements of coordinate uniqueness and smoothness (Section 5.8).

What other labels do you put on "distance" and "time" to make them acceptable in general relativity?

One is "wristwatch time" between events that can be widely separated along—and therefore connected by—a stone's worldline. Also we will still allow measured coordinate differences $\Delta x_{\text{inertial}}$ *and* $\Delta t_{\text{inertial}}$ *in a given local inertial frame, equation (7)—even though a purist will rightly criticize us because, even in special relativity, coordinate separations between events are different in rocket and laboratory frames.*

Tell me about Einstein's equations, since they are so almighty important.

Spacetime squirms in ways that neither a vector nor a simple calculus expression can describe. Einstein's equations describe this squirming with an advanced mathematical tool called a ***tensor****. (There are other mathematical tools that do the same thing.) After all the fuss, however, Einstein's equations deliver back a metric expressed in simple calculus; in this book we pass up Einstein's equations (until Chapter 22) and choose to start with the global metric.*

Okay, back to work: What meaning can you give to the phrase "the distance between two far-apart events," for example: *Event Number One:* The star emits a flash of light. *Event Number Two:* That flash hits the detector in my telescope.

Your statement tells us that the worldline of the light flash connects Event One and Event Two. On the way, this worldline may pass close to another star or galaxy and be deflected. The Universe expands a bit during this transit. Interstellar dust absorbs light of some frequencies, and also . . . .

Stop, stop! I do not want all that distraction. Just direct that lightlike worldline through an interstellar vacuum and into my telescope.

Okay, but those features of the Universe—intermediate stars, expansion, dust—will not go away. Do you see what you are doing?

No, what?

You are making a ***model****—some would call it a Toy Model—that uses a "clean" metric to describe the separation between you and that star. What you call "distance" springs from that model. Later you may add analysis of deflection, expansion, and dust to your model. Your final derived "distance" is a child of the final model and should be so labeled.*

To Hell with models! I want to know the Truth about the Universe.

Good luck with that! See the star over there? Observationally we know exactly three things about that star's location: (1) its apparent angle in the sky relative to other stars, (2) the redshift of its light, and (3) that its light follows a lightlike worldline to us. What do these observations tell us about that star? To answer this question, we must build a model *of the cosmos, including—with Einstein's help—a metric that describes how spacetime develops. Our model not only converts redshift to a calculated model-distance—note the label "model"—but also predicts the deflection of light that skims past an intermediate galaxy on its way to us, and so forth.*

What's the bottom line of this whole discussion?

The bottom line is that everyday ideas about the apparently simple words "distance" and "time" by themselves are fatally misleading in general relativity. Global coordinates connect local inertial frames, each of which we use to report all our measurements. We may give a remote galaxy the global radial coordinate $r = 10$ *billion light years (with you, the observer, at radial coordinate* $r = 0$*), but that coordinate difference is not a distance.*

Wait! Isn't that galaxy's distance from us 10 billion light years?

No! We did not say distance; we gave its global r*-coordinate. Remember, coordinates are arbitrary. Never, ever, confuse a simple coordinate difference between events with "the distance" (or "the time") between them. If you decide to apply some model to coordinate separations, always tell us what that model is and label the resulting separations accordingly. Again, "distance" by itself and "time" by itself have no place in general relativity.*

Okay, but I want to get on with learning general relativity. Are you going to bug me all the time with your picky distinctions between various kinds of "distances" and various kinds of "times" between events?

No. The topic will come up only when there is danger of misunderstanding.

2.8 ■ REVIEW EXERCISE

Euclid's Principle of Minimal Length vs. Einstein's Principle of Maximal Aging

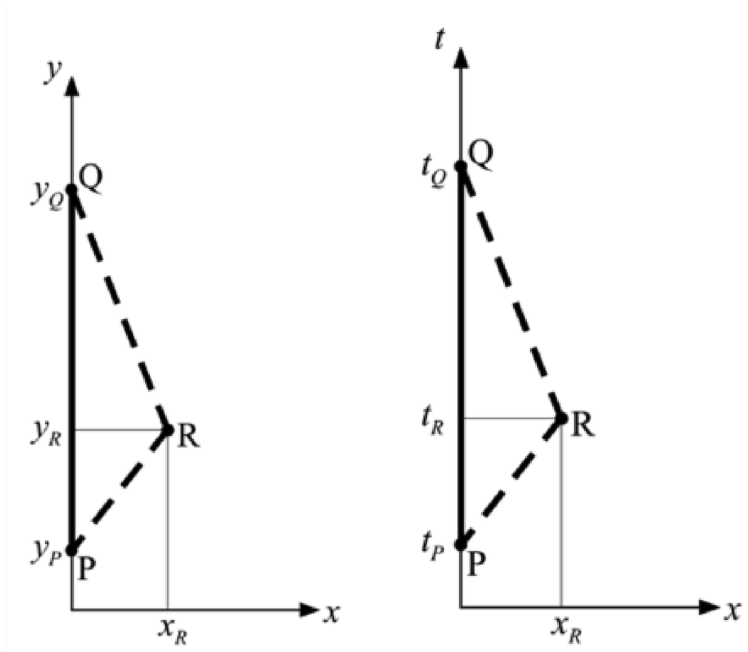


FIGURE 9 Left panel: Euclidean plane showing straight line PQ and broken line PRQ. Right panel: Spacetime diagram showing straight worldline PQ and broken worldline PRQ.

A. Consider Point P and Point Q along the y-axis of an (x, y) Cartesian coordinate system on a 2D Euclidean plane (Figure 9, left panel). Connect Point P to Point Q with a *straight* line and express the *length* of that line in terms of the coordinates of the two end points.

Now introduce an intermediate Point R slightly removed from the y-axis along the x direction, so that the line PQ is changed into a broken line PRQ in the (x, y) diagram.

B. Use the Pythagorean theorem to write an expression for the total length of broken line PRQ in terms of the coordinates of Points P, R, and Q.

C. Show that the straight line PQ is *shorter* than the broken line PRQ.

D. Describe limits, if any, on the angle that any line segment of this broken line can make with either the horizontal x or vertical y axis.

Summary: Principle of Minimal Length for Euclidean Geometry
The length of a line that connects two points is a *minimum* if the line is *straight.*

E. Next consider two Events P and Q along the t-axis of an (x, t) spacetime diagram in flat spacetime (Figure 9, right panel). Connect Point P to Point Q with a *straight* worldline and express the *wristwatch time lapse* for a stone which traverses that worldline in terms of the coordinates of the two event P and Q.

Now introduce an intermediate Event R slightly removed from the t axis in the x direction, so that the worldline PRQ is changed into a broken line in the spacetime diagram.

F. Use the interval to write the expression for the total wristwatch time of a stone that moves along the worldline PRQ in terms of the coordinates of Events P, R, and Q.

G. Show that the straight worldline PQ has a *greater* wristwatch time than the broken worldline PRQ.

H. Describe the limits, if any, on the angle that any segment of this broken worldline can make with either the horizontal x axis or the vertical t axis.

Summary: Principle of Maximal Aging for Flat Spacetime
The lapse of wristwatch time along a stone's worldline that connects two events is a *maximum* if the worldline is *straight.*

Statements in Items J, K, and L apply to both plots in Figure 9. The term *path* refers either to a Euclidean line or a spacetime worldline, and the term *extremum* refers either to a maximum or a minimum.

J. Suppose the direct path is replaced with a path with several connected straight segments. Make an argument that the straight path still has the extremum property.

K. Use the invariance principle to show that the straight path between endpoints P and Q does not need to lie along the vertical axis to satisfy the extremum property when compared with alternative paths made of several straight-line segments.

L. Show that in the "calculus limit" of a path made of an unlimited number of straight segments, alternative paths between fixed endpoints must satisfy the extremum property when compared with the straight path.

2.9 ■ REFERENCES

Initial quote of Newton's first law from *The Principia*, translation by I. Bernard Cohen and Anne Whitman, University of California Press, 1999.

Initial Einstein quote: Excerpt from an unpublished manuscript in Einstein's handwriting dating from about 1919. Quoted by Gerald Holton in *Thematic Origins of Scientific Thought*, Revised Edition, Harvard University Press, Cambridge MA, 1988, page 382. Translation from the German by Peter von Jagow.

Chapter 3. Curving

- *General relativity describes only tiny effects, right?*
- *What does "curvature of spacetime" mean?*
- *What tools can I use to visualize spacetime curvature?*
- *Do people at different r-coordinates near a black hole* age *differently? If so, can they* feel *the slowing down/speeding up of their aging?*
- *What is the "event horizon," and what weird things happen there?*
- *Do funnel diagrams describe the gravity field of a black hole?*

Download file name: Ch03Curving160511v1.pdf

CHAPTER 3

Curving

Edmund Bertschinger & Edwin F. Taylor *

In my talk ... I remarked that one couldn't keep saying "gravitationally completely collapsed object" over and over. One needed a shorter descriptive phrase. "How about black hole?" asked someone in the audience. I had been searching for just the right term for months, mulling it over in bed, in the bathtub, in my car, wherever I had quiet moments. Suddenly this name seemed exactly right. ... I decided to be casual about the term "black hole," dropping it into [a later] lecture and the written version as if it were an old family friend. Would it catch on? Indeed it did. By now every schoolchild has heard the term.

—John Archibald Wheeler with Kenneth Ford

3.1 ■ THE SCHWARZSCHILD METRIC

Spherically symmetric massive center of attraction?
The Schwarzschild metric describes the curved, empty spacetime around it.

Einstein to Schwarzschild: "splendid."

In late 1915, within a month of the publication of Einstein's general theory of relativity and just a few months before his own death from a battle-related illness, Karl Schwarzschild (1873-1916) derived from Einstein's field equations the metric for spacetime surrounding the spherically symmetric black hole. Einstein wrote to him, "I had not expected that the exact solution to the problem could be formulated. Your analytic treatment of the problem appears to me splendid."

Orbits stay in a plane.

An isolated satellite zooms around a spherically symmetric massive body. After a few orbits we discover that the satellite's motion stays confined to the initial plane determined by the satellite's position, its direction of motion, and the center of the attracting body. Why? The reason is simple: symmetry! With

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

Box 1. Metric in Polar Coordinates for Flat Spacetime

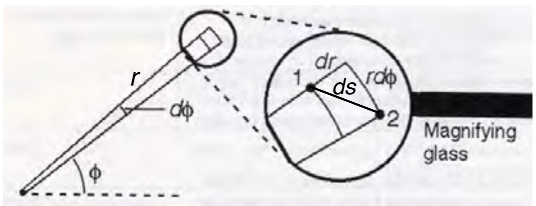


FIGURE 1 Spatial separation between two points in polar coordinates.

The metric for flat spacetime is:

$$d\tau^2 = dt^2 - ds^2 \qquad \text{(flat spacetime)} \qquad (1)$$

where ds is the spatial separation, expressed in Cartesian coordinates as

$$ds^2 = dx^2 + dy^2 \qquad \text{(flat space)} \qquad (2)$$

We look for a similar ds expression for two adjacent events numbered 1 and 2, events separated by polar coordinate increments *dr* and $d\phi$ (Figure 1).

Draw little arcs of different radii through events 1 and 2 to form a tiny box, shown in the magnified inset. The squared spatial separation between events 1 and 2 is—approximately—the sum of the squares of two adjacent sides of the little box. For a differential $d\phi$, we are entitled to express the differential space separation between event 1 and event 2 by the formula

$$ds^2 = dr^2 + r^2 d\phi^2 \qquad \text{(flat space)} \qquad (3)$$

This squared spatial separation is the space part of the squared wristwatch time differential for flat spacetime

$$d\tau^2 = dt^2 - dr^2 - r^2 d\phi^2 \qquad \text{(flat spacetime)} \qquad (4)$$

This derivation is valid only when $d\phi$ is small—vanishingly small in the calculus sense—so that the differential segment of arc $rd\phi$ is indistinguishable from a straight line. There is no such limitation to differentials for rectangular Cartesian space coordinates in flat spacetime, so each d for differential in (2) can be expanded to Δ, as it was in Section 1.10.

From Einstein's general relativity equations, Schwarzschild derived a generalization of (4) that goes beyond flat spacetime and describes curved spacetime in the vicinity of a spherically symmetric (thus non-spinning) uncharged black hole.

respect to this initial plane there is no distinction between "up out of" and "down out of" the plane, so the satellite cannot choose either and must remain in that plane. The limitation of isolated particle and light motion to a single plane greatly simplifies our analysis of physical events in this book.

We use polar coordinates (r, ϕ) for the black hole (Box 1), because polar coordinates reflect its symmetry on a plane through the black hole's center; Cartesian coordinates (x, y) do not.

Schwarzschild timelike metric

Think of two adjacent events that lie on our equatorial r, ϕ plane through the center of the black hole. These events have differential coordinate separations dt, dr, and $d\phi$. The **Schwarzschild metric** gives us the invariant $d\tau$ between this pair of events:

$$d\tau^2 = \left(1 - \frac{2M}{r}\right) dt^2 - \frac{dr^2}{\left(1 - \dfrac{2M}{r}\right)} - r^2 d\phi^2 \qquad \text{(timelike)} \qquad (5)$$

$$-\infty < t < \infty \quad \text{and} \quad 0 < r < \infty \quad \text{and} \quad 0 \le \phi < 2\pi$$

Equation (5) is the *timelike* form of the Schwarzschild metric, whose left side gives us the invariant *differential wristwatch time* $d\tau$ of a free stone that moves between a pair of adjacent events for which the magnitude of the first term on the right side is greater than the magnitude of the last two terms. In contrast, think of a pair of events for which the magnitude of the last two terms on the right predominate. Then the invariant *differential ruler distance* $d\sigma$ between these events is given by the *spacelike* form of the Schwarzschild metric:

Schwarzschild spacelike metric

$$d\sigma^2 = -d\tau^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{\left(1 - \frac{2M}{r}\right)} + r^2 d\phi^2 \quad \text{(spacelike)} \tag{6}$$

$$-\infty < t < \infty \quad \text{and} \quad 0 < r < \infty \quad \text{and} \quad 0 \leq \phi < 2\pi$$

Neither a stone nor a light flash can move between an adjacent pair of events with spacelike separation. Instead, the separation $d\sigma$ represents a differential ruler distance between two events. To make use of global metrics (5) and (6), we need to define carefully the meaning of global coordinates t, r, and ϕ. Section 3.2 shows how to measure mass in meters, so that $2M/r$ becomes unitless, as it must in order to subtract it from the unitless number one in the expression $(1 - 2M/r)$.

Meaning of "global metric"

Comment 1. Terminology: global metric

We refer to either expression (5) or (6) as a *global metric*. Professional general relativists call these expressions *line elements*; they reserve the term *metric* for the collection of coefficients of the differentials—such as $(1 - 2M/r)$, the coefficient of dt^2. We find the term *metric* to be simple, short, and clear; so in this book we use a slightly-deviant terminology and call an expression like (5) or (6) the **global metric**.

Definition: **invariant** in general relativity

DEFINITION 1. Invariant (general relativity)

Section 1.2 defined an *invariant* in special relativity as a quantity that has the same value when calculated using different *local inertial* coordinates. An **invariant** in general relativity is a quantity that has the same value when calculated using different *global* coordinate systems. Equations (5) and (6) calculate invariants $d\tau$ and $d\sigma$, respectively, using Schwarzschild global coordinates. Box 3 in Section 7.5 shows that an infinite number of global coordinate systems exist for the non-spinning black hole (indeed, for any isolated black hole). Calculation of $d\tau$ using any of these global coordinate systems delivers the same—the invariant!—value of $d\tau$ given by metrics (5) and (6).

Event horizon

Two coefficients in the Schwarzschild metric contain the expression $(1 - 2M/r)$, which goes to zero when $r \to 2M$, thus sending the first metric coefficient to zero on the right side of the metric and the magnitude of the second coefficient to infinity. This warns us about trouble at $r = 2M$, which we describe below. To the global spacetime surface at $r = 2M$ we assign the name **event horizon**, for reasons that will become clear in later sections.

It is important to realize how rare and wonderful is the Schwarzschild metric. Einstein's set of field equations is nonlinear and can be solved in

Simple global metrics are rare.

simple form only for physical systems with considerable symmetry. Schwarzschild used the symmetry of an isolated spherical non-spinning center of attraction in the derivation of his metric. This symmetry is broken—and no simple global metric exists—when we place a black hole on every street corner, although in principle a computer can provide a numerical solution of Einstein's field equations for any distribution of mass/energy/pressure. It is a measure of the scarcity of physical systems with simple metrics that almost fifty years passed before Roy Kerr found a (relatively!) simple metric for a spinning black hole in 1963 (Chapters 17 through 21).

Schwarzschild description of spacetime is **complete.**

Further investigation shows that the Schwarzschild metric plus the connectedness ("topology") of the region provides a *complete* description of spacetime external to any isolated spherically symmetric, uncharged massive body—and everywhere around such a black hole except at its central singularity (at $r = 0$), where spacetime curvature becomes infinite and general relativity fails. *Every* feature of spacetime around this kind of black hole is described or implied by the Schwarzschild metric. This one expression tells it all!

QUERY 1. Flat spacetime as $r \to \infty$

Show that as $r \to \infty$ Schwarzschild metric (5) becomes metric (4) for flat spacetime.

Ways in which the Schwarzschild metric makes sense:

We will derive the Schwarzschild metric in Chapter 22. Even now, however, we should not accept it uncritically. Here we check three ways in which it makes sense.

1. Depends only on r coordinate.

First, the expression $(1 - 2M/r)$ that appears in both the dt term and the dr term depends only on the r coordinate, not on the angle ϕ. How come? Because we are dealing with a spherically symmetric body, an object for which there is no way to tell one side from the other side or the top from the bottom. This impossibility is reflected in the absence of any direction-dependent coefficient in the metric.

2. Goes to inertial metric for zero M.

Second, the Schwarzschild metric uses coordinates that clearly show spacetime is flat when $M \to 0$, that is when there is *no* center of attraction. In this limit, the Schwarzschild metric (5) goes smoothly into the inertial metric (4) for flat spacetime.

3. Goes to local inertial metric for large r.

Third, even when M is nonzero the Schwarzschild metric (5) reduces to a local flat spacetime metric (4) very far from the black hole. The expression $(1 - 2M/r) \to 1$ when $r \to \infty$.

Schwarzschild metric applies only outside the surface.

Timelike and spacelike Schwarzschild metrics (5) and (6) describe the spacetime *external* to *any* isolated spherically symmetric, uncharged massive body. They apply with high precision to spacetime *outside* a slowly revolving massive object such as Earth or an ordinary star like our Sun. Think of a stone moving outside such an object; it makes no difference what the coordinates are inside the attracting spherical body because the stone never gets there; before it can, it collides with the surface—in the short term, our

Box 2. More About the Black Hole

John Archibald Wheeler adopted the term “black hole” in 1967 (initial quote), but the concept itself is old. As early as 1783, John Michell argued that light must “be attracted in the same manner as all other bodies” and therefore, if the attracting center is sufficiently massive and sufficiently compact, “all light emitted from such a body would be made to return toward it.” Pierre-Simon Laplace came to the same conclusion independently in 1795 and went on to reason that “it is therefore possible that the greatest luminous bodies in the universe are on this very account invisible.”

Michell and Laplace used Isaac Newton’s “action-at-a-distance” theory of gravity in analyzing the escape of light from, or its capture by, an already-existing compact object. But is such a static compact object possible? In 1939, J. Robert Oppenheimer and Hartland Snyder published the first detailed treatment of gravitational collapse within the framework of Einstein’s theory of gravitation. Their paper predicts the central features of a non-spinning black hole.

Ongoing theoretical study has shown that the black hole is the result of natural physical processes. A nonsymmetric collapsing system is not necessarily blown apart by its instabilities but can quickly—in a few seconds measured on a remote clock!—radiate away its turbulence as gravitational waves and settle down into a stable structure.

An uncharged spherically symmetric black hole is completely described by the Schwarzschild metric (plus the spacetime topology), which was derived from Einstein’s field equations by Karl Schwarzschild and published in 1916. The energy of such a non-spinning black hole cannot be milked for use outside its event horizon. For this reason, a non-spinning black hole deserves the name “dead black hole.”

In contrast to the non-spinning dead black hole, the typical black hole, like the typical star, has a spin, sometimes a large spin. The energy stored in this spin, moreover, is available for doing work: for driving jets of matter and for propelling a spaceship. In consequence, the spinning black hole deserves and receives the name “live black hole.”

The spinning black hole—or any spinning mass—drags everything in its vicinity around with it, including spacetime (Chapters 17 through 21). Near Earth this dragging is a small effect. Theory predicts that, near a rapidly-spinning black hole, such effects can be large, even irresistible, dragging along nearby spaceships no matter how powerful their rockets.

Black holes appear to be divided roughly into two groups, depending on their source: Those that result from the collapse of a single star have several times the mass our Sun. Others formed near the centers of galaxies can be monsters with millions—even billions—of times the mass of our Sun. These black holes may even shape the evolution of galaxies.

In 1963 Roy P. Kerr derived a metric for an uncharged spinning black hole. In 1967 Robert H. Boyer and Richard W. Lindquist devised a simple and convenient global coordinate system for the spinning black hole. In 2000 Chris Doran published the global coordinate system for a spinning black hole that we use in this book. In 1965 Ezra Theodore Newman and others solved the Einstein equations for the spacetime geometry around an *electrically charged* spinning black hole.

Subsequent research shows that for a steady-state black hole of specified mass, charge, and angular momentum, Kerr-Newman geometry is the *most general* solution to Einstein’s field equations. The variety, detail, and beauty of everything that forms or falls into a black hole disappears—at least according to classical (non-quantum) physics—leaving only mass, charge, and angular momentum. John Wheeler summarized this finding in the phrase, “The black hole has no hair,” which is known as the **no-hair theorem**.

Sun can be thought of as in equilibrium. The more compact the massive body, however, the larger the external region the stone can explore. Our Sun’s surface is 696 000 kilometers from its center. A cool white dwarf with the mass of our Sun has a surface r-coordinate of about 5000 kilometers, roughly that of Earth. The Schwarzschild metric describes spacetime geometry in the region external to that r-coordinate. A neutron star with the mass of our Sun has a surface r-coordinate of about 10 kilometers—the size of a typical city—so the stone can come even closer and still be “outside,” that is, in the region described correctly by the Schwarzschild metric (if the neutron star is not spinning too fast).

Box 3. Singularities: Fictitious or Real?

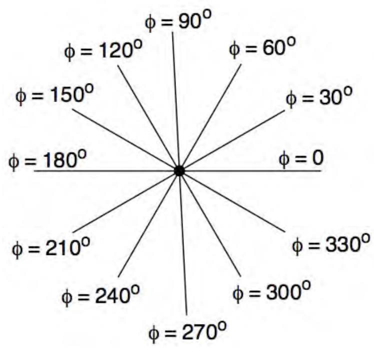


FIGURE 2 Polar coordinates on a flat Euclidean surface have a *coordinate singularity* at the center. Obviously $r = 0$ there, but what is its value of ϕ? That singularity, however, is fictitious because there is no *space* singularity at that point.

How do we know that the blow-up of the term $dr^2/(1 - 2M/r)$ at $r = 2M$ in the Schwarzschild metric does *not* signal a physical singularity? Why is this blow-up no threat to an observer falling through the event horizon—other than its one-way nature and the gradually-increasing tidal forces she feels as she descends? Einstein and others initially thought that the Schwarzschild coordinate singularity at the event horizon had a physical reality, but it does not.

Similarly, how do we know that the blow-up of the term $(1 - 2M/r)dt^2$ at $r = 0$ is lethal to all comers? How can we understand the difference between the two discontinuities in Schwarzschild coordinates?

Draw an analogy to the polar coordinate system (r, ϕ) on a flat Euclidean surface (Figure 2). The radial coordinate of the origin is clearly $r = 0$, but what is the polar angle ϕ there? *Answer:* The origin is singular in angle ϕ. *Proof:* Start at the right on the horizontal axis with label $\phi = 0$; move leftward along this axis and through the origin at $r = 0$. At this origin the axis label suddenly flips to $\phi = 180°$. There is a discontinuity of ϕ at the origin. The *coordinate* ϕ violates the requirements of uniqueness and smoothness.

The problem here is *not* Euclidean space, it is our silly (r, ϕ) coordinate system. In contrast, Cartesian coordinates $x = r\cos\phi$ and $y = r\sin\phi$ are perfectly unique and continuous at all points on the flat surface, including the origin.

Is there some way to show that there is no physical singularity at the event horizon of a non-spinning black hole? Yes, by finding a coordinate system which is perfectly smooth at the event horizon, in the same way that Cartesian coordinates in Euclidean space are perfectly smooth at the origin. In Chapter 7 we develop what we call *global rain coordinates*. At the event horizon no term blows up in the metric expressed in global rain coordinates. Global rain coordinates assign unique labels to each event and are smooth and continuous at the event horizon and all the way down to (but not including) $r = 0$.

What about the location at the center of a black hole? No coordinate system can be smooth at $r = 0$, because the so-called **Riemann curvature** is infinite there. The Riemann curvature, discovered in the 1860s by mathematician Bernhard Riemann, has a value at every spacetime event that is independent of the coordinate system. The Riemann curvature is infinite only at a physical cusp or singularity, such as the black hole singularity at $r = 0$. In contrast, the Riemann curvature is finite at $r = 2M$.

Schwarzschild describes **all** spacetime around the black hole outside the singularity.

A wonderful thing about a black hole is that it has no physical surface and no matter with which to collide. A stone can explore *all* of spacetime (except at $r = 0$) without bumping into a surface—since there is no surface at all.

?

Objection 1. *How can a black hole have "no matter with which to collide"? If it isn't made of matter, what is it made of? What happened to the star or group of stars that collapsed to form the black hole? Basically, how can something have mass without being made of matter?*

!

We think that everything that collapses into the black hole is effectively still there in some form, inducing the curvature of surrounding spacetime. This mass is crushed into a singularity at the center—along with the probe we sent in to explore it. How do we know this? We don't. What can "crushed to a singularity" possibly *mean*? We don't know. Startling? Crazy? Absurd? Welcome to general relativity!

?

Objection 2. *The global metric comes from Einstein's equations, which you say we will derive in Chapter 22. In the meantime you give us only global metrics. Why should we believe you, and why are you keeping the fundamental equations from us?*

!

Einstein's equations are most economically expressed in advanced mathematics such as *tensors*, and deriving a global metric from them is a bit tricky. In contrast, the global metric expresses itself in differentials, the working mathematics of most technical professions, and leads directly to measurable quantities: wristwatch time and ruler distance. We choose to start with the directly useful.

Next we examine the meaning of mass in units of length, so that the expression $1 - 2M/r$ in both the first and second term in the metric coefficients can have the same units, namely no units at all.

3.2 ■ MASS IN UNITS OF LENGTH

Want to reduce clutter in the metric? Then measure mass in meters!

The description of spacetime near any gravitating body is simplest when we express the mass M of that body in spatial units—in meters or kilometers. This section derives the conversion factor between, for example, kilograms and meters.

Measure mass in meters.

Earlier we wanted to measure space and time in the same unit (Section 1.2), so we used the conversion factor c, the speed of light. Conversion from kilograms to meters is not so simple. Nevertheless, here too Nature provides a conversion factor, a combination of the speed of light and Newton's **universal gravitation constant** $\boldsymbol{G}$.

Newton's theory of gravitation predicts that the gravitational force between two spherically symmetric masses M_{kg} and m_{kg} is proportional to the product of these masses and inversely proportional to the square of the Euclidean distance r between their centers:

$$F_{\text{Newtons}} = -\frac{GM_{\text{kg}}m_{\text{kg}}}{r^2} \qquad \text{(Newton, conventional units)} \tag{7}$$

In this equation G is the "constant of proportionality," whose units depend on the units with which mass and spatial separation are measured. The numerical value of G in conventional units is:

$$G = 6.67 \times 10^{-11} \frac{\text{meter}^3}{\text{kilogram second}^2} \tag{8}$$

Numerical values of G

Divide G by the square of the speed of light c^2 to find the conversion factor that translates the conventional unit of mass, the kilogram, into what we have already chosen to be the natural unit, the meter:

$$\begin{aligned}\frac{G}{c^2} &= \frac{6.67\times 10^{-11}\,\frac{\text{meter}^3}{\text{kilogram second}^2}}{8.9876\times 10^{16}\,\frac{\text{meter}^2}{\text{second}^2}} \\ &= 7.42\times 10^{-28}\,\frac{\text{meter}}{\text{kilogram}}\end{aligned} \tag{9}$$

Now convert from mass M_{kg} measured in conventional units of kilograms to mass M in meters by multiplication with this conversion factor:

$$M \equiv \frac{G}{c^2} M_{\text{kg}} = \left(7.42\times 10^{-28}\,\frac{\text{meter}}{\text{kilogram}}\right) M_{\text{kg}} \tag{10}$$

Mass in meters unclutters equations.

Why make this conversion? Because it allows us to get rid of the symbols G and c^2 that otherwise clutter up our equations.

Table 1 displays in both kilograms and meters the masses of Earth, Sun, the huge spinning black hole at the center of our galaxy, and the mass of an even larger black hole in a nearby galaxy. For each of these objects the global r-coordinate of the event horizon is twice its mass in meters. To express their masses in meters cuts planets and stars down to size!

?

Objection 3. *This is nuts! Stars and planets are not the same as space. No twisting or turning on your part can make mass and distance the same. How can you possibly propose to measure mass in units of meters?*

!

True, mass is not the same as spatial separation. Neither is time the same as space: The separation between clock ticks is different from meterstick lengths! Nevertheless, we have learned to use the conversion factor c to measure both time and space in the same unit: light-years of spatial separation and years of time, for example, or meters of spatial coordinate separation and meters of light-travel time. Payoff? The result simplifies our equations.

There are two primary birthplaces for black holes: The first is the collapse of a single star, which produces a black hole with mass equal to a modest multiple of the mass of our Sun. The second birthplace is accumulation in a galaxy, which produces a black hole with mass equal thousands to billions of the mass of our Sun. Typically, a small galaxy contains a smaller black hole, for example 50,000 times the mass of our Sun, while a large black hole, such as the last entry in Table 1, has a mass billions of times the mass of our Sun.

?

Objection 4. *You are being totally inconsistent about mass! In Chapter 1 we heard about the mass m of a stone; there you said nothing about mass in units of length. Now you define M with length units. Make up your mind!*

TABLE 3.1 Masses of some astronomical objects.

Object	Mass in kilograms	Geometric measure of mass	Equatorial r-coordinate
Earth	5.9742×10^{24} kilograms	4.44×10^{-3} meters or 0.444 centimeters	6.371×10^{6} meters or 6371 kilometers
Sun	1.989×10^{30} kilograms	1.477×10^{3} meters or 1.477 kilometers	6.960×10^{8} meters or 696 000 kilometers
Black hole at center of our galaxy	8×10^{36} kilograms (4×10^{6} Sun masses)	6×10^{9} meters	
Black hole in galaxy NGC 4889	4.2×10^{40} kilograms (21×10^{9} Sun masses)	3.1×10^{13} meters	

Excellent point. The difference between the mass M of a center of attraction and the mass m of a stone is important. First, a stone is a “free particle . . . whose mass warps spacetime too little to be measured” (inside the back cover). Second, most often we combine the stone’s mass m with another quantity in such a way that the result is a unitless ratio—for example E/m—by choosing the *same* unit in numerator and denominator. It does not matter which unit we use—joules or kilograms or electron-volts or the mass of the proton—as long as we use the *same* unit in numerator and denominator.

In contrast to the stone, the mass of a star or black hole *does* curve and warp spacetime. In this book the capital letter M *always* signals this fact. Here too we can arrange things so that M appears in a unitless ratio, such as $2M/r$, in which case M and r must have the same unit, which we choose to be meters.

Objection 5. *Okay, terrific, and this gives me a great idea: Why not simplify things even more by using unitless spacetime coordinates. Divide the Schwarzschild metric through by M^2, then define dimensionless coordinates $\tau^* \equiv \tau/M$ and $t^* \equiv t/M$ and $r^* \equiv r/M$. Here the asterisk (*) reminds us that we are using dimensionless coordinates. Now the timelike Schwarzschild metric takes the simplest possible form:*

$$d\tau^{*2} = \left(1 - \frac{2}{r^*}\right) dt^{*2} - \frac{dr^{*2}}{\left(1 - \frac{2}{r^*}\right)} - r^{*2} d\phi^2 \qquad (11)$$

(unitless coordinates)

*This notation has two big advantages: First, our equations are no longer cluttered with the symbol M, just as we have already eliminated from our equations the clutter of constants G and c. Second, metric (11) applies automatically to **all** black holes, of whatever mass M.*

Box 4. "Our Little Jugged Apocalypse"

We tend to think of a black hole as a large object, especially the "monster" at the center of our galaxy (Table 1). But the word *large* invites the question, "Large compared to what?" The diameter of the black hole in our galaxy is about 10^{-6} light year. Our galaxy, a typical one, is some 10^5 light years in diameter. Any object a factor 10^{-11} the size of a galaxy must be considered a relative dot in the galactic scheme of things. Its relatively small size allows us to call the black hole our "little jugged apocalypse," a phrase the writer John Updike uses to describe the view into the portal of a front-loading clothes-washing machine. Conveniently, spacetime curvature increases from zero far from the isolated black hole to an unlimited value at its singularity. This makes the black hole a useful example to teach large swaths—but not all—of general relativity.

Originally we used your idea for a few chapters, but then returned these chapters to our current notation, which has several advantages: (1) Keeping the M allows us to check units in every equation. An equation can be wrong if the units are correct, but it is *always* wrong if the units are incorrect! (2) We can return to flat spacetime and special relativity simply by letting $M \to 0$; a second useful check. (3) We prefer to be continually reminded of the concrete *heft*—the observed massiveness—of astronomical objects: stars and black holes. For these reasons we choose to retain coordinates in units of length and the explicit symbol M in our equations.

Newton's gravity with mass in meters

How does Newton's law of gravitation change when we express mass in meters? Think of a stone of mass m_{kg} near a center of attraction of mass M_{kg}. Rewrite Newton's second law of motion ($F = ma$) for this case, using the gravitational force equation (7), with $m_{\text{kg}} g_{\text{conv}}$ on the left, where g_{conv} is the local acceleration of gravity. The stone's mass m_{kg} cancels from both sides of the resulting equation. A minus sign signals that the acceleration is in the decreasing r direction.

$$g_{\text{conv}} = -\frac{GM_{\text{kg}}}{r^2} \qquad \text{(Newton, conventional units)} \qquad (12)$$

Now divide both sides of (12) by c^2 so as to obtain the conversion factor of equation (9). We can then write

$$g \equiv \frac{g_{\text{conv}}}{c^2} = -\frac{M}{r^2} \qquad \text{(Newton, mass in meters)} \qquad (13)$$

Newton's g_{Earth} with mass in meters

Remember that this is an equation of Newton's mechanics, not an equation of general relativity. The quantities M and r both have the unit meter, so g has the unit meter^{-1}. Substitute into (13) the values of M_{Earth} and r_{Earth} from inside the front cover to obtain the value for the acceleration of gravity g_{Earth} at Earth's surface in units of inverse meters:

$$g_{\text{Earth}} = -\frac{M_{\text{Earth}}}{r^2_{\text{Earth}}} = -1.09 \times 10^{-16}\ \text{meter}^{-1} \qquad \text{(Newton, mass in meters)} \qquad (14)$$

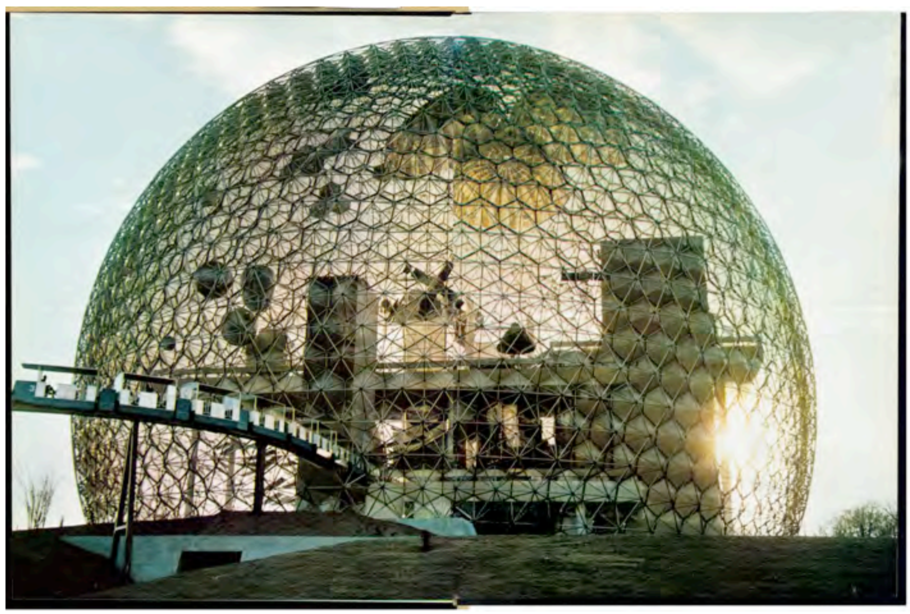


FIGURE 3 US Pavillion "geodesic dome" designed by R. Buckminster Fuller for the 1967 International and Universal Exposition in Montreal. Place a clock at every intersection of rods on the outer surface of this sphere to create a small model of our imaginary nested spherical shells concentric to a black hole. Image courtesy of the Estate of R. Buckminster Fuller.

Does this numerical value seem small? It is the same acceleration we are used to, just expressed in different units. To jump from a high place on Earth is dangerous, whatever units you use to describe your motion!

Next we continue the explanation of Schwarzschild metrics (5) and (6) with a definition of the global radial coordinate r in these equations.

3.3 ■ THE GLOBAL SCHWARZSCHILD *r*-COORDINATE

Measure the r-coordinate while avoiding the trap in the center

Why Schwarzschild global coordinates?

Section 2.5 asked, "Does the black hole care what global coordinate system we use in deriving our global spacetime metric?" and answered, "Not at all!" General relativity allows us to use *any global coordinate system whatsoever,* subject only to some requirements of smoothness and uniqueness (Section 5.9). *Next question:* Since Schwarzschild had (almost) complete freedom to choose his global coordinates t, r, and ϕ, why did he choose the particular coordinates that appear in (5) and (6)? *Next answer:* Schwarzschild's global coordinates take advantage of the spherical symmetry of a non-spinning black hole. When these coordinates are submitted to Einstein's equations, they return metrics that are (relatively!) simple. In this and the following section we introduce and describe Schwarzschild global coordinates.

Spherical shell of rods and clocks

Start with Schwarzschild's r coordinate: Take the center of attraction to be a black hole with the same mass as our Sun. In imagination, build around it a spherical shell of rods fitted together in an open mesh (Figure 3). On this shell mount a clock at every intersection of these rods. The rods and clocks of such a collection of shells provides one system of coordinates to determine the location of events that occur outside the event horizon.

We cannot measure r-coordinate directly.

How shall we define the size of the sphere formed by this latticework shell? Shall we measure directly the radial separation between the sphere's surface to its center? That won't do. Yes, in imagination we can stand on the shell. Yes, we can lower a plumb bob on a "string." But for a black hole, any string, any tape measure, any steel wire—whatever its strength—is relentlessly torn apart by the unlimited pull the black hole exerts on any object that dips close enough to its center. And even for Earth or Sun, the surface itself keeps us from lowering our plumb bob directly to the center.

Derive r-coordinate from measurement of circumference.

Therefore try another method to define the size of the spherical shell. Instead of lowering a tape measure from the shell, run a tape measure around it in a great circle. The measured distance so obtained is the *circumference* of the sphere. Divide this circumference by $2\pi = 6.283185...$ to obtain a distance that would be the directly-measured r-coordinate of the sphere *if* the space inside it were flat. But that space is *not flat,* as we shall see. Yet this procedure yields the most useful known measure of the size of the spherical shell.

The "radius" of a spherical object obtained by this method of measuring has acquired the name r**-coordinate,** because it is no genuine Euclidean radius. We call it also the **reduced circumference,** to remind us that it is derived ("reduced") from the circumference:

Definition: r**-coordinate**

$$\begin{aligned} r\text{-coordinate} &\equiv \text{reduced circumference} \\ &\equiv \frac{\text{measured circumference}}{2\pi} \end{aligned} \tag{15}$$

We sometimes use the expression Schwarzschild-r, which labels the global coordinate system of which r is a member. From now on we try not to use the word "radius" for the r-coordinate, because it can confuse results for flat spacetime with results for curved spacetime.

During construction of each shell the contractor stamps the value of its r-coordinate on it for all to see.

?

Objection 6. *Aha, gottcha! To define the r-coordinate in (15), you measure the length of the entire circumference of a spherical shell. Near a massive black hole, this circumference could be hundreds of kilometers long. Yet from the beginning you say, "Report every measurement using a local inertial frame." Near a black hole a local inertial frame is tiny compared with the length of this circumference. You do not follow your own rules for measurement.*

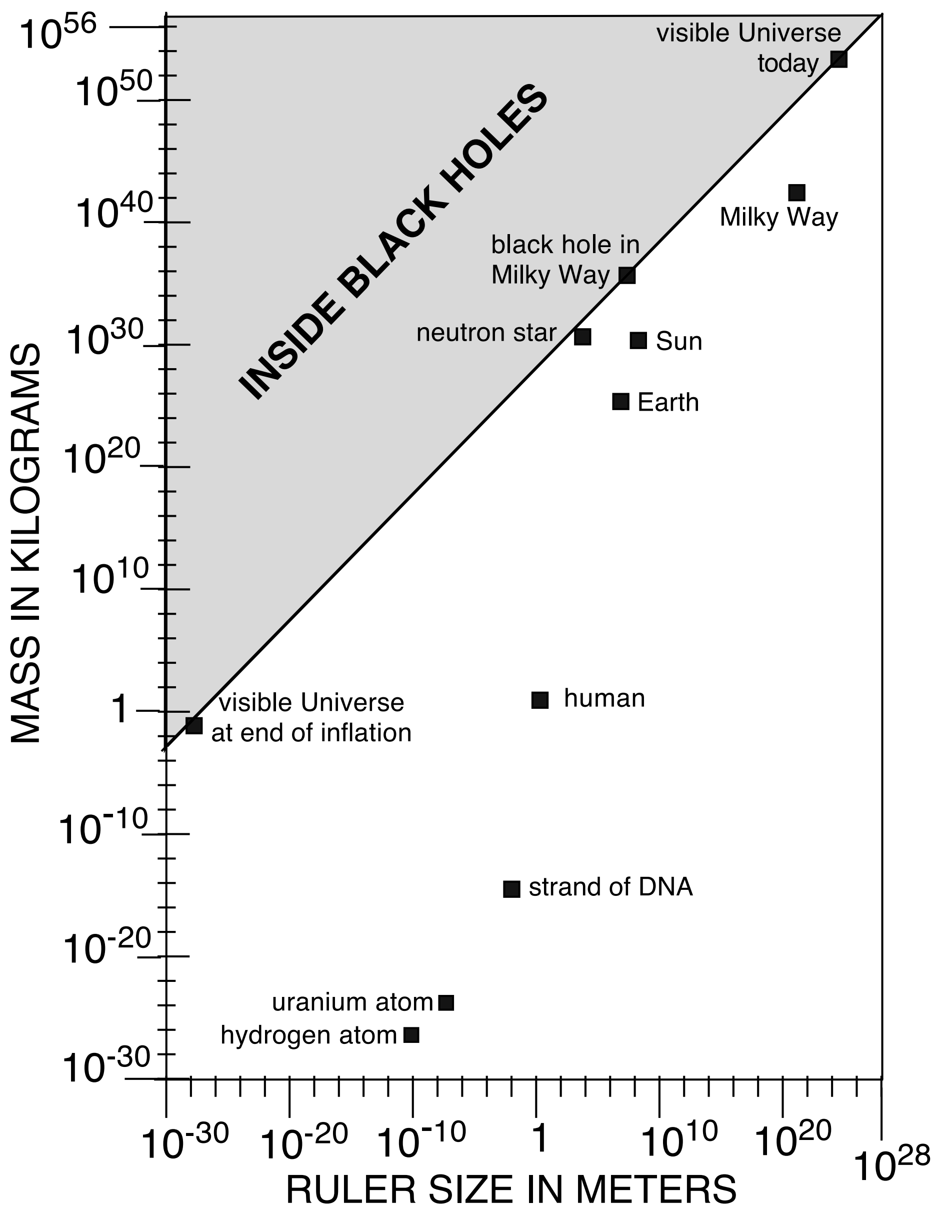


FIGURE 4 The scale of some objects described by physics. Objects close to the diagonal line are those for which correct predictions require general relativity. See Box 5. Figure adapted from the textbook *Gravity* by James Hartle.

Guilty as charged! We failed to spell out the process: Use a whole string of overlapping local inertial frames parked around the circumference of the

Box 5. When is General Relativity Necessary?

When is general relativity required to describe and predict accurately the behavior of structures and phenomena in our Universe? See Figure 4.

ORDINARY STAR. An ordinary star like our Sun does not require general relativity to account for its development, structure, or physical properties. Like all massive centers of attraction, however, it does deflect and focus passing light in ways accounted for by general relativity (Chapter 13).

WHITE DWARF. A white dwarf is the burned out cinder of an ordinary star, with a mass approximately equal to that of our Sun and r-coordinate of its surface comparable to that of Earth. General relativity is not required to account for the structure of the white dwarf but is needed to predict stability, especially near the so-called **Chandrasekhar limit** of mass—about 2.4 times the mass of our Sun—above which the white dwarf is doomed to collapse.

NEUTRON STAR. A neutron star can result from the collapse of a white dwarf star. Its mass is approximately that of our Sun with an r-coordinate of its surface about 10 kilometers, the size of a city. General relativity significantly affects the structure and oscillations of the neutron star. Emission of gravitational waves (Chapter 16) may damp out non-radial vibrations.

BLACK HOLE. "The physics of black holes calls on Einstein's description of gravity from beginning to end." (Misner, Thorne, and Wheeler)

GRAVITY WAVES. We have observed gravitational radiation predicted by general relativity.

THE UNIVERSE. Models of the Universe as a single structure employ general relativity (Chapters 14 and 15). It seems increasingly likely that general relativity correctly accounts for non-quantum features of the Universe, but it remains possible that general relativity fails over these immense spans of spacetime and must be replaced by a more general theory.

spherical shell, then define the circumference to be the summed measured distances across each of these local inertial frames. In practice this procedure is awkward, but in principle it avoids your otherwise valid objection.

Directly-measured separation between nested shells is **greater** than the difference in their r-values.

Think of building two concentric shells, a lower shell of reduced circumference $r_{\rm L}$ and a higher shell of reduced circumference $r_{\rm H}$, such that the difference in reduced circumference $r_{\rm H} - r_{\rm L}$ equals 100 meters. Stand on the higher shell and lower a plumb bob, and for the first time measure directly the radial separation perpendicularly from the higher shell to the lower one. Will we measure a 100-meter radial separation between our two shells? We would if space were flat. But outside a massive body space is *not* flat. The relation between global differential dr and measured radial differential $d\sigma$ comes from the spacelike version of the Schwarzschild metric (6) with $dt = d\phi = 0$.

$$d\sigma = \frac{dr}{\left(1 - \dfrac{2M}{r}\right)^{1/2}} \qquad \text{(radial shell separation, } dt = d\phi = 0\text{)} \qquad (16)$$

We note immediately that for the radial shell separation $d\sigma$ to be a real quantity, we must have $r > 2M$; otherwise the square root in the denominator has an imaginary value. This is an indication that shells can be built only outside the event horizon (Section 6.7).

Outside the event horizon, the magnitude of the denominator on the right side of (16) is always less than one. Hence Schwarzschild geometry tells us that every radial differential increment $d\sigma$ is *greater* than the corresponding differential increment dr of the reduced circumference. Therefore the summed (integrated) measureable distance between our two shells is greater than 100 meters, even though their circumferences differ by exactly $2\pi\times 100$ meters. This discrepancy between measured separation and difference in global r-coordinate provides striking evidence for the curvature of space. See Sample Problem 1.

Small effect near our Sun

Built around our Sun, the r-coordinate of the inner shell cannot be less than that of our Sun's surface, 695 980 kilometers. Around this inner shell we erect a second one—again in imagination—of r-coordinate 1 kilometer greater: 695 981 kilometers. The directly-measured distance between the two would be not 1 kilometer, but 2 millimeters *more* than 1 kilometer.

Get closer to the center.

How can we get closer to the center of a stellar object with mass equal to that of our Sun—but still remain external to the surface of that object? A white dwarf and a neutron star each has roughly the same mass as our Sun, but each is much smaller than our Sun. So we can—in principle—conduct a more sensitive test of the nonflatness of space much closer to the centers of these objects while staying external to them (Box 5). The effects of the curvature of space are much greater near the surface of a white dwarf than near the surface of our Sun—and greater still near the surface of a neutron star.

?

Objection 7. *Why not define the r-coordinate differently—call it $r_{\rm new}$—in terms of the directly-measured distance between two adjacent shells. For example, we could give the innermost shell at the event horizon the radial coordinate $r_{\rm new} = 2M$, and the next shell $r_{\rm new} = 2M + \Delta\sigma$, where $\Delta\sigma$ is the directly-measured separation between that shell and the innermost shell. And so on. That would eliminate the awkwardness of your quoted results.*

!

You can choose (almost) any global coordinate system you want, but the one you suggest is inconvenient. First, you cannot escape the deviation from Euclidean geometry imposed by curvature; your definition leads to a calculated circumference $2\pi r_{\rm new}$ that is different from the directly-measured one. Second, outside the event horizon your definition is awkward to carry out, since it requires collaboration between observers on different shells. Third, how is your definition applied inside the event horizon, where no shells exist? (Box 7 in Section 7.8 shows how to measure the Schwarzschild reduced circumference r inside the event horizon.) Finally, your definition of $r_{\rm new}$, when submitted with t and ϕ to Einstein's equations, results in a different metric—a more complicated one—which would be more inconvenient to use than the Schwarzschild global metric.

Huge effect near black hole

Turn attention now to a black hole of mass M. Close to it the departure from flatness is much larger than it is anywhere around a white dwarf or a neutron star. Construct an inner shell having an r-coordinate, a reduced circumference, of $3M$. Let an outer shell have an r-coordinate of $4M$. In contrast to these two r-coordinates, defined by measurements around the two shells, the directly-measured radial distance between the two shells is $1.542M$,

Sample Problem 1. "Space Stretching" Near a Black Hole

Here we verify the statement near the end of Section 3.3 that for a black hole of mass M, the directly-measured radial distance calculates as 1.542M between the lower shell at r-coordinate $r_{\rm L} = 3M$ and the higher shell at r-coordinate $r_{\rm H} = 4M$. In Euclidean geometry this measured distance would be $1.000M$, but not in curved space!

SOLUTION Equation (16) gives the radial differential $d\sigma$ between shells separated by a differential dr of the global radial coordinate r. The term $2M/r$ changes significantly over the range from $r = 3M$ to $r = 4M$, so our "summation" must be an integral. Integrating (16) from lower r-coordinate $r_{\rm L} = 3M$ to higher r-coordinate $r_{\rm H} = 4M$ yields:

$$\sigma = \int_{r_{\rm L}}^{r_{\rm H}} \frac{dr}{\left(1 - \frac{2M}{r}\right)^{1/2}}$$

$$= \int_{r_{\rm L}}^{r_{\rm H}} \frac{r^{1/2}dr}{(r-2M)^{1/2}} \qquad (17)$$

This integral is not in a common table of integrals, so make the substitution $r = z^2,$ from which $dr = 2zdz$. The resulting integral has the solution:

$$\sigma = \int_{z_{\rm L}}^{z_{\rm H}} \frac{2z^2 dz}{\left(z^2 - 2M\right)^{1/2}} \qquad (18)$$

$$= \left[z(z^2-2M)^{1/2} + 2M \ln \left| z + (z^2-2M)^{1/2} \right| \right]_{z_{\rm L}}^{z_{\rm H}}$$

Here the symbol ln (spelled "ell" "en") represents the natural logarithm (to the base *e*) and vertical-line brackets indicate absolute value. Substitute the values

$$z_{\rm L} = (3M)^{1/2} \quad \text{and} \quad z_{\rm H} = (4M)^{1/2}$$

and recall that for logarithms, $\ln(\mathrm{B}) - \ln(\mathrm{A}) = \ln(\mathrm{B/A})$. The result is

$$\sigma = 1.542M \qquad \text{(radial, exact)} \qquad (19)$$

Here the symbol σ predicts the *exact* radial separation between these shells measured by the shell observer who uses a short ruler, say one-centimeter long, laid end to end many times to find a total measured distance. This exact result is radically different from $1.000M$ predicted by Euclid.

compared to the Euclidean-geometry figure of $1.000M$ (Sample Problem 1). At this close location, the curvature of space results in measurements quite different from anything that textbook Euclidean geometry would lead us to expect. We call this effect the **stretching of space**.

?

Objection 8. *WHY is the directly-measured distance between spherical shells greater than the difference in r coordinates between these shells? Is this discrepancy caused by gravitational stretching or compression of the measuring rods?*

!

No, the quoted result assumes rigid measuring equipment. In practice, of course, a measuring rod held by the upper end will be subject to gravitational stretching (or compression if held by the lower end). Make the rod short enough; then gravitational stretching is unimportant. Now count the number of times the rod has to be moved end to end to cross from one shell to the other.

?

Objection 9. *Are you refusing to answer my question? What CAUSES the discrepancy, the fact that the directly-measured distance between spherical shells is greater than the difference in r coordinates between these shells? WHY this discrepancy?*

Sample Problem 2. Our Sun Causes Small Curvature

The Schwarzschild metric function $(1 - 2M/r)$ gauges the difference between flat and curved spacetime. How far from the center of our Sun must we be before the resulting curvature becomes extremely small or negligible?

A. As a first example, find the r-coordinate from a point mass with the mass of our Sun ($M \approx 1.5 \times 10^3$ meters) such that the metric function differs from the value one by one part in a million. Compare this r-coordinate to the actual r-coordinate of the surface of our Sun ($r_{\text{Sun}} \approx 7 \times 10^8$ meters).

B. As a second example, find the radial r-coordinate from our Sun such that the metric function differs from the value one by one part in 100 million. Compare the value of this r-coordinate with the average r-coordinate of Earth's orbit ($r \approx 1.5 \times 10^{11}$ meters).

SOLUTIONS

A. We want $(1 - 2M/r) \approx 1 - 10^{-6}$, which yields

$$r \approx \frac{2M}{10^{-6}} = 2 \times 1.5 \times 10^3 \times 10^6 \text{meters} \qquad (20)$$

$$= 3 \times 10^9 \text{meters}$$

This r-coordinate is approximately four times the r-coordinate of our Sun's surface.

B. In this case we want $(1 - 2M/r) \approx 1 - 10^{-8}$, so

$$r \approx \frac{2M}{10^{-8}} = 2 \times 1.5 \times 10^3 \times 10^8 \text{ meters} \qquad (21)$$

$$= 3 \times 10^{11} \text{ meters}$$

which is approximately twice the r-coordinate of Earth's orbit.

! A deep question! Fundamentally, this discrepancy shatters the notion of Euclidean space. We are faced with a weird measured result, which we can summarize with the statement, "Mass stretches space." Your question "Why?" is not a scientific question, and science cannot answer it. We know only observed results and their derivation from general relativity. Does the following satisfy you? *Space stretching causes the discrepancy!* Section 3.8 exhibits one way to visualize this stretching.

3.4 ■ THE GLOBAL SCHWARZSCHILD *t*-COORDINATE

Freeze global space coordinates; examine the warped t-coordinate.

To describe orbits, we need curvature of spaceTIME.

It is not enough to know the results of curvature on the r-coordinate alone. To appreciate how the grip of spacetime tells planets how to move requires us to understand how curvature affects the global t-coordinate as well. The coordinate differential dt appears on the right side of the Schwarzschild metric. Basically, Schwarzschild's definition of the t-coordinate was arbitrary, like the definition of every global coordinate.

Relation between $d\tau$ and dt

How does Schwarzschild coordinate differential dt relate to the differential wristwatch time $d\tau$ between two successive events that occur at at fixed r- and ϕ-coordinates? The coordinate differentials dr and $d\phi$ are both equal to zero for that pair of events. Then the interval between ticks is the wristwatch time derived from metric (5), that is:

$$d\tau = \left(1 - \frac{2M}{r}\right)^{1/2} dt \qquad \text{(stationary clock: } dr = d\phi = 0) \qquad (22)$$

Equation (22) shows that far from a black hole ($r \to \infty$), Schwarzschild-t coincides with the time of a shell clock located there. This is an important, but accidental, convenience of Schwarzschild's choice of global t-coordinate. It is not true for the metrics of many other global coordinate systems for the non-spinning black hole.

Slogan: "A clock at smaller r runs slower."

Now look at the prediction of equation (22) closer to a black hole—but still outside the event horizon. There the Schwarzschild coordinate differential dt will be *larger* than the differential wristwatch time $d\tau$ measured by a clock at rest on the shell at that r-coordinate. Smaller wristwatch time $d\tau$ between two standard events leads to the useful but somewhat imprecise slogan, *A clock closer to a center of attraction runs slower* (see Section 4.3).

Schwarzschild: complete description

We have now carefully defined each of the Schwarzschild global coordinates and displayed the resulting global metric handed to us by Einstein's equations, including the range of global coordinates given in equations (5) and (6). This combination—plus its connectedness (topology)–provides a *complete* description of spacetime near the isolated non-spinning black hole. These tools alone are sufficient to determine every (classical, that is non-quantum) observable property of spacetime in this region.

?

Objection 10. *Hold it! You gave us separate Sections 3.3 and 3.4 on two global coordinates, Schwarzschild-r and Schwarzschild-t, respectively. Why no section on the third global coordinate, Schwarzschild-ϕ?*

!

Good question. In answer, compare metric (4) for flat spacetime in Box 1 with the Schwarzschild metric (5) for curved spacetime. The last term is the same in both equations: $-r^2 d\phi^2$. Typical in relativity, the t-coordinate gives us the most trouble and the r-coordinate less trouble. In the non-spinning black hole metrics used in this book, the angle ϕ gives no trouble at all, due to the angular symmetry. For the spinning black hole (Chapters 17 through 21), however, even this angle becomes a troublemaker!

3.5 ■ CONSTRUCTING THE GLOBAL SCHWARZSCHILD MAP OF EVENTS

Read a road map, but don't drive on it!

"Think globally; measure locally."

In this book we choose to make every measurement and observation in a local inertial frame. But that does not suffice to describe the relation between events far from one another in the vicinity of the black hole. Suppose we know the stone's energy and momentum measured in one local inertial frame through which it passes. How can we predict the stone's energy and momentum in a second local inertial frame far from the first?

This prediction requires (a) knowledge of the stone's initial location in global coordinates, (b) analysis of the global worldline of the stone between widely-separated local frames, and (c) conversion of a piece of the global trajectory to local inertial coordinates in the remote inertial frame. This section begins that process, which we summarize with the slogan *"Think globally, measure locally."*

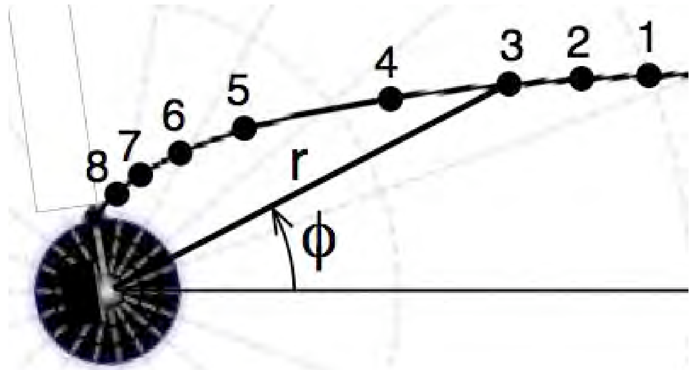


FIGURE 5 *Schwarzschild map* of the trajectory of a free stone that falls into a black hole. As it falls, it emits (numbered) flashes equally separated in time on its wristwatch. However, these flash emissions are not equally spaced along the Schwarzschild map trajectory. Each numbered event also has its Schwarzschild-t. NO ONE observes directly the entire trajectory shown on this map. *Question:* Why are numbered emission events closer together near both ends of the trajectory than in the middle of the trajectory? The answer for events 1 through 3 should be simple. The answer for events 5 through 8 appears in Section 6.5.

Global Schwarzschild coordinates locate events around a black hole similar to the way in which latitude and longitude locate places on Earth's surface (Section 2.3). A global map of Earth is nothing but a rule that assigns unique coordinates to each *point* on its surface.

The **spacetime map** assigns coordinates to every event.

By analogy, we speak of a **spacetime map**, which is nothing but a rule that assigns unique coordinates to each *event* in the region described by that map. This section describes the construction and uses of the Schwarzschild spacetime map, a task that we personalize as the work of an archivist.

Schwarzschild mapmaker

Think of Schwarzschild coordinates as an accounting system, a bookkeeping device, a spreadsheet, a tabulating mechanism, an international language, a space-and-time database created by an archivist who records every event and all motions in the entire spacetime region exterior to the surface of the Earth or Moon or Sun—or anywhere around a black hole except exactly at its center. We personify the supervisor of this record as the **Schwarzschild mapmaker**. The Schwarzschild mapmaker receives reports of actual measurements made by local shell and other inertial observers, then converts and combines them into a comprehensive description of events (in Schwarzschild coordinates) that spans spacetime around a black hole. The mapmaker makes no measurements himself and does not analyze measurements. He is a data-handler, pure and simple.

Mapmaker: the central coordinator.

The Schwarzschild mapmaker (or his equivalent) is absolutely necessary for a complete description of the motion of stones and light signals around a black hole. He has the central coordinating role in describing globally all the events that take place outside the event horizon of the black hole. He collates

Box 6. The Metric as Spacetime Micrometer

FIGURE 6 The micrometer caliper measures directly a tiny distance or thickness, bypassing x and y coordinates. The watch measures directly the invariant wristwatch time between two events, bypassing separate global coordinate increments. (Photo by Per Torphammar.)

What *is* the metric? What is it *good for*? Think of a **micrometer caliper** (Figure 6), a device used by metalworkers and other practical workers to measure a small distance. The micrometer caliper translates turns of a calibrated screw on the cylinder into the directly-measured distance across the gap between the flat ends of the little cylinders in the upper right corner of the figure. The worker *owns* the micrometer; the worker *chooses* which distance to measure with the micrometer caliper.

The metric is our "four-dimensional micrometer" that translates global coordinate separations between an adjacent pair of events into the measurable wristwatch time lapse or ruler distance between those events. You *own* the metric. You *choose* the events whose separation you wish to measure with the metric.

1. **One possible choice:** Two sequential ticks of a clock bolted to a spherical shell. Then $dr = d\phi = 0$ and the wristwatch time $d\tau$ is the time lapse read directly on the shell clock.
2. **A second possible choice:** Events with the same global t-coordinate that occur at the two ends of a stick held at rest radially between two adjacent shells, so that $dt = d\phi = 0$. Then the ruler distance $d\sigma$ is the directly-measured length of the stick—equation (16).
3. **A third possible choice:** Two sequential ticks on the wristwatch of a stone in free fall along a radial trajectory. Then $d\phi = 0$ and $d\tau$ is read directly on the wristwatch.

And so on. There are an infinite number of event-pairs near one another that you can choose for measurement using your four-dimensional micrometer—the metric.

Assembling many micrometer caliper measurements can in principle describe the geometry of space. Assembling many wristwatch and ruler measurements can in principle describe the geometry of spacetime: "The metric completely specifies local spacetime and gravitational effects within the global region in which it applies." (Inside back cover.)

What advice will the "old spacetime machinist" give to her younger colleague about the practical use of the metric? She might share the following pointers:

1. Focus on *events* and the separation between each pair of events, not fuzzy concepts like "time" or "location."
2. Do not confuse results from one pair of events with results from another pair of events.
3. Whenever possible, choose two adjacent events for which the increment of one or more map coordinates is zero.
4. Whenever possible, identify the wristwatch time or ruler distance with some observer's direct measurement.
5. When a light flash moves directly from one event to another event, the wristwatch time *and* the ruler distance between those events are both zero: $d\tau = d\sigma = 0$.

data from many local observers and combines them in various ways, for example drawing a global map such as the one plotted in Figure 5.

The Schwarzschild mapmaker can be located anywhere. How does he learn of events in his dominion? Like a taxi dispatcher, he uses radio to keep track of moving stones, light flashes, and in addition locates explosions and other events of interest, perhaps as follows:

Stamped on each spherical shell is its map r-coordinate; we mark different locations around the shell with different values of ϕ. At each location place a recording clock that reads the Schwarzschild-t (Box 6). Each clock radios to

Box 7. Where does the event horizon come from?

The event horizon—that one-way spacetime surface that lets light and stones pass inward but forbids them to cross outward—is a surprise. Who could have predicted it? Answer: Nobody did.

Newton readily predicts the gravitational consequences of a point mass, telling us immediately the initial acceleration of a stone released from rest at any r-coordinate. Twice the attracting mass, twice the stone's acceleration at that r-value; a million times the attracting mass, a million times the stone's acceleration. Newton's theory of gravity is *linear* in mass.

Not so for Einstein's general relativity, which is relentlessly *nonlinear*. In general relativity not only mass but also energy and pressure curve spacetime. A star of twice the mass typically has increased internal pressure, resulting in more than twice the gravitational effects at the same r-coordinate outside its surface. For an ordinary star the added effect of pressure is negligible; for a neutron star the added effect of pressure is important; for a black hole the added effect of pressure is catastrophic.

When a neutron star, for example, steals mass from a normal companion star, the pressure near its center increases, along with the added matter. The net result is greater than that due to the added matter alone. At a certain point, this process "runs away," resulting in collapse into a black hole.

Linear effects mean proportional response in phenomena. Nonlinear effects lead to entirely new phenomena. For the non-spinning black hole, a major outcome of nonlinearity is the event horizon. Near to the *spinning* black hole (Chapters 17 through 21), the nonlinearity of Einstein's theory leads to an even more complex geometry of spacetime and consequent radical, unexpected physical effects.

Mapmaker: top level, bureaucrat

the mapmaker the nature of an event that occurs next to it, along with its global coordinates (t, r, ϕ). After inevitable transmission delays due to the finite speed of light, the mapmaker at the control center assembles a global Schwarzschild map that gives coordinates and description of every measurement and observation. Our mapmaker acts as a top-level bureaucrat.

Using the Schwarzschild map

No one lives on a road map, but we use it to describe the territory and to plan our trip. Similarly, coordinates $r, \phi,$ and t are simply labels on a spacetime map. These coordinates uniquely locate events in the entire spacetime region outside the surface of any spherically symmetric gravitating body or anywhere around a black hole except on its singularity. The Schwarzschild map guides our navigation near a black hole, in the same way that an arbitrary set of global coordinates—made into maps—guides our travels on Earth's surface.

Map coordinate difference $\neq$ measured length or time lapse.

But never forget: In most cases Schwarzschild map coordinate separations are *not* what any local inertial observer measures directly.

Advice: It is best never to confuse a global map coordinate separation with the local inertial frame measurement of a distance or time lapse. More details in Chapter 5.

Objection 11. *Stop giving me second-hand ideas! I want* ***reality.*** *Your concept of a Schwarzschild map is nothing but an analogy to the inevitable distortions in geography when Earth's spherical surface is squashed onto a flat map. Where is the true representation of curved spacetime, corresponding to the true spherical map of Earth's surface?*

Early in the history of sea travel, mapmakers thought the world was flat. An ancient sea captain acquainted with Euclid's plane geometry (and also the much later calculus differential notation of Leibniz!) would puzzle over the metric for differential distance ds on Earth's surface, equation (3) in Section 2.3:

$$ds^2 = R^2 \cos^2 \lambda \, d\phi^2 + R^2 d\lambda^2 \qquad \text{(space metric: Earth's surface)} \quad (23)$$

The ancient sea captain asks, "What is R?" (r-coordinate of the Earth's surface). "What are λ and ϕ?" (angles of latitude and longitude). "Why does differential distance ds depend on latitude λ?" (convergence at the poles of lines of constant longitude). "Where is the edge?" (There is no edge.) Who is responsible for the captain's perplexity about a curved surface? Not Nature; not Mother Earth. Neither is Nature responsible for our perplexity about curved spacetime. Everything will be crystal clear as soon as we can visualize four-dimensional curved spacetime. But we do not know anyone who can do this; *we* certainly cannot! So we compromise, we do our best to live with our limitations and to develop intuition from the analogy to curved surfaces in space, such as the partial visualization of Schwarzschild geometry in the following sections.

Black holes just didn't "smell right"

During the 1920s and into the 1930s, the world's most renowned experts on general relativity were Albert Einstein and the British astrophysicist Arthur Eddington. Others understood relativity, but Einstein and Eddington set the intellectual tone of the subject. And, while a few others were willing to take black holes seriously, Einstein and Eddington were not. Black holes just didn't "smell right"; they were outrageously bizarre; they violated Einstein's and Eddington's intuitions about how our Universe ought to behave . . . We are so accustomed to the idea of black holes today that it is hard not to ask, "How could Einstein be so dumb? How could he leave out the very thing, implosion, that makes black holes?" Such a reaction displays our ignorance of the mindset of nearly everybody in the 1920s and 1930s . . . Nobody realized that a sufficiently compact object must *implode, and that the implosion will produce a black hole.*

—Kip Thorne

3.6 ■ THE SPACETIME SLICE

Do the best we can to visualize curved spacetime

This section introduces a method of visualizing curved spacetime—called the **spacetime slice**—that we use repeatedly throughout the book. Every such visualization of curved spacetime is partial and incomplete—it does not tell all!—but can carry us some of the way toward intuitive understanding of spacetime curvature.

DEFINITION 2. Spacetime slice

A **spacetime slice**—which we usually just call a **slice**—is a two-dimensional spacetime surface on which we plot two global coordinates of all events that lie on that surface and that have equal

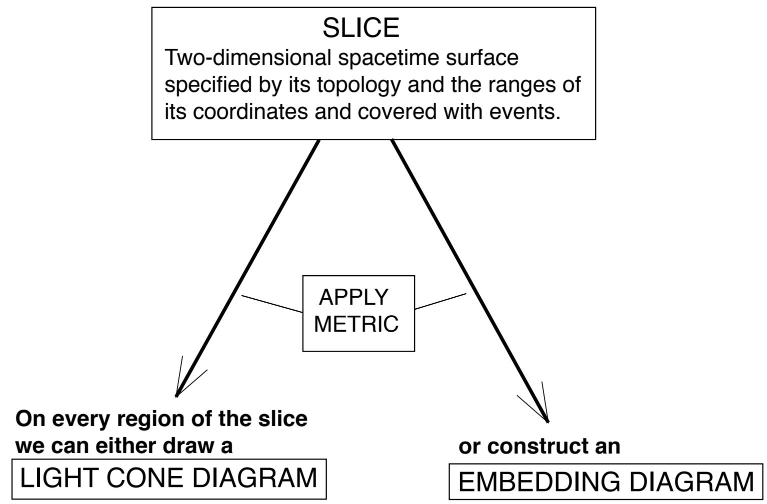


FIGURE 7 *Preview:* When we apply the global metric to a slice, then on every region of the slice we can either draw a light cone diagram or construct an embedding diagram.

Definition:
spacetime slice

values for all other global coordinates. We indicate a slice with square brackets; the three alternative slices for our Schwarzschild global coordinates are $[r,\phi]$, $[r,t]$, and $[\phi,t]$. Our definition of *slice* includes its range of coordinates and its connectedness (topology). The slice—even when populated with events—does not use the metric, so *a spacetime slice carries no information whatsoever about spacetime curvature.* This feature makes the slice useful in both special and general relativity.

On every region
of every slice:
light cone diagram or
embedding diagram

The following remarkable property of the spacetime slice will illuminate the remainder of this book: *When we apply the global metric to a spacetime slice, then on every region of every slice we can either draw worldlines or set up an embedding diagram.* Figure 7 previews the content of the following sections.

What does "every region" of the slice mean in the caption to Figure 7? For the non-spinning black hole the regions are outside and inside the event horizon. Section 3.7 shows that light cones can be drawn on both regions for the $[r,t]$ slice. Section 3.9 shows that outside the event horizon the $[r,\phi]$ slice is an embedding diagram.

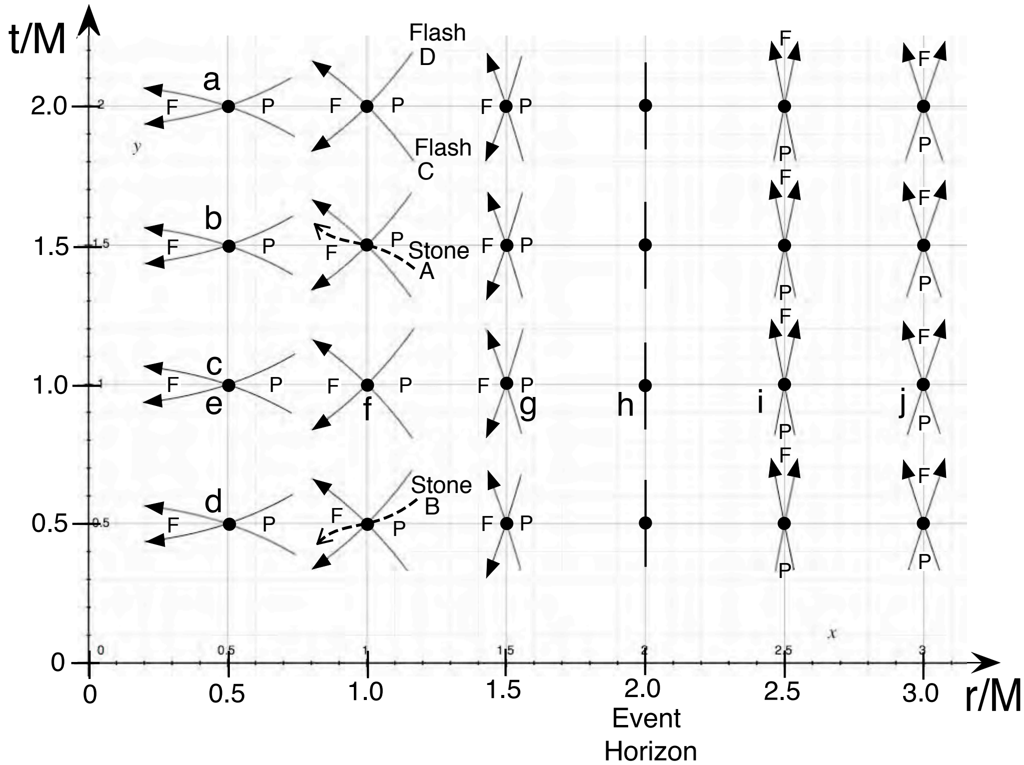


FIGURE 8 Schwarzschild light cone diagram on an $[r, t]$ slice, constructed from segments of light worldlines from equation (26), showing future (F) and past (P) of each event (filled dots). At each r-coordinate the light cone can be moved up or down vertically without change of shape, as shown.

3.7 ■ LIGHT CONE DIAGRAM ON AN [*r,t*] SLICE

The global t-coordinate can run backward along a worldline!

On an $[r, t]$ slice. . .

We can learn a lot about predictions of the Schwarzschild metric by plotting light cones. To derive the worldline of a light flash in r, t coordinates, set $d\tau = 0$ and $d\phi = 0$ in (5). The result is:

$$0 = \left(1 - \frac{2M}{r}\right) dt^2 - \left(1 - \frac{2M}{r}\right)^{-1} dr^2 \qquad \text{(light, and } d\phi = 0\text{)} \qquad (24)$$

Which leads to the equation:

$$\frac{dt}{dr} = \pm \frac{r}{r - 2M} \qquad \text{(light, radial motion)} \qquad (25)$$

Light cones

Integrate this to find the equation for the worldline of a light flash:

Box 8. A White Hole?

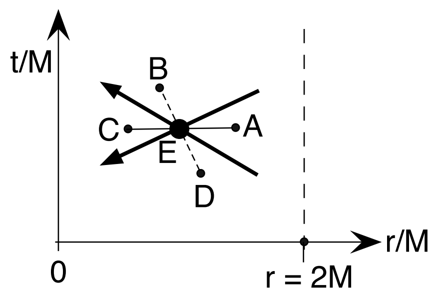


FIGURE 9 Schematic of a light cone inside the event horizon in Schwarzschild global coordinates.

Inside the event horizon, do a stone and light flash really move only toward smaller r? And does Figure 8 correctly represent this? Why do the light cones not open upward in this figure, as they do in flat spacetime and also outside the event horizon?

To answer these questions, assume that the worldline of the stone passes through event E, the intersection of the light cone worldlines in Figure 9. Then determine what worldlines through E are possible between A and C (solid line) or between D and B (dashed line). The metric tells us how the stone's wristwatch advances along its constant-ϕ worldline, From (24), it reads

$$d\tau^2 = \left(1 - \frac{2M}{r}\right) dt^2 - \left(1 - \frac{2M}{r}\right)^{-1} dr^2 \quad (27)$$

Wristwatch time in (27) is real, therefore physical, only if the right side is positive. You can show that along a worldline connecting events D and B (dashed line), the wristwatch time is imaginary. In contrast, you can show that along a worldline that connects events A and C (solid line), wristwatch time is real. *First conclusion:* Worldlines of stones that pass through event E can pass only from either the A region to the C region or from the C region to the A region. No stone worldline through event E can connect events B and D.

Next question: In which direction does the stone move between events A and C inside the event horizon? Arrows on the light cone imply that the motion is from A to C, namely to smaller r. But all differentials in (27) are squared: The metric allows motion in either direction.

We now show that motion to larger r cannot occur inside the event horizon. This means that the solution of the metric that allows motion to larger r inside the event horizon is an extraneous solution and does not correspond to the workings of Nature.

Suppose that the stone moves to larger r, from event C to event A, in which case the light cone arrows in Figure 9 would point to the right. That means that at an earlier wristwatch time the stone was at C. Now draw a light cone that crosses at event C. Then there is a still earlier event to the left of C through which the stone passed. Repeat this process until we reach $r = 0$, from which this stone must have emerged. The result is what we call a **white hole**. A white hole spews stones and light outward from its singularity, the opposite of a black hole.

Do white holes exist in Nature? We have not detected any. And if they should temporarily form, how could they possibly survive, since their central feature is to empty themselves into surrounding spacetime? The method we use here is called *reductio ad absurdum*, reduction to an absurd result.

Final conclusion: Arrows on the light cones inside the event horizon in Figure 9 point in the physically correct direction, which funnels stones and light toward the singularity. The corresponding light cones in Figure 8 do the same.

$$t - t_1 = \pm\left(r - r_1 + 2M \ln\left|\frac{r/M - 2}{r_1/M - 2}\right|\right) \qquad \text{(light, radial motion)} \quad (26)$$

where (r_1, t_1) are initial coordinates of the light flash. Figure 8 plots the resulting light cone diagram for many different values of (r_1, t_1).

Trouble at the event horizon

Figure 8 tells us a lot about physical predictions of the Schwarzschild metric. The light cone of an event tells us the past (P) and future (F) of that event. Note, first, that at the event horizon light does not change r-coordinate on this slice. Second, inside the event horizon everything moves to smaller r. The light cone corrals possible worldlines of a stone that passes through that event—such as worldlines for Stone A and Stone B in the plot. Note, third, that the t-coordinate runs backward along worldlines B and D.

Objection 12. *How can light be stuck at the event horizon, moving neither inward or outward?*

Figure 8 tells us that near the event horizon the t-coordinate changes very rapidly along a light ray, while the r coordinate changes very little. This is a problem with the Schwarzschild t coordinate that obscures observed results. We can say that the Schwarzschild t-coordinate is *diseased*, does not correctly predict observations. Chapters 6 and 7 analyze and overcome this global coordinate difficulty and show that light can fall to smaller r, but not move to larger r inside the event horizon.

Objection 13. *Oops! How can time run backward along a worldline, such as that of Stone B in Figure 8? Its arrow tends downward with respect to the t/M axis.*

Careful! Never use the word "time" by itself (Section 2.7). Only the global t-coordinate runs backward along worldlines B and D in Figure 9. Global coordinates are (almost) totally arbitrary; we choose them freely, so we cannot trust them to tell us what we will observe. Only the left side of the metric does that, for example giving us wristwatch time between two events. The wristwatch time is positive as the stone progresses along worldline B in Figure 8; and along the worldline of *every* light flash the wristwatch time is zero. Box 8 shows that the motion of both light and stones must be to smaller r inside the event horizon.

Objection 14. *Aha! I've caught you in a serious contradiction. Inside the horizon the worldline of the stone in Figure 8 is flatter than that of light. That is, the stone traverses a greater span of r coordinate per unit time than light does. The stone moves faster than light! Let's see you wiggle out of that one!*

Again you use the word "time" incorrectly and compound the error by changing r rather than moving a distance. Global coordinates are arbitrary—our choice!—and global coordinate separations are not measured quantities. This arbitrariness combines with spacetime curvature to create the distortions plotted in Figure 8. Different global coordinates give different distortions—see the same plot with different global coordinates in Figure 5, Section 7.6. For *every* global coordinate system dr/dt inside the event horizon does not *measure* the velocity of anything. We favor measurement and observation on a local flat patch, where special relativity rules. Chapter 5 has a lot more on this subject.

3.8 ■ INSIDE THE EVENT HORIZON: A LIGHT CONE DIAGRAM ON AN [r,ϕ] SLICE

Inside the event horizon, Schwarzschild-r is timelike!

On an $[r, \phi]$ slice. . .

To continue our attempt to visualize curved spacetime around a black hole, we plot light cones on an $[r, \phi]$ slice. Light plots on this slice require that $d\tau = 0$ and $dt = 0$. With these conditions, (5) becomes

$$0 = -\left(1 - \frac{2M}{r}\right)^{-1} dr^2 - r^2 d\phi^2 \qquad \text{(light, and } dt = 0) \tag{28}$$

So the trajectory of light on the $[r, \phi]$ slice satisfies the equation:

$$\frac{d\phi}{dr} = \pm\frac{1}{r^{1/2}(2M - r)^{1/2}} \qquad \text{(light, } dt = 0) \tag{29}$$

Light cones inside the event horizon

The left side of (29) is real only if $r \leq 2M$, namely at or inside the event horizon. Whoops: *The only region on the* $[r, \phi]$ *slice on which we can draw worldlines is inside the event horizon.* So what is going on *outside* the event horizon? Section 3.9 answers this question; here we plot light cones on the $[r, \phi]$ slice inside the event horizon. To integrate (29), use the substitution:

$$r = 2Mz^2 \qquad \text{so} \qquad dr = 4Mz dz \tag{30}$$

With this substitution, (29) becomes:

$$\frac{d\phi}{dz} = \pm\frac{4z}{(2z^2)^{1/2}(2 - 2z^2)^{1/2}} = \pm\frac{2}{(1 - z^2)^{1/2}} \qquad \text{(light, } dt = 0) \tag{31}$$

Integrate this to obtain:

$$\phi - \phi_1 = \pm 2\int_{z_1}^{z} \frac{dz}{(1 - z^2)^{1/2}} = \pm 2\left[\arcsin z - \arcsin z_1\right] \tag{32}$$

Substitute back from (30) to yield the integral of (29):

$$\phi - \phi_1 = \pm 2\left[\arcsin\left(\frac{r}{2M}\right)^{1/2} - \arcsin\left(\frac{r_1}{2M}\right)^{1/2}\right] \tag{33}$$

$$\text{(light, } 0 < r \leq 2M,\ 0 \leq \phi < 2\pi)$$

Light cones sprout from events at the filled dots (r_1, ϕ_1) in Figure 10. Equation (33) does not give real results for $r > 2M$. However, as r approaches $r_1 = 2M$ from below, the magnitude of the slopes of $d\phi/dr$ in (29) increases without limit, leading to the vertical lines at $r = 2M$ in the figure.

?

Objection 15. *Wait a minute! I thought we could draw light cones only on a diagram with one space axis and one time axis. Figure 10 plots light cones using two space coordinates,* r *and* ϕ*!*

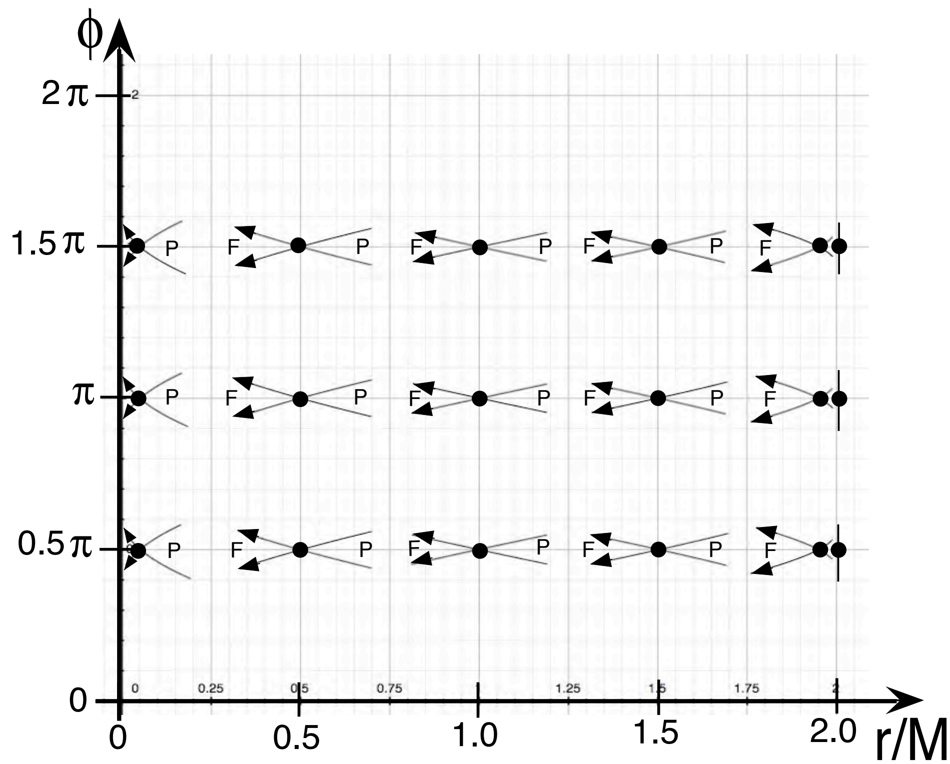


FIGURE 10 Light cones for different events (filled dots) on an $[r, \phi]$ slice inside and at the event horizon, showing the past (P) and future (F) of each event. Each light cone can be moved vertically, as shown. At $r = 2M$ the light moves neither inward nor outward, hence the vertical line. Because of the cyclic nature of ϕ, namely $\phi + 2\pi = \phi$, this diagram can be rolled up as a cylinder, on which the $\phi = 0$ axis and the $\phi = 2\pi$ line coincide.

Never assume that global coordinate separations in t, r, or ϕ tell us anything about space and time *measurements*. We favor measurement in a local inertial frame, using local coordinates—*not* global coordinates. Later we show that inside the event horizon the Schwarzschild r coordinate behaves like a time (and the Schwarzschild t coordinate behaves like a distance). So Figure 10 does describe the motion of light. The light cones in the figure fulfill one of their basic functions: For each event they divide spacetime into the past (P), the future (F), and the absolute elsewhere.

3.9 ■ OUTSIDE THE EVENT HORIZON: AN EMBEDDING DIAGRAM ON AN $[r,\phi]$ SLICE

Freeze Schwarzschild-t; examine stretched space.

On an $[r,\phi]$ slice: embedding diagram outside the event horizon

Equation (29) tells us that we cannot draw light cones on the $[r,\phi]$ slice outside the event horizon. Figure 7 predicts an alternative way to visualize curved spacetime: an **embedding diagram**. Figure 12 shows the world's most famous embedding diagram, the funnel whose form we now explain and derive.

We add a third dimension.

Think of the $[r,\phi]$ slice outside of the event horizon as an initially horizontal rubber sheet. Here's how we create the embedding diagram: Anchor a ring at $r = 2M$ on the original flat slice, then for $r > 2M$ pull the rubber sheet upward, perpendicular to that flat surface, in such a way that the curve with $d\phi = 0$, called $Z(r)$, satisfies the equation

$$d\sigma^2 = \frac{dr^2}{\left(1 - \dfrac{2M}{r}\right)} \qquad \text{(embedded surface profile)} \qquad (34)$$

Figure 11 illustrates the resulting construction. From this figure:

$$d\sigma^2 = dZ^2 + dr^2 \qquad (35)$$

From equations (34) and (35):

$$dZ^2 = d\sigma^2 - dr^2 = \frac{2M}{r-2M}dr^2 \qquad (36)$$

Take the square root of both sides of (36) and integrate the result from the lower limit at $r = 2M$:

$$Z(r) = \pm\,(2M)^{1/2}\int_{2M}^{r}\frac{dr}{(r-2M)^{1/2}} = 2^{3/2}M^{1/2}(r-2M)^{1/2} \qquad (37)$$

Parabaloid funnel

We choose the plus sign for the final expression on the right of (37) for convenience of drawing. Square both sides of (37) to obtain an equation of the form $Z^2 = Ar + B$; this shows that the funnel profile is a parabola. Rotate this curve around the vertical line $r = 0$ to create the surfaces in Figures 12 and 13. This funnel surface, with its parabola profile, is called a **paraboloid of revolution**. It is sometimes called a **gravity well** or **Flamm's paraboloid** after Ludwig Flamm, the first to identify it in 1916.

Spacetime only on funnel surface

The vertical dimension in Figures 11, 12, and 13 is an artificial construct; it is *not* a dimension of spacetime. *We ourselves added this third Euclidean space dimension to help visualize Schwarzschild geometry.* Only the embedded surface represents physical spacetime where objects and people can exist. An observer posted on this paraboloidal surface is bound to stay on that surface, not because he is physically limited in any way, but because locations off the surface in these diagrams simply do not exist in physical spacetime.

The embedding diagram in Figure 13 illustrates some analytical results derived earlier in this chapter. For example:

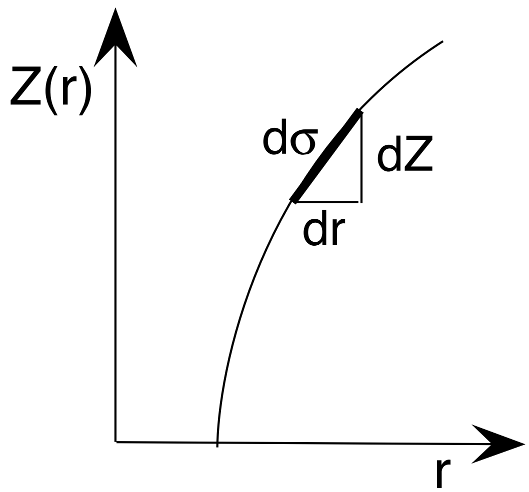


FIGURE 11 Constructing the radial profile of the funnel in Figures 12 and 13.

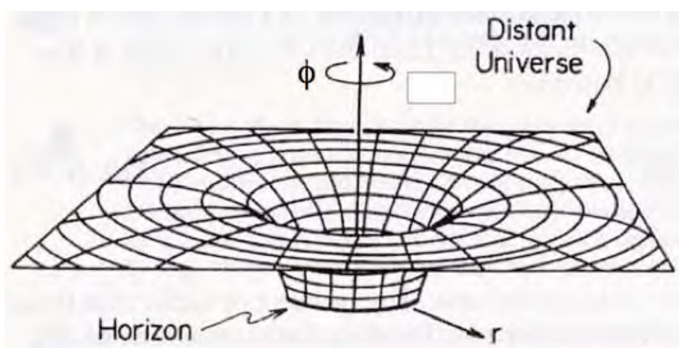


FIGURE 12 Space geometry visualized by distorting a slice through the center of a black hole, the result “embedded” in a three-dimensional Euclidean perspective. Adjacent circles represent adjacent shells. WE add the vertical dimension to show that the radial differential distance $d\sigma$ is greater than the differential dr (see Figure 13). Space stretching appears as a “bending” of the plane downwards into the shape of a funnel. At the throat of the funnel, where its slope is vertical, the r-coordinate is $r = 2M$.

1. Along the radial direction, $d\sigma$ is greater than dr, as equation (35) implies and Figure 12 illustrates.
2. The ratio $d\sigma/dr$ increases without limit as the radial coordinate decreases toward the critical value $r = 2M$ (vertical slope of the paraboloid at the throat of the funnel).

Picturing analytical results

3. The observer constrained to the paraboloid surface cannot directly measure the r-coordinate of any shell. He derives this r-coordinate—the “reduced circumference”—indirectly by measuring the circumference of the shell and dividing this circumference by 2π (Section 3.3).

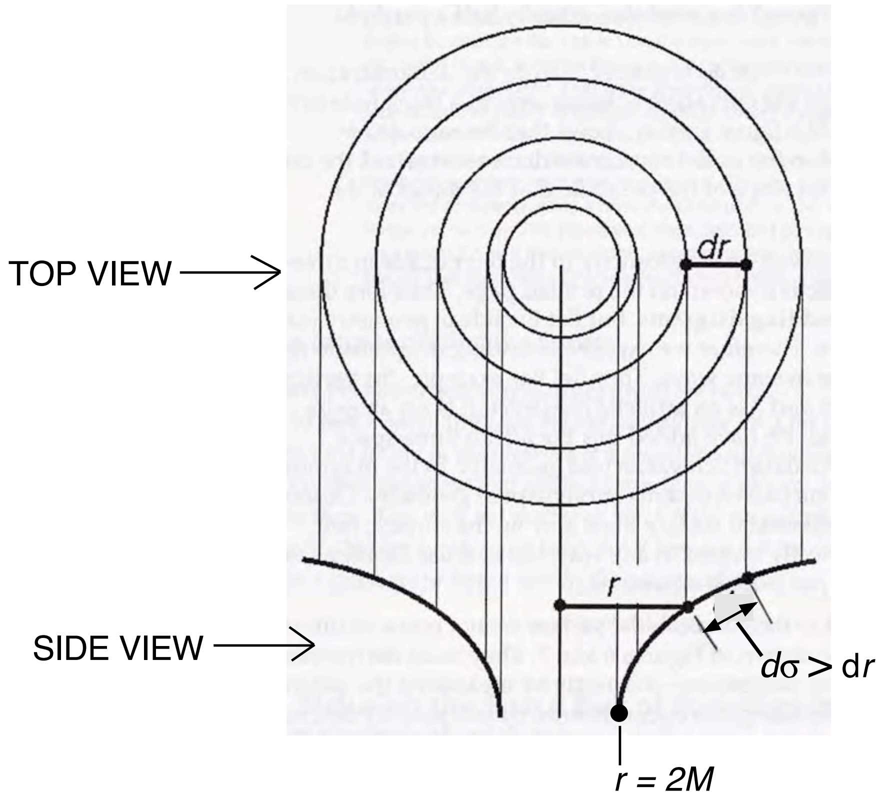


FIGURE 13 Projections of the embedding diagram of Figure 12. The thick curves in the side view are parabolas. WE choose the vertical coordinate for these curves in such a way that the increment along a parabola corresponds to the radial increment $d\sigma$ measured directly by the shell observer. A shell observer can exist only on the paraboloidal surface (shown edge-on as the thick curve). He can measure $d\sigma$ directly but not r or dr. He derives the r coordinate (*"reduced circumference"*) of a given circle by measuring its circumference and dividing by 2π. Then dr is the *computed* difference between the reduced circumferences of adjacent circles; *no* shell observer measures dr directly.

4. In contrast, the observer *can* measure the distance—call it $\sigma_{1,2}$—between adjacent shells. He finds that this directly-measured distance is greater than the difference of their r-coordinates: $\sigma_{1,2} > r_2 - r_1$.

QUERY 2. Spacelike relation of adjacent events on an embedded surface

A. Explain how on an embedded surface every adjacent pair of events—separated by differential global coordinates—has a spacelike relation to each other.

B. Argue that the answer to the question, "Can a worldline (Definition 9, Section 1.5) lie on an embedding diagram?" is a resounding "NO!"

Adjacent events on an embedding diagram have a spacelike relation.

In Query 1 you show that every pair of adjacent events on an embedded surface has a spacelike relation to one another ($d\sigma^2 > 0$). In contrast, a stone *must* move between timelike events along its worldline ($d\tau^2 > 0$). Therefore a stone *cannot* move on an embedded surface. Even light—which moves along a lightlike trajectory ($d\tau = 0$)—cannot move on an embedded surface. Hence an embedding diagram cannot display motion at all.

?

Objection 16. *In a science museum I see steel balls rolling around in a metal funnel. Is this the same as the funnel in Figure 13?*

!

No. The motion of these balls approximate Newtonian orbits provided the depth at each funnel radius is proportional to the inverse of the radius, which mimics the Newtonian potential energy. This is unrelated to the general relativistic distortion of space near a center of gravitational attraction. The cross section curve in Figure 13 is a parabola.

Comment 2. Terminology: "Except on the singularity."

Neither the Schwarzschild metric, nor any other global metric we use, is valid on the singularity of a black hole. On a singularity, by definition, spacetime curvature increases without limit, so general relativity is not valid there. In all the global coordinates we use, the non-spinning black hole has a point singularity. The spinning black hole has a ring singularity in our global coordinates (Chapter 18). We authors get tired of using—and you get tired of reading—the steady refrain "except at the singularity." So from now on that idea will mostly "go without saying." We will repeat the phrase occasionally, as a reminder, but—please!—mentally insert the phrase "except at the singularity" into every discussion of global coordinates around a black hole.

?

Objection 17. *So in summary, the space outside the event horizon of the non-spinning black hole has the shape of a funnel, right? I certainly see that funnel in textbooks and popular articles about general relativity.*

!

Here is the correct statement: "The global metric in Schwarzschild coordinates leads to a funnel embedding diagram for $r > 2M$." *Notice:* This statement describes a consequence of using Schwarzschild global coordinates. But it is not the consequence in *every* global coordinate system. Chapter 7 introduces a global coordinate system —Painlevé-Gullstrand (which we call global rain coordinates)—whose global metric leads to an embedding diagram that is **flat everywhere**, inside as well as outside the event horizon (Box 5, Section 7.6). The key idea here is that **curvature is a property of spacetime**, not of either global space coordinates alone or the global t-coordinate alone. Light cone plots and embedding diagrams help us to visualize features of curved spacetime, but no single diagram fully represents curved **spacetime.** Sorry!

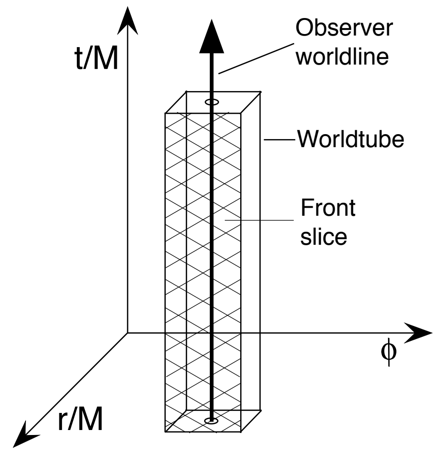


FIGURE 14 A worldtube surrounding an observer at rest in $(\phi, r/M)$ coordinates. This worldtube is bounded with slices, one of which is shaded. How "fat" the worldtube can be and still keep the the local frame of the observer inertial depends on the local spacetime curvature and the sensitivity to tides of the experiment we want to conduct.

3.10 ■ ROOM AND WORLDTUBE

Drill a hole through spacetime.

We are used to the idea of experimenting or carrying out an observation in a room. A **room** is a physical enclosure, such as (1) a laboratory, (2) a powered or unpowered spaceship, or (3) an elevator with or without its supporting cables.

Definition: **room**

DEFINITION 3. Room

A **room** is a physical enclosure of fixed spatial dimensions in which we make measurements and observations over an extended period of time.

Thus far our room is empty; we have not yet installed the rods and clocks that allow us to record and analyze events (Figure 4, Section 5.7). However, even if the room is stationary in global r and ϕ coordinates, it changes its global t-coordinate. As it does so, the room sweeps out what we call a **worldtube** in global coordinates. Figure 14 shows the worldtube of a room at rest in r and ϕ coordinates surrounding the worldline of an observer at rest in the room.

Definition: **worldtube**

DEFINITION 4. Worldtube

A **worldtube** is a bundle of worldlines of objects at rest in a room and worldlines of the structural components of that room. Think of a worldtube as sheathing the worldline of an observer at work in the room. Sometimes, but not always, we choose to bound the worldtube with spacetime slices, as in Figure 14.

Worldtube plot typically curves.

The plot of the worldtube need not be straight, since it bounds the observer's worldline, which typically curves in global coordinates. Figure 15 shows a worldtube inside the event horizon.

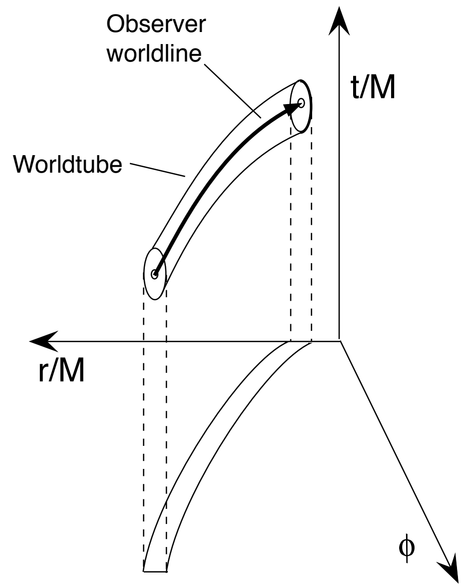


FIGURE 15 A worldtube inside the event horizon. The cross section of this particular worldtube is not rectangular; its sides are not slices in Schwarzschild coordinates. A horizontal or near-horizontal worldline is permitted inside the event horizon; see Figure 8.

In this book we prefer to make every measurement in a local inertial frame. In curved spacetime inertial frames are limited in spacetime extent. Viewed locally, each experiment takes place inside a room of limited space dimension and during a limited time lapse on clocks installed and synchronized in that room. Viewed globally, every experiment takes place within a limited segment of a worldtube.

Objection 18. *You keep saying, "In this book we prefer to make every measurement in a local inertial frame." Is this necessary? Could you describe general relativity without using local inertial frames at all?*

Yes. The timelike global metric (5) delivers, on its left side, the observed wristwatch time between two events differentially close to one another. You can integrate this differential along the worldline of a stone, for example, to find the wristwatch time between two events widely separated along this worldline. A similar distant spatial separation derives from the spacelike global metric (6). All of physics hangs on events, so all of (classical, non-quantum) physics can be analyzed without local inertial frames. Our preference for measurement in local inertial frames, where special relativity rules, is a matter of taste, clarity, and convenience for us and the reader.

3.11 ■ EXERCISES

1. Measured Distance Between Spherical Shells

A black hole has mass $M = 5$ kilometers, a little more than three times that of our Sun. Two concentric spherical shells surround this black hole. The **L**ower shell has map r-coordinate $r_{\rm L}$; the **H**igher shell has map r-coordinate $r_{\rm H} = r_{\rm L} + \Delta r$. Assume that $\Delta r = 1$ meter and consider the following four cases:

(a) $r_{\rm L} = 50$ kilometers
(b) $r_{\rm L} = 15$ kilometers
(c) $r_{\rm L} = 10.1$ kilometers
(d) $r_{\rm L} = 10.01$ kilometers
(e) $r_{\rm L} = 10.001$ kilometers

A. For each case (a) through (e), use (16) to make an estimate of the radial separation σ measured directly by a shell observer. Keep three significant digits for your estimate.
B. Next, in each case (a) through (e) use the result of Sample Problem 1 in Section 3.3 to find the exact distance between shells measured directly by a shell observer. Keep three significant digits for your result.
C. How do your estimates and exact results compare, to three significant digits, for each of the five cases? Give a criterion for the condition under which the estimate of part A will be a good approximation of the exact result of part B.

2. Grazing our Sun

Verify the statement in Section 3.4 concerning two spherical shells around our Sun. The lower shell, of reduced circumference $r_{\rm L} = 695\ 980$ kilometers, just grazes the surface of our Sun. The higher shell is of reduced circumference one kilometer greater, namely $r_{\rm H} = 695\ 981$ kilometers. Verify the prediction that the directly-measured distance between these shells will be 2 millimeters more than 1 kilometer. *Hint:* Use the approximation inside the front cover. (Outbursts and flares leap thousands of kilometers up from Sun's roiling surface, so this exercise is unrealistic—even if we could build these shells!)

3. Many Shells?

The President of the Black Hole Construction Company is waiting in your office when you arrive. He is waxing wroth. ("Tell Roth to wax [him] for awhile."— Groucho Marx)

"You are bankrupting me!" he shouts. "We signed a contract that I would build spherical shells centered on Black Hole Alpha, the shells to be 1 meter apart extending down to the event horizon. But we have already constructed the total number we thought would be required and are nowhere near finished. We are running out of materials and money!"

"Calm down a minute." you reply. "Black Hole Alpha has an event horizon r-coordinate $r = 2M = 10$ kilometers $= 10\,000$ meters. You agreed to build 1000 spherical shells starting at reduced circumference $r = 10\,001$ meters, then $r = 10\,002$ meters, then $r = 10\,003$ meters, and so forth, ending at $r = 11\,000$ meters. So what is the problem?"

"I don't know. Here is our construction method: My worker robot mounts a 1-meter rod vertically (radially) from each completed shell, measures this rod in place to be sure it is exactly 1 meter long, then welds to the top end of this rod the horizontal (tangential) beam of the next spherical shell of larger r-coordinate."

"Ah, then your company is indeed facing a large unnecessary expense," you conclude. "But I think I can tell you how you should construct the shells."

A. Explain to the President of the Black Hole Construction Company what his construction method should have been in order to fulfill his obligation to build 1000 correctly spaced spherical shells. Be specific, but do not be fussy.

B. Substitute the r-coordinate of the innermost shell into equation (16) to make a first estimate of the directly-measured separation between the innermost shell and the second shell, the one with the next-larger r-coordinate.

C. Using the r-coordinate of the second shell, the one just outside the innermost shell, make a second estimate of the directly-measured separation between the innermost shell and the second shell.

D. *Optional.* Use equation (18) to make an exact calculation of the directly-measured separation between the innermost shell and the one just outside it. How does the result of your exact calculation compare with the estimates of Parts B and C?

E. Determine the number of shells that the Black Hole Construction Company would have built if the President had completed the task according to his misunderstood plan.

4. A Dilute Black Hole

Most descriptions of black holes are apocalyptic; you get the impression that black holes are extremely dense objects. Of course a black hole is not dense throughout, because all matter quickly dives to the central crunch point. Still, one can speak of an artificial "average density," defined, say, by the total mass M divided by a spherical Euclidean volume of radius $r = 2M$. In terms of this definition, general relativity does not require that a black hole have a large average density. In this exercise you design a black hole with average density equal to that of the atmosphere you breathe on Earth, roughly 1 kilogram per cubic meter. Carry out all calculations to one-digit accuracy—we want an estimate! *Hint:* Be careful with units, especially when dealing with both conventional and geometric units.

A. From the Euclidean equation for the volume of a sphere

$$V = \frac{4}{3}\pi r^3 \qquad \text{(Euclid)}$$

find an equation for the mass M of air contained in a sphere of radius r, in terms of the density ρ in kilograms/meter^3. Use the conversion factor G/c^2 (Section 3.2) to express this mass in meters. (The volume formula used here is for Euclidean geometry, and we apply it to curved space geometry—so this exercise is only the first step in a more sophisticated analysis.)

B. Let the radius of the Euclidean spherical volume of air be equal to the map r-coordinate of the event horizon of the black hole. Assuming that our designer black hole has the density of air, what is the map r of the event horizon in terms of physical constants and air density?

C. Compare your answer to the radius of our solar system. The mean radius of the orbit of the (former!) planet Pluto is approximately 6×10^{12} meters.

D. How many times the mass of our Sun is the mass of your designer black hole?

5. Astronaut Stretching According to Newton

As you dive feet first radially toward the center of a black hole, you are not physically stress-free and comfortable. True, you detect no overall accelerating "force of gravity." But you do feel a tidal force pulling your feet and head apart and additional forces squeezing your middle from the sides like a high-quality corset. When do these tidal forces become uncomfortable? We have not yet answered this question using general relativity, but Newton is available for consultation, so let's ask him. One-digit accuracy is plenty for numerical estimates in this exercise.

A. Take the derivative with respect to r of the local acceleration g in equation (13) to obtain an expression dg/dr in terms of M and r.

We want to find the radius r_{ouch} at which you begin to feel uncomfortable. What does "uncomfortable" mean? So that we all agree, let us say that you are uncomfortable when your head is pulled upward (relative to your middle) with a force equal to the force of gravity on Earth, $\Delta g = |g_{\text{Earth}}|$, your middle is in a local inertial frame so feels no force, and your feet are pulled downward (again, relative to your middle) with a force equal to the force of gravity on Earth $\Delta g = |g_{\text{Earth}}|$.

B. How massive a black hole do you want to fall into? Suppose $M = 10$ kilometers $= 10\,000$ meters, or about seven times the mass of our Sun. Assume your head and feet are 2 meters apart. Find r_{ouch}, in meters, at which you become uncomfortable according to our criterion. Compare this radius with that of Earth's radius, namely 6.4×10^6meters.

C. Will your discomfort increase or decrease or stay the same as you continue to fall toward the center from this radius?

D. Suppose you fall from rest at infinity. How fast are you going when you reach r_{ouch} according to Newton? Express this speed as a fraction of the speed of light.

E. Take the speed in part D to be constant from that radius to the center and find the corresponding (maximum) time in meters to travel from r_{ouch} to the center, according to Newton. This will be the maximum Newtonian time lapse during which you will be—er—uncomfortable.

F. What is the maximum time of discomfort, according to Newton, expressed in seconds?

Note 1: If you carried the symbol M for the black hole mass through these equations, you found that it canceled out in expressions for the maximum time lapse of discomfort in parts E and F. In other words, your discomfort time is the same for a black hole of *any* mass when you fall from rest at infinity—according to Newton. This equality of discomfort time for all M is also true for the general relativistic analysis.

Note 2: Suppose you drop from rest starting at a great distance from the black hole. Section 7.2 analyzes the wristwatch time lapse from any radius to the center according to general relativity. Section 7.9 examines the general relativistic "ouch time."

6. Black Hole Area Never Decreases

Stephen Hawking discovered that the area of the event horizon of a black hole never decreases, when you calculate this area with the Euclidean formula $A = 4\pi r^2$. Investigate the consequences of this discovery under alternative assumptions described in parts A and B that follow.

Comment 3. Increase disorder
The rule that the area of a black hole's event horizon does not decrease is related in a fundamental way to the statistical law stating that the disorder (the so-called **entropy**) of an isolated physical system does not decrease. See Thorne, *Black Holes and Time Warps,* pages 422–426 and 445-446, and Wheeler, *A Journey into Gravity and Spacetime,* pages 218-222.

Assume that two black holes coalesce. One of the initial black holes has mass M_1 and the other has mass M_2.

A. Assume, first, that the masses of the initial black holes simply add to give the mass of the resulting larger black hole. How does the

r-coordinate of the event horizon of the final black hole relate to the r-coordinates of the event horizons of the initial black holes? How does the area of the event horizon of the final black hole relate to the areas of the event horizons of the initial black holes? Calculate the map r and area of the event horizon of the final black hole for the case where one of the initial black holes has twice the mass of the other one, that is, $M_2 = 2M_1 = 2M$; express your answers as functions of M.

B. Now make a different assumption about the final mass of the combined black hole. Listen to John Wheeler and Ken Ford (*Geons, Black Holes, and Quantum Foam,* pages 300-301) describe the coalescence of two black holes.

> *If two balls of putty collide and stick together, the mass of the new, larger ball is the sum of the masses of the balls that collide. Not so for black holes. If two spinless, uncharged black holes collide and coalesce—and if they get rid of as much energy as they possibly can in the form of gravitational waves as they combine—the* square *of the mass of the new, heavier black hole is the sum of the* squares *of the combining masses. That means that a right triangle with sides scaled to measure the [squares of the] masses of two black holes has a hypotenuse that measures the [square of the] mass of the single black hole they form when they join. Try to picture the incredible tumult of two black holes locked in each other's embrace, each swallowing the other, both churning space and time with gravitational radiation. Then marvel that the simple rule of Pythagoras imposes its order on this ultimate cosmic maelstrom.*

Following this more realistic scenario, find the r-value of the resulting event horizon when black holes of masses M_1 and M_2 coalesce. How does the area of the event horizon of the final black hole relate to the areas of the event horizons of the initial black holes?

C. Do the results of both part A and part B follow Hawking's rule that the event horizon's area of a black hole does not decrease?

D. Assume that the mass lost in the analysis of Part B escapes as gravitational radiation. What is the mass-equivalent of the energy of that gravitational radiation?

7. Zeno's Paradox

Zeno of Elea, Greece, (born about 495 BCE, died about 430 BCE) developed several paradoxes of motion. One of these states that a body in motion starting from Point A can reach a given final Point B only after having traversed half the distance between Point A and Point B. But before traversing this half it must cover half of that half, and so on *ad infinitum.* Consequently the goal can never be reached.

A modern reader, also named Zeno, raises a similar paradox about crossing the event horizon. Zeno refers us to the relation between $d\sigma$ and dr for radial separation:

$$d\sigma = \frac{dr}{\left(1 - \frac{2M}{r}\right)^{1/2}} \qquad (dt = 0, d\phi = 0) \tag{38}$$

Zeno then asserts, "As r approaches $2M$, the denominator on the right hand side of (38) goes to zero, so the distance between adjacent shells becomes infinite. Even at the speed of light, an object cannot travel an infinite distance in a finite time. Therefore nothing can arrive at the event horizon and enter the black hole." Analyze and resolve this modern Zeno's paradox using the following argument or some other method.

> As often happens in relativity, the question is: Who measures what? In order to cross the event horizon, the diving object must pass through every shell outside the event horizon. Each shell observer measures the incremental ruler length $d\sigma$ between his shell and the one below it. Then the observer on that next-lower shell measures the incremental ruler distance between that shell and the one below *it.* By adding up these increments, we can establish a measure of the "summed ruler lengths measured by shell observers from the shell at higher map r_{H} to the shell at lower map r_{L}" through which the object must move to reach the event horizon.
>
> We integrated (38) from one shell to another in Sample Problem 1 in Section 3.3. Let $r_{\mathrm{L}} \to 2M$ in that solution, and show that the resulting distance from r_{H} to r_{L}, the "summed ruler lengths," is finite as measured by the collection of collaborating shell observers. This is true even though the right side of (38) becomes infinite exactly at $r = 2M$.
>
> Will collaborating shell observers conclude among themselves that the in-falling stone reaches the event horizon? The present exercise shows that the "summed ruler lengths" is finite from any shell to the event horizon. However, motion involves not only distance but also time—and in relativity time does not follow common expectations! What can we say about the "summed shell time" for the passage of a diver through the "summed shell distance" calculated above? Chapter 6, Diving, shows that the observer on every shell measures an inertial diver to pass him with non-zero speed, a local shell speed that continues to increase as the diver gets closer and closer to the event horizon. Each shell observer therefore clocks a finite (non-infinite) time for the diver to pass from his shell to the shell below. Take the sum of these finite times—"sum" meaning an integral similar to the integral of equation (38) carried out in Sample Problem 1. When computed, this integral of shell times yields

a finite value for the total time measured by the collection of shell observers past whom the diver passes. Hence the group of shell observers agree among themselves: Someone diving radially passes them all in a finite "summed shell time" and reaches the event horizon. Thank you, Zeno!

3.12 ■ REFERENCES

Initial quote: John Archibald Wheeler with Kenneth Ford, *Geons, Black Holes and Quantum Foam*, 1998, W. W. Norton and Company, New York, pages 296-297.

The term *event horizon* was introduced by Wolfgang Rindler in 1956, reprinted in in the journal *General Relativity and Gravitation*, Volume 34, Number 1, January 2002, pages 133 through 153.

Quotes from *The Principia* by Isaac Newton translated by I. Bernard Cohen and Anne Whitman, University of California Press, 1999.

References for Box 2, Section 3, "More about the Black Hole." This box is excerpted in part from John Archibald Wheeler, "The Lesson of the Black Hole," *Proceedings of the American Philosophical Society*, Volume 125, Number 1, pages 25–37 (February 1981); J. Michell, *Philosophical Transactions of the Royal Society*, London, Volume 74, pages 35–37 (1784), cited and discussed in S. Schaffer, "John Michell and Black Holes," *Journal for the History of Astronomy*, Volume 10, pages 42–43 (1979); P.-S. Laplace, *Exposition du système du monde*, Volume 2 (Cercle-Social, Paris, 1795), modern English translation in S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge University Press, Cambridge, U.K., 1973, pages 365–368; J. R. Oppenheimer and H. Snyder, *Physical Review*, Volume 56, pages 455–459 (1939) (published the day World War II began), quoted in Stuart L. Shapiro and Saul A. Teukolsky, *Black Holes, White Dwarfs, and Neutron Stars: The Physics of Compact Objects*, John Wiley and Sons, New York, 1983, page 338; R. P. Kerr, *Physical Review Letters*, Volume 11, pages 237–238 (1963); E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence, *Journal of Mathematical Physics*, Volume 6, pages 918–919 (1965); S. W. Hawking "Black Hole Explosions?" *Nature*, Volume 248, pages 30–31 (1 March 1974); See also *Black Holes: Selected Reprints*, edited by Steven Detweiler, American Association of Physics Teachers, New York, December 1982, which includes reprints of papers by John Michell, Karl Schwarzschild, S. Chandrasekhar, J. Robert Oppenheimer, and H. Snyder, Roy P. Kerr, S. W. Hawking, and others.

Flamm's paraboloid, Figure 11 in Section 3.9: Ludwig Flamm (1916). "Beiträge zur Einstein'schen Gravitationstheorie". Physikalische Zeitschrift Volume 17, pages 448-454.

Some items in Box 5 adapted from Misner, Thorne, and Wheeler, GRAVITATION, W. H. Freeman Company, 1970, San Francisco (now New York), page 671

Quote *Black holes just didn't "smell right"* from Kip Thorne, *Black Holes and Time Warps: Einstein's Outrageous Legacy*, New York, W. W. Norton, 1994, pages 134 and 137.

Download file name: Ch03Curving170511v1.pdf

Chapter 4. Global Positioning System

- *How does the Global Positioning System [GPS] work?*
- *How accurately can I locate myself on Earth with the GPS?*
- *Why does the GPS not work when I "turn off general relativity"?*
- *What are practical uses of the GPS?*

Download file name: Ch04GlobalPositioningSystem160401.pdf

CHAPTER

4 Global Positioning System

Edmund Bertschinger & Edwin F. Taylor *

There is no better illustration of the unpredictable payback of fundamental science than the story of Albert Einstein and the Global Positioning System [GPS] . . . the next time your plane approaches an airport in bad weather, and you just happen to be wondering "what good is basic science," think about Einstein and the GPS tracker in the cockpit, guiding you to a safe landing.

—Clifford Will

4.1 ■ OPERATION OF THE GLOBAL POSITIONING SYSTEM

Relativistic effects of altitude and speed on clock rates

General relativity: Crucial to the operation of the GPS

Do you think that general relativity concerns only events far from common experience? Think again. Your hand-held Global Positioning System (GPS) receiver "listens" to overhead satellites and tells you where you are—anywhere on Earth! In this chapter you show that the operation of the GPS system depends fundamentally on general relativity.

GPS satellite system

The Global Positioning System includes a network of 24 operating satellites in circular orbits around Earth with orbital period of 12 hours, distributed in six orbital planes equally spaced in angle (Figure 1). Each satellite carries an operating atomic clock (along with several backup clocks) and emits a timed signal that also codes the satellite's location. By analyzing signals from at least four of these satellites (Box 1), your hand-held receiver on Earth displays your own location (latitude, longitude, and altitude). Consumer receivers provide a horizontal position accurate to approximately 5 meters. Among its almost endless applications, the GPS guides your driving, flying, hiking, exploring, rescuing, mapmaking, and locating your dog.

General relativity: position and motion effects

The timing accuracy required for the performance of the GPS is so great that general relativistic effects are central to its operation: First relativistic

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

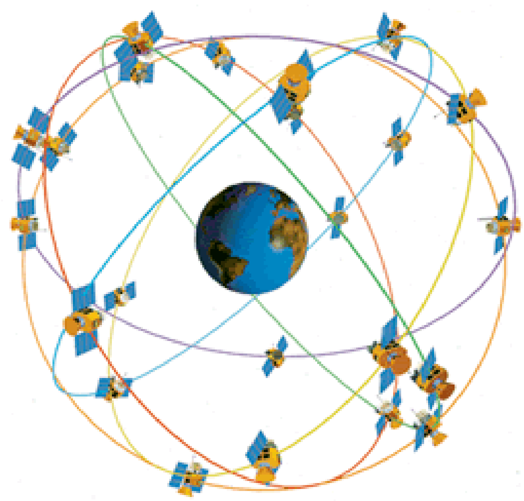


FIGURE 1 Schematic plot of GPS satellites in 12-hour orbits around Earth. Not to scale.

effect: different observed clock rates at different altitudes. Second relativistic effect, different observed rates of clocks in relative motion. (In this first analysis of the GPS, we assume all signal propagation occurs in a vacuum.)

4.2 ■ STATIONARY CLOCKS

Warping of t-coordinate at different altitudes.

Simplest case: clock on tower and no Earth rotation.

The Global Positioning System depends on the reception by a receiver on Earth's surface of microwave signals from multiple overhead satellites. Begin with the simplest possible case: Earth does not rotate and the higher clock is not in a satellite but rather sits on top of a tower at Higher r-coordinate, $r_{\rm H}$. The tower clock communicates with us on Earth, at Lower r-coordinate, $r_{\rm L}$. Calculate the radially-downward dr/dt of microwaves that move from the top to the bottom of the tower. Light and microwaves move at this same rate. For these conditions, $d\phi = 0$ and for light, $d\tau = 0$. Then the Schwarzschild metric, equation (3.5), yields the following radial motion in global coordinates:

$$\frac{dr}{dt} = -\left(1 - \frac{2M}{r}\right) \qquad \text{(light moving radially inward)} \qquad (1)$$

Map speed of light $\neq 1$.

Is equation (1) a surprise? For the first time in our study of relativity, calculated light speed differs from one meter of distance per meter of time. Ah, but the expression dr/dt in *global* coordinates is a unicorn, not measured by anyone. We need to go back and determine the observable wristwatch time lapse between two flashes emitted from the clock at the top of the tower. As usual, the metric converts from global coordinate separations (on its right side) to measured wristwatch time lapse (on its left side).

Box 1. Practical Operation of the Global Positioning System

The goal of the Global Positioning System (GPS) is to determine your position on Earth in three dimensions: east-west, north-south, and vertical—longitude, latitude, and altitude. Signals from three overhead satellites provide this information. Each satellite emits a signal that encodes its local time of emission and the satellite's position in global coordinates at that emission event, this position continually revised using data uploaded from control stations on the ground. The local clock in your hand-held GPS receiver records the local time of reception of each signal, then subtracts the emission t (encoded with the incoming signal) to determine the lapse in t-coordinate and hence how far the signal has traveled at the speed of light in global coordinates. This is the map distance the satellite was from your position when it emitted the signal. In effect, the receiver constructs three spheres from these distances, one sphere centered on the emission point of each satellite. Simple triangulation locates the point where the three spheres intersect. That point is your location in global coordinates.

Of course there is a wrinkle: The local clock in your hand-held receiver is not nearly so accurate as the atomic clocks carried in a satellite. For this reason, the signal from a fourth overhead satellite is used to correct the local clock in your receiver. This fourth signal enables your hand receiver to process GPS signals as though it contained an atomic clock.

Signals exchanged between atomic clocks at different altitudes and moving at different speeds are subject to general relativistic effects. Neglect these effects and the GPS is useless (Box 3).

Tower clock emits two downward flashes.

The clock at the top of the tower emits two flashes radially downward (emission events A and B) differentially close together in global t-coordinate: dt_{AB}. For this top tower clock, $dr = 0$ and $d\phi = 0$, the metric tells us the corresponding wristwatch time lapse $d\tau_{\text{H}}$ recorded on the tower clock:

$$d\tau_{\text{H}} = \left(1 - \frac{2M}{r_{\text{H}}}\right)^{1/2} dt_{\text{AB}} \qquad (d\phi = 0,\ dr = 0) \tag{2}$$

Map t-lapse between flashes is constant during descent.

Figure 2 traces on an $[r, t]$ slice the radially-downward global worldlines of the two flashes emitted by the tower clock at events A and B. The Earth clock receives these flashes at events C and D with t-coordinate separation dt_{CD}. Equation (1) tells us that these worldlines have identical slopes (the radial global coordinate speed of light has the same value) at every intermediate r-coordinate. As a result, the two worldlines are parallel at every r-coordinate on the $[r, t]$ slice, so the global t-coordinate separation between them maintains its initial value dt_{AB}. The two flashes arrive at the ground with the initial difference in global t-coordinate.

$$dt_{\text{CD}} = dt_{\text{AB}} \tag{3}$$

The clock on Earth's surface is also at fixed r_{L}. Therefore its wristwatch time lapse on the ground between the reception of events is similarly given by (2):

$$d\tau_{\text{L}} = \left(1 - \frac{2M}{r_{\text{L}}}\right)^{1/2} dt_{\text{CD}} = \left(1 - \frac{2M}{r_{\text{L}}}\right)^{1/2} dt_{\text{AB}} \tag{4}$$

The final step in equation (4) comes from (3). Equations (2) and (4) give us the relation between wristwatch time lapses of stationary clocks at higher and lower global r-coordinates:

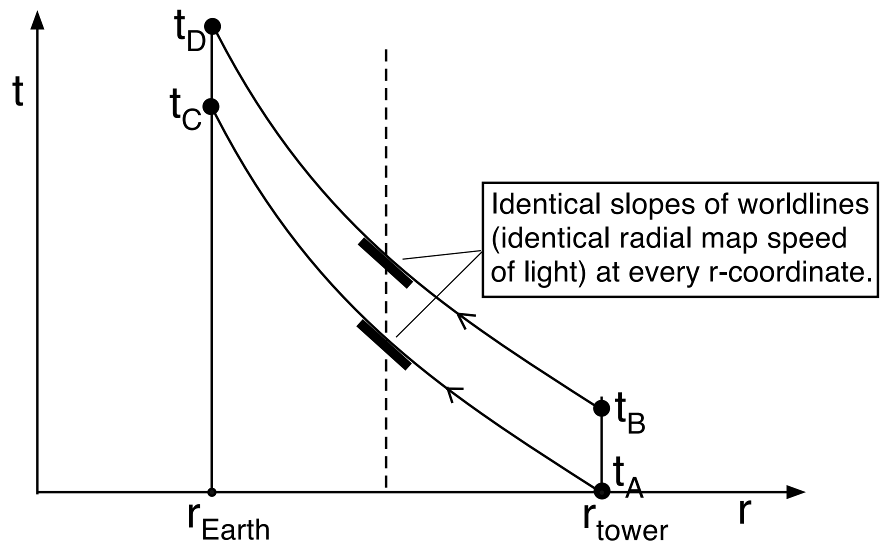


FIGURE 2 Schematic plot in Schwarzschild global coordinates (t, r) of worldlines of two sequential flashes moving downward from the top to the bottom of a tower. According to equation (1) the r-coordinate (map) speed depends only on the r-coordinate. As a result, the map t-coordinate difference between receptions of the flashes is identical to the map t-coordinate difference between emissions. However, the *wristwatch* time between the reception events C and D, measured by the bottom observer, is different from the *wristwatch* time between emission events A and B, measured by the top observer, equation (5). The figure greatly exaggerates the variation of radial map light speed with r-coordinate.

$$\frac{d\tau_{\mathrm{H}}}{d\tau_{\mathrm{L}}} = \left(\frac{1 - \dfrac{2M}{r_{\mathrm{H}}}}{1 - \dfrac{2M}{r_{\mathrm{L}}}}\right)^{1/2} \qquad \text{(stationary clocks)} \qquad (5)$$

Wristwatch time between flashes is different at different r.

The lapse dt in Schwarzschild global map t-coordinate between flashes is the same at the locations of upper and lower clocks, but the *wristwatch* time is different as recorded on these different clocks. Indeed, $r_{\mathrm{H}} > r_{\mathrm{L}}$, so equation (5) tells us that $d\tau_{\mathrm{H}} > d\tau_{\mathrm{L}}$; the lapse of wristwatch time on the higher clock is greater than the lapse of wristwatch time on the lower clock.

Gravitational red and blue shifts

The wristwatch time lapse $d\tau_{\mathrm{H}}$ on the higher clock can be extended to the measured period T_{H} of a sinusoidal signal emitted from the top of the tower. The measured period T_{L} of the signal as it reaches Earth's surface is therefore observed to be smaller. Frequency is inversely proportional to period, so the observed frequency of the signal increases as it descends. This is called the **gravitational blue shift**, and gives the lower observer the impression that clocks above him "run fast" compared with his. In contrast, for a signal rising from Earth's surface to be observed at the top of the tower, the period increases and the measured frequency decreases, an effect labeled **gravitational red shift**, which gives the higher observer the impression that clocks below him "run slow." (Impression or reality? See Box 2)

4.3 ■ APPROXIMATIONS

How small is small?

GR effects: small but crucial to GPS.

The general relativistic effects we study are small. How small? Small compared to what? When *can* we use approximations to general relativistic expressions? And when we do, which approximations are good enough? These questions are so central to the analysis of the GPS that it is useful to begin with a rough estimate of the expected effects, not worrying initially about the crudeness of this approximation.

Assume that our tower—standing on a non-rotating Earth—extends to the height of the GPS satellite and that the satellite rests without moving on the top of the tower ($v = 0$). First write (5) in the form

$$\frac{d\tau_{\rm H}}{d\tau_{\rm L}} = \left(1 - \frac{2M}{r_{\rm H}}\right)^{1/2}\left(1 - \frac{2M}{r_{\rm L}}\right)^{-1/2} \qquad \text{(stationary clocks)} \qquad (6)$$

In Query 7 you show Newton's result that, for a 12-hour circular orbit, the orbital radius (from Earth's center) is about 26.6×10^6 meters. Inside the front cover are values for the radius and mass of Earth. We now make use of an approximation also written inside the front cover:

$$(1+\epsilon)^n \approx 1 + n\epsilon + O(\epsilon^2) \quad \text{provided} \quad |\epsilon| \ll 1 \quad \text{and} \quad |n\epsilon| \ll 1 \qquad (7)$$

Our approximations are "to first order," that is, we neglect the correction term $O(\epsilon^2)$, which means "terms of second (and higher) order in ϵ."

QUERY 1. Clock rate difference due to difference in altitude.
Apply approximation (7) to the two parenthetical expressions on the right side of equation (6). Multiply out the result to show that, to first order:

$$\frac{d\tau_{\rm H}}{d\tau_{\rm L}} \approx 1 - \frac{M}{r_{\rm H}} + \frac{M}{r_{\rm L}} \qquad \text{(to first order for } v = 0 \text{ and nonrotating Earth)} \qquad (8)$$

Verify that values of both $M/r_{\rm Earth}$ and $M/r_{\rm satellite}$ satisfy the criteria for approximation (7) that leads to the result (8).

QUERY 2. Numerical approximation, stationary clocks.

In the following equation, b stands for the sum of two terms added to the number one on the right side of equation (8). Substitute numbers into equation (8) and find the numerical value of b:

$$\frac{d\tau_{\rm H}}{d\tau_{\rm L}} \approx 1 + b \qquad (v = 0 \text{ and nonrotating Earth}) \tag{9}$$

Small fractional differences in clock rates affect GPS operation.

The numerical value of b in equation (9) gives us an estimate of the fractional difference in rates of signals between stationary clocks at the position of the satellite and at Earth's surface. Is this fractional difference negligible or important to the operation of the GPS? Suppose the timing of a satellite clock is off by one nanosecond (10^{-9} second). In one nanosecond a light signal (or microwave pulse) propagates approximately 30 centimeters, approximately one English foot. So a difference of, say, hundreds of nanoseconds will render GPS results inaccurate if we need a location precision of ten meters or so.

QUERY 3. Synchronization discrepancy after one day.
As long as Earth and satellite clocks do not move and the Earth does not rotate, the wristwatch time increments in equation (9) can be as long as we want, leading to the equation

$$\tau_{\rm H} \approx (1 + b)\,\tau_{\rm L} \qquad (v = 0 \text{ and nonrotating Earth}) \tag{10}$$

There are approximately 86 400 seconds in one day. (The fractional difference in rates is so small that it does not matter which local clock records this time.) To an accuracy of one significant digit, the satellite clock and Earth clock go out of synchronism by about 50 000 nanoseconds per day due to their difference in altitude alone. Find the correct value to three-digit accuracy.

The Earth observer thinks that the satellite clock above him "runs fast" by something like 50 000 nanoseconds per day compared with his local clock, due to position effects alone. Clearly we must use general relativity to analyze the operation of the Global Positioning System, even though the *fractional* difference between clock rates at the two locations (at least the part due to difference in r-coordinate) is small.

Speed effects opposite to altitude effects.

In addition to the effect of altitude, we must include the effect due to relative motion between satellite and Earth observer. Which way will this second effect influence the discrepancy in clock rates due to altitude introduced by general relativity: to increase it or decrease it? The satellite clock now moves with respect to a string of Earth clocks. Special relativity tells us (in an imprecise summary): "Speeding clocks run slow." Therefore we expect the effect of motion to *reduce* the amount by which the satellite clock runs fast compared to the Earth clock. In brief, when speed effects are taken into account, we expect the satellite clock to run faster than the Earth clock by *less* than the estimated 50 000 nanoseconds per day. In Query 8, you check your final result against this prediction.

4.4 ■ MOVING CLOCKS

Relative speed changes relative clock rates

Earth clock and satellite clock both move in circles.

Now we take account of the effects of relative motion on the relative rates of Earth and satellite clocks. Think of the Earth clock as fixed at the equator, so it moves in a circle as the Earth rotates. The satellite clock also circles the Earth, but in its own independent circular orbit. In each case $d\tau$ is the wristwatch time between ticks, the time recorded by a given clock. Set $dr = 0$ in the Schwarzschild metric and divide through by dt^2 to obtain, for either clock in its orbit,

$$\left(\frac{d\tau}{dt}\right)^2 = \left(1 - \frac{2M}{r}\right) - r^2\left(\frac{d\phi}{dt}\right)^2 = \left(1 - \frac{2M}{r}\right) - v^2 \qquad (dr = 0) \qquad (11)$$

Here $d\tau$ is the wristwatch time between ticks of either clock and $v = rd\phi/dt$ is the instantaneous tangential speed of that clock in global coordinates.

QUERY 4. Clock rate correction formula.
First apply equation (11) to the satellite clock, then apply (11) to the Earth clock. Divide the two sides of the satellite equation by the corresponding sides of the Earth equation. Take the square root of both sides of the result. For both numerator and denominator in the resulting equation, use the approximation (7). In the numerator, set

$$\epsilon_{\rm H} = -\frac{2M}{r_{\rm H}} - v_{\rm H}^2 \qquad \text{(subscript H means satellite)} \qquad (12)$$

Now do the same for the denominator. In the denominator the formula for $\epsilon_{\rm L}$ is the same as that for $\epsilon_{\rm H}$, but with L for "lower" as subscripts. Carry out an analysis similar to that in Query 1 to retain only the dominant terms. Show that the result is

$$\boxed{\frac{d\tau_{\rm H}}{d\tau_{\rm L}} \approx 1 - \frac{M}{r_{\rm H}} - \frac{v_{\rm H}^2}{2} + \frac{M}{r_{\rm L}} + \frac{v_{\rm L}^2}{2} \qquad \text{(satellite directly overhead)} \qquad (13)}$$

Newton orbits good enough for GPS analysis.

Now we need numerical values for the quantities on the right side of (13). Chapter 8 derives the map speed of a satellite in circular orbit according to general relativity. After completing that chapter you can verify that the following Newtonian derivation of orbit radius and satellite speed in that orbit are sufficiently accurate for our analysis of the Global Positioning System.

QUERY 5. Speed of a clock on the equator
Earth's center is in free fall as Earth orbits the Sun. The Earth also rotates on its axis, completing one full rotation with respect to the distant stars in what is called a **sidereal day**, which is 86 164.1 seconds long. (Even when we require 6-digit accuracy, which of our local clocks measures this time does not matter.) With respect to Earth's center, what is the speed v of a clock at rest on Earth's surface at the equator? Use Newtonian "universal time" t. Express your answer as a fraction of the speed of light.

Box 2. OPINION: Slogans and Observations

"Moving clocks run slow." Special relativity gives us this useful slogan, a slogan that follows us into general relativity, which adds a second useful slogan, "Clocks higher in a gravitational field run faster." What do these slogans mean?

First of all, we need to specify "faster or slower" with respect to what? More precisely, special relativity says, "An observer measures a clock moving past him to run slower than a set of synchronized clocks in his frame."' General relativity adds, "An observer at lower altitude in a gravitational field may interpret signals he receives from a clock above him to mean that the higher clock runs faster than his own clock." The GPS verifies these slogans and demonstrates their usefulness.

But there is a deeper issue: When you ride in a spaceship speeding past your friend, your wristwatch does not run slow *for you*; it ticks at its accustomed pace compared, for example, with your pulse—or your aging! Similarly, when you mount a ladder to climb vertically away from your friend in a gravitational field, you notice no change in your wristwatch rate; your wristwatch does not speed up *for you* as you gain altitude.

So does a clock *really* slow down as it moves faster? Does a clock *really* speed up as it rises in a gravitational field? Welcome to relativity: *Observations depend on the observer!*

Physics does not tell us about *reality*—whatever that means. Physics formulates theories to describe our observations and to predict new ones.

What is the value of the speed v of the satellite? Newton tells us that in a circular orbit the center-directed acceleration has the magnitude v^2/r, where v is measured in conventional units, such as meters per second. Newton also tells us that the satellite mass m multiplied by this acceleration must equal Newton's gravitational force exerted by Earth:

$$F = ma = \frac{mv^2}{r} = m\frac{GM}{r^2} \qquad \text{(Newton, conventional units)} \qquad (14)$$

Newtonian orbit analysis

Equation (14) provides one relation between the speed of the satellite and the radius of its circular orbit. (This is Newtonian mechanics, where "radius" can, in principle, be directly measured.) A second relation connects satellite speed and orbit radius to the period of revolution. This period T is 12 hours for GPS satellites:

$$v = \frac{2\pi r}{T} \qquad \text{(Newton, conventional units)} \qquad (15)$$

QUERY 6. Units in meters

Convert equations (14) and (15) to units of meters, Earth mass M and satellite orbital period T to meters, and satellite speed v to the unitless fraction of light speed. Then eliminate r between these two equations to find an expression for v in terms of M and T and numerical constants.

4.5 ■ THE FINAL RECKONING

Effects of altitude AND relative speed on clock rates

Comment 1. Relative Motion Leads to Doppler Shift
Is the GPS satellite approaching the ground receiver or receding from it? If the satellite approaches, the receiver detects a Doppler increase in frequency of the clock-tick signals from the satellite. In contrast, when the satellite recedes from the Earth receiver it detects a Doppler decrease in frequency of the clock-tick signals. In this chapter we carry out calculations for satellite emissions when it is positioned directly above the Earth receiver. In this case the *change* in the detected clock-tick signal frequency as it passes overhead is due to the relative tangential motion between the satellite and ground clocks. For other relative motions of satellite and receiver, the computer in the receiver calculates the anticipated Doppler shift and adjusts the local receiver time lapses between incoming tick signals accordingly.

QUERY 7. Satellite orbital radius and speed, according to Newton.
Find the numerical value of the speed v (as a fraction of the speed of light) for a satellite in a 12-hour circular orbit. Find the numerical value of the radius r for this orbit—according to Newton and Euclid.

Now we have numerical values for all the terms in equation (13) and can estimate the difference in rate between satellite and Earth clocks.

QUERY 8. Clock rate correction, numerical
Substitute values for the various quantities in (13). Show that the satellite clock appears to run faster than the Earth clock by approximately 38 700 nanoseconds per day.

Section 4.3 described the difference in clock rates due only to difference in altitude. We predicted at the end of Section 4.3 that the full general relativistic treatment would lead to a *smaller* difference in clock rates than reckoned for the altitude effect alone. Your result in Query 8 verifies this prediction.

Before-launch settings of satellite and Earth clocks

Before launch, GPS satellite clocks are set to run at a rate of 38 700 nanoseconds per day *slower* than identical Earth clocks next to them, clocks that will remain on Earth's surface. *Result:* When the satellite clock passes overhead, the increased frequency (gravitational blue shift) of its signal received on Earth synchronizes with the Earth clock (Box 3).

An historical aside: Carroll O. Alley, a consultant to the original GPS project, had a hard time convincing the designers *not* to apply *twice* the correction given in (13): first to account for the different rate of time advance on wristwatches located at different altitudes and second to allow for the gravitational blue shift in frequency for the signal sent downward from satellite to Earth. There is only one correction; moreover there is no way to identify uniquely the "cause" of this correction. Listen to what Clifford Will says about

Gravitational shifts: no single identifiable cause

Box 3. General Relativity On/Off Switch

Launching the Global Positioning System was an immense military and civilian effort. Most participants were not skilled in general relativity and, indeed, wondered if the academic advisors were right about this strange theory. As one later publication put it:

> *There was considerable uncertainty among Air Force and contractor personnel designing and building the system whether these effects were being correctly handled, and even, on the part of some, whether the effects were real.*

The GPS prototype satellite called Navigation Technological Satellite 2 (NTS-2) was launched into a near-12-hour circular orbit on June 23, 1977, with its single atomic clock initially set (on Earth) to run at the same rate as Earth clocks. However, it had a general relativity on/off switch, leading to two possible modes of operation. In the first mode, with the switch set to ”off”, the satellite clock was simply left to run at the rate at which it had been set on Earth. It ran in this condition for 20 days. The satellite clock drifted, relative to Earth clocks, at the rate predicted by general relativity, “well within the accuracy capabilities of the orbiting clock.”

The NTS-2 satellite validated the general relativity results, so the general relativity on/off switch was flipped to “on.” This changed the satellite clock rate to a pre-arranged 38 700 nanoseconds per day slower than that of the Earth clock, also set before launch when the two clocks were side by side on Earth. Then the gravitational blue shift of the signal from an orbiting overhead satellite raised the frequency of the signal received on Earth to that of the Earth clocks. Since then, every GPS satellite goes into orbit with general relativity built into its design and construction. No more general relativity on/off switch!

the difference in rates between one clock emitting a signal from the top of a tower and a second identical clock receiving the signal on the ground:

> *A question that is often asked is, Do the intrinsic rates of the emitter and receiver or of the clock change, or is it the light signal that changes frequency during its flight? The answer is that it doesn't matter. Both descriptions are physically equivalent. Put differently, there is no operational way to distinguish between the two descriptions. Suppose that we tried to check whether the emitter and the receiver agreed in their rates by bringing the emitter down from the tower and setting it beside the receiver. We would find that indeed they agree. Similarly, if we were to transport the receiver to the top of the tower and set it beside the emitter, we would find that they also agree. But to get a gravitational red shift, we must separate the clocks in height; therefore, we must connect them by a signal that traverses the distance between them. But this makes it impossible to determine unambiguously whether the shift is due to the clocks or to the signal. The observable phenomenon is unambiguous: the received signal is blue shifted. To ask for more is to ask questions without observational meaning. This is a key aspect of relativity, indeed of much of modern physics: we focus only on observable, operationally defined quantities, and avoid unanswerable questions.*
>
> —Clifford Will

TWO COMMENTS

Comment 2. Newtonian orbit radius OK.
We assume in this chapter that the radius $r_{\rm H}$ of the circular orbit of the satellite and the speed v of the satellite in that orbit are both computed accurately enough using Newtonian mechanics. Exercise 2 in Section 8.7 validates this assumption.

Comment 3. Little latitude effect.
Our analysis considers an Earth clock fixed to the ground at the equator. One might expect that the speed-dependent correction would take on different values for an Earth clock fixed to the ground at different latitudes north or south of the equator, going to zero at the poles where there is no relative motion of the Earth clock due to rotation of Earth. In practice there is negligible latitude effect because Earth is not perfectly spherical; it bulges a bit at the equator due to its rotation, like a squashed balloon. The smaller r at the poles increases the $M/r_{\rm L}$ term in (13) by roughly the same amount that the speed term decreases. The outcome is that our calculation for the equator applies quite well to all latitudes.

QUERY 9. Orbit radius for zero time correction.
At a cocktail party one hears, “A speeding clock runs slow.” and “A higher clock runs faster.” This implies that there should be a radius at which the two effects cancel, so that two flashes received from a clock passing overhead would have the same time lapse between them as measured by an Earth observer directly below.

A. *Guess:* Do you expect the radius at which the two effects cancel to be smaller or larger than the actual orbital radius of GPS satellite orbits?
B. Use Newton to calculate the radius of the “cancellation orbit”?
C. A permanent circular orbit around Earth must be above the atmosphere. Is this true of the “cancellation orbit” you calculated in Item B?

4.6 ■ APPLICATIONS OF THE GLOBAL POSITIONING SYSTEM

GPS applications everywhere!

Uses of GPS? Look around!

Applications of the Global Positioning System have exploded. To ask how the GPS is used today is like asking about applications of the automobile or the telephone. Geologists measure the millimeters-per-year motion of the continents (motion with respect to what?); biologists track wildlife (Box 4). How is the GPS used? Look around and read the news!

QUERY 10. Aging on the International Space Station.
The International Space Station (ISS) circles the Earth at an altitude of approximately 350 kilometers, or at a radius of 6.73×10^6 meters from Earth’s center, at an orbital speed of 7 707 meters per second. An astronaut lives on the ISS for one year. When she returns to Earth’s surface, how much younger (or older?) is she than her twin sister who stayed on Earth?

Box 4. Tracking the Pack

Wolf 832F ventured out of her territory in Yellowstone's Lamar Valley. As soon as she left the park, she lost its protections, and the wolf, a 6-year-old alpha female, was shot and killed by a hunter. She had been wearing an expensive GPS tracking collar, which allowed scientists to follow her every move and gain crucial insight into the lives of gray wolves. Is this particular predator a pack leader or a lone wolf? A dedicated hunter or a mooch? How much time does it spend with its pups? Who are its associates, rivals and mates?

By using satellite and cellular tags to track free-ranging animals, biologists are providing us with intimate access to the daily lives of other species, drawing us closer to the world's wild things and making us more invested in their welfare.

Today's tags are capable of collecting months' or years' worth of data on an animal's location at a given moment, and can be used to track everything from tiny tropical orchid bees to blubbery, deep-diving elephant seals.

The Tagging of Pacific Predators project created a Web site broadcasting the movements of their subjects in real time (or close to it). While the project lasted, anyone with an Internet connection could follow the wanderings of Monty, the mako shark, Genevieve, the leatherback turtle, or Jon Sealwart and Stelephant Colbert, both northern elephant seals.

Bird lovers can follow the migrations of bald eagles through EagleTrak, run by the Center for Conservation Biology. The group provides detailed updates on the journeys of two eagles, Camellia and Azalea. Each bird has around a hundred "adoptive parents," proving how attached we can get to a wild creature when we have a name and a life story to assign to it.

We've only just scratched the surface of what's possible.

—Emily Anthes

4.7 ■ REFERENCES

Initial quote from "A Tale of Einstein, Congress and Safe Landings" by Clifford Will, an online posting by the Editorial Services, Washington University, St. Louis.

Box 3 quote about general relativity on/off switch: Carroll O. Alley, "Proper Time Experiments in Gravitational Fields With Atomic Clocks, Aircraft, and Laser Light Pulses," in *Quantum Optics, Experimental Gravity, and Measurement Theory,* edited by Pierre Meystre and Marlan O. Scully, Plenum Publishing, New York, 1983, pages 421-424.

Box 4 Tracking the Pack by Emily Anthes, New York Times, February 4, 2013. Excerpted and rearranged.

Quote about clocks on the ground and on a tower: Clifford M. Will, *Was Einstein Right? Putting General Relativity to the Test,* Second Edition, Basic Books/Perseus Group, New York, 1993, pages 49-50.

Download file name: Ch04GlobalPositioningSystem160401.pdf

Chapter 5. Global and Local Metrics

- *Why did Einstein take seven years to go from special relativity to general relativity?*
- *Why are so many different kinds of flat maps used to plot Earth's curved surface?*
- *Why use coordinates at all? Why not just measure distances directly, say with a ruler?*
- *Why does the spacetime metric use differentials?*
- *Are Schwarzschild global coordinates the only way to describe spacetime around a black hole?*

CHAPTER

5 Global and Local Metrics

Edmund Bertschinger & Edwin F. Taylor *

> *The basic demand of the special theory of relativity (invariance of the laws under Lorentz-transformations) is too narrow, i.e., that an invariance of the laws must be postulated relative to* nonlinear *transformations for the co-ordinates in the four-dimensional continuum.*
>
> *This happened in 1908. Why were another seven years required for the construction of the general theory of relativity?* ***The main reason lies in the fact that it is not so easy to free oneself from the idea that coordinates must have an immediate metrical meaning.***
>
> —Albert Einstein [boldface added]

5.1 ■ EINSTEIN'S PERPLEXITY

Why seven years between special relativity and general relativity?

Einstein's seven-year puzzle

It took Albert Einstein seven years to solve the puzzle compressed into the two-paragraph quotation above. The first paragraph complains that special relativity (with its restriction to flat spacetime coordinates) is too narrow. Einstein demands that a *nonlinear* coordinate system—that is, one that is *arbitrarily stretched*—should also be legal. *Nonlinear* means that it can be stretched by different amounts in different locations.

Stretch coordinates arbitrarily.

In the second paragraph, Einstein explains his seven-year problem: He tried to apply to a stretched coordinate system the same rules used in special relativity. Einstein's phrase **immediate metrical meaning** describes something that can be measured directly—for example, the radar-measured distance between the top of the Eiffel Tower and the Paris Opera building. Einstein says that since we can use nonlinear stretched coordinates, these coordinate separations need not be something we can measure directly, for example with a ruler.

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

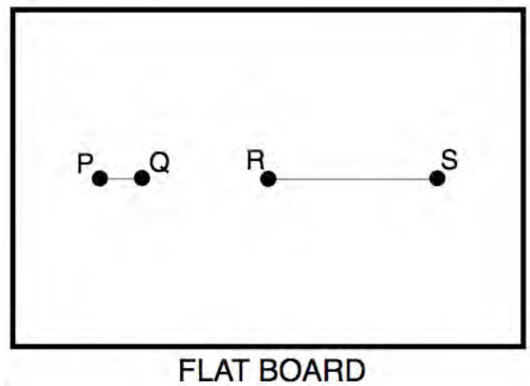


FIGURE 1 Compare distances between two different pairs of points on a flat wooden cutting board. First measure with a ruler the distance between the pair of points P and Q. Then measure the distance between the pair of points R and S. Measured distance PQ is *smaller* than the measured distance RS. We require no coordinate system whatsoever to verify this inequality; we measure distances directly on a flat surface.

Solving Einstein's puzzle leads to the global metric.

What is the relation between the coordinate separations between two points and the directly-measured distance between those two points? How does this distinction affect predictions of special and general relativity? Answering these questions reveals the unmeasurable nature of global coordinate separations, but nevertheless the central role of the *global metric* in connecting different local inertial frames in which we carry out measurements.

5.2 ■ EINSTEIN'S PERPLEXITY ON A WOODEN CUTTING BOARD

Move beyond high school geometry and trigonometry!

Simplify: From curved spacetime to a flat cutting board.

We transfer Einstein's puzzle from spacetime to space and—to simplify further—measure the distance between two points on the flat surface of a wooden cutting board (Figure 1).

Measure distance directly, with a ruler.

A pair of points, P and Q, lie near to one another on the surface. A second pair of points, R and S, are farther apart than points P and Q. How do we know that distance RS is greater than distance PQ? We measure the two distances directly, with a ruler. To ensure accuracy, we borrow a ruler from the local branch of the National Institute of Standards and Technology. Sure enough, with our official centimeter-scale ruler we verify distance RS to be greater than distance PQ. *We do not need any coordinate system whatsoever to measure distance PQ or distance RS or to compare these distances on a flat surface.*

Difference in Cartesian coordinates verifies difference in distances.

Next, apply coordinates to the flat surface. Do not draw coordinate lines directly on the cutting board; instead spread a fishnet over it (Figure 2). When we first lay down the fishnet, its narrow strings look like Cartesian square coordinate lines. Adjacent strings are one centimeter apart. The x-coordinate separation between P and Q is 1 centimeter, and the x-coordinate separation

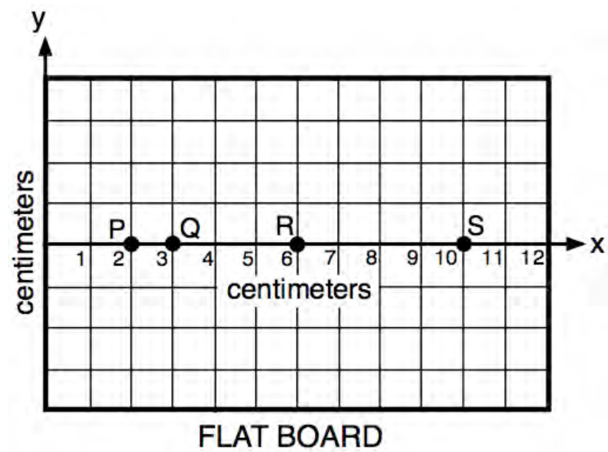


FIGURE 2 A fishnet with one-centimeter separations covers the wooden cutting board. Expressed in these coordinates, the coordinate separation PQ is 1 centimeter, while the coordinate separation RS is 4 centimeters. In this case a coordinate separation *does* have "an immediate metrical meaning" in Einstein's phrase. *Interpretation:* In this case we can derive from coordinate separations the values of directly-measured distances.

between R and S is 4 centimeters, confirming the inequality in our direct distance measurements. In this case each difference (or separation) in Cartesian coordinates, PQ and RS, *does* have "an immediate metrical meaning;" in other words, it corresponds to the *directly-measured distance.*

Stretch fishnet by variable amounts in x-direction.

Moving ahead, suppose that instead of string, we make the fishnet out of rubber bands. As we lay the rubber band fishnet loosely on the cutting board, we do something apparently screwy: As we tack down the fishnet, we stretch it along the x-direction by different amounts at different horizontal positions. Figure 3 shows the resulting "stretch" coordinates along the x-direction.

"Stretch" coordinate separation not equal to measured distance.

Now check the x-coordinate difference between P and Q in Figure 3, a difference that we call Δx_{PQ}. Then $\Delta x_{\mathrm{PQ}} = 5 - 2 = 3$. Compare this with the x-coordinate separation between R and S: $\Delta x_{\mathrm{RS}} = 10 - 9 = 1$. Lo and behold, the coordinate separation Δx_{PQ} is *greater* than the coordinate separation Δx_{RS}, even though our directly-measured distance PQ is *less* than the distance RS. This contradiction is the simplest example we can find of the great truth that Einstein grasped after seven years of struggle: *coordinate separations need not be directly measurable.*

Stretch coordinates form a legal map.

"No fair!" you shout. "You can't just move coordinate lines around arbitrarily like that." Oh yes we can. Who is to prevent us? Any coordinate system constitutes a **map**. What is a map? Applied to our cutting board, a map is simply a rule for assigning numbers that uniquely specify the location of every individual point on the surface. Our coordinate system in Figure 3 does that job nicely; it is a legal and legitimate map. However, the amount of stretching—what we call the **map scale**—varies along the x-direction.

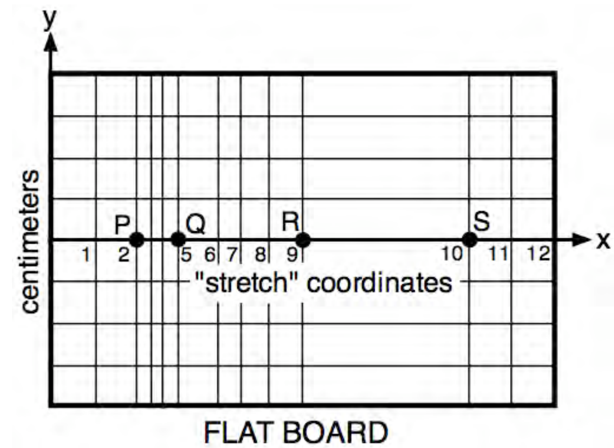


FIGURE 3 Global coordinate system that covers our entire cutting board, but in this case made with a rubber fishnet tacked down so as to stretch the x separation of fishnet cords by different amounts at different locations along the horizontal direction. The coordinate separation $\Delta x_{\text{PQ}} = 3$ between points P and Q is greater than the coordinate separation $\Delta x_{\text{RS}} = 1$ between points R and S, even though the measured *distances* between each of these pairs show the reverse inequality. Einstein was right: In this case coordinate separations do *not* have "an immediate metrical meaning;" in other words, coordinate separations do *not* tell us the values of directly-measured distances.

Of course, for convenience we usually *choose* the map scale to be everywhere uniform, as displayed in Figure 2. This choice is perfectly legal. We call this legality of Cartesian coordinates Assertion 1:

Assertion 1 for a FLAT SURFACE: CAN draw map with everywhere-uniform map scale.

Assertion 1. ON A FLAT SURFACE IN SPACE, we CAN FIND a global coordinate system such that every coordinate separation IS a directly-measured distance.

Standard Cartesian (x, y) coordinates allow us to use the power of the Pythagorean Theorem to predict the directly-measured distance s between two points anywhere on the board in Figure 2:

$$\Delta s^2 = \Delta x^2 + \Delta y^2 \qquad \text{(flat surface: \textit{Choose} Cartesian coordinates.)} \quad (1)$$

Cartesian separations: Pythagoras works!

The coordinate separations Δx and Δy and the resulting measured distance Δs can be as small or as large as we want, as long as the map scale is uniform everywhere on the flat cutting board.

In contrast, we *cannot* apply the Pythagorean Theorem using the "stretch" coordinates in Figure 3 to find the distance between a pair of points that are far apart in the x-direction. Why not? Because a large separation between two points can span regions where the map scale varies noticeably, that is, where rubber bands stretch by substantially different amounts. For example in Figure 3, the x-coordinate separation between points Q and S on

Stretch coordinates: Pythagoras fails on a flat surface.

the flat surface is $\Delta x_{\rm QS} = 5$, whereas points P and S have a much greater x-coordinate separation: $\Delta x_{\rm PS} = 8$. This is true even though the directly-measured *distance* between P and S is only slightly greater than the directly-measured *distance* between Q and S.

Stretched-fishnet coordinates of Figure 3, provide a case in which the Pythagorean Theorem (1) gives incorrect answers—coordinate separations are *not* the same as directly-measured distances. This yields Assertion 2, an alternative to Assertion 1:

Assertion 2 for a FLAT SURFACE: We are FREE to choose variable map scale over the surface.

Assertion 2. ON A FLAT SURFACE IN SPACE, we are FREE TO CHOOSE a global coordinate system for which coordinate separations ARE NOT directly-measured distances.

5.3 ■ GLOBAL SPACE METRIC FOR A FLAT SURFACE

Space metric to the rescue.

How can we predict measured distances using arbitrary coordinates? *Answer:* The metric!

Einstein tells us that we are free to stretch or contract conventional (in this case Cartesian) coordinates in any way we want. But if we do, then the resulting coordinate separations lose their "immediate metrical meaning;" that is, a coordinate separation between a pair of points no longer predicts the distance we measure between these points. If the coordinate separation can no longer tell us the distance between two points, what can? Our simple question about space on a flat cutting board is a preview of the far more profound question about spacetime with which Einstein struggled: How can we predict the measured wristwatch time τ or the measured ruler distance σ between a pair of events using the differences in *arbitrary* global coordinates between them? The answer was a breakthrough: "The metric!" Here's the path to that answer, starting with our little cutting board.

Space metric gives differential ds from differentials dx and dy.

Begin by recognizing that very close to any point on the flat surface the coordinate scale is nearly uniform, with a multiplying factor (local map scale) to correct for the local stretching in the x-coordinate. Strictly speaking, the coordinate scale is uniform only vanishingly close to a given point. *Vanishingly close?* That phrase instructs us to use the vanishingly small calculus limit: differential coordinate separations. For the coordinates of Figure 3, we find the differential distance ds from a **global space metric** of the form:

$$ds^2 = F(x_{\rm stretch})dx^2_{\rm stretch} + dy^2_{\rm stretch} \qquad \text{(variable } x\text{-stretch)} \qquad (2)$$

To repeat, we use the word *global* to emphasize that x is a valid coordinate everywhere across our cutting board covered by the stretched fishnet. In (2), $F(x)$—actually the square root of $F(x)$—is the map scale that corrects for the stretch in the horizontal coordinate *differentially close to that value of* x. If $F(x)$ is defined everywhere on the cutting board, however, then equation (2) is also valid at every point on the board.

Metric works well LOCALLY, even with stretched coordinates.

The global space metric is a tremendous achievement. On the right side of metric (2) the function $F(x)$ corrects the squared differential $dx^2_{\rm stretch}$ to give the correct squared differential distance ds^2 on the left side.

We have gained a solution to Einstein's puzzle for the simplified case of differential separations on a flat surface in space. But we seem to have suffered a great loss as well: calculus insists that the differential distance ds predicted by the space metric is vanishingly small. We cannot use our official centimeter-scale ruler to measure a vanishingly small differential distance. How can we possibly predict a measured distance—for example the distance between points P and S on our flat cutting board? We want to predict and then make *real* measurements on *real* flat surfaces!

Differential distance ds is too small to measure. . .

Differential calculus curses us with its stingy differential separations ds, but integral calculus rescues us. We can sum ("integrate") differential distances ds along the curve. The result is a predicted *total distance* along the curved path, a prediction that we can verify with a tape measure. As a special case, let's predict the distance s along the straight horizontal x-axis from point P to point S in Figure 3. Call this distance $s_{\rm PS}$. "Horizontal" means no vertical, so that $dy = 0$ in equation (2). The distance $s_{\rm PS}$ is then the sum (integral) of $ds = [F(x)]^{1/2}\,dx$ from $x = 2$ to $x = 10$, where the scale function $[F(x)]^{1/2}$ varies with the value of x:

. . . but we can predict measured distance from summed (integrated) ds.

$$s_{\rm PS} = \int_{x=2}^{x=10} [F(x_{\rm stretch})]^{1/2}\,dx_{\rm stretch} \qquad \text{(horizontal distance: P to S)} \qquad (3)$$

When we evaluate this integral, we can once again use our official centimeter-scale ruler to verify by direct measurement that the total distance $s_{\rm PS}$ between points P and S predicted by (3) is correct.

The example of metric (2) leads to our third important assertion:

Assertion 3 for a FLAT SURFACE: Metric gives us ds, whose integral predicts measured distance s.

<u>Assertion 3.</u> ON A <u>FLAT</u> SURFACE IN SPACE when using a global coordinate system for which coordinate separations ARE NOT directly-measured distances, a space metric is REQUIRED to give the differential distance ds whose integrated value predicts the measured distance s between points.

5.4 ■ GLOBAL SPACE METRIC FOR A CURVED SURFACE

Squash a spherical map of Earth's surface onto a flat table? Good luck!

In Sections 5.2 and 5.3, we chose variably-stretched coordinates on a flat surface. Then we corrected the effects of the variable stretching using a metric. This is a cute mathematical trick, but who cares? We are not *forced* to use stretched coordinates on a flat cutting board, so why bother with them at all? To answer these questions, apply our ideas about maps to the curved surface of Earth. Chapter 2 derived a global metric—equation (3), Section 2.3—for the spherical surface of Earth using angular coordinates λ for latitude and ϕ

for longitude, along with Earth's radius R. Here we convert that global metric to coordinates x and y:

$$\begin{aligned} ds^2 &= R^2 \cos^2 \lambda \, d\phi^2 + R^2 d\lambda^2 \qquad (0 \leq \phi < 2\pi \text{ and } -\pi/2 \leq \lambda \leq \pi/2) \qquad (4) \\ &= \cos^2 \left(\frac{R\lambda}{R}\right) (Rd\phi)^2 + (Rd\lambda)^2 \qquad (\text{metric: Earth's surface}) \\ &= \cos^2 \left(\frac{y}{R}\right) dx^2 + dy^2 \qquad (0 \leq x < 2\pi R \text{ and } -\pi R/2 \leq y \leq \pi R/2) \end{aligned}$$

On a sphere, we define $y \equiv R\lambda$ and $x \equiv R\phi$ (the latter from the definition of radian measure).

Undistorted flat maps of Earth impossible.

Compare the third line of (4) with equation (2). The y-dependent coefficient of dx^2 results from the fact that as you move north or south from the equator, lines of longitude converge toward a single point at each pole. That coefficient of dx^2 makes it impossible to cover Earth's spherical surface with a flat Cartesian map without stretching or compressing the map at some locations.

A curved surface *forces* us to use stretched coordinates.

Throughout history, mapmakers have struggled to create a variety of flat projections of Earth's spherical surface for one purpose or another. But each projection has *some* distortion. *No uniform projection of Earth's surface can be laid on a flat surface without stretching or compression in some locations.* If this is impossible for a spherical Earth with its single radius of curvature, it is certainly impossible for a general curved surface—such as a potato—with different radii of curvature in different locations. In brief, it is impossible to completely cover a curved surface with a single Cartesian coordinate system. (Is a cylindrical surface curved? No; technically it is a flat surface, like a rolled-up newspaper, which Cartesian coordinates can map exactly.) We bypass formal proof and state the conclusion:

Assertion 4 for a CURVED SURFACE: Everywhere-uniform map scale is IMPOSSIBLE.

Assertion 4. ON A CURVED SURFACE IN SPACE, it is IMPOSSIBLE to find a global coordinate system for which coordinate separations EVERYWHERE on the surface are directly-measured distances.

The dy on the third line of equation (4) is still a directly-measured distance: the differential distance northward from the equator. That is true for a sphere, whose constant R-value allows us to define $y \equiv R\lambda$. But Earth is not a perfect sphere; rotation on its axis results in a slightly-bulging equator. Technically the Earth is an **oblate spheroid**, like a squashed balloon. In that case neither x or y coordinate separations are directly-measured distances. And most curved surfaces are more complex than the squashed balloon. Einstein was right: In most cases coordinate separations *cannot* be directly-measurable distances.

Metric required on curved surface.

No possible uniform map scale over the entire surface of Earth? Then there is an inevitable distinction between a coordinate separation and measured distance. The space metric is no longer just an option, but has become the indispensable practical tool for predicting distances between two points from their coordinate separations.

Assertion 5 for a CURVED SURFACE: Metric REQUIRED to calculate distance.

Assertion 5. ON A CURVED SURFACE IN SPACE, a global space metric is REQUIRED to calculate the differential distance ds between a pair of adjacent points from their differential coordinate separations.

As before, integrating the differential ds yields a measured total distance s along a path on the curved surface, whose predicted length we can verify directly with a tape measure.

Space summary

SPACE SUMMARY: *On a flat surface in space we can choose Cartesian coordinates, so that the Pythagorean theorem—with no differentials—correctly predicts the distance s between two points far from one another. On a curved surface we cannot. But on any curved surface we can use a space metric to calculate ds between a pair of adjacent points from values of the differential coordinate separations between them. Then we can integrate these differentials ds along a given path in space to predict the directly-measured length s along that path.*

"Connectedness" = **topology.**

The combination of global coordinates plus the global metric is even more powerful than our summary implies. Taken together, the two describe a curved surface completely. In principle we can use the global coordinates plus the metric to reconstruct the curved surface exactly. (Strictly speaking, the global coordinate system must include information about ranges of its coordinates, ranges that describe its "connectedness"—technical name: its **topology**.)

5.5 ■ GLOBAL SPACETIME METRIC

Visit a neutron star with wristwatch, tape measure—and metric—in your back pocket.

To distorted space add warped t. Result? Trouble for Einstein!

What does all this curved-surface-in-space talk have to do with Einstein's perplexity during his journey from special relativity to general relativity? As usual, we express the answer as an analogy between a curved surface in space and a curved region of spacetime. Spacetime around a black hole multiplies the complications of the curved surface: not only is space distorted compared with its Euclidean description but the fourth dimension, the t-coordinate, is warped as well. All this complicates our new task, which is to predict our measurement of ruler distance σ or wristwatch time τ between a *pair of events in spacetime.*

Here we simply state, for flat and curved regions of spacetime, five assertions similar to those stated earlier for flat and curved surfaces in space.

Assertion A for FLAT SPACETIME: Everywhere-uniform map scale possible.

Assertion A. IN A FLAT REGION OF SPACETIME, we CAN FIND a global coordinate system in which every coordinate separation IS a directly-measured quantity.

In Chapter 1 we introduced a pair of expressions for flat spacetime called the *interval*, similar to the Pythagorean Theorem for a flat surface. One form of the interval predicts the wristwatch time τ between two events with a timelike relation. The second form tells us the ruler distance σ between two events with a spacelike relation:

$$\begin{aligned} \Delta\tau_{\text{lab}}^2 &= \Delta t_{\text{lab}}^2 - \Delta s_{\text{lab}}^2 \qquad \text{(flat spacetime, timelike-related events)} \\ \Delta\sigma^2 &= \Delta s_{\text{lab}}^2 - \Delta t_{\text{lab}}^2 \qquad \text{(flat spacetime, spacelike-related events)} \end{aligned} \tag{5}$$

In *flat* spacetime, each space coordinate separation Δs_{lab} and time coordinate separation Δt_{lab} measured in the laboratory frame can be as small or as great as we want. On to our second assertion:

Assertion B for FLAT SPACETIME: We are free to choose a variable map scale over the region.

Assertion B. IN A FLAT REGION OF SPACETIME we are FREE TO CHOOSE a global coordinate system in which coordinate separations ARE NOT directly-measured quantities.

In this case we can choose not only stretched space coordinates but also a system of scattered clocks that run at different rates. If we choose such a "stretched" (but perfectly legal) global spacetime coordinate system, the interval equations (5) are no longer valid, because any of these coordinate separations may span regions of varying spacetime map scales. So we again retreat to a differential version of this equation, adding coefficients similar to that of space metric (2). A simple timelike metric might have the general form:

$$d\tau^2 = J(t,y,x)dt^2 - K(t,y,x)dy^2 - L(t,y,x)dx^2 \tag{6}$$

Spacetime metric delivers $d\tau$ from differentials dt, dy, and dx.

Here each of the coefficient functions J, K, and L may vary with x, y, and t. (The coefficient functions are not entirely arbitrary: the condition of flatness imposes differential relations between them, which we do not state here.) Given such a metric for flat spacetime, we are free to use this metric to convert differentials of global coordinates (right side of the metric) to measured quantities (left side of the metric). This leads to our third assertion:

Assertion C for FLAT SPACETIME: Variable map scale requires metric to calculate $d\tau$ or $d\sigma$.

Assertion C. IN A FLAT REGION OF SPACETIME, when we choose a global coordinate system in which coordinate separations are not directly-measured quantities, then a global spacetime metric is REQUIRED to calculate the differential interval, $d\tau$ or $d\sigma$, between two adjacent events using their differential global coordinate separations.

On the other hand, in a region of curved spacetime—analogous to the situation on a curved surface in space—we *cannot* set up a global coordinate system with the same map scale everywhere in the region.

Assertion D for CURVED SPACETIME: Everywhere-uniform map scale is IMPOSSIBLE.

Assertion D. IN A CURVED REGION OF SPACETIME it is IMPOSSIBLE to find a global coordinate system in which coordinate separations EVERYWHERE in the region are directly-measured quantities.

Assertion E for CURVED SPACETIME: Metric REQUIRED to calculate $d\tau$ or $d\sigma$.

Assertion E. IN A CURVED REGION OF SPACETIME, a global spacetime metric is REQUIRED to calculate the differential interval, $d\tau$ or $d\sigma$, between a pair of adjacent events from their differential global coordinate separations.

Spacetime summary

> **SPACETIME SUMMARY:** *In flat spacetime we can choose coordinates such that the spacetime interval—with no differentials—correctly predicts the wristwatch time (or the ruler distance) between two events far from one another. In curved spacetime we cannot. But in curved spacetime we can use a spacetime metric to calculate $d\tau$ or $d\sigma$ between adjacent events from the values of the differential coordinate separations between them. Then we can integrate $d\tau$ along the worldline of a particle, for example, to predict the directly-measured time lapse τ on a wristwatch that moves along that worldline.*

"Connectedness" = **topology.**

As in the case of the curved surface, a complete description of a spacetime region results from the combination of global spacetime coordinates and global metric—along with the connectedness (topology) of that region. For example, we can in principle use Schwarzschild's global coordinates and his metric to answer all questions about spacetime around the black hole.

5.6 ■ ARE WE SMARTER THAN EINSTEIN?

Did Einstein fumble his seven-year puzzle?

Einstein's struggle

We have now solved the puzzle that troubled Einstein for the seven years it took him to move from special relativity to general relativity. Surely Einstein would understand in a few seconds the central idea behind cutting-board examples in Figures 1 through 3. However, the extension of this idea to the four dimensions of spacetime was not obvious while he was struggling to create a brand new theory of spacetime that is curved, for example, by the presence of Earth, Sun, neutron star, or black hole. Is it any wonder that during this intense creative process Einstein took a while to appreciate the lack of "immediate metrical meaning" of differences in global coordinates?

One co-author didn't get it.

It is embarrassing to admit that one co-author of this book (EFT) required more than two years to wake up to the basic idea behind the present chapter, even though this central result is well known to every practitioner of general relativity. Even now EFT continues to make Einstein's original mistake: He confuses global coordinate separations with measured quantities.

You too will probably find it difficult to avoid Einstein's mistake.

FIRST ADVICE FOR THE ENTIRE BOOK

> **FIRST STRONG ADVICE FOR THIS ENTIRE BOOK**
>
> **To be safe, it is best to assume that global coordinate separations do not have any measured meaning. Use global coordinates only with the metric in hand to convert a mapmaker's fantasy into a surveyor's reality.**

Global coordinate systems come and go; wristwatch ticks and ruler lengths are forever!

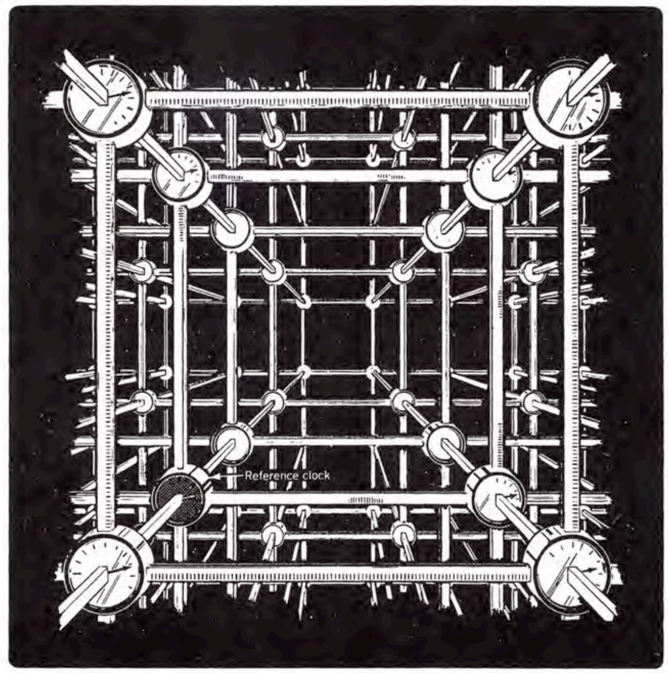


FIGURE 4 On a flat patch we build an inertial Cartesian latticework of meter sticks with synchronized clocks. This is an instrumented room (defined in Section 3.10), on which we impose a local coordinate system—a frame—limited in both space and time. Limited by what? Limited by the sensitivity to curvature of the measurement we want to carry out in that local inertial frame.

5.7 ■ LOCAL MEASUREMENT IN A ROOM USING A LOCAL FRAME

Where we make real measurements

> *Of all theories ever conceived by physicists, general relativity has the simplest, most elegant geometric foundation. Three axioms: (1) there is a global metric; (2) the global metric is governed by the Einstein field equations; (3) all special relativistic laws of physics are valid in every local inertial frame, with its (local) flat-spacetime metric.*
>
> —Misner, Thorne, and Wheeler (edited)

> *No phenomenon is a physical phenomenon until it is an observed phenomenon.*
>
> —John Archibald Wheeler

Special relativity assumes that a measurement can take place throughout an unlimited space and during an unlimited time. Spacetime curvature denies us this scope, but general relativity takes advantage of the fact that almost everywhere on a curved surface, space is locally flat; remember "flat Kansas" in Figure 3, Section 2.2. Wherever spacetime is smooth—namely close to every event except one on a singularity—general relativity permits us to approximate the gently curving stage of spacetime with a local inertial frame. This section sets up the command that we shout loudly everywhere in this book:

Spacetime is locally flat almost everywhere.

SECOND ADVICE FOR THE ENTIRE BOOK

<u>SECOND STRONG ADVICE FOR THIS ENTIRE BOOK</u>

In this book we choose to make every measurement in a local inertial frame, where special relativity rules.

We ride in a *room,* a physical enclosure of fixed spatial dimensions (defined in Section 3.10) in which we make our measurements, each measurement limited in local time. We assume that the room is sufficiently small—and the duration of our measurement sufficiently short—that these measurements can be analyzed using special relativity. This assumption is correct on a *patch.*

Definition: **patch**

DEFINITION 1. Patch

A **patch** is a spacetime region purposely limited in size and duration so that curvature (tidal acceleration) does not noticeably affect a given measurement.

Important: The definition of patch depends on the scope of the measurement we wish to make. Different measurements require patches of different extent in global coordinates. On this patch we lay out a local coordinate system, called a *frame.*

Definition: **frame**

DEFINITION 2. Frame

A **frame** is a local coordinate system of our choice installed onto a spacetime patch.This local coordinate system is limited to that single patch.

Among all possible local frames, we choose one that is inertial:

Definition: **inertial frame**

DEFINITION 3. Inertial frame

An **inertial** or **free-fall frame** is a local coordinate system—typically Cartesian spatial coordinates and readings on synchronized clocks (Figure 4)—for which special relativity is valid. In this book we report every measurement using a local inertial frame.

In general relativity every inertial frame is local, that is limited in spacetime extent. Spacetime curvature precludes a global inertial frame.

Who makes all these measurements? The observer does:

Definition: **observer**

DEFINITION 4. Observer = Inertial Observer

An **observer** is a person or machine that moves through spacetime making measurements, each measurement limited to a local inertial frame. Thus an observer moves through a series of local inertial frames.

Box 1. What moves?

A story—impossible to verify—recounts that at his trial by the Inquisition, after recanting his teaching that the Earth moves around the Sun, Galileo muttered under his breath, "Eppur si muove," which means "And yet it moves."

According to special and general relativity, what moves? We quickly eliminate coordinates, events, patches, frames, and spacetime itself:

- Coordinates do not move. Coordinates are number-labels that locate an event; it makes no sense to say that a coordinate number-label moves.
- An event does not move. An event is completely specified by coordinates; it makes no sense to say that an event moves.
- A flat patch does not move. A flat patch is a region of spacetime completely specified by a small, specific range of map coordinates; it makes no sense to say that a range of map coordinates moves.
- A local frame does not move. A frame is just a set of local coordinates—numbers—on a patch; it makes no sense to say that a set of local coordinates move.
- Spacetime does not move. *Spacetime* labels the arena in which events occur; it makes no sense to say that a label moves.

You cannot drop a frame. You cannot release a frame. You cannot accelerate a frame. It makes no sense to say that you can even move a frame. You cannot carry a frame around, any more than you can move a postal zip code region by carrying its number around.

What does move? Stones and light flashes move; observers and rooms move. Whatever moves follows a worldline or worldtube through spacetime.

- A stone moves. Even a stone at rest in a shell frame moves on a worldline that changes global t-coordinate.
- A light flash moves; it follows a *null worldline* along which both r and ϕ can change, but $\Delta\tau = 0$.
- An observer moves. Basically the observer is an instrumented stone that makes measurements as it passes through local frames.
- A room moves. Basically a room is a large, hollow stone.

Why do almost all teachers and special relativity texts—including our own physics text *Spacetime Physics* and Chapter 1 of this book!—talk about "laboratory frame" and "rocket frame"? Because it is a tradition; it leads to no major confusion in special relativity. But when we specify a local rain frame in curved spacetime using (for example) a small range of Schwarzschild global coordinates t, r, and ϕ, then it makes no sense to say that this local rain frame—this range of global coordinates—moves. Stones move; coordinates do not.

The observer, riding in a room (Definition 3, Section 3.10), makes a sequence of measurements as she passes through a series of local inertial frames. As it passes through spacetime, the room drills out a *worldtube* (Definition 4, Section 3.10). Figure 5 shows such a worldtube.

?

Objection 1. *In Definition 4 you say that the observer moves through a series of local inertial frames. But doesn't a shell observer stay in one local frame?*

!

No! The shell observer is *not* stationary in the global t-coordinate, but moves along a worldline (Figure 5). By definition, a local inertial frame spans a given lapse of frame time $\Delta t_{\rm shell}$, as well as a given frame volume of space. In Figure 5 the first measurement takes place in Frame #1. When the first measurement is over, global t/M has elapsed and the observer leaves Frame #1. A second measurement takes place in Frame #2. The range of r/M and ϕ global coordinates of Frame #2 may be the same as in Frame #1. The shell observer makes a series of measurements, each measurement in a *different* local inertial frame.

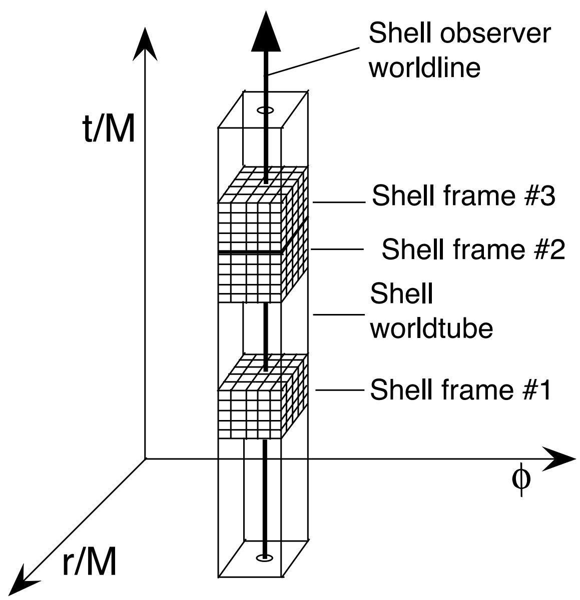


FIGURE 5 A shell worldtube (Section 3.10) that embraces three sample shell frames outside the event horizon. The shell observer carries out an experiment while passing through Frame #1 in the figure. He may then repeat the same experiment or carry out another one in Frames #2 and #3 at greater t coordinates. For simplicity each shell frame is shown as a cube. Each frame is *nailed* to a particular event at map coordinates $(\bar{t}/M, \bar{r}/M, \bar{\phi})$.

Comment 1. Euclid's curved space vs. Einstein's curved spacetime
Figure 5 shows a case in which a shell observer stands at constant r and ϕ coordinates while he passes, with changing map t-coordinate, through a series of local frames, each frame defined over a range of r, ϕ, and t-coordinates. Figure 5 in Section 2.2 showed the Euclidean space analogy in which a traveler passes across a series of local flat maps on her way along the curved surface of Earth from Amsterdam to Vladivostok. Each of these flat maps is essentially a set of numbers: local space coordinates we set up for our own use. Similarly, each local frame of Figure 5 is just a set of numbers, local space and time coordinates we set up for our own use. A frame is not a room; a frame does not fall; a frame does not move; it is just a set of numbers—coordinates—that we use to report results of local measurements (Box 1). Figure 5 shows multiple shell frames, two of them adjacent in t-coordinate. Shell frames can also overlap, analogous to the overlap of adjacent local Euclidean maps in Figure 5, Section 2.2.

Objection 2. *Whoa! Can a frame exist inside the event horizon?*

Definitely. A frame is a set of coordinates—numbers! Numbers are not things; they can exist anywhere, even inside the event horizon. In contrast, the diver in her unpowered spaceship is a "thing." Even inside the event

> horizon the she-thing continues to pass through a series of local frames. Inside the event horizon, however, she is doomed to continue to the singularity as her wristwatch ticks inevitably forward.

By definition, we use the flat-spacetime metric to analyze events in a local inertial frame. We write this metric for a local shell frame in a rather strange form which we then explain:

$$\Delta\tau^2 \approx \Delta t_{\text{shell}}^2 - \Delta y_{\text{shell}}^2 - \Delta x_{\text{shell}}^2 \tag{7}$$

Local flat spacetime $\rightarrow$ local inertial metric.

Choose the increment Δy_{shell} to be vertical (radially outward), and the Δx_{shell} increment to be horizontal (tangential along the shell).

Instead of an equal sign, equation (7) has an approximately equal sign. This is because near a black hole or elsewhere in our Universe there is always *some* spacetime curvature, so the equation cannot be exact. The upper case Delta, Δ, also has a different meaning in (7) than in special relativity. In special relativity (Section 1.10) we used Δ to emphasize that in flat spacetime the two events whose separation is described by (7) can be very far apart in space or time and their coordinate separations still satisfy (7) with an equals sign. In equation (7), however, both events must lie in the local frame within which the coordinate separations Δt_{shell}, Δy_{shell}, and Δx_{shell} are defined.

Connect global and local metrics

How do we connect local metric (7) to the Schwarzschild global metric? We do this by considering a local frame over which global coordinates t, r, and ϕ vary only a little. Small variation allows us to replace r with its average value $\bar{r}$ over the patch and write the Schwarzschild metric in the approximate form:

$$\Delta\tau^2 \approx \left(1 - \frac{2M}{\bar{r}}\right)\Delta t^2 - \frac{\Delta r^2}{\left(1 - \dfrac{2M}{\bar{r}}\right)} - \bar{r}^2\Delta\phi^2 \quad \text{(spacetime patch)} \tag{8}$$

Equation (8) is no longer global. The value of $\bar{r}$ depends on *where* this patch is located, leading to a local wristwatch time lapse $\Delta\tau$ for a given change Δr. The value of $\bar{r}$ also affects how much $\Delta\tau$ changes for a given change in Δt or $\Delta\phi$. Equation (8) is approximately correct only for limited ranges of Δt, Δr, and $\Delta\phi$. In contrast to the differential global Schwarzschild metric, (8) has become a *local* metric. That is the bad news; now for some good news.

Local shell coordinates

Coefficients in (8) are now constants. So simply equate corresponding terms in the equations (8) and (7):

$$\Delta t_{\text{shell}} \equiv \left(1 - \frac{2M}{\bar{r}}\right)^{1/2}\Delta t \tag{9}$$

$$\Delta y_{\text{shell}} \equiv \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2}\Delta r \tag{10}$$

$$\Delta x_{\text{shell}} \equiv \bar{r}\Delta\phi \tag{11}$$

FIGURE 6 Flat triangular segments on the surface of a Buckminster Fuller geodesic dome. A single flat segment is the geometric analog of a locally flat patch in curved spacetime around a black hole; we add local coordinates to this patch to create a local frame. (Figure 3 in Section 3.3 shows a complete geodesic dome with six-sided segments.)

Substitutions (9), (10), and (11) turn approximate metric (8) into approximate metric (7), which is—approximately!—the metric for flat spacetime. *Payoff:* We can use special relativity analyze local measurements and observations in a shell frame near a black hole.

?

Objection 3. *What is the meaning of equations (9) through (11)? What do they accomplish? How do I use them?*

!

These equations are fundamental to our application of general relativity to Nature. On the left are measured quantities: Δt_{shell} is the measured shell time between two events, Δy_{shell} and Δx_{shell} are their measured separations in local space shell coordinates. These equations, plus the local metric (7) unleash special relativity to analyze local measurements in curved spacetime. In this book we choose to report every measurement using a local inertial frame.

Comment 2. Left-handed $(\Delta y_{\text{shell}}, \Delta x_{\text{shell}})$ local space coordinates
We find it convenient to have the local Δy_{shell} point along the outward global Schwarzschild r-coordinate and the local Δx_{shell} point along the direction of increasing angle $\Delta\phi$, on the $[r, \phi]$ slice through the center of the black hole. This earns the label **left-handed** for the space part of these local coordinates, which differs from most physics usage.

Figure 6 shows a geometric analogy to a local flat patch: the local flat plane segments on the curved exterior surface of a Buckminster Fuller geodesic dome.

Summary: local notation

We summarize here the new notation introduced in equation (7) and equations (9) through (11):

$\approx$	equality is not exact, due to residual curvature and coordinate conversion (Section 5.8)	(12)
Δ	coordinate separation of two events within the local frame	(13)
$\bar{r}$	average r-coordinate across the patch	(14)

Objection 4. *How large—in Δt_{shell}, Δy_{shell}, and Δx_{shell}—am I allowed to make my local inertial frame? If you cannot tell me that, you have no business talking about local inertial frames at all!*

You are right, but the answer depends on the measurement you want to make. Some measurements are more sensitive than others to tidal accelerations; each measurement sets its own limit on the maximum extent of the local frame in order that it remain inertial for that measurement. If the local frame is too extended in both the Δx_{shell} and Δy_{shell} directions to be inertial, then it may be necessary to restrict the frame time Δt_{shell} during which it is carried out. *Result:* Different measurements prevent us from setting a universal, one-fits-all size for a local inertial frame. Sorry.

Objection 5. *What happens when we choose the size of the local frame too great, so the frame is no longer inertial? How do we know when we exceed this limit?*

There are two answers to these questions. The first is spacetime curvature: Section 1.11 entitled Limits on Local Inertial Frames describes this situation using Newtonian intuition. If two stones initially at rest near Earth are separated radially, the stone nearer the center accelerates downward at a faster rate. If two stones, initially at rest, are separated tangentially, their accelerations do not point in the same directions, Figure 8, Section 1.11. These effects go under the name *tidal accelerations*, because ocean tides on Earth result from differences in gravitational attraction of Moon and Sun at different locations on Earth. If these tidal accelerations exceed the achievable accuracy of an experiment, then the local frame cannot be considered inertial.

The second answer to the question results from the global coordinate system itself and the process by which the local inertial frame is derived from it. This part is treated in Section 5.8.

Box 2. Who cares about local inertial frames?

Sections 5.1 through 5.6 make no reference to local inertial frames. Nor are they necessary. The left side of the global metric predicts differentials $d\tau$ or $d\sigma$ (or $d\tau = d\sigma = 0$) between adjacent events. Of course we cannot measure differentials directly, because they are, by definition, vanishingly small. We need to integrate them; for example we integrate wristwatch time along the worldline of a stone. The resulting predictions are sufficient to analyze results of any experiment or observation. No local inertial frames are required, and most general relativity texts do not use them.

Our approach in this book is different; we *choose* to predict, carry out, and report all measurements with respect to a local inertial frame. *Payoff:* In each local inertial frame we can unleash all the concepts and tools of special relativity, such as directly-measured space and time coordinate separations, measurable energy and momentum of a stone; Lorentz transformations between local inertial frames.

We may report local-frame measurements in the calculus limit, as we often do on Earth. For example, we record the motion of a light flash in our local inertial frame. Rewrite (7) as

$$\Delta\tau^2 \approx \Delta t_{\text{shell}}^2 - \Delta s_{\text{shell}}^2 \tag{15}$$

where Δs_{shell} is the distance between two events measured in the shell frame. Now let a light flash travel directly between the two events in our local frame. For light $\Delta\tau = 0$ and we write its speed (a positive quantity) as:

$$\left|\frac{\Delta s_{\text{shell}}}{\Delta t_{\text{shell}}}\right| \approx 1 \qquad \text{(speed of light flash)} \tag{16}$$

Can take calculus limit in local frame.

We may want to know the instantaneous speed, which requires the calculus limit. In this process all increments shrink to differentials and $\bar{r} \to r$. For the light flash the result is:

$$v_{\text{shell}} \equiv \lim_{\Delta t_{\text{shell}} \to 0} \left|\frac{\Delta s_{\text{shell}}}{\Delta t_{\text{shell}}}\right| = 1 \qquad \text{(instantaneous light flash speed)} \tag{17}$$

Equation (17) reassures us that the speed of light is exactly one when measured in a local shell frame at any r (outside the event horizon, where shells can be constructed). The measured speed of a stone is always less than unity:

$$v_{\text{shell}} \equiv \lim_{\Delta t_{\text{shell}} \to 0} \left|\frac{\Delta s_{\text{shell}}}{\Delta t_{\text{shell}}}\right| < 1 \qquad \text{(instantaneous stone speed)} \tag{18}$$

5.8 ■ THE TROUBLE WITH COORDINATES

Coordinates, as well as spacetime curvature, limit accuracy.

Can use global metric exclusively.

We need global coordinates and cannot apply general relativity without them. Only global coordinates can connect widely separated local inertial frames in

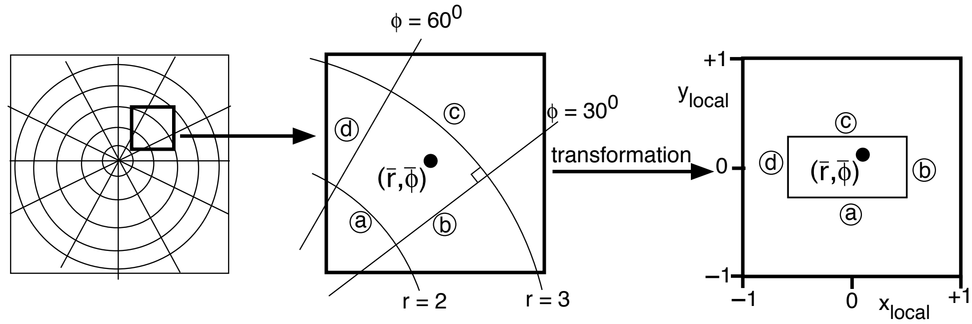


FIGURE 7 Inaccuracies due to polar coordinates on a flat sheet of paper. Coordinates in the middle frame are curved.

which we make measurements. Indeed, we can choose to use only global coordinates to apply general relativity (Box 2). Instead, in this book we *choose* to design and carry out measurements in a local inertial frame in order to unleash the power and simplicity of special relativity. In this process we fix average values of global coordinates to make constant the coefficients in the global metric. This allows us to write down the relation between global and local coordinates, equations (9) through (11), in order to generate a local flat spacetime metric (7).

We *choose* to use local coordinates.

But our choice has a cost that has nothing to do with spacetime curvature, illustrated by analogy to a flat geometric surface in Figure 7. The left frame shows polar coordinates laid out on the entire flat sheet. Choose a small area of the sheet (expanded in the second frame). That small area is, a *patch* (Definition 1) with a small section of *global* coordinates superimposed. This is a *frame* (Definition 2) whose local coordinate system is derived from global coordinates. The third frame shows Cartesian coordinates that cover the same patch, converting it to a local Cartesian frame, analogous to an inertial frame (Definition 3). What is the relation between the second frame and the third frame?

Approximation due to coordinate conversion

The exact differential separation between adjacent points is

$$ds^2 = dr^2 + r^2 d\phi^2 \tag{19}$$

In order to provide some "elbow room" to carry out local measurements on our small patch, we expand from differentials to small increments with the approximations:

$$\begin{aligned} \Delta s^2 &\approx \Delta r^2 + \bar{r}^2 \Delta\phi^2 \\ &\approx \Delta x^2 + \Delta y^2 \end{aligned} \tag{20}$$

Because of the average $\bar{r}$ due to curved coordinates, equation (20) is not exact. The approximation of this result has nothing to do with curvature, since the surface in the left panel is flat. A similar inexactness haunts the relation

Approximate due to (1) residual curvature plus (2) coordinate conversion.

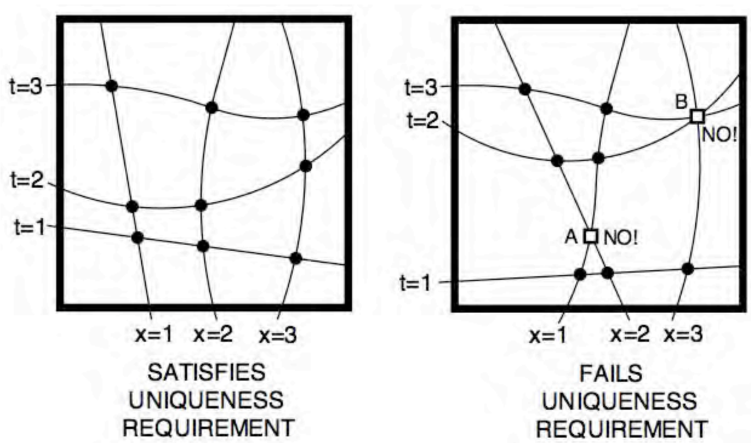


FIGURE 8 Left panel. Example of global coordinates that satisfy the uniqueness requirement: every event shown (filled circles) has a unique value of x and t. Right panel: Example of a global coordinate system that fails to satisfy the uniqueness requirement; Event A has two x-coordinates: $x = 1$ and $x = 2$; Event B has two t-coordinates: $t = 2$ and $t = 3$.

between global and local coordinates in equations (9) through (11). These equations are approximate for two reasons: (1) the residual curvature of spacetime across the local frame and (2) the conversion between global and local coordinates. In this book we emphasize the first of these, but the second is ever-present.

5.9 ■ REQUIREMENTS OF GLOBAL COORDINATE SYSTEMS

Which coordinate systems can we use in a global metric?

Some restrictions on global coordinates

Thus far we have put no restrictions on global coordinate systems for global metrics in general relativity. The basic requirements are a global coordinate system that (a) uniquely specifies the spacetime location of every event, and (b) when submitted to Einstein's equations results in a global metric. Here are three technical requirements, quoted from advanced theory without proof.

FIRST REQUIREMENT: UNIQUENESS

Unique set of coordinates for each event

The global coordinate system must provide a unique set of coordinates for each separate event in the spacetime region under consideration.

The uniqueness requirement seems reasonable. A set of global coordinates, for example t, r, ϕ, must allow us to distinguish any given event from every other event. That is, no two distinct events can have every global coordinate the same; nor can any given event be labelled by more than one set of coordinates. The left panel in Figure 8 shows an example of global coordinates that satisfy the uniqueness requirement; the right panel shows an example of global coordinates that fail this requirement.

Box 3. Find a particular local inertial frame.

How can we locate and label a particular local inertial frame on a shell around a black hole?

Ask a simpler question: How do we label and find one particular flat triangular surface on a Buckminster Fuller geodesic dome (Figure 6)? One way is simply to number each flat surface: triangle #523 next to triangle #524 next to triangle #525. Carry out this procedure for every flat triangle on the geodesic dome. The result is a huge catalog that we must consult to locate a given local flat segment on these nested Buckminster Fuller geodesic domes.

We could use a similar sequential numbering scheme to label and find a local inertial shell frame around a black hole, sequential in both space and time. But we already have a simpler way to index a single local inertial frame:

Equations (9) through (11) provide a much simpler indexing scheme: the average values $\bar{t}$, $\bar{r}$, and $\bar{\phi}$. Average $\bar{r}$ gives us the shell, average $\bar{\phi}$ locates the position of the local frame along the shell, and average $\bar{t}$ tells us the global t-coordinate of the frame at that location—local in time as well as space. Three numbers, for example $\bar{t}$, $\bar{r}$, and $\bar{\phi}$, specify precisely the local inertial shell frame in spacetime surrounding a black hole.

In addition to the uniqueness requirement, we must be able to set up a local inertial frame everywhere around the black hole (except on its singularity. To allow this possibility, we add the second, smoothness requirement:

Smooth coordinates

SECOND REQUIREMENT: SMOOTHNESS
The coordinates must vary smoothly from event to neighboring event. In practice, this means there must be a differentiable coordinate transformation that takes the global metric to a local inertial metric (except on a physical singularity).

The third and final requirement seems obvious to us in everyday life but is often the troublemaker in curved spacetime.

Every event has coordinates.

THIRD REQUIREMENT: COVERING OR EXTENSIBILITY
Every event must have coordinates. Coordinates must cover all spacetime.

Good and **bad** coordinates

Coordinates that satisfy all three requirements we will call **good coordinates**. Coordinates that fail to satisfy all three coordinates we will call **bad coordinates**. In flat spacetime we can find *good* coordinates that satisfy all three requirements. *In curved spacetime there are frequently no good coordinates.*

Frequently: no good coordinates in curved spacetime

The third requirement is often the first to be violated, because in many curved spacetimes a single coordinate system cannot cover the entire spacetime while preserving the first two conditions. A simple example is the sphere, which requires two good coordinate systems because latitude and longitude coordinates violate the second requirement at the poles. We usually ignore this while using polar coordinates, even though these coordinates are bad at $r = 0$ (Box 3, Section 3.1).

Comment 3. The (almost) complete freedom of general relativity
There are an unlimited number of valid global coordinate systems that describe spacetime around a stable object such as a star, white dwarf, neutron star, or black hole (Box 3, Section 7.5). Who chooses which global coordinate system to use? We do!

Near every event (except on a singularity) there are an unlimited number of possible local inertial frames in an unlimited number of relative motions. Who chooses the single local frame in which to carry out our next measurement? We do!

Nature has no interest whatsoever in which global coordinates we choose or how we derive from them the local inertial frames we employ to report our measurements and to check our predictions. Choices of global coordinates and local frames are (almost) completely free human decisions. Welcome to the wild permissiveness of general relativity!

5.10 ■ EXERCISES

5.1. Rotation of vertical

The inertial metric (7) assumes that the Δx_{shell} and Δy_{shell} are both straight-line separations that are perpendicular to one another. How many kilometers along a great circle must you walk before both the horizontal and vertical directions "turn" by one degree

A. on Earth.

B. on the Moon (radius 1 737 kilometers).

C. on the shell at map coordinate $r = 3M$ of a black hole of mass five times that of our Sun.

5.2. Time warping

In a given global coordinate system, two identical clocks stand at rest on different shells directly under one another, the lower clock at map coordinate r_{L}, the higher clock at map coordinate r_{H}. By *identical clocks* we mean that when the clocks are side by side the measured shell time between sequential ticks is the same for both. When placed on shells of different map radii, the measured time lapses between adjacent ticks are $\Delta t_{\text{shell H}}$ and $\Delta t_{\text{shell L}}$, respectively.

A. Find an expression for the fractional measured time difference f between the shell clocks, defined as:

$$f \equiv \frac{\Delta t_{\text{shell H}} - \Delta t_{\text{shell L}}}{\Delta t_{\text{shell L}}} \qquad (21)$$

This expression should depend on only the map r-values of the two clocks and on the mass M of the center of attraction.

B. Fix r_{L} of the lower shell clock. For what higher r_{H}-value does the fraction f have the greatest magnitude? Derive the expression f_{max} for this maximum fractional magnitude.

C. Evaluate the numerical value of the greatest magnitude $f_{\max}$ from Item B when r_L corresponds to the following cases:

(a) Earth's surface (numerical parameters inside front cover)

(b) Moon's surface (radius 1 737 kilometers, mass 5.45×10^{-5} meters)

(c) on the shell at $r_L = 3M$ of a black hole of mass $M = 5M_{\text{Sun}}$ (Find the value of M_{Sun} inside front cover)

D. Find the higher map coordinate r_H at which the fractional difference in clock rates is 10^{-10} for the cases in Item C.

E. For case (c) in item C, what is the directly-measured distance between the shell clocks?

F. What is the value of $f_{\max}$ in the limit $r_L \to 2M$? What is the value of f in the limit $r_L \to 2M$ and $r_H = 2M(1+\epsilon)$, where $0 < \epsilon \ll 1$. What does this result say about the ability of a light flash to move outward from the event horizon?

G. Which items in this exercise have different answers when the upper clock and the lower clock do *not* lie on the same radial line, that is when the upper clock is *not* directly above the lower clock?

5.3. The International Space Station as a local inertial frame

The International Space Station (ISS) orbits at an altitude of $d = 400$ kilometers above Earth's surface. Astronauts inside the ISS are (almost) in free float, because the ISS approximates an inertial frame. It is approximate, that is a *local* inertial frame because Earth's gravity causes tidal accelerations, tiny differences in gravitational accelerations at different locations.

The size of the ISS along the radial direction is $h = 20$ meters. Inside the ISS, at a point farthest from Earth, an astronaut releases a small wooden ball from rest. Simultaneously in the local ISS frame, along the same radial line but at a point nearest to Earth, another astronaut releases a small steel ball from rest. If the ISS did not depart from the specifications for an inertial frame, the two balls would remain at rest relative to each other.

A. Use a qualitative argument to show that tidal acceleration causes the two balls to move *apart* in the local ISS frame.

B. Use Newtonian mechanics to show that in the local ISS frame the wooden ball moves away from the steel ball with a relative acceleration given by the equation:

$$a = \frac{2GM_E h}{(R_E + d)^3} \approx 5.1 \times 10^{-5} \quad \text{meter/second}^2 \tag{22}$$

Here the subscript E refers to Earth, and G is the universal gravitational constant. How many seconds elapse in the ISS frame for the distance between the two balls to increase by 1 centimeter?

5.4. Diving inertial frame

Think of a local inertial frame constructed in a free capsule that dives past a local shell frame with local radial velocity v_{rel} measured by the shell observer. Use Lorentz transformations from Chapter 1 to find expressions similar to equations (9) through (11) that give coordinate increments Δt_{dive}, Δy_{dive}, and Δx_{dive} between a pair of events in the diving frame in terms of $\bar{r}$, v_{rel}, and global coordinate increments Δt, Δr, and $\Delta\phi$.

5.5. Tangentially moving inertial frame

Think of a local inertial frame constructed in a capsule that moves instantaneously in a tangential direction with tangential speed v_{rel} measured by the shell observer. Use Lorentz transformations from Chapter 1 to find expressions similar to equations (9) through (11) that give coordinate increments Δt_{tang}, Δy_{tang}, and Δx_{tang} between a pair of events in the tangentially-moving frame in terms of $\bar{r}$, v_{rel}, and global coordinate increments Δt, Δr, and $\Delta\phi$.

5.11 ■ REFERENCES

Albert Einstein quotes from "Autobiographical Notes," in *Albert Einstein: Philosopher-Scientist*, edited by Paul Arthur Schilpp, Volume VII of The Library of Living Philosophers, MJF Books, New York 1970, page 67.

Misner, Thorne, and Wheeler quote from Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler, *GRAVITATION*, W. H. Freeman Company, San Francisco [now New York], 1971, pages 302-303.

Wheeler on a phenomenon: Quoted in Robert J. Scully, *The Demon and the Quantum* (2007), page 191.

DownloadFileName: Ch05GlobalLocalMetrics170328v2.pdf

Chapter 6. Diving

- *Am I comfortable as I fall toward a black hole?*
- *How fast am I going when I reach the event horizon? Who measures my speed?*
- *How long do I live, measured on my wristwatch, as I fall into a black hole?*
- *How much does the mass of a black hole increase when a stone falls into it? when I fall into it?*
- *How close to a black hole can I stand on a spherical shell and still tolerate the "acceleration of gravity"?*

Download File Name: Ch06Diving170329v1.pdf

CHAPTER 6

Diving

Edmund Bertschinger & Edwin F. Taylor *

Many historians of science believe that special relativity could have been developed without Einstein; similar ideas were in the air at the time. In contrast, it's difficult to see how general relativity could have been created without Einstein – certainly not at that time, and maybe never.

—David Kaiser

6.1 ■ GO STRAIGHT: THE PRINCIPLE OF MAXIMAL AGING IN GLOBAL COORDINATES

"Go straight!" spacetime shouts at the stone.
The stone's wristwatch verifies that its path is straight.

Section 5.7 described how an observer passes through a sequence of local inertial frames, making each measurement in only one of these local frames. Special relativity describes motion in each such local inertial frame. The observer is just a stone that acts with purpose. Now we ask how a (purposeless!) free stone moves in global coordinates.

Section 1.6 introduced the Principle of Maximal Aging that describes motion in a single inertial frame. To describe global motion, we need to extend this principle to a *sequence* of adjacent local inertial frames. Here, without proof, is the simplest possible extension, to a *single adjacent pair* of local inertial frames.

Definition: **Principle of Maximal Aging** in curved spacetime

DEFINITION 1. Principle of Maximal Aging (curved spacetime)
The *Principle of Maximal Aging* states that a free stone follows a worldline through spacetime such that its wristwatch time (aging) is a maximum when summed across every adjoining pair of local inertial frames along its worldline.

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

Box 1. What Then Is Time?

What then is time? If no one asks me, I know what it is. If I wish to explain it to him who asks me, I do not know.

****

The world was made, not in time, but simultaneously with time. There was no time before the world.

—St. Augustine (354–430 C.E.)

Time takes all and gives all.

—Giordano Bruno (1548–1600 C.E.)

Everything fears Time, but Time fears the Pyramids.

—Anonymous

Philosophy is perfectly right in saying that life must be understood backward. But then one forgets the other clause—that it must be lived forward.

—Søren Kierkegaard

As if you could kill time without injuring eternity.

****

Time is but the stream I go a-fishing in.

—Henry David Thoreau

Although time, space, place, and motion are very familiar to everyone, . . . it is useful to distinguish these quantities into absolute and relative, true and apparent, mathematical and common.

—Isaac Newton

Time is defined so that motion looks simple.

—Misner, Thorne, and Wheeler

Nothing puzzles me more than time and space; and yet nothing troubles me less, as I never think about them.

—Charles Lamb

Either this man is dead or my watch has stopped.

—Groucho Marx

“What time is it, Casey?”

“You mean right now?”

—Casey Stengel

It’s good to reach 100, because very few people die after 100.

—George Burns

The past is not dead. In fact, it’s not even past.

—William Faulkner

Time is Nature’s way to keep everything from happening all at once.

—Graffito, men’s room, Pecan St. Cafe, Austin, Texas

What time does this place get to New York?

—Barbara Stanwyck, during trans-Atlantic crossing on the steamship *Queen Mary*

Objection 1. *Now you have gone off the deep end! In Chapter 1, Speeding, you convinced me that the Principle of Maximal Aging was nothing more than a restatement of Newton’s First Law of Motion, the observation that in flat spacetime the free stone moves at constant speed along a straight line in space. But in curved spacetime the stone’s path will obviously be curved. You have violated your own Principle.*

On the contrary, we have changed the Principle of Maximal Aging as little as possible in order to apply it to curved spacetime. We require the free stone to move along a straight worldline across *each one* of the pair of adjoining local inertial frames, as demanded by the special relativity Principle of Maximal Aging in each frame. We allow the stone only the choice of one map coordinate of the event, at the boundary between these two frames. That single generalization extends the Principle of Maximal Aging from flat to curved spacetime. And the result is a single kink in the worldline. When we shrink all adjoining inertial frames along the worldline to the calculus limit, then the result is what you predict: a curved worldline in global coordinates.

Now we can use the more general Principle of Maximal Aging to discover a constant of motion for a free stone, what we call its *map energy.*

6.2 ■ MAP ENERGY FROM THE PRINCIPLE OF MAXIMAL AGING

The global metric plus the Principle of Maximal Aging leads to map energy as a constant of motion.

Map energy: a constant of motion

This section uses the Principle of Maximal Aging from Section 6.1, plus the Schwarzschild global metric to derive the expression for map energy of a free stone near a nonspinning black hole. For a free stone, map energy is a constant of motion; its value remains the same as the stone moves. Our derivation uses a stone that falls inward along the r-direction, but at the end we show that the resulting expression for map energy also applies to a stone moving in any direction; energy is a *scalar*, which has no direction.

?

Objection 2. *Here is a fundamental objection to the Principle of Maximal Aging: You nowhere derive it, yet you expect us readers to accept this arbitrary Principle. Why should we believe you?*

!

Guilty as charged! Our major tool in this book is the metric, which—along with the topology of a spacetime region—tells us everything we can know about the shape of spacetime in that region. But the shape of spacetime revealed by the metric tells us nothing whatsoever about how a free stone moves in this spacetime. For that we need a second tool, the Principle of Maximal Aging which, like the metric, derives from Einstein's field equations. In this book the metric plus the Principle of Maximal Aging—both down one step from the field equations—are justified by their immense predictive power. Until we derive the metric in Chapter 22, we apply the slogan, "Handsome is as handsome does!"

Find maximal aging: find natural motion.

The Principle of Maximal Aging maximizes the stone's total wristwatch time across *two adjoining* local inertial frames. Figure 1 shows the Above Frame A (of average map coordinate $\bar{r}_{\mathrm{A}}$) and adjoining Below Frame B (of average map coordinate $\bar{r}_{\mathrm{B}}$). The stone emits initial flash 1 as it enters the top of Frame A, emits middle flash 2 as it transits from Above Frame A to Below Frame B, and emits final flash 3 as it exits the bottom of Below Frame B. We use the three *flash emission events* to find maximal aging.

Outline of the method: Fix the r- and ϕ-coordinates of all three flash emissions and fix the t-coordinates of upper and lower events 1 and 3. Next vary the t-coordinate of the middle flash emission 2 to maximize the total *wristwatch time* (aging) of the stone across both frames.

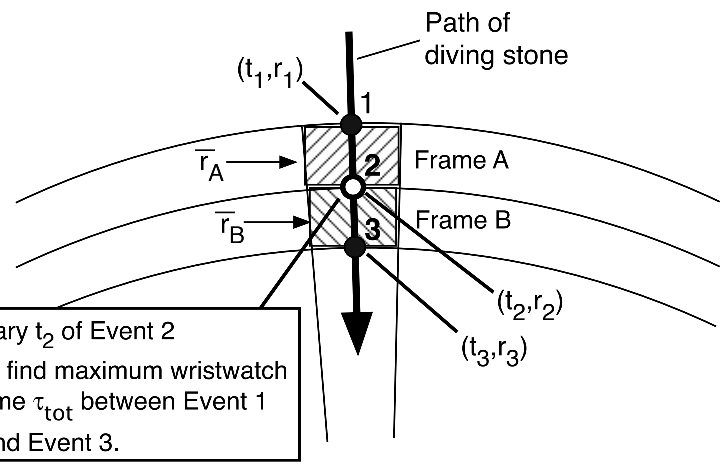


FIGURE 1 Use the Principle of Maximal Aging to derive the expression for Schwarzschild map energy. The diving stone first crosses the Above Frame A, then crosses the Below Frame B, emitting flashes at events 1, 2, and 3. Fix all three coordinates of events 1 and 3; but fix only the r- and ϕ-coordinates of intermediate event 2. Then vary the t-coordinate of event 2 to maximize the *total wristwatch time* (aging) across both frames between fixed end-events 1 and 3. This leads to expression (8) for the stone's map energy, a constant of motion.

So much for t-coordinates. How do we find *wristwatch times* across the two frames? The Schwarzschild metric ties the increment of wristwatch time to changes in r- and t-coordinates for a stone that falls inward along the r-coordinate. Write down the approximate form of the global metric twice, first for Above frame A (at average $\bar{r}_A$) and second for the Below frame B (at average $\bar{r}_B$). Take the square root of both sides:

Approximate the Schwarzschild metric for each frame.

$$\tau_A \approx \left[\left(1 - \frac{2M}{\bar{r}_A}\right)(t_2 - t_1)^2 + (\text{terms without } t\text{-coordinate})\right]^{1/2} \qquad (1)$$

$$\tau_B \approx \left[\left(1 - \frac{2M}{\bar{r}_B}\right)(t_3 - t_2)^2 + (\text{terms without } t\text{-coordinate})\right]^{1/2} \qquad (2)$$

We are interested only in those parts of the metric that contain the map t-coordinate, because we take derivatives with respect to that t-coordinate. To prepare for the derivative that leads to maximal aging, take the derivative of τ_A with respect to t_2 of the intermediate event 2. The denominator in the resulting derivative is just τ_A:

$$\frac{d\tau_A}{dt_2} \approx \left(1 - \frac{2M}{\bar{r}_A}\right)\frac{(t_2 - t_1)}{\tau_A} \qquad (3)$$

The corresponding expression for $d\tau_B/dt_2$ is:

$$\frac{d\tau_{\mathrm{B}}}{dt_2} \approx -\left(1-\frac{2M}{\bar{r}_{\mathrm{B}}}\right)\frac{(t_3-t_2)}{\tau_{\mathrm{B}}} \tag{4}$$

Add the two wristwatch times to obtain the summed wristwatch time τ_{tot} between first and last events 1 and 3:

$$\tau_{\mathrm{tot}} = \tau_{\mathrm{A}} + \tau_{\mathrm{B}} \tag{5}$$

Maximize aging summed across both frames.

Recall that we keep constant the total t-coordinate separation across both frames. To find the maximum total wristwatch time, take the derivative of both sides of (5) with respect to t_2, substitute from (3) and (4), and set the result equal to zero in order to find the maximum:

$$\frac{d\tau_{\mathrm{tot}}}{dt_2} = \frac{d\tau_{\mathrm{A}}}{dt_2} + \frac{d\tau_{\mathrm{B}}}{dt_2} \approx \left(1-\frac{2M}{\bar{r}_{\mathrm{A}}}\right)\frac{(t_2-t_1)}{\tau_{\mathrm{A}}} - \left(1-\frac{2M}{\bar{r}_{\mathrm{B}}}\right)\frac{(t_3-t_2)}{\tau_{\mathrm{B}}} \approx 0 \tag{6}$$

From the last approximate equality in (6),

$$\left(1-\frac{2M}{\bar{r}_{\mathrm{A}}}\right)\frac{(t_2-t_1)}{\tau_{\mathrm{A}}} \approx \left(1-\frac{2M}{\bar{r}_{\mathrm{B}}}\right)\frac{(t_3-t_2)}{\tau_{\mathrm{B}}} \tag{7}$$

The expression on the left side of (7) depends only on parameters of the stone's motion across the Above Frame A; the expression on the right side depends only on parameters of the stone's motion across the Below Frame B. Hence the value of either side of this equation must be independent of *which* adjoining pair of frames we choose to look at: this pair can be *anywhere* along the worldline of a stone. Equation (7) displays a quantity that has the same value on *every* local inertial frame along the worldline. We have found the expression for a quantity that is a constant of motion.

Map energy of a stone in Schwarzschild coordinates

Now shrink differences $(t_2 - t_1)$ and $(t_3 - t_2)$ in (7) to their differential limits. In this process the average r-coordinate becomes exact, so $\bar{r} \to r$. Next use the result to *define* the stone's **map energy per unit mass**:

$$\frac{E}{m} \equiv \left(1-\frac{2M}{r}\right)\frac{dt}{d\tau} \qquad \text{(map energy of a stone per unit mass)} \tag{8}$$

Far from the black hole, map energy takes special relativity form.

Why do we call the expression on the right side of (8) *energy* (per unit mass)? Because when the mass M of the center of attraction becomes very small—or when the stone is very far from the center of attraction—the limit $2M/r \to 0$ describes a stone in flat spacetime. That condition reduces (8) to $E/m = dt/d\tau$, which we recognize as equation (23) in Section 1.7 for E/m in flat spacetime. Hence we take the right side of (8) to be the general-relativistic generalization, near a nonspinning black hole, of the special relativity expression for E/m.

Map energy E same unit as m

Note that the right side of (8) has no units; therefore both E and m on the left side must be expressed in the *same* unit, a unit that we may choose for our convenience. *Both* numerator and denominator in E/m may be expressed in kilograms or joules or electron-volts or the mass of the proton, or any other common unit.

Map energy expression valid for *any* motion of the stone.

Our derivation of map energy employs only the t-coordinate in the metric. It makes no difference in the outcome for map energy—expression (8)—whether dr or $d\phi$ is zero or not. This has an immediate consequence: The expression for map energy in Schwarzschild global coordinates is valid for a free stone moving on *any* orbit around a spherically symmetric center of attraction, not just along the inward r-direction. We will use this generality of (8) to predict the general motion of a stone in later chapters.

6.3 ■ UNICORN MAP ENERGY VS. MEASURED SHELL ENERGY

Map energy is like a unicorn: a mythical beast

Map energy E/m is a unicorn: a mythical beast.

The expression on the right side of equation (8) is like a unicorn: a mythical beast. Nobody measures directly the r- or t-coordinates in this expression, which are Schwarzschild global map coordinates: entries in the mapmaker's spreadsheet or accounting form. Nobody measures E/m on the left side of (8) either; the map energy is also a unicorn. If this is so, why do we bother to derive expression (8) in the first place? Because E/m has an important virtue: It is a constant of motion of a free stone in Schwarzschild global coordinates; it has the same value at every event along the global worldline of a stone. The value of E/m helps us to predict its global motion (Chapters 8 and 9). But it does not tell us the value of the energy measured by an observer in a local inertial frame.

What is the stone's energy measured by the shell observer? The special relativity energy expression is valid for the shell observer. Equation (9) in Section 5.7 gives us:

$$\Delta t_{\text{shell}} = \left(1 - \frac{2M}{\bar{r}}\right)^{1/2} \Delta t \tag{9}$$

Then:

$$\frac{E_{\text{shell}}}{m} = \lim_{\Delta\tau \to 0} \frac{\Delta t_{\text{shell}}}{\Delta\tau} = \lim_{\Delta\tau \to 0} \left(1 - \frac{2M}{\bar{r}}\right)^{1/2} \frac{\Delta t}{\Delta\tau} \tag{10}$$

As we shrink increments to the differential calculus limit, the average r-coordinate becomes exact: $\bar{r} \to r$. The result is:

$$\frac{E_{\text{shell}}}{m} = \left(1 - \frac{2M}{r}\right)^{1/2} \frac{dt}{d\tau} \quad \text{(shell energy of a stone per unit mass)} \tag{11}$$

Into this equation substitute expression (8) for the stone's map energy to obtain:

$$\frac{E_{\text{shell}}}{m} = \frac{1}{(1 - v_{\text{shell}}^2)^{1/2}} = \left(1 - \frac{2M}{r}\right)^{-1/2} \frac{E}{m} \qquad (12)$$

Shell energy

where we have added the special relativity expression (23) in Section 1.7. Equation (12) tells us how to use the map energy—a unicorn—to predict the frame energy directly measured by the shell observer as the stone streaks past.

Expression (12) for shell energy E_{shell} applies to a stone moving in any direction, not just along the r-coordinate. Why? Energy—including map energy E—is a *scalar*, a property of the stone independent of its direction of motion.

Different shell observers compute same map energy.

The shell observer knows only his local shell frame coordinates, which are restricted in order to yield a local inertial frame. He observes a stone zip through his local frame and disappear from that frame; he has no global view of the stone's path. However, equation (12) is valid for a stone in *every* local shell frame and for *every* direction of motion of the stone in that frame. The shell observer uses this equation and his local r—stamped on every shell—to compute the map energy E/m, then radios his result to every one of his fellow shell observers, for example, "The green-colored free stone has map energy $E/m = 3.7$." A different shell observer, at different map r, measures a different value of shell energy E_{shell}/m of the green stone as it streaks through his own local frame, typically in a different direction. However, armed with (12), every shell observer verifies the constant value of map energy of the green stone, for example $E/m = 3.7$.

In brief, each local shell observer carries out a real measurement of shell energy; from this result plus his knowledge of his r-coordinate he derives the value of the map energy E/m, then uses this map energy—a constant of motion—to predict results of shell energy measurements made by shell observers distant from him. The result is a multi-shell account of the entire orbit of the stone.

The entire scheme of shell observers depends on the existence of local shell frames, which cannot be built inside the event horizon. Now we turn to the experience of the diver who passes inward across the event horizon.

6.4 ■ RAINDROP CROSSES THE EVENT HORIZON

Convert t-coordinate to raindrop wristwatch time.

How to get inside the event horizon?

The Schwarzschild metric satisfies Einstein's field equations everywhere in the vicinity of a nonrotating black hole (except on its singularity at $r = 0$). Map coordinates alone may satisfy Schwarzschild and Einstein, but they do not satisfy us. We want to make every measurement in a local inertial frame. Shell frames serve this purpose nicely outside the event horizon, but we cannot construct stationary shells inside the event horizon. Moreover, the expression $(1 - 2M/r)^{-1/2}$ in energy equation (12) becomes imaginary inside the event horizon, which provides one more indication that shell energy does not apply there.

Yet everyone tells us that an unfortunate astronaut who crosses inward through the event horizon at $r = 2M$ inevitably arrives at the lethal central singularity $r = 0$. In the following chapter we build a local frame around a falling astronaut. To prepare for such a local diving frame, we start here as simply as possible: We ask the stone wearing a wristwatch that began our study of relativity (Section 1.1) to take a daring dive, to drop from initial rest far from the black hole and plunge inward to $r = 0$. We call this diving, wristwatch-wearing stone a **raindrop**, because on Earth raindrops fall from rest at a great height. By definition, the raindrop has no significant spatial extent; it has no frame, it is just a stone wearing a wristwatch.

Raindrop defined: stone dropped from rest far away

DEFINITION 2. Raindrop

A **raindrop** is a stone wearing a wristwatch, that freely falls inward starting from initial rest far from the center of attraction.

Map energy of a raindrop

Examine the map energy (8) of a raindrop. Far from the black hole $r \gg 2M$ so that $(1 - 2M/r) \rightarrow 1$. For a stone at rest there, $dr = d\phi = 0$ and the Schwarzschild metric tells us that $d\tau \rightarrow dt$. As a result, (8) becomes:

$$\frac{E}{m} \equiv \left(1 - \frac{2M}{r}\right)\frac{dt}{d\tau} \rightarrow 1 \quad \text{(raindrop: released from rest at } r \gg 2M) \quad (13)$$

The raindrop, released from rest far from the black hole, must fall inward along a radial line. In other words, $d\phi = 0$ along the raindrop worldline. Formally we write:

$$\frac{d\phi}{d\tau} = 0 \qquad \text{(raindrop)} \qquad (14)$$

The raindrop-stone, released from rest at a large r map coordinate, begins to move inward, gradually picks up speed, finally plunges toward the center. As the raindrop hurtles inward, the value of $E/m\,(= 1)$ remains constant. Equation (12) then tells us that as r decreases, $2M/r$ increases, and so E_{shell} must also increase, implying an increase in v_{shell}. The local shell observer measures this increased speed directly. Equation (12) with $E/m = 1$ for the raindrop yields:

Shell energy of the raindrop

$$\frac{E_{\text{shell}}}{m} = \left(1 - v_{\text{shell}}^2\right)^{-1/2} = \left(1 - \frac{2M}{r}\right)^{-1/2} \qquad \text{(raindrop)} \qquad (18)$$

It follows immediately that:

$$v_{\text{shell}} = -\left(\frac{2M}{r}\right)^{1/2} \qquad \text{(raindrop shell velocity)} \qquad (19)$$

where the negative value of the square root describes the stone's inward motion. Equation (19) shows that the shell-measured speed of the

Box 2. Slow speed + weak field $\Longrightarrow$ Mass + Newtonian KE and PE

"If you fall, I'll be there." —Floor

The map energy E/m may be a unicorn in general relativity, but it is a genuine race horse in Newtonian mechanics. We show here that the map energy E/m of a stone moving at non-relativistic speed in a weak gravitational field reduces to the mass of the stone plus the familiar Newtonian energy (kinetic + potential). Rearrange (12) to read:

$$\frac{E}{m} = \left(1 - \frac{2M}{r}\right)^{1/2} \left(1 - v_{\text{shell}}^2\right)^{-1/2} \qquad (15)$$

For $r \gg 2M$ (weak gravitational field) and $v_{\text{shell}}^2 \ll 1$ (non relativistic stone speed) use the approximation inside the front cover twice:

$$\left(1 - \frac{2M}{r}\right)^{1/2} \approx 1 - \frac{M}{r} \qquad (r \gg 2M) \qquad (16)$$

$$\left(1 - v_{\text{shell}}^2\right)^{-1/2} \approx 1 + \frac{1}{2} v_{\text{shell}}^2 \qquad (v_{\text{shell}}^2 \ll 1)$$

Substitute these into (15) and drop the much smaller product $(M/2r)v_{\text{shell}}^2$. The result is an approximation:

$$E \approx m + \frac{1}{2} m v_{\text{shell}}^2 - \frac{Mm}{r} \qquad (17)$$

$$\left(r \gg 2M,\ \ v_{\text{shell}}^2 \ll 1\right)$$

In this equation $-Mm/r$ is the gravitational potential energy of the stone. (In conventional mks units it would read $-GM_{\text{kg}}m_{\text{kg}}/r$.) We recognize in (17) Newtonian's kinetic energy (KE) plus his potential energy (PE) of a stone, with the added stone's mass m.

As a jockey in curved spacetime, you must beware of riding the unicorn map energy E/m; gravitational potential energy is a fuzzy concept in general relativity. Dividing energy into separate kinetic and potential forms works only under special conditions, such as those given in equation (16).

Except for these special conditions, we expect the map constant of motion E to differ from E_{shell}: The local shell frame is inertial and excludes effects of curved spacetime. In contrast, map energy E—necessarily expressed in map coordinates—includes curvature effects, which Newton attributes to a "force of gravity."

The approximation in (17) is quite profound. It reproduces a central result of Newtonian mechanics without using the concept of force. In general relativity, we can always eliminate gravitational force (see inside the back cover).

raindrop—the magnitude of its velocity—increases to the speed of light at the event horizon. This is a limiting case, because we cannot construct a shell—even in principle—at the exact location of the event horizon.

?

Objection 3. *I am really bothered by the idea of a material particle such as a stone traveling across the event horizon as a particle. The shell observer sees it moving at the speed of light, but it takes light to travel at light speed. Does the stone—the raindrop—become a flash of light at the event horizon?*

!

No. Be careful about limiting cases. No shell can be built *at* the event horizon, because the initial gravitational acceleration increases without limit there (Section 6.7). An observer on a shell just outside the event horizon clocks the diving stone to move with a speed *slightly less* than the speed of light. Any directly-measured stone speed less than the speed of light is perfectly legal in relativity. So there is no contradiction.

Raindrop dr/dt

Compare the shell velocity (19) of the raindrop with the value of dr/dt at a given r-coordinate. To derive dr/dt, solve the right-hand equation in (13) for $d\tau$ and substitute the result into the Schwarzschild metric with $d\phi = 0$. Result for the raindrop:

Sample Problems 1. The Neutron Star Takes an Aspirin

Neutron Star Gamma has a total mass 1.4 times that of our Sun and a map $r_0 = 10$ kilometers. An aspirin tablet of mass one-half gram falls from rest at a large r coordinate onto the surface of the neutron star. An advanced civilization converts into useful energy the entire kinetic energy of the aspirin tablet, measured in the local surface rest frame. Estimate how long this energy will power a 100-watt bulb. Repeat the analysis to find the useful energy for the case of an aspirin tablet falling from a large r coordinate onto the surface of Earth.

SOLUTION

From the value of the mass of our Sun (inside the front cover), the mass of the neutron star is $M \approx 2 \times 10^3$ meters. Hence $2M/r_0 \approx 2/5$. Far from the neutron star the total map energy of the aspirin tablet equals its rest energy, namely its mass, hence $E/m = 1$. From (18), the shell energy of the aspirin tablet just before it hits the surface of the neutron star rises to the value

$$\frac{E_{\text{shell}}}{m} = \left(1 - \frac{2M}{r_0}\right)^{-1/2} \approx 1.3 \qquad \text{(Neutron Star)} \tag{20}$$

The shell *kinetic energy* of the half-gram aspirin tablet is 0.3 of its rest energy. The rest energy is $m = 0.5$ gram = 5 $\times 10^{-4}$ kilogram or $mc^2 = 4.5 \times 10^{13}$ joules. The fraction 0.3 of this is 1.35×10^{13} joules. One watt is one joule/second; a 100-watt bulb consumes 100 joules per second. At that rate, the bulb can burn for 1.35×10^{11} seconds on the kinetic energy of the aspirin tablet. One year is about 3×10^7 seconds. Result: The kinetic energy of the half-gram aspirin tablet falling to the surface of Neutron Star Gamma from a large r coordinate provides energy sufficient to light a 100-watt bulb for approximately 4500 years!

What happens when the aspirin tablet falls from a large r coordinate onto Earth's surface? Set the values of M and r_0 to those for Earth (inside front cover). In this case $2M \ll r_{\text{E}}$, so equation (20) becomes, to a very good approximation:

$$\frac{E_{\text{shell}}}{m} \approx \left(1 + \frac{M}{r_0}\right) \approx 1 + 6.97 \times 10^{-10} \qquad \text{(Earth)} \tag{21}$$

Use the same aspirin tablet rest energy as before. The lower fraction of kinetic energy yields 3.14×10^4 joules. At 100 joules per second the kinetic energy of the aspirin tablet will light the 100-watt bulb for 314 seconds, or a little more than 5 minutes.

$$\frac{dr}{dt} = -\left(1 - \frac{2M}{r}\right)\left(\frac{2M}{r}\right)^{1/2} \qquad \text{(raindrop)} \tag{22}$$

Raindrop dr/dt: a unicorn!

Equation (22) shows an apparently outrageous result: as the raindrop reaches the event horizon at $r = 2M$, its Schwarzschild dr/dt drops to zero. (This result explains the strange spacing of event-dots along the orbit approaching the event horizon in Figure 3.6.) Does any local observer witness the stone coasting to rest? No! Repeated use of the word "map" reminds us that map velocities are simply spreadsheet entries for the Schwarzschild mapmaker and need not correspond to direct measurements by any local observer. Figure 2 shows plots of both shell speeds and map $dr/d\tau$ of the descending raindrop. Nothing demonstrates more clearly than the diverging lines in Figure 2 the radical difference between (unicorn) map entries and the results of direct measurement.

Does the raindrop cross the event horizon or not? To answer that question we need to track the descent with its directly-measured wristwatch time, not the global t-coordinate. Use equation (13) to convert global coordinate differential dt to wristwatch differential $d\tau$. With this substitution, (22) becomes:

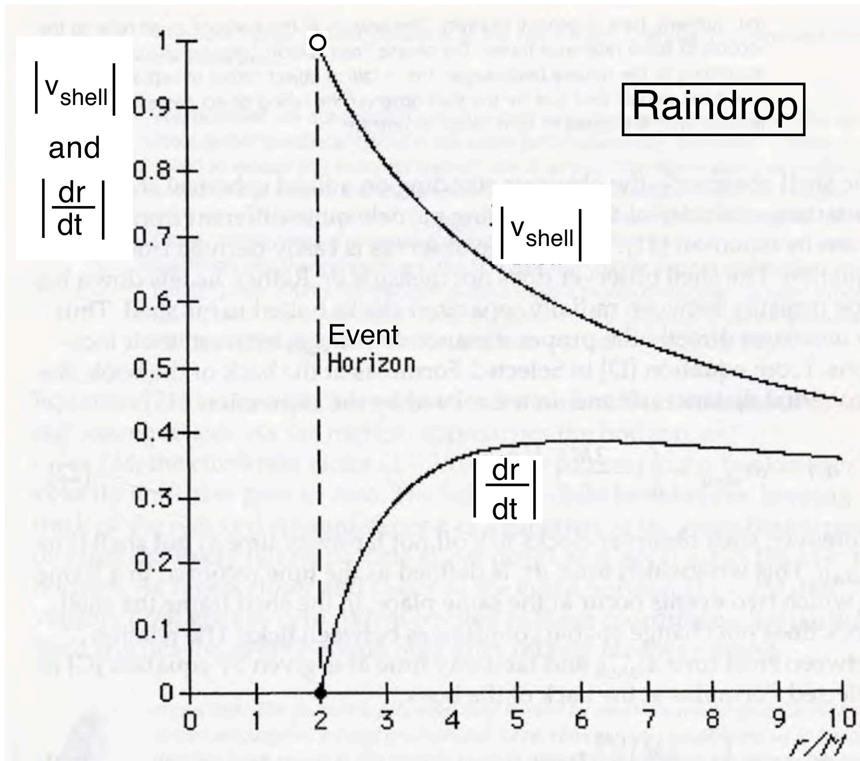


FIGURE 2 Computer plot of the speed $|v_{\rm shell}|$ of a raindrop directly measured by shell observers at different r-values, from (19), and its Schwarzschild map speed $|dr/dt|$ from (22). Far from the black hole the raindrop is at rest, so both speeds are zero, but both speeds increase as the raindrop descends. Map speed $|dr/dt|$ is not measured but computed from spreadsheet records of the Schwarzschild mapmaker. At the event horizon, the measured shell speed rises to the speed of light, while the computed map speed drops to zero. The upper open circle at $r = 2M$ reminds us that this is a limiting case, since no shell can be constructed *at* the event horizon. (Why not? See the Appendix, Section 6.7.)

$$\frac{dr}{d\tau_{\rm raindrop}} = -\left(\frac{2M}{r}\right)^{1/2} \tag{23}$$

Raindrop crosses the event horizon.

Expression (23) combines a map quantity dr with the differential advance of the wristwatch $d\tau_{\rm raindrop}$. It shows that the raindrop's r-coordinate decreases as its wristwatch time advances, so the raindrop passes inward through the event horizon. Indeed, inside the event horizon the magnitude of $dr/d\tau_{\rm raindrop}$ becomes greater than one, and increases without limit as $r \to 0$. But this need not worry us: Both r and dr are map quantities, so $dr/d\tau$ is just an entry on the mapmaker's spreadsheet, not a quantity measured by anyone.

Comment 1. How do we find the value of dr inside the event horizon?
The numerator dr on the left side of (23) has a clear meaning only *outside* the

Box 3. Newton Predicts the Black Hole?

It's remarkable how well much of Newton's mechanics works—sort of—on the stage of general relativity. One example is that Newton appears to predict the r-coordinate of the event horizon $r = 2M$. Yet the meaning of that barrier is strikingly different in the two pictures of gravity, as the following analysis shows.

A stone initially at rest far from a center of attraction drops inward. Or a stone on the surface of Earth or of a neutron star is fired outward along r, coming to rest at a large r coordinate. In either case, Newtonian mechanics assigns the same total energy (kinetic plus potential) to the stone. We choose the gravitational potential energy to be zero at the large r coordinate, and the stone out there does not move. From (17), we then obtain

$$\frac{E}{m} - 1 = \frac{v^2}{2} - \frac{M}{r} = 0 \quad \text{(Newton)} \qquad (24)$$

From (24) we derive the diving (or rising) speed at any r-coordinate:

$$|v| = \left(\frac{2M}{r}\right)^{1/2} \quad \text{(Newton)} \qquad (25)$$

which is the same as equation (19) for the shell speed of the raindrop. One can predict from (25) the r-value at which the speed reaches one, the speed of light, which yields $r = 2M$, the black hole event horizon. For Newton the speed of light is the **escape velocity** from the event horizon.

Newton assumes a single universal inertial reference frame and universal time, whereas (19) applies only to shell separation divided by shell time. A quite different expression (22) describes dr/dt—map differential dr divided by map differential dt—for raindrops.

Does Newton correctly describe black holes? No. Newton predicts that a stone launched radially outward from the event horizon with a speed less than that of light will rise to higher r, slow, stop without escaping, then fall back. In striking contrast, Einstein predicts that nothing, not even light, can be successfully launched outward from inside the event horizon, and that light launched outward *exactly at* the event horizon hovers there, balanced as on a knife-edge (Box 4).

event horizon, where every shell displays the stamped value of r. Box 7 in Section 7.8 describes one practical method by which a descending rain observer can measure map r, both outside and inside the event horizon.

6.5 ■ GRAVITATIONAL MASS

A new way to measure total energy

Mass m of the stone

This book uses the word *mass* in two different ways. Symbol m in equations (8) and (11) represents the inertial mass of a test particle, which we call a *stone*. By definition, the mass of a stone is too small to curve spacetime by a detectable amount. Expression (8) measures the stone's map energy E and mass m in the same units.

Mass M of the center of attraction

The mass M of the center of attraction is quite different: It is the gravitational mass that curves spacetime, as reflected in the global metric expression $(1 - 2M/r)$.

Drop a stone of mass m into a star of mass M.

What happens when a stone of mass m falls into a black hole of mass M? Does the swallowed mass m increase the black hole's mass? Our new understanding of energy helps us to calculate how much the mass of a black hole grows when it swallows matter—and yields a surprising result. To begin, start with a satellite orbiting close to a star. How can we measure the total gravitational mass of the star-plus-satellite system? We make this measurement using the initial acceleration of a distant test particle so remote

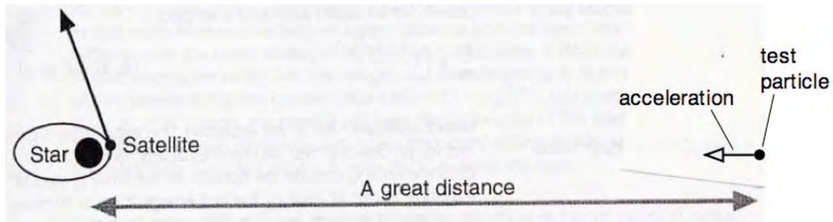


FIGURE 3 Measure the total mass-energy M_{total} of a central star-satellite system using the acceleration of a test particle at a large r coordinate, analyzed using Newtonian mechanics.

that Newtonian mechanics gives a correct result (Figure 3). In units of inverse meters, Newton's expression for this acceleration is:

$$a = -\frac{M_{\mathrm{total}}}{r^2} \qquad \text{(Newton)} \qquad (26)$$

Newton says, "Add m to M_{star}."

What is M_{total}? In Newtonian mechanics total mass equals the mass M_{star} of the original star plus the mass m of the satellite orbiting close to it:

$$M_{\mathrm{total}} = M_{\mathrm{star}} + m \qquad \text{(Newton)} \qquad (27)$$

Could this also be true in general relativity? The answer is no, but proof requires a sophisticated analysis of Einstein's equations.

Birkhoff's theorem

A mathematical theorem of general relativity due to G. D. Birkhoff in 1923 states that the spacetime outside any spherically symmetric distribution of matter and energy is completely described by the Schwarzschild metric with a *constant* gravitational mass M_{total}, no matter whether that spherically symmetric source is at rest or, for example, moving inward or outward along the r-coordinate.

M_{total} includes contracting bubble of dust.

In order to apply Birkhoff's theorem, we approximate the moving satellite of Figure 3 by the inward-falling uniform spherical bubble of Figure 4, a bubble composed of unconnected particles—dust—whose total mass m is the same as that of the satellite in Figure 3. (We use the label "bubble" instead of "shell" to avoid confusion with the stationary concentric shells we construct around a black hole on which we make measurements and observations.) This falling uniform dust bubble satisfies the condition of Birkhoff's theorem, so the Schwarzschild metric applies outside this inward-falling bubble.

How does dust bubble increase M_{total}?

Unfortunately, Birkhoff's theorem does not tell us how to calculate the value of M_{total}, only that it is a constant for any spherically symmetric configuration of mass/energy. What property of the dust bubble remains constant as it falls inward? Its inertial mass m? Not according to special relativity! Inertial mass is *not* conserved; it can be converted into energy. We had better look for a conserved energy for our infalling dust bubble. Equation (12) is our guide: At a given r-coordinate every particle of dust in the collapsing bubble falls inward at the same rate, so the measure of the total

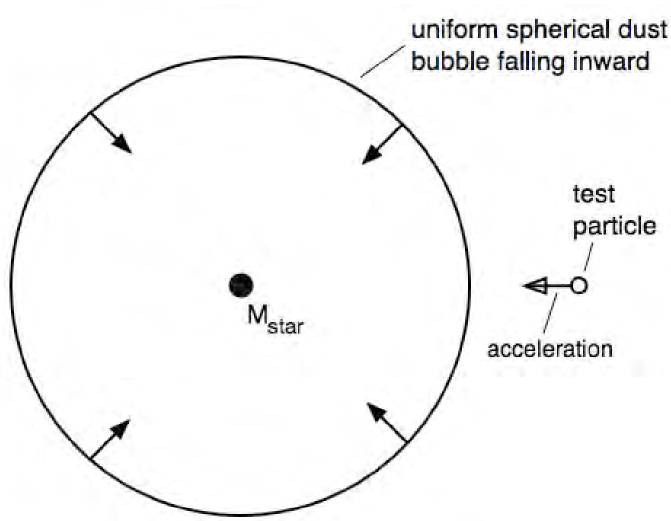


FIGURE 4 Replace the moving satellite of Figure 3 with an inward-falling uniform spherical bubble of dust that satisfies the condition of Birkhoff's theorem, so the Schwarzschild metric applies outside the contracting dust bubble.

shell energy E_{shell} of the bubble at a given r-coordinate is the sum over the individual particles of the dust bubble. Clearly from (12), successive shell observers at successively smaller r-coordinates measure successively higher values of E_{shell} as the collapsing dust bubble falls past them, so we cannot use shell energy in the Birkhoff analysis.

Einstein: "Add dust bubble E to M_{star} to find M_{total}."

However, the Schwarzschild map energy E *does* remain constant during this collapse. So instead of the Newtonian expression (27) we have the trial general relativity replacement:

$$M_{\text{total}} = M_{\text{star}} + E \qquad \text{(Einstein)} \tag{28}$$

Check validity of (28).

How do we know whether or not the total map energy E of the dust bubble is the correct constant to add to M_{star} in order to yield the total mass M_{total} of the system? One check is that when the satellite/dust bubble is far from the star ($r \gg 2M_{\text{total}}$) but the remote test particle is still exterior to the dust bubble, then $E \to E_{\text{shell}}$ from (12). In addition, for a slow-moving satellite/dust bubble, $E \to E_{\text{shell}} \to m$, and we recover Newton's formula (27), as we should in the limits $r \gg 2M$ and $v^2_{\text{shell}} \ll 1$. And when the satellite/dust bubble falls inward so that our stationary shell observer measures $E_{\text{shell}} > m$, then equation (28) remains valid, because $E(\approx m)$ does not change. Note that Birkhoff's Theorem is satisfied in this approximation.

Result: Convert stone map E into gravitational mass.

If (28) is correct, then general relativity merely replaces Newton's m in (27) with total map energy E, a constant of motion for the satellite/bubble. Thus the mass of a star or black hole grows by the value of the map energy E of a stone or collapsing bubble that falls into it. *The map energy of the stone is converted into gravitational mass.* Earlier we called map energy E "a unicorn, a mythical beast." Now we must admit that this unicorn can add its mass-equivalence to the mass of a star into which it falls.

?

Objection 4. *You checked equation (28) only in the Newtonian limit, where the remote dust bubble is at rest or falls inward with small kinetic energy. Is (28) valid for all values of E? Suppose that the dust bubble in Figure 4 is launched inward (or outward) at relativistic speed. In this case does total E still simply add to $M_{\rm star}$ to give total mass $M_{\rm total}$ for the still more distant observer?*

!

Yes it does, but we have not displayed the proof, which requires solution of Einstein's equations. Let a massive star collapse, then explode into a supernova. If this process is spherically symmetric, then a distant observer will detect no change in gravitational attraction in spite of the radical conversions among different forms of energy in the explosion. Indeed, the distant observer has no way to know about these transformations before the outward-blasting bubble of radiation and neutrinos passes her. As they pass, she detects a decline in the gravitational acceleration of the local test particle, because some of the original energy of the central attractor is carried to an r-value greater than hers.

Gravity waves carry off energy.

Is the Birkhoff restriction to spherical symmetry important? It can be: A satellite orbiting around or falling into a star or black hole will emit gravitational waves that carry away some energy, decreasing $M_{\rm total}$. Chapter 16 notes that a spherically symmetric distribution cannot emit gravitational waves, no matter how that spherical distribution pulses in or out. As a result, equation (28) is okay to use only when the emitted gravitational wave energy is very much less than $M_{\rm total}$. When that condition is met, the cases shown in Figures 3 and 4 are observationally indistinguishable.

Measuring E from far away.

As long as gravitational wave emission is negligible and we are sufficiently far away, we can, in principle, use (28) to measure the map energy E of *anything* circulating about, diving into, launching itself away from, or otherwise interacting with a center of attraction. Simply use Newtonian mechanics to carry out the measurement depicted in Figure 3, first with the satellite absent, second with the satellite in orbit near the star. Subtract the second value from the first for the acceleration (26) and use (28) to determine the value of $E = M_{\rm total} - M_{\rm star}$. As in Box 2, this example shows that E (and not $E_{\rm shell}$) includes effects of curved spacetime.

Box 4. Event Horizon vs. Particle Horizon

The *event horizon* around any black hole separates events that can affect the future of observers outside the event horizon from events that cannot do so. Barring quantum mechanics, the event horizon never reveals what is hidden behind it. (For a possible exception, see Box 5 on Hawking radiation.)

We can now define a black hole more carefully: *A black hole is a singularity cloaked by an event horizon.*

In Chapter 14 we learn about another kind of horizon, called a **particle horizon**. Some astronomical objects are so far from us that the light they have emitted since they were formed has not yet reached us. In principle more and more such objects swim into our distant field of view every day, as our cosmic particle horizon sweeps past them. In contrast to the event horizon, the particle horizon yields up its hidden information to us—gradually!

In order to avoid confusion among these different kinds of horizons, we try to be consistent in using the full name of the *event horizon* that cloaks a black hole.

6.6 ■ OVER THE EDGE: ENTERING THE BLACK HOLE

No jerk. No jolt. A hidden doom.

Except for the singularity at $r = 0$, no feature of the black hole excites more curiosity than the event horizon at $r = 2M$. It is the point of no return beyond which no traveler can find the way back—or even send a signal—to the outside world. What is it like to fall into a black hole? No one from Earth has yet experienced it. Moreover, we predict that future explorers who do so will not be able to return to report their experiences or to transmit messages about their experience to us outsiders—so we believe! In spite of the impossibility of receiving a final report, there exists a well-developed and increasingly well-verified body of theory that makes clear predictions about our experience as we approach and cross the event horizon of a black hole. Here are some of those predictions.

Predict what no one can verify.

We are not "sucked into" a black hole. Unless we get close to its event horizon, a black hole will no more grab us than our Sun grabs Earth. If our Sun should suddenly collapse into a black hole without expelling any mass, Earth and the other planets would continue on their courses undisturbed (even though, after eight minutes, continuous night would prevail for us on Earth!). The Schwarzschild solution (plus the Principle of Maximal Aging) would still continue to describe Earth's worldline around our Sun, just as it does now. In Section 6.7 you show that for an orbit at r-coordinate greater than about $300M$, Newtonian mechanics predicts gravitational acceleration with an accuracy of about 0.3 percent. We also find (Section 8.5) that no stable circular orbit is possible at r less than $6M$. Even at an r-value between $6M$ and the event horizon at $2M$, we can always escape the grip of the black hole, given sufficient rocket power. Only when we reach or cross the event horizon are we irrevocably swallowed, our fate sealed.

We are not sucked into a black hole.

We detect no special event as we fall inward through the event horizon. Even when we drop across the event horizon at $r = 2M$, we experience no shudder, jolt, or jar. True, tidal forces are ever-increasing as we fall inward, and this increase continues smoothly as we cross the event horizon.

No jolt as we cross the event horizon.

Box 5. Escape from the Black Hole? Hawking Radiation

Einstein's field equations predict that nothing, not even a light signal, escapes from inside the event horizon of a black hole. In 1973, Stephen Hawking demonstrated an exception to this conclusion using quantum mechanics. For years quantum mechanics had been known to predict that particle-antiparticle pairs—such as an electron and a positron—are continually being created and recombined in "empty" space, despite the frigidity of the vacuum. These processes have indirect, but significant and well-tested, observational consequences. Never in cold flat spacetime, however, do such events present themselves to direct observation. For this reason the pairs receive the name "virtual particles." When such a particle-antiparticle pair is produced near, but outside, the event horizon of a black hole, Hawking showed, one member of the pair will occasionally be swallowed by the black hole, while the other one escapes to a large r coordinate—now a *real* particle. Escaped particles form what is called **Hawking radiation.** Before particle emission, we had just the black hole; after particle emission, we have the black hole plus the distant real particle outside the event horizon. Where did the energy of this distant particle come from? In order to conserve mass/energy, the mass of the black hole must decrease in this process. This loss of mass causes the black hole to "evaporate." As the mass of the black hole decreases, the loss rate grows until eventually it becomes explosive, destroying the black hole. For a black hole of several solar masses, however, Hawking's theory predicts that the Earth-time required to achieve this explosive state exceeds the age of the Universe by a fantastic number of powers of ten. For this reason, we ignore Hawking radiation in our description of black holes.

We are not suddenly squashed or torn apart at $r = 2M$, because the event horizon is not a *physical* singularity, as explained in Box 3, Section 3.1. There is no sudden discontinuity in our experience as we pass through the event horizon.

No shell frames inside the event horizon.

Inside the event horizon no shell frames are possible. Outside the event horizon we have erected, in imagination, a set of nested spherical shells concentric to the black hole. We say "in imagination" because no known material is strong enough to withstand the "pull of gravity," which increases without limit as we approach the event horizon from outside (Section 6.7). Locally such a stationary shell can be replaced by a spaceship with rockets blasting in the inward direction to keep it at the same r and ϕ coordinates. Inside the event horizon, however, nothing remains at rest. No shell, no rocket ship can remain at constant r-coordinate there, however ferocious the blast of its engines. The material composing the original star, no matter how strong, was itself unable to resist the collapse that formed the black hole. The same irresistible collapse forbids any stationary structure or object inside the event horizon.

Packages can move inward, not outward.

"Outsiders" can send packages to "insiders." Inside the event horizon, different local frames can still move past one another with measurable relative speeds. Here are some examples. *First local frame:* One traveler may drop from rest just outside the event horizon. *Second local frame:* An unpowered spaceship may fall in from far away. *Third local frame:* Another unpowered spaceship may be hurled inward from outside the event horizon. Light and radio waves can carry messages inward to us. We who have fallen inside the event horizon can still see the stars, though with directions, colors, and intensities that change as we fall (Chapters 11 through 13). Packages and communications sent inward across the event horizon? Yes. How about moving

Box 6. Baked on the Shell?

As you stand on a spherical shell close to the event horizon of a black hole, you are crushed by an unsupportable local gravitational acceleration directed downward toward the center (Section 6.7). If that is not enough, you are also enveloped by an electromagnetic radiation field. William G. Unruh used quantum field theory to show that the temperature *T* of this radiation field (in degrees Kelvin) experienced on the shell is given by the equation

$$T = \frac{h g_{\rm conv}}{4\pi^2 k_{\rm B} c} \qquad (29)$$

Here $g_{\rm conv}$ is the local acceleration of gravity expressed in conventional units, meters/second^2; h is Planck's constant; c is the speed of light; and $k_{\rm B}$ is **Boltzmann's constant,** which has the value 1.381×10^{-23} kilogram-meters^2/(second^{2}degree Kelvin). The quantity $k_{\rm B}T$ has the unit joules and gives the average ambient thermal energy of this radiation field. (The same radiation field surrounds you when you accelerate at the rate $g_{\rm conv}$ in flat spacetime.)

Section 6.7 derives an expression for the local gravitational acceleration on a shell at r. Equation (46) gives the magnitude of this acceleration, expressed in the unit meter^{-1}:

$$g_{\rm shell} = \frac{g_{\rm conv}}{c^2} = \frac{M}{r^2}\left(1 - \frac{2M}{r}\right)^{-1/2} \qquad (30)$$

Substitute $g_{\rm conv}$ from (30) into (29) to obtain

$$T = \frac{hc}{4\pi^2 k_{\rm B}} \frac{M}{r^2}\left(1 - \frac{2M}{r}\right)^{-1/2} \qquad (31)$$

with M in meters. This temperature increases without limit as you approach the event horizon. Therefore one would expect the radiation field near the event horizon to shine brighter than any star when viewed by a distant observer. Why doesn't this happen? In a muted way it does happen. Remember that radiation is gravitationally red-shifted as it moves away from any center of attraction. Every frequency is red-shifted by the factor $(1 - 2M/r)^{1/2}$, which cancels the corresponding factor in (31). For radiation coming from near the event horizon, let $r \rightarrow 2M$ in the resulting equation. The distant viewer sees the radiation temperature

$$T_H = \frac{hc}{16\pi^2 k_{\rm B} M} \qquad \text{(distant view of event horizon)} \qquad (32)$$

with M in meters. The temperature $T_{\rm H}$ is called the **Hawking temperature** and characterizes the Hawking radiation from a black hole (Box 5). Notice that this temperature *increases* as the mass M of the black hole *decreases*. Even for a black hole whose mass is only a few times that of our Sun, this temperature is extremely low, so from far away such a black hole really looks *almost* black.

The radiation field described by equations (29) through (32), although perfectly normal, leads to strange conclusions. Perhaps the strangest is that this radiation goes entirely undetected by a free-fall observer. The free-fall diving traveler observes no such radiation field, while for the shell observer the radiation is a surrounding presence. This paradox cannot be resolved using the classical general relativity theory presentged in this book; see Kip Thorne's *Black Holes and Time Warps: Einstein's Outrageous Legacy*, page 444.

How realistic is the danger of being baked on a shell near the event horizon of a black hole? In answer, compute the local acceleration of gravity for a shell where the radiation field reaches a temperature equal to the freezing point of water, 273 degrees Kelvin. From (29) you can show that $g_{\rm conv} = 6.7 \times 10^{22}$ meters/second^2, or almost 10^{22} times the acceleration of gravity on Earth's surface. Evidently we will be crushed by gravity long before we are baked by radiation!

---

outward through the event horizon? No. Box 4 tells us—and Section 7.7 demonstrates—that when a diver fires a light flash radially outward at the instant she passes inward through the event horizon, that light flash hovers at the same r-coordinate at the event horizon. Nothing moves faster than light, so if light cannot move outward through the event horizon, then packages and stones definitely cannot move outward there either.

Surf a collapsing galaxy group.

Inside the event horizon life goes on—for a while. Make a daring dive into an already mature black hole? No. We and our exploration team want to be still more daring, to follow a black hole as it forms. We go to a multiple-galaxy system so crowded that it teeters on the verge of gravitational collapse. Soon after our arrival at the outskirts, it starts the collapse, at first slowly, then more and more rapidly. Soon a mighty avalanche thunders

Box 7. General relativity is a classical (non-quantum) theory.

Newton's laws describe the motion of a stone in flat spacetime at speeds very much less than the speed of light. For higher speeds we need relativity. Newton's laws correctly describe slow-speed motion of a "stone" more massive than, say, a proton. To describe behavior of smaller particles we need quantum physics.

Does this mean that we have no further use for Newton's laws of motion? Not at all! Newton's laws are *classical*, that is non-quantum. In this book we repeatedly use Newton's mechanics as a simple, intuitive first cut at prediction and observation. And with it we check every prediction of relativity in the limit of slow speed and vanishing spacetime curvature. We expect that Newton's laws of motion will be scientifically useful as long as humanity survives.

General relativity is also a *classical*—non-quantum—theory. General relativity does not predict Hawking radiation (Box 5) or the Hawking temperature (Box 6). These are predictions of quantum field theory, predictions that we mention as important asides to our classical analysis.

General relativity does not correctly represent every property of the black hole, any more than Newton's mechanics correctly predicts the motion of fast-moving particles. We still expect that general relativity—along with Newton's mechanics—will be scientifically useful during the long future of humanity.

(silently!) toward the center from all directions, an avalanche of objects and radiation, a cataract of momentum-energy-pressure. The matter of the galaxies and with it our group of enterprising explorers pass smoothly across the event horizon at Schwarzschild $r = 2M$.

"Publish *and* perish."

From that moment onward we lose all possibility of signaling to the outer world. However, radio messages from that outside world, light from familiar stars, and packages fired after us at sufficiently high shell speed continue to reach us. Moreover, communications among us explorers take place now as they did before we crossed the event horizon. We use the familiar categories of space and time to share our findings. With our laptop computers we turn out an exciting journal of observations, measurements, and conclusions. (Our motto: "Publish *and* perish.")

Killer tides.

Tides become lethal. Nothing rivets our attention more than the tidal forces that pull heads up and feet down with ever-increasing tension (Sections 1.11 and 7.9). Before much time has passed on our wristwatch, we can predict, this differential pull will reach the point where we can no longer survive. Moreover, we can foretell still further ahead and with certainty the instant of total crunch. That crunch swallows up not only the stars beneath us, not only we explorers, but time itself. All worldlines inside the event horizon terminate on the singularity. For us an instant comes after which there is no "after." Chapters 7 and 21 give more details of life inside the event horizon.

After crunch there is no "after."

6.7 ■ APPENDIX: INITIAL SHELL GRAVITATIONAL ACCELERATION FROM REST

Unlimited gravitational acceleration on a shell near the event horizon.

Is gravity real or fictitious?

When you stand on a shell near a black hole, you experience gravity—a pull downward—just as you do on Earth. On the shell this gravity can be great: near the event horizon it increases without limit, as we shall see. On the other hand, "In general relativity . . . gravity is *always* a fictitious force which we can eliminate by changing to a local frame that is in free fall . . ." (inside the back cover). So is this "gravity" real? Falls kill and injure many people every year. Anything that can kill you is definitely real, not fictitious! Here we avoid philosophical issues by asking a practical question: "When the shell observer drops a stone from rest, what *initial* acceleration does he measure?"

Practical experiment to define gravity

To begin, we behave like an engineer: Use a thought experiment to define what we mean by the initial gravitational acceleration of a stone dropped from rest on a shell at r_0. Use the heavy machinery of general relativity to find the magnitude of the newly-defined acceleration experienced by a shell observer.

Figure 5 presents our method to measure quantities used to define initial gravitational acceleration on a shell. The shell is at map r_0. At a shell distance $|\Delta y_{\rm shell}|$ below the shell lies a stationary platform onto which the shell observer drops a stone. The time lapse $\Delta t_{\rm shell}$ for the drop is measured as follows:

Specific instructions for experiment to define gravity

1. The shell observer starts his clock at the instant he drops the stone.
2. When the stone strikes the platform, it fires a laser flash upward to the shell clock.
3. The shell observer determines the shell time lapse between drop and impact, $\Delta t_{\rm shell}$, by deducting flash transit shell time from the time elapsed on his clock when he receives the laser flash.

To calculate the "flash transit shell time" in Step 3, the shell observer divides the shell distance $|\Delta y_{\rm shell}|$ by the shell speed of light. (In an exercise of Chapter 3, you verified that the shell observer measures light to move at its conventional speed—value one—in an inertial frame.)

Define $g_{\rm shell}$

The shell observer substitutes $\Delta y_{\rm shell}$ and $\Delta t_{\rm shell}$ into the expression that defines uniform acceleration $g_{\rm shell}$:

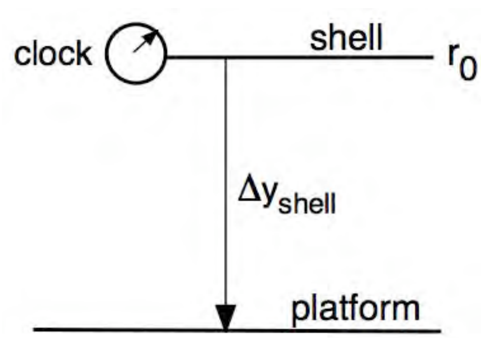


FIGURE 5 Notation for thought experiment to define initial gravitational acceleration from rest in a shell frame. The shell observer at r_0 releases a stone from rest and measures its shell time of fall $\Delta t_{\rm shell}$ onto a lower stationary platform that he measures to be a distance $|\Delta y_{\rm shell}|$ below the shell. From these observations he defines and calculates the value of the stone's initial acceleration $g_{\rm shell}$, equation (33).

$$\Delta y_{\text{shell}} = -\frac{1}{2} g_{\text{shell}} \Delta t_{\text{shell}}^2 \qquad (\text{uniform } g_{\text{shell}}) \tag{33}$$

Thus far our engineering definition of g_{shell} has little to do with general relativity. The fussy procedure of this thought experiment reflects the care required when general relativity is added to the analysis, which we do now.

Mapmaker demands constant map energy for falling stone.

What does the Schwarzschild mapmaker say about the acceleration of a dropped stone? She insists that, whatever motion the free stone executes, its map energy E/m must remain a constant of motion. So start with the map energy of a stone bolted to the shell at r_0. From map energy equation (15) with $v_{\text{shell}} = 0$ and $r = r_0$, we have:

$$\frac{E}{m} = \left(1 - \frac{2M}{r_0}\right)^{1/2} \qquad (\text{stone released from rest at } r_0) \tag{34}$$

Now release the stone from rest. The mapmaker insists that as the stone falls its map energy remains constant, so equate the right sides of (34) and (8), square the result, and solve for $d\tau^2$:

$$d\tau^2 = \left(1 - \frac{2M}{r_0}\right)^{-1} \left(1 - \frac{2M}{r}\right)^2 dt^2 \tag{35}$$

Substitute this expression for $d\tau^2$ into the Schwarzschild metric for radial motion ($d\phi = 0$), namely

$$d\tau^2 = \left(1 - \frac{2M}{r}\right) dt^2 - \left(1 - \frac{2M}{r}\right)^{-1} dr^2 \tag{36}$$

Divide corresponding sides of equations (36) and (35), then solve the resulting equation for $(dr/dt)^2$:

$$\left(\frac{dr}{dt}\right)^2 = \left(1 - \frac{2M}{r_0}\right)^{-1} \left(1 - \frac{2M}{r}\right)^2 \left(\frac{2M}{r} - \frac{2M}{r_0}\right) \qquad (\text{from rest at } r_0) \tag{37}$$

We want the acceleration of the stone in Schwarzschild map coordinates. Take the derivative of both sides with respect to the t-coordinate and cancel the common factor $2(dr/dt)$ from both sides of the result to obtain:

$$\frac{d^2 r}{dt^2} = -\left(\frac{M}{r^2}\right)\left(1 - \frac{2M}{r}\right)\left(1 - \frac{2M}{r_0}\right)^{-1}\left(\frac{4M}{r_0} + 1 - \frac{6M}{r}\right) \tag{38}$$

This equation gives the map acceleration at r of a stone released from rest at r_0. This acceleration depends on r, so is clearly *not* uniform as the stone falls, but *decreases* as r gets smaller, going to zero as r reaches the event horizon. We know that map acceleration is a unicorn, a result of Schwarzschild map coordinates, not measured by any inertial observer. We are interested in the *initial* acceleration at the instant of release from rest. Set $r = r_0$ in equation (38), which then reduces to the relatively simple form:

$$\left(\frac{d^2r}{dt^2}\right)_{r_0} = -\frac{M}{r_0^2}\left(1 - \frac{2M}{r_0}\right) \qquad \text{(initial, from rest at } r_0\text{)} \tag{39}$$

Acceleration in map coordinates

What is the meaning of this acceleration in Schwarzschild map coordinates? It is only a spreadsheet entry, an accounting analysis by the mapmaker, not the result of a direct observation by anyone. Observation requires an experiment on the shell, which we have already designed, leading to the expression (33). What is the relation between our engineering definition of acceleration and acceleration (39) in Schwarzschild coordinates? To compare the two expressions, expand the Schwarzschild r-coordinate of the dropped stone close to the radial position r_0 using a Taylor series for a short lapse Δt:

$$r = r_0 + \left(\frac{dr}{dt}\right)_{r_0}\Delta t + \frac{1}{2}\left(\frac{d^2r}{dt^2}\right)_{r_0}(\Delta t)^2 + \frac{1}{6}\left(\frac{d^3r}{dt^3}\right)_{r_0}(\Delta t)^3 + ... \tag{40}$$

Because Δt is small, we disregard terms higher than quadratic in Δt. This allows us to approximate uniform gravity (constant acceleration) and to compare mapmaker accounting entries with observed shell acceleration. Since the stone drops from rest at r_0, the initial map speed is zero: $(dr/dt)_{r_0} = 0$. With these considerations, insert (39) into (40) and obtain:

$$r - r_0 = \Delta r \approx -\frac{1}{2}\left[\left(1 - \frac{2M}{r_0}\right)\frac{M}{r_0^2}\right](\Delta t)^2 \tag{41}$$

This equation has a form similar to that of our experimental definition (33) of shell gravitational acceleration, except the earlier equation employs vertical shell separation Δy_{shell} and shell time lapse Δt_{shell}. Convert these to Schwarzschild quantities using standard transformations—equations (5.8) and (5.9):

$$\Delta y_{\text{shell}} = \left(1 - \frac{2M}{r_0}\right)^{-1/2}\Delta r \quad \text{and} \quad \Delta t_{\text{shell}}^2 = \left(1 - \frac{2M}{r_0}\right)(\Delta t)^2 \tag{44}$$

With these substitutions, and after rearranging terms, equation (33) becomes:

$$\Delta r = -\frac{1}{2}\left[\left(1 - \frac{2M}{r_0}\right)^{3/2} g_{\text{shell}}\right](\Delta t)^2 \tag{45}$$

As we go to the limit $\Delta t \to 0$, the extra terms in (40) become increasingly negligible, so (41) approaches an equality and we can equate square-bracket expressions in (41) and (45). Replacing the notation r_0 with r yields the magnitude of the initial acceleration of a stone dropped from rest on a shell at any r-coordinate:

Initial shell acceleration

Sample Problems 2. Initial Gravitational Acceleration on a Shell

1. On a shell at $r/M = 4$ near a black hole, the initial gravitational acceleration from rest is how many times that predicted by Newton?
2. On a shell at $r/M = 2.1$ near a black hole, the initial gravitational acceleration is how many times that predicted by Newton?
3. What is the minimum value of r/M so that, at or outside of that r-coordinate, Newton's formula for gravitational acceleration yields values that differ from Einstein's by less than ten percent? by less than one percent?
4. Compute the weight in pounds of a 100-kilogram astronaut on the surface of a neutron star with mass equal to $1.4M_{\text{Sun}}$ and $M/r_0 = 2/5$.

SOLUTIONS

1. At $r/M = 4$ the factor $(1 - 2M/r)^{-1/2}$ in (46) predicts a gravitational acceleration $2^{1/2} = 1.41$ times that predicted by Newton.
2. Even at $r/M = 2.1$ the gravitational acceleration is still the relatively mild multiple of 4.6 times the Newtonian prediction.
3. Setting $(1 - 2M/r)^{-1/2} = 1.1$ yields $r/M = 11.5$. At or outside this r-coordinate, Newton's prediction will be in error (it will be too low) by less than ten percent. At or outside $r/M = 100$ Newton's prediction will be too low by less than one percent.
4. The Newtonian acceleration in conventional units is:

$$g_{\text{Newton conv}} = \left(\frac{GM_{\text{kg}}}{c^2 r_0^2}\right) c^2 = \left(\frac{M}{r_0^2}\right) c^2 \qquad (42)$$
$$= \left(\frac{M}{r_0}\right)^2 \frac{c^2}{M} = \left(\frac{2}{5}\right)^2 \frac{c^2}{1.4 \times M_{\text{Sun}}}$$

Insert values of c^2 and M_{Sun} (in meters) to yield $g_{\text{Newton conv}} \approx 7.0 \times 10^{12}$ meters/second^2. From (46),

$$\text{weight} = mg_{\text{shell}} = \left(1 - \frac{4}{5}\right)^{-1/2} mg_{\text{Newton}} \qquad (43)$$
$$\approx 16 \times 10^{14} \text{ Newtons}$$

One Newton = 0.225 pounds, so our astronaut weighs approximately 3.5×10^{14} pounds, or 350 trillion pounds (USA measure of weight). It is surprising that, even at the surface of this neutron star, the general relativity result in (43) is greater than Newton's by the rather small factor $5^{1/2} = 2.24$.

$$g_{\text{shell}} = \left(1 - \frac{2M}{r}\right)^{-1/2} \frac{M}{r^2} \qquad \text{(initial, drop from rest)} \qquad (46)$$

Sample Problems 2 explore initial shell accelerations under different conditions. It is surprising how accurate Newton's expression $g_{\text{Newton}} = M/r^2$ is even quite close to the event horizon of a black hole—an intellectual victory for Newton that we could hardly have anticipated.

QUERY 1. Gravitational acceleration on Earth's surface
Use values for the constants M_{E} and r_{E} for the Earth listed inside the front cover to show that equation (46) correctly predicts the value of the gravitational acceleration g_{E} at Earth's surface. Check your calculated values against those also listed inside the front cover.

A. Show that in units of length this acceleration has the value $g_{\text{E}} = 1.09 \times 10^{-16}$ meter^{-1}.
B. Show that in conventional units this acceleration has the value $g_{\text{E,conv}} = 9.81$ meters/second^2.

A GRAVITYLESS DAY

I am sitting here 93 million miles from the sun on a rounded rock which is spinning at the rate of 1,000 miles an hour, and roaring through space to nobody-knows-where, to keep a rendezvous with nobody-knows-what . . . and my head pointing down into space with nothing between me and infinity but something called gravity which I can't even understand, and which you can't even buy anyplace so as to have some stored away for a gravityless day . . .

—Russell Baker

6.8 ■ EXERCISES

1. Schwarzschild Metric at Earth's Surface

The Newtonian formula for gravitational acceleration for a non-rotating ("stat") Earth: $g_{\text{Newton,stat}} = M/r_{\text{Earth}}^2 = 1.09 \times 10^{-16}$ meter^{-1} with $r_{\text{Earth}} = 6.38 \times 10^6$ meters.

A. Use Newton's formula for the centrifugal pseudo-force ("cf") at the equator $F_{\text{cf}} = mr_{\text{Earth}}\omega^2$, to calculate the Newtonian correction factor to $g_{\text{Newton,stat}}$ at the equator due to Earth's rotation.

B. Use (46) and assume a non-rotating Earth to calculate the Schwarzschild correction factor to $g_{\text{Newton,stat}}$.

C. Which of the two correction factors – centrifugal force or Schwarzschild correction – is larger and by how many orders of magnitude?

2. Diving from Rest Far Away

Black Hole Alpha has a mass $M = 10$ kilometers. A stone starting from rest far away falls radially into this black hole. In the following, express all speeds as a decimal fraction of the speed of light.

A. What is the speed of the stone measured by the shell observer at $r = 50$ kilometers?

B. Write down an expression for $|dr/dt|$ of the stone as it passes $r = 50$ kilometers?

C. What is the speed of the stone measured by the shell observer at $r = 25$ kilometers?

D. Write down an expression for $|dr/dt|$ of the stone as it passes $r = 25$ kilometers?

E. In two or three sentences, explain why the change in the speed between Parts A and C is qualitatively different from the change in $|dr/dt|$ between Parts B and D.

3. Maximum Raindrop $|dr/dt|$

A stone is released from rest far from a black hole of mass M. The stone drops radially inward. Mapmaker records show that the the value of $|dr/dt|$ of the stone initially increases but declines toward zero as the stone approaches the event horizon. The value of $|dr/dt|$ must therefore reach a maximum at some intermediate r. Find this r-value for this maximum. Find the numerical value of $|dr/dt|$ at that r-value. Who measures this value?

4. Hitting a Neutron Star

A particular nonrotating neutron star has a mass $M = 1.4$ times the mass of our Sun and $r = 10$ kilometers. A stone starting from rest far away falls onto the surface of this neutron star.

A. If this neutron star were a black hole, what would be the map r-value of its event horizon? What fraction is this of the r-value of the neutron star?
B. With what speed does the stone hit the surface of the neutron star as measured by someone standing (!) on the surface?
C. With what value of $|dr/dt|$ does the stone hit the surface?
D. With what kinetic energy per unit mass does the stone hit the surface according to the surface observer?

Earlier it was thought that astronomical gamma-ray bursts might be caused by stones (asteroids) impacting neutron stars. Carry out a preliminary analysis of this hypothesis by assuming that the stone is made of iron. The impact kinetic energy is very much greater than the binding energy of iron atoms in the stone, greater than the energy needed to completely remove all 26 electrons from each iron atom, and greater even than the energy needed to shatter the iron nucleus into its component 26 protons and 30 neutrons. So we neglect all these binding energies in our estimate. The result is a vaporized gas of 26 electrons and 56 nucleons (protons and neutrons) per incident iron atom. We want to find the average energy of photons (gamma rays) emitted by this gas.

E. Explain briefly why, just after impact, the electrons have very much less kinetic energy than the nucleons. So in what follows we neglect the initial kinetic energy of the electron gas just after impact.
F. The hot gas emits thermal radiation with characteristic photon energy approximately equal to the temperature. What is the characteristic energy of photons reaching a distant observer, in MeV?

NOTE: It is now understood that astronomical gamma-ray bursts release much more energy than an asteroid falling onto a neutron star. Gamma ray bursts are now thought to arise from the birth of new black holes in distant galaxies.

5. A Stone Glued to the Shell Breaks Loose

A stone of mass m glued to a shell at r_0 has map energy given by equation (34). Later the glue fails so that the stone works loose and drops to the center of the black hole of mass M.

A. By what amount ΔM does the mass of the black hole increase?
B. A distant observer measures the mass of black hole plus stone at rest at r_0 using the method of Figure 3. How will the value of this total mass change after the stone has fallen into the black hole?
C. Apply your result of Part A to find the numerical value of the constant K in the equation $\Delta M = Km$ for the three cases: (a) $r_0 \gg 2M$, (b) $r_0 = 8M$ and (c) r_0 is just outside the event horizon. In all cases the observer in Figure 3 is much farther away than r_0.

6. Wristwatch Time to the Center

An astronaut drops from rest off a shell at r_0. How long a time elapses, as measured on her wristwatch, between letting go and arriving at the center of the black hole? If she drops off the shell just outside the event horizon, what is her event-horizon-to-crunch wristwatch time?
Several hints: The first goal is to find $dr/d\tau$, the rate of change of r-coordinate with wristwatch time τ, in terms of r and r_0. Then form an integral whose variable of integration is r/r_0. The limits of integration are from $r/r_0 = 1$ (the release point) to $r/r_0 = 0$ (the center of the black hole). The integral is

$$\frac{\tau}{M} = -\frac{1}{2^{1/2}}\left(\frac{r_0}{M}\right)^{3/2}\int_1^0 \frac{(r/r_0)^{1/2}\,d\,(r/r_0)}{(1-r/r_0)^{1/2}} \tag{47}$$

Solve this integral using tricks, nothing but tricks: Simplify by making the substitution $r/r_0 = \cos^2\psi$ (The “angle” ψ is not measured anywhere; it is simply a variable of integration.) Then $(1 - r/r_0)^{1/2} = \sin\psi$ and $d(r/r_0) = -2\cos\psi\sin\psi\,d\psi$ The limits of integration are from $\psi = 0$ to $\psi = \pi/2$. With these substitutions, the integral for wristwatch time becomes

$$\frac{\tau}{M} = 2^{1/2}\left(\frac{r_0}{M}\right)^{3/2}\int_0^{\pi/2} \cos^2\psi d\psi \tag{48}$$
$$= 2^{1/2}\left(\frac{r_0}{M}\right)^{3/2}\left[\frac{\psi}{2} + \frac{\sin 2\psi}{4}\right]\Bigg|_0^{\pi/2}$$

Both sides of (48) are unitless. Complete the formal solution. For a black hole 20 times the mass of our Sun, how many seconds of wristwatch time elapse between the drop from rest just outside the event horizon to the singularity?

7. Release a stone from rest

You release a stone from rest on a shell of map coordinate r_0.

A. Derive an expression for $|dr/dt|$ of the stone as a function of r. Show that when the stone drops from rest far away, $|dr/dt|$ reduces to the expression (22) for a raindrop. Find the r-value at which map speed is *maximum* and the expression for that maximum map speed. Verify that in the limit in which the stone is dropped from rest far away, these expressions reduce to those found in Exercise 6.2 for the raindrop.

B. Derive an expression for the *shell velocity* of the stone as a function of r. Show that in the limit in which the stone drops from rest far away, the shell velocity reduces to the expression (19) for a raindrop.

C. Sketch graphs of shell speed *vs.* r similar to Figure 2 for the following values of r_0:

   (a) $r_0/M = 10$
   (b) $r_0/M = 6$
   (c) $r_0/M = 3$

8. Hurl a stone inward from far away

You hurl a stone radially inward with speed $v_{\rm far}$ from a remote location. (At a remote r where spacetime is flat, $|dr/dt|$ equals shell speed.)

A. Derive an expression for dr/dt of the stone as a function of r. Show that when you launch the stone from rest, dr/dt reduces to the expression (22) for a raindrop. Find the value of r at which $|dr/dt|$ is *maximum* and the expression for $|dr/dt|$. Verify that in the limit in which the stone is dropped from rest far away, these expressions reduce to those found in Exercise 6.2 for the raindrop.

B. Derive an expression for the *shell velocity* of the stone as a function of r. Show that in the limit in which the stone drops from rest far away, the shell velocity reduces to the expression (19) for a raindrop.

C. Sketch graphs of shell speed *vs.* r similar to Figure 2 for the following values of $v_{\rm far}$:

   (a) $v_{\rm far} = 0.20$
   (b) $v_{\rm far} = 0.60$
   (c) $v_{\rm far} = 0.90$

9. All Possible Shell Speeds

Think of a shell observer at any $r > 2M$. Consider the following three launch methods for a stone that passes him moving radially inward:(a) released at rest from a shell at $r_0 \geq r$, (b) released from rest far away, and (c) hurled radially inward from far away with initial speed $0 < |v_{\rm far}| < 1$. Show that, taken together, these three methods can result in all possible speeds $0 \leq |v_{\rm shell}| < 1$ measured by this shell observer at $r > 2M$.

10. Only One Shell Speed—with the Value One—at the Event Horizon

Show that the three kinds of radial launch of a stone described in Exercise 8 yield the *same* shell speed, namely $|v_{\rm shell}| = 1$, as a limiting case when the stone moves inward across the event horizon. Your result shows that at the event horizon (as a limiting case): (a) You cannot make the shell-observed speed of a stone *greater* than that of light, no matter how fast you hurl it inward from far away. (b) You cannot make the shell-observed speed of the stone *less* than that of light, no matter how close to the event horizon you release it from rest.

11. Energy from garbage using a black hole

Define an **advanced civilization** as one that can carry out any engineering task not forbidden by the laws of physics. An advanced civilization wants to use a black hole as an energy source. Most useful is a "live" black hole, one that spins (Chapters 17 through 21), with rotation energy available for use. Unfortunately the nonrotating black hole that we study in this chapter is "dead:" no energy can be extracted from it (except for entirely negligible Hawking radiation, Box 5). Instead, our advanced civilization uses the dead (nonspinning) black hole to convert garbage to useful energy, as you analyze in this exercise.

A bag of garbage of mass m drops from rest at a power station located at r_0, onto a shell at r; a machine at the lower r brings the garbage to rest and converts all of the *shell kinetic energy* into a light flash. Express all energies requested below as fractions of the mass m of the garbage.

A. What is the energy of the light flash measured on the shell where it is emitted?

B. The machine now directs the resulting flash of light radially outward. What is the energy of this flash as it arrives back at the power station?

C. Now the conversion machine at r releases the garbage so that it falls into the black hole. What is the increase ΔM in the mass of the black hole? What is its increase in mass if the conversion machine is located—as a limiting case—exactly at the event horizon?

D. Find an expression for the efficiency of the resulting energy conversion, that is (output energy at the power station)/(input garbage mass m) as a function of the converter r and the r_0 of the power station. What is the efficiency when the power station is far from the black hole, $r_0 \rightarrow \infty$, and the conversion machine is on the shell at $r = 3M$? (Except for matter-antimatter collisions, the efficiency of mass-to-energy conversions in nuclear reactions on Earth is never greater than a fraction of one percent.)

E. *Optional:* Check the conservation of *map* energy in all of the processes analyzed in this exercise.

Comment 2. Decrease disorder with a black hole vacuum cleaner?
Suppose that the neighborhood of a black hole is strewn with garbage. We tidy up the vicinity by dumping the garbage into the black hole. This cleanup reduces disorder in the surroundings of the black hole. But wait! Powerful principles of thermodynamics and statistical mechanics demand that the disorder—technical name: **entropy**—of an isolated system (in this case, garbage plus black hole) cannot decrease. Therefore the disorder of the black hole itself must increase when we dump disordered garbage into it. Jacob Bekenstein and Stephen Hawking quantified this argument to define a measure of the entropy of a black hole, which turns out to be proportional to the Euclidean-calculated spherical "area" of the event horizon. See Kip S. Thorne, *Black Holes and Time Warps*, pages 422–448.

12. Temperature of a Black Hole

A Use equation (32) to find the temperature, when viewed from far away, of a black hole of mass five times the mass of our Sun.

B. What is the mass of a black hole whose temperature, viewed from far away, is 1800 degrees Kelvin (the melting temperature of iron)? Express your answer as a fraction or multiple of the mass of Earth. (Equation (32) tells us that "smaller is hotter," which leads to increased emission by a smaller black hole and therefore shorter life. If this analysis is correct, small black holes created in the Big Bang must have evaporated by now.)

6.9 ■ REFERENCES

Initial quote David Kaiser, personal communication.

This chapter owes a large intellectual debt in ideas, figures, and text to *Gravitation* by Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler, W. H. Freeman and Company, San Francisco (now New York), 1973. Other quotations from *A Journey into Gravity and Spacetime* by John Archibald Wheeler, W. H. Freeman and Company, New York, 1990. In addition, our treatment was helped by reference to "Nonrotating and Slowly Rotating Holes" by Douglas A. Macdonald, Richard H. Price, Wai–Mo Suen, and Kip S. Thorne in the book *Black Holes: The Membrane Paradigm*, edited by Kip S. Thorne, Richard H. Price, and Douglas A. Macdonald, Yale University Press, New Haven, 1986.

More advanced treatment of the Principle of Maximal Aging in Misner, Thorne, and Wheeler, page 315 and following.

Newton quotation in Box 1: I. Bernard Cohen and Anne Whitman, *Isaac Newton: The Principia*, 1999, University of California Press, page 408

References for Box 6, "Baked on the Shell?" Historical background in Kip S. Thorne, Richard H. Price, and Douglas A. MacDonald *Black Holes: The Membrane Paradigm*, Yale University Press, 1986, pages 280–285 and 35–36; W. G. Unruh, "Notes on Black-Hole Evaporation", Physical Review D, Volume 14, Number 4, 15 August 1976, pages 870–892; William G. Unruh and Robert M. Wald, "What Happens When an Accelerating Observer Detects a Rindler Particle," Physical Review D, Volume 29, Number 6, 15 March 1984, pages 1047–1056.

Russell Baker quotation excerpted from *New York Times,* May 18, 1975, *New York Times Magazine*, page 6. Reprinted in *A Geological Miscellany* by E. J. Jones and G. Y. Craig, editors, 1985, Princeton University Press, Princeton, pages 191-192.

Download File Name: Ch06Diving170329v1.pdf

Chapter 7. Inside the Black Hole

- *Why would anyone volunteer to dive to the center of a black hole?*
- *Why does everything inside the event horizon inevitably move to smaller r?*
- *How massive must a black hole be so that 20 years pass on my wristwatch between crossing the event horizon and arrival at the crunch point?*
- *How can I construct a local inertial frame that is valid inside the event horizon?*
- *What do I see ahead of me and behind me as I approach the crunch point?*
- *Is my death quick and painless?*

Download File Name: Ch07InsideTheBlackHole170511v1.pdf

CHAPTER 7

Inside the Black Hole

Edmund Bertschinger & Edwin F. Taylor *

Alice had not a moment to think about stopping herself before she found herself falling down what seemed to be a very deep well. 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 . . . So many out-of-the-way things had happened lately that Alice had begun to think that very few things indeed were really impossible.

—Lewis Carroll, *Alice in Wonderland*

7.1 ■ INTERVIEW OF A DIVING CANDIDATE

Few things are really impossible.

So you are applying to be a member of the black hole diving research group.

Yes.

Have you personally had experience diving into black holes?

This question is a joke, right?

Why do you want to be part of this diving group, since your research results cannot be reported back to us outside the event horizon?

We want to see for ourselves whether or not our carefully-studied predictions are correct. You know very well that 27 percent of qualified Galaxy Fleet personnel volunteered for this mission.

Tell me, why doesn't the black-hole diving group use local shell coordinates to make measurements inside the event horizon?

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

Inside the event horizon no one can build a spherical shell that stays at constant r in Schwarzschild coordinates. Instead we make measurements in a series of local inertial frames that can exist anywhere except on the singularity.

Then how will you measure your r-coordinate without a spherical shell?

We track the decreasing value of r by measuring the decreasing separation between us and a test particle beside us that is also diving radially inward.

What clocks will you use in your experiments?

Our wristwatches.

When does your diving group cross the event horizon?

As measured on whose clock?

You *are* savvy. When does your diving group cross the event horizon as read on your wristwatches?

Zeroing different clocks in different locations is arbitrary. The Astronautics Commission has a fancy scheme for coordinating the various clock readings, mostly for convenience in scheduling. Want more details?

Not now. Is there any service that we on the outside can provide for your diving group once you are inside the event horizon?

Sure. We will welcome radio and video bulletins of the latest news plus reports of scientific developments outside the event horizon.

And will your outgoing radio transmissions from inside the event horizon change frequency during their upward transit to us?

Another joke, I see.

Yes. Does your personal—ah—end seem mercifully quick to you?

The terminal "spaghettification" will take place in a fraction of a second as recorded on my wristwatch. Many of you outside the event horizon would welcome assurance of such a quick end.

What will you personally do for relaxation during the trip?

I am a zero-g football champion and grandmaster chess player. Also, my fiancé has already been selected as part of the team. We will be married before launch.

[We suppress the transcript of further discussion about the ethical and moral status of bringing children into the diving world.]

Box 1. Eggbeater Spacetime?

Being "spaghettified" as you approach the center of a black hole is bad enough. But according to some calculations, your atoms will be scrambled by violent, chaotic tidal forces before you reach the center—especially if you fall into a young black hole.

The first theory of the creation of a black hole by J. Robert Oppenheimer and Hartland Snyder (1939) assumed that the collapsing structure is spherically symmetric. Their result is a black hole that settles quickly into a placid final state. A diver who approaches the singularity at the center of the Oppenheimer-Snyder black hole is stretched with steadily increasing force along the r-direction and compressed steadily and increasingly from all sides perpendicular to the r-direction.

In Nature an astronomical collapse is rarely spherically symmetric. Theory shows that when a black hole forms, the asymmetries exterior to the event horizon are quickly radiated away in the form of gravitational waves—in a few seconds measured on a distant clock! Gravitational radiation captured inside the event horizon, however, evolves and influences spacetime inside the black hole.

So what happens? There is no way to verify any predictions about events inside the event horizon (Objection 1), but that does not stop us from making them! Vladimir Belinsky, Isaac Markovich Khalatnikov, Evgeny Mikhailovich Lifshitz, and independently Charles Misner discovered that Einstein's equations predict more than one kind of singularity.

Their theory says that as a diving observer approaches the center point, spacetime can oscillate chaotically, squeezing and stretching the poor traveler in random directions like an electric mixer (eggbeater). These oscillations increase in both amplitude and frequency as the astronaut approaches the singularity of the black hole. Any physical object, no matter what stresses it can endure, is necessarily utterly destroyed at an eggbeater singularity.

However, there is some theoretical evidence that eggbeater oscillations will die away, so an astronaut who waits a while to dive after the black hole has formed may not encounter them. Before these eggbeater oscillations die away—if they do—spacetime in the chaotic regions is definitely NOT described by the Schwarzschild metric!

In the present chapter we assume the non-spinning black hole under exploration is an ancient one and that we can ignore possible eggbeater oscillations of spacetime. We predict (and hope!) that as our astronaut colony approaches the center, the "spacetime weather" is clear and calm.

Objection 1. *What kind of science are these people talking about? Obviously nothing more than science fiction! No one who crosses the event horizon of a black hole can report observations to the scientific community outside the event horizon. Therefore all observations carried out inside the event horizon—and conclusions drawn from them—remain private communications. Private communication is not science!*

Yours is one sensible view of science, but if the "spacetime weather is clear and calm" inside the event horizon (Box 1), then the diving research group may have decades of life ahead of them, as recorded on their wristwatches. They can receive news and science updates from friends outside, view the ever-changing pattern of stars in the heavens (Chapter 12), carry out investigations, discuss observations among themselves, and publish their own exciting research journal.

We recognize that the event horizon separates two communities of investigators with a one-way surface or "membrane." Outsiders cannot receive reports of experiments that test their predictions about life inside the event horizon. They must leave it to insiders to verify or disprove these predictions with all the rigor of a lively in-falling research community. Later chapters on the *spinning* black hole raise the possibility that an explorer might navigate in such a way as to reemerge from the event horizon, possibly into a different spacetime region.

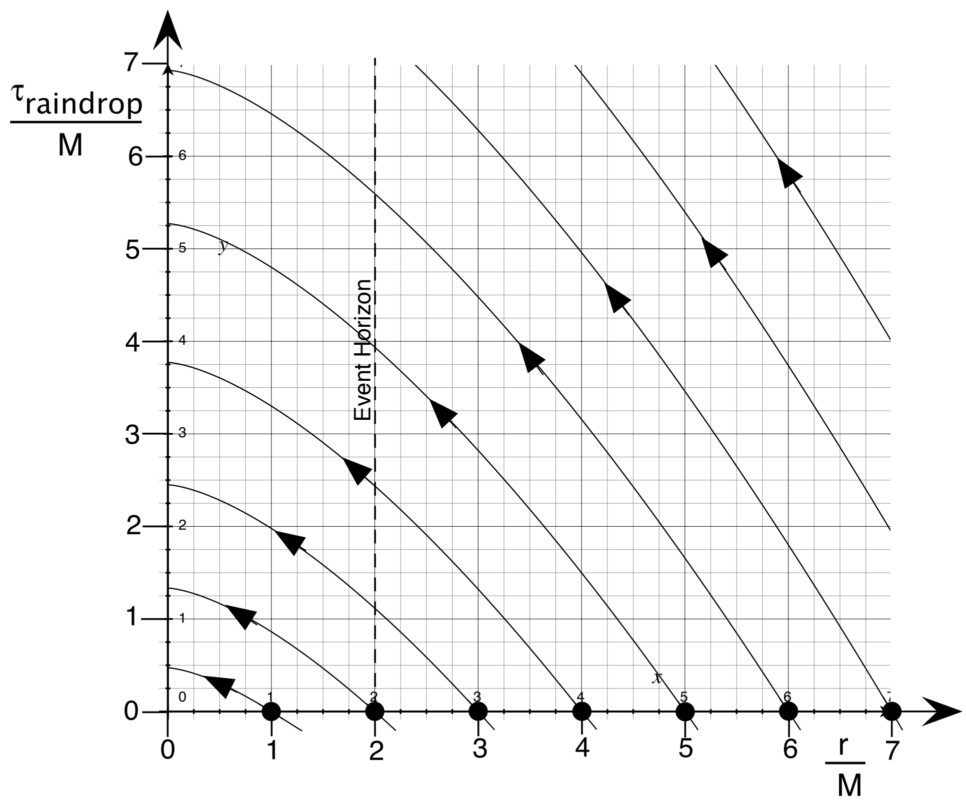


FIGURE 1 Raindrop wristwatch time vs. global r-coordinate from (2) for a series of raindrops that pass $r = M, 2M, ..., 9M$ at $\tau_{\rm raindrop} = 0$ (little filled circles along the horizontal axis). The shapes of these curves are identical, just displaced vertically with respect to one another. Every raindrop moves smoothly across the event horizon when clocked on its wristwatch, but not when tracked with global Schwarzschild t-coordinate (Figure 2).

John A. Wheeler: radical conservatism

Comment 1. Wheeler's "radical conservatism"
John Archibald Wheeler (1911-2008), who co-author the first edition of *Exploring Black Holes*, rescued general relativity from obscurity in the 1950s and helped to jump-start the present golden age of gravitational physics. He was immensely inventive in research and teaching; for example he adopted and publicized the name *black hole* (initial quote, Chapter 3). Wheeler's professional philosophy was **radical conservatism**, which we express as: *Follow well-established physical principles while pushing each to its extreme limits. Then develop a new intuition!* The black hole—both outside and inside its event horizon—is a perfect structure on which to apply Wheeler's radical conservatism, as we do throughout this book.

Comment 2. Non-spinning vs. spinning black hole
Chapters 2 through 13 describe spacetime around a black hole: a black hole that does not rotate. We call this a **non-spinning black hole**. The Universe is full of black holes that spin; many of them spin very fast, with deep

consequences for their structure and for spacetime around them. We call each of these a **spinning black hole**, the subject of Chapters 17 through 21.

7.2 ■ RAINDROP WORLDLINE

"Raindrops keep fallin' on my head . . ." song by Hal David and Burt Bacharach

We start with the raindrop of Chapter 6. The *raindrop* is a stone (wearing a wristwatch) that drops from initial rest very far from the black hole (Definition 2, Section 6.4). From equation (23) of that section:

$$d\tau_{\text{raindrop}} = -\left(\frac{r}{2M}\right)^{1/2} dr \tag{1}$$

In the following Queries you integrate (1) and apply the result to a raindrop that first falls past an Above r-coordinates r_{A} then falls past a sequence of lower r-coordinates (Figure 1).

QUERY 1. Raindrop wristwatch time lapse between the above r_{A} and lower r

A. Integrate (1) to determine the elapsed raindrop wristwatch time from the instant the raindrop falls past the Above coordinate r_{A} until it passes a sequence of smaller r-coordinates. Express this elapsed raindrop wristwatch time with the notation $[r_{\text{A}} \rightarrow r]$ for r-limits.

$$\tau_{\text{raindrop}}\left[r_{\text{A}} \rightarrow r\right] = \frac{4M}{3}\left[\left(\frac{r_{\text{A}}}{2M}\right)^{3/2} - \left(\frac{r}{2M}\right)^{3/2}\right] \tag{2}$$

Figure 1 plots this equation for a series of raindrops after each passes through a different given r_{A} at $\tau_{\text{raindrop}} = 0$.

B. What happens to the value of the raindrop wristwatch time lapse in (2) when the initial r_{A} becomes very large? Explain why you are not disturbed by this result.

QUERY 2. Raindrop wristwatch time lapse from event horizon to crunch.
Suppose you ride the raindrop, and assume (incorrectly) that you survive to reach the center. This Query examines how long (on your wristwatch) it takes you to drop from the event horizon to the singularity—the crunch point.

A. Adapt your result from Query 1 to show that:

$$\tau_{\text{raindrop}}[2M \rightarrow 0] = \frac{4M}{3} \qquad \text{(event horizon to crunch, in meters)} \tag{3}$$

B. *Predict:* Does every curve in Figure 1 satisfy (3)?

C. Use constants inside the front cover to find the event horizon-to-crunch raindrop wristwatch time *in seconds* for a black hole of mass M/M_{Sun} times the mass of our Sun:

$$\tau_{\text{raindrop}}[2M \rightarrow 0] = 6.57 \times 10^{-6}\frac{M}{M_{\text{Sun}}} \qquad \text{(event horizon to crunch, in seconds)} \tag{4}$$

D The monster black hole at the center of our galaxy has mass $M \approx 4 \times 10^6 M_{\text{Sun}}$: its mass is about four million times the mass of our Sun. Assume (incorrectly) that this black hole does not spin. How long, in seconds on your wristwatch, will it take you—riding on the raindrop—to fall from the event horizon of this monster black hole to its singularity?

E. **Discussion question:** How can the r-value of the event horizon $r = 2M$ possibly be greater than the wristwatch time $4M/3$ that it takes the raindrop to fall from the event horizon to the singularity? Does the raindrop move faster than light inside the event horizon? (*Hint:* Do global coordinate separations dependably predict results of our measurements?)

QUERY 3. Mass of the "20-year black hole."

The black hole we feature in this chapter has a mass such that it takes 20 years—recorded on the wristwatch of the raindrop—to fall from event horizon to singularity.

A. Find the mass of the "20-year black hole" (a) in meters, (b) as a multiple of the mass of our Sun, and (c) in light-years.

B. An average galaxy holds something like 10^{11} stars of mass approximately equal to that of our Sun. The "20-year black hole" has the mass of approximately how many average galaxies?

C. What is the value of the r-coordinate at the event horizon of the "20-year black hole" in light-years?

Now turn attention to the motion of the raindrop in global Schwarzschild coordinates. Equation (22) in Section 6.4 gives the Schwarzschild map velocity of the raindrop:

$$\begin{aligned}\frac{dr}{dt} &= -\left(1 - \frac{2M}{r}\right)\left(\frac{2M}{r}\right)^{1/2} \qquad \text{(raindrop map velocity)} \\ &= \left(\frac{r}{2M}\right)^{-3/2}\left(1 - \frac{r}{2M}\right)\end{aligned} \tag{5}$$

We want to find $r(t)$, the r-coordinate of the raindrop as a function of the t-coordinate. This function defines a worldline (Section 3.10).

To find the worldline of the raindrop, manipulate (5) to read:

$$dt = \frac{\left(\dfrac{r}{2M}\right)^{3/2} dr}{1 - \dfrac{r}{2M}} = \frac{4Mu^4 du}{1 - u^2} \qquad \text{(raindrop)} \tag{6}$$

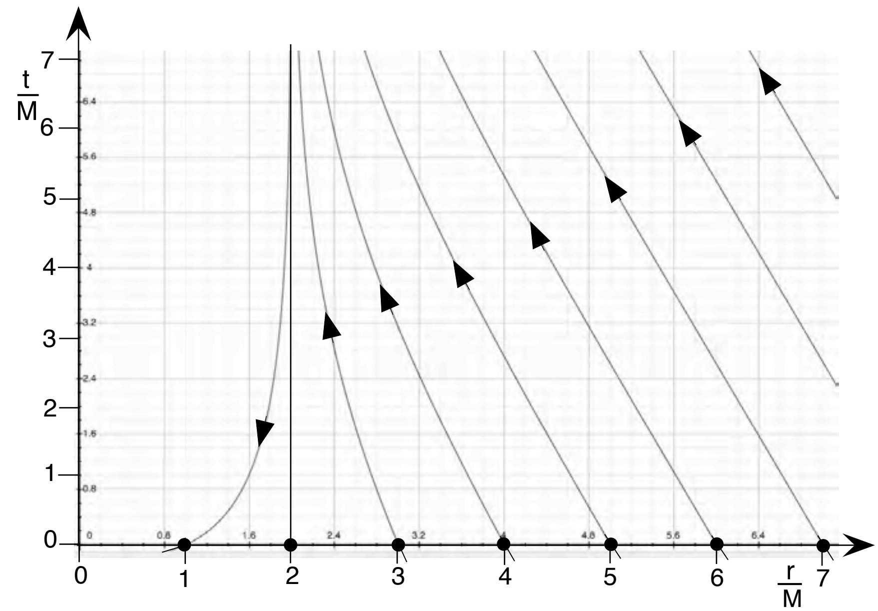


FIGURE 2 Schwarzschild worldlines of raindrops from (10), plotted on the $[r, t]$ slice. These particular raindrops pass $t_A/M = 0$ at different values of r_A/M (filled dots along the horizontal axis). The curves for $r_A/M > 2$ are identical in shape, simply displaced vertically with respect to one another. These worldlines are not continuous across the event horizon (compare Figure 1).

where the expression on the right side of (6) results from substitutions:

$$u \equiv \left(\frac{r}{2M}\right)^{1/2} \quad \text{so} \quad du = \frac{1}{4M}\left(\frac{r}{2M}\right)^{-1/2} dr \quad \text{and} \quad dr = 4Mu\,du \qquad (7)$$

From a table of integrals:

$$\int \frac{u^4 du}{1-u^2} = -\frac{u^3}{3} - u + \frac{1}{2}\ln\left|\frac{1+u}{1-u}\right| \qquad (8)$$

Integrate (6) from u_A to u, where A stands for Above. The integral of (6) becomes:

$$t - t_A = 4M\left[\frac{u_A^3}{3} - \frac{u^3}{3} + u_A - u + \frac{1}{2}\ln\left|\frac{1+u}{1-u}\right| - \frac{1}{2}\ln\left|\frac{1+u_A}{1-u_A}\right|\right] \quad \text{(raindrop)} \qquad (9)$$

Substitute the expression for u from (7) into (9):

$$t - t_A = \frac{4M}{3}\left[\left(\frac{r_A}{2M}\right)^{3/2} - \left(\frac{r}{2M}\right)^{3/2} + 3\left(\frac{r_A}{2M}\right)^{1/2} - 3\left(\frac{r}{2M}\right)^{1/2}\right. \qquad (10)$$

$$\left. + \frac{3}{2}\ln\left|\frac{1+\left(\frac{r}{2M}\right)^{1/2}}{1-\left(\frac{r}{2M}\right)^{1/2}}\right| - \frac{3}{2}\ln\left|\frac{1+\left(\frac{r_A}{2M}\right)^{1/2}}{1-\left(\frac{r_A}{2M}\right)^{1/2}}\right|\right] \quad \text{(raindrop)}$$

Equation (10) is messy, but the computer doesn't care and plots the curves in Figure 2 for $t_{\rm A} = 0$ and $r_{\rm A}/M = 1$ through 9.

Objection 2. *Wait! The t/M versus r/M worldline in Figure 2 tells us that the raindrop takes an unlimited t to reach the event horizon. Do you mean to tell me that our raindrop does not cross the event horizon?*

Recall our "strong advice" in Section 5.6: "To be safe, it is best to assume that global coordinate separations do not have any measured meaning." The worldlines in Figure 2 that rise without limit in t-coordinate as $r/M \to 2^+$ (from above) do not tell us directly what any observer measures. In contrast, the observer riding on the raindrop reads and records her wristwatch time τ as she passes each shell. At each such instant on her wristwatch, she also her direct reading of the r-coordinate stamped on the shell she is passing. When timed on her wristwatch, the raindrop passes smoothly inward across the event horizon (Figure 1).

Objection 3. *There is still a terrible problem with Figure 2. Why does the worldline of the raindrop that somehow makes it inward across $r/M = 2$ run* ***backward*** *in the Schwarzschild t-coordinate?*

We saw this earlier in the light-cone diagram of Figure 8 in Section 3.7, in which the t-coordinate can run backward along a worldline. However, that has no measurable consequence. You cannot grow younger by falling into a black hole. Sorry! The global t-coordinate is *not* time. Does the idea of backward-running global t-coordinate along a worldline make you uncomfortable? Get used to it! In contrast, the time you read on your wristwatch *always* runs forward along your worldline, in particular along the raindrop worldline (Figure 1). In Section 7.4 we develop a set of **global rain coordinates** that not only labels each event—which we require of every set of global coordinates—but also yields predictions more comfortable to our intuition about the "respectable sequence" of global coordinates along a worldline. This (arbitrary) choice of global coordinates is purely for our own convenience: Nature doesn't care!

7.3 ■ THE LOCAL RAIN FRAME IN SCHWARZSCHILD COORDINATES

Carry out experiments as you pass smoothly through the event horizon

Does passing through the event horizon disturb local experiments that we may be conducting during this passage? To answer this questions go step by step from the raindrop to the local inertial rain frame and (in Section 7.4) from the local frame to a new global description. Start with local shell coordinates outside the horizon and use special relativity to derive local rain frame coordinates.

Equations (9) through (11) in Section 5.7 give us local shell coordinates expressed in Schwarzschild global coordinates:

$$\Delta t_{\text{shell}} \equiv \left(1 - \frac{2M}{\bar{r}}\right)^{1/2} \Delta t \qquad \text{(from Schwarzschild)} \tag{11}$$

$$\Delta y_{\text{shell}} \equiv \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2} \Delta r \tag{12}$$

$$\Delta x_{\text{shell}} \equiv \bar{r}\Delta\phi \tag{13}$$

Now use the Lorentz transformation equations of Section 1.10 to find to the local time lapse Δt_{rain} in an inertial frame in which the raindrop is at rest. With respect to the local shell frame, the raindrop moves with velocity v_{rel} in the $-\Delta y_{\text{shell}}$ direction. From equation (18) in Section 6.4:

$$v_{\text{rel}} = -\left(\frac{2M}{r}\right)^{1/2} \quad \text{so} \quad \gamma_{\text{rel}} \equiv \frac{1}{(1 - v_{\text{rel}}^2)^{1/2}} = \left(1 - \frac{2M}{r}\right)^{-1/2} \tag{14}$$

Then from the first of equations (41) in Section 1.10:

$$\begin{aligned} \Delta t_{\text{rain}} &= \gamma_{\text{rel}} \left(\Delta t_{\text{shell}} - v_{\text{rel}} \Delta y_{\text{shell}}\right) \qquad (15) \\ &= \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2} \left[\left(1 - \frac{2M}{\bar{r}}\right)^{1/2} \Delta t + \left(\frac{2M}{\bar{r}}\right)^{1/2} \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2} \Delta r\right] \end{aligned}$$

so that

$$\Delta t_{\text{rain}} = \Delta t + \left(\frac{2M}{\bar{r}}\right)^{1/2} \left(1 - \frac{2M}{\bar{r}}\right)^{-1} \Delta r \tag{16}$$

And from the second of equations (41) in Section 1.10:

$$\begin{aligned} \Delta y_{\text{rain}} &= \gamma_{\text{rel}} \left(\Delta y_{\text{shell}} - v_{\text{rel}} \Delta t_{\text{shell}}\right) \qquad (17) \\ &= \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2} \left[\left(1 - \frac{2M}{\bar{r}}\right)^{-1/2} \Delta r + \left(\frac{2M}{\bar{r}}\right)^{1/2} \left(1 - \frac{2M}{\bar{r}}\right)^{1/2} \Delta t\right] \end{aligned}$$

so that

$$\Delta y_{\text{rain}} = \left(1 - \frac{2M}{\bar{r}}\right)^{-1} \Delta r + \left(\frac{2M}{\bar{r}}\right)^{1/2} \Delta t \tag{18}$$

Finally, from the third of equations (41) in Section 1.10, the shell and rain coordinates transverse to the direction of relative motion have equal values:

$$\Delta x_{\text{rain}} = \Delta x_{\text{shell}} = \bar{r}\Delta\phi \tag{19}$$

The right sides of local rain frame equations (16) and (18) suffer from the same disease as their parent global Schwarzschild coordinates: They blow up at the event horizon. Nevertheless, we can use these *local* coordinate equations to derive a new set of *global* rain coordinates that, at long last, cures this disease and allows us to predict observations made in the local rain frame as we ride smoothly inward across the event horizon and all the way to the singularity.

7.4 ■ GLOBAL RAIN COORDINATES

Convert from Schwarzschild-t to global rain T.

Schwarzschild coordinates are completely legal, but they make us uncomfortable because they do not describe the motion of a stone or light flash inward across the event horizon in a finite lapse of the t-coordinate. So in this section we find a new global coordinate—that we label the T-coordinate—which advances smoothly along the global worldline of a descending stone, even when the stone crosses the event horizon.

Global rain coordinates or "Painlevé-Gullstrand coordinates."

The result is a new set of global coordinates with the old global coordinates r and ϕ but a new T-coordinate. We call this new set of coordinates **global rain coordinates**. They are often called **Painlevé-Gullstrand coordinates** after Paul Painlevé and Alvar Gullstrand who independently developed them in 1921 and 1922, respectively. The present section develops global rain (Painlevé-Gullstrand) coordinates. Section 7.5 uses these new global coordinates to derive the **global rain metric**

?

Objection 4. *Hold on! How can we have two different global coordinate systems for the same spacetime?*

!

For the same reason that a flat Euclidean plane can be described by either Cartesian coordinates or polar coordinates. More than one global coordinate system can describe the same spacetime. Indeed, an unlimited number of global coordinate systems exist for any configuration of mass-energy-pressure (Box 3).

?

Objection 5. *I am awash in arbitrary global coordinates here. What makes practical sense of all this formalism? What can I hang onto and depend upon?*

!

Primary goal: Predict result of local measurements

The answer comes from invariant wristwatch time and invariant ruler distance. These direct observables are outputs of the global metric. In John Wheeler's words, *"No phenomenon is a real phenomenon until it is an observed phenomenon."* Spacetime is effectively flat on a local patch; on that flat patch we use the approximate global metric to derive local coordinates that we choose to be inertial (Section 5.7). The observer makes a measurement and expresses the result in those local coordinates. **Every set of global coordinates must lead to the same predicted result of a given measurement.** That is what makes sense of the formalism; that is what you can hang onto and depend on.

The troublemaker in Schwarzschild coordinates is the t-coordinate, which Figure 2 shows to be diseased at the event horizon. The cure is a new global rain coordinate which we call capital T. The other two global rain coordinates, r and ϕ remain the same as the corresponding Schwarzschild coordinates.

Comment 3. Why conversion from Schwarzschild to global rain?
Most often in this book we simply display a global metric with its global coordinate system without derivation. In what follows, we make a "conversion" of Schwarzschild coordinates to global rain coordinates that leads to the global rain metric. Why this conversion? Why don't we simply display the global rain metric and its coordinate system? We do this to show that there are two ways to derive a valid global metric. The first way is to submit the (almost) arbitrary global coordinates to Einstein's equations, which return the correct global metric. The second way is simply to transform the already-validated global metric directly. That conversion does not require Einstein's equations.

You cannot derive global coordinates from local coordinates. . . .

To find the new global T coordinate we do something apparently illegal: We derive it from *local* rain coordinate Δt_{rain} in equation (16). From the beginning of this book we have emphatically declared that you cannot derive global coordinates from local coordinates. Why not? Because in considering any flat local inertial frame, we lose details of the global curvature of the spacetime region. The local Δt_{rainA} from equation (16) is unique to Frame A which depends on the average value $\bar{r}_{\text{A}}$; adjacent Frame B has a different Δt_{rainB} which depends on a different average value $\bar{r}_{\text{B}}$. This leads to a discontinuity at the boundary between these two local frames. The local frame with its local coordinates is useful for us because it allows us to apply special relativity to an experiment or observation made in a limited spacetime region in globally curved spacetime. But the local frame does have this major drawback: We cannot connect adjacent flat frames smoothly to one another in curved spacetime. That keeps us from generalizing from local frame coordinates to global coordinates.

. . . except when local coordinates lead to an exact differential.

But there is an exception. To understand this exception, pause for a quick tutorial in the mathematical theory of calculus (invented in the late 1600s by both Isaac Newton and Gottfried Wilhelm Leibniz). In calculus we use differentials, for example dt and dr. The technical name for the kind of differential we use in this book is **exact differential**, sometimes called a **perfect differential**. Formally, an exact differential (as contrasted with an inexact differential or a partial differential) has the form dQ or dT, where Q and T are **differentiable functions**. What is a differentiable function? It is simply a function whose derivative exists at every point in its domain.

Global coordinates are differentiable.

Question: Are global coordinates differentiable? *Answer:* We choose global coordinates ourselves, then submit them to Einstein's equations which are differential equations that return to us the global metric. So if one of our chosen global coordinates is not differentiable, it is our own fault. *Conclusion:*

For a global coordinate to be useful in general relativity, it must be differentiable and thus have an exact differential (except at a physical singularity).

So we purposely choose a set of global coordinates that are differentiable. If we then make a transformation between sets of global coordinates, the transformed global coordinate is also a differentiable function and therefore has an exact differential.

A close look at equation (16) shows that we can turn it into an exact differential, as follows:

$$\lim_{\Delta t \to 0} \Delta t_{\text{rain}} = dt + \frac{df(r)}{dr} dr = d[t + f(r)] \equiv dT \tag{20}$$

where, from (16), $T = t + f(r)$ is a differentiable function with

$$\frac{df(r)}{dr} = \left(\frac{2M}{r}\right)^{1/2} \left(1 - \frac{2M}{r}\right)^{-1} \tag{21}$$

Equation (20) is immensely significant. It tells us that there is an exact differential of what we call a new global T-coordinate. This means that T is a differentiable function of global rain coordinates. This coordinate $T = t + f(r)$ is global because (a) Schwarzschild's t and r (along with ϕ) already label every event for $r > 0$, and (b) Schwarzschild coordinates and metric satisfy Einstein's equations. From (20) and (21):

$$dT = dt + d[f(r)] = dt + \left(\frac{2M}{r}\right)^{1/2} \left(1 - \frac{2M}{r}\right)^{-1} dr \tag{22}$$

To validate the new global rain coordinate T, we need to express it in the already-validated Schwarzschild coordinates. To start this process integrate (21):

$$f(r) \equiv \int_0^r \left(\frac{2M}{r}\right)^{1/2} \left(1 - \frac{2M}{r}\right)^{-1} dr = -(2M)^{1/2} \int_0^r \frac{r^{1/2} dr}{2M - r} \tag{23}$$

A table of integrals helps us to integrate the right side of (23). The result is:

$$f(r) = 4M \left(\frac{r}{2M}\right)^{1/2} - 2M \ln \left| \frac{1 + (2M/r)^{1/2}}{1 - (2M/r)^{1/2}} \right| \tag{24}$$

Finally, from (22) and (24):

$$T = t + f(r) = t + 4M \left(\frac{r}{2M}\right)^{1/2} - 2M \ln \left| \frac{1 + (2M/r)^{1/2}}{1 - (2M/r)^{1/2}} \right| \tag{25}$$

where we have arbitrarily set some constants of integration equal to zero. Equation (25) is the final proof that the new global rain T is valid, since it transforms directly from valid Schwarzschild t and r global coordinates.

?

Objection 6. *I still think that your definition of T derived from a local coordinate increment $\Delta t_{\rm rain}$ in equations (22) through (25) violates your own prohibition in Section 5.7 of a local-to-global transformation.*

!

No, we do *not* transform from local rain coordinates to global rain coordinates, which would be illegal. Instead, we start with the global Schwarzschild expression for $\Delta t_{\rm rain}$ in (16), take its differential limit in (20), which converts it to the global rain differential dT. The key step in that conversion is to recognize that equation (20) contains an exact differential in global Schwarzschild coordinates. Note that this does not happen for shell time in Schwarzschild coordinates:

$$\lim_{\Delta t \to 0} \Delta t_{\rm shell} \to \left(1 - \frac{2M}{r}\right)^{1/2} dt \qquad (26)$$

The right side of (26) cannot be expressed as $d[T(r,t)]$, the exact differential of a coordinate $T(r,t)$.

The exact differential in (20) allows us to complete the derivation and validation of the global rain T-coordinate in equation (25). Is this magic? No, but it does require sophisticated use of calculus.

QUERY 4. Differentiate T (*Optional*)
Take the differential of (25) to show that the result is (22).

Coordinate transformation to global rain coordinates

We worked in Schwarzschild coordinates to find a function $T(t,r) = t + f(r)$ whose differential matches local rain frame time. We can free this T of its origin and regard it as a new label for each spacetime event. But after that addition there is no need to use both T and t to label events. We can now eliminate t from the list, because we can always find it if we know the value of T (along with r) through $t = T - f(r)$. So we perform the **coordinate transformation** (25), in which we replace one set of global coordinates (t, r, ϕ) with a new set of global coordinates (T, r, ϕ). We call the result **global rain coordinates**—or historically, **Painlevé-Gullstrand coordinates**.

Payoff: "Intuitive" global rain metric

Why go to all this bother? In order to derive (in the next section) a global rain metric that describes the motion of a stone or light flash across the event horizon in a manner more comfortable to our intuition. The global rain metric encourages our unlimited, free exploration of *all* spacetime outside, at, and inside the event horizon.

An immediate payoff of global rain coordinates is local shell coordinates expressed in global rain coordinates (Box 2).

Comment 4. Global T is defined everywhere.
Although we started from local shell coordinates which exist only for $r > 2M$, our new T coordinate is defined everywhere, even at and inside the event

Box 2. Local shell coordinates expressed in global rain coordinates

We now derive local shell coordinates as functions of global rain coordinates. Equation (11) gives Δt_{shell} as a function of the Schwarzschild t-coordinate increment:

$$\Delta t_{\text{shell}} \equiv \left(1 - \frac{2M}{\bar{r}}\right)^{1/2} \Delta t \tag{27}$$

The Schwarzschild t-coordinate does increase without limit along the worldline of a stone that approaches $r = 2M$, but shells exist only outside this event horizon, where (27) is well-behaved. Write equation (22) in approximate form:

$$\Delta t \approx \Delta T - \left(\frac{2M}{\bar{r}}\right)^{1/2} \left(1 - \frac{2M}{\bar{r}}\right)^{-1} \Delta r \tag{28}$$

Substitute (28) into (27) to yield:

$$\Delta t_{\text{shell}} \equiv \left(1 - \frac{2M}{\bar{r}}\right)^{1/2} \Delta T - \left(\frac{2M}{\bar{r}}\right)^{1/2} \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2} \Delta r \tag{29}$$

Equations for Δy_{shell} and Δx_{shell} do not depend on Δt, so we copy them directly from (12) and (13).

$$\Delta y_{\text{shell}} \equiv \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2} \Delta r \tag{30}$$

$$\Delta x_{\text{shell}} \equiv \bar{r}\Delta\phi \tag{31}$$

Note: Expressions for Δt_{shell} and Δy_{shell} are real only for $r > 2M$, consistent with the conclusion that no shell can exist inside the event horizon.

horizon, because its defining equation (25) contains only global Schwarzschild coordinates, which span all of spacetime.

Should we worry that the last term on the right side of (25) blows up as $r \to 2M$? No, because Schwarzschild t in (25) blows up in the opposite direction in such a way that T is continuous across $r = 2M$. Our only legitimate worry is continuity of the resulting metric, which (32). Section 7.5 shows this metric to be continuous across $r = 2M$.

7.5 ■ THE GLOBAL RAIN METRIC

Move inward across the event horizon.

Section 7.4 created a global T-coordinate, validated it by direct transformation from already-approved global Schwarzschild coordinates, and installed it in a new set of global rain coordinates. To derive the **global rain metric**, solve the dT-transformation (22) for dt, substitute the result into the Schwarzschild metric, and collect terms to yield:

$$d\tau^2 = \left(1 - \frac{2M}{r}\right) dT^2 - 2\left(\frac{2M}{r}\right)^{1/2} dT\,dr - dr^2 - r^2 d\phi^2 \tag{32}$$

$$-\infty < T < +\infty, \quad 0 < r < \infty, \quad 0 \le \phi < 2\pi \quad \text{(global rain metric)}$$

Metric (32), with its connectedness (topology), provides a complete description of spacetime around a non-spinning black hole, just as the

Schwarzschild metric does. In addition, all Schwarzschild-based difficulties with worldlines that cross the event horizon disappear.

QUERY 5. Global rain metric
Substitute dt from (22) into the Schwarzschild metric to verify the global rain metric (32).

QUERY 6. Flat spacetime far from the black hole
Show that as $r \to \infty$ metric (32) becomes the metric for flat spacetime in global coordinates T, r, ϕ.

Use global rain metric from now on.

Comment 5. USE THE GLOBAL RAIN METRIC FROM NOW ON.
From now on in this book we use the global rain metric—and expressions derived from it—to analyze events in the vicinity of the non-spinning black hole.

Principle of General Covariance

Question: When all is said and done, which set of coordinates is the "correct" one for the non-spinning black hole: Schwarzschild coordinates or rain coordinates or some other set of global coordinates? *Answer: Every global coordinate system is valid provided it is either (a) submitted to Einstein's equations, which return a global metric or (b) transformed from an already-validated global coordinate system.* This reflects a fundamental principle of general relativity with the awkward technical name **Principle of General Covariance**. In this book we repeat over and over again that no single observer measures map coordinates directly (back cover). To overlook the Principle of General Covariance by attaching physical meaning to global coordinates is wrong and sets oneself up to make fundamental errors in the predictions of general relativity.

EVERY GENERATION MUST LEARN ANEW

One of the fundamental principles of general relativity (the principle of general covariance) states that all [global] spacetime coordinate systems are equally valid for the description of nature and that metrics that are related by a coordinate transformation are physically equivalent. This principle sounds simple enough, but one repeatedly finds in the literature arguments that amount to advocacy for the interpretation based on one set of [global] coordinates to the exclusion of the interpretation that is natural when using another set of [global] coordinates for the same set of events. It has been said that each generation of physicists must learn anew (usually the hard way) the meaning of Einstein's postulate of general covariance.

—Richard C. Cook and M. Shane Burns

Box 3. An Unlimited Number of Global Coordinate Systems

General relativity uses two methods to derive a global metric: *Method 1:* Choose an arbitrary set of global coordinates and submit them to Einstein's equations, which return the global metric expressed in those coordinates. Karl Schwarzschild did this to derive the Schwarzschild metric. *Method 2:* Transform an already-validated set of global coordinates to another set, then substitute the new coordinates into the already-verified metric of Method 1 to yield a metric in the new global coordinates. In the present chapter we use Method 2 to go from the Schwarzschild metric to the global rain metric.

How many different global coordinate systems are there for the non-spinning black hole? An unlimited number! Any one-to-one transformation from one valid set of global coordinates to another set of global coordinates makes the second set valid as well, provided it meets the usual criteria of uniqueness and smoothness (Section 5.8).

As a simple case, make the transformation:

$$T^{\dagger} \equiv KT \tag{33}$$

where K is any real number—or even a simple minus sign! With this substitution, global rain metric (32) becomes

$$d\tau^2 = \left(1 - \frac{2M}{r}\right)\left(\frac{dT^{\dagger}}{K}\right)^2 - 2\left(\frac{2M}{r}\right)^{1/2}\frac{dT^{\dagger}}{K}dr - dr^2 - r^2 d\phi^2 \tag{34}$$

If K is negative, then global $T^{\dagger}$ runs backward along the worldline of every stone. This does not bother us; Schwarzschild t does the same along some worldlines (Figure 2). Equations for local shell and local rain coordinates are similarly modified, which does not change any prediction or the result of any measurement in these local frames.

In a similar manner we can let $r^{\dagger} \equiv Qr$ and/or $\phi^{\dagger} \equiv -\phi$. You can write down new metrics with any one or any pair of these new coordinates. These changes may seem trivial, but they are not. For example, choose $K = Q = M^{-1}$. The result is the variables T/M and r/M, so-called **unitless coordinates**, in which curves are plotted in many figures of this book. Plots with unitless coordinates are correct for *every* non-spinning black hole, independent of its mass M.

Result: Global rain metric (32) is only one of an unlimited set of alternative, equally-valid global metrics for the non-spinning black hole, each one expressed in a different global coordinate system. The examples in this box may be no more useful than the original global rain metric, but there are other global metrics that highlight some special property of the non-spinning black hole. Look up **Eddington-Finkelstein coordinates** and **Kruskal-Szekeres coordinates** and their metrics.

In particular: Fixation on an interpretation based on one set of arbitrary coordinates can lead to the mistaken belief that global coordinate differences correspond to measurable quantities (Section 2.7).

Reminder of Einstein's error

Comment 6. You know some relativity that Einstein missed!
Chapter 5 says that Einstein took seven years to appreciate that global coordinate separations have no measurable meaning. But even then he did not fully understandd this fundamental idea. At a Paris conference in 1922—seven years after he completed general relativity—Einstein worried about what would happen at a location where the denominator of the dr^2 term in the Schwarzschild metric, namely $(1 - 2M/r)$, goes to zero. He said it would be "an unimaginable disaster for the theory; and it is very difficult to say *a priori* what would occur physically, because the theory would cease to apply." (In 1933 the Belgian priest Georges Lemaître recognized that the apparent singularity in Schwarzschild coordinates at $r = 2M$ is "fictional.") Einstein was also baffled by the $dTdr$ cross term in the global rain metric (32) presented to him by Paul Painlevé and rejected Painlevé's solution out of hand. This led to the eclipse of this metric for decades.

Congratulations: You know some relativity that Einstein did not understand!

QUERY 7. Map energy in global rain coordinates

A. Use the global rain metric and the Principle of Maximal Aging to derive the map energy of a stone in global rain coordinates. You can model the procedure on the derivation of the Schwarzschild-coordinate expression for E/m in Section 6.2, but will need to alter some of the notation. Demonstrate this result:

$$\frac{E}{m} = \left(1 - \frac{2M}{r}\right)\frac{dT}{d\tau} - \left(\frac{2M}{r}\right)^{1/2}\frac{dr}{d\tau} \qquad \text{(global rain coordinates)} \qquad (35)$$

B. Substitute for dT from (22) and show that the result yields the Schwarzschild expression for map energy, equation (8) in Section 6.2.

C. Show that the map energy equation (35) applies to a stone in orbit around the black hole, not just to one that moves along the r-coordinate line.

D. Find an expression for dr/dT of the raindrop in rain map coordinates? Start with the first line of (5) and multiply both sides by dt/dT from (22). Show that the result is:

$$\frac{dr}{dT} = \frac{dr}{dt}\frac{dt}{dT} = -\left(\frac{2M}{r}\right)^{1/2} \qquad \text{(raindrop)} \qquad (36)$$

E. Show that the map worldline of a raindrop is given by the equation:

$$T - T_{\text{A}} = \frac{4M}{3}\left[\left(\frac{r_{\text{A}}}{2M}\right)^{3/2} - \left(\frac{r}{2M}\right)^{3/2}\right] \qquad \text{(raindrop)} \qquad (37)$$

where $(r_{\text{A}}, T_{\text{A}})$ locates the initial event on the plotted curve (Figure 3).

?

Objection 7. *Wait! The right side of (37) is identical to the right side of (2). The two equations are both correct if and only if the left sides are equal:*

$$T - T_{\text{A}} = \tau_{\text{raindrop}}[r_{\text{A}} \rightarrow r] \qquad (38)$$

But these two quantities are completely different—apples and oranges! The left side is the difference in a global coordinate, while the right side is the lapse of wristwatch time of a particular falling stone.

!

For our own convenience, we *chose* global rain coordinates so that equation (38) is valid. This is playing with fire, of course, because it is dangerous to assume that any given global coordinate corresponds to a measurable quantity (Section 5.8). But we are adults now, able to see the pitfalls of mature life. The goal, as always, is correct prediction of measurements and observations.

Raindrop equation (36) for dr/dT looks quite different from raindrop equation (5) for dr/dt, but the two must predict the same raindrop wristwatch time from r_A to a smaller r-coordinate given by (2). Let's check this: For a raindrop, set $E/m = 1$ in (35) and multiply through by $d\tau_{\text{raindrop}}$:

$$d\tau_{\text{raindrop}} = \left(1 - \frac{2M}{r}\right) dT - \left(\frac{2M}{r}\right)^{1/2} dr \tag{39}$$

Solve (36) for dT and substitute into (39).

$$\begin{aligned} d\tau_{\text{raindrop}} &= -\left(1 - \frac{2M}{r}\right)\left(\frac{2M}{r}\right)^{-1/2} dr - \left(\frac{2M}{r}\right)^{1/2} dr \\ &= -\left(\frac{r}{2M}\right)^{1/2} dr \end{aligned} \tag{40}$$

which is the same as equation (1), so its integral from r_A to r must be (2), as required. This is an example of an important property of different global metrics: *Every global metric must predict the same result of a given measurement or observation* (Objection 5).

Box 4 carries out a Lorentz transformation from local shell coordinates in Box 2 to local rain coordinates, then verifies that local rain coordinates are valid everywhere, not just outside the event horizon.

Comment 7. An observer passes through a sequence of local frames. The rain observer rides on a raindrop (Definition 4, Section 7.7). In curved spacetime, local inertial frames are limited in both space and time. During her fall, the rain observer passes through a series of local rain frames as shown in Figure 3. Equations (42) through (44) contain an $\bar{r}$, assumed to have the same value everywhere in that local frame. Although absent from the equations, similar average $\bar{T}$ and $\bar{\phi}$ are implied by all three local rain coordinate equations. *Result:* Each rain observer passes through a sequence of local inertial frames. Similar statements also apply to, and may seem more natural for, local shell frames with local coordinates (29) through (31).

Box 4 derives *local* rain coordinates expressed in *global* rain coordinates. This simplifies local rain coordinates compared with those expressed in Schwarzschild coordinates in equations (16) through (19).

Figure 3 displays several rain observer worldlines on the $[r, T]$ slice. We surround one worldline with a worldtube—shown in cross section—that contains local rain frames through which this rain observer passes in sequence. With $\Delta t_{\text{rain}} = \Delta T = 0$, equation (43) tells us that Δy_{rain} coordinate lines are horizontal in this figure. Finally, Δx_{rain} coordinate lines, which are perpendicular to both Δt_{rain} and Δy_{rain} coordinate lines, project outward, perpendicular to the page in Figure 3.

We could cover the worldtube in Figure 3 with adjacent or overlapping local rain frames. The resulting figure would be analogous to Figure 5 in Section 2.2, which places overlapping local flat maps along the spatial path from Amsterdam to Vladivostok along Earth's curved surface.

Box 4. Local rain coordinates expressed in global rain coordinates

Apply the Lorentz transformation to local shell coordinates in Box 2 to derive local rain coordinates. Relative velocity in the Lorentz transformation lies along the common Δy_{shell} and Δy_{rain} line. With this change, Lorentz transformation equations of Section 1.10 become:

$$\begin{aligned}\Delta t_{\text{rain}} &= \gamma_{\text{rel}}\left(\Delta t_{\text{shell}} - v_{\text{rel}}\Delta y_{\text{shell}}\right) \qquad (41)\\ \Delta y_{\text{rain}} &= \gamma_{\text{rel}}\left(\Delta y_{\text{shell}} - v_{\text{rel}}\Delta t_{\text{shell}}\right)\\ \Delta x_{\text{rain}} &= \Delta x_{\text{shell}}\end{aligned}$$

Substitute v_{rel} and γ_{rel} from (14), along with local shell coordinates from Box 2, into equations (41) to obtain expressions for local rain coordinates as functions of global rain coordinates:

$$\boxed{\Delta t_{\text{rain}} \equiv \Delta T} \qquad (42)$$

$$\boxed{\Delta y_{\text{rain}} \equiv \Delta r + \left(\frac{2M}{\bar{r}}\right)^{1/2}\Delta T} \qquad (43)$$

$$\boxed{\Delta x_{\text{rain}} \equiv \bar{r}\Delta\phi} \qquad (44)$$

Coefficients on the right sides of these equations remain real inside the event horizon, so local rain coordinates are valid there (Figure 3).

Wait! How can we justify our derivation of rain coordinates from local shell coordinates, which are valid only outside the event horizon? To do so, we need to show that local rain coordinates lead back to the global rain metric, which is valid everywhere outside the singularity.

$$\Delta\tau^2 \approx \Delta t_{\text{rain}}^2 - \Delta y_{\text{rain}}^2 - \Delta x_{\text{rain}}^2 \qquad (45)$$

Substitute into (45) from (42) through (44):

$$\begin{aligned}\Delta\tau^2 \approx \Delta T^2 &- \left[\Delta r + \left(\frac{2M}{\bar{r}}\right)^{1/2}\Delta T\right]^2 \qquad (46)\\ &- \bar{r}^2\Delta\phi^2\end{aligned}$$

Multiply out:

$$\begin{aligned}\Delta\tau^2 \approx &\left(1 - \frac{2M}{\bar{r}}\right)\Delta T^2 \qquad (47)\\ &- 2\left(\frac{2M}{\bar{r}}\right)^{1/2}\Delta T\Delta r - \Delta r^2 - \bar{r}^2\Delta\phi^2\end{aligned}$$

In the calculus limit, equation (47) becomes the global rain metric (32). The global rain metric is valid everywhere down to the singularity; therefore local rain coordinates can be constructed down to the singularity.

Box 5 derives the global rain embedding diagram, Figure 4, from the global rain metric (32) and compares it with the embedding diagram for Schwarzschild coordinates.

7.6 ■ TETRAD FORMS OF THE GLOBAL RAIN METRIC

A difference of squares hides the cross term.

Global rain metric (32) has a cross term. The metric for any local inertial frame derived from this global metric does not have a cross term. For example:

$$\Delta\tau^2 \approx \Delta t_{\text{shell}}^2 - \Delta y_{\text{shell}}^2 - \Delta x_{\text{shell}}^2 \qquad (49)$$

$$\Delta\tau^2 \approx \Delta t_{\text{rain}}^2 - \Delta y_{\text{rain}}^2 - \Delta x_{\text{rain}}^2 \qquad (50)$$

Why this difference between global and local metrics? Can we use a set of local inertial coordinates to create a form of the global metric that consists of the sum and difference of squares? Try it! From expressions for local shell and

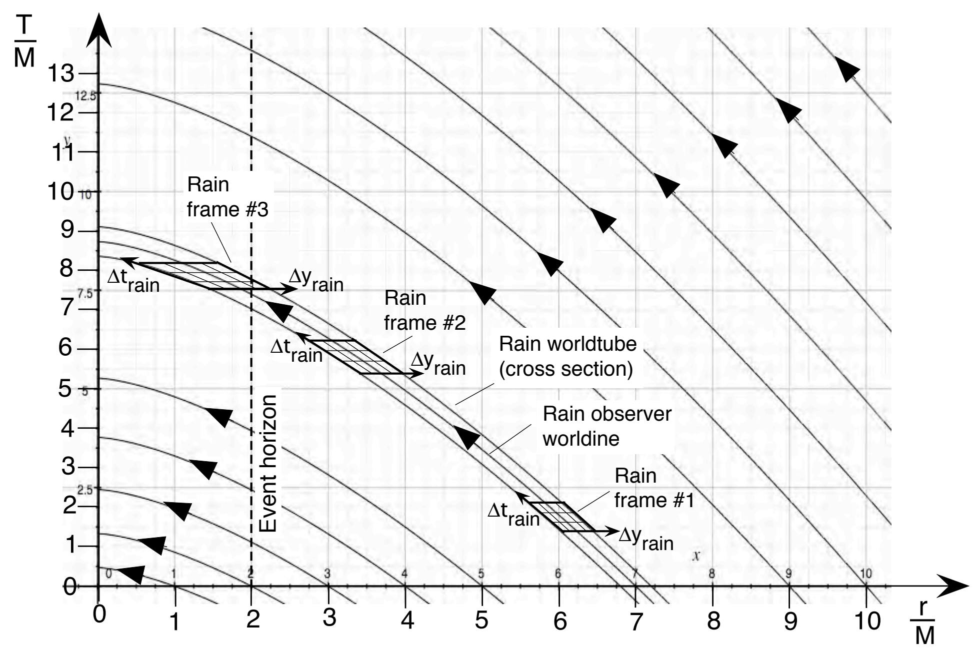


FIGURE 3 Raindrop worldlines plotted on an $[r, T]$ slice. Note that these worldlines are continuous through the event horizon (compare Figure 2). All these worldlines have the same shape and are simply displaced vertically with respect to one another. Around one of these worldlines we construct a worldtube (shown in cross section on this slice) that bounds local rain frames through which that rain observer passes.

rain coordinates in Boxes 2 and 4, respectively, simply write down two differential forms of the global metric. From local shell coordinates (Box 2):

$$d\tau^2 = \left[\left(1-\frac{2M}{r}\right)^{1/2} dT - \left(\frac{2M}{r}\right)^{1/2}\left(1-\frac{2M}{r}\right)^{-1/2} dr\right]^2 - \left(1-\frac{2M}{r}\right)^{-1} dr^2 - r^2 d\phi^2 \qquad \text{(global rain metric)} \qquad (51)$$

$$-\infty < T < +\infty, \quad 0 < r < \infty, \quad 0 \le \phi \le 2\pi$$

And from local rain coordinates (Box 4):

$$d\tau^2 = dT^2 - \left[dr + \left(\frac{2M}{r}\right)^{1/2} dT\right]^2 - r^2 d\phi^2 \qquad \text{(global rain metric)} \qquad (52)$$

$$-\infty < T < +\infty, \quad 0 < r < \infty, \quad 0 \le \phi \le 2\pi$$

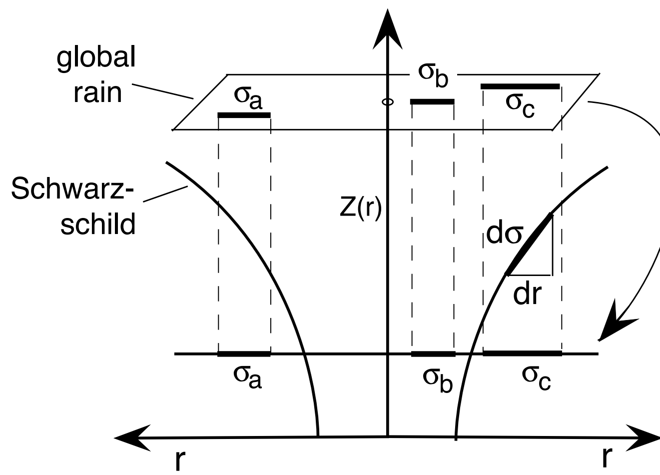


FIGURE 4 Figure for Box 5. Compare the embedding diagram outside the event horizon for Schwarzschild coordinates (the funnel) in Figures 11 through 13 in Section 3.9 with the flat embedding diagram of global rain coordinates (the flat surface shown in perspective across the top). The function $Z(r)$ is the fictional dimension we add in order to visualize these surfaces.

Box 5. Embedding diagrams for Schwarzschild and global rain coordinates.

Set $dT = 0$ in global rain metric (32), which then retains terms that include only r and ϕ:

$$d\sigma^2 = ds^2 = dr^2 + r^2 d\phi^2 \qquad (dT = 0) \qquad (48)$$

Surprise! The differential ruler distance $d\sigma$ obeys Euclidean flat-space geometry, which leads to the flat embedding diagram at the top of Figure 4 (point at $r = 0$ excluded). Because the global rain embedding diagram is flat, we can simply sum increments $d\sigma$ to draw arbitrary lines or curves with measured lengths $\sigma_{\rm a}$, $\sigma_{\rm b}$, and $\sigma_{\rm c}$ (for simplicity, drawn as parallel straight lines in Figure 4).

Figure 4 also repeats the embedding diagram outside the event horizon in Schwarzschild global coordinates from Figures 11 through 13 in Section 3.9.

Outside the event horizon both embedding diagrams are valid for what they describe. And the flat global rain embedding diagram is valid inside the event horizon as well.

What's going on here? Is space flat or funnel-shaped around this black hole? *That depends on our choice of global coordinates! Spacetime* as a unity is curved; but Nature does not care how we share the description of spacetime curvature among the terms of the global metric. In global rain coordinates the dT^2 and $dTdr$ terms describe spacetime curvature, which leaves the $[r, \phi]$ embedding diagram to show flat space. In contrast, for the Schwarzschild metric the dt^2 term and the dr^2 term share the description of spacetime curvature, which yields a funnel outside the event horizon on the $[r, \phi]$ embedding diagram. Each embedding diagram and global light cone diagram is a child of our (arbitrary!) choice of global coordinates in which they are expressed.

Recall Herman Minkowski's declaration (Section 2.7): "Henceforth space by itself and time by itself are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality." Each global metric displays that union in a different way.

Are (51) and (52) valid global metrics? Yes! In Query 8 you multiply out these global metrics to show that they are algebraically equivalent to the original global rain metric (32).

QUERY 8. The same global metric

Expand the right side of (51) and the right side of (52). Show that in both cases the result is equal to the right side of the original global rain metric (32).

Conclusion: All three forms of the global rain metric, (32), (51), and (52) are simply algebraic rearrangements of one another. So what? Why bother? Here's why: Suppose we are first given either metric (51) or metric (52). In that case we can immediately write down the expressions Δt_{frame}, Δy_{frame}, and Δx_{frame} for *some* local inertial frame. (We may not know right away *which* local inertial frame it is.)

A metric of the form (51) or (52) is called the **tetrad form** of the global metric. "Tetra" means "four," in this case the four dimensions of spacetime. A tetrad form rearranges the global metric in a form more useful to us.

DEFINITION 1. A tetrad form of a global metric

A **tetrad form of a global metric** consists of a sum and difference of squares, with no additional terms.

DEFINITION 2. Tetrad

A **tetrad** is a set of four differential expressions, each of which is squared in a "tetrad form" of a global metric.

Example: Equation (52) is the global rain metric in the tetrad form that leads to the local rain frame coordinates in Box 4.

Objection 8. *You said the tetrad consists of* four *differential expressions. I see only three.*

You're right. In most of this book we use only two global space dimensions, those on a slice through the center of the black hole. A third global space dimension would add a fourth Δz_{rain} component to the tetrad. We retain the professional terminology *tetrad* in spite of our simplification.

Comment 8. Tetrad as link

A tetrad is the link between a global metric and a local inertial frame. To specify a particular tetrad, give both the local inertial frame and the global metric. The local inertial frame stretches differentials d in the global metric to deltas Δ in the local inertial frame. This stretch creates elbow room to make measurements in the local inertial frame.

The global metric in tetrad form immediately translates into expressions for local inertial coordinates (Box 6). The remainder of this book will make primary use of such global tetrad metrics.

Box 6. A Brief History of Tetrads

The concept of a tetrad originates from the "repère mobile" (*moving frame*), introduced by Élie Cartan in the 1930s. Cartan showed that a sequence of local inertial coordinate systems, grouped along a set of curves such as worldlines, can provide a complete global description of a curved spacetime. He did so by introducing new calculus concepts on curved spaces, which extended the foundations of differential geometry laid by Bernhard Riemann in 1854.

Cartan's moving-frame theory was incomprehensible to Einstein, but later physicists found it useful and even necessary to study elementary particle physics in curved spacetime.

In this book we simplify the moving frame, or tetrad, to its most basic element: a set of local inertial frames in which motion is described using local inertial coordinates, with each frame and each local coordinate system related to the global coordinates provided by the metric. The tetrad is the bridge between the global metric and the local inertial metric in which we carry out all measurements and observations.

Because this metric is the sum and difference of the squares of the tetrad—such as in equations (51) and (52)—general relativists sometimes call the tetrad the **square root of the metric.**

7.7 ■ RAIN WORLDLINES OF LIGHT

The light flash is ingoing or outgoing—or "outgoing."

In the present chapter we consider only r-motions, motions that can be either ingoing or outgoing along r-coordinate lines. (Chapter 11 analyzes the general motion of light in global coordinates.) The r-motion of light is easily derived from global rain metric (53), which you show in Query 8 to be equivalent to global rain metric (32).

QUERY 9. Light r-motion in global rain coordinates

A. Multiply out the right side of (53) to show that it is equivalent to the global rain metric (32):

$$d\tau^2 = -\left[dr + \left\{1 + \left(\frac{2M}{r}\right)^{1/2}\right\} dT\right]\left[dr - \left\{1 - \left(\frac{2M}{r}\right)^{1/2}\right\} dT\right] - r^2 d\phi^2 \qquad (53)$$

(global rain metric)

B. For light ($d\tau = 0$) that moves along the r-coordinate line ($d\phi = 0$), show that (53) has two solutions for the rain map velocity of light, which are summarized by equation (54):

$$\frac{dr}{dT} = -\left(\frac{2M}{r}\right)^{1/2} \pm 1 \qquad \text{(light flash that moves along the } r\text{-coordinate line)} \qquad (54)$$

$(- =$ incoming light; $+ =$ outgoing or "outgoing" light$)$

C. Look separately at each element of the plus-or-minus sign in (54). Show that the solution with the lower sign ($-$) describes an *incoming* flash (light moving inward along the r-coordinate line). Then show that the solution with the upper sign ($+$) describes an *outgoing* flash (light moving outward along the r-coordinate line) outside the event horizon, but an "*outgoing*" flash inside the event horizon (Definition 3).

DEFINITION 3. "Outgoing" light flash.

The **"outgoing" light flash**—with quotes—is a flash inside the event horizon whose r-motion is described by (54) with the plus sign. Inside the event horizon dr is negative as T advances (positive dT), so that even the "outgoing" light flash moves to smaller r-coordinate. The light cone diagram Figure 5 shows this, and Figure 6 displays longer worldlines of light flashes emitted sequentially by a plunging raindrop. Compare the global rain worldline in Figure 6 with the Schwarzschild worldlines in Figure 2.

We want to plot worldlines of incoming and outgoing (and "outgoing") light flashes. Rewrite (54) to read

$$dT = \frac{dr}{-\left(\dfrac{2M}{r}\right)^{1/2} \pm 1} = \frac{r^{1/2}dr}{-(2M)^{1/2} \pm r^{1/2}} \qquad (r\text{-moving flash}) \qquad (55)$$

We carry the plus-or-minus sign along as we integrate to find expressions for light that moves in either r-direction. Make the substitution:

$$u \equiv -(2M)^{1/2} \pm r^{1/2} \qquad (56)$$

From (56),

$$r^{1/2} = \pm[u + (2M)^{1/2}] \quad \text{and} \quad dr = +2\left[u + (2M)^{1/2}\right] du \qquad (57)$$

With these substitutions, equation (55) becomes

$$\begin{aligned} dT &= \pm 2\frac{\left[u + (2M)^{1/2}\right]^2 du}{u} \qquad (r\text{-moving flash}) \qquad (58)\\ &= \pm 2\left[u + 2\,(2M)^{1/2} + \frac{2M}{u}\right] du \end{aligned}$$

Integrate the second line of (58) from an initial u_0 to a final u. The result is:

$$\pm(T - T_0) = u^2 - u_0^2 + 4(2M)^{1/2}\,(u - u_0) + 4M \ln\left|\frac{u}{u_0}\right| \quad (r\text{-moving flash}) \qquad (59)$$

To restore global rain coordinates in (59), reverse the substitution in (56). (*Hint:* To save time—and your sanity—replace $\pm$ in (59) with a symbol such as Q.) There are two cases: *First case:* an incoming flash from a larger r-coordinate $r_0 = r_A$ at coordinate $T_0 = T_A$ to a lower r at coordinate T. For

Incoming flash

this incoming flash, take the lower minus signs in (56) and (59). Multiply both sides of the result by minus one to obtain:

$$\begin{aligned} T - T_{\rm A} = (r_{\rm A} - r) - 4M\left[\left(\frac{r_{\rm A}}{2M}\right)^{1/2} - \left(\frac{r}{2M}\right)^{1/2}\right] \\ + 4M \ln\left[\frac{1 + (r_{\rm A}/2M)^{1/2}}{1 + (r/2M)^{1/2}}\right] \quad \text{(incoming flash)} \end{aligned} \tag{60}$$

QUERY 10. Horizon-to-crunch global T-coordinate lapse for light

A. Verify that for $r = r_{\rm A}$ in (60), the elapsed global T-coordinate $T - T_{\rm A}$ is zero, as it must be for light.

B. Show that from the event horizon $r_{\rm A} = 2M$ to the crunch point $r = 0$, the elapsed T-coordinate $T - T_{\rm A} = 0.773M$.

C. Compare the result of Item B with the event-horizon-to-crunch wristwatch time $\tau_{\rm raindrop} = (4/3)M$ in equation (3). Why is the result for a light flash in (60) less than the result for the raindrop? Are the plots in Figure 6 consistent with this inequality?

Outgoing flash

Second case: the outgoing and "outgoing" light flashes from an initial r-coordinate $r_0 = r_{\rm L}$ at coordinate $T_{\rm L}$ to a final r at coordinate T. In this case we must take the upper, plus signs in (56) and (59). The result is:

$$\begin{aligned} T - T_{\rm L} = (r - r_{\rm L}) + 4M\left[\left(\frac{r}{2M}\right)^{1/2} - \left(\frac{r_{\rm L}}{2M}\right)^{1/2}\right] \\ + 4M \ln\left[\frac{1 - (r/2M)^{1/2}}{1 - (r_{\rm L}/2M)^{1/2}}\right] \quad \text{(outgoing flash)} \end{aligned} \tag{61}$$

Inside event horizon, "outward" means worldline with $dr < 0$.

What does equation (61) predict when $r_{\rm L} < 2M$? Worldlines of light that sprout upward from the raindrop worldline in Figure 6 show that after the diver falls through the event horizon, even the "outward" flash moves to smaller r in global rain coordinates.

Figure 5 uses equations (60) and (61) to plot light cones for a selection of events inside and outside of the event horizon.

QUERY 11. Motion to smaller r only

Use the dashed worldline of a stone in Figure 5 to explain, in one or two sentences, why "everything moves to smaller r-coordinate" inside the event horizon. *Hint:* Think of the connection between worldlines of stones and future light cones.

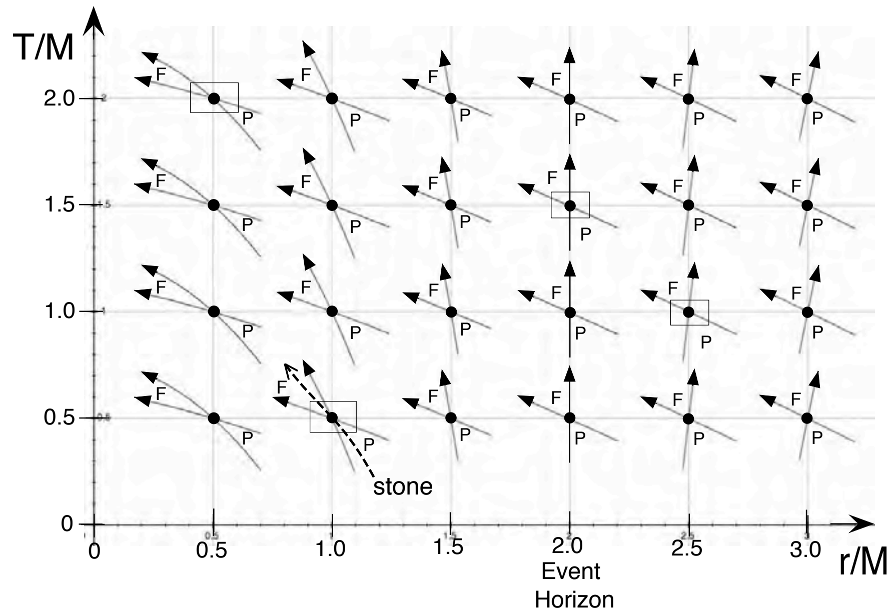


FIGURE 5 Light cone diagram on the $[r, T]$ slice, plotted from equations (60) and (61) for events both outside and inside the event horizon. Past and future events of each filled-dot-event are corralled inside the past (P) and future (F) light cone of that event. At each r-coordinate, the light cone can be moved up or down vertically without change of shape, as shown. Inside the event horizon, light and stones can move only to smaller r-coordinate. A few sample boxes show locally flat patches around a single event. The global rain T-coordinate conveniently runs forward along every worldline (in contrast to the Schwarzschild t-coordinate along some worldlines in the light cone diagram of Figure 8, Section 3.7).

QUERY 12. Detailed derivation *Optional*
Show details of the derivation of equations (60) and (61) from equation (59). Recall the hint that follows equation (59).

Time to celebrate! The raindrop worldline in Figure 6 is continuous and smooth as it moves inward across the horizon. Global rain coordinates yield predictions that are natural and intuitive for us. With global rain coordinates, we no longer need to reconcile the awkward contrast between the discontinuous Schwarzschild worldlines of Figure 2 and the smooth advance of raindrop wristwatch time in Figure 1 (even though both of these plots are valid and technically correct). The simplicity of results for global rain coordinates leads us to use them from now on to describe the non-spinning black hole. Farewell, Schwarzschild metric!

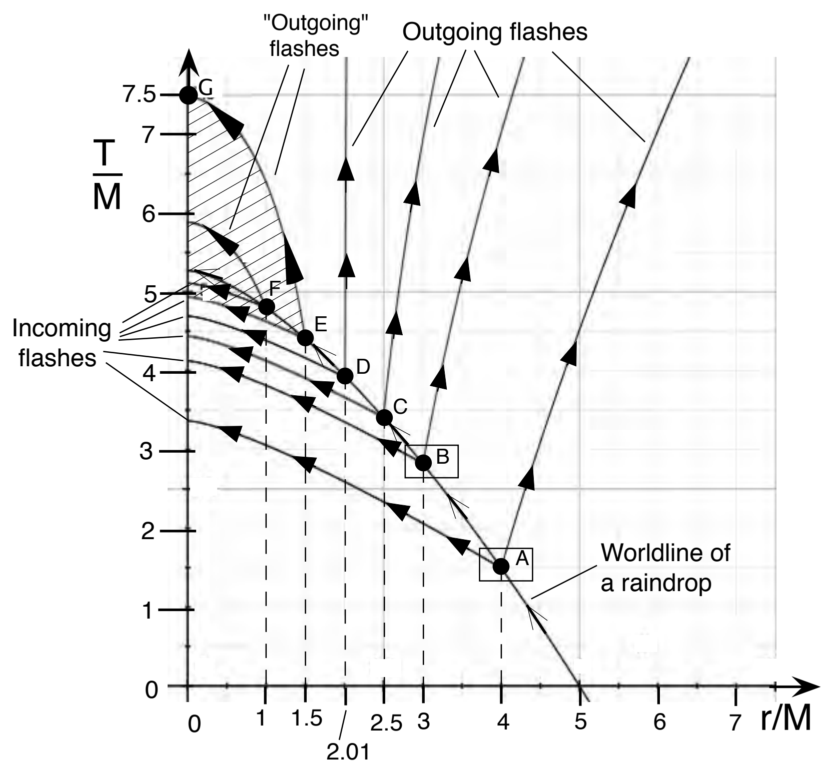


FIGURE 6 A raindrop passes $r/M = 5$ at $T/M = 0$ and thereafter emits both incoming and outgoing flashes at events A through F. "Outgoing" flashes—with quotes—from events E and F move to smaller global r-coordinates, along with everything else inside the event horizon. Little boxes at A and B represent two of the many locally flat patches through which the rain observer passes as she descends. When the rain diver reaches event E, her "range of possible influence" consists of events in the shaded region, for example event F.

7.8 ■ THE RAIN OBSERVER LOOKS—AND ACTS

Which distant events can the rain observer see? Which can she influence?

You ride a raindrop; in other words, you fall from initial rest far from the black hole. What do you see radially ahead of you? behind you? Of all events that occur along this r-line, which ones can you influence from where you are? Which of these events can influence you? When can you no longer influence any events? To answer these questions we give the raindrop some elbow room, turn her into a **rain observer** who makes measurements and observations in a series of local rain frames through which she falls. This definition specializes the earlier general definition of an observer (Definition 4, Section 5.7).

Definition: **Rain observer**

DEFINITION 4. Rain observer

A **rain observer** is a person or a data-collecting machine that rides a raindrop. As she descends, the rain observer makes a sequence of measurements, each measurement limited to a local inertial rain frame (Box 4).

Objection 9. *Wait: Go back! You have a fundamental problem that ruins everything. The global rain metric (32) contains the r-coordinate, but you have not defined the r-coordinate inside the event horizon. Section 3.3 defined the r-coordinate as "reduced circumference," that is, the circumference of a shell divided by 2π. But you cannot build a shell inside the event horizon, so you cannot define global coordinate r there. Therefore you have no way even to describe the worldline of the rain observer once she crosses the event horizon.*

Guilty as charged! Box 4 defined local rain coordinates and justified their validity inside the event horizon, but we have not formally defined the r-coordinate inside the event horizon, or how an observer might determine its value there. Here is one way (Box 7): As the rain observer drops from rest far from the black hole, she simultaneously releases a stone test particle from rest beside her and perpendicular to her direction of motion. Thereafter she uses radar or a meter stick to measure the distance to the stone. In this way she monitors her r-coordinate as she descends inside the event horizon.

Box 4 introduced local rain frames in which we can carry out and record measurements using special relativity. Small boxes in Figures 5 and 6 represent effectively flat patches on which we can construct local inertial frames. In this chapter we allow the rain observer to look only at events that lie before and behind her along her worldline. She can also send light flashes and stones to influence (as much as possible) this limited set of events. (Chapter 11 allows the raindrop observer to look all around her.)

QUERY 13. Observe ingoing and outgoing light flashes in a local rain frame. How do light flashes that we describe as ingoing, outgoing, and "outgoing" in global rain coordinates (Definition 3) move when observed entirely within a local rain frame? Answer this question with the following procedure or some other method.

A. From (42) and (43) show that:

$$\Delta y_{\text{rain}} = \left[\frac{\Delta r}{\Delta T} + \left(\frac{2M}{\bar{r}}\right)^{1/2}\right]\Delta t_{\text{rain}} \qquad \text{(light flash that moves along the } r\text{-coordinate line)} \tag{66}$$

B. Use an approximate version of (54) to replace the square bracket expression in (66):

$$\frac{\Delta y_{\text{rain}}}{\Delta t_{\text{rain}}} = \pm 1 \qquad \text{(light flash that moves along the } r\text{-coordinate line)} \tag{67}$$

C. Is equation (67) a surprise—or obvious? What does each sign mean for measurement of light velocity inside a rain frame?

Box 7. Define the Value of r Inside the Event Horizon

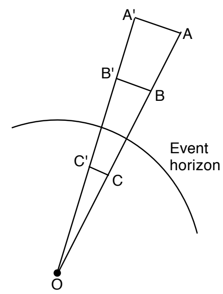


FIGURE 7 The rain observer measures her r-coordinate inside the event horizon.

Question: How can a rain observer inside the event horizon determine her current r-coordinate? *Answer:* To adapt the Chapter 3 definition of the r-coordinate—reduced circumference of the shell outside the event horizon—she measures only a tiny arc inside the event horizon.

The rain observer takes the path ABCO in Figure 7; the stone that accompanies her takes the converging path A'B'C'O. Draw a circular arc AA' and similar circular arcs BB' and CC'. The angle AOA' is the same for every arc, so the length of each arc represents the same fraction of the circumference of its corresponding circle. In equation form,

$$\frac{\begin{pmatrix}\text{length of}\\ \text{arc AA'}\end{pmatrix}}{\begin{pmatrix}\text{circumference}\\ \text{of shell thru A}\end{pmatrix}} = \frac{\begin{pmatrix}\text{length of}\\ \text{arc BB'}\end{pmatrix}}{\begin{pmatrix}\text{circumference}\\ \text{of shell thru B}\end{pmatrix}} \qquad (62)$$

The values of r-coordinates r_{A} and r_{B} are stamped on the shells outside the event horizon, so the denominators of the two sides of the equation become $2\pi r_{\mathrm{A}}$ and $2\pi r_{\mathrm{B}}$, respectively, and we cancel the common factor 2π.

If the angle at the center is small enough, we can replace the length of each circular arc with the straight-line distance measured between, say, A and A' shown in Figure 7. Call this measured distance AA'. And call BB' the corresponding straight-line distance measured between B and B'. Then (62) becomes,

$$\frac{\mathrm{AA}'}{r_{\mathrm{A}}} \approx \frac{BB'}{r_{\mathrm{B}}} \qquad (63)$$

The rain observer monitors the distance to her accompanying stone as she descends, with radar or—if the stone lies near enough—directly with a meter stick. While she is outside the event horizon, the rain observer reads the value of the r-coordinate r_{A} stamped on that spherical shell as she passes it and the measured distance AA' between the two rain frames, and later BB' as she passes and reads off r_{B}. She verifies that this direct reading with the value of r_{B} is the same as that calculated with the equation:

$$\mathrm{r_B} \approx \frac{\mathrm{BB}'}{\mathrm{AA}'} r_{\mathrm{A}} \qquad (64)$$

At any point C inside the event horizon, the observer measures distance CC' and *defines* her instantaneous r-coordinate r_{C} as:

$$r_{\mathrm{C}} \equiv \frac{\mathrm{CC}'}{\mathrm{AA}'} r_{\mathrm{A}} \qquad \text{(definition)} \qquad (65)$$

This definition of the r-coordinate inside the event horizon is a direct extension of its definition outside the event horizon and is valid for any observer falling along an r-coordinate line.

D. From observations inside a rain frame, is there any difference between a light flash we describe as *outgoing* and one we describe as *"outgoing"*? More generally, can observations carried out entirely inside a rain frame tell us whether that rain frame is outside of, at, or inside the event horizon?

"Range of possible influence"

As our rain observer arrives at any of the emission points A through F in Figure 6, she can try—by firing an ingoing or outgoing stone or light flash—to influence a later event located within the region embraced by the worldlines of the incoming and outgoing (or "outgoing") flashes from that event. The farther toward the singularity the rain observer falls, the smaller is this "range of possible influence.” When she arrives at event E, for example, she can influence only events in the shaded region in Figure 6, including event F.

QUERY 14. Future events that the rain observer can still influence.
Make four photocopies of Figure 6. On each copy, choose one emission event A through D or F.

A. Shade the spacetime region in which a rain observer can influence future events once she has arrived at that emission event.

B. Which of these emission points is the last one from which the rain observer can influence events that occur at $r > 2M$?

As she crosses the event horizon, how long will it be on her wristwatch before she reaches the singularity? Equation (3) tells us this wristwatch time is $4M/3$ meters.

Objection 10. *Ha! I can live a lot longer inside the event horizon than your measly $4M/3$ meters of time. All I have to do, once I get inside the event horizon, is to turn on my rockets and boost myself radially outward. For example, I can fire super-powerful rockets at event E and follow the “outgoing” photon flash from E that reaches the singularity at Event G (top left corner of that figure). That final T-value is much greater than the T-value where the raindrop worldline reaches the singularity.*

Be careful! You want to maximize wristwatch time, not the span of global T which, remember, is usually *not* measureable time. The wristwatch time is zero along the worldline of a light flash, so the closer you come to that worldline the smaller will be your wristwatch time during descent from Event E to an event just below G in Figure 6.

Objection 11. *Okay, then! I’ll give up the rocket blast, but I still want to know what is the longest possible wristwatch time for me to live after I cross the event horizon.*

Part B of Exercise 3 at the end of this chapter shows how to extend your lifetime to πM meters after you cross the event horizon, which is a bit longer than the raindrop $4M/3$ meters. You will show that the way to achieve this is to drip from the shell just outside the event horizon; that is, you release yourself from rest in global coordinates at $r = 2M^+$.

Objection 12. *Can I increase my lifetime inside the event horizon by blasting rockets in either ϕ direction to add a tangential component to my global velocity?*

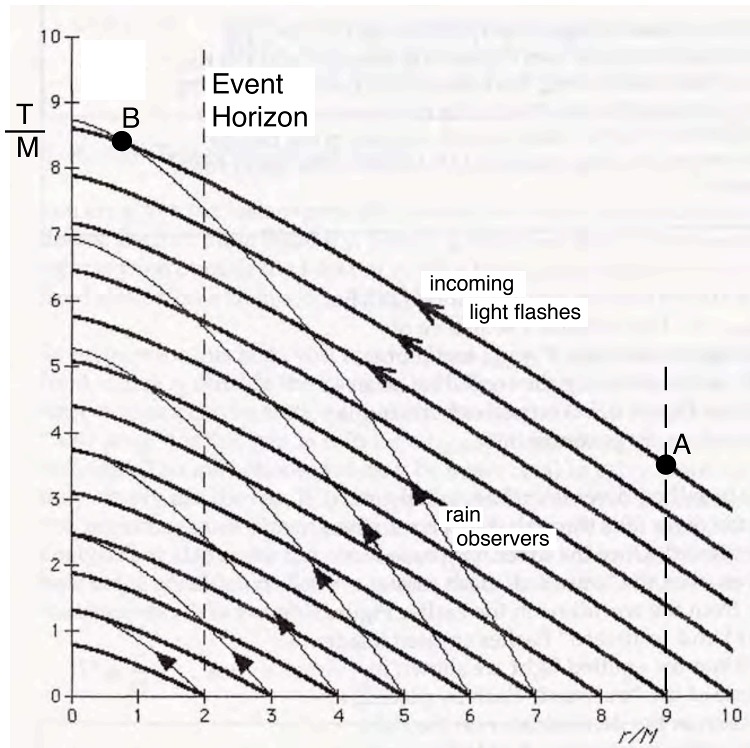


FIGURE 8 Worldlines of rain observers (thin curves) and incoming light flashes (thick curves) plotted on the $[r, T]$ slice. All rain observer worldlines have the same form and can be moved up and down without change in shape. Light-flash worldlines also have the same form. A shell observer at $r/M = 9$ emits a signal from Event A that a rain observer receives inside the event horizon at Event B.

!

The present chapter analyzes only r-motion. In the exercises of Chapter 8 you will show that the answer to your question is no; a tangential rocket blast *decreases* your lifetime inside the event horizon. A wristwatch time lapse of πM is the best you can do. Sorry.

Figure 8 displays both global worldlines of light flashes—thick curves derived from (60)—and worldlines of raindrops—thin curves derived from (2). It is evident from Figure 8 that news bulletins—incoming or outgoing electromagnetic radio bursts fired outside the event horizon—can be scheduled to catch up with the diver community at any predetermined r-coordinate.

QUERY 15. Can you see a rain diver ahead of you?

Compare Figures 6 and 8. Label as #1 the rain diver whose worldline is plotted in Figure 6. Label as #2 a second rain diver who falls along the *same* r-coordinate line in space as rain diver #1, but at a T-coordinate greater by ΔT. Use worldlines of Figure 8 to answer the following questions. (*Optional:*

Derive analytic solutions to these questions and compare the results with your answers derived from the figure.)

- **A.** Over what range of delays ΔT will rain diver #2 be able to see the flash from Event E emitted by diver #1 but not the flash from Event F?
- **B.** Over what range of delays ΔT will rain diver #2 be able to see the flash from Event D emitted by diver #1 but not the flash from Event E?
- **C.** Answer this question decisively: Can any later diver #2 see a flash emitted by diver #1 at say, $r/M = 0.1$, just before diver #1 reaches the singularity?

QUERY 16. Can you see a rain diver behind you?
Extend the results of Query 15 to analyze a third rain diver labeled #3 who falls along the same r-coordinate line in space as the earlier two rain divers, but at a T-coordinate that is smaller by $-\Delta T$. Diver #3 looks outward at light pulses emitted by diver #1. Use the worldlines of Figure 6 to answer the following questions. (*Optional:* Derive analytic solutions to these questions and compare the results with your answers derived from the figure.)

- **A.** Over what range of earlier launches $-\Delta T$ will rain diver #3 be able to see the flash from Event D emitted by diver #1 but not the flash from Event F?
- **B.** Over what range of earlier launches $-\Delta T$ will rain diver #3 be able to see the flash from Event A emitted by diver #1 but not the flash from Event D?
- **C.** Answer this question decisively: Can any earlier diver #3 see a flash emitted by diver #1 at say, $r/M = 0.1$, just before diver #1 reaches the singularity?

QUERY 17. Can you see the crunch point ahead of you?
You are the rain diver whose worldline is plotted in Figure 6. By some miracle, you survive to reach the center of the black hole. Show that you *cannot* see the singularity ahead of you before you arrive there. (What a disappointment after all the training, preparation, and sacrifice!)

QUERY 18. Rain frame energy
The rain frame is inertial. Therefore the expression for energy of a stone in rain frame coordinates is that of special relativity, namely $E_{\text{rain}}/m = dt_{\text{rain}}/d\tau$ (Section 1.7). Recall also from the differential version of (42) that $dt_{\text{rain}} = dT$.

A. Use (35) together with the special relativity expression for the rain frame energy of the stone and the above identification of global rain T with rain frame time from (42) to show that:

Box 8. The River Model

FIGURE 9 In the river model of a black hole, fish that swim at different rates encounter a waterfall. The fastest fish represents a photon. "The fish upstream can make way against the current, but the fish downstream is swept to the bottom of the waterfall." The event horizon corresponds to that point on the waterfall at which the upward-swimming photon-fish stands still. [From Hamilton and Lisle, see references]

Andrew Hamilton and Jason Lisle created a **river model** of the black hole. In their model, water "looks like ordinary flat space, with the distinctive feature that space itself is flowing inward at the Newtonian escape velocity. The place where the infall velocity hits the speed of light . . . marks the event horizon. . . . Inside the event horizon, the infall velocity exceeds the speed of light, carrying everything with it." At every r-coordinate near a black hole the river of space flows past at the speed of a raindrop, namely a stone that falls from initial rest far from the waterfall.

Envision flat spacetime distant from a black hole as still water in a large lake with clocks that read raindrop wristwatch time τ_{raindrop} floating at rest with respect to the water. At one side of the lake the water drifts gently into a river and that carries the raindrop clocks with it. River water moves faster and faster as it approaches and flows over the brink of the waterfall. Each jet of falling water narrows as it accelerates downward. Fish represent objects that move in the river/space; the fastest fish represents a photon. At some point below the lip of the waterfall, not even the photon-fish can keep up with the downward flow and is swept to the bottom of the falls (Figure 9) The black hole event horizon corresponds to the point at which the upward-swimming "photon-fish" stands still.

The river model helps us to visualize many effects observed near the black hole. Hamilton and Lisle write, "It explains why light cannot escape from inside the event horizon, and why no star can come to rest within the event horizon. It explains how an extended object will be stretched radially by the inward acceleration of the river, and compressed transversely by the spherical convergence of the flow. It explains why an object that falls through the event horizon appears to an outsider redshifted and frozen at the event horizon: as the object approaches the event horizon, light [a photon-fish] emitted by it takes an ever-longer global time to forge against the onrushing current of space and eventually to reach the outside observer."

Hamilton and Lisle show that the river model is consistent with the results of general relativity. In that sense the river model is *correct* and *complete*.

The river model is a helpful visualization, but that visualization comes at a price. It carries two misleading messages: First, that space itself—represented by the river—is observable. We easily observe various flows of different rivers on Earth, but no one—and no instrument—registers or observes any "flow of space" into a black hole. Second, the river model embodies global rain coordinates, but we have seen that there are an unlimited number of global coordinates for the black hole, many of which cannot be envisioned by the river model.

$$\frac{E_{\text{rain}}}{m} \equiv \lim_{\Delta\tau\to 0}\frac{\Delta t_{\text{rain}}}{\Delta\tau} \equiv \frac{dT}{d\tau} = \left(1-\frac{2M}{r}\right)^{-1}\left[\frac{E}{m}+\left(\frac{2M}{r}\right)^{1/2}\frac{dr}{d\tau}\right] \qquad (68)$$

B. Is it possible for E_{rain}/m to become negative inside the event horizon? Would any observer complain if it did?

C. Same questions as Item B for E/m, the global map energy per unit mass.

D. Perform a Lorentz transformation (Section 1.10) with v_{rel} and γ_{rel} from (14) to obtain E_{shell} in terms of E. Compare with E and E_{shell} in Schwarzschild coordinates from Sections 6.2 and 6.3.

Box 9. The Planck length

General relativity is a classical—non-quantum—theory (Box 7, Section 6.7). One of its beauties is that, when applied to the black hole, general relativity points to its own limits. The Schwarzschild metric plus the Principle of Maximal Aging predict that everything which moves inward across the horizon will end up on the singularity, a point. We know that this is does not satisfy quantum mechanics: The Heisenberg uncertainty principle of quantum mechanics tells us that a single electron confined to a point has unlimited momentum. So not even a single electron—much less an entire star gobbled up by the black hole—can be confined to the singularity. In this book we assume that classical general relativity is valid until very close to the singularity. How close? One estimate is the so-called **Planck length**, derived from three fundamental constants:

$$\text{Planck length} = \left(\frac{hG}{2\pi c^3}\right)^{1/2} = 1.616\,199 \times 10^{-35} \tag{69}$$

in meters, with an uncertainty of ± 97 in the last two digits. The presence in this equation of Planck's constant $h = 1.054\,571\,726 \times 10^{-34}$ Joule-second (± 15 in the last two digits) tells us that we have entered the realm of quantum mechanics, where classical general relativity is no longer valid. Cheer up! Before any part of you arrives at the Planck distance from the singularity, you will no longer feel any discomfort.

What happens when a single electron arrives at a Planck length away from the singularity? Nobody knows!

7.9 ■ A MERCIFUL ENDING?

How long does the "terminal spaghettification" process last?

Is death near the singularity painful?

To dive into a black hole is to commit suicide, which may go against religious, moral, or ethical principles—or against our survival instinct. Aside from such considerations, no one will volunteer for your black-hole diver research team if she predicts that as she approaches the crunch point her death will be painful. Your task is to estimate the **ouch time** τ_{ouch}, defined as the lapse of time on the wristwatch of the diver between her first discomfort and her arrival at the singularity, $r = 0$.

QUERY 19. Preliminary: Acceleration g in units of inverse meters.
Newton's expression for gravitational force in conventional units:

$$F_{\text{conv}} \equiv m_{\text{conv}} g_{\text{conv}} = -\frac{GM_{\text{conv}} m_{\text{conv}}}{r^2} \qquad \text{(Newton)} \tag{70}$$

A. Verify the resulting gravitational acceleration in units of inverse meters:

$$g \equiv \frac{g_{\text{conv}}}{c^2} = -\frac{M}{r^2} \qquad \text{(Newton)} \tag{71}$$

B. Show that at Earth's surface the Newtonian acceleration of gravity has the value given inside the front cover, namely

$$|g_{\text{Earth}}| \equiv |g_{\text{E}}| = \left|-\frac{M_{\text{Earth}}}{r^2_{\text{Earth}}}\right| = 1.09 \times 10^{-16}\ \text{meter}^{-1} \qquad \text{(Newton)} \tag{72}$$

Rain frame observer feels tides.

The rain observer is in free fall and does not feel any net force as a result of local acceleration. However, she does feel radially stretched due to a *difference* in acceleration between her head and her feet, along with a compression from side to side. We call these differences **tidal accelerations**.

QUERY 20. Tidal acceleration along the r-coordinate line

We want to know how much this acceleration *differs* between the head and the feet of an in-falling rain observer. Take the differential of g in (71). Convert the result to increments over a patch of average $\bar{r}$. Show that

$$\Delta g \approx \frac{2M}{\bar{r}^3}\Delta r \qquad \text{(Newton)} \tag{73}$$

What does Einstein say about tidal acceleration? Section 9.7 displays the correct general relativistic expressions for the variation of local gravity with spatial separation. *Surprise:* The expression for tidal acceleration in *any* inertial frame falling along the r-coordinate line has a form identical to the Newtonian result (73), and thus for the local rain frame becomes:

$$\Delta g_{\text{rain}} \approx \frac{2M}{\bar{r}^3}\Delta y_{\text{rain}} \tag{74}$$

Define "discomfort."

What are the criteria for discomfort? Individual rain observers will have different tolerance to tidal forces. To get a rough idea, let the rain observer's body be oriented along the r-coordinate line as she falls, and assume that her stomach is in free fall, feeling no stress whatever. Assume that the rain observer first becomes uncomfortable when the *difference* in local acceleration between her free-fall stomach and her head (or her feet), that stretches her, is equal to the acceleration at Earth's surface. By this definition, the rain observer becomes uncomfortable when her feet are pulled downward with a force equal to their weight on Earth and her head is pulled upward with a force of similar magnitude. Let her height be h in her frame, and the ruler distance between stomach and either her head or her feet be half of this, that is, $\Delta y_{\text{rain}} = h/2$.

QUERY 21. The r-value for the start of "ouch."

From our criteria above for discomfort, we have:

$$\Delta g_{\text{rain ouch}} \equiv g_{\text{E}} \qquad \text{(stomach-to-foot distance)} \tag{75}$$

Show that the r-value for the start of "ouch," namely r_{ouch}, is:

$$r_{\text{ouch}} = \left(\frac{Mh}{g_{\text{E}}}\right)^{1/3} \qquad (h = \text{head-to-foot height}) \tag{76}$$

QUERY 22. Three cases for the start of "ouch."
Approximate h, the head-to-foot distance, as 2 meters. Find the value of the ratio $r_{\text{ouch}}/(2M)$ for these cases:

A. A black hole with ten times the mass of our Sun.
B. The "20-year black hole" in Query 3.
C. Suppose that the ouch r-coordinate is at the event horizon. What is the mass of the black hole as a multiple of the Sun's mass?

QUERY 23. The wristwatch ouch time τ_{ouch}

A. Use equation (2) to show that the raindrop ouch time τ_{ouch} (the wristwatch time between initial ouch and arrival at the singularity) is *independent of the mass of the black hole*:

$$\tau_{\text{ouch}} = \frac{1}{3}\left(\frac{2h}{g_{\text{E}}}\right)^{1/2} \qquad \text{(raindrop wristwatch ouch time in meters)} \qquad (77)$$

Here, recall, h is the height of the astronaut, about 2 meters, and the value of g_{E} is given in (72). Show that the wristwatch ouch time in seconds, the same for *all* non-spinning black holes, is:

$$\tau_{\text{ouch}} = \frac{1}{3c}\left(\frac{2h}{g_{\text{E}}}\right)^{1/2} \qquad \text{(raindrop wristwatch ouch time in seconds)} \qquad (78)$$

B. Substitute numbers into equation (78). Show that the duration of raindrop wristwatch ouch time is about 2/9 of a second for *every* non-spinning black hole, independent of its mass M. *Guess:* Will pain signals travel from your extremities to your brain during this brief wristwatch ouch time?

MUTABILITY OF PHYSICAL LAWS

By 1970, I had become convinced not only that black holes are an inevitable consequence of general relativity theory and that they are likely to exist in profusion in the universe, but also that their existence implies the mutability of physical law. If time can end in a black hole, if space can be crumpled to nothingness at its center, if the number of particles within a black hole has no meaning, then why should we believe that there is anything special, anything unique, about the laws of physics that we discover and apply? These laws must have come into existence with the Big Bang as surely as space and time did.

—John Archibald Wheeler

7.10 ■ EXERCISES

1. Crossing the Event Horizon

Pete Brown disagrees with the statement, "No special event occurs as we fall through the event horizon." He says, "Suppose you go feet first through the event horizon. Since your feet hit the event horizon before your eyes, then your feet should disappear for a short time on your wristwatch. When your eyes pass across the event horizon, you can see again what's inside, including your feet. So tie your sneakers tightly or you will lose them in the dark!" Is Pete correct? Analyze his argument without criticizing him.

2. Equations of Motion of the Raindrop

From equations in Chapter 6 we can derive the equations of motion for a raindrop in Schwarzschild coordinates. From the definition of the raindrop,

$$\frac{E}{m} = 1 \quad \text{and} \quad \frac{d\phi}{d\tau} = 0 \qquad \begin{array}{c}\text{(raindrop in Schwarzschild coordinates} \\ \text{and in global rain coordinates)}\end{array} \tag{79}$$

In addition, equation (23) in Section 6.4 tells us that

$$\frac{dr}{d\tau} = -\left(\frac{2M}{r}\right)^{1/2} \qquad \begin{array}{c}\text{(raindrop in Schwarzschild coordinates} \\ \text{and in global rain coordinates)}\end{array} \tag{80}$$

From equation (13) in Section 6.4, you can easily show that

$$\frac{dt}{d\tau} = \left(1 - \frac{2M}{r}\right)^{-1} \qquad \begin{array}{c}\text{(raindrop in Schwarzschild coordinates} \\ \text{and in global rain coordinates)}\end{array} \tag{81}$$

Now derive the raindrop equations of motion in global rain coordinates. First, show that both $dr/d\tau$ and $d\phi/d\tau$ have the same form in global rain coordinates as in Schwarzschild coordinates, as stated in the labels of equations (80) and (81). Second, use equations (80) and (81) plus equation (35) to show that

$$\frac{dT}{d\tau} = 1 \qquad \text{(raindrop in global rain coordinates)} \tag{82}$$

Comment 9. Simple definition of global rain T
Equation (82) can be used as the definition of global rain coordinate differential dT. In other words, we *choose* dT equal to the differential lapse of wristwatch time on a falling raindrop.

3. Different masses for the "20-year black hole."

This chapter describes a "20-year black hole," defined as one for which the wristwatch on a raindrop registers a 20-year lapse between its crossing of the event horizon and its arrival at the singularity. But the wristwatch may be on a hailstone, flung radially inward from far away; or on a drip, dropped from rest from a shell outside the event horizon. What is the required mass of the "20-year black hole" in these two cases?

A. We fling an incoming **hailstone** inward along the r-coordinate line with initial shell speed $|v_{\text{far}}|$ from far away from the black hole. A lengthy derivation of the wristwatch time from event horizon to the singularity yields the result:

$$\tau_{\text{Aail}}[2M \to 0] = M\left[\frac{2}{v_{\text{far}}^2\gamma_{\text{far}}} - \frac{1}{v_{\text{far}}^3\gamma_{\text{far}}^3}\ln\left(\frac{1+v_{\text{far}}}{1-v_{\text{far}}}\right)\right] \tag{83}$$

where, remember, $\gamma \equiv (1-v^2)^{-1/2}$ and we treat v_{far} as a (positive) speed. Answer questions in the following items:

a. *Guess:* In the case of the hailstone, will the mass of "20-year black hole" be greater or less than that for the raindrop?

b. Consider $\gamma_{\text{far}} = 2$. What is the value of v_{far}?

c. Show that for this particular value $\gamma_{\text{far}} = 2$, the first term inside the square bracket in (83) alone gives the same result as the raindrop in equation (3).

d. Was your guess in Item **a** correct or incorrect?

e. What is the mass of the "20-year black hole" for that hailstone? How does it compare to the mass of the "20-year black hole" for the raindrop?

B. A **drip** drops from rest on a shell of global coordinate $r_0 > 2M$. Another lengthy derivation of the wristwatch time from event horizon to the singularity yields the result:

$$\tau_{\text{drip}}[2M \to 0] = \tag{84}$$
$$2M\left(\frac{2M}{r_0}\right)^{-3/2}\left[-\left(\frac{2M}{r_0}\right)^{1/2}\left(1-\frac{2M}{r_0}\right)^{1/2} + \arctan\left(\frac{2M/r_0}{1-2M/r_0}\right)^{1/2}\right]$$

a. *Guess:* In the case of a drip, will the mass of "20-year black hole" be greater or less than that for the raindrop?

b. Next, take the limiting case $r_0 \to 2M$. Show that in this limit arctan takes the value $\pi/2$.

c. Show that in this case $\tau_{\text{drip}}[2M \to 0] \to \pi M$.

d. Was your guess in Item **a** correct or incorrect?

e. What is the mass of the "20-year black hole" for that drip? How does it compare to the mass of the "20-year black hole" for the raindrop?

C. *Fascinating but optional:* The next to last paragraph in Box 8 states that every stone that passes inward across the event horizon at $r = 2M$ moves at that r-value with shell velocity $v_{\rm shell} = -1$, the speed of light (as a limiting case). Since this holds for *all* r-diving stones, how can the masses of "20-year black holes" possibly differ for raindrops, hailstones, and drips?

4. Map energy of a drip released from r_0

A. Derive the following expression for E/m in global rain coordinates for a drip released from rest with respect to the local shell frame at $r_0 > 2M$:

$$\frac{E}{m} = \left(1 - \frac{2M}{r_0}\right)^{1/2} \quad \text{(drip released from rest at } r_0\text{)} \qquad (85)$$

Compare with equation (33) of Chapter 6. Are you surprised by what you find? Should you be?

B. What are the maximum and minimum values of E/m in (85) as a function of r_0? How can the minimum value possibly be less than the rest energy m of the stone measured in an inertial frame?

C. Is expression (85) consistent with the value $E/m = 1$ for a raindrop?

D. Is expression (85) valid for $r_0 < 2M$? What is the physical reason for your answer?

E. For $r_0 > 2M$, is expression (85) still valid when that stone arrives inside the event horizon?

5. Map energy of a hailstone

A. Derive the following expression for E/m in global rain coordinates for a hailstone hurled radially inward with speed $v_{\rm far}$ from a shell very far from the black hole.

$$\frac{E}{m} = \gamma_{\rm far} \equiv \left(1 - v_{\rm far}^2\right)^{-1/2} \quad \text{(hailstone)} \qquad (86)$$

Compare this expression with results of Exercise 7 in Chapter 6. Are you surprised by what you find? Should you be?

B. Is expression (86) for the hailstone consistent with expression (85) for the drip? consistent with $E/m = 1$ for a raindrop?

6. Motion of outgoing light flash outside and at the event horizon

Find the maximum value of r/M at which the "outgoing" flash moves to larger r, that is $dr > 0$, at each of these global map velocities:

A. $dr/dT = 0.99$
B. $dr/dT = 0.9$
C. $dr/dT = 0.5$
D. $dr/dT = 0$

7. Motion of the "outgoing" flash inside the event horizon.

Find the value of r/M at which the"outgoing" flash moves to smaller r, that is $dr < 0$, at each of these global map velocities:

A. $dr/dT = -0.1$
B. $dr/dT = -0.5$
C. $dr/dT = -1$
D. $dr/dT = -9$

8. Motion of the incoming flash

At each value of r/M found in Exercises 6 and 7, find the value of dr/dT for the *incoming* flash.

7.11 ■ REFERENCES

Box 1, Eggbeater Spacetime comes in part, from Chapter 13 of *Black Holes and Time Warps* by Kip Thorne. The original description of a spherically symmetric collapse into an uncharged, non-spinning black hole was by J. R. Oppenheimer and H. Snyder, *Physical Review,*Volume 56, pages 455–459 (1939). Vladimir Belinsky, Isaac Markovich Khalatnikov, Evgeny Mikhailovich Lifshitz, *Advances in Physics*, Volume 31, page 639 (1982). Charles W. Misner, *Physical Review Letters*, Volume 22, page 1071.

Wristwatch time from event horizon to singularity: Geraint F. Lewis and Juliana Kwan, "No Way Back: Maximizing survival time below the Schwarzschild event horizon," arXiv:0705.1029v1 [physics.ed-ph] 8 May 2007

Box 8. The River Model: Andrew J. S. Hamilton and Jason P. Lisle, "The river model of black holes," *American Journal of Physics*, Volume 76, Number 6, June 2008, pages 519-532.

Global rain coordinates and the global rain metric: Paul Painlevé, "La mécanique classique et la théorie de la relativité", C. R. Acad. Sci. (Paris)

173, 677-680(1921). Allvar Gullstrand, “Allgemeine Lösung des statischen Einkörperproblems in der Einsteinschen Gravitationstheorie”, Arkiv. Mat. Astron. Fys., Volume 16, Number 8, 1-15 (1922). For a more modern derivation and references, see “Regular coordinate systems for Schwarzschild and other spherical spacetimes,” by Karl Martel and Eric Poisson, *American Journal of Physics,*Volume 69, Number 4, April 2001, pages 476-480.

Comment 6: “Einstein in Paris: Einstein Presents and Discusses His Theory” by Charles Nordmann, translated by Shawna Halevy, *21st Century Science & Technology,* Summer 2011, pages 19-37.

Realization that the apparent singularity at $r = 2M$ in the Schwarzschild metric is “fictional”: G. Lemaître (1933). *Annales de la Société Scientifique de Bruxelles.* Volume **A53** pages 51−85.

Quote on mutable physical law: John Archibald Wheeler with Kenneth Ford, *Geons, Black Holes, and Quantum Foam*, New York, W. W. Norton Company, 1998, page 351.

Download File Name: Ch07InsideTheBlackHole170511v1.pdf

Chapter 8. Circular Orbits

- *How do orbits around a black hole differ from planetary orbits around our Sun?*
- *How close to a black hole can a free stone move in a circular orbit?*
- *Can a stone reach the speed of light in a circular orbit around a black hole?*
- *Can I use a black hole circular orbit to travel forward in time? backward in time?*
- *What is the source of the energy that the so-called QUASAR radiates outward in such prodigious quantity?*

CHAPTER 8

Circular Orbits

Edmund Bertschinger & Edwin F. Taylor *

How happy is the little Stone
That orbits a Black Hole alone*
And doesn't care about Careers
And Exigencies never fears –
Whose Coat of elemental Brown
A passing Universe put on
And independent as the Sun
Associates or glows alone
Fulfilling absolute Decree
In casual simplicity –

—Emily Dickinson

*Line two in the original reads:
That rambles in the Road alone

8.1 ■ STEP OR ORBIT?

"Go straight!" shouts spacetime. The Principle of Maximal Aging interprets that command

Nature shouts at the stone "Go straight!"

A stone in orbit streaks around a black hole—or around Earth. What tells the stone how to move? Spacetime grips the stone, giving it the simplest possible command: "Go straight!" or in the more legalistic language of the Principle of Maximal Aging, "Follow the worldline of maximal aging across the next two adjoining local inertial frames." From instant to instant this directive is enough to tell the stone what to do next, the next step to take in its motion.

This chapter: circular orbits

This command for its next step is sufficient for the stone, but we want more: We seek a description of the entire orbit of the stone through spacetime—its worldline in global coordinates. The present chapter uses the global metric and the Principle of Maximal Aging to predict circular orbits of a stone around any spherically symmetric center of attraction. This prediction

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

Constants of motion: map energy and map angular momentum

uses two map quantities that do not change as the motion progresses: map energy and map angular momentum. In Query 7, Section 7.5, you derived the map energy of a stone in global rain coordinates. Section 8.2 in the present chapter derives an expression for map angular momentum in global rain coordinates. Sections 8.4 shows how to use map angular momentum—together with map energy—to forecast circular orbits. We find that a *free* stone can move (a) in a *stable* circular orbit only at an r-coordinates greater than $r = 6M$, or (b) in an *unstable* circular orbit from $r = 6M$ down to $r = 3M$. No circular orbit for a free stone exists for $r < 3M$.

Comment 1. Global quantities are unicorns
Expressions for global quantities such as map energy and map angular momentum are specific to the global coordinates in which they are expressed. They are unicorns—mythical beasts—unmeasured by a local inertial observer, except by some quirk of the global coordinates (Section 6.3).

The circular orbit is a special case of an *orbit.* We have not yet carefully defined an orbit. Here is that definition.

DEFINITION 1. Orbit
An **orbit** is the path of a free stone through spacetime described by a given set of global coordinates. The path of a radially-plunging stone, with $d\phi = 0$ is a special case of the orbit.

Comment 2. Orbit vs. worldline
The *orbit* of a stone is different from its *worldline*. The worldline of a stone (Definition 9, Section 1.5) is its (free or driven) path through spacetime described by its wristwatch time. The description of a worldline does not require either coordinates or the metric.

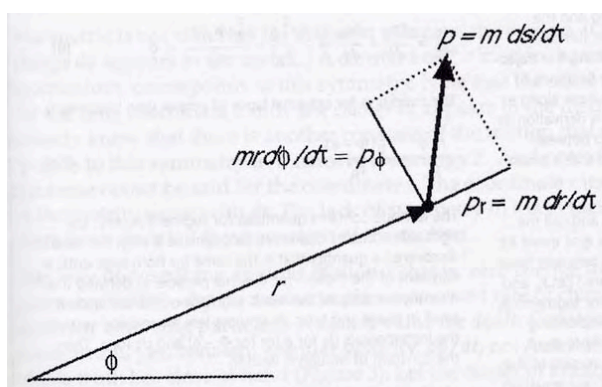


FIGURE 1 In flat spacetime, angular momentum L is the product of r and the ϕ-component of linear momentum $p_\phi = mrd\phi/d\tau$, which yields $L = mr^2 d\phi/d\tau$. Here $d\tau$ is the differential advance of wristwatch time of the stone. Box 1 shows that the same expression, written in global (either Schwarzschild or rain) coordinates, is a constant of motion around a non-spinning black hole.

8.2 ■ MAP ANGULAR MOMENTUM OF A STONE FROM MAXIMAL AGING

Vary the map angle of an intermediate event on a worldline to find map angular momentum.

Here we derive the expression for map angular momentum using global rain coordinates with its T-coordinate. The resulting expression for map angular momentum is also valid in Schwarzschild coordinates. Why? Because both global coordinate systems have the same r and ϕ coordinates, and the global t- or T-coordinate—different in the two global coordinate systems—does not appear in the expression for map angular momentum.

Start with the global rain metric, equation (15) in Section 7.4. Write down its approximation at the average r-coordinate $\bar{r}$:

$$\Delta\tau^2 \approx \left(1 - \frac{2M}{\bar{r}}\right)\Delta T^2 - 2\left(\frac{2M}{\bar{r}}\right)^{1/2}\Delta T\Delta r - \Delta r^2 - \bar{r}^2\Delta\phi^2 \qquad (9)$$

Box 1 uses the now-familiar Principle of Maximal Aging to derive the expression for map angular momentum in global rain coordinates. Box 1 tells us that $r^2 d\phi/d\tau$ is a constant of motion for a free stone moving around the non-spinning black hole. Can we recognize this constant as something familiar? Figure 1 shows that in flat spacetime the angular momentum of the stone (symbol L) has the form $L = mr^2 d\phi/d\tau$. So we identify our new constant of motion as the **map angular momentum per unit mass** of the stone: $L/m = r^2 d\phi/d\tau$.

Map angular momentum

$$\frac{L}{m} \equiv r^2\frac{d\phi}{d\tau} \qquad \text{(map angular momentum)} \qquad (10)$$

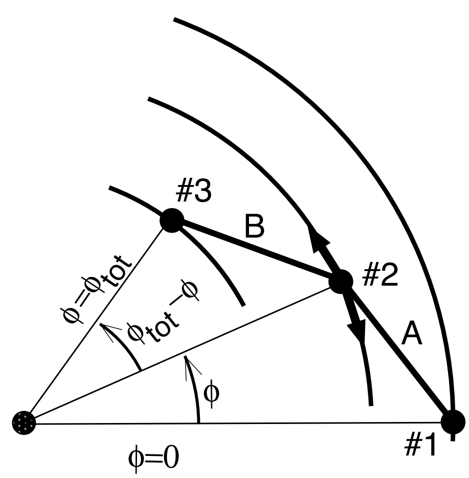


FIGURE 2 [Figure for Box 1.] Derivation of map angular momentum. Find the intermediate map angle ϕ that maximizes the stone's wristwatch time between Events #1 and #3.

Box 1. Derive the Expression for Map Angular Momentum

Strategy: Apply the Principle of Maximal Aging to maximize the wristwatch time of a free stone that moves along two adjoining worldline segments labeled A and B—for Above and Below—in Figure 2. The stone emits flashes at Events #1, #2, and #3 that mark off the segments. Fix the global rain T- and r-coordinates of all three flashes and the ϕ-coordinates of flashes #1 and #3. Vary the ϕ-coordinate of Event #2 by sliding it along a circle (double-headed arrow in Figure 2) to maximize the total wristwatch time between flashes #1 and #3. Then identify the resulting constant of motion as the map angular momentum per unit mass of the stone. Now the details.

Set the fixed ϕ-coordinate of Event #1 equal to zero and call ϕ_{tot} the fixed final ϕ-coordinate for Event #3. To change the angle ϕ of Event #2, move it in either direction along its circle (double-headed arrow in the figure). Let $\bar{r}_{\text{A}}$ and $\bar{r}_{\text{B}}$ be appropriate average values of the r-coordinate for segments A and B, respectively, and let τ_{A} and τ_{B} be the corresponding lapses of wristwatch time of the stone moving along these segments. With these substitutions, and for a small value of τ_{A}, the approximate global rain metric (9) for higher Segment A becomes:

$$\tau_{\text{A}} \approx [-\bar{r}_{\text{A}}^2\phi^2 + (\text{terms without } \phi)]^{1/2} \qquad (1)$$

To prepare for the derivative that leads to maximal aging, take the derivative of this expression with respect to ϕ:

$$\frac{d\tau_{\text{A}}}{d\phi} \approx -\frac{\bar{r}_{\text{A}}^2\phi}{\tau_{\text{A}}} \qquad (2)$$

Similarly for lower Segment B,

$$\tau_{\text{B}} \approx [-\bar{r}_{\text{B}}^2(\phi_{\text{tot}} - \phi)^2 + (\text{terms without } \phi)]^{1/2} \qquad (3)$$

$$\frac{d\tau_{\text{B}}}{d\phi} \approx \frac{\bar{r}_{\text{B}}^2(\phi_{\text{tot}} - \phi)}{\tau_{\text{B}}} \qquad (4)$$

The total wristwatch time for both segments is $\tau = \tau_{\text{A}} + \tau_{\text{B}}$. Take the derivative of this expression with respect to ϕ, substitute from (2) and (4), and set the resulting derivative equal to zero in order to apply the Principle of Maximal Aging:

$$\frac{d\tau}{d\phi} = \frac{d\tau_{\text{A}}}{d\phi} + \frac{d\tau_{\text{B}}}{d\phi} \approx -\frac{\bar{r}_{\text{A}}^2\phi}{\tau_{\text{A}}} + \frac{\bar{r}_{\text{B}}^2(\phi_{\text{tot}} - \phi)}{\tau_{\text{B}}} = 0 \quad (5)$$

The condition for maximal lapse of wristwatch time becomes

$$\frac{\bar{r}_{\text{A}}^2\phi}{\tau_{\text{A}}} \approx \frac{\bar{r}_{\text{B}}^2(\phi_{\text{tot}} - \phi)}{\tau_{\text{B}}} \qquad (6)$$

or in our original Δ notation:

$$\frac{\bar{r}_{\text{A}}^2\Delta\phi_{\text{A}}}{\Delta\tau_{\text{A}}} \approx \frac{\bar{r}_{\text{B}}^2\Delta\phi_{\text{B}}}{\Delta\tau_{\text{B}}} \qquad (7)$$

The left side contains quantities for Segment A only; the right side quantities for Segment B only. We have discovered a quantity that has the same value for both segments, a *global constant of motion* for the free stone across *every* pair of adjoining segments along the worldline of the free stone. In deriving this quantity, we assumed that each segment of the worldline is small. To yield an equality in (7), go to the calculus limit in (7), for which $\bar{r} \rightarrow r$; the constant of motion becomes

$$\lim_{\Delta\tau\rightarrow 0}\left(\bar{r}^2\frac{\Delta\phi}{\Delta\tau}\right) = r^2\frac{d\phi}{d\tau} = \text{a constant of motion} \quad (8)$$

where r and τ are in units of meters. The text identifies this constant of motion as *L/m*, the map angular momentum of the stone per unit mass.

Since r and τ are in units of meters, therefore L/m is also in units of meters.

8.3 ■ EQUATIONS OF MOTION FOR A STONE IN GLOBAL RAIN COORDINATES

The stone's wristwatch ticks off $d\tau$. From $d\tau$ find the resulting changes $d\phi$, dr, and dT.

Equations of motion

We now have in hand the tools needed to calculate the step-by-step advance of the free stone in global rain coordinates. Map energy and map angular momentum—global constants of motion—plus the global metric give us three equations in the three global rain unknowns dT, dr, and $d\phi$, expressed as functions of the advance $d\tau$ of the stone's wristwatch. Starting from an arbitrary initial event, the computer advances wristwatch time and calculates the consequent advance of all three map coordinates, then sums the results of these steps to plot the stone's worldline in global coordinates. We now spell out this process.

First equation of motion

The first equation of motion comes from (10):

$$\frac{d\phi}{d\tau} = \frac{L}{mr^2} \tag{11}$$

The second equation of motion comes from the expression for E/m, equation (35) in Section 7.5:

$$\frac{E}{m} = \left(1 - \frac{2M}{r}\right)\frac{dT}{d\tau} - \left(\frac{2M}{r}\right)^{1/2}\frac{dr}{d\tau} \qquad \text{(global rain coordinates)} \tag{12}$$

Solve (12) for $dT/d\tau$:

$$\frac{dT}{d\tau} = \left[\frac{E}{m} + \left(\frac{2M}{r}\right)^{1/2}\frac{dr}{d\tau}\right]\left(1 - \frac{2M}{r}\right)^{-1} \tag{13}$$

Take the differential limit of (9), divide through by $d\tau^2$, and substitute into it from (11) and (13):

$$1 = \left[\frac{E}{m} + \left(\frac{2M}{r}\right)^{1/2}\frac{dr}{d\tau}\right]^2\left(1 - \frac{2M}{r}\right)^{-1} \tag{14}$$

$$- 2\left(\frac{2M}{r}\right)^{1/2}\frac{dr}{d\tau}\left[\frac{E}{m} + \left(\frac{2M}{r}\right)^{1/2}\frac{dr}{d\tau}\right]\left(1 - \frac{2M}{r}\right)^{-1} - \left(\frac{dr}{d\tau}\right)^2 - \left(\frac{L}{mr}\right)^2$$

Second equation of motiont

Multiply out and collect terms. Solve the resulting quadratic equation in $dr/d\tau$ to yield our second equation of motion for the stone:

$$\frac{dr}{d\tau} = \pm\left[\left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{m^2r^2}\right)\right]^{1/2} \qquad \text{(stone)} \tag{15}$$

Third equation of motion

The third equation of motion shows how dT varies with stone wristwatch time lapse $d\tau$. Substitute for $dr/d\tau$ from (15) into (13) and solve for $dT/d\tau$:

$$\frac{dT}{d\tau} = \left(1 - \frac{2M}{r}\right)^{-1}\left\{\frac{E}{m} \pm \left(\frac{2M}{r}\right)^{1/2}\left[\left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{m^2r^2}\right)\right]^{1/2}\right\} \tag{16}$$

Comment 3. Plotting the orbit

To plot any orbit of the stone—not just a circular orbit— you (or your computer) can integrate the derivative $d\phi/dr = (d\phi/d\tau)(d\tau/dr)$ using equations (11) and (15).

Equations of motion in global rain coordinates

Taken together, equations (11), (15), and (16) are the *equations of motion of the stone* in global rain coordinates. Their integration yields the worldline of the stone in global rain coordinates T, r, and ϕ. Interactive software GRorbits carries out this process, plots the orbit in r and ϕ, and outputs a spreadsheet of events along the worldline of the stone.

QUERY 1. Crossing the event horizon in global rain coordinates.
A first glance at equation (16) might lead to the conclusion that $dT/d\tau$ blows up at the event horizon, so that a stone requires an unlimited lapse in the T-coordinate to cross there. Set $r = 2M(1+\epsilon)$ in this equation to show that as $\epsilon \to 0$ the right side does *not* blow up.

8.4 ■ EFFECTIVE POTENTIAL

Grasp orbit features at a single glance!

Effective potential: the r-component of motion

The orbit computation in Section 8.3 puts into our hands powerful tools to describe any motion of the free stone in the equatorial plane of a spherically symmetric center of attraction. Indeed, the wealth of possible orbits is so great that we need some classification scheme with which to sort orbits at a glance. One classification scheme uses the so-called **effective potential** that focuses on the r-component of motion. Clearer even than our computed orbits, the effective potential plot instantly shows many central features of our stone's motion.

Vicious gravitational effects close to a black hole dominate the effective potential there. In addition to the attractive potential of gravity at large r-coordinates and the effective repulsion due to map angular momentum at intermediate r-values, at still smaller r-coordinates Einstein adds a pit in the potential, shown at the left of Figures 3 and 4.

Pit in the potential

The potential? A pit in this potential? Can we get this potential from principles that are simple, clear, and solid? Yes, starting from map energy and map angular momentum, both of them global constants of motion.

To begin this process, square both sides of (15).

$$\left(\frac{dr}{d\tau}\right)^2 = \left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{m^2r^2}\right) \tag{17}$$

Effective potential for a stone

Define a function $(V_{\rm L}(r)/m)^2$ to replace the second term on the right side of (17). Call this function the square of the **effective potential**.

$$\left(\frac{V_{\rm L}(r)}{m}\right)^2 \equiv \left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{m^2r^2}\right) \qquad \text{(squared effective potential)} \tag{18}$$

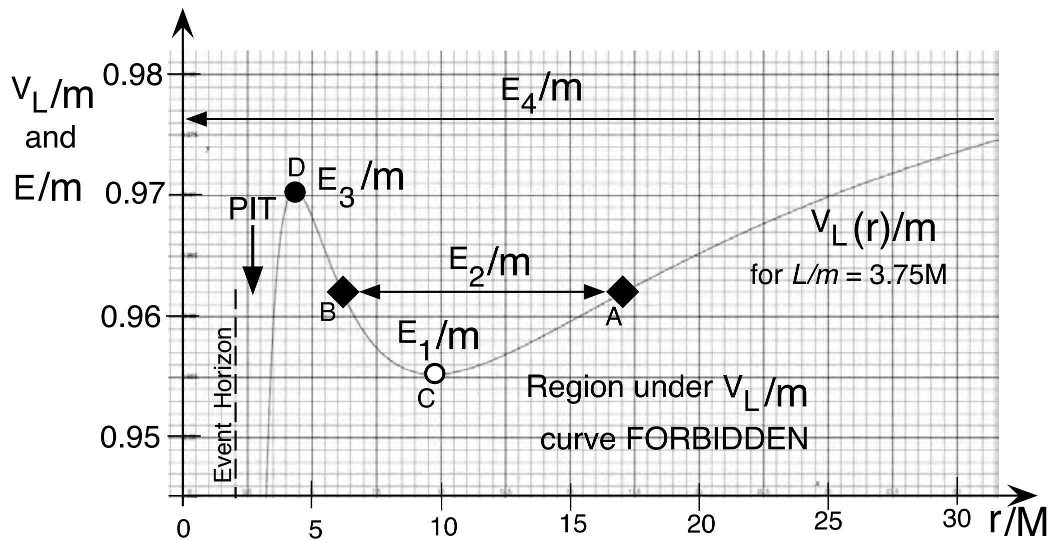


FIGURE 3 Effective potential for a stone that orbits the black hole with map angular momentum $L/m = 3.75M$. When the stone's map energy equals the minimum of the effective potential energy (little open circle at C), the stone is in a stable circular orbit. A stone with somewhat greater map energy, E_2/m, (line with double arrow) oscillates back and forth in r between turning points (little black rotated squares) labeled A and B. When the stone's map energy equals the maximum of the effective potential energy (little filled circle at D), the stone is in an unstable circular orbit. When the map energy E_4/m of an inward-moving stone is greater than the peak of the effective potential (upper horizontal line), the approaching stone crosses the event horizon and plunges to the singularity at $r \to 0$.

Subscript L on $V_\mathrm{L}(r)$ reminds us that this effective potential is different for different values of the map angular momentum L. Substitute (18) into (17) and take the square root of both sides:

$$\frac{dr}{d\tau} = \pm\left[\left(\frac{E}{m}\right)^2 - \left(\frac{V_\mathrm{L}(r)}{m}\right)^2\right]^{1/2} \qquad (19)$$

The squared effective potential $(V_\mathrm{L}(r)/m)^2$ is what we subtract from the squared map energy term $(E/m)^2$ to obtain $(dr/d\tau)^2$. The plus sign in (19) describes increase in r-coordinate, the minus sign describes decreasing r.

Figure 3 plots effective potential $V_\mathrm{L}(r)/m$ from (18) and shows the r-range for motion of stones with three different map energies.

Forbidden map energy region; turning points

Note that $dr/d\tau$ in equation (19) is real only where $(E/m)^2$ has a value greater than $(V_\mathrm{L}(r)/m)^2$. This has important consequences: The stone cannot exist with a map energy in the region under the effective potential curve: that is the **forbidden map energy region**. As a result, the horizontal map energy line labeled E_2/m in Figure 3 terminates wherever it meets the $V_L(r)/m$ curve. At these points, called **turning points** in r, the map energy and the effective potential are equal: $E/m = V_\mathrm{L}/m$, so that $dr/d\tau = 0$ in (19). At a turning point the r-component of map motion goes to zero (while the stone continues to sweep around in the ϕ-direction). In Figure 3 the stone's r map position oscillates back and forth between turning points in r labeled A and B. Earth and each solar planet oscillates back and forth with an r-component of motion similar to that labeled E_2/m in Figure 3, each around a minimum of its own solar effective potential that depends on its map angular momentum.

Definition: **Forbidden map energy region**

DEFINITION 2. Forbidden map energy region

The **forbidden map energy region** is a region in a $V_{\mathrm{L}}(r)/m$ vs. r/M plot in which equations of motion of the stone (Section 8.3) become imaginary or complex. Hence the stone cannot move—or even exist—with map energy in the forbidden map energy region.

QUERY 2. Demonstrate forbidden map energy regions

Verify the statement in Definition 2 that "In the forbidden map energy region, the equations of motion of a stone (Section 8.3) become imaginary or complex." for *each* equation of motion in Section 8.3.

DEFINITION 3. Turning point, circle point, and bounce point

Figures 3 and 4 show little filled circles, little open circles, and little rotated filled squares, each one located on the effective potential curve. These points are called *turning points*. (Section 8.5 defines the meaning of the "half-black" circle numbered one in Figure 4.)

Definition: **Turning point**

A **turning point** is a value of r for which $E = V_{\mathrm{L}}(r)$. At a turning point, $dr/d\tau = 0$. Examples of turning points: points A through D in Figure 3 and points 1 through 5 in Figure 4. We distinguish two kinds of turning points: circle point and bounce point:

Definition: **Circle point**

A **circle point** is a turning point at a maximum or minimum of the effective potential. At a circle point $dr/d\tau$ equals zero and remains zero, at least temporarily, so a stone at a circle point is in either an unstable or a stable circular orbit. We plot a circle point as either a little filled circle (at an unstable circular orbit r-value) or a little open circle (at a stable circular orbit r-value). Examples of bounce points: C and D in Figure 3 and points labeled 1 through 5 in Figure 4.

Definition: **Bounce point**

A **bounce point** is a turning point that is *not* at a maximum or minimum of the effective potential. At a bounce point, $dr/d\tau$ for a free stone reverses sign. We plot a bounce point as a little filled rotated square. Examples of bounce points: A, and B in Figure 3. A stone that moves between bounce points—such as the stone with map energy E_2/m in Figure 3, is in a bound orbit that is *not* circular (Chapter 9).

Three payoffs of effective potential

Here are four important payoffs of the effective potential. First, it gives $dr/d\tau$ in terms of E, L, and r. Second, at every r it shows us the map energy region that is forbidden to the stone. Third, it fixes r-values of the turning points for given E and L. Fourth, and most important, it helps us to categorize—at a glance—different kinds of orbits, including circular orbits.

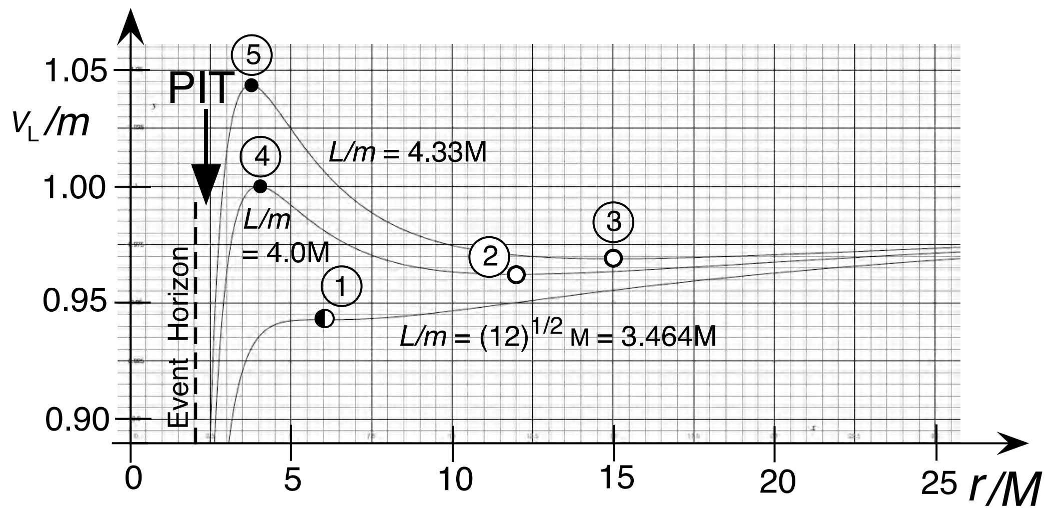


FIGURE 4 The r-coordinates of stable and unstable (knife-edge) circular orbits at points of zero slope of the effective potentials for three values of L/m. Unstable circular orbits (little filled circles numbered 4 and 5) lie between $r = 3M$ and $r = 6M$. Stable circular orbits, little open circles numbered 2 and 3, lie at r greater than $r = 6M$. Orbit numbered 1 (little half-black circle) is the limiting case, stable for increase in r; unstable for decrease in r. Section 8.5 discusses this "half-stable orbit." A forbidden map energy region (Definition 2) lies under the curve for each value of L/m.

QUERY 3. Compare Newtonian and general-relativistic orbital motion *(optional)*
The right side of (17) tells us a great deal about the difference between the stone's global motion described in global rain coordinates and its motion described by Newton.

A. Multiply out the right side of (17) and divide through by 2 to yield

$$\frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 = \frac{1}{2}\left[\left(\frac{E}{m}\right)^2 - 1\right] - \left(-\frac{M}{r} + \frac{L^2}{2m^2r^2} - \frac{ML^2}{m^2r^3}\right) \qquad \text{(global rain coordinates)} \quad (20)$$

B. Newton's expression for angular momentum, with Newton's "universal time t" is:

$$\frac{L}{m} \equiv r^2\frac{d\phi}{dt} \qquad \text{(Newton, universal time } t\text{)} \quad (21)$$

Show that Newton's expression for the square of the velocity of the stone is:

$$v^2 = \left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\phi}{dt}\right)^2 = \left(\frac{dr}{dt}\right)^2 + \frac{L^2}{m^2r^2} \qquad \text{(Newton)} \quad (22)$$

C. Now, Newton's expression for gravitational potential energy per unit mass (chosen to go to zero far from the center of attraction) is $U(r) = -M/r$. Write down Newton's conservation of energy equation, solve it for the radial velocity, and show the result:

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 = \frac{E}{m} - \left(-\frac{M}{r} + \frac{L^2}{2m^2r^2}\right) = \frac{E}{m} - \frac{V_{\rm NewtL}(r)}{m} \qquad \text{(Newton)} \qquad (23)$$

where **Newton's effective potential** is $V_{\rm NewtL}(r)/m$.

D. Sketch for the Newtonian case a diagram like that of Figure 3: a plot of $V_{\rm NewtL}(r)$ with horizontal lines for different values of E. Describe the resulting orbits and contrast them to those for motion in curved spacetime.

Of course the general relativity expression (20) is not just another version of Newton's equation (23). But look at the basic similarity of the right sides of these two equations: a constant term from which we subtract a function of the r-coordinate—the "effective potential"—that varies with the value of map angular momentum L.

Conclusion of this analysis: It is the negative third term in the effective potential on the right side of (20), with r^3 in its denominator, that drives the effective potential downward as r becomes smaller as it approaches the event horizon—thereby creating the PIT in the potential labeled in Figures 3 and 4. This third term is the child of spacetime curvature.

In a stable circular orbit the stone's map energy rests at the minimum of the effective potential; the stone rides round and round the black hole without changing r-coordinate.

Stable orbit at effective potential minimum

DEFINITION 4. Stable circular orbit

A stone in a stable circular orbit has map energy E/m equal to the *minimum* of the effective potential $V_{\rm L}(r)/m$, for example the map energy labeled 1 in Figure 3 and energies labeled 2 and 3 in Figure 4. Any incremental change in the r-coordinate at constant E/m puts the stone into the forbidden map energy region under the effective potential curve, where a stone cannot go.

We use a little open circle to locate a stable circular orbit on an effective potential energy curve. The point labeled 1 in Figure 4 is the stable circular orbit of minimum r-value analyzed in Section 8.5.

Einstein opens up a second set of r-coordinates where the effective potential also has zero slope, illustrated by point D in Figure 3 and points 4 and 5 in Figure 4. Each of these is a *maximum* of the effective potential curve; at this r-coordinate the stone experiences no tendency to move either to larger or smaller r-coordinate, so will stay at the same r-coordinate, riding round and round the black hole at constant r-coordinate. We call these **unstable** or **knife-edge** circular orbits, because slight departure from the knife-edge r-coordinate leads to decisive motion either to larger r, or else—horrors!—to smaller r that leads to the event horizon.

DEFINITION 5. Unstable (or knife-edge) circular orbit

A stone in an unstable (or knife-edge) circular orbit has map energy

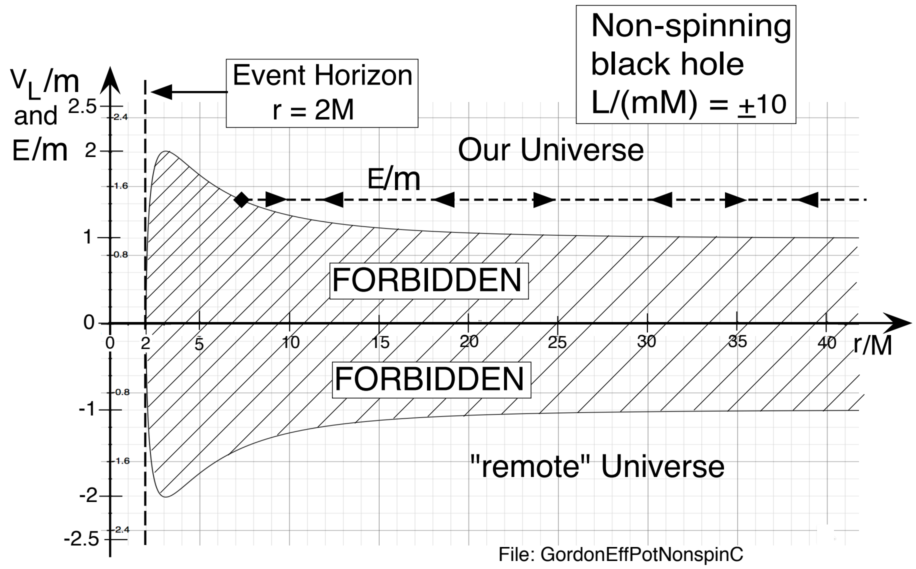


FIGURE 5 Plot of equation (17) for stone map angular momentum $L/(mM) = \pm 10$. In our Universe map energy is positive; it is negative in a "remote" Universe below the forbidden region. We cannot travel between our Universe and the "remote Universe" because the worldlines that connect them must pass inward through the event horizon, then back out again. (Diagonal lines emphasize impenetrability.) So where *is* this "remote" Universe? See Chapter 21.

Unstable (or knife-edge) orbit at effective potential maximum

E/m equal to the *maximum* of the effective potential $V_{\rm L}/m$, so that any incremental r-displacement in either direction puts the stone into a region with a gap between E/m and $V_{\rm L}/m$ such that this displacement increases.

We use a little filled circle to locate an unstable circular orbit on an effective potential energy curve.

Comment 4. How long on a knife edge?
Suppose that our spaceship is in a knife-edge orbit, technically an *unstable orbit*. Slight cosmic wind, firing of a projectile, or ejection of the day's trash may give our spaceship a tiny r-motion. Once displacement from the effective potential peak occurs, the slope of the effective potential urges the spaceship farther away from the point of zero slope, either outward toward larger r-coordinate or inward toward the event horizon. Sooner or later—who knows when?—a stone inevitably falls off the effective potential maximum of an unstable circular orbit.

"Why, oh why," our captain cries, "didn't I carry along a booster rocket? A tiny rocket boost to push us outward could have reversed our initially slow inward motion and allowed us to escape. But now it's too late!"

Strange results follow from equation (19), which requires that $(E/m)^2 \geq (V_{\rm L}/m)^2$ in order that $dr/d\tau$ be real. A consequence of this condition is that either $E/m \geq +V_{\rm L}/m$ or $E/m \leq -V_{\rm L}/m$. Figure 5 shows this condition. A stone cannot move, or even exist, with E/m in the region $+V_{\mathrm{L}}/m > E/m > -V_{\mathrm{L}}/m$. This is a forbidden map energy region, because $dr/d\tau$ would be imaginary there. *Result:* The forbidden map energy region divides spacetime outside the event horizon into two isolated regions: one for positive map energy and the other for negative map energy. The stone cannot travel directly between them. This definition of a forbidden map energy region is consistent with that given in Definition 2.

Figure 3 shows only positive values of map E/m. This is the region we live in, where we carry out our measurements and observations, the upper region of positive map energy in Figure 5. What is the meaning of negative E/m in the lower region of Figure 5? Can we carry out measurements and observations there? Remember that map energy is a global map quantity, not a quantity that we can measure; its negative value tells us nothing about permitted measurements. In the exercises you show that we can construct local inertial frames in the negative map energy region, so we can carry out measurements and observations there, just as we can in the region above the forbidden map energy region.

Can we travel from the upper (positive map energy) region in Figure 5 to the lower (negative map energy) region? Our own worldline, just like the worldline of a stone, cannot pass directly through that forbidden map energy region. Figure 5 shows that the forbidden map energy region ends at the event horizon, $r = 2M$. Can we make an end run around the forbidden map energy region by moving in through the event horizon and back out again? No, sorry: Once inside the event horizon, we cannot come out again; instead we move relentlessly inward to the singularity. See exercise 11 in Section 8.7.

?

Objection 1. *Can light move between the upper and lower regions?*

!

Nope. Figure 11 in Section 11.8 shows that a corresponding forbidden region for light separates upper and lower regions. Both for stones and for light, the two regions are physically isolated.

?

Objection 2. *Wait! Where is this lower region? It has the same r-values as the upper region but you tell me that it lies "somewhere else," in a negative map energy region we cannot reach. Where is it?*

!

The answer is subtle and deep. Later we will understand that global rain coordinates do not include all of spacetime. We must find other global coordinates that include such regions. Chapter 21 treats these matters. Keep on reading!

Chapters 17 through 21 examine the spinning black hole. We will find that for the spinning black hole we may be able to travel between the corresponding upper and lower regions by dropping through the event horizon from the upper region, using rocket thrusts while inside the event horizon, then emerging outward through the event horizon into the lower region. Luc Longtin summarizes: "The non-spinning black hole is like the spinning black hole, but with its gate to other universes closed. For the spinning black hole, the gate is ajar." (initial quote, Chapter 21)

8.5 ■ PROPERTIES OF CIRCULAR ORBITS

Details! Details!

A series of Queries helps you to explore some properties of circular orbits in the everyday positive map energy region around the non-spinning black hole.

QUERY 4. Map r-values of circular orbits

A. A circular orbit is possible at every r-coordinate where the effective potential has zero slope. Take the r-derivative of both sides of (18) for a fixed L/m, set this derivative equal to zero, and show the following result:

$$r^2 - \frac{L^2}{Mm^2}r + 3\frac{L^2}{m^2} = 0 \qquad \text{(circular orbit)} \tag{24}$$

B. Equation (24) is linear in $(L/m)^2$. Solve it to find:

$$\left(\frac{L}{m}\right)^2 = \frac{Mr^2}{r - 3M} \qquad \text{(circular orbit, } r > 3M\text{)} \tag{25}$$

Note that this expression is valid for both stable and unstable circular orbits and is invalid for $r < 3M$, where L/m would be imaginary. Circular orbits cannot exist for $r < 3M$, and for $r = 3M$ the circular orbit is a limiting case (Item D in Query 8)

B. Equation (24) is quadratic in r. Solve it to find:

$$r = \frac{L^2}{2m^2M}\left[1 \pm \left(1 - \frac{12M^2m^2}{L^2}\right)^{1/2}\right] \qquad \text{(circular orbit, } r > 3M\text{)} \tag{26}$$

Refer to Figure 4. Make the argument that the + sign in (26) corresponds to the minimum of the effective potential, that is to a stable circular orbit; and that the − sign corresponds to the maximum of the effective potential, that is to the unstable (knife-edge) circular orbit.

C. *Optional:* Take the second derivative of (18) and verify that the ± signs in (26) correspond, respectively, to a minimum and maximum of the effective potential.

Look more closely at equation (26) and the effective potential curve in Figure 4 with the "half-black" little circle labeled number 1. In order for the r-coordinate to be real, the square root expression in (26) must be real. This occurs only when $|L/m| \geq (12)^{1/2}M = 3.4641M$. You can show that for the minimum map angular momentum, the global r-coordinate of the circular orbit is $r = 6M$. This is called the **innermost stable circular orbit** and is located at $r_{\text{ISCO}} = 6M$.

Definition
ISCO

DEFINITION 6. Innermost stable circular orbit (ISCO)
The **innermost stable circular orbit (ISCO)**, located at $r_{\text{ISCO}} = 6M$, divides r-values for unstable circular orbit in the region $3M < r < 6M$ from r-values for stable circular orbits in the region $r > 6M$. We can call the ISCO "half stable:" An increase in r at the same map energy puts the stone into a forbidden map energy region (like a stable circular orbit); a decrease in r at the same map energy puts the stone into a legal map energy region (like an unstable circular orbit).

Section 8.6 describes a so-called *toy model* of a quasar, the brightest steady source of light in the heavens. This emission comes from the loss of map energy of a stone that enters a circular orbit at large r and tumbles down through a series of "stable" circular orbits of smaller and smaller r. When the stone reaches the innermost stable circular orbit and continues to lose map energy, it spirals inward across the event horizon, after which we can no longer detect its radiation.

QUERY 5. Shell speed of a stone in a circular orbit

Compute the speed of the stone in a circular orbit measured by a shell observer, as follows.

A. Consider two ticks of the orbiting stone's clock, separated by wristwatch time $\Delta\tau$ and by zero distance measured in the stone's local frame, but separated by shell time Δt_{shell} and by shell distance $\Delta x_{\text{shell}} = \bar{r}\Delta\phi$. The relation between Δt_{shell} and $\Delta\tau$ is just the special relativity expression

$$\Delta t_{\text{shell}} = \gamma_{\text{shell}}\Delta\tau = (1 - v_{\text{shell}}^2)^{-1/2}\Delta\tau \tag{27}$$

where γ_{shell} has an obvious definition. From the value of map angular momentum, we can use (27) to calculate shell speed:

$$\begin{aligned} v_{\text{shell}} &= \lim_{\Delta t_{\text{shell}}\to 0}\left(\frac{\bar{r}\Delta\phi}{\Delta t_{\text{shell}}}\right) = (1 - v_{\text{shell}}^2)^{1/2}\ \frac{r^{\,2}d\phi}{r d\tau} \\ &= (1 - v_{\text{shell}}^2)^{1/2}\ \frac{L}{mr} \qquad (\phi - \text{motion}) \end{aligned} \tag{28}$$

From this equation, show that

$$v_{\text{shell}}^2 = \left[1 + \left(\frac{mr}{L}\right)^2\right]^{-1} \qquad (\phi - \text{motion}) \tag{29}$$

From equation (25) show that

$$\left(\frac{mr}{L}\right)^2 = \frac{r}{M} - 3 \qquad \text{(circular orbit)} \tag{30}$$

Substitute this into (29) to find

$$v_{\text{shell}}^2 = \frac{M}{r-2M} = \left(\frac{M}{r}\right)\left(1 - \frac{2M}{r}\right)^{-1} \qquad \text{(circular orbit, } r > 3M\text{)} \tag{31}$$

Equation (31) is valid for both stable and unstable (knife-edge) circular orbits.

B. What is the value of the shell speed v_{shell} in the ISCO, the innermost stable circular orbit at $r = 6M$?

C. Verify that the minimum map r-coordinate for a circular orbit is $r = 3M$. (*Hint:* What is the upper limit of the shell speed of a stone?)

D. From (25) show that, as a limiting case, the map angular momentum L/m increases without limit for the knife-edge circular orbit of minimum r-coordinate.

Circular orbit of light

Comment 5. Unlimited map angular momentum?

How can the map angular momentum possibly increase indefinitely (Item D of Query 6)? It does so only as a limiting case. According to (10), the map angular momentum is equal to $L/m = r^2 d\phi/d\tau$. The relation between wristwatch time $d\tau$ and shell time dt_{shell} is given by (27), the usual time-stretch formula of special relativity. As the stone's speed approaches the speed of light, the advance of wristwatch time becomes smaller and smaller compared with the advance of shell time. In the limit, it takes zero wristwatch time for the stone to circulate once around the black hole. Because $d\tau$ is in the denominator of the expression for angular momentum, the map angular momentum L/m increases without limit. The speed of light is the limiting speed of a stone, so the speed-of-light orbit is a limiting case, reached by a stone only after an unlimited lapse of the Schwarzschild t-coordinate. This limiting case tells us, however, that light can travel in a circular orbit at $r = 3M$ (Chapter 11).

QUERY 6. Global map energy of a stone in circular orbit

Find an expression for map energy E/m in global rain coordinates for the stone in a circular orbit, as follows:

A. Use (25) and (15) with $dr = 0$ for a circular orbit. Show that the result is:

$$\frac{E}{m} = \frac{r-2M}{r^{1/2}(r-3M)^{1/2}} \qquad \text{(circular orbit, } r > 3M\text{)} \tag{32}$$

B. Does (32) go to the values you expect in three cases: Case 1: $r \gg M$? Case 2: $r \to 3M^+$ (r decreases from above)? Case 3: $r < 3M$?

QUERY 7. Map energy and map angular momentum of a stone in the ISCO

A. Show that the map angular momentum of the ISCO is $L_{\text{ISCO}}/(mM) = 3.464\ 101\ 615$.

B. Show that the map energy of the ISCO is $E_{\text{ISCO}}/m = 0.942\ 809\ 042$.

QUERY 8. Shell energy of a stone in a circular orbit

A. Use the special relativity relation $E_{\text{shell}}/m = (1 - v_{\text{shell}}^2)^{-1/2}$ for the local shell frame plus (31) for v_{shell}^2 to show that

$$\boxed{\frac{E_{\text{shell}}}{m} = \left(\frac{r-2M}{r-3M}\right)^{1/2} \qquad \text{(circular orbit, } r > 3M\text{)}} \tag{33}$$

B. From (32) and (33), verify that

$$\frac{E_{\text{shell}}}{m} = \left(1 - \frac{2M}{r}\right)^{-1/2} \frac{E}{m} \qquad \text{(circular orbit } r > 2M\text{)} \tag{34}$$

This agrees with equation (12) in Section 6.3 for a diving stone.

C. Far from the black hole, that is for $r \gg M$, set $\epsilon = M/r$. Use our standard approximation (inside the front cover) to show that at large r-coordinate equation (33) becomes:

$$\frac{E_{\text{shell}}}{m} \approx 1 + \frac{M}{2r} \qquad \text{(circular orbit, } r \gg M\text{)} \tag{35}$$

D. Take (31) to the same limit and show that (35) becomes:

$$E_{\text{shell}} \approx m + \frac{1}{2} m v_{\text{shell}}^2 \qquad \text{(circular orbit, } r \gg M\text{)} \tag{36}$$

Would Newton be happy with your result? Would Einstein?

QUERY 9. Orbiter wristwatch time for one circular orbit

A. From (25) and (11) verify the following wristwatch time for one circular orbit ($\Delta\phi = 2\pi$),

$$\frac{\Delta\tau}{M} = \frac{2\pi(r/M)^2}{L/(mM)} = 2\pi\frac{r}{M}\left(\frac{r-3M}{M}\right)^{1/2} \qquad \text{(one circular orbit)} \qquad (37)$$

B. Explain why $\Delta\tau \to 0$ as $r \to 3M$.

C. For a black hole with $M = 10M_{\text{Sun}}$, find the wristwatch time in seconds for one circular orbit for the three values $r/M = 10, 6, 4$.

D. For a non-spinning black hole of mass $M \approx 4 \times 10^6 M_{\text{Sun}}$ equal to the black hole at the center of our galaxy, find the wristwatch time in seconds for one circular orbit for the three values $r/M = 10, 6, 4$.

E. *Optional:* Solve (37) for $(r/M - 3)$ and put $r \approx 3M$ in the expression on the right side of your result. Find the value of $(r/M - 3)$ when $\tau = 1$ microsecond for a black hole of mass $M = 10M_{\text{Sun}}$. What is the numerical value of the observed distance $2\pi r$ around this circumference in meters—a directly-measurable distance (Section 3.3). So now we have an astronaut who traverses this large, measurable circumference in a microsecond. To do this, she must move at many times the speed of light. Can this be right? Explain your answer.

QUERY 10. Shell time for one circular orbit

Verify the following expressions for the periods of one circular orbit.

A. From equations (27), (31), and (37), show that the local shell time for one circular orbit is:

$$\Delta t_{\text{shell}} = 2\pi r\left(\frac{r-2M}{M}\right)^{1/2} \qquad \text{(one circular orbit)} \qquad (38)$$

For the minimum (knife-edge) orbit, with $r = 3M$, explain why the shell period is equal to the circumference of the orbit.

B. For a circular orbit of very large r-coordinate, explain why global rain ΔT, shell Δt_{shell}, and orbiter wristwatch time $\Delta\tau$ all have the same value for one orbit, namely $2\pi r^{3/2}/M^{1/2}$.

8.6 ■ TOY MODEL OF A QUASAR

Beacon of the heavens

Quasar

A **quasar** ("quasi-stellar object") is an astronomical object that pours out electromagnetic radiation of many frequencies at a prodigious rate. The quasar is the brightest steady source of light in the heavens, so we can see it farther away than any other steady source. At the center of a quasar is, almost certainly, a spinning black hole (Chapters 17 through 21), but here we make a first quick model of a quasar using a non-spinning black hole. This sort of rough, preliminary analysis is called a **toy model**.

Toy model

Accretion disk

Radiating away change in map energy

A simple model of quasar emission postulates an **accretion disk**, a gas disk that swirls around the black hole in its equatorial plane. Interactions among the molecules and atoms in this gas cloud are complicated. We assume simply that interactions among neighboring atoms and ions heats the accretion disk to high temperature and that the resulting electromagnetic emission is what we observe far from the quasar. The radiated energy we observe comes from the change in orbital map energy of each atom as it moves sequentially from a large-r circular orbit to smaller-r circular orbits with smaller and smaller map energy. We also assume that significant map energy change takes place over many orbits, so sequential orbits are nearly circular, each with its nearly constant values of E/m and L/m.

Comment 6. Losing map angular momentum

Transport map angular momentum outward

What is the mechanism of this orbit change? The sequence of circled numbers 3-2-1 in Figure 4 shows that "our circulating atom" decreases both its map energy and its map angular momentum as it occupies a set of circular orbits of decreasing r-values. Map angular momentum of an isolated system is conserved, so the lost map angular momentum of our circulating atom must be transported outward, away from the black hole. What mechanism can transport map angular momentum outward? Equation (31) tells us that atoms in adjacent circular orbits have slightly different shell speeds, with atoms in the higher orbit moving more slowly. One might think (incorrectly) that *friction* between our atom and atoms in a higher orbit increases the velocity—and therefore the map angular momentum—of atoms in the higher orbit, and so on outward. However, direct friction turns out to be far too small to account for the outward transport of map angular momentum. The mechanism may depend on our model that the accretion disk consists of highly ionized atoms, a plasma, threaded with magnetic field lines. Magnetic fields greatly increase interactions between ions, so might account for the outward transport of map angular momentum in a quasar. We simply do not know.

Innermost stable circular orbit

Eventually our atom's circular orbit drops to $r = 6M$, the innermost stable circular orbit (Definition 6 in Section 8.5). At this point our atom continues to lose map angular momentum, so that it drops out of the last stable circular orbit and spirals inward across the event horizon. Once our atom crosses the event horizon, any further radiation moves only inward and cannot reach us, the external observers.

QUERY 11. Map energy given up by "our atom."

The prodigious radiation we observe from quasars is all emitted before orbiting atoms and ions cross the event horizon.

A. Start with an atom in a circular orbit at large r-coordinate, moving slowly so its initial map energy is approximately equal to its mass, $E/m \approx 1$, from (32). Now think of its map energy later, as the atom moves in the stable circular orbit of minimum r-coordinate, $r = 6M$. Using (32), find the map energy E/m of the atom in this minimum-r circular orbit to three significant digits. How much map energy has the atom given up during the process of dropping gradually from large r-coordinate to the smallest stable circular orbit? [My answers: $E_{\text{final}} = 0.943m$ so $\Delta E = 0.057m$.]

B. Suppose that the atom emits as electromagnetic radiation all the map energy it gives up (from Item A) as it spirals down to the circular orbit at $r = 6M$. Show that the map energy of that total amount of radiation emitted is $\Delta E = 0.057m$. Since initially we had $E/m = 1$, therefore 0.057, or 5.7%, is also the fraction of initial map energy that is radiated as the atom spirals inward to the lowest stable circular orbit.

Measure map energy at far from the black hole

Map energy E/m is a constant of motion, independent of position. Suppose that the map energy radiated by the atom during its descent finds its way outward. Then the same map energy ΔE arrives at the distant r-coordinate from which the atom departed earlier with $E/m \approx 1$. Moreover, very far from the black hole spacetime is flat; so map energy is equal to shell energy there, equation (34). Therefore the group of shell frame observers far from the black hole see—can in principle measure—a total radiated energy of $\Delta E = 0.057m$, which is 5.7 percent of the stone's initial map energy.

Comment 7. How much emitted energy?
No nuclear reaction on Earth—except particle-antiparticle annihilation—releases as much as one percent of the rest energy of its constituents. Chapter 18 shows that for a black hole of maximum spin, the fraction of initial mass radiated away by a stone that spirals down from a large r-coordinate to an innermost stable circular orbit is 42 percent of its rest energy. No wonder quasars are such bright beacons in the heavens!

Rate of emitted radiation

Now let our atom drop into the black hole from the innermost stable circular orbit at $r = 6M$. How much does the mass of the black hole increase? Equation (28) in Section 6.5 says that the total mass of the black hole increases by the map energy E/m of the object falling into it. This allows us to connect the rate of increase of the mass of a quasar and its brightness to the rate at which it is swallowing matter from outside. Let dm/dT be the rate at which mass falls into the black hole from far away and dM/dT be the rate at which the mass of the black hole increases. Then Item B in Query 11 tells us that the rate of radiated energy is

$$\text{Rate of radiated energy} \approx 0.057\,\frac{dm}{dT} \qquad (\,dm = \text{mass falling in}) \tag{39}$$

so that the mass M of the black hole increases at the rate:

$$\frac{dM}{dT} = (1 - 0.057)\frac{dm}{dT} = 0.943\frac{dm}{dT} \qquad (M = \text{ mass of black hole}) \tag{40}$$

QUERY 12. Power output of a quasar
During every Earth-year, a distant quasar swallows $m = 10M_{\text{Sun}}$ = ten times the mass of our Sun. Recall that watts equals joules/second and, from special relativity, $\Delta E[\text{joules}] = \Delta m[\text{kilograms}]c^2[\text{meters}^2/\text{second}^2]$.

A. How many watts of radiation does this quasar emit, according to our toy model?
B. Our Sun emits radiation at the rate of approximately 4×10^{26} watts. The quasar is how many times as bright as our Sun?
C. Compare your answer in Item B to the total radiation output of a galaxy, approximately 10^{11} Sun-like stars.

QUERY 13. How long does a quasar shine?
We see most quasars with large redshifts of their light, which means they began emission not long after the Big Bang, about 14×10^9 years ago. A typical quasar is powered by a black hole of mass less than 10^9 solar masses. Explain, from the results of Query 12, what this says about the lifetime during which the typical quasar shines.

?

Objection 3. *We have talked about t and T global coordinates and different kinds of local times near a black hole: shell time, diver time, orbiter time. Is it possible for me to travel to a black hole and use it, in some way, to live longer than I can live here on Earth?"*

!

As with many profound questions, the answer is both "Yes," and "No." With or without a black hole, you may live longer—on your wristwatch—between any two events than your twin—on her wristwatch—who takes a different worldline through spacetime between these two events (Twin "Paradox," Section 1.6). However, you cannot escape the iron rule that *your aging is identical to the total time lapse on your wristwatch,* no matter where you travel or at what rates you move along the way. When your wristwatch says 100 years after birth, you have aged 100 years. This is the total time lapse that you *experience.* In this sense, relativity does not provide a way for you to burst the bonds of human aging. Sorry!

8.7 ■ EXERCISES

1. Newtonian and relativistic equations of motion

The general-relativistic equation of motion of a stone (20) looks drastically different from the corresponding Newtonian equation of motion (23). For example, E is squared in (20) but has the first power in (23). Show that in the limit of $r \gg M$ and slow motion of the stone in the local inertial frame, the two equations of motion, (20) and (23) become the same, provided that the symbol E has *different* definitions in the two equations.

2. Shell time for one orbit

An observer in a circular orbit at a given map r-coordinate moves at speed v_{shell} past the shell observer. Equation (31) gives the value of this shell speed.

Query 9 gives the wristwatch time for one orbit. What is the shell time for one orbit?

A. Show that this shell time for one orbit is

$$\frac{\Delta t_{\text{shell}}}{M} = \frac{2\pi r/M}{v_{\text{shell}}} = 2\pi \frac{r}{M}\left(\frac{r-2M}{M}\right)^{1/2} \qquad \text{(one circular orbit)} \quad (41)$$

(*Hint:* Recall the definition in Section 3.3 of r—the "reduced circumference"—as the measured circumference of a concentric shell divided by 2π.)

B. Compare Δt_{shell} for one orbit in (41) with $\Delta\tau_{\text{shell}}$ for one orbit from (37). Which is longer at a given r-value? Give a simple explanation.

C. What is the map angular momentum L of the orbiter, written as r times an expression involving v_{shell}? (The answer is *not* mrv_{shell}.)

D. The text leading up to Definition 4 in Section 8.5 shows that the smallest r-coordinate for a stable circular orbit is $r = 6M$; equation (31) determines that in this orbit the orbiter's shell speed $v_{\text{shell}} = 0.5$, half the speed of light. Assume the central attractor to be Black Hole Alpha, with $M = 5000$ meters. The following equation gives, to one significant digit, the values of some measurable quantities for the innermost stable circular orbit. Find these values to three significant digits.

$$\begin{aligned} \Delta t_{\text{shell}} &\approx 4 \times 10^5 \text{ meters} && \text{(shell time for one orbit)} \quad (42) \\ \Delta\tau_{\text{orbiter}} &\approx 3 \times 10^5 \text{ meters} && \text{(wristwatch time for one orbit)} \\ L/m &\approx 2 \times 10^4 \text{ meters} && \end{aligned}$$

E. The orbiter of Item D completes one circuit of the black hole in approximately one millisecond on her wristwatch. If you ignore tidal effects, does this extremely fast rotation produce *physical discomfort* for the orbiter? If she closes her eyes, does she get dizzy as she orbits?

3. When are Newton's Circular Orbits Almost Correct?

Your analysis of the Global Positioning System (GPS) in Chapter 4 calculated values of r-coordinate and orbital speed of a GPS satellite in circular orbit using Newton's mechanics, with the prediction that the general relativistic analysis gives essentially the same values of r-coordinate and speed for this application. Under what circumstances are circular orbits predicted by Newton indistinguishable from circular orbits predicted by Einstein? Answer this question using the following outline or some other method.

A. Find Newton's expression similar to equation (26) for the r-coordinate of a stable circular orbit, starting with equation (23).

B. Recast equation (26) for the general-relativistic prediction of r for stable orbits in the form

$$r = r_{\text{Newt}}(1-\epsilon) \tag{43}$$

where r_{Newt} is the r-coordinate of the orbit predicted by Newton and ϵ is the small fractional deviation of the orbit from Newton's prediction. This expression neglects differences between the Newtonian and relativistic values of L when expressed in the same units. Use the approximation inside the front cover to derive a simple algebraic expression for ϵ as a function of r_{Newt}.

C. Set your expression for ϵ equal to 0.001 as a criterion for good-enough equality of the r-coordinate according to both Newton and Einstein. Find an expression for r_{min}, the smallest value of the r-coordinate for which this approximation is valid.

D. Find a numerical value for r_{min} in meters for our Sun. Compare the value of r_{min} with the r-coordinate of the Sun's surface.

E. What is the value of ϵ for the r-coordinate of the orbit of the planet Mercury, whose orbit has an average r-coordinate 0.387 times that of Earth?

F. What is the value of ϵ for the r-coordinate of a 12-hour orbit of GPS satellites around Earth?

4. Map ΔT for one orbit

Convert lapse of wristwatch time $\Delta\tau$ for one circular orbit from (37) to lapse ΔT for one circular orbit using the following outline or some other method:

A. Show that for a circular orbit, equation (13) becomes:

$$\frac{\Delta T}{\Delta\tau}(\text{one orbit}) = \frac{E}{m}\left(1-\frac{2M}{r}\right)^{-1} = \frac{E}{m}\frac{r}{(r-2M)} \tag{44}$$

B. Into this equation, substitute for E/m from (32) to obtain

$$\frac{\Delta T}{\Delta\tau}(\text{one orbit}) = \left(\frac{r}{r-3M}\right)^{1/2} \tag{45}$$

C. Use this result plus (37) to show that

$$\Delta T(\text{one orbit}) = \Delta\tau\frac{\Delta T}{\Delta\tau} = 2\pi\frac{r^{3/2}}{M^{1/2}} \tag{46}$$

Does any observer measure this lapse Δt for one orbit?

5. Kepler's Laws of Planetary Motion

Johannes Kepler (1571-1630) provided a milestone in the history of astronomy: his **Three Laws of Planetary Motion**, deduced from a huge stack of planetary observations made by his mentor Tycho Brahe (1546-1601) and expressed in our notation.

1. A planet orbits around the Sun in an elliptical orbit with the Sun at one focus of the ellipse.
2. The r-coordinate vector from the Sun to the planet sweeps out equal areas in equal lapses of T-coordinate.
3. The square of the period of the planet is proportional to the cube of the planet's mean r-coordinate from the Sun.

A. Show by a simple symmetry argument that Kepler's Second Law describes circular orbits around a black hole.

B. From equation (46) show that Kepler's Third Law is also valid for *circular* orbits around a black hole (when expressed in global rain coordinates).

C. Kepler's Third Law is sometimes called the **1-2-3 Law** from the exponents in the following equation. Use equation (46) to show that for circular orbits, in our regular notation using meters,

$$M \equiv M^1 = \omega^2 r^3 \tag{47}$$

where $\omega \equiv 2\pi/\Delta T$, with ΔT for one orbit.

Comment 8. Is Kepler's First Law Valid?
Figure 4 in Section 9.3 shows that Kepler's First Law is definitely *not* valid for non-circular orbits near a non-spinning black hole. Chapter 11 shows that the orbit of the planet Mercury differs *slightly* from the planetary orbit analyzed by Newton. The predicted value of this deviation of Mercury's orbit was an early validation of Einstein's general relativity.

6. Longest Life Inside the event Horizon

Objection 12 in Section 7.8 asked, "Can I increase my lifetime inside the event horizon by blasting rockets in either ϕ direction to add a ϕ-component to my global velocity?" You are now able to answer this question using your new knowledge of map angular momentum. Suppose that you ride on a stone that moves between the event horizon and the singularity.

A. What equation in the present chapter leads to the following expression for your wristwatch lifetime inside the horizon?

$$\tau\,[2M \to 0] = \int_0^{2M} \left[\left(\frac{E}{m}\right)^2 + \left(\frac{2M}{r} - 1\right)\left(1 + \frac{L^2}{m^2 r^2}\right)\right]^{-1/2} dr \tag{48}$$

Note, first, that the square-bracket expression on the right side of (48) is in the denominator of the integrand. Second, note that this equation describes any motion of the observer whatsoever, free-fall or not. Free-fall motion has constant E and L. For motion that is not free-fall, the value of E or L (or both) can change along the worldline of the stone.

B. Can any non-zero value of L along your worldline increase your wristwatch lifetime inside the event horizon?

C. What value of E gives you the maximum wristwatch lifetime inside the event horizon?

D. By what practical maneuvers can you achieve the value of E determined in Item C?

E. Show that the maximum value of wristwatch time from the event horizon to the singularity is πM meters. *Hint:* Make the substitution $(r/2M)^{1/2} = \sin\theta$.

F. Chapter 7 found the mass of a "20-year black hole" for a raindrop. Find the numerical value of (*fraction*) in the following equation:

$$\begin{aligned}&(\text{mass of "20-year black hole" in Item E}) \\ &= (fraction) \times (\text{mass of "20-year black hole" for a raindrop})\end{aligned} \tag{49}$$

7. Forward Time Travel Using a *Stable* Circular Orbit

You are on a panel of experts asked to evaluate a proposal from the Space Administration to "travel forward in time" using the difference in rates between a clock in a stable circular orbit around a black hole and our clocks remote from the black hole. Give your advice about the feasibility of the scheme, based on the following analysis or one of your own.

A. Consider two sequential ticks of the clock of a satellite in a stable circular orbit around a black hole. Use a result of Exercise 1 to show that

$$\frac{\Delta\tau_{\text{orbiter}}}{\Delta T} = \left(\frac{r-3M}{r}\right)^{1/2} \tag{50}$$

B. What is the value of the ratio $\Delta\tau_{\text{orbiter}}/\Delta T$ in the stable circular orbit of smallest r-coordinate, $r = 6M$?

C. What rocket speed in flat spacetime gives the same ratio of rocket clock time to "laboratory" time as the stable circular orbit of smallest r-coordinate?

D. Does the proposed time travel method require rocket fuel to put the rocket in orbit and to escape the black hole?

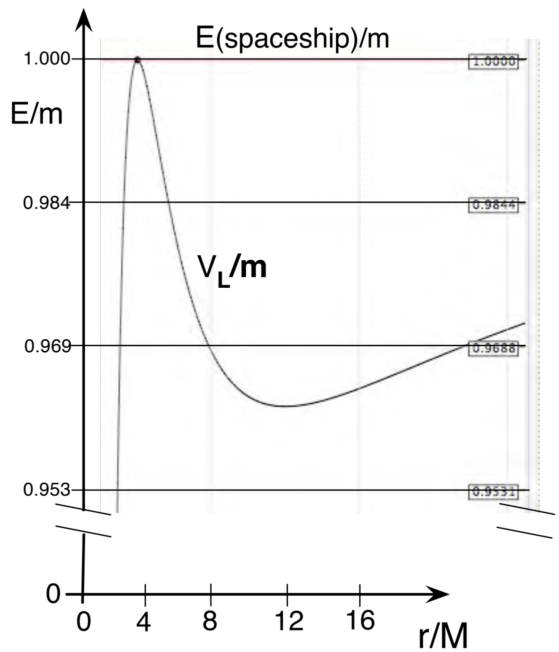


FIGURE 6 Insertion into a knife-edge orbit at $r = 4M$ with map energy $E/m \approx 1$, equal to that of a spaceship moving slowly at large r-coordinate in a direction chosen to give it the value of L/m required to establish the peak value for V_{L}/m.

E. Based on this analysis, do you recommend in favor of—or against—the Space Administration's proposal for forward time travel using stable circular orbits around a black hole?

8. Forward Time Travel Using a *Knife-Edge* Circular Orbit

Whatever your own vote on the forward time travel proposal of Exercise 6, the majority on your panel rejects the proposal because it requires extra rocket thrust for insertion into and extraction from the circular orbit at $r = 6M$. The Space Administration returns with a new proposal that uses a knife-edge circular orbit, assuming that an automatic device can fire small rockets to balance the satellite safely on the knife-edge of the effective potential. The Space Administration notes that such an orbit can be set up to require *very small* rocket burns, both for insertion into and extraction from a knife-edge circular orbit. As an example, they present Figure 6 for the case of nonrelativistic distant velocity, so that the map energy of the satellite is $E/m \approx 1$. While still far from the black hole, the spaceship captain uses rockets to achieve the value of L required so that $V_{\rm L}(r)/m = E/m = 1$ on the peak shown in Figure 6. They boast that the time stretch factor is increased enormously by high satellite shell speed in the knife-edge orbit without the need for rocket burns to achieve that speed.

A. The condition shown in Figure 6 means that $V_{\rm L}(r)/m = 1$ at the peak of the effective potential (18). This equation plus equation (26) are two equations in the two unknowns r and L. Solve them to find $r = 4M$ and $L/m = 4M$. *Optional:* Describe in words how the commander of the spaceship sets the desired value of L while still far away, without changing the remote non-relativistic speed $v_{\rm far}$.
B. What is the factor $d\tau/dt_{\rm shell}$ for the spaceship in this orbit? What speed in flat spacetime gives the same time-stretch ratio?
C. Does the spaceship require a significant rocket burn to leave its knife-edge circular orbit and return to a remote position? What will be its shell speed at that distant location?

9. "Free" data-collection orbit

After its long interstellar trip, the spaceship approaches the black hole at relativistic speed, that is $E/m > 1$. The commander does not want to use a rocket burn to change spaceship map energy, but rather only its direction of motion (hence its map angular momentum) to enter a knife-edge circular orbit with the same map energy it already has.

A. Draw a figure similar to Figure 6 for this case.
B. Show that the astronauts can find a knife-edge circular orbit on which to perch, no matter how large the incoming far-away speed with respect to the black hole.

Once in an unstable circular orbit, small rocket thrusts keep the spaceship balanced at the peak of the effective potential. After they finish collecting data, the astronauts push-off outward and return toward home base at the same speed at which they approached, even if this speed is relativistic. In summary, once launched toward a black hole the explorers need little rocket power to go into an unstable circular orbit, to balance in that orbit while they study the black hole, then to return home. Further details in Chapter 9.

10. Nandor Bokor disproves relativity.

Nandor Bokor looks at Exercise 1 and shouts, "Aha, now I can disprove relativity!" Parts A through D below are steps in Nandor's reasoning, not separate questions to be answered. Resolve Nandor's disproof without criticizing him.

A. Nandor Bokor says, "Before I begin my disproof of relativity, recall that we have always had a choice about the shell frame. *First choice:* In order to be inertial, the local shell frame must be in free fall. In this case we drop the local shell frame from rest as we begin the experiment and must complete the experiment so quickly that the shell frame's r-coordinate changes a negligible amount. *Second choice:* The local shell frame is at rest and therefore has a local gravitational acceleration. In that case we must complete our experiment or observation so quickly that local gravity does not affect the outcome. Usually our choice does not change the experimental result, but I am being super-careful here and will take the first choice, so that shell and orbiter frames are both inertial.

B. "Assume, then, that the shell frame is inertial," Nandor continues. "Equation (42) says that during one revolution of the orbiter its measured time lapse is $\Delta t_{\text{orbiter}} \approx 3 \times 10^5$ meters, while the measured shell clock time lapse is $\Delta t_{\text{shell}} \approx 4 \times 10^5$ meters. Note that these are both observed readings—measurements—and they are *different.* When the orbiter returns after one orbit the two inertial frames—orbiter and shell—overlap again.

C. "Now we have two truly equivalent inertial reference frames that overlap twice so we can compare their clock readings directly. (This is different from special relativity, in which one of the two frames—in the Twin Paradox, Section 1.6—is not inertial during their entire separation.) In the present orbiting case, neither observer can tell which of the two inertial frames s/he is in from inside his or her inertial frame."

D. Nandor concludes, "You tell me, Dude, which of the two equivalent inertial clocks—the orbiter's frame clock or the shell observer's frame clock—runs slow compared with the clock in the other frame. You can't! Equation (42) claims a difference where no difference is possible. Good-bye relativity!"

11. Equations of motion in Schwarzschild global coordinates

Start with the Schwazschild metric, equation (5) in Section 3.1, and show that equations (11), are (15) are the same in both global coordinate systems, but (16) takes the simpler form:

$$\frac{dt}{d\tau} = \left(1 - \frac{2M}{r}\right)^{-1} \frac{E}{m} \qquad \text{(stone, Schwarzschild)} \qquad (51)$$

Comment 9. Why not Schwarzschild?
Why don't we take advantage of the simpler equation (51) by using Schwarzschild coordinates to describe the motion of the free stone? Because we already know—equation (22) in Section 6.4—that neither light nor a stone moves inward through the event horizon in a finite lapse of the Schwarzschild t-coordinate. In theory, Schwarzschild coordinates would not cause a problem with circular orbits in the present chapter because these orbits exist only outside the event horizon—indeed, only in the region $r > 3M$. But Chapter 9 treats more general trajectories of a stone, some of which move inward across the event horizon.

12. Life under the forbidden map energy region

If we could find some way to travel from our normal upper, positive map energy region in Figure 5 to the lower, negative map energy region (which extends outward far from the black hole), could we live a normal life there? What does "normal life" mean? We reduce "normal life" to essentials: that the equations of motion for a stone are real! Limit attention to motion outside the event horizon:

A. Show that the first two equations of motion (11) and (15) are the same for E/m under the forbidden region as for E/m above the forbidden region.
B. Show that the third equation of motion (16) tells us that $dT/d\tau$ is negative under the forbidden region, so that global T runs backward along the worldline of the stone. But T is a unicorn, not a measured quantity, so the third equation of motion is also valid under the forbidden region.

Where are we when we are under the forbidden map energy region in Figure 5? This is our first hint that our everyday lives may not have access to all regions of spacetime. Alice had it right: Wonderland—and black holes—become "curiouser and curiouser."

8.8 ■ REFERENCES

Initial Emily Dickinson poem from R. W. Franklin, *The Poems of Emily Dickinson, Variorum Edition* 1998, The Belknap Press of Harvard

University. This poem is variation E of the poem with Franklin number 1570, written about 1882. Reprinted and modified with permission of Harvard University.

GRorbits interactive software program that displays orbits of a stone and light flash is available at `http://stuleja.org/grorbits/`

Last sentence of the final exercise: *Alice in Wonderland* by Lewis Carroll, first sentence of Chapter 2.

Download File Name: Ch08CircularOrbits170511v1.pdf

Chapter 9. Orbiting the Black Hole

- *As I approach a black hole from far away, how can I put my spaceship into a circular orbit?*
- *How can I transfer from one circular orbit to another one?*
- *Why am I uncomfortable in some orbits near a black hole?*
- *Can I enter a circular orbit without firing a rocket?*
- *How do I move a probe from a circular orbit inward across the event horizon?*

Download File Name: Ch09Orbiting160404v1.pdf

CHAPTER 9

Orbiting the Black Hole

Edmund Bertschinger & Edwin F. Taylor *

I want to know how God created this world. I am not interested in this or that phenomenon, in the spectrum of this or that element. I want to know his thoughts. The rest are details.

****

What really interests me is whether God could have created the world any differently; in other words, whether the requirement of logical simplicity admits a margin of freedom.

—Albert Einstein

9.1 ■ OBSERVE THE BLACK HOLE FROM A SEQUENCE OF CIRCULAR ORBITS

The sequence of orbits in our exploration plan

Observe the black hole from circular orbits.

Chapter 8 introduced circular orbits of a free stone around a black hole. The present chapter describes how the captain of an approaching spaceship can insert it into a circular orbit, then transfer to progressively smaller circular orbits in order to get closer looks at the black hole. Our exploration program includes several maneuvers:

EXPLORATION PROGRAM FOR THE BLACK HOLE

Exploration program

Step 1. Insert the approaching spaceship into a stable circular orbit at $r = 20M$.

Step 2. Transfer a probe from this initial orbit to the innermost stable circular orbit at $r_{\mathrm{ISCO}} = 6M$.

Step 3. Transfer the probe from the ISCO to an *unstable* circular orbit at $r = 4M$.

Step 4. Tip the probe off the unstable circular orbit at $r = 4M$ so that it spirals inward across the event horizon.

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

To describe this sequence of orbits, use equations from previous chapters, summarized here in global rain coordinates, T, r, ϕ. Both the unpowered spaceship and the unpowered probe move in the same way as a free stone.

Free motion

GENERAL FREE MOTION OF UNPOWERED SPACESHIP OR PROBE

$$\frac{E}{m} \equiv \left(1 - \frac{2M}{r}\right)\frac{dT}{d\tau} - \left(\frac{2M}{r}\right)^{1/2}\frac{dr}{d\tau} \qquad \text{(free: (35) in Sec. 7.5)} \qquad (1)$$

$$\frac{L}{m} \equiv r^2\frac{d\phi}{d\tau} \qquad \text{(free: (10) in Sec. 8.2)} \qquad (2)$$

$$\frac{E_{\text{shell}}}{m} = \frac{1}{(1 - v_{\text{shell}}^2)^{1/2}} = \frac{E/m}{\left(1 - \dfrac{2M}{r}\right)^{1/2}} \qquad \text{(free: (17) in Sec. 6.3)} \qquad (3)$$

$$\left(\frac{dr}{d\tau}\right)^2 = \left(\frac{E}{m}\right)^2 - \left(\frac{V_{\text{L}}(r)}{m}\right)^2 \qquad \text{(free: (21) in Sec. 8.4)} \qquad (4)$$

$$\left(\frac{V_{\text{L}}(r)}{m}\right)^2 \equiv \left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{m^2r^2}\right) \qquad \text{(free: (20) in Sec. 8.4)} \qquad (5)$$

Motion in a circular orbit

CIRCULAR-ORBIT MOTION OF UNPOWERED SPACESHIP OR PROBE ($r > 3M$)

$$\left(\frac{L}{m}\right)^2 = \frac{Mr^2}{r - 3M} \qquad \text{(circle: (28) in Sec. 8.5)} \qquad (6)$$

$$r = \frac{L^2}{2m^2M}\left[1 \pm \left(1 - \frac{12m^2M^2}{L^2}\right)^{1/2}\right] \qquad \text{(circle: (27) in Sec. 8.5)} \qquad (7)$$

$$\frac{E}{m} = \frac{r - 2M}{\left[r\,(r - 3M)\right]^{1/2}} \qquad \text{(circle: (34) in Sec. 8.5)} \qquad (8)$$

$$\frac{E_{\text{shell}}}{m} = \left(\frac{r - 2M}{r - 3M}\right)^{1/2} \qquad \text{(circle: (35) in Sec. 8.5)} \qquad (9)$$

$$v_{\text{shell}}^2 = \frac{M}{r - 2M} \qquad \text{(circle: (33) in Sec. 8.5)} \qquad (10)$$

Figure 1 previews some kinds of orbits we discuss in this chapter.

9.2 ■ INSERT THE APPROACHING SPACESHIP INTO A CIRCULAR ORBIT

Approach from far away and enter a circular orbit.

Insert into a circular orbit.

How does the captain insert her approaching spaceship into an initial circular orbit from which to observe the black hole? Here's one possible method: While still far from the black hole, the captain uses speed- and direction-changing

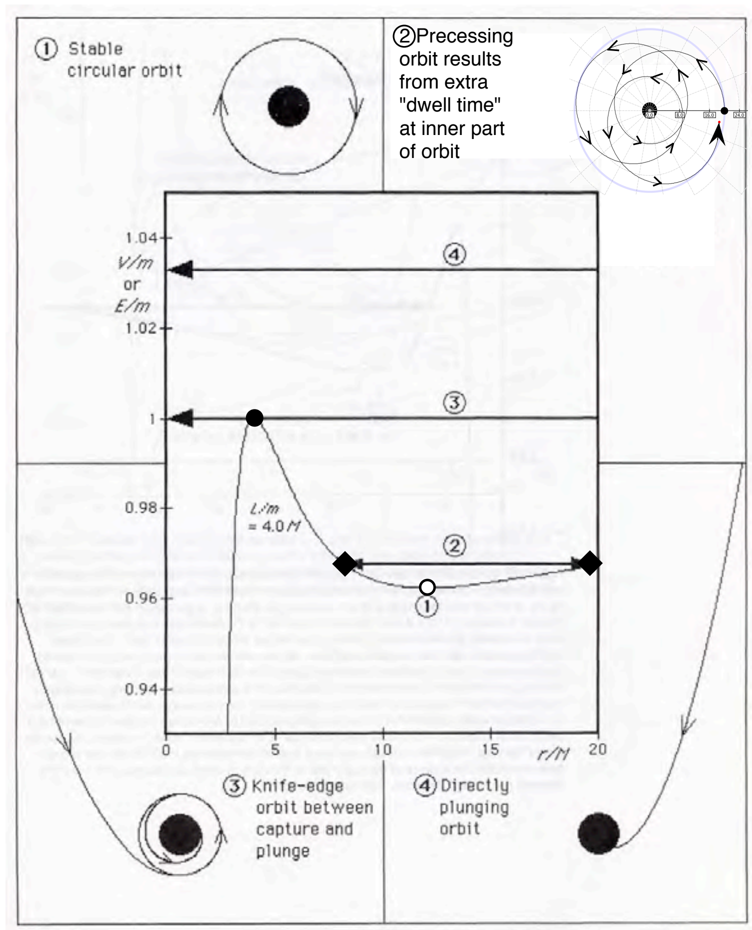


FIGURE 1 Preview: Some kinds of orbits discussed in this chapter, shown here for a single value of map angular momentum L/m but several different values of map energy E/m. A glance at the central plot allows us to make quick predictions about the motion of a stone that orbits or is captured by a black hole. Four different energies numbered on this plot correspond to orbits that appear in the four outer corners of the figure. Adapted from Misner, Thorne, and Wheeler.

rocket thrusts to put the spaceship into a free-fall insertion orbit whose minimum r-value matches that of the desired circular orbit (Figure 2). At that minimum, when the spaceship moves tangentially for an instant, the captain fires a rocket to slow down the spaceship to the tangential speed of the stable circular orbit at that r.

Insertion orbit

With what values of map E/m and L/m will an unpowered spaceship approaching from far away end up moving tangentially for an instant at the desired r-coordinate? To find out, substitute (5) into (4), set $dr/d\tau = 0$, and solve the resulting equation for L/m:

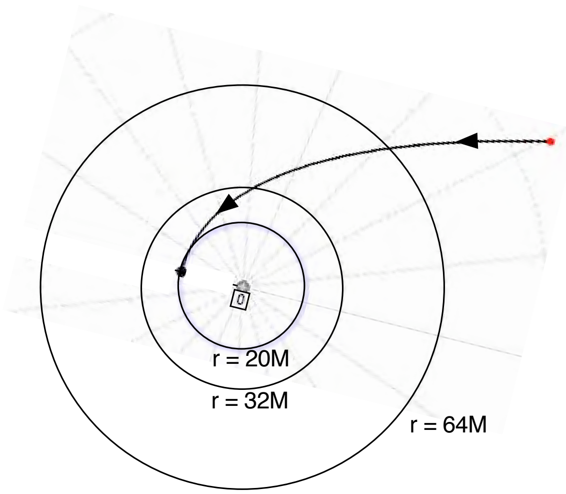


FIGURE 2 Insertion orbit for unpowered spaceship that approaches from far away. At the instant of tangential motion at $r = 20M$, the spaceship fires a tangential rocket thrust to reduce the locally-measured shell velocity to that for a circular orbit (Figure 3).

$$\frac{L}{m} = \pm r^2 \left[\frac{(E/m)^2}{1-(2M/r)} - 1\right]^{1/2} \qquad \text{(tangential motion)} \qquad (11)$$

The $\pm$ sign in (11) distinguishes between two possible directions of motion at the r-value in equation (11). We choose positive angular momentum—that is, in the counterclockwise direction of increasing ϕ. Equation (11) is valid whe $dr/d\tau = 0$, including turning points of all orbits as well as everywhere along a circular orbit.

Choose circular orbit at $r = 20M$.

The captain chooses her circular orbit at $r = 20M$. While still far from the black hole, she maneuvers the incoming spaceship to move with arbitrarily-chosen map energy $E/m = 1.001$ and the positive value of L/m that results from equation (11)—both entered in Table 1. Then she turns off the rockets and lets the spaceship coast. Figure 2 shows the resulting orbit, which corresponds to the incoming horizontal arrow at $E/m = 1.001$ in Figure 3.

Definitions: Subscripts in Table 1

DEFINITION 1. Subscripts in Table 1

Here are definitions of the subscripts in the left-hand column of Table 1. All definitions describe the motion of a free stone or unpowered spaceship or unpowered probe.

insert: for free motion from far away to instantaneous tangential motion at r

circle: for free motion in a circular orbit at r

transfer: for free motion that is instantaneously tangential at both values of r

shell: for values measured in the local inertial frame at r

TABLE 9.1 Numerical values at $r = 20M$ and $r_{\text{ISCO}} = 6M$

Values of	$r = 20M$	$r_{\text{ISCO}} = 6M$
$(L/m)_{\text{insert}}$	$6.733\,036\,31M$	———
$(E/m)_{\text{insert}}$	1.001	———
$v_{\text{x,shell,insert}}$	0.319 056 897	———
$(L/m)_{\text{circle}}$	$4.850\,712\,50M$	$3.464\,101\,62M$
$(E/m)_{\text{circle}}$	0.976 187 060	0.942 809 042
$v_{\text{x,shell,circle}}$	0.235 702 260	0.5
$(L/m)_{\text{transfer}}$	$3.787\,166\,42M$	$3.787\,166\,42M$
$(E/m)_{\text{transfer}}$	0.965 541 773	0.965 541 773
$v_{\text{x,shell,transfer}}$	0.186 052 102	0.266 880 257

NOTE: All shell velocities in this table are tangential, in the positive shell x-direction.

Long numbers in tables

Comment 1. Significant digits
In this chapter we analyze several unstable (knife-edge) circular orbits. Interactive software, such as GRorbits, requires accurate inputs to display the orbit of an unpowered probe that stays in an unstable circular orbit for more than one revolution. To avoid clutter, we put numbers with many significant digits into tables.

Comment 2. Long subscripts
In Table 1 the symbols $v_{\text{x,shell,insert}}$, $v_{\text{x,shell,circle}}$, and $v_{\text{x,shell,transfer}}$ have long, ungainly subscripts. We need long subscripts to fully describe these velocity components: that they are x-components measured in a local shell frame and whether they describe insertion speed into a circular orbit, speed in that circular orbit, or transfer between circular orbits.

Impulse rocket thrusts

Comment 3. Impulse rocket thrusts
We assume that each change in vehicle speed results from a quick rocket thrust, an impulse. In practice there is no hurry; some efficient rocket engines provide low thrust, which carries the vehicle through a series of intermediate orbits. To analyze the outcome of a slow burn complicates calculations and does not add to our understanding. So our vehicles use quick rocket thrusts to transfer from one orbit to another.

Comment 4. Which direction is the "rocket thrust"?
What is the meaning of the phrase *outward rocket thrust*? The rocket fires in one direction; the probe or spaceship that carries the rocket changes speed in the opposite direction. We define *outward rocket thrust* to mean that the rocket burn tends to move the rocket to larger r. Similarly, the *inward rocket thrust* tends to move the rocket to smaller r.

Insert into circular orbit

When the spaceship moves tangentially for an instant at $r = 20M$, the spaceship fires a tangential rocket thrust to put it into the stable circular orbit at that r. What change in tangential velocity must this rocket thrust provide? Tangential velocity in *which* frame? Our policy: make every measurement in a local inertial frame; for that purpose, choose the local *shell* frame. Box 2 in Section 7.4 gives shell frame coordinates from which we derive shell components of velocity. For reasons that will become apparent, we start with

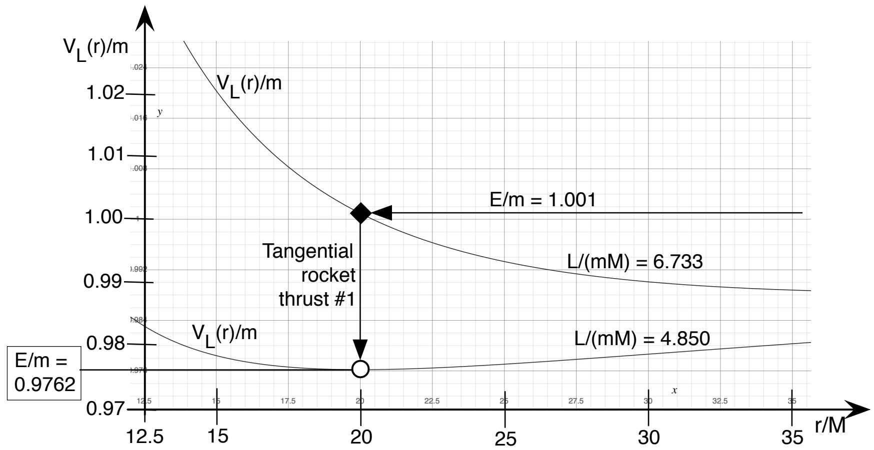


FIGURE 3 At the instant when the incoming spaceship moves tangentially at the radial turning point $r = 20M$ (Figure 2), it fires tangential rocket thrust #1 that changes its map energy and map angular momentum to insert it into a stable circular orbit.

definitions of $dt_{\text{shell}}/d\tau$, $dy_{\text{shell}}/d\tau$, and $dx_{\text{shell}}/d\tau$, each with wristwatch time differential $d\tau$ in the denominator.

$$\frac{dt_{\text{shell}}}{d\tau} = \lim_{\Delta\tau \to 0} \frac{\Delta t_{\text{shell}}}{\Delta\tau} \tag{12}$$

$$= \left(1 - \frac{2M}{r}\right)^{-1/2} \left[\left(1 - \frac{2M}{r}\right)\frac{dT}{d\tau} - \left(\frac{2M}{r}\right)^{1/2} \frac{dr}{d\tau}\right] \tag{13}$$

$$= \left(1 - \frac{2M}{r}\right)^{-1/2} \frac{E}{m} \tag{14}$$

The last step uses equation (1). Similarly:

$$\frac{dy_{\text{shell}}}{d\tau} = \lim_{\Delta\tau \to 0} \frac{\Delta y_{\text{shell}}}{\Delta\tau} = \left(1 - \frac{2M}{r}\right)^{-1/2} \frac{dr}{d\tau} \tag{15}$$

To find an expression for $dr/d\tau$ in this equation, combine equations (4) and (5):

$$\frac{dr}{d\tau} = \pm\left[\left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{m^2r^2}\right)\right]^{1/2} \tag{16}$$

And finally:

$$\frac{dx_{\text{shell}}}{d\tau} = \lim_{\Delta\tau \to 0} \frac{\Delta x_{\text{shell}}}{\Delta\tau} = r\frac{d\phi}{d\tau} = \frac{L}{mr} \tag{17}$$

Shell velocity components

The last step uses equation (2). To complete the derivation of shell velocity components, note, for example, that $v_{\text{y,shell}} = (dy_{\text{shell}}/d\tau)(d\tau/dt_{\text{shell}})$, so from (15) and (14):

$$v_{\rm y,shell} = \frac{dr/d\tau}{E/m} = \pm\left[1 - \left(\frac{E}{m}\right)^{-2}\left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{m^2r^2}\right)\right]^{1/2} \quad (18)$$

$$v_{\rm x,shell} = \left(1 - \frac{2M}{r}\right)^{1/2}\frac{L}{rE} \quad (19)$$

Use the first two entries in Table 1 plus equation (19) to calculate the value of $v_{\rm x,shell,insert}$ at $r = 20M$ (where the shell y-component $v_{\rm y,shell,insert} = 0$) and check the result in the third line of Table 1.

QUERY 1. Tangential shell velocity in a circular orbit

A. What is the tangential shell velocity of the spaceship in the circular orbit at r? Combine equations (6) and (8) to find an expression for L/E and substitute the result into (19):

$$v_{\rm shell,circle} = \left(\frac{M}{r}\right)^{1/2}\left(1 - \frac{2M}{r}\right)^{-1/2} \qquad \text{(circular orbit, } r > 3M) \quad (20)$$

B. Show that your derivation is not valid unless $r > 3M$.

C. Use (20) to calculate a value for $v_{\rm shell,circle}$ at $r = 20M$. Check your answer with the entry in Table 1.

Table 1 tells us that the shell frame velocity $v_{\rm x,shell,insert}$ of the spaceship in its insertion orbit is greater than its shell frame velocity $v_{\rm x,shell,circle}$ in the circular orbit. Therefore a rocket thrust must bring the spaceship's shell velocity down to that of the circular orbit.

Use velocity addition laws.

Einstein shouts, "Look out! To calculate the needed change in spaceship velocity to be provided by the rocket thrust, you do *not* use the difference between $v_{\rm x,shell,insert}$ and $v_{\rm x,shell,circle}$." Why not? Because in special relativity (which rules in every local inertial frame), velocities do not simply add or subtract.

In what local inertial frame can we measure directly the change in velocity provided by the rocket thrust? That would be the local inertial frame in which the spaceship is initially at rest just before the thrust. Just before the rocket thrust, the spaceship moves at velocity $v_{\rm x,shell,insert}$ in the shell frame. We call the local inertial frame in which the spaceship is at rest the **instantaneous initial rest frame** or IIRF.

Definition: Instanteous Initial Rest Frame (IIRF)

DEFINITION 2. Instantaneous Initial Rest Frame (IIRF)
The instantaneous initial rest frame (IIRF) is the local inertial frame in which a rocket is at rest just before it fires a rocket thrust to change its velocity with respect to that frame. We use the subscript IIRF to indicate

TABLE 9.2 Rocket Thrusts in Instantaneous Initial Rest Frames (IIRF)

Thrust	at $r =$	$\Delta v_{\rm IIRF}$ component	Description
#1	$20M$	$\Delta v_{\rm x,IIRF1} = -0.090\ 132\ 846\ 2$	into circular orbit
#2	$20M$	$\Delta v_{\rm x,IIRF2} = -0.051\ 927\ 321\ 7$	into transfer orbit
#3	$6M$	$\Delta v_{\rm x,IIRF3} = -0.269\ 017\ 469$	into ISCO
#4	$6M$	$\Delta v_{\rm x,IIRF4} = 0.060\ 908\ 153\ 8$	into transfer orbit
#4	$6M$	$\Delta v_{\rm y,IIRF4} = -\ 0.228\ 989\ 795$	into transfer orbit

NOTE: After thrust #4, the probe coasts into the unstable circular orbit at $r = 4M$.

quantities in this rest frame, as in the symbols $\Delta v_{\rm x,IIRF}$ and $\Delta v_{\rm y,IIRF}$ for the change in velocity components in the IIRF frame caused by that rocket impulse. We describe four different IIRF thrusts, listed with an additional number 1 through 4 added to the subscript (Table 2).

IIRF1 transfer velocity change

Special relativity addition of velocities gives us our first, tangential, IIRF rocket-thrust change $\Delta v_{\rm x,IIRF1}$ with the number 1 added to the subscript. This rocket thrust must reduce the shell speed of the spaceship. From equation (54) of Section 1.13,

$$\Delta v_{\rm x,IIRF1} = \frac{v_{\rm x,shell,circle} - v_{\rm x,shell,insert}}{1 - v_{\rm x,shell,insert}v_{\rm x,shell,circle}} \tag{21}$$

$$= -0.090\ 132\ 846\ 2 \qquad \text{(into circular orbit at } r = 20M)$$

Put this numerical value into Table 2. This rocket-thrust velocity change (−27 021 kilometers/second) inserts the incoming spaceship into the circular orbit at $r = 20M$.

?

Objection 1. *Wait! The two velocities, $v_{\rm x,shell,circle}$ and $v_{\rm x,shell,insert}$ are measured in the same local inertial shell frame. The difference in x-components is the measured difference in x-components; why confuse things with complicated equation (21)?*

!

Remember in special relativity the *law of addition of velocities* between two inertial frames in relative motion (Part A of Exercise 17 , Section 1.13)? Equation (21) could be called the *law of subtraction of velocities*—Part B of that earlier exercise. The complication of equation (21) does not require general relativity.

?

Objection 2. *Wow, that is quite a long vertical line in Figure 3. How fast does the probe move along that line? That quick transition must violate the light-speed limit!*

!

No, the probe does not change any global coordinate, T, r, or ϕ, as it traverses the (idealized) vertical line. That transition results from a rocket thrust; it simply changes L and E almost instantaneously (Comment 3).

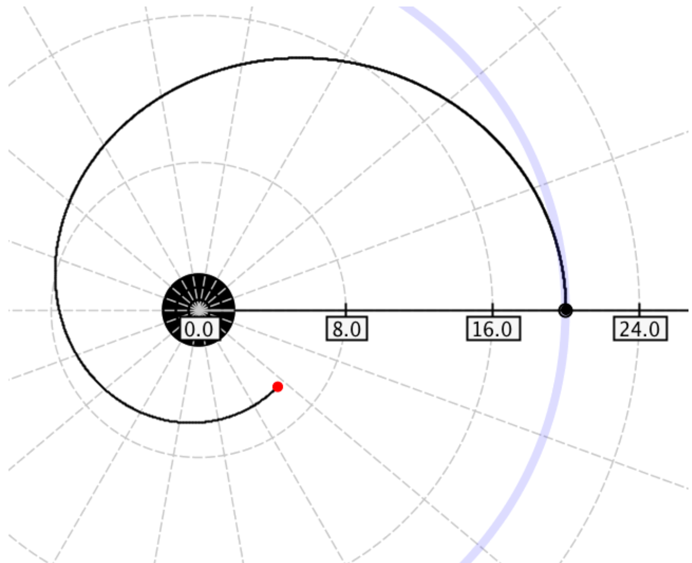


FIGURE 4 Transfer orbit in which the unpowered probe coasts from tangential motion at $r_A = 20M$ to tangential motion at $r_{ISCO} = 6M$. Figure 5 shows the effective potential for this transfer and change in tangential speed required to put the probe into this transfer orbit.

?

Objection 3. *Your analysis of insertion into a circular orbit takes no account of mass loss due to required rocket thrusts. Whenever spaceship mass changes, its map energy and map angular momentum also change.*

!

Right you are. However, constants of motion in our equations are map energy and map angular momentum *per unit mass.* Map energy E/m and map angular momentum $L/(mM)$ are unitless. Therefore the initial mass of the spaceship (before a rocket thrust) and the final spaceship mass (after the rocket thrust) do not affect these equations.

9.3 ■ TRANSFER TO THE ISCO

Get closer

The spaceship completes observations from the stable circular orbit at $r = 20M$ and its captain wants to make further observations from a smaller circular orbit—still outside the event horizon. To take the entire spaceship to this smaller orbit requires a large amount of rocket fuel; instead the captain launches a small probe toward the smaller orbit.

Transfer to circular orbit at $r_{ISCO} = 6M$.

What r-value shall we choose for the inner circular orbit? Be bold! Take the probe all the way down to the so-called Innermost Stable Circular Orbit at $r_{ISCO} = 6M$ (Definition 6, Section 8.5).

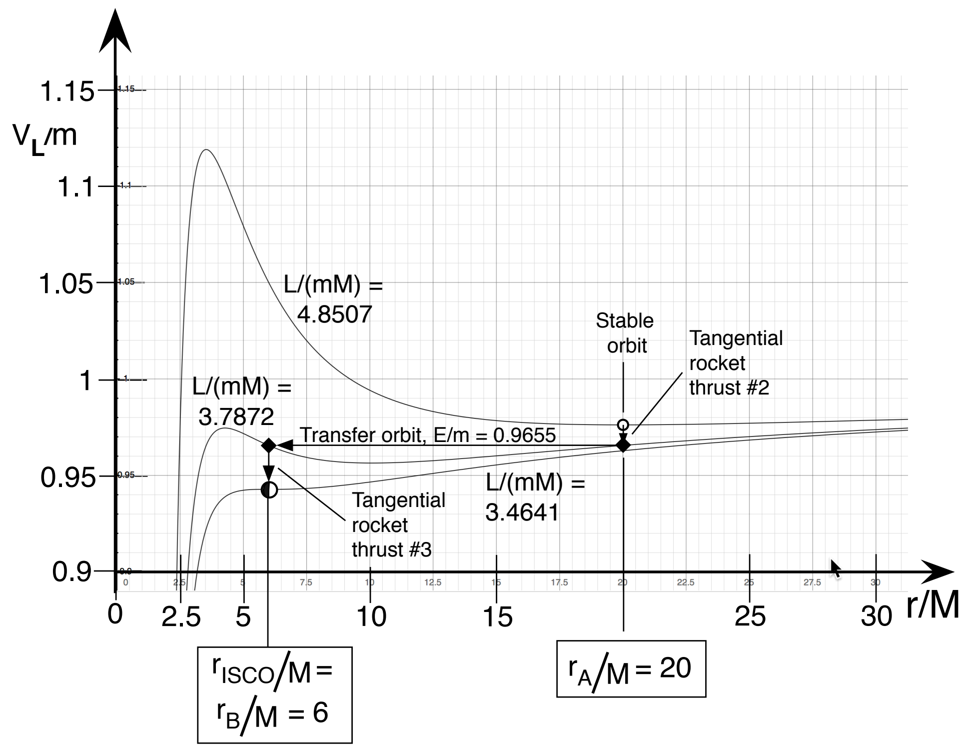


FIGURE 5 Transfer orbit between sequential tangential rocket thrusts #2 and #3. This maneuver moves the probe from the stable circular orbit at $r = 20M$ to the half-stable ISCO at $r_{\rm ISCO} = 6M$. Figure 4 plots this transfer orbit on the $[r, \phi]$ slice.

Comment 5. ISCO as a limiting case
The ISCO is hazardous because it's "half stable" and may lead to a death spiral inward through the event horizon. To prevent this, the inner circular orbit r-value should be slightly greater than $r_{\rm ISCO}$ to make it fully stable. In what follows we ignore this necessary small r-adjustment.

Figure 4 shows a transfer orbit, tangential at both $r_{\rm A} = 20M$ and $r_{\rm B} = r_{\rm ISCO} = 6M$. Recall that these radii are called **radial turning points,** because at both r-values $dr/d\tau = 0$, so the orbiter instantaneously sweeps around only tangentially. Figure 5 displays the corresponding map energy on the effective potential plot.

QUERY 2. Profile of transfer orbit
In 1925 Walter Hohmann described a transfer orbit between two planetary orbits around our Sun as "half an ellipse." Half an ellipse would have maxima of $r_{\rm A}$ and $r_{\rm B}$ on opposite sides of the center of attraction. The orbit plot in Figure 4 does not look like half an ellipse. Why is this different from Hohmann's prediction?

We seek a transfer orbit between the specified Above circular orbit at r_A/M and the Below circular orbit at r_B/M; Figure 5 shows this transfer. In equation (4), $dr/d\tau = 0$ at the two turning points r_A/M and r_B/M, which yields:

$$\left(\frac{E}{m}\right)^2 = \left(\frac{V_L(r_A)}{m}\right)^2 = \left(\frac{V_L(r_B)}{m}\right)^2 \qquad \text{(at turning points)} \qquad (22)$$

Transfer orbit map L and E

Look first at the right equality in (22), in which the square of the effective potential (5) has the same value at two different r. Write down this equality and solve the resulting equation for $(L/m)^2$. The result is equation (23). Next look at the left equality in (22), in which the square of the map energy $(E/m)^2$ is equal to the square of the effective potential at either r. Write down this equality and solve the resulting equation for $(E/m)^2$. The result is equation (24).

$$\left(\frac{L}{m}\right)^2_{\text{transfer}} = \frac{2Mr_A^2 r_B^2 (r_A - r_B)}{r_A^3(r_B - 2M) - r_B^3(r_A - 2M)} \qquad \text{(between circular orbits)} \quad (23)$$

$$\left(\frac{E}{m}\right)^2_{\text{transfer}} = \frac{(r_A - 2M)(r_B - 2M)(r_A^2 - r_B^2)}{r_A^3(r_B - 2M) - r_B^3(r_A - 2M)} \qquad \text{(between circular orbits)} \quad (24)$$

QUERY 3. Transfer either way

Show that equations (23) and (24) are both symmetrical in r_A and r_B. In other words, show that the same values of $(L/m)_{\text{transfer}}$ and $(E/m)_{\text{transfer}}$ apply, irrespective of the direction of transfer between the circular orbits. Is this result obvious?

Substitute values $r_A = 20M$ and $r_B = r_{\text{ISCO}} = 6M$ into equations (23) and (24); enter resulting values of L/m and E/m into Table 1. Then equations (18) and (20) give us values of $v_{\text{x,shell,transfer}}$ and $v_{\text{x,shell,circle}}$. These results allow us to compute the rocket thrust needed to put the probe into the transfer orbit. This is our second, also tangential, instantaneous initial rest frame IIRF thrust (Definition 2) with the number 2 added to the subscript, $\Delta v_{\text{x,IIRF2}}$.

IIRF2 transfer velocity change

$$\Delta v_{\text{x,IIRF2}} = \frac{v_{\text{x,shell,transfer}} - v_{\text{x,shell,circle}}}{1 - v_{\text{x,shell,transfer}} v_{\text{x,shell,circle}}} \qquad \text{(into transfer orbit} \qquad (25)$$

$$= -0.051\ 927\ 321\ 7 \qquad \text{from } r = 20M \text{ to } r_{\text{ISCO}})$$

Enter this numerical result into Table 2. This rocket-thrust velocity change (−15 567 kilometers/second) inserts the probe into a transition orbit that carries it from tangential motion at $r = 20M$ down to tangential motion at $r_{\text{ISCO}} = 6M$.

?

Objection 4. *You talk about moving into a circular orbit and transferring between orbits. But what will our orbiting observers **see**? You have told us nothing about what they see as they look around.*

!

Guilty as charged! Section 7.7 showed only what a raindrop diver sees radially inward and radially outward as she plunges to the center of the black hole. Beyond that, we have made no predictions whatsoever about what any observer sees. For example: In what local frame direction must an observer look to see a particular star? What must we know to make such predictions? Chapters 13 answers these questions. The cosmic trip planner must read beyond the present chapter!

IIRF3 transfer velocity change

When the probe reaches $r_{\text{ISCO}} = 6M$, it travels tangentially for an instant at shell velocity $v_{\text{x,shell,transfer}}$. Then a third insertion rocket thrust changes this shell velocity to $v_{\text{x,shell,circle}}$ for the circular orbit at r_{ISCO}. Table 1 has values of both of these velocities. What insertion rocket thrust does this? As before, it is a tangential thrust in the instantaneous inertial rocket frame IIRF (Definition 2), with the number 3 added to the subscript, $\Delta v_{\text{x,IIRF3}}$.

$$\begin{aligned}\Delta v_{\text{x,IIRF3}} &= \frac{v_{\text{x,shell,transfer}} - v_{\text{x,shell,circle}}}{1 - v_{\text{x,shell,transfer}} v_{\text{x,shell,circle}}} \\ &= -0.269\ 017\ 469 \qquad (\text{into circular orbit at } r_{\text{ISCO}} = 6M)\end{aligned} \tag{26}$$

Enter the numerical result in Table 2. This rocket-thrust velocity change (−86 494 kilometers/second) inserts the probe into the circular orbit at $r_{\text{ISCO}} = 6M$.

9.4 ■ TRANSFER TO AN UNSTABLE CIRCULAR ORBIT

Put the probe at risk!

Transfer to unstable orbit at $r = 4M$

Thus far we have inserted our spaceship into a stable circular orbit at $r = 20M$, then transferred a probe down to the half-stable circular orbit at $r_{\text{ISCO}} = 6M$. Now the spaceship captain wants to make observations even closer to the black hole. She decides to transfer the probe from $r_{\text{ISCO}} = 6M$ to the unstable circular orbit at $r = 4M$, a maneuver shown in Figures 6 and 7.

QUERY 4. Unstable circular orbit at $r = 4M$

A. Show that the unstable circular orbit at $r = 4M$ has map angular momentum $L/m = 4M$.
B. Show that the unstable circular orbit at $r = 4M$ has map energy $E/m = 1$.
C. Make an argument that the transfer orbit from $r = 6M$ to $r = 4M$ in Figures 6 and 7 must have the same values of map energy and map angular momentum given in the first two items of this Query.

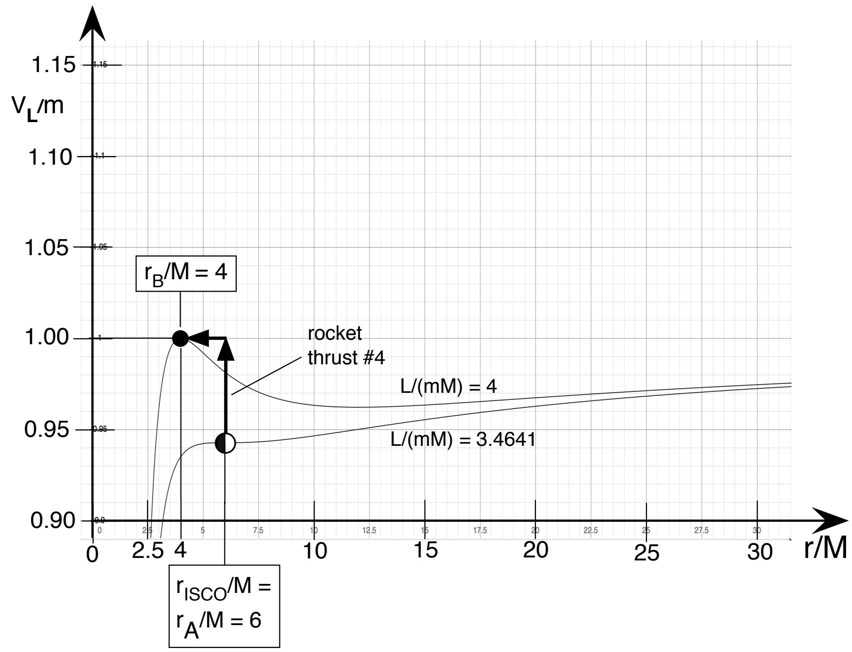


FIGURE 6 Probe transfer orbit between half-stable orbit at $r_{\rm ISCO} = 6M$ and unstable circular orbit at $r = 4M$. See Figure 7.

D. Verify the bottom right hand entry in Table 3, namely that at $r = 4M$, $v_{\rm x,shell,circle} = v_{\rm x,shell,transfer} = |v_{\rm shell,transfer}|$

Transfer orbits have radial turning points where $E/m = V_{\rm L}(r)$. Usually these turning points are not at an extremum of the effective potential, so they are not at r-values of circular orbits. In this case, however, we need a rocket thrust to *create* the extremum for a circular orbit at that r-value.

No rocket thrust needed for insertion into unstable orbit.

At a maximum of the effective potential, the turning point occurs at the r-value of the circular orbit, so we need no rocket thrust to put the probe into that circular orbit. Figure 6 shows this special case: The probe moves to smaller r along the horizontal arrow in Figure 6. As it does so it reaches the effective potential maximum at $r = 4M$ where it automatically enters the unstable circular orbit at that r-value. So we need only a single rocket thrust at $r = 6M$ to change map energy and map angular momentum to that of the circular orbit at $r = 4M$ (Figure 7).

?

Objection 5. *Once the rocket thrust #4 shoots the probe upward in Figure 6 to map energy $E/m = 1$, why should the probe go left in that figure, to smaller r? Why doesn't it go right, to larger r?*

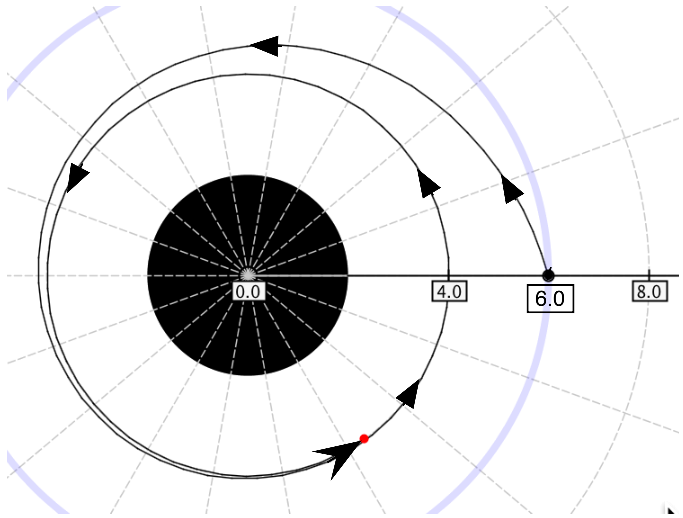


FIGURE 7 Transfer orbit from $r_{\mathrm{ISCO}} = 6M$ to the unstable circular orbit at $r = 4M$ (Figure 6). This requires a velocity $v_{\mathrm{shell,transfer}}$ inward from 90° by 19.471 degrees, with shell velocity components and magnitude given in Table 3.

Figure 7 and Table 3 show the answer: The rocket thrust is not tangential but has an inward r-component.

Need two thrust components for transfer orbit

Query 4 already tells us the map values $E/m = 1$ and $L/m = 4M$ of the leftward horizontal arrow in Figure 6. Because the rocket thrust is not tangential, we need to apply the full set of equations (18) and (19) to find the shell components of the velocity in the transfer orbit. Enter these results for $v_{\mathrm{y,shell,transfer}}$ and $v_{\mathrm{x,shell,transfer}}$ in Table 3.

To start this transfer from r_{ISCO} we use the fourth rocket thrust measured in the instantaneous initial rest frame. This thrust requires two components, which we call $\Delta v_{\mathrm{x,IIRF4}}$ and $\Delta v_{\mathrm{y,IIRF4}}$, with the number 4 added to the subscript. In this case we must adapt both velocity addition equations (54) in Section 1.13.

$$\Delta v_{\mathrm{x,IIRF4}} = \frac{v_{\mathrm{x,shell,transfer}} - v_{\mathrm{x,shell,circle}}}{1 - v_{\mathrm{x,shell,circle}} v_{\mathrm{x,shell,transfer}}} \quad \text{(into the transfer orbit...} \qquad (27)$$

$$\Delta v_{\mathrm{y,IIRF4}} = \frac{v_{\mathrm{y,shell,transfer}}}{\gamma_{\mathrm{x,shell,circle}}(1 - v_{\mathrm{x,shell,circle}} v_{\mathrm{x,shell,transfer}})} \quad \text{...from } r = 6M \qquad (28)$$

$$\text{where} \quad \gamma_{\mathrm{x,shell,circle}} = (1 - v^2_{\mathrm{x,shell,circle}})^{-1/2} \qquad \text{...to } r = 4M) \qquad (29)$$

Substitute into these equations from $r = r_{\mathrm{ISCO}} = 6M$ values in Tables 1 and 3 and enter the resulting components into Table 2. This rocket thrust, which corresponds to the vertical arrow in Figure 6, causes a velocity change of magnitude, $|\Delta v_{\mathrm{IIRF4}}| = 0.236\ 951\ 745 = 71\ 036$ kilometers/second.

TABLE 9.3 Numerical values for transfer from $r_{\text{ISCO}} = 6M$ to $r = 4M$

Values of	$r_{\text{ISCO}} = 6M$	$r = 4M$
$(L/m)_{\text{transfer}}$	$4M$	$4M$
$(E/m)_{\text{transfer}}$	1	1
$v_{\text{x,shell,transfer}}$	0.544 331 054	0.707 106 781
$v_{\text{y,shell,transfer}}$	−0.192 450 090	0
$\lvert v_{\text{shell,transfer}}\rvert$	0.577 350 269	0.707 106 781
$\theta_{\text{x,shell}}$	−19.471 220 6°	0
$v_{\text{x,shell,circle}}$	0.500 000 000	0.707 106 781

Good-bye probe!

Our probe coasts to the unstable circular orbit at $r = 4M$, an effective potential peak close to the black hole. After it completes measurements there, the captain decides to dispose of the probe. To do this, she commands the probe to fire a tiny inward rocket thrust to tip it off the effective potential peak and send it spiraling inward across the event horizon. Good job!

Section 9.5 applies some of what we have learned to analyze Larry Niven’s short story “Neutron Star.”

9.5 ■ “NEUTRON STAR” BY LARRY NIVEN

Close to a neutron star? Look out!

Why did earlier explorers die?

Larry Niven’s science fiction short story “Neutron Star” describes the trip by spaceship pilot Beowulf Schaeffer to discover why two earlier pilots died while orbiting a neutron star. Sponsors of Beowulf’s trip are aliens called puppeteers, who manufacture spaceship hulls that are utterly indestructable and—so they claim—impenetrable. Naturally, the death of two pilots in an “impenetrable” puppeteer spaceship hull has reduced sales. The puppeteers want to know what deadly force has managed to enter their high-tech hulls.

Passage through closest approach

As Beowulf approaches the neutron star, the long axis of his spaceship inexorably orients along a radial line to the star (Why?). Beowulf suddenly realizes that he must position himself at the point in the spaceship where at least one part of his body feels no gravity in order to be in free-fall motion around the neutron star. Here is Niven’s description of his passage through the r-coordinate of closest approach:

“Giants’ hands gripped . . .”

> *My time was up. A red disk leapt up at me; the ship swung around me; I gasped and shut my eyes tight. Giants’ hands gripped my arms and legs and head, gently but with great firmness, and tried to pull me in two. In that moment it came to me that Peter Laskin had died like this. He’d made the same guesses I had, and he’d tried to hide in the access tube. But he’d slipped . . . as I was slipping . . . From the control room came a multiple shriek of tearing metal. I tried to dig my feet into the hard tube walls. Somehow they held.*

Close-call survival

According to Niven's story, Beowulf is (barely!) able to cling to the point of zero local gravity, though the skin on his extremities is injured. After returning to base, he reports to the puppeteers that the deaths of earlier explorers were due to their slipping from this gravity zero point and falling to the front (or back) of the spaceship.

Objection 6. *What in (or out of) this world is happening to Beowulf? His orbit around the neutron star is similar to those we use to insert our spaceship into a circular orbit. Why is Beowulf in danger, and why did earlier explorers die?*

"All politics is local," said politician Tip O'Neill. A monster may lurk at opposite ends of your spaceship. In "Neutron Star" the monster is *tidal acceleration*, which can be lethal.

Killer tides

Tidal acceleration is nothing new for us. Section 7.9 introduced it for the radial fall into the black hole, and in the present chapter Section 9.7, Appendix: Killer Tides, gives expressions for radial and tangential tidal accelerations. This information allows us to answer the question, "Can Beowulf Schaffer survive his transit past the neutron star?"

Survival?

We need numerical values from "Neutron Star" in order to apply tidal acceleration expressions from Section 9.7. Larry Niven tells us that (a) the neutron star's mass is 1.3 times the mass of our Sun, (b) the minimum r-coordinate of approach is approximately 10.5 kilometers, so that $r_{\text{min}} \approx 5.5M$. (The neutron star is also spinning, but too slowly to have a significant effect on Beowulf's global orbit or local safety.)

QUERY 5. Einstein predicts Beowulf Schaeffer's fate

Use the parameters in the preceding paragraph to find out whether or not Beowulf Schaeffer survives tidal accelerations during his encounter with the neutron star. Assume that the distant speed of approach to the neutron star is nonrelativistic, so that $E/m \approx 1$.

A. Use (3) to determine v_{shell} at the closest approach r_{min}.

B. By what multiple is the radial tidal effect (in the local spaceship Δy_{ship} direction) larger than the Newtonian prediction?

C. At the moment of closest approach to the neutron star, Beowulf Schaeffer extends his arm one meter radially inward. What happens to him next?

D. Give a definitive answer to the question, "Can Beowulf Schaeffer survive the trip described in "Neutron Star"? (When our class sent numerical results to Larry Niven, he replied, "Thank you for the calculations. I'm not sure how I will use them, but thanks anyway.")

E. If you conclude that Beowulf cannot survive the "Neutron Star" trip, find an r-coordinate of closest approach to the neutron star at which Beowulf Schaeffer can survive. State your criteria for *survival.* On the way to this result, give a specific numerical value for $\Delta g/\Delta y_{\text{ship}}$ that, in your estimate, is survivable.

QUERY 6. Blackmail

Discussion question: Beowulf Schaefer blackmails the secretive puppeteers by threatening to reveal that they come from a moonless world. How does he know that?

QUERY 7. *Optional:* **Swimming in spacetime?**

A massive mother ship is in a circular orbit with its long dimension tangential with respect to the black hole. Astronauts inside extend a mechanical arm radially inward toward the black hole. The "hand" on this arm experiences a radially inward force.

A. Can such a maneuver be used to change the orbit of the mother ship?

B. Can similar maneuvers provide a method for balancing a spaceship in a circular knife-edge orbit without using rockets?

C. Using repeated "calisthenics," can a freely-floating astronaut "swim" around the mother ship? (See "Swimming in Spacetime" in the references.)

D. Do such maneuvers violate the laws of conservation of map energy or map angular momentum?

E. Do similar maneuvers work in flat spacetime?

9.6 ■ A COMFORTABLE CIRCULAR ORBIT

How close to the black hole?

Meaning of "comfortable"?

Up to this point, our description of circular orbits has a serious flaw: We do not answer the question, "What is the minimum r-value of a circular orbit in which the astronaut will be comfortable?" Our answer to this question has three parts:

- Part I. What are the tidal accelerations in a circular orbit of given r-coordinate? To answer this question, we consult Section 9.7, Appendix: Killer Tides.
- Part II. What is the maximum tidal acceleration for which a human is comfortable?
- Part III. What is the minimum r-coordinate of a circular orbit (Part I) for which a human is comfortable (Part II)?

Instead of choosing an orbit that is comfortable for a human, we can replace the human with a probe hardened to withstand hundreds or thousands of times the tidal accelerations that would injure or kill a person.

Part I: Tidal acceleration in circular orbit

In order to apply tidal equations (46) through (48) to a circular orbit, we need the square of the tangential shell velocity in (10).

Tidal acceleration in circular orbit

Think of an astronaut in a circular orbit with the long axis of his body oriented along the radial direction. His height is larger than his width, so we carry out our calculations for the radial tidal component only, knowing that the other components will be smaller. Half his height provides a value for Δy_{local} in equation (46). Substitute (10) into (46) and rearrange so the right side of the equation contains only expressions in r.

$$\Delta g_{\text{local},y} \approx \frac{M}{\bar{r}^3}\left(\frac{2\bar{r}-3M}{\bar{r}-3M}\right)\Delta y_{\text{local}} \qquad \text{(circular orbit)} \tag{30}$$

Part II: Define human comfort.

Tidal acceleration for human comfort

How large a tidal acceleration is comfortable for a human being? The answer is different for people of different heights. Here we treat our human astronaut gently, using the definition employed in Section 7.9 under the assumption that he is oriented along a radial line, with head above feet. Then with his stomach in free fall, the astronaut remains comfortable if his head is accelerated upward with the acceleration it would experience on Earth—call it g_{E}—and his feet are accelerated downward with the same magnitude of Earth acceleration.

Assume the astronaut is approximately two meters tall, so his measured distance between head and stomach is one meter, the same as the separation between stomach and feet. Then $\Delta y_{\text{local}} = 1\,\text{meter}$ in equation (30).

Part III: Minimum-r circular orbit for human comfort

Minimum r for comfort?

The acceleration g_{E} at Earth's surface has the numerical value $g_{\text{E}} = 1.09 \times 10^{-16}\ \text{meter}^{-1}$ (inside the front cover). We want to insert g_{E} into (30) when the circling astronaut's "half height" is $\Delta y_{\text{local}} = 1$ meter:

$$g_{\text{E}} = \Delta g_{\text{local},y} \approx \frac{M}{\bar{r}^3_{\text{comfort}}}\left(\frac{2\bar{r}_{\text{comfort}}-3M}{\bar{r}_{\text{comfort}}-3M}\right)\times 1\,\text{meter} \quad \text{(human comfort limit)} \tag{31}$$

$$g_{\text{E}} \approx \frac{M^{-2}}{\left(\bar{r}_{\text{comfort}}/M\right)^3}\left(\frac{2\bar{r}_{\text{comfort}}/M-3}{\bar{r}_{\text{comfort}}/M-3}\right)\times 1\,\text{meter} \tag{32}$$

In this equation, $\bar{r}_{\text{comfort}}$ refers to the smallest r-value of the circular orbit in which the observer is comfortable. Multiply the left and right sides of (32) by M^2 and divide by g_{E}. The result is

$$M^2 \approx \frac{1}{\left(\bar{r}_{\text{comfort}}/M\right)^3}\left(\frac{2\bar{r}_{\text{comfort}}/M-3}{\bar{r}_{\text{comfort}}/M-3}\right)\frac{1\,\text{meter}}{g_{\text{E}}} \quad \text{(human comfort limit)} \tag{33}$$

We can rearrange (33) to give the mass of the black hole in number of Suns, M/M_{Sun}, as a function of the minimum r-value, r_{comfort}, of the circular orbit in which a human astronaut will be comfortable:

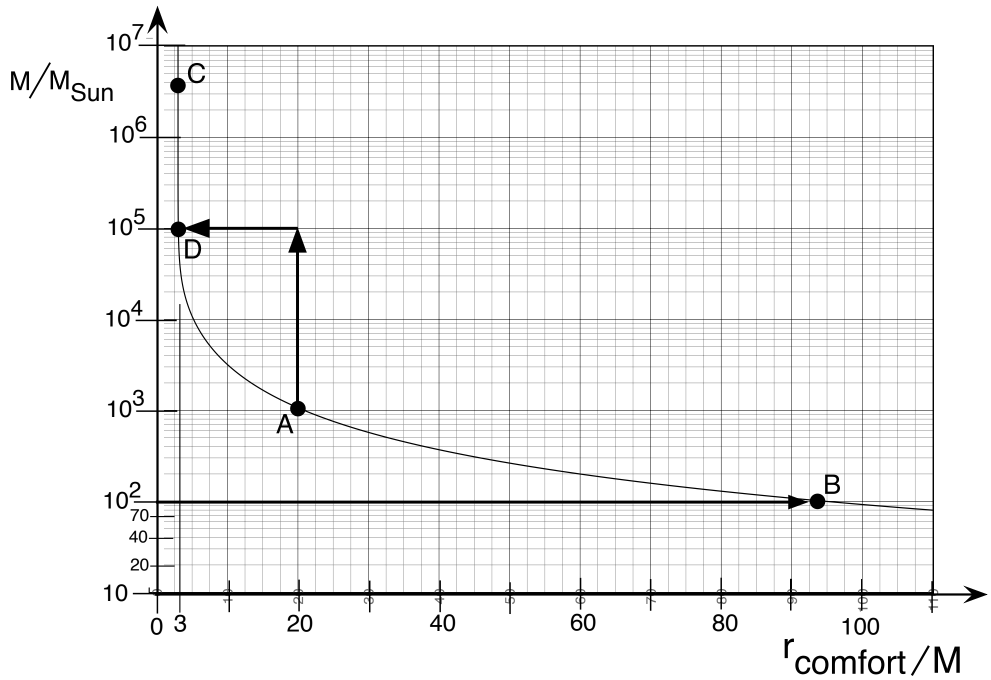


FIGURE 8 The horizontal axis, $r_{\rm comfort}/M$, gives the minimum-r circular orbit in which a human will be comfortable. On the vertical axis, $M/M_{\rm Sun}$ is a number equal to the mass of the black hole in units of the mass of our Sun. Arrows and little filled circles illustrate solutions of Sample Problems 1A through 1D.

$$\frac{M}{M_{\rm Sun}} = \frac{1}{M_{\rm Sun}}\left(\frac{1\,{\rm meter}}{g_{\rm E}}\right)^{1/2}\left[\frac{1}{(\bar{r}_{\rm comfort}/M)^3}\left(\frac{2\bar{r}_{\rm comfort}/M-3}{\bar{r}_{\rm comfort}/M-3}\right)\right]^{1/2} \tag{34}$$

$$= 6.47\times 10^4\left[\frac{1}{(\bar{r}_{\rm comfort}/M)^3}\left(\frac{2\bar{r}_{\rm comfort}/M-3}{\bar{r}_{\rm comfort}/M-3}\right)\right]^{1/2} \tag{35}$$

(minimum-r circular orbit for human comfort)

The last step substitutes values of $M_{\rm Sun}$ and $g_{\rm E}$ from inside the front cover. Verify that both sides of this equation are unitless. Figure 8 plots the curve of this equation. Sample Problems 1 explain the arrows.

Sample Problems 1. Minimum-r Circular Orbit for Human Comfort

PROBLEM 1A

What is the numerical value of M/M_{Sun} for which $r_{\text{comfort}}/M = 20$ is the minimum circular orbit in which a human feels comfortable? What is the value of r_{comfort} in meters?

SOLUTION 1A

Figure 8 shows that at $r_{\text{comfort}}/M = 20$, $M/M_{\text{Sun}} \approx 10^3$, indicated by point A in the figure. The value of r_{comfort} in meters is $r_{\text{comfort}} = 20 \times M \text{ meters} = 20 \times (M/M_{\text{Sun}}) \times M_{\text{Sun}} \text{ meters} \approx 20 \times 10^3 \times 1.48 \times 10^3$ meters $\approx 3 \times 10^7 \text{meters} \approx 3 \times 10^4$ kilometers.

PROBLEM 1B

I approach the black hole of mass value $N_{\text{Suns}} = 10^2$. What is the minimum r_{comfort} of the circular orbit in which I will feel comfortable?

SOLUTION 1B

The long horizontal arrow to the right at $N_{\text{Suns}} = 10^2$ in Figure 8 crosses the "comfort curve" at $r_{\text{comfort}}/M \approx 93$, indicated by point B in Figure 8.

PROBLEM 1C

I approach the monster black hole in the center of our galaxy, for which $N_{\text{Suns}} \approx 4 \times 10^6$. Assume (incorrectly) that this monster black hole is not spinning. What is the approximate value of r_{comfort} for this circular orbit?

SOLUTION 1C

The number $M/M_{\text{Sun}} = 4.1 \times 10^6$ is point C on the curve in Figure 8. You will be comfortable in an orbit of approximately $r_{\text{comfort}}/M = 3$

PROBLEM 1D

The robot satellite released by the spaceship at $r_{\text{comfort}}/M = 20$ in Problem 1A is made small and hardened in various ways to withstand tidal accelerations 10^4 times as great as that for which a human will be comfortable. What is the value of r_{comfort} of the circular orbit in which this probe will continue to operate?

SOLUTION 1D

Look at equation (34). The black hole remains the same, so the ratio M/M_{Sun} on the left side remains the same. Therefore the right side must remain the same. When g_{E} in the denominator on the right side *increases* by a factor of 10^4, then its square root contribution to the right side *decreases* by the factor 10^2. To compensate, the square root of the square-bracket expression must *increase* by the factor 10^2. The vertical arrow in the figure extends upward by this factor of 10^2. The leftward horizontal arrow finds r_{conf}/M, for the "comfort orbit" of the robot. This $r_{\text{comfort}}/M \approx 3$ for the robot is at almost the minimum r-value for an unstable circular orbit.

9.7 ■ APPENDIX: KILLER TIDES

Avoid spaghettification!

Size of local inertial frame limited by tides.

The dangers experienced by Beowulf and other explorers near a neutron star should not surprise us. Objects near to one another in curved spacetime can experience relative accelerations. Section 1.11 described these "tidal accelerations" that limit the size of a local inertial frame. At locations near to one another on Earth's surface, these relative accelerations are too small for us to notice in everyday life. In contrast, near a neutron star or a black hole relative tidal accelerations at different locations on a single human body can injure or kill. We call such different accelerations **killer tides**.

In principle, you can derive the following tidal accelerations using only basic tools for the motion of a stone: the metric plus the Principle of Maximal Aging. This process, however, is an algebraic nightmare, so we simply quote results obtained with the use of a more advanced general-relativistic formalism.

TIDES DURING RADIAL MOTION

Radial motion: Newton's tidal accelerations are valid.

Surprise! For the special cases of an observer either at rest in global coordinates near a black hole or moving radially toward or away from it, local tidal effects predicted by general relativity are identical to those predicted by Newton. Write Newton's expression for gravitational acceleration in the radially outward or local y-direction due to a point or spherically symmetric source. In unitless coordinates:

$$g_y = -\frac{M}{r^2} \qquad \text{(Newton)} \tag{36}$$

Take the differential of this to measure radial tidal effects and write the result in the approximate form for local frame measurements:

$$\Delta g_{\text{local},y} \approx \frac{2M}{\bar{r}^3}\Delta r \approx \frac{2M}{\bar{r}^3}\Delta y_{\text{local}} \qquad \text{(Newton)} \tag{37}$$

The final step, equating Δr to Δy_{local}, makes sense only for Newton; in general relativity the relation between global increment Δr and local frame increment Δy_{local} depends on the position and motion of the local frame in global coordinates. Nevertheless—surprise again!—the full general relativity analysis also yields the last expression in (37). To show this is difficult. The following boxed three equations tell us the tidal accelerations in the three directions in the inertial frame.

$$\Delta g_{\text{local},y} \approx \frac{2M}{\bar{r}^3}\Delta y_{\text{local}} \tag{38}$$

$$\Delta g_{\text{local},x} \approx -\frac{M}{\bar{r}^3}\Delta x_{\text{local}} \tag{39}$$

$$\Delta g_{\text{local},z} \approx -\frac{M}{\bar{r}^3}\Delta z_{\text{local}} \tag{40}$$

Subscript "local" means *any* local frame at rest or moving radially inward or outward in global rain coordinates.

Spaghettification: radial stretch plus tangential compression

A radially-diving observer suffers not only stretching in the radial direction, but also compression in tangential directions as her descending body funnels into an ever-narrowing local space. Negative signs in (39) and (40) reflect this compression. We give the light-hearted name **spaghettification** to the physical result of these combined stretch and compression tidal effects: lengthwise extension combined with transverse compression. Sample Problem 2 carries out a Newtonian analysis of gravity gradients (tides), whose results turn out to be identical in form to general relativistic results (38) through (40).

Expressions (38) through (40) shrink to become calculus expressions (44) at a point. Every approximate equation in this section can lead to a similar calculus expression. We keep the Δ notation, however, to remind us that we deal here with a local frame of finite extent.

Now apply equations (38) through (40) to a local *inertial* frame. A liquid drop of nearly incompressible fluid, such as water or mercury, has a surface tension that tends to minimize surface area, which makes the droplet spherical

Sample Problem 2. Newton's tidal components

Derive expressions similar to (38) through (40) for Newton's case, in the calculus limit.

SOLUTION:
This is one of only two places in this book where we use vector expressions and partial derivatives. Represent unit vectors in the x, y, and z directions by $\hat{x}$, $\hat{y}$, and $\hat{z}$, respectively. Use this notation to write (36) as a vector equation:

$$\mathbf{g} = -\frac{M\,(x\hat{x} + y\hat{y} + z\hat{z})}{(x^2 + y^2 + z^2)^{3/2}} \qquad \text{(Newton)} \qquad (41)$$

Each component of this vector has the algebraic form:

$$g_q = -\frac{Mq}{(x^2 + y^2 + z^2)^{3/2}} \qquad (42)$$

where q stands for any coordinate x, y, or z. Take the partial derivates similar to the general relativistic equations (38) through (40). You can show that the results also have the same form for all three components:

$$\frac{\partial g_q}{\partial q} = -\frac{M}{r^3} + \frac{3Mq^2}{r^5} \qquad (q \rightarrow x, y, z) \qquad (43)$$

We want expressions for these partial derivatives at global coordinate r in flat spacetime. Take y to be along the radial direction, so at that point $y = r$, while $x = z = 0$. Equations (43) become:

$$\frac{\partial g_x}{\partial x} = -\frac{M}{r^3} \qquad \text{(Newton)} \qquad (44)$$
$$\frac{\partial g_y}{\partial y} = -\frac{M}{r^3} + \frac{3M}{r^3} = +\frac{2M}{r^3}$$
$$\frac{\partial g_z}{\partial z} = -\frac{M}{r^3}$$

Inspection shows that equations (44) have the same form as equations (38) through (40).

All radial speeds give same local tidal accelerations.

in an inertial frame. Equations (38) through (40) show us that for radial motion, the drop will be distorted into the shape of a throat lozenge or smooth potato—technical term: **prolate spheroid**—shown in Figure 9.

Equations (38) through (40) are valid for *all possible* radial speeds—including rest—for example a local inertial frame launched in any of the following ways:

- **Local rain frame:** Local inertial frame dropped from rest far away (Box 4, Section 7.4).
- **Local hail frame:** Local inertial frame hurled radially inward from far away with any initial local shell speed.
- **Local drip frame:** Local inertial frame dropped from rest at any initial $r_0 > 2M$.

Radial free-fall frames

All of these are radially-moving *local free-fall frames* (Section 2.1). Taken together, free-fall frames result in every possible inward or outward radial speed of the radially moving frame as measured by a shell observer at any given average $\bar{r}$. General relativity provides results independent of radial speed in (38) through (40), but the tools developed in this book are not sufficient to explain the reason for this result.

Notice that equations (38) through (40) satisfy the equation

$$\frac{\Delta g_{\text{local},y}}{\Delta y_{\text{local}}} + \frac{\Delta g_{\text{local},x}}{\Delta x_{\text{local}}} + \frac{\Delta g_{\text{local},z}}{\Delta z_{\text{local}}} \approx 0 \qquad (45)$$

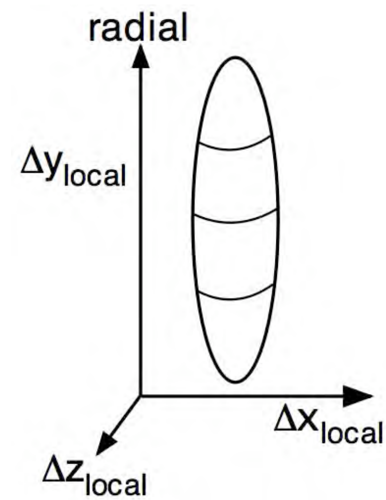


FIGURE 9 Schematic diagram of tide-induced shape for an incompressible liquid drop with surface tension restoring force, observed in a local inertial frame instantaneously at rest or moving radially with respect to a black hole. From the symmetry of the black hole with respect to radial motion, it follows that the tidal squeeze is symmetric perpendicular to the radial direction. Result: the shape is that of an oblong throat lozenge or smooth potato.

Relation among tidal components

This is a general result for tides analyzed by general relativity. In the calculus limit, the approximate equality in (45) becomes mathematically exact, and applies to partial derivatives in (44).

Comment 6. Tides preserve volume.
In the calculus limit, equation (45) expresses a simple and powerful result: The volume of a tiny cloud of free, non-interacting dust particles remains constant as tidal accelerations act on the cloud. This central result is valid even for the far more complicated tidal accelerations near a spinning black hole (Chapter 19).

Tidal effects are continuous across event horizon.

Notice that equations (38) through (40) are continuous across the event horizon at $r/M = 2$. This result provides additional evidence for our repeated claim that an observer falling through the event horizon experiences a steady increase in tidal effects but no sudden jar or jolt there. Indeed, from evidence internal to her local frame the diver cannot tell when she passes radially inward through the event horizon.

TIDES DURING TANGENTIAL MOTION

Tangential motion: tidal accelerations differ from Newton's.

An observer moving in the r, ϕ plane streaks through a local shell frame in the tangential, or Δx_{shell}, direction with shell velocity $v_{\text{shell},x}$. In the following equations, only the factor $M/\bar{r}^3$ reminds us of the corresponding Newtonian analysis in equation (37). *For motion along the tangential* $\pm\Delta x_{\text{shell}}$ *directions:*

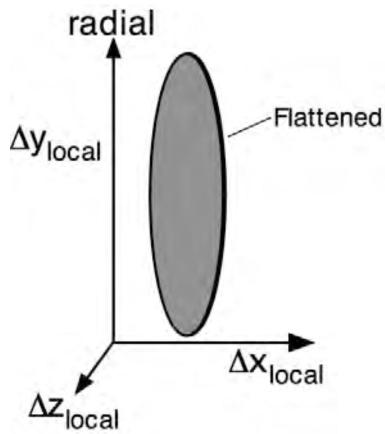


FIGURE 10 Schematic diagram of tide-induced shape for an incompressible liquid drop with surface tension restoring force, observed in a local *inertial* frame that moves in either direction along a $\Delta x_{\rm shell}$ tangential line. This figure shows results for high tangential speed $v_{{\rm shell},x}$: both the tidal stretch in the $\Delta y_{\rm shell}$ direction and the tidal squeeze in the $\Delta z_{\rm shell}$ direction are huge, much greater than the tidal squeeze in the $\Delta x_{\rm local}$ direction. The resulting shape: a thin ribbon with rounded ends lying in the $\Delta x_{\rm shell}$, $\Delta y_{\rm shell}$ plane.

$$\Delta g_{{\rm local},y} \approx \left(\frac{1+v^2_{{\rm shell},x}/2}{1-v^2_{{\rm shell},x}}\right)\frac{2M}{\bar{r}^3}\Delta y_{\rm local} \qquad (46)$$

$$\Delta g_{{\rm local},x} \approx -\frac{M}{\bar{r}^3}\Delta x_{\rm local} \qquad (47)$$

$$\Delta g_{{\rm local},z} \approx -\left(\frac{1+2v^2_{{\rm shell},x}}{1-v^2_{{\rm shell},x}}\right)\frac{M}{\bar{r}^3}\Delta z_{\rm local} \qquad (48)$$

Subscript "local" means *any* local frame moving tangentially in either direction in global coordinates.

Limiting cases for tangential motion

Notice that equation (47) is the same as equation (39) for radial motion, while the equations for the other two directions simply multiply the radial results by coefficients that depend on $v^2_{{\rm shell},x}$. In the low-speed limit ($v^2_{{\rm shell},x} \ll 1$), these equations also reduce to the radial ones (38) and (40). Finally, note that as $v_{{\rm shell},x}$ increases toward the speed of light, the y component leads to radical stretching, while the z component leads to much greater tangential compression than that in the $\Delta x_{\rm local}$ direction.

Expressions (46) through (48) also satisfy the general relation (45) among the local components of gravity gradient, which preserves the volume of a tiny dust cloud moving in the map tangential direction.

For a local *inertial frame*, the result is the tidal distortion of a drop of water or liquid mercury into a flat ribbon with rounded ends, shown in Figure

10 for tangential motion. Equations (46) through (48) are correct for *any* value of $v_{\text{shell},x}$, not just the value of a stone's local shell speed when it is in a circular orbit. For example, a stone that approaches a black hole from far away and returns to far away will travel tangentially at its point of closest approach; these three equations apply at this point.

Section 9.3 applies these results to find the minimum-r circular orbit for human comfort.

QUERY 8. Departure from Newton's gravity gradient

Expressions in parentheses on the right sides of (46) and (48) are a measure of the departure of Einstein's gravity gradients from those predicted by Newton. Temporarily call these expressions **Einstein multipliers**.

A. For what value of $v_{\text{shell},x}$ does the largest of the Einstein multipliers become "significant," which we define as the value 1.1?

B. For what value of $v_{\text{shell},x}$ does the largest of the Einstein multipliers become "large," which we define as the value 10?

C. Exercise 5 in Chapter 1 analyzes the highest energy cosmic ray so far detected, with an energy of 3×10^{20} electron volts. Let this cosmic ray be a speeding proton (mass $= 1.63 \times 10^{-27}$ kilogram $= 9.38 \times 10^{8}$ electron-volts) that streaks tangentially past Earth just above its atmosphere, about 100 kilometers above the surface. Estimate the value of the largest Einstein multiplier in this case. Hint: Define $v_{\text{shell},x} \equiv 1 - \delta$, then use our approximation formula from inside the front cover to redefine the Einstein multipliers in terms of δ.

D. The proton is a quantum particle; its "radius" is not a classical quantity. Nevertheless, estimate the tidal stress on the proton cosmic ray of Part C: Assume this proton radius to be 10^{-15} meter. What are the tidal accelerations at the surface of the "fastest proton" moving tangentially above Earth's atmosphere?

E. Repeat Part D for the "fastest proton" skimming past the surface of a neutron star with $r/M = 10$ kilometers.

9.8 ■ EXERCISES

1. Smallest circular orbit for a hardened probe around the black hole

We harden a probe so that it can withstand K times the maximum comfortable tidal acceleration of a human (Section 9.6). The probe enters a circular orbit around the black hole of mass M in which the tidal acceleration has this maximum. What is the r-value of this circular orbit?

2. The photon (Star Trek) rocket

An advanced civilization develops the **photon rocket engine**, one that combines matter and antimatter in a controlled way to yield only photons

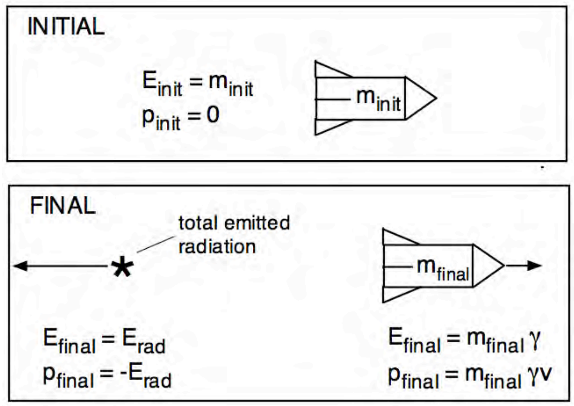


FIGURE 11 Exercise 2. Diagram showing initial and final states of a photon rocket that emits only radiation.

(high-energy gamma rays), all of which it directs out the rear of the rocket. The photon rocket engine is the most efficient in the sense that it produces the greatest possible change of velocity for a given fractional change in mass of the rocket ship. Analyze the photon rocket using special relativity, including the definition $\gamma \equiv (1 - v^2)^{-1/2}$.

A. Write down the energy and momentum conservation laws using Figure 11.

B. Combine the conservations laws, show that $\gamma v = \left(\gamma^2 - 1\right)^{1/2}$, and derive the equation for the *mass ratio*:

$$\frac{m_{\text{init}}}{m_{\text{final}}} = \gamma + \left(\gamma^2 - 1\right)^{1/2} \quad \text{(photon rocket, flat spacetime)} \qquad (49)$$

where m_{init} is the initial mass of the rocket ship.

C. Find the mass ratio for $\gamma = 10$

D. Show that the result of Part C is an example of the approximation

$$\frac{m_{\text{init}}}{m_{\text{final}}} \approx 2\gamma \quad (\text{when } \gamma^2 \gg 1) \quad \text{(photon rocket, flat spacetime)} \quad (50)$$

3. Newton's Tangential Tidal Displacement Near Earth.

Brave Monica Sefner "walks the plank" at the top of the 828-meter-tall Dubai Tower, Burj Khalifa (Figure 12), on which she moves horizontally outward to a point that clears the base of the tower. Then she steps off the plank attached to a bungee cord and falls freely for 600 meters, at which point the cord "takes hold" and slows her to a stop before she reaches the ground. As she leaves the plank, Monica stretches out her arms and releases from rest two marbles

FIGURE 12 Exercise 3. DubaiTower, 828 meters high.

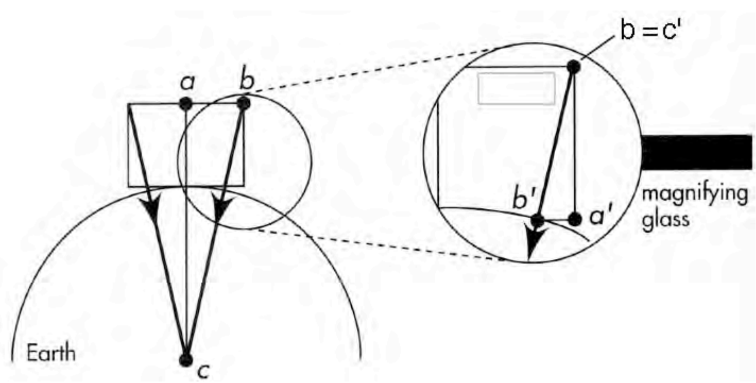


FIGURE 13 Exercise 3. Construction to analyze tangential tidal acceleration of radially falling marbles in Newton's mechanics. Not to scale, and with gross differences in relative scale of different parts of the diagram.

initially 2 meters apart horizontally. Just before the end of her 600-meter free fall, how much will the measured separation between these marbles have decreased? Will Monica be able to measure this decrease in separation? To answer these questions, use the following method of similar triangles (Figure 13) or your own method.

> Assume that the air neither slows down nor deflects either marble from its straight-line course. Then each marble falls from rest toward the center of Earth, as indicated by arrows in Figure 13. Solve the problem using the ratio of sides of similar triangles abc and $a'b'c'$. These triangles

are upside down with respect to one another, but they are similar because their respective sides are parallel. We know the lengths of some of these sides (some greatly exaggerated in the figure): Side $b'c' = 600$ meters; side bc is effectively equal to the r-coordinate of Earth; side $ab = 1$ meters equals half of the original separation of the marbles; side $a'b'$ equals *half the change* in their separation after a drop of 600 meters.

A. Use the ratio of sides of similar triangles to find the "half change" in separation as the two marbles fall 600 meters. From this result, find the entire change in separation between the marbles.

B. Suppose that, as she steps off the plank, Monica releases the two marbles from rest with a *vertical* separation of 2 meters. From Newton's equations (36) and (37), find the increase in separation of two marbles after they fall 600 meters, under the assumption that the marbles fall in a vacuum.)

C. Re-derive your result of Part A using the simpler Part B plus equation (45).

4. Measure your global radial coordinate r near a black hole?

You are the captain of a spaceship with rockets blasting as you descend slowly toward a black hole along a radial line. In effect, you stand for a minute on each shell, then step downward sequentially to the next shell below. From earlier observations you know the value of the black hole mass M and would like to measure your map r-coordinate in order to be sure you are not near the event horizon.

A. Describe how you can determine r from the initial acceleration of a test particle as you descend.

B. Oops! Is there a paradox here? You have measured a map quantity, r, using observations on a local shell. Isn't that illegal?

5. Spaceship approach at relativistic speed

The present chapter assumes that the approaching spaceship moves slowly—not at relativistic speed—with respect to the black hole, so that $E/m \approx 1$. But the captain of the approaching spaceship does not want to waste valuable rocket fuel to slow down in order to apply the analysis of this chapter. She decides not to reduce the large value of her map energy E/m (with respect to the black hole) and instead to use her main thrusters to adjust the value of her map angular momentum $L/(mM)$ so that she moves directly to a knife-edge orbit. If the rocket thrust that increase L/m also increases E/m, no problem: Just use the final value of E/m in what follows.

A. For a large value of map energy $E/m \gg 1$, the r-value of the knife-edge orbit is only slightly greater than 3M. Set $r/M = 3(1+\delta)$ in (8). Show that:

$$\frac{E}{m} \approx \frac{1}{3\delta^{1/2}} \qquad (E/m \gg 1, \text{ knife-edge orbit}) \tag{51}$$

so that for the given large value of E/m,

$$\delta^{1/2} \approx \frac{m}{3E} \qquad (E/m \gg 1, \text{ knife-edge orbit}) \tag{52}$$

B. Show that for this case, equation (6) for the knife-edge orbit becomes:

$$\frac{L}{mM} \approx \left(\frac{3}{\delta}\right)^{1/2} = 3^{3/2}\frac{E}{m} \qquad (E/m \gg 1, \text{ knife-edge orbit}) \tag{53}$$

C. When observations are complete, how does the commander move away from the black hole? Give a general description of this maneuver; don't sweat the details.

6. Swoop Orbit

Figure 14 shows the effective potential for a so-called **swoop orbit** of a stone whose map energy E/m is slightly smaller than that of the effective potential peak at small r-value.

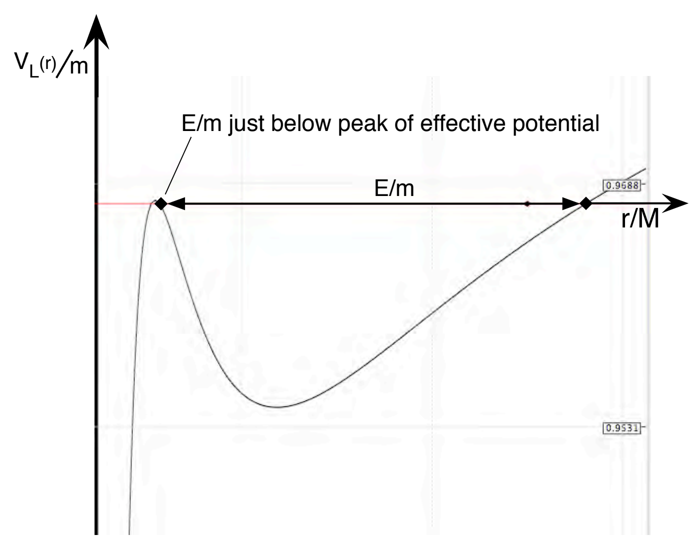


FIGURE 14 Exercise 6: Effective potential for the **swoop orbit** of a stone with map energy E/m just below the (left-hand peak) of the effective potential.

A. Make a rough sketch of the swoop orbit on the $[r, \phi]$ slice. *Optional:* Use interactive softward GRorbits to create and print this swoop orbit.

Luc Longtin is a junior engineer at the Space Agency. He claims that with a small rocket thrust he can put the entire incoming spaceship into a swoop orbit that oscillates between $r = 4M$ and $r = 100M$. This will allow direct observations from the spaceship at r-values between these two limits, completely eliminating the need for probes.

The Space Agency rejects Luc's plan as too risky. Luc invites you, the Chief Engineer, to a bar where he tries to convince you to that the Space Agency should reverse its decision and use his plan. Luc lays out his proposal as follows:

B. Luc begins, "Look at the effective potential for $L/(mM) = 4$ in Figure 6. The inner peak of this effective potential is at $r = 4M$ with $E/m = 1$ and the spaceship approaches from far away with $E/m = 1 + \epsilon$, where $\epsilon = 0.001$. My plan is that when the spaceship reaches, say $r = 20$, it uses a tiny rocket thrust to flip its map energy to $E/m = 1 - \epsilon$ without changing its angular momentum (so the effective potential does not change). Let engineers worry about details of that thrust; just look at the result. The spaceship enters a swoop orbit that bounces off the effective potential peak just outside $r = 4M$. At that bounce, $dr/d\tau = 0$, so equation (17) in Section 8.4 becomes"

$$\frac{dr}{d\tau} = 0 = \left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{m^2 r^2}\right) \qquad (54)$$

$$0 = (1-\epsilon)^2 - \left(1 - \frac{2M}{r}\right)\left(1 + \frac{16M^2}{r^2}\right) \qquad (55)$$

$$0 = 32\left(\frac{M}{r}\right)^3 - 16\left(\frac{M}{r}\right)^2 + 2\left(\frac{M}{r}\right) - \left[1 - (1-\epsilon)^2\right] \qquad (56)$$

Fill in the steps between (55) and (56).

C. Luc continues, "We set up equation (56) for the bounce point near $r = 4M$. But this equation has only global map quantities in it, so is also correct for the bounce point at the large r-value at the outward end of the swoop orbit. At this large r-value, the first term on the right of (56) is small compared to the other terms, so neglect this first term. What remains is a quadratic in the small quantity M/r. Solve this quadratic to show that the only acceptable solution for large r/M is $M/r = \epsilon$ or $r = M/\epsilon = 100M$ for the right-hand bounce point of the swoop orbit."

Verify Luc's calculations.

C. Luc concludes, "So a very small rocket thrust installs the entire incoming spaceship in a swoop orbit that moves in and out between

$r = 100M$ and an r-value slightly greater than $r = 4M$. No need for those silly probes. Astronauts can make observations in this orbit as long as they want as they move in and out. When they finish, a small rocket thrust similar to that described in Item B (during the outgoing portion of its orbit) flips the spaceship map energy back to $E/m = 1 + \epsilon$, so the spaceship escapes the black hole."

Do you agree with this part of Luc's plan?

Will you recommend Luc's program to the Space Agency?

9.9 ■ REFERENCES

Initial Einstein quote from *The New Quotable Einstein*, Alice Caprice, Editor, Princeton University Press, 2005, page 19.

Figure 1 from Misner, Thorne, and Wheeler blahblah, page XX

Hohmann transfer orbits: "The Attainability of Heavenly Bodies," NASA Technical Translation F-44, page 76. Original publication: Walter Hohmann, "Die Erreichbarkeit der Himmelskörper" in R. Oldenbourg (Munich-Berlin), 1925.

Larry Niven, *Neutron Star*, quote from page 26, A Del Ray Book, Ballantine Books, New York, 1968.

"Swimming in Spacetime: Motion by Cyclic Changes in Body Shape," Jack Wisdom, *Science*, 21 March 2003, pages 1865-1869.

Exercise 3 on tidal displacement near Earth and Figure 11 adapted from Edwin F. Taylor and John Archibald Wheeler, *Spacetime Physics*, second edition, 1990, W. H. Freeman, New York, page 46.

Dubai Tower image from Skidmore, Owings and Merrill at `http://www.emporis.com/application/?nav=building&lng=3&id=182168*`

Most of the orbit plots in this chapter were made using the interactive software GRorbits.

Download File Name: Ch09Orbiting160404v1.pdf

Chapter 10
Advance of Mercury's Perihelion

- *What does "advance of the perihelion" mean?*
- *You say Newton does not predict any advance of Mercury's perihelion in the absence of other planets. Why not?*
- *The advance of Mercury's perihelion is tiny. So why should we care?*
- *Why pick out Mercury? Doesn't the perihelion of every planet change with Earth-time?*
- *You are always shouting at me to say whose time measures various motions. Why are you so sloppy about time in analyzing Mercury's orbit?*

Download file name: Ch10AdvanceOfMercurysPerihelion160401v1.pdf

CHAPTER 10 Advance of Mercury's Perihelion

Edmund Bertschinger & Edwin F. Taylor *

This discovery was, I believe, by far the strongest emotional experience in Einstein's scientific life, perhaps in all his life. Nature had spoken to him. He had to be right. "For a few days, I was beside myself with joyous excitement." Later, he told Fokker that his discovery had given him palpitations of the heart. What he told de Haas is even more profoundly significant: when he saw that his calculations agreed with the unexplained astronomical observations, he had the feeling that something actually snapped in him.

—Abraham Pais

10.1 ■ JOYOUS EXCITEMENT

Tiny effect; large significance.

"Perihelion precession"?

What discovery sent Einstein into "joyous excitement" in November 1915? It was his calculation showing that his brand new (not quite completed) theory of general relativity gave the correct value for one detail of the orbit of the planet Mercury that had not been previously explained, an effect with the technical name **precession of Mercury's perihelion**.

Mercury (and every other planet) circulates around the Sun in a not-quite-circular orbit. In this orbit it oscillates in and out radially while it circles tangentially. A full Newtonian analysis predicts an elliptical orbit. Newton tells us that if we consider only the interaction between Mercury and the Sun, then the time for one 360-degree trip around the Sun is *exactly* the same as the time for one in-and-out radial oscillation. Therefore the orbital point closest to the Sun, the so-called **perihelion,** stays in the same place; the elliptical orbit does not shift around with each revolution—according to Newton. You will begin by verifying his nonrelativistic prediction for the simple Sun-Mercury system.

Newton: Sun-Mercury perihelion fixed.

However, observation shows that Mercury's orbit does indeed change. The perihelion moves forward in the direction of rotation of Mercury; it *advances*

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

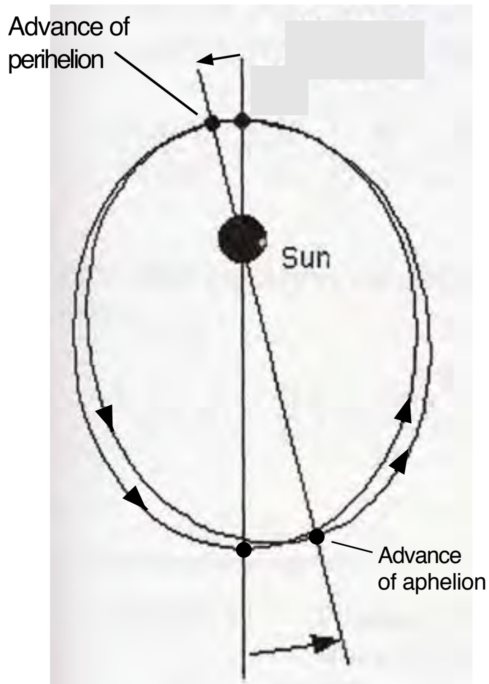


FIGURE 1 Exaggerated view of the advance, during one century, of Mercury's perihelion (and aphelion). The figure shows two elliptical orbits. One of these orbits is the one that Mercury traces over and over again in the year, say, 1900. The other is the elliptical orbit that Mercury traces over and over again in the year, say, 2000. The two are shifted with respect to one another, a rotation called *the advance (or precession) of Mercury's perihelion.* The unaccounted-for precession in one Earth-century is about 43 arcseconds, less than the thickness of a line in this figure.

Observation: perihelion advances.

with each orbit (Figure 1). The long ("major") axis of the ellipse rotates. We call this rotation of the axis the **advance (or precession) of the perihelion**.

The **aphelion** is the point of the orbit farthest from the Sun; it advances at the same angular rate as the perihelion (Figure 1).

Newton: Influence of other planets, predicts most of the perihelion advance . . .

Observation shows that the perihelion of Mercury precesses at the rate of 574 arcseconds (0.159 degree) *per Earth-century.* (One degree equals 3600 arcseconds.) Newton's mechanics accounts for 531 seconds of arc of this advance by computing the perturbing influence of the other planets. But a stubborn 43 arcseconds (0.0119 degree) per Earth-century, called a **residual**, remains after all these effects are accounted for. This residual (though not its modern value) was computed from observations by Urbain Le Verrier as early as 1859 and more accurately later by Simon Newcomb (Box 1). Le Verrier attributed the residual in Mercury's orbit to the presence of an unknown inner planet, tentatively named Vulcan. We know now that there is no planet Vulcan. (Sorry, Mr. Spock!)

. . . but leaves a *residual.*

Box 1. Simon Newcomb

FIGURE 2 Simon Newcomb
Born 12 March 1835, Wallace, Nova Scotia.
Died 11 July 1909, Washington, D.C.
(Photo courtesy of Yerkes Observatory)

From 1901 until 1959 and even later, the tables of locations of the planets (so-called **ephemerides**) used by most astronomers were those compiled by Simon Newcomb and his collaborator George W. Hill.

By the age of five Newcomb was spending several hours a day making calculations, and before the age of seven was extracting cube roots by hand. He had little formal education but avidly explored many technical fields in the libraries of Washington, D. C. He discovered the *American Ephemeris and Nautical Almanac,* of which he said, "Its preparation seemed to me to embody the highest intellectual power to which man had ever attained."

Newcomb became a "computer" (a person who computes) in the American Nautical Almanac office and by stages rose to become its head. He spent the greater part of the rest of his life calculating the motions of bodies in the solar system from the best existing data. Newcomb collaborated with Q. M. W. Downing to inaugurate a worldwide system of astronomical constants, which was adopted by many countries in 1896 and officially by all countries in 1950.

The advance of the perihelion of Mercury computed by Einstein in 1914 would have been compared to entries in the tables of Simon Newcomb and his collaborator.

Einstein correctly predicts residual precession.

Newton's mechanics says that there should be *no residual* advance of the perihelion of Mercury's orbit and so cannot account for the 43 seconds of arc per Earth-century which, though tiny, is nevertheless too large to be ignored or blamed on observational error. But Einstein's general relativity accounted for the extra 43 arcseconds on the button. Result: joyous excitement!

Method: Compare in-and-out time with round-and-round time for Mercury.

Preview, Newton: This chapter begins with Newton's approximations that lead to his no-precession conclusion (in the absence of other planets). Mercury moves in a near-circular orbit; Newton calculates the time for one orbit. The approximation also describes the small radial in-and-out motion of Mercury as if it were a harmonic oscillator moving back and forth about a potential energy minimum (Figure 3). Newton calculates the time for one in-and-out radial oscillation and compares it with the time for one orbit. The orbital and radial oscillation T-values are exactly equal (according to Newton), provided one considers only the Mercury-Sun interaction. He concludes that Mercury circulates around once in the same time that it oscillates radially inward and back out again. The result is an elliptical orbit that closes on itself. In the absence of other planets, Mercury repeats this exact elliptical path forever—according to Newton.

Preview, Einstein: In contrast, our general relativity approximation shows that these two times—the orbital round-and-round and the radial in-and-out T-values—are *not quite equal.* The radial oscillation takes place more slowly, so that by the time Mercury returns to its inner limit, the

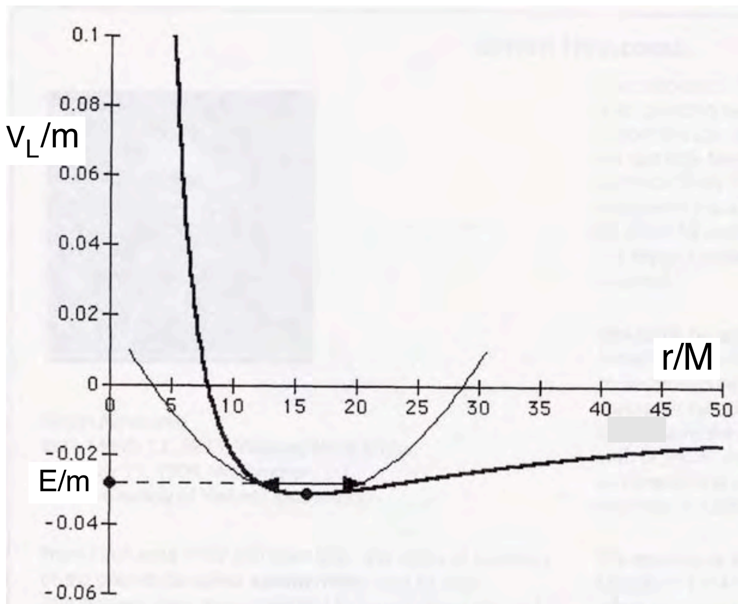


FIGURE 3 Newton's effective potential, equation (5) (heavy curve), on which we superimpose the parabolic potential of the simple harmonic oscillator (thin curve) with the shape given by equation (3). Near the minimum of the effective potential, the two curves closely conform to one another.

circular motion has carried it farther around the Sun than it was at the preceding minimum r-coordinate. From this difference Einstein reckons the residual angular rate of advance of Mercury's perihelion around the Sun and shows that this predicted difference is close to the observed residual advance. Now for the details.

Comment 1. Relaxed about Newton's time and coordinate T
In this chapter we speak freely about Newton's time or Einstein's change in global T-value, without worrying about which we are talking about. We get away with this sloppiness for two reasons: (1) All observations are made from Earth's surface. Every statement about time should in principle be followed by the phrase, "as observed on Earth." (2) For this system, the effects of spacetime curvature on the rates of local clocks are so small that all time or T-measures give essentially the same rate of precession, as summarized in Section 10.11.

10.2 ■ NEWTON'S SIMPLE HARMONIC OSCILLATOR

Assume radial oscillation is sinusoidal.

Why does the planet oscillate in and out radially? Look at the effective potential in Newton's analysis of motion, the heavy line in Figure 3. This heavy line has a minimum, the location at which the planet can ride around at constant r-value, tracing out a circular orbit. But with a slightly higher energy, it not only moves tangentially, it also oscillates radially in and out, as shown by the two-headed arrow in Figure 3.

How long does it take for one in-and-out oscillation? That depends on the shape of the effective potential curve near the minimum shown in Figure 3. But if the amplitude of the oscillation is small, then the effective part of the curve is very close to this minimum, and we can use a well-known mathematical theorem: If a continuous, smooth curve has a local minimum, then near that minimum a parabola approximates this curve. Figure 3 shows such a parabola (thin curve) superimposed on the (heavy) effective potential curve. From the diagram it is apparent that the parabola is a good approximation of the potential, at least near that local minimum.

In-and-out motion in parabolic potential . . .

. . . predicts simple harmonic motion.

From introductory Newtonian mechanics, we know how a particle moves in a parabolic potential. The motion is called **simple harmonic oscillation**, described by the following expression:

$$x = \mathrm{A} \sin \omega t \qquad (1)$$

Here A is the amplitude of the oscillation and ω (Greek lower case omega) tells us how rapidly the oscillation occurs in radians per unit time. The potential energy per unit mass, V/m, of a particle oscillating in a parabolic potential follows the formula

$$\frac{V}{m} = \frac{1}{2}\omega^2 x^2 \qquad (2)$$

To find the rate of oscillation ω of the harmonic oscillator, take the second derivative with respect to x of both sides of (2).

$$\frac{d^2\left(V/m\right)}{dx^2} = \omega^2 \qquad (3)$$

10.3 ■ NEWTON'S ORBIT ANALYSIS

Round and round vs. in and out

Newton's equilibrium r_0

The in-and-out radial oscillation of Mercury does not take place around $r = 0$ but around the r-value of the effective potential minimum. What is the r-coordinate of this minimum (call it r_0)? Start with Newton's equation (23) in Section 8.4:

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 = \frac{E}{m} - \left(-\frac{M}{r} + \frac{L^2}{2m^2r^2}\right) = \frac{E}{m} - \frac{V_{\mathrm{L}}(r)}{m} \qquad \text{(Newton)} \qquad (4)$$

This equation defines the effective potential,

$$\frac{V_{\rm L}(r)}{m} \equiv -\frac{M}{r} + \frac{L^2}{2m^2r^2} \qquad \text{(Newton)} \qquad (5)$$

To locate the minimum of this effective potential, set its derivative equal to zero:

$$\frac{d(V_{\rm L}/m)}{dr} = \frac{M}{r^2} - \frac{L^2}{m^2r^3} = 0 \qquad \text{(Newton)} \qquad (6)$$

Solve the right-hand equation to find r_0, the r-value of the minimum:

$$r_0 = \frac{L^2}{Mm^2} \qquad \text{(Newton, equlibrium radius)} \qquad (7)$$

Newton: In-and-out time equals round-and-round time.

We want to compare the rate $\omega_{\rm r}$ of in-and-out radial motion of Mercury with its rate ω_ϕ of round-and-round tangential motion. Use Newton's definition of angular momentum, with increment dt of Newton's universal time, similar to equation (10) of Section 8.2:

$$\frac{L}{m} \equiv r^2\frac{d\phi}{dt} = r^2\omega_\phi \qquad \text{(Newton)} \qquad (8)$$

where $\omega_\phi \equiv d\phi/dt$. Equation (8) gives us the angular velocity of Mercury along its almost-circular orbit.

Queries 1 and 2 show that for Newton the radial in-and-out angular velocity $\omega_{\rm r}$ is equal to the orbital angular velocity ω_ϕ.

QUERY 1. Newton's angular velocity ω_ϕ of Mercury in orbit.
Set $r = r_0$ in (8) and substitute the result into (7). Show that at the equilibrium radius, $\omega_\phi^2 = M/r_0^3$ for Newton.

QUERY 2. Newton's radial oscillation rate ω_r for Mercury's orbit
We want to use (3) to find the angular rate of radial oscillation. Accordingly, take the second derivative of $V_{\rm L}$ in (5) with respect to r. Set $r = r_0$ in the resulting expression and substitute your value for L^2 in (7). Use (3) to show that at Mercury's orbital radius, $\omega_r^2 = M/r_0^3$, according to Newton.

Important result: *For Newton, Mercury's perihelion does not advance when one considers only the gravitational interaction between Mercury and the Sun.*

10.4 ■ EFFECTIVE POTENTIAL: EINSTEIN

Extra effective potential term advances perihelion.

Now we repeat the analysis of radial and tangential orbital motion for the general relativistic case. Chapter 9 predicts the radial motion of an orbiting satellite. Multiply equations (4) and (5) of Section 9.1 through by 1/2 to obtain an equation similar to (4) above for the Newton's case:

$$\begin{aligned}\frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 &= \frac{1}{2}\left(\frac{E}{m}\right)^2 - \frac{1}{2}\left(1-\frac{2M}{r}\right)\left(1+\frac{L^2}{m^2r^2}\right) \qquad (9)\\ &= \frac{1}{2}\left(\frac{E}{m}\right)^2 - \frac{1}{2}\left(\frac{V_\mathrm{L}(r)}{m}\right)^2 \qquad \text{(Einstein)}\end{aligned}$$

Set up general relativity effective potential.

Equations (4) and (9) are of similar form, and we use this similarity to make a general relativistic analysis of the harmonic radial motion of Mercury in orbit. In this process we adopt the *algebraic manipulations* of Newton's analysis in Sections 10.2 and 10.3 but apply them to the general relativistic expression (9).

Different time rates of different clocks do not matter.

Before we proceed, note three characteristics of equation (9). First, $d\tau$ on the left side of (9) is the differential wristwatch time $d\tau$, not the differential dt of Newton's universal time t. This different reference time is not necessarily fatal, since we have not yet decided which relativistic measure of time should replace Newton's universal time t. You will show in Section 10.11 that for Mercury the choice of which time to use (wristwatch time, global map T-coordinate, or even shell time at the r-value of the orbit) makes a negligible difference in our predictions about the rate of advance of the perihelion.

Note, second, that in equation (9) the relativistic expression $(E/m)^2$ stands in the place of the Newtonian expression E/m in (4). However, both are constant quantities, which is all that matters in the analysis.

Evidence that we are on the right track results when we multiply out the second term of the first line of (9), which is the square of the effective potential, equation (18) of Section 8.4, with the factor one-half. Note that we have assigned the symbol $(1/2)(V_\mathrm{L}/m)^2$ to this second term.

$$\begin{aligned}\frac{1}{2}\left(\frac{V_\mathrm{L}(r)}{m}\right)^2 &= \frac{1}{2}\left(1-\frac{2M}{r}\right)\left(1+\frac{L^2}{m^2r^2}\right) \qquad \text{(Einstein)} \qquad (10)\\ &= \frac{1}{2} - \frac{M}{r} + \frac{L^2}{2m^2r^2} - \frac{ML^2}{m^2r^3}\end{aligned}$$

Details of relativistic effective potential

The heavy curve in Figure 4 plots this function. The second line in (10) contains the two effective potential terms that made up the Newtonian expression (5). The final term on the right of the second line of (10) describes an added attractive potential from general relativity. For the Sun-Mercury case at the r-value of Mercury's orbit, this term leads to the slight precession of the elliptical orbit. As r becomes small, the r^3 in the denominator causes this term to overwhelm all other terms in (10), which results in the downward plunge in the effective potential at the left side of Figure 4.

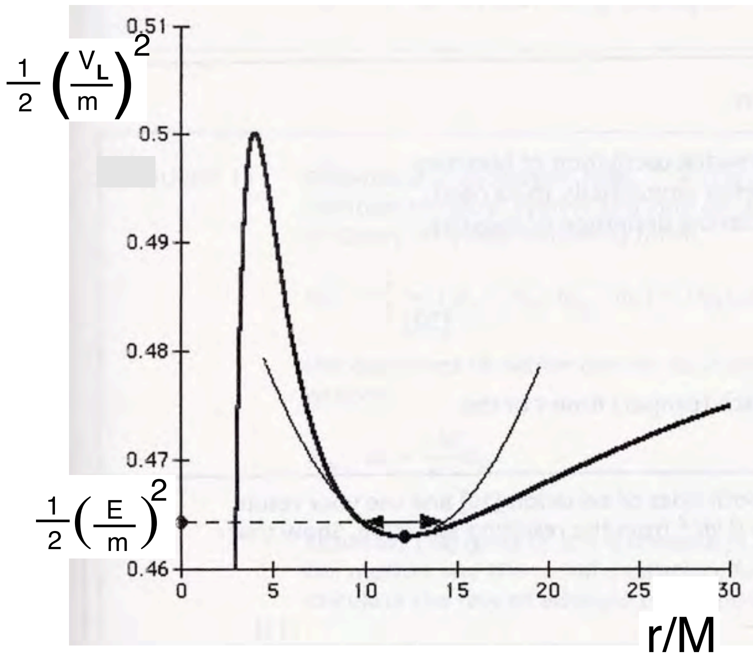


FIGURE 4 General-relativistic effective potential $(V_L/m)^2/2$ (heavy curve) and its approximation at the local minimum by a parabola (light curve) in order to analyse the radial excursion (double-headed arrow) of Mercury as simple harmonic motion. The effective potential curve is for a black hole, not for the Sun, whose effective potential near the potential minimum would be indistinguishable from the Newton's effective potential on the scale of this diagram. However, this minute difference accounts for the tiny residual precession of Mercury's orbit.

Finally, note third that the last term $(1/2)(V_L/m)^2$ in relativistic equation (9) takes the place of the Newton's effective potential V_L/m in equation (4).

In summary, we can manipulate general relativistic expressions (9) and (10) in nearly the same way that we manipulated Newton's expressions (4) and (5) in order to analyze the radial component of Mercury's motion and small perturbations of Mercury's elliptical orbit brought about by general relativity.

10.5 ■ EINSTEIN'S ORBIT ANALYSIS

Einstein tweaks Newton's solution.

Now analyze the radial oscillation of Mercury's orbit according to Einstein.

QUERY 3. Local minimum of Einstein's effective potential

Take the first derivative of the squared effective potential (10) with respect to r, that is find $d[(1/2)(V_L/m)^2]/dr$. Set this first derivative aside for use in Query 4. As a separate calculation, equate this derivative to zero, set $r = r_0$, and solve the resulting equation for the unknown quantity $(L/m)^2$ in terms of the known quantities M and r_0.

QUERY 4. Einstein's radial oscillation rate ω_r for Mercury in orbit.
We want to use (3) to find the rate of oscillation ω_r in the radial direction.

A. Take the second derivative of $(1/2)(V_\mathrm{L}/m)^2$ from (10) with respect to r. Set the resulting $r = r_0$ and substitute the expression for $(L/m)^2$ from Query 3 to obtain

$$\left[\frac{d^2}{dr^2}\left(\frac{1}{2}\frac{V_\mathrm{L}^2}{m^2}\right)\right]_{r=r_0} = \omega_r^2 = \frac{M}{r_0^3}\frac{\left(1-\dfrac{6M}{r_0}\right)}{\left(1-\dfrac{3M}{r_0}\right)} \qquad \text{(Einstein)} \qquad (11)$$

$$\approx \frac{M}{r_0^3}\left(1-\frac{6M}{r_0}\right)\left(1+\frac{3M}{r_0}\right) \qquad (12)$$

$$\approx \frac{M}{r_0^3}\left(1-\frac{3M}{r_0}\right) \qquad (13)$$

where we have made repeated use of the approximation inside the front cover in order to find a result to first order in the fraction M/r.

B. For our Sun, $M \approx 1.5 \times 10^3$ meters, while for Mercury's orbit $r_0 \approx 6 \times 10^{10}$ meters. Does the value of M/r_0 justify the approximations in equations (12) and (13)?

Note that the coefficient M/r_0^3 in these three equations equals Newton's expression for ω_r^2 derived in Query 1.

Now compare ω_r, the in-and-out oscillation of Mercury's orbital r-coordinate with the angular rate ω_ϕ with which Mercury moves tangentially in its orbit. The rate of change of azimuth ϕ springs from the definition of angular momentum in equation (10) in Section 8.2:

$$\frac{L}{m} = r^2\frac{d\phi}{d\tau} \qquad \text{(Einstein)} \qquad (14)$$

Note the differential wristwatch time $d\tau$ for the planet.

QUERY 5. Einstein's angular velocity
Square both sides of (14) and use your result from Query 3 to eliminate L^2 from the resulting equation. Show that at the equilibrium r_0 the result can be written

$$\omega_\phi^2 \equiv \left(\frac{d\phi}{d\tau}\right)^2 = \frac{M}{r_0^3}\left(1 - \frac{3M}{r_0}\right)^{-1} \qquad \text{(Einstein)} \tag{15}$$

$$\approx \frac{M}{r_0^3}\left(1 + \frac{3M}{r_0}\right) \tag{16}$$

where again we use our approximation inside the front cover. Compare this result with equation (13) and with Newton's result in Query 1.

10.6 ■ PREDICT MERCURY'S PERIHELION ADVANCE

Simple outcome, profound consequences

Einstein: in-out rate differs from circulation rate.

According to Einstein, the advance of Mercury's perihelion springs from the difference between the frequency with which the planet sweeps around in its orbit and the frequency with which it oscillates in and out in r. In Newton's analysis these two frequencies are equal (for the interaction between Mercury and the Sun). But Einstein's theory shows that these two frequencies are *slightly* different; Mercury reaches its minimum r (its perihelion) at an incrementally greater angular position in each successive orbit. *Result:* the advance of Mercury's perihelion. In this section we compare Einstein's prediction with observation. But first we need to define what we are calculating.

What do we mean by the phrase "the period of a planet's orbit"? The period with respect to what? Here we choose what is technically called the **synodic period** of a planet, defined as follows:

Definition: synodic period

DEFINITION 1. Synodic period of a planet

The **synodic period** of a planet is the lapse in time (Newton) or lapse in global T-value (Einstein) for the planet to revolve once around the Sun with respect to the fixed stars.

"Fixed" stars?

Comment 2. Fixed stars?

What are the "fixed stars"? Chapter 14 The Expanding Universe shows that stars are anything but fixed. With respect to our Sun, stars move! However, stars that we now know to be very distant do not change angle rapidly from our point of view. Over a few hundred years—the lifetime of the field of astronomy itself—these stars may be called *fixed*.

The value T_r to make a complete in-and-out radial oscillation is

$$T_r \equiv \frac{2\pi}{\omega_r} \qquad \text{(period of radial oscillation)} \tag{17}$$

In global coordinate lapse T_r, Mercury goes around the Sun, completing an angle

$$\omega_\phi T_{\rm r} = \frac{2\pi\omega_\phi}{\omega_{\rm r}} = (\text{Mercury revolution angle in } T_{\rm r}) \tag{18}$$

which exceeds one complete revolution in radians by:

$$\omega_\phi T_{\rm r} - 2\pi = T_{\rm r}\left(\omega_\phi - \omega_{\rm r}\right) = (\text{excess angle per revolution}) \tag{19}$$

QUERY 6. Difference in Einstein's oscillation rates

The two angular rates ω_ϕ and ω_r are *almost* identical in value, even in the Einstein analysis. Therefore we can write approximately:

$$\omega_\phi^2 - \omega_{\rm r}^2 = (\omega_\phi + \omega_{\rm r})(\omega_\phi - \omega_{\rm r}) \approx 2\omega_\phi(\omega_\phi - \omega_{\rm r}) \tag{20}$$

A. Substitute equations (13) and (16) into the left side of (20):

$$\omega_\phi^2 - \omega_{\rm r}^2 \approx \frac{M}{r_0^3}\left[\left(1 + \frac{3M}{r_0}\right) - \left(1 - \frac{3M}{r_0}\right)\right] = \frac{M}{r_0^3}\frac{6M}{r_0} \tag{21}$$

B. Equation (20) becomes:

$$\omega_\phi^2 - \omega_{\rm r}^2 \approx \frac{M}{r_0^3}\frac{6M}{r_0} \approx \omega_\phi^2\frac{6M}{r_0} \approx 2\omega_\phi(\omega_\phi - \omega_{\rm r}) \tag{22}$$

C. Simplify the right-hand equation in (22), write the result as:

$$\omega_\phi - \omega_{\rm r} \approx \frac{3M}{r_0}\omega_\phi \qquad \text{(angular rates, Einstein)} \tag{23}$$

Equation (23) shows the difference in angular velocity between the tangential motion and the radial oscillation. From this rate difference we will calculate the advance of the perihelion of Mercury in one Earth-century.

Comment 3. What is X?

Symbols ω in (23) express rotation rates in radians per unit of—what? *Question:* What is X in the denominator of $d\phi/dX \equiv \omega$? Does X equal global coordinate T? planet wristwatch time τ? shell time $t_{\rm shell}$ at the average r-value of the orbit? *Answer:* It does not matter which of these quantities X represents, as long as this measure is the *same* on both sides of any resulting equation. Comment 1 told us to be relaxed about time. In the following Queries you use (23) to calculate the precession rate of Mercury in radians/second, then to convert this result to arcseconds/Earth-century.

10.7 ■ COMPARE PREDICTION WITH OBSERVATION

Check out Einstein!

Now compare our approximate relativistic prediction with observation.

QUERY 7. Mercury's angular velocity

The synodic period of Mercury's orbit is 7.602×10^6 seconds. To one significant digit, $\omega_\phi \approx 8 \times 10^{-7}$ radian/second. What is its value to three significant digits?

QUERY 8. Calculated coefficient

The mass M of the Sun is 1.477×10^3 meters and r_0 of Mercury's orbit is 5.80×10^{10} meters. To one significant digit, the coefficient $3M/r_0$ in (23) is 1×10^{-7}. Find this result to three significant digits.

QUERY 9. Advance of Mercury's perihelion in radians/second

From equation (23) and results of Queries 7 and 8, derive a numerical prediction of the advance of the perihelion of Mercury's orbit in radians/second. To one significant digit the result is 6×10^{-14} radians/second. Find the result to three significant digits.

QUERY 10. Advance of Mercury's perihelion in arcseconds per Earth-century.

Estimate the general relativity prediction of advance of Mercury's perihelion in arcseconds per century. Use results from preceding queries plus conversion factors inside the front cover plus the definition that 3600 arcseconds equals one degree. To one significant digit, the answer is 40 arcseconds/century. Find the result to three significant digits.

Observation and careful calculation agree.

A more accurate relativistic analysis predicts 42.980 arcseconds (0.011939 degrees) per Earth-century (Table 10.1). The observed rate of advance of the perihelion is in perfect agreement with this value: 42.98 ± 0.1 arcseconds per Earth-century. By what percentage did your prediction differ from observation?

10.8 ■ ADVANCE OF THE PERIHELIA OF THE INNER PLANETS

Help from a supercomputer.

All planet orbits precess.

Do the *perihelia* (plural of *perihelion*) of other planets in the solar system also advance as described by general relativity? Yes, but these planets are farther from the Sun, and their orbits are less eccentric, so the magnitude of the predicted advance is less than that for Mercury. In this section we compare our

TABLE 10.1 Advance of the perihelia of the inner planets

Planet	Advance of perihelion in seconds of arc per Earth-century (JPL calculation)	r-value of orbit in AU*	Period of orbit in years
Mercury	42.980 ± 0.001	0.38710	0.24085
Venus	8.618 ± 0.041	0.72333	0.61521
Earth	3.846 ± 0.012	1.00000	1.00000
Mars	1.351 ± 0.001	1.52368	1.88089

*Astronomical Unit (AU): average r-value of Earth's orbit; inside front cover.

estimated advance of the perihelia of the inner planets Mercury, Venus, Earth, and Mars with results of an accurate calculation.

Computer analysis of precessions.

The Jet Propulsion Laboratory (JPL) in Pasadena, California, supports an active effort to improve our knowledge of the positions and velocities of the major bodies in the solar system. For the major planets and the moon, JPL maintains a database and set of computer programs known as the Solar System Data Processing System. The input database contains the observational data measurements for current locations of the planets. Working together, more than 100 interrelated computer programs use these data and the relativistic laws of motion to compute locations of planets at in the past and the future. The equations of motion take into account not only the gravitational interaction between each planet and the Sun but also interactions among all planets, Earth's moon, and 300 of the most massive asteroids, as well as interactions between Earth and Moon due to nonsphericity and tidal effects.

JPL multi-program computation.

To help us with our project on perihelion advance, Myles Standish, Principal Member of the Technical Staff at JPL, kindly used the numerical integration program of the Solar System Data Processing System to calculate orbits of the four inner planets over four centuries, from A.D. 1800 to A.D. 2200. In an overnight run he carried out this calculation twice, first with the full program including relativistic effects and second "with relativity turned off." Standish "turned off relativity" by setting the speed of light to 10^{10} times its measured value, making light speed effectively infinite.

For each of the two runs, the perihelia of the four inner planets were computed for the four centuries. The results from the nonrelativistic run were subtracted from those of the relativistic run, revealing advances of the perihelia per Earth-century accounted for only by general relativity. The second column of Table 10.1 shows the results, together with the estimated computational error.

QUERY 11. Approximate advances of the perihelia of the inner planets

Compare the JPL-computed advances of the perihelia of Venus, Earth, and Mars in Table 10.1 with approximate results calculated using equation (23).

10.9 ■ CHECK THE STANDARD OF TIME

Whose clock?

We have been casual about whose time tracks the advance of the perihelion of Mercury and other planets; we even treated the global T-coordinate as a time, which is against our usual rules. Does this invalidate our approximations?

QUERY 12. Difference between shell time and Mercury's wristwatch time.
Use special relativity to find the fractional difference between planet Mercury's wristwatch time increment $\Delta\tau$ and the time increment Δt_{shell} read on shell clocks at the same average r_0 at which Mercury moves in its orbit at the average velocity 4.8×10^4 meters/second. By what fraction does a change of time from $\Delta\tau$ to Δt_{shell} change the total angle covered in the orbital motion of Mercury in one century? Therefore by what fraction does it change the predicted angle of advance of the perihelion in that century?

QUERY 13. Difference between shell time and global rain map T.
Find the fractional difference between shell time increment Δt_{shell} at r_0 and global map increment ΔT for r_0 equal to the average r-value of the orbit of Mercury. By what fraction does a change from Δt_{shell} to a lapse in global T alter the predicted angle of advance of the perihelion in that century?

QUERY 14. Does the time standard matter?
From your results in Queries 12 and 13, say whether or not the choice of a time standard—wristwatch time of Mercury, shell time, or map t—makes a detectable difference in the numerical prediction of the advance of the perihelion of Mercury in one Earth-century. Would your answer differ if the time were measured with clocks on Earth's surface?

DEEP INSIGHTS FROM MORE THAN THREE CENTURIES AGO

> *Newton himself was better aware of the weaknesses inherent in his intellectual edifice than the generations that followed him. This fact has always roused my admiration.*
>
> —Albert Einstein

We agree with Einstein. In the following quote from the end of his great work *Principia*, Isaac Newton summarizes what he knows about gravity and what he does not know. We find breathtaking the scope of what Newton says—and the integrity with which he refuses to say what he does not know. In the following, "feign" means "invent," and since Newton's time "experimental philosophy" has come to mean "physics."

"I do not 'feign' hypotheses."

> *Thus far I have explained the phenomena of the heavens and of our sea by the force of gravity, but I have not yet assigned a cause to gravity. Indeed, this force arises from some cause that penetrates as far as the centers of the sun and planets without any diminution of its power to act, and that acts not in proportion to the quantity of the* surfaces *of the particles on which it acts (as mechanical causes are wont to do) but in proportion to the quantity of* solid *matter, and whose action is extended everywhere to immense distances, always decreasing as the squares of the distances. Gravity toward the sun is compounded of the gravities toward the individual particles of the sun, and at increasing distances from the sun decreases exactly as the squares of the distances as far as the orbit of Saturn, as is manifest from the fact that the aphelia of the planets are at rest, and even as far as the farthest aphelia of the comets, provided that those aphelia are at rest. I have not as yet been able to deduce from phenomena the reason for these properties of gravity, and I do not "feign" hypotheses. For whatever is not deduced from the phenomena must be called a hypothesis; and hypotheses, whether metaphysical or physical, or based on occult qualities, or mechanical, have no place in experimental philosophy. In this experimental philosophy, propositions are deduced from the phenomena and are made general by induction. The impenetrability, mobility, and impetus of bodies, and the laws of motion and the law of gravity have been found by this method. And it is enough that gravity really exists and acts according to the laws that we have set forth and is sufficient to explain all the motions of the heavenly bodies and of our sea.*
>
> —Isaac Newton

10.10 ■ REFERENCES

Initial quote: Abraham Pais, *Subtle Is the Lord: The Science and the Life of Albert Einstein*, Oxford University Press, New York, 1982, page 253.

Einstein's analysis of precession of Mercury's orbit in *Sitzungberichte der Preussischen Akademie der Wissenschaften zu Berlin,* Volume 11, pages 831-839 (1915). English translation by Brian Doyle in *A Source Book in Astronomy and Astrophysics, 1900-1975,* pages 820-825. Our analysis influenced by that of Robert M. Wald (*General Relativity,* University of Chicago Press, 1984, pages 142–143)

Myles Standish of the Jet Propulsion Laboratory ran the programs on the inner planets presented in Section 10. He also made useful comments on the project as a whole for the first edition.

Precession in binary neutron star systems: For B1913+16: J. M. Weisberg and J. H. Taylor, ApJ, 576, 942 (2002). For J0737-3039: R. P. Breton et al, Science, Vol. 321, 104 (2008)

Final Newton quote from I. Bernard Cohen and Anne Whitman, *Isaac Newton, the Principia, A New Translation*, University of California Press, Berkeley CA, 1999, page 943

Download File Name: Ch10AdvanceOfMercurysPerihelion160401v1.pdf

Chapter 11 Orbits of Light

-

- *What variety of orbits does light follow around a black hole?*
- *Can a black hole reverse the direction of a light flash?*
- *Can light go into a circular orbit around a black hole? if so, is this circular orbit stable?*
- *How many different orbits can light take from a single star to my eye?*

Download File Name: Ch11OrbitsOfLight160511v1.pdf

CHAPTER 11 Orbits of Light

Edmund Bertschinger & Edwin F. Taylor *

Then the sun god Ra emerged out of primal chaos.
—Egyptian creation story

And at once Kiho made his eyes to glow with flame—and the darkness became light.
—Tuamotuan (Polynesian) creation story

And God said, Let there be light: and there was light.
—first Biblical act of creation, Genesis 1:3

He bringeth them out of darkness unto light by His decree . . .
—Qur'an 5:16

Along with death came the Sun the Moon and the stars . . .
—Inuit creation story

11.1 ■ TURN A STONE INTO A LIGHT FLASH

Faster and faster, less and less mass

So far, observers are blind.

Thus far in this book almost all observers have been blind. Chapter 5 defined the shell observer but did not predict what he sees when he looks at stars or other objects outside his local inertial frame. The rain diver as she descends to the singularity (Chapter 7) peers in just two opposite directions—radially inward and radially outward. The explorer in her circular orbit around a black hole (Chapter 8) does not report what she sees—neither the starry heavens around her nor the black hole beneath her. In the present chapter we lay the groundwork to cure this blindness: we plot orbits of light in global map coordinates.

No local observation in this chapter

But this chapter still does not describe what any observer *sees*. Recall that we make every measurements and observation in a local inertial frame. The present chapter describes only map "starlight orbits," for example the orbit

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

that connects remote Star X with an observer at (or passing through) map location Y. The following Chapter 12 will tell us in what direction an observer at Y looks to see Star X.

Seeing is *not* believing.

What can we say about the global motion of light around, past, or into a spherically symmetric, nonspinning black hole? We ask here no small question: Almost every message from events in space comes to us by way of electromagnetic radiation of different frequencies. *Exceptions:* cosmic rays, neutrinos, and gravitational waves. A starlight orbit may deflect as it passes close to a massive object. Near a black hole this deflection can be radical; starlight can even go into a circular orbit. This and the following chapter make clear that for an observer near a black hole, seeing is definitely *not* believing!

Find orbits of light.

How do we plot the global orbit of light around a black hole? This is a new question; up until now we plotted light cones with short legs that sprout from a single event. Now we want to "connect the dots," the events along an entire orbit of light that stretches from a specified distant star to a given local observer near a black hole.

Constant(s) of motion for light?

The *free stone* has two global constants of motion along its worldline: map energy E and map angular momentum L. Chapters 3 and 8 used the Principle of Maximal Aging to derive map expressions for each of these global constants of motion. Can we use the Principle of Maximal Aging to find constant(s) of motion for a light flash?

Principle of Maximal Aging does not apply directly to light.

The Principle of Maximal Aging says that a stone chooses a path across an adjoining tiny pair of segments along its worldline such that its wristwatch time is a maximum between a fixed initial event as the stone enters the pair and a fixed final event as it leaves the pair. But the Principle of Maximal Aging cannot apply directly to light, and for a fundamental reason: *The aging of a light flash along its worldline in a vacuum is automatically zero!* Aging $d\tau$ equals *zero* along every differential increment of the light flash worldline. *Question:* How can we possibly apply the Principle of Maximal Aging to light, whose aging is automatically zero?

Adapt Principle of Maximal Aging to light.

Answer: Sneak up on it! Start in flat spacetime far from a black hole. Think of a series of faster and faster stones, each stone with a smaller mass than the previous one. Let this series occur in such a way that the map energy E remains constant. Far from the black hole, map energy equals the measurable energy in a local inertial shell frame, in which the stone has squared speed v_{shell}^2. Take the limit of equation (28) in Section 1.7 as $m \to 0$ and $v_{\text{shell}} \to 1$:

$$E = \lim_{v_{\text{shell}} \to 1}^{m \to 0} \frac{m}{(1 - v_{\text{shell}}^2)^{1/2}} = \text{constant} \qquad (\text{light, } r/M \gg 1) \qquad (1)$$

Stone $\to$ light as $m \to 0$ and $v \to 1$

The present chapter analyzes consequences of this limit-taking process in (1).

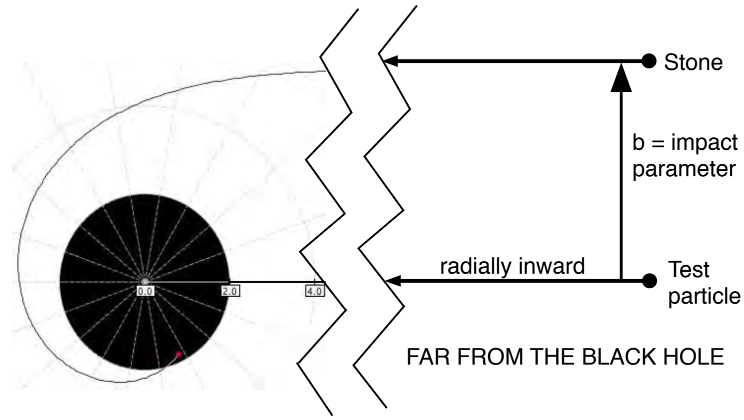


FIGURE 1 Impact parameter b of a stone that approaches the black hole from a far away. Far from the black hole, we define b as the perpendicular offset between the line of motion of the approaching stone and the parallel line of motion of a test particle that makes a dive at constant ϕ into the black hole. Values of b and M determine whether or not the black hole captures the incoming stone.

11.2 ■ IMPACT PARAMETER b

Impact parameter from map angular momentum and map energy

Chapter 8 analyzed circular orbits of a stone around the black hole. Now we want to describe more general orbits of both a stone and a light flash, so we define an orbit.

Definition: orbit

> **DEFINITION 1. Orbit: Stone or light flash**
> An orbit is the worldline of a stone or light flash described by global coordinates. An orbit need not be circular around an origin, it need not be closed, it need not even remain in a bounded region of space.

A starlight orbit is a special case of the orbit:

Definition: starlight orbit

> **DEFINITION 2. Starlight orbit**
> A starlight orbit is the orbit (Definition 1) of a light flash emitted by a star.

Think first about the orbit of a free stone far from the black hole—the right side of Figure 1. Far from the black hole this orbit is straight. How do we measure this orbit to verify that it is straight? As always, carry out measurements in a local inertial frame. We choose a shell frame (Section 5.7). Sufficiently far from the black hole this "local" shell frame can be quite large in the sense that over a significant range of r and ϕ special relativity correctly describes this orbit as a *straight line*. Now find a parallel straight line orbit that—by trial and error—moves without deflection to the center of the black hole (verified by measurement in a series of shell frames on both sides of Figure 1).

"Straight line" verified in local shell frame.

In a local inertial shell frame far from the black hole, we can measure perpendicular distances between parallel orbits. This leads to the definition of the **impact parameter**, with the symbol b. In a preliminary definition, we define the impact parameter of a stone far from the black hole:

Preliminary definition: impact parameter

DEFINITION 3. Impact parameter b of a stone (preliminary)
The impact parameter b of a stone is the perpendicular distance—measured far from the black hole—between the straight orbit of the free stone and the parallel straight orbit of a second stone (test particle) that plunges at constant ϕ into the black hole.

QUERY 1. Every moving stone has an impact parameter
Show that every distant stone that changes global coordinates r or ϕ (or both) has an impact parameter—even a stone that moves away from the black hole.

Thus far the definition of the impact parameter is purely geometric. However, the right side of Figure 1 can be used to define angular momentum. The angular momentum of the stone takes the simple form:

$$L_{\text{far}} \equiv b_{\text{far}} p_{\text{far}} \qquad \text{(stone in distant—flat—spacetime)} \qquad (2)$$

Map angular momentum L

where p_{far} is the momentum of special relativity (Section 1.8). Equation (2) determines the value of L where $r/M \gg 1$, that is where spacetime is flat. However L is a map constant of motion, the same everywhere around the black hole. Therefore its value, calculated from (2) far from the black hole, is the same close to the black hole.

Recall equation (39) for a stone in Section 1.9, with p defined in (2):

$$m^2 = E^2 - p^2 = E^2 - \left(\frac{L}{b}\right)^2 \qquad \text{(stone, flat spacetime)} \qquad (3)$$

Impact parameter of a stone

Solve this equation for b, in which b and L are either both positive or both negative:

$$\boxed{b \equiv \frac{L}{(E^2 - m^2)^{1/2}} \qquad \text{(impact parameter for a stone, everywhere)} \quad (4)}$$

Both map energy E and map angular momentum L are map constants of motion and m is an invariant quantity. Therefore equation (4) is valid close to the black hole as well as far away. Even though it was derived assuming flat spacetime, we take (4) to define b everywhere. Close to the black hole, b is no longer the perpendicular distance of Definition 3. But every orbit has an L and an E and therefore can be assigned a unique value of b.

For light, carry out the limit-taking process demanded in (1), with constant E but decreasing m. The limit $m \to 0$ defines the impact parameter for light:

Impact parameter of a light flash

$$b \equiv \frac{L}{E} \qquad \text{(impact parameter of light, everywhere)} \qquad (5)$$

This leads to the final definition of the impact parameter for a stone or a light flash around a black hole:

Definition: impact parameter b

DEFINITION 4. Impact parameter b

The **impact parameter** b for a stone is given by (4) and for a light flash by (5).

?

Objection 1. *You use two perfectly good constants of motion, L and E and give a geometric interpretation for a combination of them. So what? I can define a thousand combinations of L and E. Who cares? I didn't need any such combination for a stone. Why are you wasting my time?*

!

We introduce b because neither L alone or E alone will be helpful when $m \rightarrow 0$. Equations of motion for light derived below depend only on the fraction L/E and no other combination. Global motion of a stone depends on two constants of motion, L and E. Global motion of light is simpler, completely described by one constant of motion, $b \equiv L/E$. Rejoice!

We have defined impact parameter, but we have not yet predicted the global motion of a light flash near the black hole. To obtain equations of motion for light, we again apply the limit-taking process of equation (1), in this case to the equations of motion for a stone from Chapter 8.

11.3 ■ EQUATIONS OF MOTION FOR LIGHT

A single constant of motion for light, namely b

Flat starlight wavefront approaching the black hole . . .

Light spreads out from a star as a spherical wave. We assume that every star is so far away that as its starlight approaches our black hole—but still travels in flat spacetime—it forms a flat wavefront (right side of Figure 2).

. . . is equivalent to a bundle of parallel straight orbits.

We already have another powerful way to describe starlight in flat spacetime: as a bundle of parallel straight orbits. Figure 2 displays four starlight orbits from a single star, each with a different impact parameter b, as these orbits approach the black hole. Far from the black hole (right side of the figure) these starlight orbits remain parallel to one another. Close to the black hole (left side of the figure) they diverge: Only the orbit with $b/M = 0$ remains straight. Starlight Orbit 1 deflects but escapes; Starlight Orbit 2 enters a circular orbit; Starlight Orbit 3 plunges to the center of the black hole.

Close to the black hole, orbits from the star are neither parallel nor straight.

Critical impact paraneter

Starlight Orbit 2 in Figure 2 is unique; it enters a circular orbit at $r = 3M$. We call this orbit *critical* and its impact parameter the *critical impact parameter*, b_{critical}. In Query 3 you show that the critical impact parameter has the value $b_{\text{critical}} = (27)^{1/2}M$.

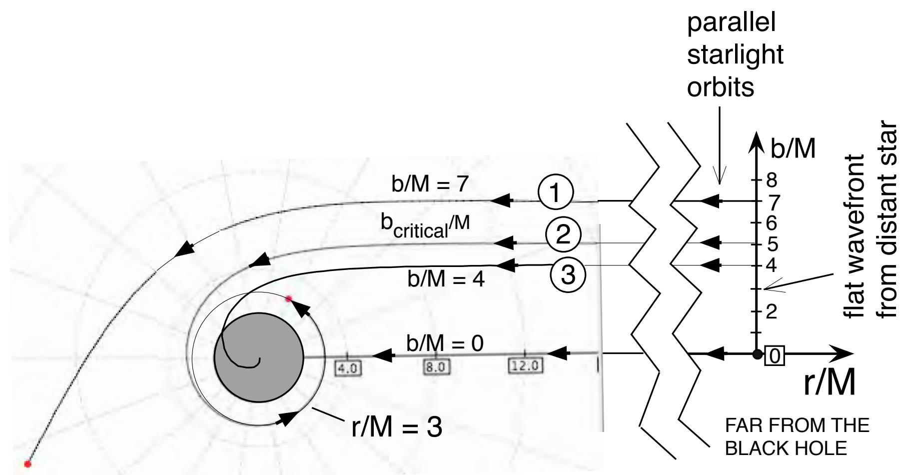


FIGURE 2 Jagged lines separate flat spacetime far from the black hole (on the right) from curved spacetime near the black hole (on the left). The right side of this plot shows two ways to visualize starlight orbits far from the black hole: first as as a set of straight parallel orbits, second as a flat wavefront. On the left side of this plot, near the black hole, only the starlight orbit with $b/M = 0$ remains straight, while starlight orbits 1 through 3, originally parallel, diverge: Starlight Orbit 1 with the impact parameter $b/M = 7$ deflects but escapes. Starlight Orbit 2 with the so-called *critical impact parameter* b_{critical}/M, equation (28), becomes an unstable circular orbit at $r/M = 3$. Starlight Orbit 3 with $b/M = 4$ crosses the event horizon and ends at the singularity.

We need general equations of motion of light, which we now derive using the limiting process of equation (1). Start with equations of motion of a stone from Section 8.3, written in slightly altered form:

$$\frac{dr}{d\tau} = \pm\left[\left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{m^2r^2}\right)\right]^{1/2} \qquad \text{(stone)} \qquad (6)$$

$$\frac{d\phi}{d\tau} = \frac{L}{mr^2} \qquad \text{(stone)} \qquad (7)$$

$$\frac{d\tau}{dT} = \frac{\left(1 - \dfrac{2M}{r}\right)}{\dfrac{E}{m} \pm \left(\dfrac{2M}{r}\right)^{1/2}\left[\left(\dfrac{E}{m}\right)^2 - \left(1 - \dfrac{2M}{r}\right)\left(1 + \dfrac{L^2}{m^2r^2}\right)\right]^{1/2}} \qquad (8)$$

Comment 1. Choice of signs for the motion of a stone
We choose the stone's wristwatch time to advance as the stone moves along its worldline. Therefore the upper $(+)$ sign in (6) is for a stone with increasing r and the lower $(-)$ sign is for a stone with decreasing r. The $\pm$ sign in the denominator of equation (8) has the same meaning.

In order to describe the motion of light, we need to eliminate $d\tau$ from these equations, because adjacent events along the worldline of a light flash have zero wristwatch time lapse between them: $d\tau = 0$. Multiply both sides of (6) by the corresponding sides of (8), then factor out and cancel (E/m) from the resulting numerator and denominator.

$$\begin{aligned}\frac{dr}{dT} &= \frac{dr}{d\tau}\frac{d\tau}{dT} \qquad \text{(stone)} \qquad (9)\\ &= \pm\frac{\left(1-\dfrac{2M}{r}\right)\left[1-\left(\dfrac{m}{E}\right)^2\left(1-\dfrac{2M}{r}\right)\left(1+\dfrac{L^2}{m^2r^2}\right)\right]^{1/2}}{1\pm\left(\dfrac{2M}{r}\right)^{1/2}\left[1-\left(\dfrac{m}{E}\right)^2\left(1-\dfrac{2M}{r}\right)\left(1+\dfrac{L^2}{m^2r^2}\right)\right]^{1/2}}\end{aligned}$$

Equation (1) requires that for light $m \to 0$ while E remains constant. Apply these requirements to (9). The result is our first equation of motion for light:

$$\frac{dr}{dT} = \pm\frac{\left(1-\dfrac{2M}{r}\right)\left[1-\left(1-\dfrac{2M}{r}\right)\left(\dfrac{L}{rE}\right)^2\right]^{1/2}}{1\pm\left(\dfrac{2M}{r}\right)^{1/2}\left[1-\left(1-\dfrac{2M}{r}\right)\left(\dfrac{L}{rE}\right)^2\right]^{1/2}} \qquad \text{(light)} \quad (10)$$

Carry out a similar procedure on equations (7) and (8): multiply their corresponding sides $d\phi/dT = (d\phi/d\tau)(d\tau/dT)$, factor out E/m in the denominator, cancel m with one in the numerator, then let $m \to 0$. The result is our second equation of motion for light:

$$\frac{d\phi}{dT} = \frac{\dfrac{L}{r^2E}\left(1-\dfrac{2M}{r}\right)}{1\pm\left(\dfrac{2M}{r}\right)^{1/2}\left[1-\left(1-\dfrac{2M}{r}\right)\left(\dfrac{L}{rE}\right)^2\right]^{1/2}} \qquad \text{(light)} \quad (11)$$

To construct our third equation of motion for light, combine (10) with (11):

$$\frac{dr}{d\phi} = \left(\frac{dr}{dT}\right)\left(\frac{dT}{d\phi}\right) = \pm\frac{r^2E}{L}\left[1-\left(1-\frac{2M}{r}\right)\left(\frac{L}{rE}\right)^2\right]^{1/2} \qquad \text{(light)} \quad (12)$$

Equations (10) through (12) are the equations of motion for light. The choice of signs in these equations is the same as for a stone, given in Comment 1.

Our three equations of motion for light contain a wonderful surprise: The only quantity we need to describe the orbit of light is the ratio L/E. *Meaning:* The orbit of light near a black hole is completely determined by the single value of the ratio L/E instead of by the separate values of the map constants of motion L and E. And equation (5) tells us that this ratio equals the impact parameter for light.

Light motion depends on only $L/E = b$.

Substitute the expression $b = E/L$ into equations (10) through (12):

$$\frac{dr}{dT} = \pm\frac{\left(1-\dfrac{2M}{r}\right)\left[1-\left(1-\dfrac{2M}{r}\right)\left(\dfrac{b}{r}\right)^2\right]^{1/2}}{1\pm\left(\dfrac{2M}{r}\right)^{1/2}\left[1-\left(1-\dfrac{2M}{r}\right)\left(\dfrac{b}{r}\right)^2\right]^{1/2}} \qquad \text{(light)} \qquad (13)$$

$$\frac{d\phi}{dT} = \frac{\dfrac{b}{r^2}\left(1-\dfrac{2M}{r}\right)}{1\pm\left(\dfrac{2M}{r}\right)^{1/2}\left[1-\left(1-\dfrac{2M}{r}\right)\left(\dfrac{b}{r}\right)^2\right]^{1/2}} \qquad \text{(light)} \qquad (14)$$

$$\frac{dr}{d\phi} = \pm\frac{r^2}{b}\left[1-\left(1-\frac{2M}{r}\right)\left(\frac{b}{r}\right)^2\right]^{1/2} \qquad \text{(light)} \qquad (15)$$

An identical square-bracket expression appears multiple times in these equations. To simplify them, define a new function $F(b,r)$:

$$F(b,r) \equiv \left[1-\frac{b^2}{r^2}\left(1-\frac{2M}{r}\right)\right]^{1/2} \qquad \text{(light)} \qquad (16)$$

Equations of motion for light

so that equations of motion for light become:

$$\frac{dr}{dT} = \pm\frac{\left(1-\dfrac{2M}{r}\right)F(b,r)}{1\pm\left(\dfrac{2M}{r}\right)^{1/2}F(b,r)} \qquad \text{(light)} \qquad (17)$$

$$\frac{d\phi}{dT} = \frac{\dfrac{b}{r^2}\left(1-\dfrac{2M}{r}\right)}{1\pm\left(\dfrac{2M}{r}\right)^{1/2}F(b,r)} \qquad \text{(light)} \qquad (18)$$

$$\frac{dr}{d\phi} = \pm\frac{r^2}{b}F(b,r) \qquad \text{(light)} \qquad (19)$$

The $\pm$ signs in equations (17) through (19) have the same interpretation as in (6) through (8) and also (10) through (12), namely the upper (+) sign describes light with increasing r and the lower ($-$) describes light with decreasing r.

Chapters 9 and 10 use interactive software GRorbits to plot orbits of a stone. GRorbits also integrates equations (17) through (19) for light. Given the value of b and initial location, the software plots the orbit and outputs a spreadsheet with global coordinates (T, r, ϕ) of events along the orbit.

Equations of motion for light look complicated. We now derive a simple way to visualize the global r-motion of light using the effective potential, modeled after the effective potential for a stone in Section 8.4.

11.4 ■ EFFECTIVE POTENTIAL FOR LIGHT

Describe global motion of light at a glance.

The present section sets up an effective potential for a light orbit in order to visualize its r-component of motion simply and directly. Recall equation (21) in Section 8.4 that relates the r-motion of a stone to its effective potential:

$$\left(\frac{dr}{d\tau}\right)^2 = \left(\frac{E}{m}\right)^2 - \left(\frac{V_{\rm L}(r)}{m}\right)^2 \qquad \text{(stone)} \qquad (20)$$

The key idea of this equation is that the first term on the right is a constant of the stone's motion—independent of location—while the second term is a function of r—independent of the properties or motion of the stone. We defined the second term to be the effective potential for a stone.

To make similar predictions about the r-motion of light, we seek an equation with the same form as (20). To find this equation, square both sides of (17), rearrange the results, and multiply through by $(M/r)^2$ to obtain:

$$\left(\frac{M}{b}\right)^2\left(1-\frac{2M}{r}\right)^{-2}\left[1 \pm \left(\frac{2M}{r}\right)^{1/2} F(b,r)\right]^2\left(\frac{dr}{dT}\right)^2 = \left(\frac{M}{b}\right)^2 F^2(b,r) \qquad (21)$$

On the left side of (21) we define the function

$$A^2(b,r) \equiv \left(\frac{M}{b}\right)^2\left(1-\frac{2M}{r}\right)^{-2}\left[1 \pm \left(\frac{2M}{r}\right)^{1/2} F(b,r)\right]^2 \qquad \text{(light)} \qquad (22)$$

and on the right side of (21) we substitute for $F^2(b,r)$ from (16).

$$\left(\frac{M}{b}\right)^2 F^2(b,r) = \frac{M^2}{b^2} - \frac{M^2}{b^2}\frac{b^2}{r^2}\left(1-\frac{2M}{r}\right) \qquad \text{(light)} \qquad (23)$$

effective potential for light

Substitute the left sides of (22) and (23) into (21) and write the result as:

Box 1. Use of the effective potential for a stone and for a light flash

Compare and contrast the forms and uses of effective potentials for a stone and for a light flash:

$$\left(\frac{dr}{d\tau}\right)^2 = \left(\frac{E}{m}\right)^2 - \left(\frac{V_{\rm L}(r)}{m}\right)^2 \quad \text{(stone)} \quad (26)$$

$$A^2\left(\frac{dr}{dT}\right)^2 = \left(\frac{M}{b}\right)^2 - \left(\frac{V(r)}{m}\right)^2 \quad \text{(light)} \quad (27)$$

For a stone:

- $V_{\rm L}$ depends on both L and r.
- The turning point occurs where $V_{\rm L} = \pm E$.
- $|E| < |V_{\rm L}|$ is forbidden
- When $|E| \geq |V_{\rm L}|$, equation (26) gives $|dr/d\tau|$ in terms of r, L, E.

For a light flash:

- V depends on r alone.
- The turning point occurs where $V = \pm 1/b = \pm E/L$, not E alone.
- $|E| < |V|$ is forbidden
- When $|1/b| \geq |V|$, equation (27) gives $|dr/dT|$ in terms of r, b.

What's the difference between the two cases?

For light, L has been removed from the effective potential and combined with E; only $b = L/E$ remains. Impact parameter b can be taken completely out of the effective potential, so V depends only on r. This makes orbits of light *simpler* than orbits of a stone. Only *one* constant of motion is needed, not two.

$$A^2(b,r)\left(\frac{dr}{dT}\right)^2 = \left(\frac{M}{b}\right)^2 - \left(\frac{V(r)}{M}\right)^2 \qquad \text{(light)} \qquad (24)$$

where (25) defines the square of the **effective potential for light**

$$\left(\frac{V(r)}{M}\right)^2 \equiv \frac{M^2}{r^2}\left(1 - \frac{2M}{r}\right) \qquad \text{(light)} \qquad (25)$$

Figure 3 plots positive values of the effective potential for light. In Query 2 you show that the coefficient $A^2(b,r)$ in equation (22) is well behaved when light descends to the event horizon, provided $b \neq 0$.

Box 1 compares and contrasts effective potentials for light and for stones.

QUERY 2. Approaching the event horizon

What happens to the left side of (24) as $r/M \to 2^+$, that is as light approaches the event horizon from above? Just above the event horizon set $r/M = 2(1+\epsilon)$ where $0 < \epsilon \ll 1$ and use our standard approximation (inside the front cover) to show that coefficient $A^2(b,r)$ in (24) is well behaved even as light descends to the event horizon, provided $b \neq 0$.

Quick predictions with the effective potential

With the effective potential we can predict—at a glance—the r-component of light motion. The first term, $(M/b)^2$, on the right side of (24) is a constant of motion, the same everywhere along the orbit. The second term is a function of r and does not include b. Figure 3 and its caption also contain a preview of *turning points*, which we analyze more fully in Section 11.4.

Same effective potential for light of EVERY energy (EVERY wavelength)

Huge payoff: The right side of (24) does not include the energy or angular momentum of light. *One effective potential applies to light orbits of every energy and every angular momentum.* In particular, it applies to electromagnetic radiation of all wavelengths: radio waves; microwaves; infrared, visible, and ultraviolet light; X-rays; and gamma rays! (This result assumes that the wavelength of light is small compared with the coordinate separations over which spacetime curvature changes appreciably.)

QUERY 3. Critical impact parameter

A. Show that the peak of the effective potential occurs at $r/M = 3$.

B. Verify that the so-called **critical value** of the impact parameter at $r/M = 3$ is

$$\frac{b_{\text{critical}}}{M} = (27)^{1/2} = 5.196\ 152\ 42 \qquad \text{(light, critical impact parameter)} \qquad (28)$$

C. From Figure 3, read off approximate values of b/M and r/M for the circular orbit. Compare these values with the analytic results of Items A and B.

Effective potentials reveals turning points.

Both the effective potential for light and effective potentials for stones enable us to find the r-coordinate at which the r-component of motion goes to zero, which occurs for a circular orbit and also at what we call a *turning point* (Section 8.4 and Section 11.5).

DEFINITION 5. Plunge Orbit, Bounce Orbit, Trapped Orbit

Figure 3 sorts all light orbits near a black hole into three categories, which we give names to simplify our analysis:

- **Plunge Orbit:** A plunge orbit is an incoming or outgoing orbit with $|b| < b_{\text{critical}}$ that passes above the peak of the effective potential curve in Figure 3. A starlight Plunge Orbit is—by definition—an incoming orbit that *plunges* through the event horizon to the singularity. Outside the event horizon light can, in principle, move in either direction along the plunge orbit shown. We call this a plunge orbit, whether r decreases or increases.
- **Bounce Orbit:** A bounce orbit is an incoming or outgoing orbit with $|b| > b_{\text{critical}}$. The bounce orbit exists only to the right of the effective potential in Figure 3 and below its peak. A starlight Bounce Orbit is—by definition—an orbit that initially moves inward, then reverses its r-component of motion—its r-coordinate *bounces*—at a turning point on the outer edge of the effective potential, while its ϕ-component of motion continues. After the bounce, the light moves outward on the same horizontal line in the figure, and escapes to infinity. A Bounce Orbit cannot reach the singularity.

Definitions:
Plunge Orbit
Bounce Obit
Trapped Orbit

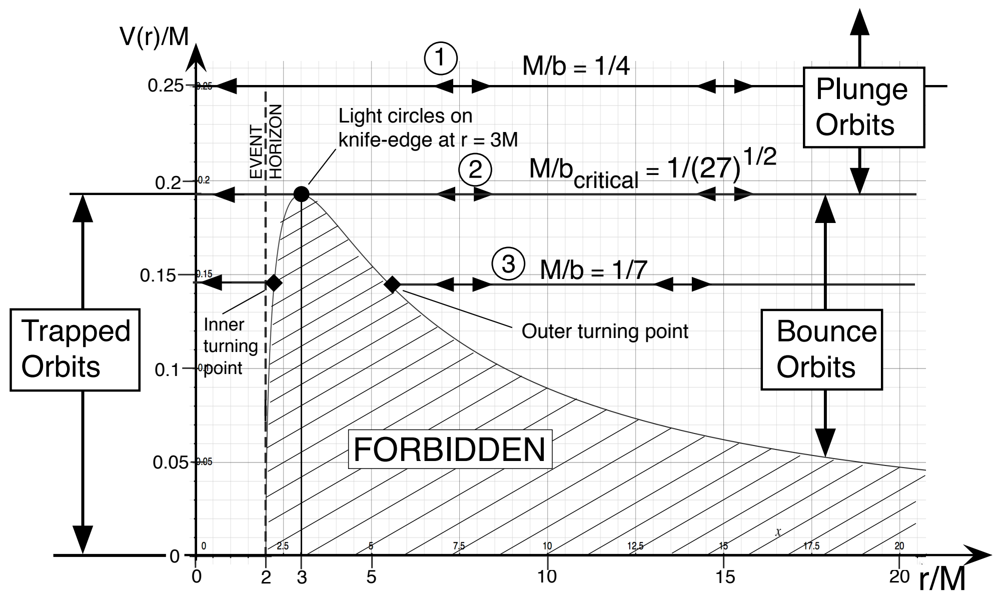


FIGURE 3 Examples of the three categories into which we sort all orbits (Definition 5). *Horizontal Line (1):* a Plunge Orbit with $M/b = 1/4$ that enters the black hole. *Horizontal Line (2):* the orbit with $M/b_{\rm critical} = 1/(27)^{1/2}$ that reaches the peak of the effective potential—marked with a little filled circle—and enters an unstable circular orbit there. *Horizontal Line (3):* a Bounce Orbit with $M/b = 1/7$ approaches the black hole, reverses its r-motion at the *outer turning point* (Section 11.6), and moves away from the black hole. The Trapped Orbit with $M/b = 1/7$ originates in the narrow horizontal region between the event horizon and the effective potential curve and moves inward through the event horizon.

- **Trapped Orbit:** A trapped orbit is an orbit with $|b| > b_{\rm critical}$ to the left of the effective potential in Figure 3 and below its peak. No starlight orbit can be a Trapped Orbit. An initially outgoing Trapped Orbit outside the event horizon reverses its r-component of motion at the inner turning point on the inner edge of the effective potential. *Every* Trapped Orbit reaches the singularity unless intercepted.

The horizontal line for $M/b_{\rm critical}$ in Figure 3 is the dividing line between these different categories of orbits. Figure 4 shows Plunge and Bounce Orbits; Figure 5 shows two Trapped Orbits.

11.5 ■ TURNING POINTS

The r-motion of light can reverse at a turning point.

At a turning point the r-component of motion goes to zero, while the ϕ-component of motion continues. Little filled squares in Figures 3 through 5 mark what we call **outer and inner turning points**.

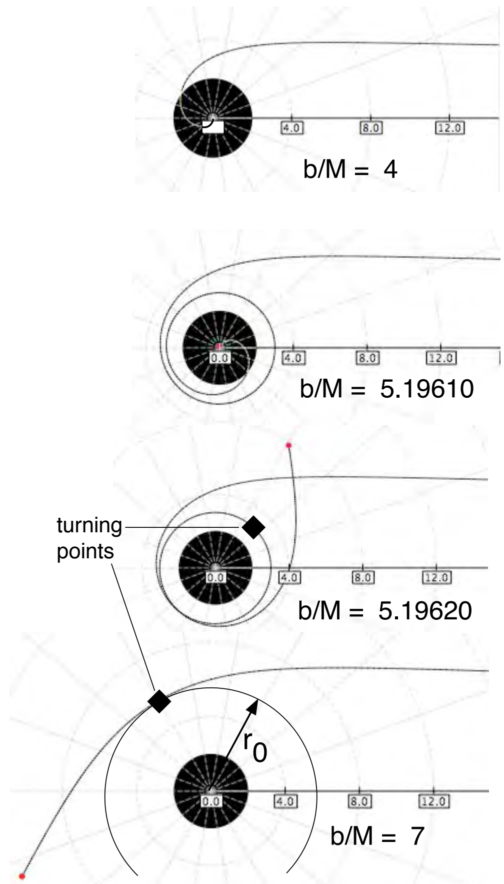


FIGURE 4 Top two panels: Plunge Orbits. Bottom two panels: Bounce Orbits, each with a little filled square at the turning point (Section 11.4). Middle two panels: b-values straddle $b_{\rm critical}/M = 5.19615...$, for which the orbit enters a knife-edge circular orbit.

DEFINITION 6. Turning Point

A turning point is the r-value at which the right side of equation (24) equals zero, where M/b equals the value of the effective potential.

Definitions:
Turning point
Outer turning point
Inner turning point
Circular orbit poin

- An **outer turning point** is to the right and below the peak of the effective potential (see Figure 3).
- An **inner turning point** is to the left and below this peak.The peak itself is the location of the unstable (knife-edge) circular orbit of light.
- A **circular orbit point** is the r-value at which the effective potential is maximum. This is the r-location of an unstable (knife-edge) circular orbit for light.

Turning point
subscript: tp

We use the subscript **tp** to label the r-coordinate of a turning point. *Example:* In Figure 3, Orbit 3 with $|b/M| = 7$ reverses its r-motion at

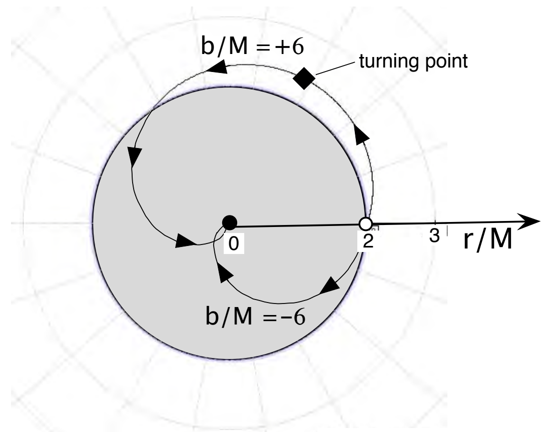


FIGURE 5 Two Trapped Orbits that originate from the same point just outside the event horizon at $r/M = 2^+$ (little open circle). One orbit has $b/M = +6$ with an inner turning point (little filled square); the other has $b/M = -6$ and no turning point. Both orbits reach the singularity at $r/M = 0$. Figure 6 adds labels to this plot.

$r_{\text{tp}} = 5.617M$. Any outgoing light with $|b/M| = 7$ that arrives at the inner turning point at $r_{\text{tp, inner}} = 2.225M$ thereafter moves with $dr < 0$ and enters the black hole.

Equations (24) and (25) tell us that the turning point r_{tp}, the r-coordinate at which $dr/dT = 0$ and motion is purely tangential, occurs for the value of b given by:

$$b/M = \pm \frac{r_{\text{tp}}/M}{\left(1 - \dfrac{2M}{r_{\text{tp}}}\right)^{1/2}} \qquad \text{(given } r_{\text{tp}},\ \text{find } b) \qquad (29)$$

Turning points only for $b^2 > b^2_{\text{critical}}$

Comment 2. No turning point inside the event horizon
Equation (29) guarantees that there can be no turning point for light inside the event horizon, because b/M on the left side is necessarily a real quantity, while the right side of (29) is imaginary for $r_{\text{tp}} < 2M$.

Derive r_{tp} from b.

Equation (29) gives us the value of b when we know the r-coordinate r_{tp} of the turning point. More often, we know the value of b and want to find the r-coordinate of the turning point. In that case, convert (29) into a cubic equation in r_{tp}:

$$r_{\text{tp}}^3 - b^2 r_{\text{tp}} + 2Mb^2 = 0 \qquad \text{(given } b,\ \text{find } r_{\text{tp}}) \qquad (30)$$

QUERY 4. *Optional:* Some consequences of turning points.

A. From equations (24) and (25) show that a light orbit with a given value of b cannot exist in a range of r-coordinates determined by the following inequality:

$$r^3 - b^2 r + 2Mb^2 < 0 \qquad \text{(region with no light orbits)} \qquad (31)$$

B. Show that inequality (31) describes the shaded region under the effective potential curve in Figure 3. In other words, light cannot penetrate the effective potential curve.

Find the turning points

Equation (30) is cubic—includes a third power of $r_{\rm tp}$. Cubic equations can be difficult to solve. Here are analytic solutions of (30). The first two yield r values of the outer and inner turning points, respectively, such as those in Figure 3. In Query 4 you show that the third solution is real but negative, so cannot represent the always-positive map r-coordinate:

$$r_{\rm tp} = 3M\left[\frac{1}{2} - \cos\left(\psi - 120^\circ\right)\right]^{-1} \tag{32}$$

(Outer turning points lie at $r > 3M$.)

$$r_{\rm tp,\,inner} = 3M\left[\frac{1}{2} - \cos\left(\psi + 120^\circ\right)\right]^{-1} \tag{33}$$

(Inner turning points lie between $r/M = 2$ and $r/M = 3$.)

$$r_{\rm NO} = 3M\left[\frac{1}{2} - \cos\psi\right]^{-1} \tag{34}$$

(Yields negative r: not physical.)

For all three solutions, ψ depends on b as follows:

$$\psi \equiv \frac{1}{3}\arccos\left(\frac{54M^2}{b^2} - 1\right) \qquad \left(|b| \geq b_{\rm critical},\ 0 \leq \psi \leq \pi\right) \tag{35}$$

We take what is called the *principle value* of the $\arccos z$, that is the angle between 0 and π radians whose cosine is z. Recall that the magnitude of the cosine is never greater than one. Therefore turning points exist only when the arccos function (35) exists, that is when $b^2 \geq b^2_{\rm critical}$ or when the horizontal line for $(M/b)^2$ in Figure 3 is at or below the peak of the effective potential. This makes graphical, as well as analytic, sense.

QUERY 5. Unphysical third solution

Show that the third solution (34) yields a negative value for r, which cannot represent the non-negative r-coordinate.

QUERY 6. Examples of turning points

A. For the outer and inner turning points of the orbit with $|b/M| = 7$, derive the numerical values $r_{\rm tp} = 5.617M$ and $r_{\rm tp,\,inner} = 2.225M$. Use Figure 3 to verify these r-coordinates approximately.

B. Show that $F(b, r) = 0$ at the turning points.

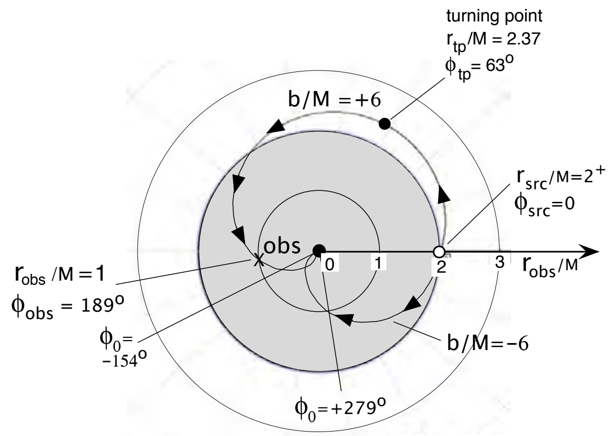


FIGURE 6 Elaboration of Figure 5. Two Trapped Orbits originate from just outside the event horizon at $r_{\rm src}/M = 2^+$, $\phi_{\rm src} = 0$. The counterclockwise orbit, with $b/M = +6$, rises to a turning point at $(r_{\rm tp}/M = 2.37, \phi_{\rm tp} = 63°)$, then falls back through the event horizon to arrive at the singularity at map angle $\phi_0 = +279°$. The clockwise orbit with $b/M = -6$ crosses the horizon immediately and reaches the singularity at the map angle $\phi_0 = -154°$. The event X locates a falling observer that intercepts the counterclockwise light orbit at $(r_{\rm obs}/M = 1, \phi_{\rm obs} = 189°)$.

C. An orbit with impact parameter $|b/M| \approx b_{\rm critical}/M = (27)^{1/2}$ circles at $r \approx 3M$ for a while. Then it "falls off the knife-edge," either spiraling inward or returning outward to $r/M \gg 1$. In the second case the turning r-coordinate is $r_{\rm tp}/M \approx 3$, but *where on that circle* is the turning point?

QUERY 7. Infinite impact parameter

A. From equation (29), find two different conditions that lead to $|b/M| \to \infty$.

B. In Figure 3, what horizontal line corresponds to $(M/b)^2 \to 0$ or $|b/M| \to \infty$? Point out two places on the graph (one a limiting case) where $(V(r)/M)^2$ reaches this line.

11.6 ■ STARLIGHT ORBIT: FROM STAR TO OBSERVER

Starlight orbit must reach me.

Which orbit(s) connect(s) the star with the observer?

Which light orbit(s) connect(s) a particular star to a given map location near the black hole? This question is important because sooner or later we want to predict in what direction one of the many possible inertial observers at that map location looks to see a particular star. But an observer cannot see light that does not reach him or her. The central goal of this chapter is to find the global path of an orbit that connects distant Star X to a given map location Y, whatever the motion may be of an observer at rest or moving through that location.

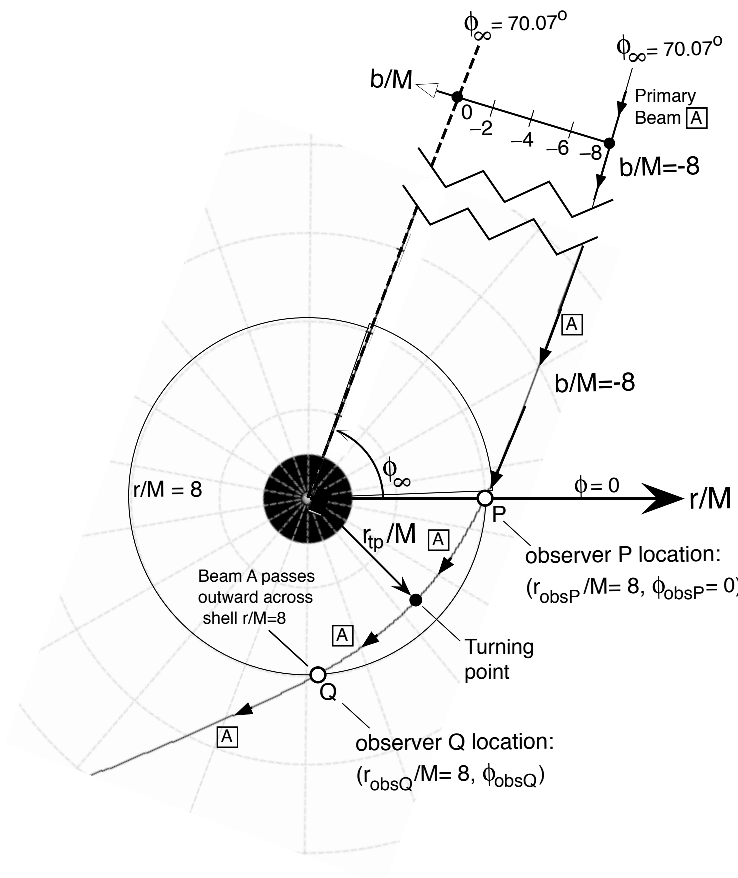


FIGURE 7 Starlight orbit A with impact parameter $b/M = -8$ moves in a clockwise direction to connect the star at map angle $\phi_\infty = 70.07^\circ$ to observer P located at $(r_{\rm obsP}/M = 8, \phi_{\rm obsP} = 0)$. The starlight orbit proceeds to observer Q, crossing outward through the shell at the same $r_{\rm obsQ}/M = r_{\rm obsP}/M = 8$ but at a different value $\phi_{\rm obsP}$, to be determined.

?

Objection 2. *Ha, gotcha! You say that the observer can be at any coordinate $r_{\rm obs}$. But inside the event horizon nothing can stand still in global coordinates. Therefore you cannot have an observer at $r_{\rm obs} < 2M$.*

!

You are correct: No observer can remain constant r inside the event horizon. However Chapters 6, 7, and 12 describe the rain observer who starts from rest far from the black hole and drops to its center. This rain observer receives starlight even inside the event horizon. To predict the spectacular, ever-changing rain observer's pre-doom panoramas (Chapter 12), we must know which orbit(s) from every star reach(es) her there.

The orbit labeled A in Figure 7 connects a distant star to a point with map location $(r_{\rm obs}/M = 8, \phi_{\rm obs} = 0)$ where we will later place one of many possible

Map angle ϕ_∞ to a star

observers. This figure introduces the **map angle** ϕ_∞ of the distant star. The subscript infinity, ∞, reminds us that the star lies far from the black hole.

$$\phi_\infty \equiv \text{(map angle to a distant star, this angle measured counterclockwise from the direction } \phi = 0) \qquad (36)$$

Primary orbit

Section 11.7 shows that many orbits—in principle an infinite number of orbits—from each star arrive at the map location of any observer. How do we choose which orbit to follow? *Answer:* We discover that there is a single most-direct orbit between star and observer, an orbit whose spatial path is the least deflected in map coordinates. We call this the **primary orbit** and give it most of our attention, often simply calling it "the orbit."

Primary orbit between star and map location of the observer

What primary orbit connects the star at given map angle ϕ_∞ most directly with the observer at map location $(r_{\text{obs}}, \phi_{\text{obs}} = 0)$? This is an important question with a complicated answer. So start with an example.

Figure 7 shows the interactive software GRorbits plot of a primary Bounce Orbit between a star at map angle $\phi_\infty = 70.07°$ and an observer at map location $(r_{\text{obs}} = 8M, \phi_{\text{obs}} = 0)$. *Result:* The orbit with impact parameter $b/M = -8$ connects this observer with the star at map angle $\phi_\infty = 70.07°$.

Incoming orbit may move out again across the same shell.

The incoming orbit in Figure 7 sweeps clockwise past the observer at $r/M = 8$, reaches a turning point at smaller r-coordinate, then crosses the $r/M = 8$ shell a second time, now in an outgoing direction. Two observers located at *different* points along the same shell can see the same orbit from the *same* star.

11.7 ■ INTEGRATE THE STARLIGHT ORBIT

An exact and immediate result

Goal: To plot $\phi_\infty - \phi_{\text{obs}}$ for starlight

Our goal is to plot $\phi_\infty - \phi_{\text{obs}}$ for starlight as a function of r_{obs} for a given value of the impact parameter b. To accomplish this, integrate $d\phi/dr$ directly. Figure 7 shows two cases. Case I: The orbit reaches the observer before the turning point. Case II: The obit reaches the observer after the turning point. Both cases integrate equation (19).

$$\phi_\infty - \phi_{\text{obs}} = \int_{r=\infty}^{\text{r}_{\text{obs}}} \frac{b}{r^2} F^{-1}(b,r)dr \qquad (37)$$

(Case I: observer before turning point)

$$\phi_\infty - \phi_{\text{obs}} = \int_{r=\infty}^{r_{\text{tp}}} \frac{b}{r^2} F^{-1}(b,r)dr + \int_{r_{\text{tp}}}^{\text{r}_{\text{obs}}} \frac{b}{r^2} F^{-1}(b,r)dr \qquad (38)$$

(Case II: observer after turning point)

Figure 8 displays the result of these integrals. The vertical axis "unrolls" the ϕ-angle.

Objection 3. *How do you carry out these integrals? Function $F(b, r)$ in (16) is complicated; these integrations must be difficult.*

Modern numerical methods evaluate these integrals to high accuracy. We do not pause here to describe these methods.

Plunge Orbit has small $|b|$. Bounce Orbit has large $|b|$.

Figure 3 previewed the summary message of Figure 8: An incoming orbit with small magnitude of $|b|$ plunges through the event horizon to the singularity. An incoming orbit with a large magnitude of $|b|$ deflects and returns outward again. An incoming orbit with the particular intermediate value $\pm b_{\text{critical}}$ circles temporarily at $r = 3M$, then either continues ingoing or becomes outgoing.

Objection 4. *You are not telling us the whole story! Orbits in most figures of this chapter have arrows on them. Every arrow tells us the direction of motion of light at that place along the orbit. But motion involves increments in the T-coordinate. Your equations that lead to these figures do not contain global T. Therefore these equations can give us only the curves themselves, without arrows.*

Yes and no. Equation (5) defines b as L/E, so the sign of the impact parameter is the same as the sign of L. This means that the motion of light is counterclockwise for positive values of b and clockwise for negative values. So equations (38) and (39) do give us the directions of motion (arrow directions) simply from the signs of b/M in those equations. Indeed, these equations do not tell us the map position of each light flash as a function of the T-coordinate. But we are interested in the plot of a steady starlight orbit, which does not vary with T.

Sample Problems 2 illustrate uses of Figure 8.

Comment 3. Every black hole redirects to every observer multiple orbits from every star.

You can use Figure 8 to find the value b of an orbit that connects *any* distant star $(-180° < \phi_\infty \leq +180°)$ to a map location on *some* circle of *any* r-coordinate around the black hole. Whoa! Does this mean that the black hole never obscures any star in the heavens for an observer near it? Yes, and more: The following section and Figure 10 show that every black hole in the visible Universe redirects multiple orbits from every single star in the heavens to an observer at every single map location.

11.8 ■ MULTIPLE STARLIGHT ORBITS FROM EVERY STAR

An infinite number of orbits that appear fainter and fainter to an observer.

One star: Infinite images?

It is remarkable that every map location near a black hole receives multiple orbits—in principle an infinite number of orbits—from a single star, and thus

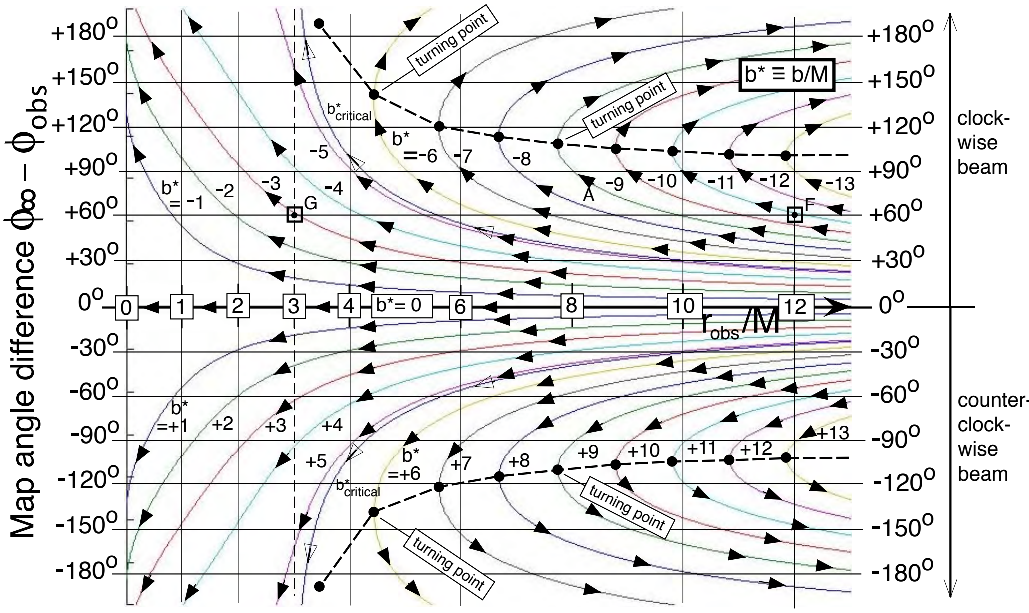


FIGURE 8 Difference in map angles between a distant star and the observer at map location $(r_{obs}/M, \phi_{obs})$ derived for an orbit of impact parameter b/M from that star. To reduce clutter, we define $b^* \equiv b/M$. Arrows on the curves tell whether the starlight is incoming or outgoing; at a turning point the orbit changes from incoming to outgoing.

from *every* star in the heavens. Figure 9 replots the primary orbit of Figure 7 and adds two additional orbits, called **higher-order orbits** from the same star. By trial and error, the interactive software program GRorbits finds values $b/M = +5.4600$ and $b/M = -5.2180$ for these additional orbits from the same star.

Higher-order orbits

In Figure 9, the higher-order orbit with $b/M = +5.4600$ moves around the black hole counterclockwise and approaches the map location $(r/M = 8,\ \phi = 0)$ from below. This orbit lacks 70.07° of making a complete circuit around the black hole. Therefore the *total* angle to the same star is $\phi_\infty = -(360° - 70.07°) = -289.93°$.

The next higher-order orbit with $b/M = -5.2180$ moves around the black hole clockwise and approaches the map location $(r/M = 8,\ \phi = 0)$ from above. This orbit makes a complete circuit around the black hole, *plus* 70.07°, for a total of 430.07°. Therefore the *total* angle to the same star is $\phi_\infty = +(360° + 70.07°) = +430.07°$.

Each observer receives many orbits from every star.

Figure 10 extends the vertical scale of Figure 8 to show orbits with b-values close to the critical value that circle several times around the black hole before they either escape outward or plunge on inward. The upward and downward vertical scales in Figure 10 extend indefinitely, leading to more and more orbits with b-values on either side of $b_{critical}/M = (27)^{1/2} = 5.196152....$ *Conclusion:* An observer at each r-coordinate r_{obs} receives multiple orbits—in principle an infinite number of orbits—from every star in the heavens.

Sample Problems 1. Orbits that reach $r/M = 3$

Think of orbits with different b-values that reach the observer map location at $(r_{\rm obs}/M = 3, \phi_{\rm obs} = 0)$. Use Figure 8 to provide approximate answer the following questions.

A. What is the b-value of the orbit that comes from the star at map angle $\phi_\infty = +60°$? **Solution A:** Look at the vertical dashed line at $r_{\rm obs}/M = 3$. This line intersects with the horizontal line $\phi_\infty = +60°$ very close to the curve $b/M = -3$, at the point marked G. So this is the b-value of the Plunge Orbit that connects the star at map angle $\phi_\infty = +60°$ with the observer at $(r_{\rm obs}/M = 3, \phi_{\rm obs} = 0)$.

B. What is the b-value of the orbit that comes from the star at map angle $\phi_\infty = +90°$? **Solution B:** The vertical dashed line at $r_{\rm obs}/M = 3$ intersects the horizontal line $\phi_\infty = +90°$ very close to the Plunge Orbit $b/M = -4$.

C. What is the b-value of the orbit that comes from the star at map angle $\phi_\infty = +30°$? **Solution C:** The vertical dashed line $r_{\rm obs}/M = 3$ intersects with the horizontal line $\phi_\infty = +30°$ about six-tenths of the separation between the curves $b/M = -1$ and $b/M = -2$. Therefore the Plunge Orbit with $b \approx -1.6$ connects the star at map angle $\phi_\infty = +30°$ with the map location $(r_{\rm obs}/M = 3, \phi_{\rm obs} = 0)$.

D. What is the b-value of the orbit that comes from the star at negative map angle $\phi_\infty = -90°$? **Solution D:** The vertical dashed line $r_{\rm obs}/M = 3$ intersects the horizontal line $\phi_\infty = -90°$ very close to the curve $b/M = +4$. The positive b-value means that the orbit moves counterclockwise around the black hole.

E. an orbit comes from the opposite side of the black hole, at $\phi_\infty = 180°$. What is the b-value of this orbit? **Solution E:** Both $\phi_\infty = +180°$ and $\phi_\infty = -180°$ are map angles to a star on the other side of the black hole. The vertical dashed line $r_{\rm obs}/M = 3$ intersects the horizontal lines $\phi_\infty = \pm 180°$ approximately half way between $b/M = \pm 5$ and $b/M = \pm(27)^{1/2} = \pm 5.196$. Therefore the b-values of these two Plunge Orbits are approximately $b \approx \pm 5.1$. *Optional:* Sketch this orbit.

Sample Problems 2. Orbits from a single star that reach observers at different r-coordinates

Orbits with different b-values from the star at map angle $\phi_\infty = +60°$ reach observers at different r-coordinates along the line $\phi = 0$. What are these b-values at r-coordinates $r_{\rm obs}/M = 12, 8, 4, 2, \text{and } 1$? In each case say whether the orbit is a Plunge Orbit, a Bounce Orbit, or a Trapped Orbit.

Solution: All of the orbits are from a star; therefore none of them can be a Trapped Orbit. In Figure 8, look at the intersections of horizontal line $\phi_\infty = +60°$ with vertical lines at these different r-coordinates. We estimate the b-values to one decimal place.

- At $r_{\rm obs}/M = 12$, $b/M \approx -10.9$, the point marked F in the figure; a Bounce Orbit
- At $r_{\rm obs}/M = 8$, $b/M \approx -7.3$, a Bounce Orbit
- At $r_{\rm obs}/M = 4$, $b/M \approx -3.8$, a Plunge Orbit
- At $r_{\rm obs}/M = 2$, $b/M \approx -2.0$, a Plunge Orbit
- At $r_{\rm obs}/M = 1$, $b/M \approx -1.2$, a Plunge Orbit

Look at the little square white boxes on the vertical line at $r/M = 8$ in Figure 10. Three of the little white boxes on the vertical line at $r/M = 8$ correspond to the three starlight orbits displayed in Figure 9. Other little boxes represent more of the multiple higher-order orbits between this star and this observer. Each little box is offset vertically by $\pm 360°$ from its nearest neighbor.

QUERY 8. *Optional:* Classify primary and higher-order orbits from a star.

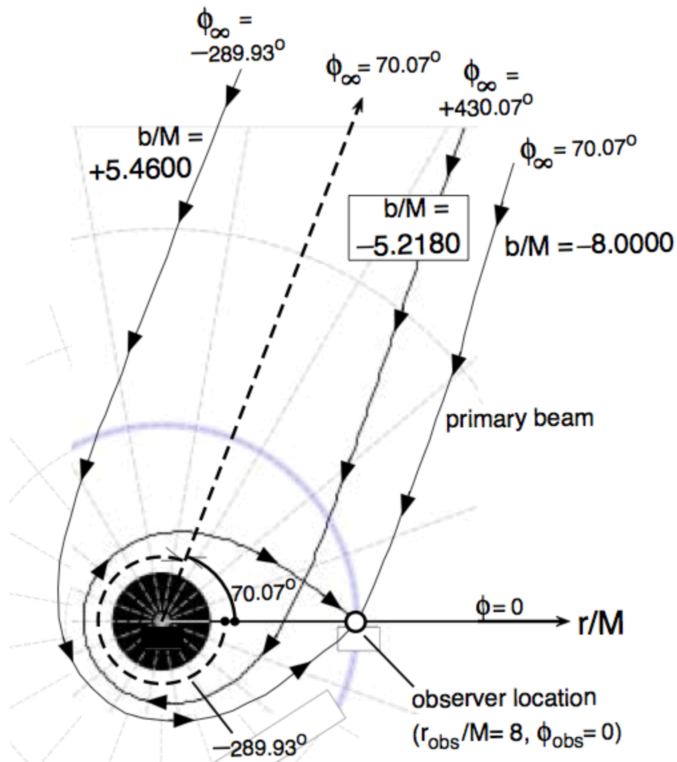


FIGURE 9 Three of the infinite number of orbits of light that, in principle, arrive at the same observer from a single star. For the primary orbit with $b/M = -8$, the star angle is $\phi_\infty = 70.07^\circ$ (as in Figure 7). For the second orbit, with $b/M = +5.4600$, the star angle (dashed arc) is $\phi_\infty = -(360^\circ - 70.07^\circ) = -289.93^\circ$. For the third orbit, with $b/M = -5.2180$, the star angle (angle-arc not shown) is $\phi_\infty = (360^\circ + 70.07^\circ) = +430.07^\circ$. All three orbits come from the same star, but the observer sees three different images in three different directions.

Classify the primary and higher-order starlight orbit as a Plunge Orbit or a Bounce Orbit. Figure 10 may be useful. *Reminder:* This analysis says nothing about the state of motion of the observer at that map location: he may be at rest there; she may dive or orbit past that map location.

A. Show that for every observer inside $r/M = 3$, *all* starlight orbits are Plunge Orbits.
B. Show that for every observer outside $r/M = 3$, starlight orbits are either Plunge Orbits or Bounce Orbits.
C. At any $r/M > 3$, what is the value of b/M that divides Plunge Orbits from Bounce Orbits?
D. Find an equation for the *maximum* magnitude of the impact parameter b/M of a Bounce Orbit that an observer on the shell of a given r-coordinate $r/M > 3$ can see?
E. Show that for every observer at $r/M > 3$, every higher-order orbit is an outgoing Bounce Orbit.
F. Can a primary or higher-order starlight orbit be a Trapped Orbit? Explain your answer.

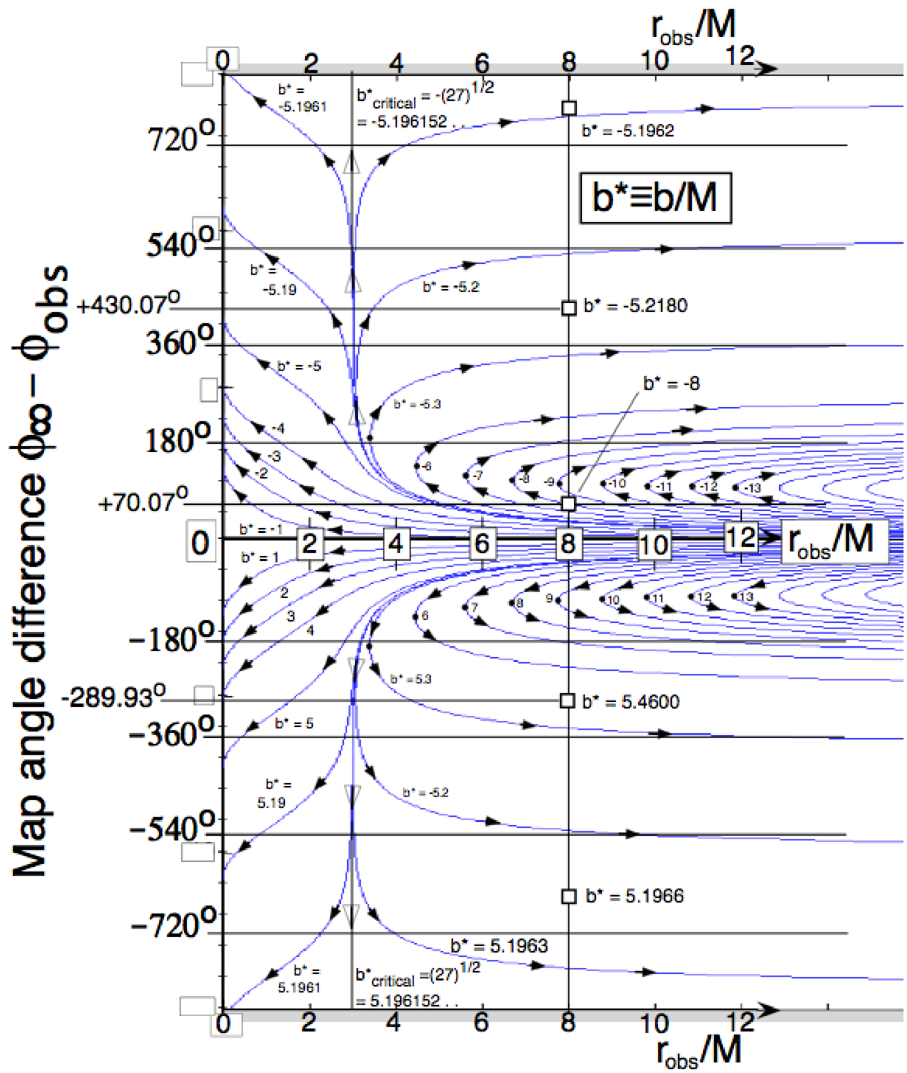


FIGURE 10 Expanded vertical scale for starlight orbits of Figure 8. The observer is at map location $(r_{\rm obs}/M, \phi_{\rm obs})$. *New feature of this plot:* Orbits with $b^* \approx \pm b_{\rm critical}/M$ follow the vertical line at $r/M = 3$ (they circulate at $r/M = 3$) before they either return to $r/M \gg 1$ or plunge into the black hole. *Result:* Multiple orbits—in principle an infinite number of orbits—from every star arrive at each observer, cross every possible vertical line in the figure. *Example:* Three of the little white boxes on the vertical line at $r/M = 8$ correspond to the three starlight orbits displayed in Figure 9.

Higher-order orbits have fainter, smeared images.

Higher-order orbits that go around the black hole more and more times are less and less intense when they arrive at the observer. There is always *some* spread in the orbit, so the more times an orbit circles the black hole, the more it spreads out transverse to its direction of motion and the smaller the

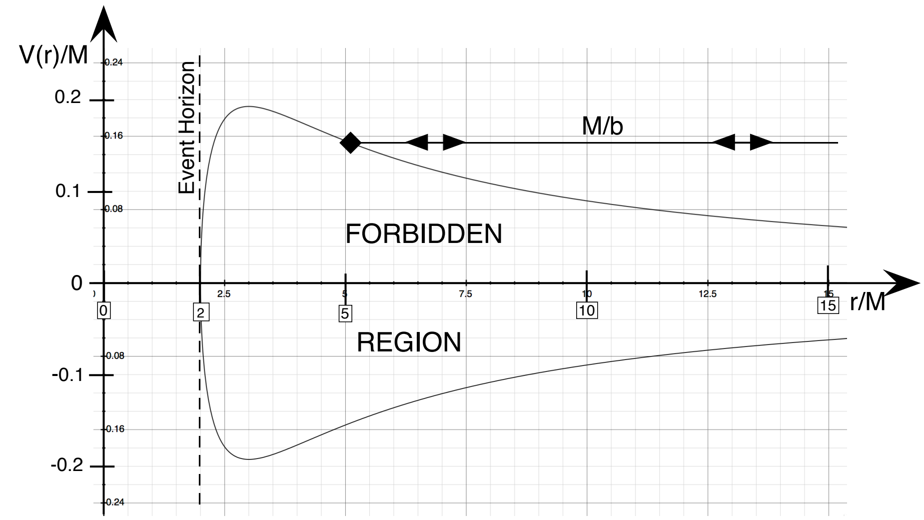


FIGURE 11 Forbidden region for light. Near the non-spinning black hole, this forbidden region separates our world, above the forbidden region, from another world, below the forbidden region.

fraction of photons in the initial orbit that enter the detector at the final map location. Chapter 12 shows that the shell observer also sees higher-order orbits bunched closer and closer together in the observed direction. *Overall result:* Higher and higher order orbits lead to images that get fainter and fainter and smear into one another. As a result, an observer sees separately only a few of the infinite number of orbits that, in principle, arrive from each star.

Strange results follow from equation (24), which expresses $(dr/dT)^2$ in terms of the difference $(M/b)^2 - (V(r)/M)^2$. Differentials dr and dT are both real, so dr/dT must be real. In other words $(dr/dT)^2$ must be positive. *Conclusion:* $(M/b)^2 - (V(r)/M)^2$ must be positive. A consequence of this condition is that either $M/b > +V(r)/M$ or $M/b < -V(r)/M$. The result is a forbidden region where light cannot exist, as shown in Figure 11. Compare corresponding Figure 5 in Section 8.4 for the stone and review the text that accompanies that figure. Near the black hole the forbidden region for light separates our world (above the forbidden region) from another world (below the forbidden region). We can move between these worlds only by entering and then exiting the event horizon—not possible for a non-spinning black hole. However, we will find that for the spinning black hole a trip from the corresponding upper region to the corresponding lower region may be possible. John Archibald Wheeler's radical conservatism says,"Follow the equations wherever they lead, no matter how strange the result."

Two worlds, separated for the non-spinning black hole

11.9 ■ EXERCISES

Note: In the exercises the word *approximately* means that the requested number may be estimated from a figure in this chapter.

1. Thought question: Shadow of a Black Hole?

According to legend, a vampire has no reflection in a mirror and casts no shadow. When illuminated from one side by a distant incoming flat wave, does a black hole cast a shadow on the other side? Think of a possible shadow on a flat plane located far away from the black hole where spacetime is flat.

2. Values of b for orbits that arrive at $r_{\text{obs}}/M = 6$.

Repeat parts A through E of Sample Problems 2 for orbits that reach the observer at map location $(r_{\text{obs}}/M = 6, \phi_{\text{obs}} = 0)$. Classify each orbit as incoming, outgoing, or tangential.

3. Orbits that reach observers at different r-coordinates from the star at map angle $\phi_\infty = -120°$.

Repeat Sample Problems 2 for a star at map angle $\phi_\infty = -120°$.

4. The visual size of a black hole

Figure 10 shows the b-values of beams that escape or are captured by the black hole. The smallest b-value of a beam that can escape is $|b_{\text{critical}}| = (27)^{1/2}M$. Some light from every star circles temporarily on this unstable orbit at $r = 3M$. Because this is a knife-edge orbit, it continually sheds light beams that "fall off" to move either inward or outward.

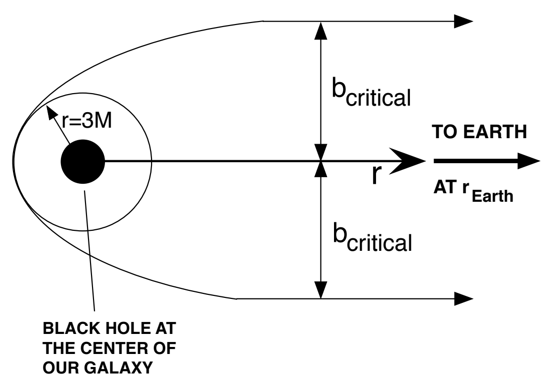


FIGURE 12 Schematic diagram showing the visual size of the black hole Sagittarius A* located at the center of our galaxy, assumed (incorrectly) to be non-spinning. The text shows that all possible parallel straight beams form a three-dimensional cylinder directed toward the observer on Earth.

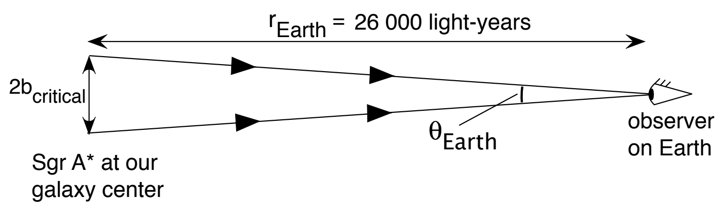


FIGURE 13 Critical beams from Sgr A* form a long cone as seen from Earth

Consider outward light beams that enter the eye of a distant observer on Earth. Figure 12 shows two such beams on one $[r, \phi]$ slice through the center of the black hole. But the same distant observer sees a similar pair of beams that lie on each of an infinite number of similar slices rotated around the r-axis in Figure 12. The resulting set of beams form a cylinder observed by the Earth observer.

To speak more carefully, the beams we see on Earth do not move *exactly* on a cylinder, but rather on a very long cone with its apex at the Earth (Figure 13). As a result, we on Earth see the black hole as a ring. What angle does this ring subtend at our eye on Earth?

Answer this question for the monster black hole called Sagittarius A* (abbreviation: SgrA*) with mass $M_{\text{SgrA}} \approx 4 \times 10^6 M_{\text{Sun}}$ that lies at the center of our galaxy, about 26 000 light-year from Earth. Label this distance r_{Earth}. Assume (incorrectly) that SgrA* is a nonspinning black hole. Derive and justify an expression for the angular size θ_{Earth} of this black hole observed from Earth. (An exercise in Chapter 20 carries out a more realistic analysis that takes account of the spin of this black hole.)

A. From Figure 13, derive the following expression for the very small angle θ_{Earth}.

$$\theta_{\text{Earth}} \approx \frac{2(27)^{1/2} M_{\text{SgrA}}}{r_{\text{Earth}}} \qquad (r \gg M_{\text{SgrA}}) \qquad (39)$$

B. Insert into (39) values for M_{SgrA} and Earth's r-coordinate separation from the black hole of r_{Earth} light years. The following are results to one significant digit. Find each result to two significant digits:

$$\begin{aligned} \theta_{\text{Earth}} &\approx 2 \times 10^{-10} && \text{radian} \qquad (40) \\ &\approx 1 \times 10^{-8} && \text{degree} \\ &\approx 5 \times 10^{-5} && \text{arcsecond} \\ &\approx 50 && \text{microarcseconds} \end{aligned}$$

Comment 4. Microwaves, not visible light
Dust between Earth and the spinning black hole at the center of our galaxy absorbs visible light. Microwaves pass through this dust, so our detectors on Earth are microwave dishes distributed over the surface of Earth.

5. The "incoming map floodlight"

Define an **incoming map floodlight** as a lamp at a given r-coordinate r_{inlamp} that emits all light beams that are ingoing at that r—that is, all beams with a negative r-coordinate differential, $dr < 0$.

A. An incoming map floodlight at $r_{\text{inlamp}}/M = 12$ emits light that might have come from stars with approximately what range of map angles ϕ_∞?

B. An incoming map floodlight at $r_{\text{inlamp}}/M = 6$ emits light that might have come from stars with approximately what range of map angles ϕ_∞?

C. An incoming map floodlight at $r_{\text{inlamp}}/M = 3$ emits light that may have come from stars with approximately what range of map angles ϕ_∞?

D. An incoming map floodlight at $r_{\text{inlamp}}/M = 1$ emits light that may have come from stars with approximately what range of map angles ϕ_∞?

E. Can the incoming map floodlight at $r_{\text{inlamp}}/M = 6$ be at rest in global coordinates? Can the incoming map floodlight at $r_{\text{inlamp}}/M = 1$ be at rest in global coordinates?

6. The "outgoing map floodlight"

Define an **outgoing map floodlight** as a lamp at a given r-coordinate, r_{outlamp}, that emits all light beams that are outgoing at that r-coordinate—that is, all beams with a positive r-coordinate differential, $dr > 0$.

A. An outgoing map floodlight at $r_{\text{outlamp}}/M = 8$ emits light that might have come from stars with approximately what range of map angles ϕ_∞?

B. An outgoing map floodlight at $r_{\text{outlamp}}/M = 5$ emits light that may have come from stars with approximately what range of map angles ϕ_∞?

C. An outgoing map floodlight at $r_{\text{outlamp}}/M = 3$ emits light that may have come from stars with approximately what range of map angles ϕ_∞?

D. Is there a range of r-coordinates in which the outgoing map floodlight is useless? *Hint:* look at Figure 10.

7. Newton's plot of map angle difference.

Make a *rough* sketch (don't sweat the details) of Figure 8 for orbits of light in Newtonian mechanics, in which spacetime is flat around the center of attraction and light is fast particle. What "Newtonian assumptions" do you make about the path of light under this attraction? (We have no record that Newton himself made any prediction about the effect of his "gravitational force" on the orbits of light.)

11.10 ■ REFERENCES

Initial quotes:

Egyptian creation quote from
M=http:/www.aldokkan.com/religion/creation.htm/=

Tuamotuan creation quote from *The Myths of Creation* by Charles H. Long, George Braziller, New York, 1963, pages 173 and 179.

Inuit creation quote from *The Power of Stars—How Celestial Observations Have Shaped Civilization* by Bryan E. Penprasem, New York, Springer 2011, page 97.

The interactive GRorbits program that plots orbits of light is available at website http://stuleja.org/grorbits/

Download File Name: Ch11OrbitsOfLight170511v1.pdf

Chapter 12. Diving Panoramas

- *In which local direction (or directions) does a local inertial rain observer look to see a given star as she passes coordinate r?*
- *In which direction (or directions) does a shell observer stationary at r and ϕ coordinates look to see the same star?*
- *How does the panorama of the heavens change for the local rain observer as she descends?*
- *Is gravitationally blue-shifted starlight lethal for the rain observer as she approaches the singularity? Is this starlight more dangerous than killer tides?*
- *How close to the singularity will the rain observer survive?*
- *What is the last thing the local rain observer sees?*

Download File Name: Ch12DivingPanoramas190403v1.pdf

CHAPTER 12

Diving Panoramas

Edmund Bertschinger & Edwin F. Taylor *

Tell all the truth but tell it slant –
Success in Circuit lies
Too bright for our infirm Delight
The Truth's superb surprise
As Lightning to the Children eased
With explanation kind
The Truth must dazzle gradually
Or every man be blind –

—Emily Dickinson

12.1 ■ FALLING INTO THE BLACK HOLE

See the same beam in two different directions.

"What's it like to fall into a black hole?"

"What is it like to fall into a black hole?" Our book thus far can be thought of as preparation to answer this question. The simplest possible answer has two parts: "What do I *feel* as I descend?" and "What do I *see* as I descend?" You *feel* tidal accelerations that—sorry about this—"spaghettify" you before you reach the singularity (Query 23, Section 7.9). As you descend, you *see* a changing panorama of the starry heavens, developed in this chapter and narrated in Section 12.7.

Where does a rain observer look to see a given beam?

The preceding Chapter 11 plotted trajectories of starlight in *global* rain coordinates and told us which beams (plural!) connect a given distant star to the *map location* of an observer. But that chapter said nothing about the direction in which that observer looks to see each beam or the beam energy she measures. These are the goals of the present chapter.

Comment 1. The rain diver
To dive—to be a diver—means to free-fall radially inward toward a center of attraction. A diver can drop from rest on—or be hurled radially inward from—any shell, including a shell far from the black hole. Among divers, the rain observer is a special case: a diver that drops from rest far away. In this chapter the word *diver* most often means the *rain diver*.

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

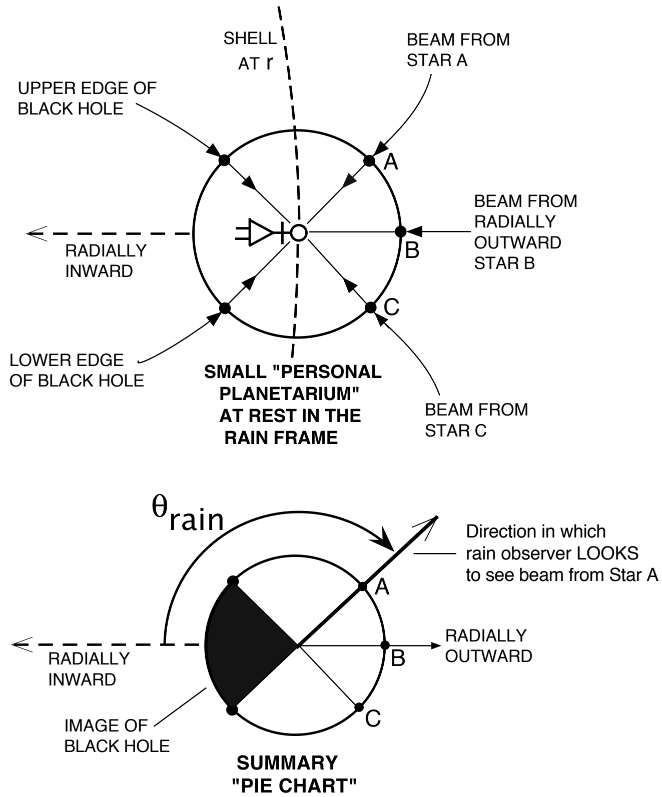


FIGURE 1 *Upper panel:* The **personal planetarium** is a small transparent sphere that encloses—and falls with—the local rain observer. The observer marks on the inside of this sphere the point-images of stars. She also draws a circle around the visual edge of the black hole. *Lower panel:* The **pie chart** summarizes rain observer markings; a black “pie slice” spans the visual image of the black hole. To see a particular beam, the rain observer *looks* in the direction $\theta_{\rm rain}$, which she measures clockwise from the radially inward direction.

12.2 ■ THE PERSONAL PLANETARIUM

Enjoy the view in weightless comfort.

“Personal planetarium” to record local observations

How does the local rain observer view and record starlight beams? One practical answer to this question is the **personal planetarium**: a transparent sphere at rest in the observer’s local inertial frame with the observer’s eye at its center (upper panel of Figure 1). Light beams run straight with respect to this local inertial frame, as shown in the figure. The observer marks on the inside of the transparent sphere the points of light she sees from stars in all directions; she also draws on the inside of the sphere a circle around the visual edge of the black hole.

We call the lower panel in figure 1 the **pie chart**. The pie chart takes its name from the standard graphical presentation whose black “pie slice” shows the fraction of some quantity as the proportion of the whole. In our pie chart the pie slice shows the range of visual angles covered by the black hole.

On the personal planetarium sphere the observer locates a star with the angle $\theta_{\rm rain}$ between the center of the black hole image and the dot she has placed on the image of that star. For simplicity, we omit the coordinate subscript "obs" for "observer" used in Chapter 11.

Definition: angle $\theta_{\rm rain}$

DEFINITION 1. Angle $\theta_{\rm rain}$

The observer in the personal planetarium looks in the direction $\theta_{\rm rain}$ to see the to see any given star. We define angle $\theta_{\rm rain} = 0$ to be radially inward, from the observer's eye toward the center of the black hole and the positive angle $\theta_{\rm rain}$ to be clockwise from this direction measured in her local rain frame (lower panel, Figure 1).

Full 3-dimensional rain observer's panorama

Can the local rain observer see a star that lies out of the plane of this page? Of course: Every star lies on *some* slice determined by three points: the star, the rain observer's eye, and the map coordinate $r = 0$. To encompass all stars in the heavens, rotate each of the circles in Figure 1 around its horizontal radial line. This rotation turns the pie chart into a sphere and the black hole "pie slice" into a cone. From inside her planetarium, the local rain observer sees the full panorama of stars in the heavens.

Objection 1. *You say, "From inside her planetarium, the local rain observer sees the full panorama of stars in the heavens." Why isn't that the end of the story? What more does the rain observer need to know?*

If she is satisfied to describe a general view of the heavens, that is sufficient for her. However, she may want to know, for example, where to look to see Alpha Centauri, one of Earth's nearest neighbors, as she plunges past $r = 4M$. The present chapter tells her the angle $\theta_{\rm rain}$ in which she looks to see the star located at global angle ϕ_∞. Finding $\theta_{\rm rain}$ of a star is a two-step process: The present chapter says in what local direction $\theta_{\rm rain}$ the rain observer looks to see a beam with a given value of b. The analysis from Chapter 11 then tells us the global angle ϕ_∞ of the star that emits the beam with that value of b. Angle $\theta_{\rm rain}$ depends on the observer's instantaneous global coordinate r, so varies with r as she descends. *Result:* a changing panorama, described in Section 12.8.

12.3 ■ RAIN FRAME VIEW OF LIGHT BEAMS

"I spy with my little eye . . . "

From stone to light

As she falls past r, the local rain observer sees a beam from a distant star at the angle $\theta_{\rm rain}$ with respect to the radially inward direction. We want an expression for this observation angle as a function of r and the b-value of that light beam. Chapter 11 defines b as the ratio $b \equiv L/E$. Equation (35) in Section 7.5 gives the expression for the map energy of a stone:

$$\frac{E}{m} = \left(1 - \frac{2M}{r}\right)\frac{dT}{d\tau} - \left(\frac{2M}{r}\right)^{1/2}\frac{dr}{d\tau} \qquad \text{(stone)} \qquad (1)$$

The expression for map angular momentum of a stone comes from equation (10) in Section 8.2:

$$\frac{L}{m} = r^2\frac{d\phi}{d\tau} \qquad \text{(stone)} \qquad (2)$$

b in global rain coordinates

Divide both sides of (2) by the corresponding sides of (1) and divide numerator and denominator of the result by dT. The symbol m cancels on the left side to yield the ratio $b \equiv L/E$ for light:

$$b \equiv \frac{L}{E} = \frac{r\left(\dfrac{rd\phi}{dT}\right)}{\left(1 - \dfrac{2M}{r}\right) - \left(\dfrac{2M}{r}\right)^{1/2}\dfrac{dr}{dT}} \qquad \text{(light)} \qquad (3)$$

Equation (3) expresses b in rain coordinates. But the local planetarium observer measures visual angles in her local inertial rain frame. Box 4 in Section 7.5 expressed local rain frame coordinates in global coordinates:

$$\Delta t_{\text{rain}} \equiv \Delta T \qquad (4)$$

$$\Delta y_{\text{rain}} \equiv \Delta r + \left(\frac{2M}{\bar{r}}\right)^{1/2}\Delta T \qquad (5)$$

$$\Delta x_{\text{rain}} \equiv \bar{r}\Delta\phi \qquad (6)$$

Light moves with speed unity in the local inertial rain frame, $\Delta s_{\text{rain}}/\Delta t_{\text{rain}} = 1$, so the Pythagorean Theorem provides labels for legs of the right triangle in Figure 2:

$$\Delta x_{\text{rain}}^2 + \Delta y_{\text{rain}}^2 = \Delta s_{\text{rain}}^2 = \Delta t_{\text{rain}}^2 \qquad \text{(light, in local inertial rain frame)} \; (7)$$

Find rain observer angle to see light with b

At what angle θ_{rain} does the rain observer look in her local frame in order to see the incoming beam with impact parameter b? The incoming beam in Figure 2 represents any one of the multiple beams arriving at the rain observer from a single star. In Figure 2 the symbols A, B, and C stand for the (positive) lengths of the sides of the right triangle. In contrast, the inward-moving light has negative components of motion radially and tangentially in the figure. The local frame time lapse Δt_{rain} is positive along the worldline. Expressing these results in local rain coordinates (4) through (6) leads to the following expressions for sine and cosine of the angle θ_{rain}:

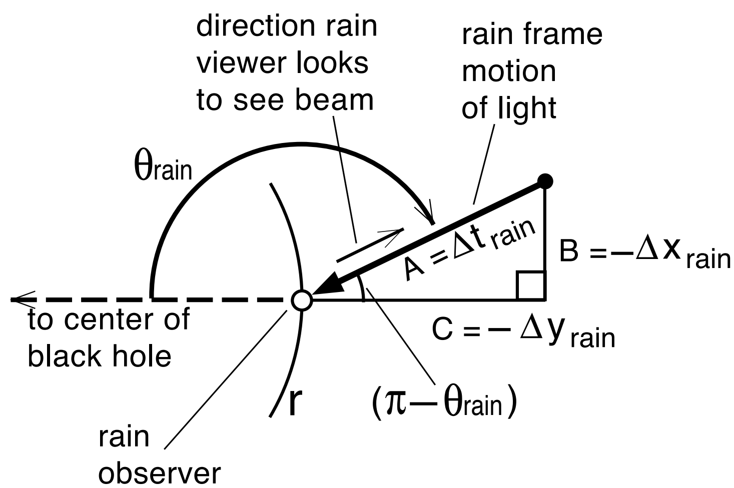


FIGURE 2 The beam from a distant star reaches the local rain observer as she dives inward past the shell at r. She measures the observation angle $\theta_{\rm rain}$ clockwise with respect to the radially inward direction. Letters A, B, and C are the (positive) lengths of the legs of the right triangle in the local rain frame. As the light approaches the observer, it has a clockwise tangential component in the local rain frame so $\Delta x_{\rm rain}$ is negative (and $-\Delta x_{\rm rain}$ is positive, equal to side B, as shown). The light also has a radially inward component, so $\Delta y_{\rm rain}$ is also negative (and $-\Delta y_{\rm rain}$ is positive, equal to C as shown). The hypotenuse is $\Delta s_{\rm rain} = \Delta t_{\rm rain}$ from (7) is positive and lies along the worldline of the beam $\Delta t_{\rm rain}$.

$$\begin{aligned}\sin\theta_{\rm rain} &= \sin(\pi - \theta_{\rm rain}) \qquad (8)\\ &= \lim_{A\to 0}\frac{B}{A} = \lim_{\Delta t_{\rm rain}\to 0}\frac{-\Delta x_{\rm rain}}{\Delta t_{\rm rain}} = \lim_{\Delta T\to 0}\frac{-\bar{r}\Delta\phi}{\Delta T} \qquad \text{so that}\\ \sin\theta_{\rm rain} &= -\frac{r d\phi}{dT} \qquad (9)\end{aligned}$$

A similar procedure leads to an expression for $\cos\theta_{\rm rain}$:

$$\begin{aligned}\cos\theta_{\rm rain} &= -\cos(\pi - \theta_{\rm rain}) \qquad (10)\\ &= -\lim_{A\to 0}\frac{C}{A} = -\lim_{\Delta t_{\rm rain}\to 0}\frac{-\Delta y_{\rm rain}}{\Delta t_{\rm rain}} = \lim_{\Delta T\to 0}\frac{\Delta r}{\Delta T} + \left(\frac{2M}{r}\right)^{1/2}\\ &= \frac{dr}{dT} + \left(\frac{2M}{r}\right)^{1/2} \qquad \text{so that}\\ \frac{dr}{dT} &= \cos\theta_{\rm rain} - \left(\frac{2M}{r}\right)^{1/2} \qquad (11)\end{aligned}$$

Substitute expressions (9) and (11) into (3), square both sides of the result, then eliminate the remaining sine squared with the identity $\sin^2\theta = 1 - \cos^2\theta$. Rearrange the result to yield the following quadratic equation in $\cos\theta_{\rm rain}$:

$$\left(1+\frac{b^2}{r^2}\frac{2M}{r}\right)\cos^2\theta_{\text{rain}} - 2\frac{b^2}{r^2}\left(\frac{2M}{r}\right)^{1/2}\cos\theta_{\text{rain}} - \left(1-\frac{b^2}{r^2}\right) = 0 \qquad (12)$$

Rain observer angle to see light with b

Solve the quadratic equation (12) to find an expression for $\cos\theta_{\text{rain}}$:

$$\cos\theta_{\text{rain}} = \frac{\dfrac{b^2}{r^2}\left(\dfrac{2M}{r}\right)^{1/2} \pm F(b,r)}{1+\dfrac{b^2}{r^2}\dfrac{2M}{r}} \qquad \text{(light)} \qquad (13)$$

Sign in numerator for global motion of light: $+$ for $dr > 0$, $-$ for $dr < 0$

where, from equation (16) in Section 11.2,

$$F(b,r) \equiv \left[1-\frac{b^2}{r^2}\left(1-\frac{2M}{r}\right)\right]^{1/2} \qquad \text{(light)} \qquad (14)$$

QUERY 1. Equations for $\cos\theta_{\text{rain}}$ (*Optional*)

A. Use equations (3), (9), and (11) to derive quadratic equation (12) for $\cos\theta_{\text{rain}}$.

B. Solve quadratic equation (12) to derive (13) for $\cos\theta_{\text{rain}}$.

There is an ambiguity in (13) because $\cos\theta = \cos(-\theta)$. To remove this ambiguity, substitute (9) and (11) into (3), then use (13) to substitute for $\cos\theta_{\text{rain}}$. After considerable manipulation, the result is:

$$\sin\theta_{\text{rain}} = \frac{b}{r}\left[\frac{-1 \pm \left(\dfrac{2M}{r}\right)^{1/2} F(b,r)}{1+\dfrac{b^2}{r^2}\dfrac{2M}{r}}\right] \qquad \text{(light)} \qquad (15)$$

which provides a stand-by correction of sign in (13). As in that equation, the plus sign is for $dr > 0$ and the minus sign for $dr < 0$.

QUERY 2. Derivation of $\sin\theta_{\text{rain}}$ (*Optional*)

Carry out the derivation of the expression for $\sin\theta_{\text{rain}}$ in (15).

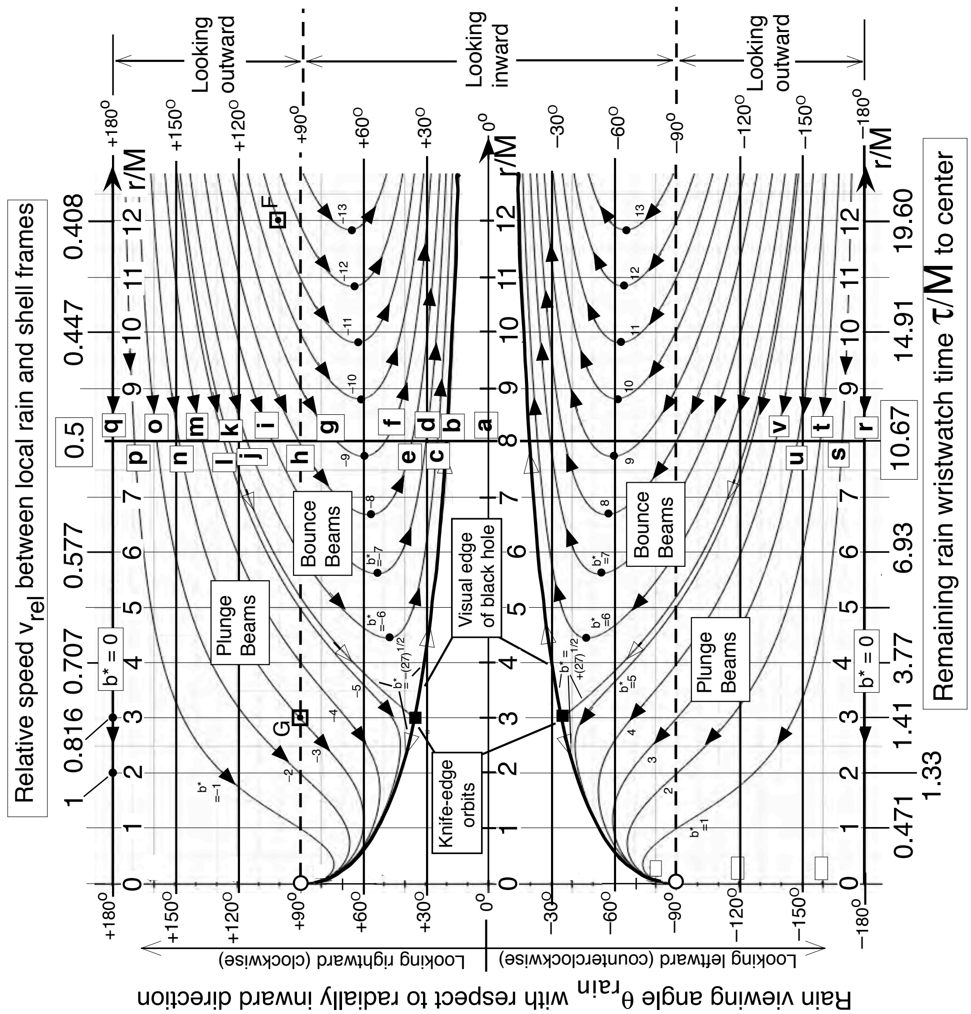


FIGURE 3 Angle θ_{rain} at which the local rain observer, passing the global location $(r, \phi = 0)$, looks to see starlight beams for several values of b/M, equation (13). To save space, we set $b^* \equiv b/M$. Arrows on the curve for each beam tell whether that beam is incoming or outgoing. A little black dot marks a turning point, the r-coordinate at which an incoming beam reverses dr to become an outgoing beam. Upper and lower three-branch curves with open arrowheads represent light with impact parameter $\pm b_{\text{critical}}/M = \pm(27)^{1/2}$. Two little black squares at $r = 3M$ represent circular knife-edge orbits of these critical beams on the tangential light sphere. As she descends, the rain observer sees the visual edge of the black hole at angles shown by the heavy curve. Reminder Figure 4 recalls the meaning of labels "Plunge Beams" and "Bounce Beams" in this figure. The text uses the lowercase labels **a** through **v** to explain details of this figure.

Plot rain observer angle to see light with b

Figure 3 plots equation (13) for several values of $b^* \equiv b/M$; it tells us in which directions θ_{rain} the local rain observer looks to see starlight beams with various impact parameters b. You can think of the vertical axis as unrolling the polar angle of the local rain frame.

Comment 2. Mirror image of upper and lower parts of Figure 3
Equation (13) is a function of b^2, so cannot distinguish between positive and negative values of b, which leads to the mirror symmetry of Figure 3 above an below the horizontal $\theta_{\text{rain}} = 0$ axis.

Numbers along the top of Figure 3 show shell speeds of the rain frame at various r-coordinates. You can check these numbers with equation (23) in Section 6.4:

$$v_{\text{rel}} = \left(\frac{2M}{r}\right)^{1/2} \qquad \text{(shell speed of rain diver, } r \geq 2M) \qquad (16)$$

Numbers along the bottom of Figure 3 tell the remaining wristwatch time τ the rain observer has before arriving at the singularity, from equation (2) in Section 7.2:

$$\tau[r \to 0] = \frac{2^{1/2}M}{3}\left(\frac{r}{M}\right)^{3/2} \qquad \text{(rain diver } \tau \text{ from } r \text{ to center)} \qquad (17)$$

Figure 4 reminds us of the meaning of labels in Figure 3.

Figure 3 carries an immense amount of information. We list here a few examples:

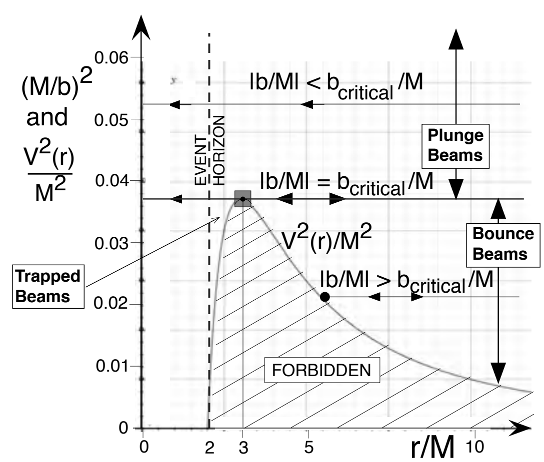


FIGURE 4 Reminder figure of the meaning of labels in Figure 3 for starlight beams—from Figure 3 in Section 11.3. There are no trapped starlight beams.

WHAT THE RAIN VIEWER SEES AS SHE PASSES $r = 8M$

1. **Visual Edge of the Black Hole**
   Two heavy curves straddle the r-axis in Figure 3; they form the outline of a trumpet. As the falling rain observer passes r, she sees ahead of her a black circle whose edges lie between angles $\pm\theta_{\text{rain}}$ given by these two heavy curves. As she descends further, the image of the black hole grows until, at $r = 0$, the black hole covers the entire forward hemisphere from $\theta_{\text{rain}} = -90°$ to $\theta_{\text{rain}} = +90°$.

Does a starlight beam escape?

2. **Dive or Escape?**
   Figure 3 shows that the starlight beam with $|b| < b_{\text{critical}}$ (plunge beam) is incoming, with $dr < 0$ along its entire length. In contrast, the starlight beam with $|b| > b_{\text{critical}}$ (bounce beam) is initially incoming, then reaches a turning point after which it becomes outgoing and escapes. Every starlight beam that escapes has a turning point, marked with the black dot in Figure 4.

Where does $r = 8M$ shell observer look to see beams with different values of b?

3. **Direction in which the Rain Viewer Looks to See a Star**
   Here is a detailed account of what the local rain observer sees as she falls past $r = 8M$: In Figure 3, look at points labeled with lower-case letters **a** through **v** on the vertical line at $r = 8M$. At point **a** on the r-axis, the rain observer looks radially inward at the center of the black hole, where she sees no starlight. When she looks somewhat to the right (point **b**), she sees the visual edge of the black hole at $\theta_{\text{rain}} \approx +23°$. This image is brought to her by the outgoing beam with $b = -b_{\text{critical}}$. Farther to her right, at points **c, d, e,** and **f** she sees outgoing beams with $b/M = -6, -7, -8$ and -9, respectively. She sees beams **a** through **g** by looking *inward*—that is, at angles $0 \leq \theta_{\text{rain}} < 90°$, as labeled on the right side of the figure. At point **h** she sees the beam with $b/M = -8$ at $\theta_{\text{rain}} = 90°$.

   Comment 3. Look inward to see a beam coming from behind?
   At point **g** in Figure 3, the rain viewer sees *incoming beam* $b/M = -9$ ahead of her at approximate angle $70°$. *Question:* How can she look *inward* to see a beam that the plot shows is coming from behind her? *Answer:* Aberration (Section 12.9). In Query 6 you explain this paradox for $r = 8M$.

   To see beams **i** through **q**, the local rain observer looks *outward*, at rain frame angles $90° < \theta_{\text{rain}} \leq 180°$, in which directions she sees incoming beams with values of b from about $b/M = -9$ to $b/M = 0$.

   Point **q** at the top of the diagram, for which $\theta_{\text{rain}} = +180°$, represents the radially outward direction, and is the same as point **r** at the bottom of the diagram, for which $\theta_{\text{rain}} = -180°$. Points **s** through **v** represent directions in which the rain observer looks to the left of the radially inward direction to see beams with $b/M = +1$ through $b/M = +4$ as she turns her gaze back toward the center of the black hole back at point **a**.

Important: The rain observer sees all of these beams *simultaneously* in her local frame as she falls inward past $r = 8M$.

4. **Three-Dimensional Panorama**
   Where does the local rain observer look to see beams from a star that does not lie in the plane of Figure 3? We know the answer to this question: Rotate Figure 3 around the central r-axis until the candidate star lies on the resulting global symmetry plane that contains the star, the observer's eye, and $r = 0$. You can use this result to construct the three-dimensional rain observer's view of every star in the heavens as she passes every r-coordinate.

3-dimensional panorama

Star images swing forward, then back to $\theta_{\rm rain} = \pm 90^o$.

Figure 3 also reveals something complex but fascinating: Look at incoming Plunge Beams at the top and bottom of Figure 3. As long as she remains outside $r = 3M$, the rain observer sees Plunge Beams move steadily to smaller visual angles $\theta_{\rm rain}$, even while the visual edges of the black hole are moving steadily to larger visual angles (heavy "trumpet" lines). The rain observer sees *only* Plunge Beams after she passes inward through $r = 3M$; after that she watches images of remote stars that emitted these Plunge Beams swing inward to a minimum visual angle, then back out again to final angles $\theta_{\rm rain} = \pm 90°$. In other words, after the rain observer passes inward through $r = 3M$, she sees all the multiple images of stars in the heavens swing forward to meet the expanding edge of the black hole, then remain at this edge as the black hole continues to grow visually larger.

12.4 ■ CONNECT STAR MAP ANGLE TO RAIN VIEWING ANGLE.

The rain observer views the heavens

Want relation between ϕ_∞ and $\theta_{\rm rain}$.

Focus on primary beam.

Can we now predict the sequence of panoramas of stars enjoyed by the rain observer as she descends? Figure 3 is powerful: It tells us where each rain observer looks to see a starlight beam with any given value of the impact parameter b/M. This anchors the receiving end of the beam at the rain observer. Now we need to anchor the sending end of the same beam at the star—that is, to find the map angle ϕ_∞ of the star that emits this beam.

To anchor both ends of each beam, we use our graphical relations among values of b/M, $\theta_{\rm rain}$, and ϕ_∞ for an observer located at $(r, \phi = 0)$. Figure 3 shows the rain angle $\theta_{\rm rain}$ at which a local rain observer looks to see a beam with given b-value. Figures 8 and 10 in Section 11.8 show the relation between the impact parameter b and the map angle ϕ_∞ to the star. Taken together, the figures in these two chapters (and their generating equations) solve our problem: They provide the two-step procedure to go between ϕ_∞ and $\theta_{\rm rain}$. To begin, we focus on the **primary image**—the image due to the primary beam, the most direct beam from star to observer. But the following procedure is valid for any beam whose b-value connects a star to a rain observer.

FROM STAR MAP ANGLE ϕ_∞ TO RAIN VIEWING ANGLE $\theta_{\rm rain}$

From ϕ_∞, find $\theta_{\rm rain}$

At what angle $\theta_{\rm rain}$ does a local rain observer located at map coordinates $(r, \phi = 0)$ look to see the primary image of a star at given map angle ϕ_∞? Here is the two-step procedure:

Step A. Figures 8 and 10 in Section 11.8 tell us the b-value of the primary beam that connects the star at map angle ϕ_∞ to this observer's location and whether that beam is incoming or outgoing.

Step B. For that b-value—and knowledge of whether the beam is incoming or outgoing—Figure 3 gives the local viewing angle $\theta_{\rm rain}$ in which the rain observer looks to see his primary image of that star.

Comment 4. Reminder: Two angles, ϕ_∞ and $\theta_{\rm rain}$
We measure the map angle ϕ_∞ to a star—as we measure all map angles—*counterclockwise* with respect to the radially *outward* direction. In contrast—and for our own convenience—we measure the rain observing angle $\theta_{\rm rain}$ *clockwise* from the radially *inward* direction.

From $\theta_{\rm rain}$, find ϕ_∞.

We can also run this process backward, from rain observation angle $\theta_{\rm rain}$ to map angle ϕ_∞. The rain observer sees a star at rain angle $\theta_{\rm rain}$. Find the map angle ϕ_∞ to that star using Step B above, followed by Step A: From the value of $\theta_{\rm rain}$, Figure 3 tells us the b-value of the beam and whether it is incoming or outgoing. From this information, Figures 8 and 10 in Section 11.8 give us the map angle ϕ_∞ of the star from which this beam comes.

Automate rain panorama plots.

Comment 5. Automate rain panorama plots.
To plot panoramas of the rain viewer, we read numbers from curves in figures, which yield only approximate values. Nothing stops us from converting the data in these figures (and the equations from which they come) into look-up tables or mathematical functions directly used by a computer. Then from the map angle ϕ_∞ to every star in the heavens, the computer automatically projects onto the inside of the personal planetarium (Figure 1) the visual panorama seen by the rain observer.

The following Queries and Sample Problems provide examples and practice connecting star map angle ϕ_∞ with the angle $\theta_{\rm rain}$ in which a local rain observer looks to see that star.

QUERY 3. Sequential changes in rain viewing angles of different stars
A rain viewer first looks at a given star when she is far from the black hole; later she looks at the same star as she hurtles in turn past each of the r-values $r/M = 12$, 5, 2, 1, and just before she reaches the singularity. Do the following Items twice: once using plots in the figures, and second *optional* using equations (13) and (14).

Find the viewing angle $\theta_{\rm rain}$ at each r-coordinate for the star at each of the following map angles. At each r, find the value b/M of the beam that the rain observer sees from that star.

Sample Problems 1. What can the local rain observer at $r = 6M$ see?

In the following cases the local rain observer is passing one of our standard locations, $(r, \phi = 0)$.

A. What is the range of b-values of starlight beams that the rain observer at $r = 6M$ can see? When this observer looks inward, $0 \leq |\theta_{\rm rain}| < 90°$, what is the range of b-values of beams that she can see? **SOLUTION:** In Figure 3, look at the vertical line at $r = 6M$. Beams with b-values in the range $0 \leq |b/M| \leq\approx 7.3$, either incoming or outgoing, cross that vertical line. These are the beams that she can see. Beams that the rain observer can see looking inward, $-90° < \theta_{\rm rain} < +90°$, have b-values in the range $b_{\rm critical}/M \leq |b/M| \leq\approx 7.3$.

B. In Part A, the largest value of $|b|$ for light seen by a rain observer at r occurs for a beam whose turning point is at that r-coordinate. What is that maximum value of $|b|$ for $r = 6M$? **SOLUTION:** Use equation (36) in Section 11.4. For $r_{\rm tp} = 6$, the answer is $b/M = \pm 7.348$, the correct value compared with the approximate value we read off the plot in Figure 3.

C. At what rain angle $\theta_{\rm rain}$ will the rain viewer passing $r = 6M$ look to see a star at map angle $\phi_\infty = 180°$, exactly on the opposite side of the black hole from her? **SOLUTION:** Begin with Figure 10, Section 11.7. Look at the intersection of the vertical line at $r = 6M$ and top and bottom horizontal lines at $\phi_\infty = \pm 180°$. The b values of these beams is $b/M \approx \pm 6.6$; the figure tells us that these are outgoing beams. The plus or minus refers to beams that come around opposite sides of the black hole. Now return to Figure 3 and find the intersection of vertical line $r = 6M$ with outgoing beams whose impact parameters are $b/M = \pm 6.6$. These intersections correspond to $\theta_{\rm rain} \approx \pm 110°$. These angles are greater than $\pm 90°$, so the rain observer looks somewhat benind her to see the star on the opposite side of the black hole.

D. The local rain observer passing $r = 6M$ sees a star at angle $\theta_{\rm rain} = +74°$. What is the map angle ϕ_∞ to that star? Is this beam incoming or outgoing? **SOLUTION:** In Figure 3 the point $(r = 6M, \theta_{\rm rain} = +74°)$ lies on the curve for the incoming beam with $b/M = -7$. Now go to Figure 10, Section 11.7 to find the intersection of $r = 6M$ with the incoming beam with $b/M = -7$. This occurs at the map angle $\phi_\infty \approx +90°$.

E. Can the rain observer at $r = 6M$ see a star that lies outside of the plane of Figure 3, for example? **SOLUTION:** Sure: just rotate every relevant figure around its horizontal axis until the desired star lies in the resulting plane, then carry out the analysis as before.

A. The star at $\phi_\infty = 30°$
B. The star at $\phi_\infty = 90°$
C. The star at $\phi_\infty = 150°$

QUERY 4. Given ϕ_∞, find $\theta_{\rm rain}$.
Each of the following items lists the map angle ϕ_∞ to a star and the r-coordinate of a rain observer at map coordinates $(r, \phi = 0)$ who looks at that star. In each case find the rain angle $\theta_{\rm rain}$ (with respect to the radially inward direction) at which the local rain observer looks to see that star.

A. $\phi_\infty = +30°$, $r = 6M$
B. $\phi_\infty = -120°$, $r = 10M$
C. $\phi_\infty = +90°$, $r = 2.5M$
D. $\phi_\infty = -180°$, $r = 12M$
E. The rain observer is at the turning point of the beam that has $b/M = -7$. In what rain direction $\theta_{\rm rain}$ does she look to see that star? At what map angle ϕ_∞ is the star that she sees?

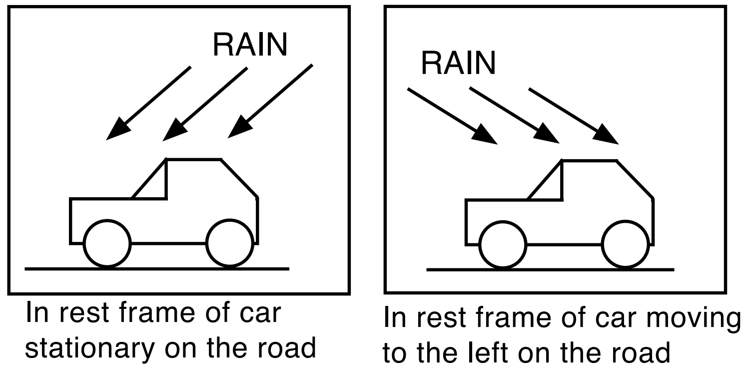


FIGURE 5 Aberration of rain as an analogy of the aberration of light. In the left panel no rain falls directly on the windshield; in the right panel the driver sees the rain coming from a forward direction.

QUERY 5. Given θ_{rain}, find ϕ_∞.

Each of the following items lists the r-coordinate of a rain observer at map coordinates $(r, \phi = 0)$ and the rain angle θ_{rain} at which she looks to see a given star. In each case find the map angle ϕ_∞ of that star.

A. $r = 4M,\ \theta_{\text{rain}} = -115°$
B. $r = 10M,\ \theta_{\text{rain}} = +80°$
C. $r = 2.5M,\ \theta_{\text{rain}} = -145°$
D. $r = 6M,\ \theta_{\text{rain}} = +155°$
E. The rain observer sees the visual edge of the black hole at $\theta_{\text{rain}} = +70°$. What is the map angle of the star that he sees at this visual edge?

12.5 ■ ABERRATION

Rain on the windshield

Aberration: spectacular consequences

Aberration is the difference in direction in which light moves as observed in overlapping inertial frames in relative motion. Aberration has spectacular consequences for what the local rain observer sees as she approaches and crosses the black hole's event horizon. Figure 5 shows an analogy: Rain that falls on a stationary and on a fast-moving car comes from different directions as viewed by a rider in the car. Light moves differently than rain, but the general idea is the same.

We deal here with different viewing directions in overlapping local inertial frames, so special relativity suffices for this analysis. Exercise 18 in Chapter 1 derived expressions for aberration between laboratory and rocket frames in special relativity. We need to modify these equations in four ways:

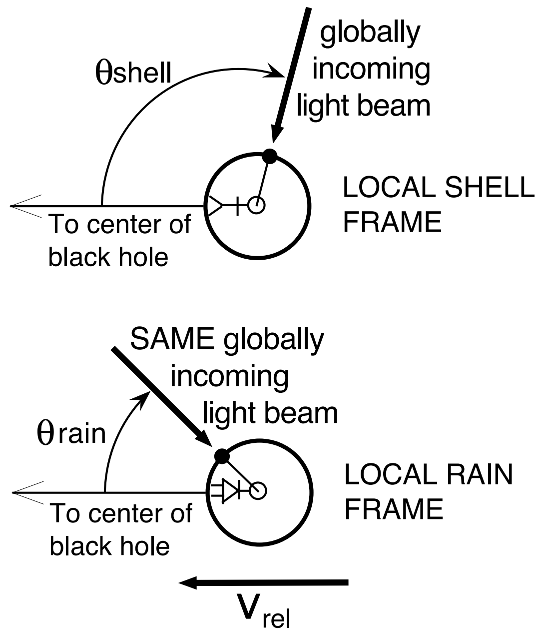


FIGURE 6 Example of light aberration for shell and local rain observers from equation (18). The shell observer at $r = 3M$ looks at the angle $\theta_{\rm shell} = 105°$ to see the beam from a star. The rain observer who passes this shell sees the beam at $\theta_{\rm rain} = 45°$.

Modify special relativity aberration equation for local rain and shell frames.

Modify Chapter 1 aberration equations.

1. Choose the shell frame (outside the event horizon) to be the laboratory frame and the local rain frame to be the rocket frame.
2. The direction of relative motion is along the common $\Delta y_{\rm frame}$ line instead of along the common $\Delta x_{\rm frame}$ line in Chapter 1.
3. The local rain frame moves in the negative $y_{\rm shell}$ direction, so $v_{\rm rel}$ in the aberration equations must be replaced by $-v_{\rm rel}$. Equation (16) gives $v_{\rm rel} = (2M/r)^{1/2}$.
4. The original special relativity aberration equations describe the direction (angle ψ) in which light *moves*. In contrast, our angles $\theta_{\rm shell}$ and $\theta_{\rm rain}$ refer to the direction in which the observer *looks* to see the beam, which is the opposite direction. Because of this, $\cos\psi$ in Chapter 1 becomes $360^{\rm o} - \cos\theta$ in the present chapter.

When all these changes are made, the aberration equation (54) in exercise 18 of Chapter 1 becomes:

$$\cos\theta_{\rm shell} = \frac{\cos\theta_{\rm rain} - \left(\dfrac{2M}{r}\right)^{1/2}}{1 - \left(\dfrac{2M}{r}\right)^{1/2}\cos\theta_{\rm rain}} \qquad \text{(light)} \qquad (18)$$

QUERY 6. Resolve the paradox in Comment 3 First of all, note that the question posed in Comment 3 in Section 12.3 is bogus. The beam "comes from behind" only in global coordinates; but we must not trust global coordinates to tell us about measurements or observations. Instead, you can use our equations to show that the descending local rain observer at $r = 8M$ sees the incoming beam with $b/M = -9$ at the inward angle $\theta_{\text{rain}} \approx 72°$, as follows:

A. Substitute the data from point **g** into equation (14) to show that $F(b, r) = F(-9M, 8M) = 0.225$.
B. Plug the results of Item A plus $r = 8M$ and $b/M = -9$ into equation (13) to show that $\cos\theta_{\text{rain}} = 0.310$.
C. From Item B, show that $\theta_{\text{rain}} = 72°$. Does this result match the vertical location of point **g** in Figure 3?

QUERY 7. Rain view at the event horizon.

A. When the rain observer passes through $r = 2M$, at what angle θ_{rain} does she see the edge of the black hole?
B. Can the rain observer use her panorama of stars to detect the moment when she crosses the event horizon?
C. Does your answer to Item B violate our iron rule that a diver cannot detect when she crosses the horizon?

12.6 ■ RAIN FRAME ENERGY OF STARLIGHT

Falling through gravitationally blue-shifted light.

Rain observer in danger from starlight?

Section 7.9 predicted that as she approaches the singularity, tidal forces will end the experience of the rain diver during a fraction of a second, as measured on her wristwatch—independent of black hole mass. But starlight increases its frequency, and hence its locally-measured energy, as it falls toward the black hole. Will the rain observer receive a lethal dose of high-energy starlight before she reaches the singularity? To engage this question, we analyze the energy of light E_{rain} measured in the local rain frame compared to its map energy E. Equation (1) gives the map energy E of a stone in global rain coordinates. The special relativity expression for rain frame energy with the substitution $\Delta t_{\text{rain}} \equiv \Delta T$ from (4) yields

$$\frac{E_{\text{rain}}}{m} = \lim_{\Delta\tau\to 0}\left(\frac{\Delta t_{\text{rain}}}{\Delta\tau}\right) = \lim_{\Delta\tau\to 0}\left(\frac{\Delta T}{\Delta\tau}\right) = \frac{dT}{d\tau} \qquad \text{(stone)} \qquad (19)$$

Modify this expression to describe light. Substitute $dT/d\tau$ from (19) into (1) and solve for E_{rain}:

$$E_{\text{rain}} = \left(1 - \frac{2M}{r}\right)^{-1} E\left[1 + \left(\frac{2M}{r}\right)^{1/2} \frac{m}{E}\frac{dr}{d\tau}\right] \qquad \text{(stone)} \qquad (20)$$

Recall equation (15) in Section 8.3:

$$\frac{dr}{d\tau} = \pm\left[\left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{m^2 r^2}\right)\right]^{1/2} \qquad \text{(stone)} \qquad (21)$$

Use equation (21) to replace the expression $(m/E)(dr/d\tau)$ just inside the right-hand square bracket in (20):

$$\frac{m}{E}\frac{dr}{d\tau} = \pm\left[1 - \left(1 - \frac{2M}{r}\right)\left(\frac{m^2}{E^2} + \frac{L^2}{E^2 r^2}\right)\right]^{1/2} \qquad \text{(stone)} \qquad (22)$$

This equation is for a stone. Turn it into an equation for light by going to the limit of small mass and high speed: $m \to 0$ and $L/E \to b$. Plug the result into (20) and divide through by E, which then becomes:

$$\frac{E_{\text{rain}}}{E} = \frac{1 \pm \left(\frac{2M}{r}\right)^{1/2}\left[1 - \frac{b^2}{r^2}\left(1 - \frac{2M}{r}\right)\right]^{1/2}}{1 - \frac{2M}{r}} \qquad \text{(light)} \qquad (23)$$

Use (14) to write this as:

$$\frac{E_{\text{rain}}}{E} = \frac{1 \pm \left(\frac{2M}{r}\right)^{1/2} F(b,r)}{1 - \frac{2M}{r}} \qquad \text{(light)} \qquad (24)$$

Numerator: minus sign for incoming starlight, plus sign for outgoing light

where equation (14) defines $F(b,r)$. In (23)—and therefore in (24)—the impact parameter b is squared; therefore the beam has the same map energy whether it moves clockwise or counterclockwise around the black hole, as we expect.

In the exercises you derive an expression E_{rain}/E for the stone.

Rain-frame energies of beams

Figure 7 plots results of equation (24). We expect the rain frame energy of an *incoming* beam to depend on two competing effects: the gravitational blue shift (increase in local frame energy) of the falling light, reduced for the diving observer by her inward motion. The result is the Doppler downshift in energy of the light viewed by this local rain observer—compared to the same light

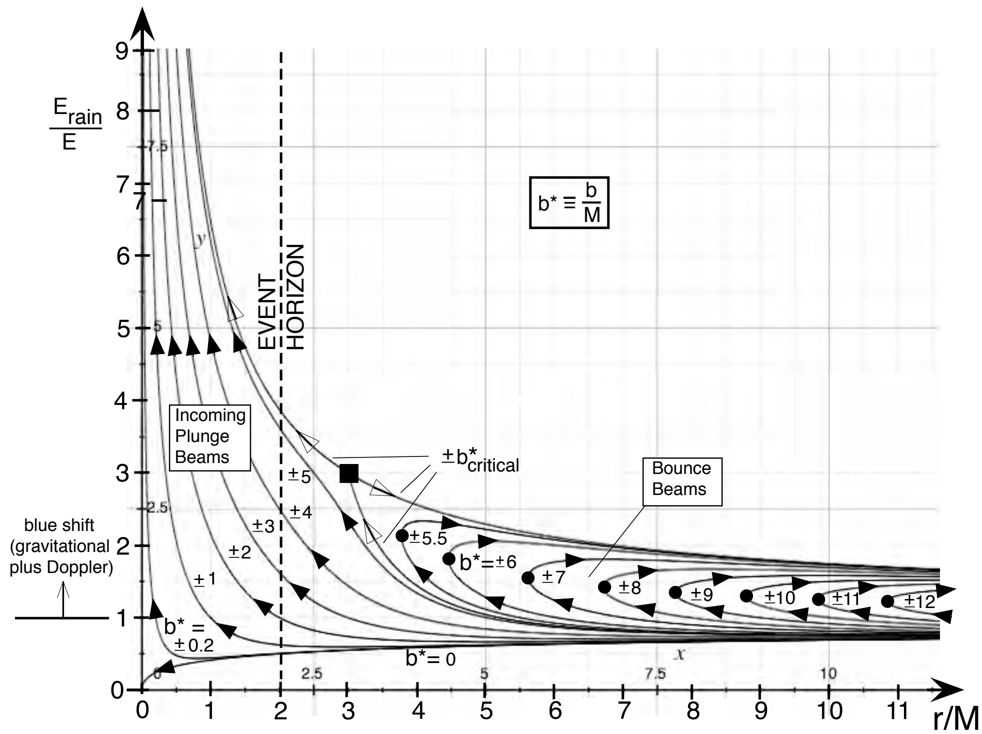


FIGURE 7 Ratio $E_{\rm rain}/E$ of starlight measured by the local rain observer, from equation (24). The curve rising out of each turning point describes an outgoing beam. Beams with $0 \leq |b/M| \leq 5$ are incoming plunge beams.

beam viewed by the local shell observer. In contrast, starlight that has passed its turning point and heads outward again in global coordinates moves opposite to the incoming rain observer, so she will measure its energy to be Doppler up-shifted. Figure 7 shows that these two effects yield a net blue shift for some beams and parts of other beams, and a net red shift for still other beams.

QUERY 8. Rain energy of light at large r.

What happens to the value of $E_{\rm rain}$ as $r \to \infty$? Show that

$$\lim_{r\to\infty} E_{\rm rain} = E \qquad \text{(light)} \tag{25}$$

QUERY 9. Rain energies of radially-moving starlight.

Find an expression for the rain energy of light with $b = 0$ that moves radially inward (for any value of r) or outward (for $r \gg 2M$). Show that in this case equation (24) becomes

$$\frac{E_{\text{rain}}}{E} = \frac{1 \pm \left(\dfrac{2M}{r}\right)^{1/2}}{1 - \dfrac{2M}{r}} = \frac{1 \pm \left(\dfrac{2M}{r}\right)^{1/2}}{\left[1 + \left(\dfrac{2M}{r}\right)^{1/2}\right]\left[1 - \left(\dfrac{2M}{r}\right)^{1/2}\right]} = \frac{1}{1 \mp \left(\dfrac{2M}{r}\right)^{1/2}} \qquad (\text{light, } b = 0) \quad (26)$$

where the upper sign is for outgoingbeams. But starlight with $b = 0$ cannot be outgoing, so:

$$\frac{E_{\text{rain}}}{E} = \frac{1}{1 + \left(\dfrac{2M}{r}\right)^{1/2}} \qquad (\text{starlight, } b = 0) \qquad (27)$$

This is the curve displayed at the bottom of Figure 7.

A. Show that E_{rain}/E has the value 1/2 at $r = 2M$, and that this result is consistent with the $b = 0$ curve in Figure 7.

B. A shell observer remote from a black hole shines radially inward a laser of map energy E_{laser}, measured in his local frame, which is also global E in flat spacetime. A local rain observer moving inward along the same radial line looks radially outward at this laser beam as she descends. Write a short account about the ratio $E_{\text{rain}}/E_{\text{laser}}$ of this laser light that she measures outside the event horizon, when she is at the event horizon, and as she approaches the singularity. If she is given the value of E_{laser}, can she detect when she crosses the event horizon? Could you design an "event horizon alarm" for our black hole explorations? Does your design violate our iron rule that a diver cannot detect when she crosses the event horizon?

QUERY 10. Details of Figure 7

Without equations, provide qualitative explanations of rain frame beam energies in Figure 7.

A. Show that E_{rain} approaches the value E at large r, as demonstrated in (25).

B. Why do beam energies not depend on the sign of b?

C. Why do the outgoing Bounce Beams at any given r have greater rain frame energy than incoming Bounce Beams?

D. For the Plunge Beams, $0 \le |b| \le b_{\text{critical}}$ at any given r, why do beams with larger values of $|b|$ have greater rain frame energy than beams with smaller values of $|b|$?

QUERY 11. *Optional:* Trouble at the event horizon?

The denominator $1 - 2M/r$ in (24) goes to zero at the event horizon. Does this mean that at $r = 2M$ starlight has infinite rain frame energy for every value of b? To answer, use our standard approximation (inside the front cover); set $r = 2M(1 + \epsilon)$, where $0 < \epsilon \ll 1$ and verify that E_{rain}/E is finite at the event horizon, provided that the beam is not an outgoing Plunge Beam. Show that your approximation at $r/M = 2$ correctly predicts values of E_{rain}/E for two or three of the curves in Figure 7.

QUERY 12. Killer starlight?
What is the energy of starlight measured by the local rain observer as she approaches the singularity? Let $r/M = \epsilon$, where $0 < \epsilon \ll 1$ and show that for starlight (incoming Plunge Beams: minus sign in (24)):

$$\lim_{r\to 0} \frac{E_{\text{rain}}}{E} = \frac{|b|}{r} \qquad \text{(incoming Plunge Beams)} \tag{28}$$

Lethal starlight is bad, but tides are worse.

In Query 12 you show that close to the singularity the energy of starlight measured by the plunging local rain observer increases as the *inverse first power* of the decreasing r. The other mortal danger to the rain observer comes from tidal accelerations. Section 7.9 showed that the rain observer "ouch time" from first discomfort to arrival at the singularity is two-ninths of a second, independent of the mass of the black hole. Equations (38) through (40) in Section 9.7 tell us that tidal accelerations increase as the *inverse third power* of the decreasing r-coordinate, which is proportionally faster than the inverse first power increase in the rain frame energy of incoming starlight.

Which will finally be lethal for the rain observer: killer starlight or killer tides? Inverse third power tidal acceleration appears to be the winning candidate. Analyzing tidal acceleration is straightforward: its effects are simply mechanical. In contrast, we have trouble predicting results for light: they depend not only on the rain frame energy of the light but also on its intensity and the rain observer's wristwatch exposure time. This book says nothing about the focusing properties of curved spacetime near the black hole—an advanced topic—so we lack the tools to predict the (short-term!) consequences of the rain observer's accumulated exposure to starlight as she descends.

Assume tides are lethal

We have a lot of experience protecting humans against radiation of different wavelengths. Perhaps a specially-designed personal planetarium (Section 12.2) will allow the rain observer to survive killer starlight all the way down to her tidal limit. In contrast, we know nothing that can shield us from tidal effects. In the description of the final fall in Section 12.7 we assume that it is killer tides that prove lethal for the rain diver.

12.7 ■ THE FINAL FALL

Free-fall to the center

We celebrate with the final parade of an all-star cast. Let's follow general relativists Richard Matzner, Tony Rothman, and Bill Unruh (see the references) looking at the starry heavens as we free-fall straight down into a

Time of a movie inside the event horizon

non-spinning black hole so massive, so large that even after crossing the event horizon we have nearly two hours of existence ahead of us—roughly the length of a movie—to behold the whole marvelous ever-changing spectacle. Almost everything we have learned about relativity—both special and general—contributes to our appreciation of this mighty sequence of panoramas.

Panoramas Seen by the Rain Frame Observer

—Adapted from Matzner, Rothman, and Unruh. Some numerical values calculated by Luc Longtin.

Fall into a one-billion-solar-mass black hole.

Imagine a free-fall journey into a billion-solar-mass black hole ($M = 10^9 M_{\text{Sun}} = 1.477 \times 10^9$ kilometers $= 1.6 \times 10^{-4}$ light-years—about one-third of the estimated mass of the black hole at the center of galaxy M87). The map r-coordinate of the event horizon—double the above figure—is about the size of our solar system. We adjust our launch velocity to match the velocity which a rain frame, falling from far away, would have at our shell launch point at $r = 5\,000M$. Our resulting inward shell launch velocity, $v = -(2M/r)^{1/2}$ with respect to a local shell observer, is equal to two percent of the speed of light. We record each stage in the journey by giving both the time-to-crunch on our wristwatch and our current map r.

26 years to the end

The beginning of the journey, 26 years before the end. At this point the black hole is rather unimpressive. There is a small region (about 1 degree across—i.e., twice the size of the Moon seen from Earth) in which the star pattern looks slightly distorted and within it (covering about one-tenth of a degree) a disk of total blackout. Careful examination shows that a few stars nearest the rim of the blacked-out region have second images on the opposite side of the rim. Had these images not been pointed out to us, we probably would have missed the black hole entirely.

300 days

Three hundred days before the end, at $r = 500M$. Some noticeable change has occurred. The dark circular portion of the sky has now grown to one full degree in width.

One week

One week before the end, at $r = 41M$. The image has grown immensely. There is now a pure dark patch ahead with a diameter of about 22 degrees (approximately the size of a dinner plate held at arm's length). The original star images that lay near the direction of the black hole have been pushed away from their original positions by about 15 degrees. Further, between the dark patch itself and these images lies a band of second images of each of these stars. Looking at the edge of this darkness with the aid of a telescope, we can even see faint second images of stars that lie behind us! This light has looped around the black hole on its way to our eye (Figure 9, Section 11.8). From this point on, Doppler shift and gravitational blue shift radically change the observed frequencies of light that originate from different stars.

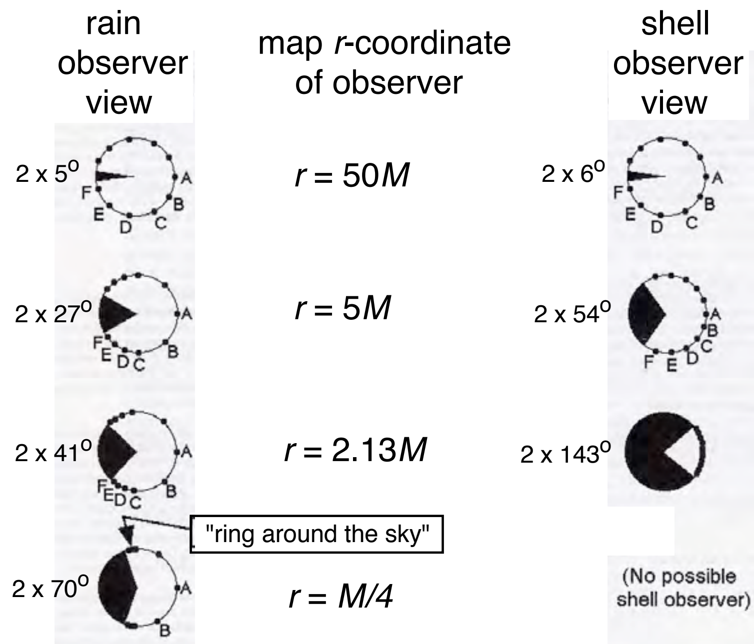


FIGURE 8 Pie charts showing rain viewing directions of stars and the visual edge of the black hole seen in sequence by a rain frame viewer (left-hand column) and by a set of stationary shell observers at different radii (right-hand column). Dots labeled A through F represent directions of stars plotted by rain and shell observers on their personal planetariums. In the final instants of her journey (at smaller radii than shown here), the sky behind the rain observer is black, nearly empty of stars, and the black hole covers the sky ahead of her. Cleaving the forward half of the firmament from the backward half is a bright ring around the sky. This figure does not show multiple images of stars due to one or several orbits of starlight around the black hole. (Figure based on the work of M. Sikora, courtesy of M. Abramowicz.)

12 hours

Twelve hours before the end, at $r = 7M$. A sizeable portion of the sky ahead of us is now black; the diameter of the black hole image covers a 44-degree angle, over 10 percent of the entire visual sphere.

3.3 hours

3.3 hours before the end. As we pass inward through $r = 3M$, we see all the stars in the heavens swing forward to meet the expanding edge of the black hole, then remain at this edge as the black hole continues to grow visually larger.

2 hours

Two hours before the end. We are now at $r = 2.13M$, just outside the event horizon and our speed is 97 percent that of light as measured in the local shell frame that we are passing. Changes in viewing angle (aberrations) are now extremely important. Anything we see after an instant from now will be a secret taken to our grave, because we will no longer be able to send any information out to our surviving colleagues. Although we will be “inside” the black hole, not all of the sky in front of us appears entirely dark. Our high speed causes light beams to arrive at our eyes at extreme forward angles. Even so, a disk subtending a total angle of 82 degrees in front of us is fully black—a substantial fraction of the forward sky.

Secondary images

Behind us we see the stars grow dim and spread out; for us their images are not at rest, but continue to move forward in angle to meet the advancing edge of the black hole. This apparent star motion is again a forward-shift due to our increasing speed. But there is a more noticeable feature of the sky: We can now see second images of all the stars in the sky surrounding the black hole. These images are squeezed into a band about 5 degrees wide around the image of the black hole. These second images are now brighter than were the original stars. Surrounding the ring of second images are the still brighter primary images of stars that lie ahead of us, behind the black hole. The band of light caused by both the primary and secondary images now shines with a brightness ten times that of Earth's normal night sky.

2 minutes

Approximately two minutes before oblivion: $r = M/7$. The black hole now subtends a total angle of 150 degrees from the forward direction—almost the entire forward sky. Behind us star images are getting farther apart and rushing forward in angle. Only 20 percent of star images are left in the sky behind us. In a 10-degree-wide band surrounding the outer edges of the black hole, not only second but also third and some fourth images of the stars are now visible. This band running around the sky now glows 1000 times brighter than the night sky viewed from Earth.

Final seconds

The final seconds. The sky is dark everywhere except in that rapidly thinning band around the black disk. This luminous band—glowing ever brighter—runs completely around the sky perpendicular to our direction of motion. At 3 seconds before oblivion it shines brighter than Earth's Moon. New star images rapidly appear along the inner edge of the shrinking band as higher and higher-order star images become visible from light wrapped many times around the black hole. The stars of the visible Universe seem to brighten and multiply as they compress into a thinner and thinner ring transverse to our direction of motion.

Awesome ring bisects the sky.

Only in the last 2/9 of a second on our wristwatch do tidal forces become strong enough to end our journey and our view of that awesome ring bisecting the sky.

12.8 ■ EXERCISES

1. Impact parameter at a turning point

From equation (29) in Section 11.5, show that $b/M \to \infty$ not only as $r_{\rm tp} \to \infty$ but also as $r_{\rm tp}/M \to 2^+$, where the subscript tp means turning point. Since b/M is finite for values between these two limits, therefore there must be at least one minimum in the b vs. $r_{\rm tp}$ curve. Verify the map location and value of this minimum, shown in Figure 9. Remember that beams for which $r_{\rm tp} < 3M$ cannot represent starlight.

2. Direction of a star seen by the local shell observer.

Exercise 18 in Section 1.13 shows the relation between the directions in which light moves in inertial laboratory and rocket frames. Replace laboratory with local shell frame and rocket with local rain frame. The direction of relative motion is along the local y-axes, and the rain frame moves moves in the negative $\Delta y_{\rm shell}$ direction, so the sign of the relative velocity $v_{\rm rel}$ must be reversed in the special relativity formulas. Equation (56) in Section 1.13 becomes

$$\cos\psi_{\rm shell} = \frac{\cos\psi_{\rm rain} - v_{\rm rel}}{1 - v_{\rm rel}\cos\psi_{\rm rain}} \qquad (\psi = \text{direction of light motion}) \qquad (29)$$

where $\psi_{\rm shell}$ is the direction the light moves in the shell frame and $\psi_{\rm rain}$ its direction of motion in the rain frame.

A. In the notation of Chapter 1, angles ψ are the directions in which the light *moves*; in the notation of the present chapter angles θ are the angles in which an observer *looks* to see the beam. The cosines of two angles that differ by 360° are the same. Show that equation (18) becomes, in the notation of our present chapter:

$$\cos\theta_{\rm shell} = \frac{\cos\theta_{\rm rain} - v_{\rm rel}}{1 - v_{\rm rel}\cos\theta_{\rm rain}} \qquad (\theta = \text{direction viewer looks}) \qquad (30)$$

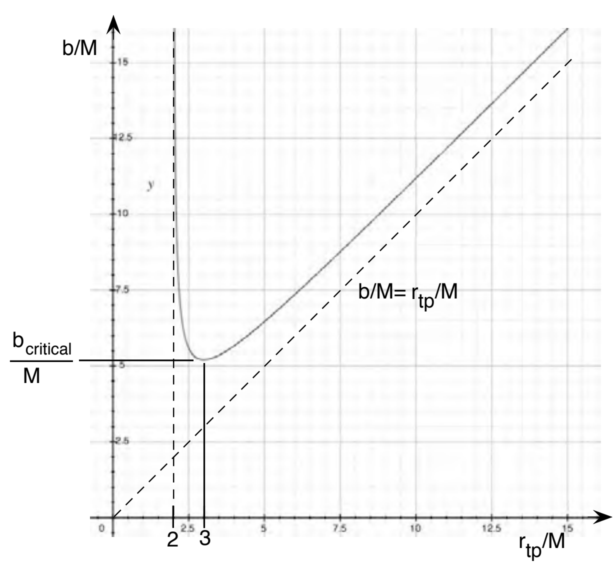


FIGURE 9 Plot of b/M vs. $r_{\rm tp}/M$ from equation (29) in Section 11.5.

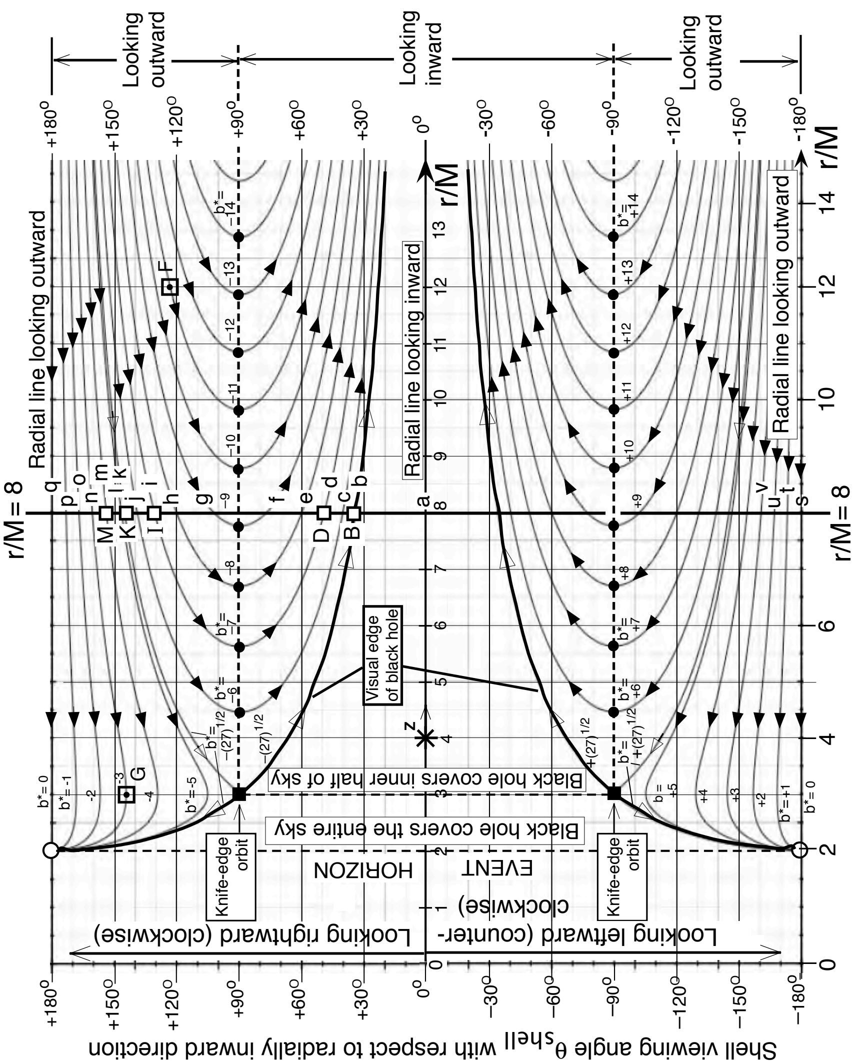


FIGURE 10 Angle θ_{shell} at which the shell observer located at global coordinates $(r, \phi = 0)$ looks to see starlight beam with values of $b^* \equiv b/M$. Arrows on each curve tell us whether that beam is incoming or outgoing. A black dot marks a turning point, the r-coordinate at which an incoming beam reverses its dr to become an outgoing beam. Upper and lower three-branch curves with open arrowheads represent light with impact parameter $b/M = \pm b_{\text{critical}}/M = \pm(27)^{1/2}$. Two little black squares at $r = 3M$ represent circular knife-edge orbits of these critical beams on the tangential light sphere. When viewed by starlight, the shell observer sees the black hole at shell angles inside the heavy curve, which represents its visual edge. The shell observer near the event horizon looks radially outward to see the entire heavens contracted to a narrow cone (Figure 8).

B. Why can't equations (18) and (31) be used inside the event horizon?

C. A shell observer at a given r and a local rain observer who passes through that map location both view the same beam. Items (a) through (c) below give the value of b and r in each case, and whether the beam is incoming or outgoing. For each case, find θ_{rain} from Figure 3; use (31) with v_{rel} from (16) to convert to shell angle θ_{shell}; then check your result in Figure 10.

(a) Outgoing beam with $b/M = -12$ observed at $r = 12M$.

(b) Incoming beam with $b/M = -7$ observed at $r = 6M$.

(c) Incoming beam with $b/M = -4$ observed at $r = 3M$.

D. Look at the list "What the Rain Viewer Sees as She Passes $r = 8M$" in Section 12.3. Use the lowercase bold letters on the $r = 8M$ vertical line in Figure 10 to write a similar analysis of what the local shell observer at $r = 8M$ sees.

3. Expression E_{rain}/E for a stone.

Section 12.6 derives the expression E_{rain}/E for *light.* Derive the same expression for a *stone.*

4. Direction of a star seen by an orbiting observer

In what direction does the observer in circular orbit look to see the same beam? Special relativity can answer this question, because it requires a simple aberration transformation—similar to (31)—first from θ_{rain} to θ_{shell} and then from θ_{shell} to θ_{orbiter}. Equation (16) gives the relative speed in the radial direction between the rain diver and the shell observer, while equation (31) in Section 8.5 gives the relative speed in the tangential direction between shell and orbiting observers: $v_{\text{rel}} = v_{\text{shell}} = (r/M - 2)^{-1/2}$. The resulting transformations, although messy, use nothing but algebra and trigonometry. The results are plotted in Figure 11, which is similar to Figure 3.

A. Sketch a figure similar to Figure 6 for the relative motion of the shell and orbiter observers, including θ_{shell}, θ_{orbiter}, and the arrow for v_{shell}. Adapt equation (31) for your figure and show that the aberration between θ_{shell} and θ_{orbiter} is:

$$\sin\theta_{\text{orbiter}} = \frac{\sin\theta_{\text{shell}} + \left(\dfrac{r}{M} - 2\right)^{-1/2}}{1 + \left(\dfrac{r}{M} - 2\right)^{-1/2} \sin\theta_{\text{shell}}} \quad (\theta = \text{angle viewer looks}) \tag{31}$$

B. Why can't equation (31) be used inside the event horizon, $r < 2M$?

C. Why can't equation (31) even be used for, $r < 3M$?

D. Figures 3 and 20 depict the viewing angle for the rain and shell observers, respectively. Because of cylindrical symmetry, we can use those two figures to create full three-dimensional panoramas for the rain and shell observers, respectively (see explanation in Section 12.3)

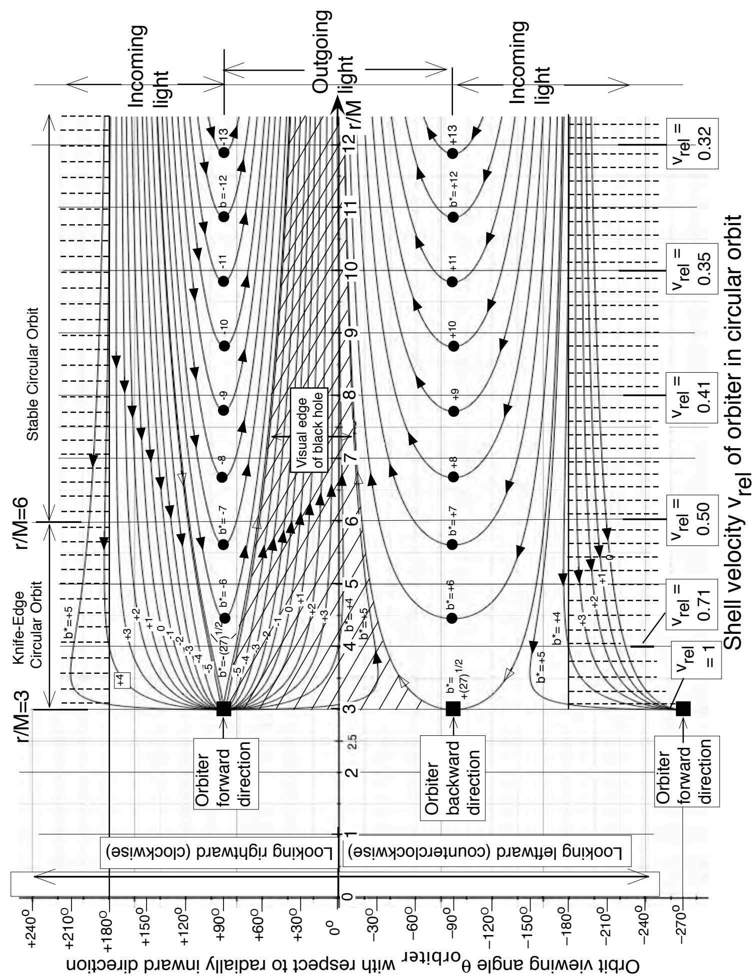


FIGURE 11 Observation angles $\theta_{\rm orbiter}$ at which an orbiter looks to see beams with different impact parameters $b^* = b/M$. The edges of the diagonally shaded region are the orbiter's visual edges of the black hole. Values of $v_{\rm rel}$ along the bottom are the relative velocities of the orbiter with respect to the local shell frame. In the limiting case of the orbit at $r = 3M$ (moving at the speed of light in the shell frame), the inner half of the sky is black for the orbiter. Upper and lower regions shaded by vertical dashed lines include some of the curves between $\theta_{\rm orbiter} = -180°$ and $\theta_{\rm orbiter} = +180°$.

Can we use Figure 11 to similarly create a full three-dimensional panorama for the orbiter?

E. A shell observer at a given r-coordinate and a local orbiter who passes through that location both view the same beam. Items (a) through (d) below give the values of r and $\theta_{\rm shell}$ for the four different cases. For each case, use Figure 10 to determine the value of b^* and whether the beam is incoming or outgoing. Then use Figure 11 to find $\theta_{\rm orbiter}$ for each case. Finally, in each case determine the relative speed $v_{\rm rel}$

between shell orbiter, and use equation (31) to calculate $\sin\theta_{\rm orbiter}$ that you found from Figure 11. Are the calculated values of $\sin\theta_{\rm orbiter}$ in agreement with the values of $\theta_{\rm orbiter}$ that you found from Figure 11?

(a) $r = 5.5M$, $\theta_{\rm shell} = 60^{\rm o}$.
(b) $r = 5.5M$, $\theta_{\rm shell} = 120^{\rm o}$.
(c) $r = 5.5M$, $\theta_{\rm shell} = -73^{\rm o}$.
(d) $r = 5.5M$, $\theta_{\rm shell} = -60^{\rm o}$.

12.9 ■ REFERENCES

Initial quote: Emily Dickinson, poem number 143, version A, about 1860, *The Poems of Emily Dickinson, Variorum Edition*, Edited by R. W. Franklin, Cambridge Massachusetts, The Belknap Press of Harvard University, 1998, Volume I, page 183.

Description of final dive (Section 12.7) and Figure 8 are adapted from Richard Matzner, Tony Rothman, and Bill Unruh, "Grand Illusions: Further Conversations on the Edge of Spacetime," in *Frontiers of Modern Physics: New Perspectives on Cosmology, Relativity, Black Holes and Extraterrestrial Intelligence*, edited by Tony Rothman, Dover Publications, Inc., New York, 1985, pages 69–73. Luc Longtin provided corrections for "times before oblivion" in in the Section The Final Fall and calculated numbers for Figure 8.

Download File Name: Ch12DivingPanoramas190403v1.pdf

Chapter 13. Gravitational Mirages

- *How can Newton's mechanics predict the deflection of light by the Sun?*
- *Does Einstein predict a different value of light deflection than Newton? If so, which prediction do we observe?*
- *Does the amount of deflection depend on the energy/wavelength of the light?*
- *Can a center of gravitational attraction act like a lens? Can it create a mirage?*
- *Can a gravitational lens magnify distant objects?*
- *Can a planet around a distant star act as a mini-gravitational lens?*
- *How can we use light deflection to detect and measure mass that we cannot see?*

CHAPTER 13

Gravitational Mirages

Edmund Bertschinger & Edwin F. Taylor *

Einstein was discussing some problems with me in his study when he suddenly interrupted his explanation and handed me a cable from the windowsill with the words, "This may interest you." It was the news from Eddington [actually from Lorentz] *confirming the deviation of light rays near the sun that had been observed during the eclipse. I exclaimed enthusiastically, "How wonderful, this is almost what you calculated." He was quite unperturbed. "I knew that the theory was correct. Did you doubt it?" When I said, "Of course not, but what would you have said if there had not been such a confirmation?" he retorted, "Then I would have to be sorry for dear God. The theory is correct." ["Da könnt' mir halt der liebe Gott leid tun. Die Theorie stimmt doch."]*

—Ilse Rosenthal-Schneider

13.1 ■ GRAVITY TURNS STARS AND GALAXIES INTO LENSES

Euclid overthrown

1919: Einstein's prediction of light deflection verified.

Arthur Eddington's verification of Albert Einstein's predicted deflection of starlight by the Sun during the solar eclipse of 1919 made Einstein an instant celebrity, because this apparently straightforward observation replaced Newton's two-centuries-old mechanics and Euclid's twenty-two-centuries-old geometry with Einstein's revolutionary new general relativity theory.

Long history of deflection predictions

Einstein's predicted deflection of light by a star has a long history, traced out in the following timeline. His prediction also implied that a gravitating structure can act as a lens, bending rays around its edge to concentrate the light and even to form one or more distorted images. We call the result **gravitational lensing**. Gravitational lensing has grown to become a major tool of modern astronomy.

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

Timeline: Deflection of Starlight

Timeline of starlight deflection

History of predicting, discovering, and employing the deflection of starlight for cosmological research:

1801 Johann George von Soldner makes a Newtonian calculation of the deflection of starlight by the Sun (Section 13.2). He predicts a deflection half as great as Einstein later derives and observation verifies. Soldner's predicted result, 0.84 arcsecond, was not followed up by astronomers. (One arcsecond is 1/3600th of one degree.)

1911 Einstein recalculates Soldner's result (obtaining 0.83 arcsecond) without knowing about Soldner's earlier work.

1914 Einstein moves to Berlin and asks astronomer Erwin Freundlich if the tiny predicted result can be measured. At dinner, Einstein covers Mrs. Freundlich's prize table cloth with equations. She later laments, "Had I kept it unwashed as my husband told me, it would be worth a fortune." Freundlich points out that the measurement Einstein seeks can be made during a total solar eclipse predicted for the Crimea in Russia on August 21, 1914. Freundlich organizes an expedition to that location. World War I breaks out between Germany and Russia on August 1; Freundlich and his team are arrested as spies, so cannot make the observation. (They are quickly exchanged for Russian prisoners; the sky is cloudy anyway.)

1915 In November Einstein applies the space curvature required by his theory (Section 13.5) to recalculate light deflection by the Sun, finds 1.7 arcsecond, double the previous value. A week later Einstein completes the logical structure of his theory of general relativity. (In January 1916 astronomer Karl Schwarzschild, serving as a German artillery officer on the Russian front, finds a solution to Einstein's equations—the Schwarzschild metric—for a spherically symmetric center of attraction.)

1919 Arthur Eddington leads a post-war group to the island of Principe, off the coast of West Africa, to measure starlight deflection during a total eclipse. He reports a result that verifies Einstein's prediction. "Lights All Askew in the Heavens" headlines the *New York Times* (Figure 1). Later observations, using both light and radio waves, validate Einstein's prediction to high accuracy.

1936 R. W. Mandl, a German engineer and amateur astronomer, asks Einstein if the chance alignment of two stars could produce a ring of light from gravitational deflection. Einstein writes it up for the journal *Science* and remarks condescendingly to the editor, "Thank you for your cooperation with the little publication which Herr Mandl squeezed out of me. It is of little value, but it makes the poor guy happy." In the paper Einstein says, "Of course, there is no hope of observing this phenomenon directly."

1937 Fritz Zwicky, controversial Swiss-American astronomer, says that Einstein is wrong, because galaxies can produce observable deflection of light. He also anticipates the use of gravitational lenses to measure the mass of galaxies.

1979 Forty-two years after Zwicky's insight, the first gravitational lens is discovered and analyzed (Figure 14).

NOW: Gravitational lensing becomes a major tool of astronomers and astrophysicists.

Gravitational lens

Gravitational mirage

In the present chapter we re-derive Soldner's expression for deflection of starlight by the Sun, re-derive Einstein's general-relativistic prediction, then apply results to any astronomical object, such as a galaxy, that acts as a **gravitational lens**. This lens deflects rays from a distant source to form a distorted image for an observer on the opposite side of, and distant from, the lensing object. A gravitational lens can yield multiple images, arcs, or rings. It can also magnify distant objects. We adopt the descriptive French term for such a distorted image: **gravitational mirage**. Astronomers use gravitational mirages to study fundamental components of the Universe, such as the presence and abundance of dark matter, and to detect planets orbiting around distant stars.

Several features of applied gravitational lensing simplify our study of gravitational mirages:

Simplifying conditions

1. The source of light is a long way from the deflecting structure.
2. The observer is a long way from (and on the opposite side of) the deflecting structure.
3. Light ray deflection by ordinary stars and galaxies is *very* small. The Sun's maximum deflection of starlight is 1.75 arcsecond = 1.75/3600 degree = 0.000486 degree. Multiple rays may connect emitter and observer, but we are safe in treating every deflection as very small.

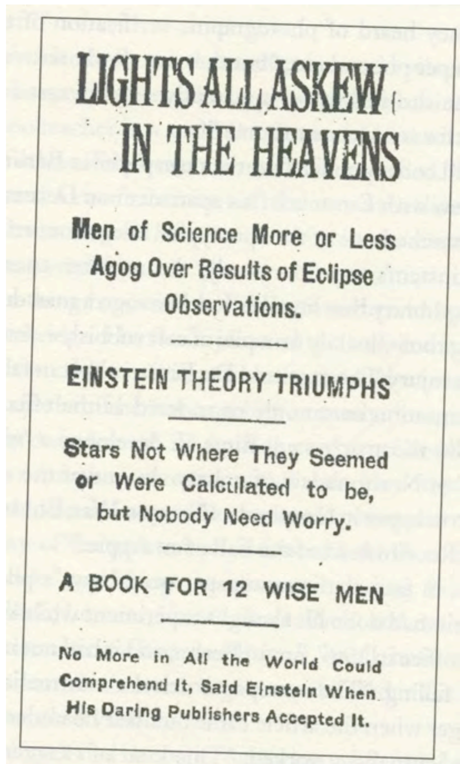
13.2 ■ NEWTONIAN STARLIGHT DEFLECTION (SOLDNER)

Pretend that a stone moves with the speed of light.

Newtonian analysis of light deflection

In 1801 Johann von Soldner extended Newton's particle mechanics to a "particle of light" in order to predict its deflection by the Sun. His basic idea was to treat light as a very fast Newtonian particle. Soldner's analysis yields an incorrect prediction. Why do we repeat an out-of-date Newtonian analysis of light deflection? Because the result highlights the radical revolution that Einstein's theory brought to spacetime (Section 13.5). However, we do not follow Soldner's original analysis, but adapt one by Joshua Winn.

Suppose the particle of light moves in the x direction tangentially past the attracting object (Figure 2). We want to know the y-component of velocity that this "fast Newtonian particle" picks up during its passage. Integrate Newton's second law to determine the change in y-momentum. We use



LIGHTS ALL ASKEW IN THE HEAVENS

Men of Science More or Less Agog Over Results of Eclipse Observations.

EINSTEIN THEORY TRIUMPHS

Stars Not Where They Seemed or Were Calculated to be, but Nobody Need Worry.

A BOOK FOR 12 WISE MEN

No More in All the World Could Comprehend It, Said Einstein When His Daring Publishers Accepted It.

FIGURE 1 Headline in the *New York Times* November 10, 1919. The phrase "12 wise men" refers to the total number of people reported to understand general relativity at that time.

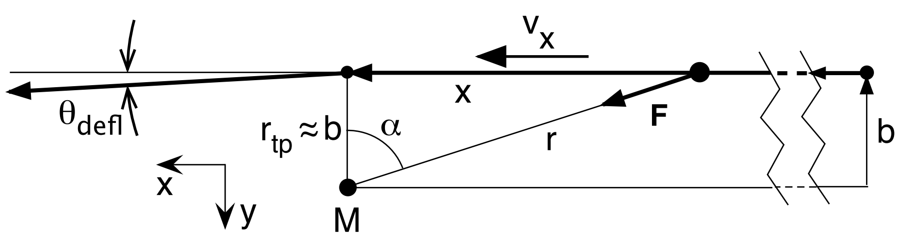


FIGURE 2 Symbols for the Newtonian calculation of the deflection of light that treats a photon as a very fast particle. This approximation assumes that the deflection is very small and occurs suddenly at the turning point $r_{\rm tp}$. Not to scale.

conventional units in order to include the speed of light c explicitly in the analysis.

$$\int_{-\infty}^{+\infty} F_{\rm y} dt = \Delta p_{\rm y} = m_{\rm kg} \Delta v_{\rm y} \qquad \text{(Newton)} \qquad (1)$$

The y-component of the gravitational force on the particle is:

$$F_{\mathrm{y}} = F\cos\alpha = \frac{GM_{\mathrm{kg}}m_{\mathrm{kg}}}{r^2}\cos\alpha \qquad \text{(Newton, conventional units)} \qquad (2)$$

Assume that the speed of the "particle" is the speed of light: $v = (v_{\mathrm{x}}^2 + v_{\mathrm{y}}^2)^{1/2} = c$. We expect the deflection to be extremely small, $v_{\mathrm{y}} \ll c$, so take $v_{\mathrm{x}} \approx c$ to be constant during the deflection. From Figure 2:

$$\frac{b}{r} = \cos\alpha \quad \text{so that} \quad \frac{1}{r^2} = \frac{\cos^2\alpha}{b^2} \quad \text{and} \qquad \text{(Newton)} \qquad (3)$$

$$dt = \frac{dx}{v_{\mathrm{x}}} \approx \frac{dx}{c} \quad \text{and} \quad x = b\tan\alpha \quad \text{so} \quad dx = \frac{b\,d\alpha}{\cos^2\alpha} \qquad \text{(Newton)} \qquad (4)$$

As the particle flies past the center of attraction, the angle α swings from $-\pi/2$ to (slightly more than) $+\pi/2$. Substitute from equations (2) through (4) into (1), cancel the "Newtonian photon mass m_{kg}" from both sides of the resulting equation, and integrate the result:

$$\frac{GM_{\mathrm{kg}}}{bc}\int_{-\pi/2}^{+\pi/2}\cos\alpha\,d\alpha = \frac{2GM_{\mathrm{kg}}}{bc} = \Delta v_{\mathrm{y}} \quad \text{(Newton, conventional units)} \quad (5)$$

The integral in (5) has the value 2. Because the deflection angle θ_{defl} is very small, we write:

$$\theta_{\mathrm{defl}} \approx \frac{\Delta v_{\mathrm{y}}}{c} = \frac{2}{b}\left(\frac{GM_{\mathrm{kg}}}{c^2}\right) \to \frac{2M}{b} \approx \frac{2M}{r_{\mathrm{tp}}} \qquad \text{(Newton)} \qquad (6)$$

The next-to-last step in (6) reintroduces mass in units of meters from equation (10) in Section 3.2. The last step, which equates the turning point r_{tp} with impact parameter b, follows from the tiny value of the deflection, as spelled out in equation (12).

Apply (6) to deflection by the Sun. The smallest possible turning point r_{tp} is the Sun's radius, leading to a maximum deflection:

$$\theta_{\mathrm{defl,max}} \approx \frac{2M}{r_{\mathrm{Sun}}} = 4.25\times10^{-6}\,\text{radian} = 2.44\times10^{-4}\,\text{degree} \quad \text{(Newton)} \quad (7)$$

Soldner's predicted deflection

Multiply this result by 3600 arcseconds/degree to find the maximum deflection $\theta_{\mathrm{defl,max}} \approx 0.877$ arcsecond. (Soldner predicted 0.84 arcsecond.) Section 13.4 shows that the correct prediction—verified by observation—is twice as large: $\theta_{\mathrm{defl,max}} \approx 4M/r_{\mathrm{Sun}} = 1.75$ arcseconds. (Einstein predicted 1.7 arcseconds.) All of these predicted and observed deflection angles are much, much smaller than even the tiny angle θ_{defl} shown in Figure 2.

Both Soldner's result (6) and Einstein's—equation (18) in Box 1—tell us that the deflection angle is inversely proportional to the turning point r_{tp}. That is, maximum bending occurs for a ray that passes closest to the deflecting object (right panel, Figure 3). Contrast this with the conventional glass *optical*

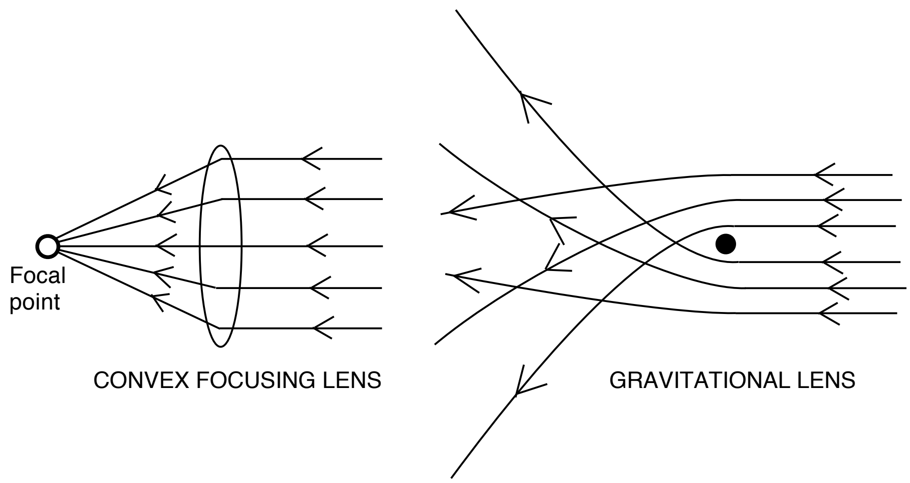


FIGURE 3 Schematic comparison of deflection of light by glass lens and gravitational lens. **Left panel:** In a conventional convex glass focusing lens, deflection increases farther from the center in such as way that incoming parallel rays converge to a *focal point.* **Right panel:** Deflection by a gravitational lens is greatest for rays that pass closest to the center—equation (6). *Result:* no focal point, which guarantees image distortion by a gravitational lens.

A gravitational lens *must* distort the image.

focusing lens, such as the lens in a magnifying glass, whose edge bends light more than its center (left panel, Figure 3). Parallel rays incident on a good optical lens converge to a single point, called the *focal point.* The existence of a focal point leads to an undistorted image. A gravitational lens—with maximum deflection for the closest ray—does not have a focal point, which results in image distortion. The base of a stem wineglass acts similarly to a gravitational lens, and shows some of the same distortions (Figure 4).

QUERY 1. Quick Newtonian analysis

Show that you obtain the same Newtonian result (6) if you assume (Figure 5) that transverse acceleration takes place only across a portion of the trajectory equal to twice the turning point and that this transverse acceleration is uniform downward and equal to the acceleration at the turning point.

13.3 ■ LIGHT DEFLECTION ACCORDING TO EINSTEIN AFTER 1915

In from infinity, out to infinity

Analysis of large deflection

Now we analyze the deflection of starlight by a center of attraction predicted by general relativity. Any incoming light with impact parameter $|b| > b_{\mathrm{critical}}$ does not cross the event horizon, but rather escapes to infinity (Figure 3 in Section 12.3). Figure 6 shows resulting rays from the star lying at map angle $\phi_\infty = 0$ with positive impact parameter b. This light approaches the black hole

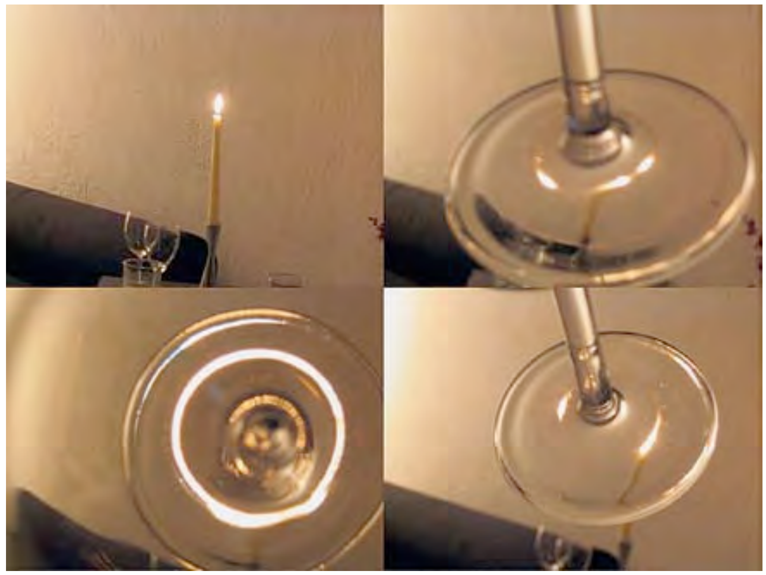


FIGURE 4 The base of a stem wine glass has lensing properties similar to that of a gravitational lens. The source of light, at the top left, is a candle. Tilting the wine glass base at different angles with respect to the source produces multiple images similar to those seen in gravitational mirages. The bottom right panel shows a double image and the top right four images. The bottom left—looking down the optical axis of the wine glass—shows an analogy to the full *Einstein ring* (Figure 12). Images courtesy of Phil Marshall.

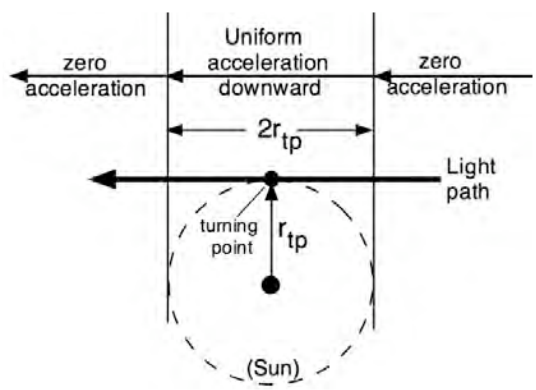


FIGURE 5 Figure for Query 1: Alternative derivation of Newtonian light deflection. Dashed circle: outline of our Sun, with light ray skimming past its edge, so that its turning point r_{tp} equals the Sun's radius r_{Sun}.

from a distant source, deflects near the black hole, then recedes from the black hole to be seen by a distant observer. The ray is symmetric on the two sides of the turning point, so the total change in direction along the ray is twice the

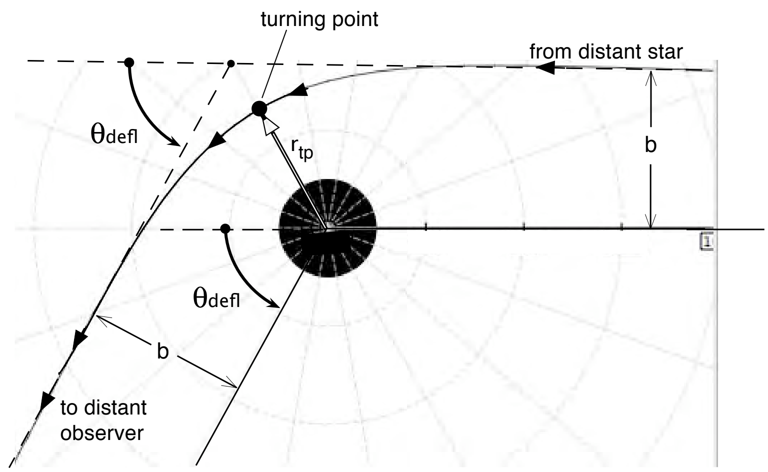


FIGURE 6 Total deflection θ_{defl} of a ray with impact parameter b that originates at a distant star. This ray deflects near the center of attraction, then runs outward to a distant observer. Deflection θ_{defl}, is the *change* in direction of motion of the flash. As usual, the positive direction of rotation is counterclockwise.

change in direction between the turning point and either the distant source or the distant observer.

Reach back into Chapter 11 for the relevant equations. Start with equation (40) of Section 11.6, which applies to an observer *after* the turning point. In the present situation, the map angle of the distant star is $\phi_\infty = 0$, the observer is far from the center of attraction, so $r_{\text{obs}} \to \infty$, and from the definition of the deflection angle, $\phi_{\text{obs}} = \pi + \theta_{\text{defl}}$. Substitute these into equation (40) of Section 11.6 to obtain:

$$\theta_{\text{defl}} + \pi = 2\int_{r_{\text{tp}}}^{\infty} \frac{b}{r^2}\left[1 - \frac{b^2}{r^2}\left(1 - \frac{2M}{r}\right)\right]^{-1/2} dr \qquad (8)$$

The integrand in (8) is a function of b, and the lower integration limit is r_{tp}. To remove these complications, make the substitution

$$u \equiv \frac{r_{\text{tp}}}{r} \quad \text{so that} \quad dr = -\frac{r_{\text{tp}}}{u^2}du \qquad (9)$$

where r_{tp} and b are constants. The variable u has the value $u = 0$ at the distant star and the value $u = 1$ at the turning point. With substitutions (9), equation (8) becomes:

$$\theta_{\text{defl}} + \pi = \frac{2b}{r_{\text{tp}}}\int_0^1 \left[1 - \left(\frac{b}{r_{\text{tp}}}\right)^2 u^2\left(1 - \frac{2Mu}{r_{\text{tp}}}\right)\right]^{-1/2} du \qquad (10)$$

Both b and r_{tp} are parameters (constants) in the integrand of (10). Use equation (27) in Section 11.4 to convert r_{tp} to b. Then b is the only parameter

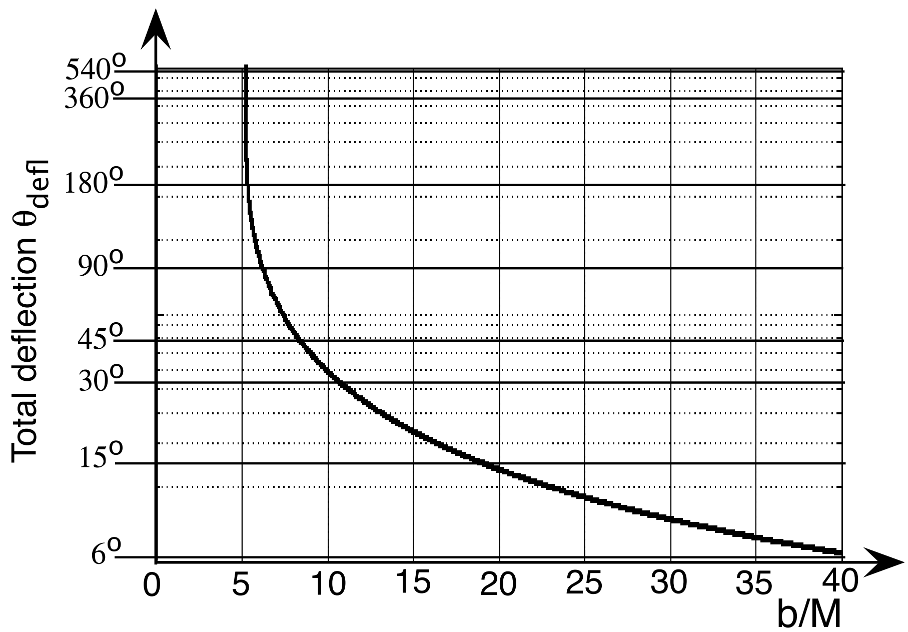


FIGURE 7 Map deflection θ_{defl} as a function of positive values of b from the numerical integration of (10). The vertical scale is logarithmic, which allows display of both small and large values of deflection.

in (10). Figure 7 plots results of a numerical integration for positive values of b. The magnitude of θ_{defl} covers a wide range; the semi-log plot makes it easier to read both small and large values of this angle.

Comment 1. Infinite deflection?
Why does the total deflection in Figure 7 appear to increase without limit as the impact parameter b drops to a value close to five? In answer, look at Figure 1 in Section 11.2. When the impact parameter of an approaching ray takes on the value $b_{\text{critical}} = 3(3)^{1/2} = 5.196$, then the ray goes into a knife-edge orbit at $r = 3M$. In effect the deflection angle becomes infinite, which is consistent with the plot in Figure 7.

13.4 ■ LIGHT DEFLECTION THROUGH SMALL ANGLES

Einstein's prediction

Simplification for small deflection

Equation (10) becomes much simpler when the deflection is very small, that is when r_{tp} of the turning point is much larger than $2M$. When a ray passes our Sun, for example, the r-value of its turning point must be greater than or equal to the Sun's radius r_{Sun} if the ray is to make it past the Sun at all. Box 1 approximates the deflection of a light ray with turning point $r_{\text{tp}} \gg 2M$.

Box 1. Starlight Deflection: Small-Angle Approximation

We seek an approximate expression for deflection θ_{defl} in equation (10) when a ray passes sufficiently far from the center of attraction to satisfy the condition:

$$\frac{2M}{r_{\text{tp}}} \equiv \epsilon \ll 1 \tag{11}$$

Equation (7) reminds us that the maximum deflection by our Sun is very small: $\epsilon_{\text{Sun}} = 4.253 \times 10^{-6}$ radian. In the following we make repeated use of the first order approximation inside the front cover. From (11) plus equation (27) in Section 11.4 for the turning point,

$$\frac{b}{r_{\text{tp}}} = \left(1 - \frac{2M}{r_{\text{tp}}}\right)^{-1/2} = (1-\epsilon)^{-1/2} \approx 1 + \frac{\epsilon}{2} \tag{12}$$

Then approximate one expression in the integrand of (10) as:

$$\begin{aligned} &\left(\frac{b}{r_{\text{tp}}}\right)^2 u^2 \left(1 - \frac{2Mu}{r_{\text{tp}}}\right) \qquad (13)\\ &\approx \left(1 + \frac{\epsilon}{2}\right)^2 u^2 \left(1 - u\epsilon\right)\\ &\approx (1+\epsilon)u^2(1-u\epsilon)\\ &\approx u^2 + (1-u)u^2\epsilon \end{aligned}$$

At each step we neglect terms in ϵ^2. With this substitution, the square bracket expression in (10) becomes

$$\begin{aligned} &\left[1 - u^2 - (1-u)u^2\epsilon\right]^{-1/2} \qquad (14)\\ &= (1-u^2)^{-1/2}\left[1 - \frac{(1-u)}{1-u^2}u^2\epsilon\right]^{-1/2}\\ &= (1-u^2)^{-1/2}\left[1 - \frac{u^2\epsilon}{1+u}\right]^{-1/2}\\ &\approx \frac{1}{(1-u^2)^{1/2}}\left[1 + \frac{u^2\epsilon}{2(1+u)}\right] \end{aligned}$$

The coefficient of the integral in (10) is $2b/r_{\text{tp}} = 2 + \epsilon$ from (12). Combine this with the last line of (14) to find the expressions that we want to integrate, again to first order in ϵ.

$$\frac{2+\epsilon}{(1-u^2)^{1/2}} + \frac{u^2\epsilon}{(1+u)(1-u^2)^{1/2}} \tag{15}$$

From a table of integrals:

$$\int_0^1 \frac{du}{(1-u^2)^{1/2}} = \frac{\pi}{2} \tag{16}$$

This occurs twice in (15), one integral multiplied by 2, the other by ϵ. The integral of the second term in (15) becomes:

$$\int_0^1 \frac{u^2 du}{(1+u)(1-u^2)^{1/2}} = 2 - \frac{\pi}{2} \tag{17}$$

In (15) this is multiplied by ϵ. Combine the results of (15) through (17) to write down the expression for θ_{defl} in (10), not forgetting the term π on the left side, which cancels an equal term on the right side:

$$\theta_{\text{defl}} \approx 2\epsilon \approx \frac{4M}{r_{\text{tp}}} \ll 1 \tag{18}$$

where r_{tp} is the turning point, the r-value of closest approach. Equation (18) is Einstein's general relativistic prediction for deflection when (11) is satisfied. Section 13.2 showed that Soldner's Newtonian analysis predicts a value of the light deflection half as great as (18).

Important note: Our derivation of (18) assumes that the deflecting structure is either (a) a point—or a spherically symmetric object—with $r_{\text{tp}} \geq$ the r-value of this structure, or (b) a black hole approached by light with impact parameter $b/M \gg 1$. It is not valid when the the light passes close to a non-spherical body, like a galaxy.

Deflection when $R \gg M$

From the result of Box 1 we predict that the largest deflection of starlight by the Sun occurs when the light ray skims past the edge of the Sun. From (18), this maximum deflection is:

$$\begin{aligned} \theta_{\text{defl,max}} = \frac{4M}{r_{\text{Sun}}} &= 8.49 \times 10^{-6} \text{ radian} \qquad (19)\\ &= 4.87 \times 10^{-4} \text{ degree}\\ &= 1.75 \text{ arcsecond} \end{aligned}$$

QUERY 2. Value of r_{tp}/M for various deflections

Compute eight values of the turning point $r_{\rm tp}$, namely the four deflection angles of Items A through D below for each of two cases. *Case I:* The mass of a star like our Sun. *Case II:* The total mass of the visible stars in a galaxy, approximately 10^{11} Sun masses.

A. $\theta_{\rm defl} \approx$ one degree
B. $\theta_{\rm defl} \approx$ one arcsecond
C. $\theta_{\rm defl} \approx 10^{-3}$ arcsecond
D. $\theta_{\rm defl} \approx 10^{-6}$ arcsecond
E. For Case I, compare the values of $r_{\rm tp}$ with the r-value of the Sun.
F. For Case II, compare the values of $r_{\rm tp}$ with the r-value of a typical galaxy, 30 000 light years.
G. Which cases can occur that lead to a deflection of one arcsecond? 10 arcseconds? Are these turning points inside the radius of a typical galaxy? If so, we cannot correctly use deflection equation (18), which was derived for either a point lens or for a turning point $r_{\rm tp}$ outside a spherically symmetric lens.

Starting with Section 13.6, the remainder of this chapter describes multiple uses of the single deflection equation (18) in astronomy, astrophysics, and cosmology. But first we take a detour to outline the profound revolution that Einstein's prediction of the Sun's deflection of starlight made in our understanding of spacetime geometry.

13.5 ■ DETOUR: EINSTEIN DISCOVERS SPACE CURVATURE

Einstein discovers a factor of two and topples Euclid.

Einstein's initial error in light deflection

Here we take a detour to show how Einstein, in effect, used an incomplete version of the Schwarzschild metric to make an incorrect prediction of the Sun's gravitational deflection of starlight, a prediction that he himself corrected before observation could prove him wrong. His original misconception was to pay attention to spacetime curvature embodied in the t-coordinate, but not to realize at first that the r-coordinate is also affected by spacetime curvature.

Global gravitational potential

In 1911, as he developed general relativity, Einstein predicted the deflection of starlight that reaches us by passing close to the Sun. Einstein recognized that in Newtonian mechanics the expression $-M/r$ is potential energy per unit mass, called the **gravitational potential**, symbolized by Φ. Einstein's initial metric generalized the gravitational potential of Newton around a point mass M:

$$\Phi = -\frac{M}{r} \qquad \text{(Newton)} \qquad (20)$$

Einstein initially warped only t-coordinate.

Einstein's 1911 analysis was equivalent to adopting a global metric that we now recognize as incomplete:

$$d\tau^2 = (1+2\Phi)\,dt^2 - dr^2 - r^2 d\phi^2 = \left(1 - \frac{2M}{r}\right) dt^2 - ds^2 \qquad (21)$$

("half-way to Schwarzschild")

"Half way to Schwarzschild"

The r, ϕ part of metric (21) is flat, as witnessed by the Euclidean expression $ds^2 = dr^2 + r^2 d\phi^2$. In contrast, the r-dependent coefficient of dt^2 shows that the t-coordinate has a position-dependent warpage. Thus metric (21) is "half way to Schwarzschild," even though in 1911 Einstein did not yet appreciate the centrality of the metric, and the derivation of the Schwarzschild metric was almost five years in the future.

For light, set $d\tau = 0$ in (21), which then predicts that the map speed of light decreases as it approaches the Sun:

$$\left|\frac{ds}{dt}\right|_{\text{light}} = \left(1 - \frac{2M}{r}\right)^{1/2} \approx 1 - \frac{M}{r} \quad (M/r \ll 1) \quad \text{(Einstein 1911)} \qquad (22)$$

Equation (22) reduces the problem of light deflection to the following exercise in geometric optics: "Light passes through a medium in which its speed varies with position according to equation (22). Use Fermat's Principle to find a light ray that grazes the edge of the Sun as it travels between a distant source and a distant observer on opposite sides of the Sun."

Fermat's principle

Fermat's Principle, derived from standard classical electromagnetic theory of light, says that light moves along a trajectory that minimizes the total time of transit from source to observer. (This is classical physics, so space is flat, as (21) assumes.) Einstein used Fermat's Principle—geometric optics plus equation (22)—to calculate a deflection, $\theta_{\text{defl}} = 2M/r_{\text{tp}}$, equal to half of the observed value.

Gravity also warps r-coordinate.

In his initial prediction, however, Einstein failed to understand that gravity also curves the r, ϕ part of spacetime near a center of attraction. We can now see this initial error and correct it ourselves. The Schwarzschild metric (3.5) shows that the r, ϕ portion of the metric is not flat; the term dr^2 in (21) should be $dr^2/(1 - 2M/r)$. The difference arises from the contribution of the $dr, d\phi$ components to curvature.

We can derive the radial component of map light velocity from the correct Schwarzschild metric:

$$\left|\frac{dr}{dt}\right|_{\text{light}} = 1 - \frac{2M}{r} \qquad \text{(radial motion, Schwarzschild)} \qquad (23)$$

Radial motion of light further slowed by space curvature.

Compare the results of equations (22) and (23). Light that grazes the surface of the Sun in its trajectory between a distant star and our eye travels *almost* radially during its approach to and recession from the Sun. Fermat's Principle still applies, but the angle of deflection predicted from the change in coordinate speed of light in (23) is twice that of the preliminary prediction derived from (22).

Profound insight: Space also curved

Einstein's realization that the r, ϕ part of global coordinates must be curved, along with the t part, was a profound shift in understanding, from which his field equations emerged. Einstein's doubled prediction of light deflection was tested by Eddington, and the currently-predicted value has since been validated to high accuracy.

Most accurate deflection results from radio astronomy

Radio astronomy, which uses radio waves instead of visible light, provides much more accurate results than the deflection of starlight observed by optical telescopes. Each October the Sun moves across the image of the quasar labeled 3C279 seen from Earth. Radio astronomers use this so-called **occultation** to measure the change in direction of the signal as—from our viewpoint on Earth—the source approaches the Sun, crosses the edge of the Sun, and moves behind the Sun. They employ an experimental technique called **very long baseline interferometry** (VLBI) that effectively uses two or more widely separated antennas as if they were a single antenna. This wide separation substantially increases the accuracy of observation. VLBI observations by E. Fomalont and collaborators measure a gravitational deflection to be a factor 1.9998 ± 0.0006 times the Newtonian prediction, in agreement with general relativity's prediction of 2 times the Newtonian result.

Since 1919, the gravitational deflection of light has become a powerful observational tool, as described in the remainder of this chapter.

PUTTING EINSTEIN TO THE TEST

No matter how revolutionary it was, no matter how beautiful its structure, our guide had to be experimentation. Equipped with new measuring tools provided by the technological revolution of the last twenty-five years, we put Einstein's theory to the test. What we found was that it bent and delayed light just right, it advanced Mercury's perihelion at just the observed rate, it made the Earth and the Moon fall the same [toward the Sun], and it caused a binary system to lose energy to gravitational waves at precisely the right rate. What I find truly amazing is that this theory of general relativity, invented almost out of pure thought, guided only by the principle of equivalence and by Einstein's imagination, not by need to account for experimental data, turned out in the end to be so right.

—Clifford M. Will

Galaxies not spherically symmetric.

Galaxies are not generally spherically symmetric, and light that passes through or close to a galaxy is *not* deflected according to Einstein's result (18). Light that passes *far* from a galaxy is deflected by an amount that can be *approximated* by Einstein's equation.

13.6 ■ GRAVITATIONAL MIRAGES

Derive and apply the gravitational lens equation.

Many rays can form an image.

So far this chapter has analyzed the deflection of a *single* ray of light by a point mass or a spherically symmetric center of attraction. But our view of the

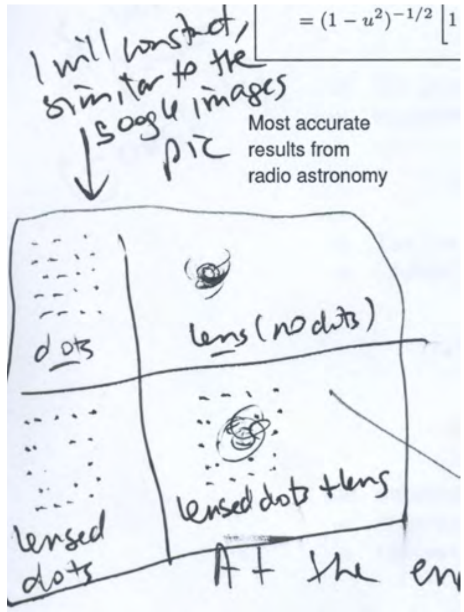


FIGURE 8 Prediction: A square array of distant light sources (upper left panel) is imaged by a galaxy (upper right panel) acting as a gravitational lens. The result is a distorted image of the square array (lower left panel). The lower right panel superposes gravitational lens and distorted image of the square array. [EB will provide final figure.]

Gravitational image: *always* distorted.

heavens is composed of *many* rays. Many rays from a single source—taken together—can form an image of that source. We now examine gravitational lensing, the imaging properties of a spherically symmetric center of attraction. We already know that a gravitational lens differs radically from a conventional focusing lens (Figure 3); this difference leads to a distorted image we call a *gravitational mirage*. Figure 8 previews the image of a square array of distant sources produced by a galaxy that lies between those sources and us.

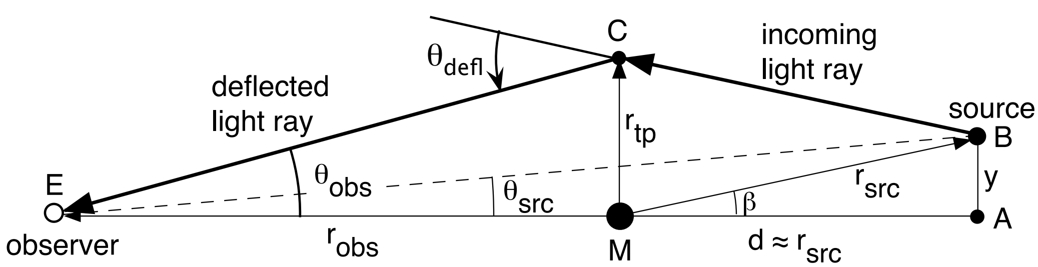


FIGURE 9 Construction for derivation of the gravitational lens equation in Box 2. View is far from a small deflecting mass M, so Euclidean geometry is valid. We assume that all angles are small and light deflection takes place at the turning point, $r_{\rm tp} \gg M$. Not to scale.

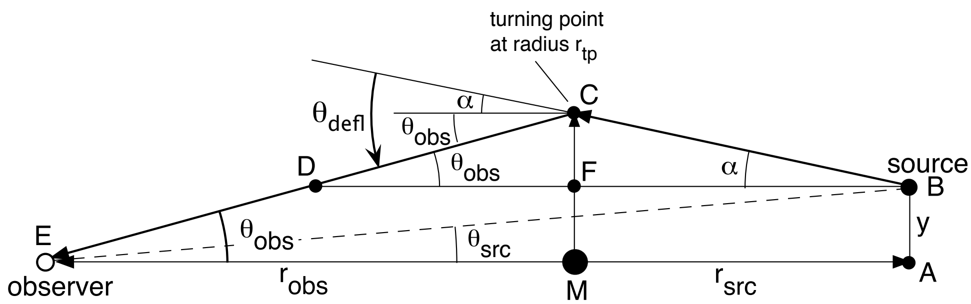


FIGURE 10 Added construction for derivation of the gravitational lens equation in Box 2. Not to scale.

Box 2. Gravitational Lens Equation

Here we derive the lens equation for a ray passing outside a spherically symmetric center of attraction of mass M. We put into this equation the angle $\theta_{\rm obs}$ at which the observer sees the source in the presence of an intermediate gravitational lens and it tells us the angle $\theta_{\rm src}$ at which the observer would see the source in the absence of that lens. The equation also contains the radial coordinate separations $r_{\rm obs}$ and $r_{\rm src}$ of observer and source from the lens.

Use the notation and construction lines in Figures 9 and 10. Assume that spacetime is flat enough so that we can use Euclidean space geometry (except in the immediate vicinity of the point mass M), assume that deflection takes place at the single turning point $r_{\rm tp}$, and finally assume that all angles are extremely small. By "extremely small," we mean, for instance, that in Figure 9, $\beta \approx y/r_{\rm src} \ll 1$. This means that the Euclidean distance d has the value

$$d = r_{\rm src}\cos\beta \approx r_{\rm src}(1-\beta^2/2) \approx r_{\rm src} \qquad (24)$$

to first order in β. Similarly, we can approximate $\sin\theta_{\rm src} \approx \theta_{\rm src}$ and $\sin\theta_{\rm obs} \approx \theta_{\rm obs}$. From triangle ABE in Figure 9

$$y = \theta_{\rm src}\left(r_{\rm obs} + r_{\rm src}\right) \qquad (25)$$

From triangle CEM in that figure:

$$r_{\rm tp} = \theta_{\rm obs} r_{\rm obs} \qquad (26)$$

In both Figures 9 and 10, the radial separation between C and M is $r_{\rm tp}$. In Figure 10, define $\overline{CF}$ as the coordinate separation between C and F, and use (26):

$$y = r_{\rm tp} - \overline{CF} = \theta_{\rm obs} r_{\rm obs} - \alpha r_{\rm src} \qquad (27)$$

We need to find an expression for the angle α. From triangle BCD in Figure 10 and the angles to the left of point C:

$$\alpha = \theta_{\rm defl} - \theta_{\rm obs} \qquad (28)$$

Equate the two expressions for y in (25) and (27) and substitute for α from (28)

$$\theta_{\rm src}\left(r_{\rm src} + r_{\rm obs}\right) = \theta_{\rm obs} r_{\rm obs} + \left(\theta_{\rm obs} - \theta_{\rm defl}\right) r_{\rm src} \qquad (29)$$

From (18) and (26):

$$\theta_{\rm defl} = \frac{4M}{r_{\rm obs}\theta_{\rm obs}} \qquad (30)$$

Substitute (30) into (29) and solve for $\theta_{\rm src}$:

$$\theta_{\rm src} = \theta_{\rm obs} - \frac{4M r_{\rm src}}{\theta_{\rm obs} r_{\rm obs}(r_{\rm src} + r_{\rm obs})} \quad \text{(lens equation)} \qquad (31)$$

This is called the **lens equation** for a point mass M. The lens equation takes a simple form when expressed in terms of the Einstein angle $\theta_{\rm E}$ defined in equation (33):

$$\theta_{\rm src} = \theta_{\rm obs} - \frac{\theta_{\rm E}^2}{\theta_{\rm obs}} \quad \text{(lens equation)} \qquad (32)$$

In geometrical optics, the **lens equation** predicts the path of every ray that passes through a lens. For a spherically symmetric center of attraction—note this restriction!—every single ray that forms a gravitational image obeys Einstein's simple deflection equation (18). The present section uses this result to derive the **gravitational lens equation**. We assume that

Gravitational lens equation

the observer is in flat interstellar space far from the lensing structure, so his frame is inertial. (An observer on Earth is sufficiently inertial for this purpose; Earth's atmosphere distorts incoming starlight far more than does Earth's gravitational deflection of light.)

The **source**

In practice, the source of light may be any radiant object, including all or part of a galaxy. So instead of "star" or "galaxy," we simply call this emitting object the **source** (subscript: **src**). The purpose of the gravitational lens equation is to find the unknown (not measured) angle $\theta_{\rm src}$ from the angle $\theta_{\rm obs}$ at which the observer sees the source. Box 2 carries out this derivation.

Einstein ring

When the source is exactly behind the imaging center of attraction—in other words when $y = 0$ and $\theta_{\rm src} = 0$ in Figures 9 and 10—then the deflection is identical on all sides of the lens, so the observer's image of the source is a ring, called the **Einstein ring** (lower left panel in Figure 4, Figures 11 and 12). In this case the observation angle $\theta_{\rm obs}$ takes the name **Einstein ring angle**, whose square is:

$$\theta_{\rm E}^2 \equiv \frac{4Mr_{\rm src}}{r_{\rm obs}(r_{\rm src}+r_{\rm obs})} \qquad \text{(Einstein ring angle)} \qquad (33)$$

$$\equiv \frac{4GM_{\rm kg}r_{\rm src}}{r_{\rm obs}c^2(r_{\rm src}+r_{\rm obs})} \qquad \text{(Einstein ring angle, conventional units)} \qquad (34)$$

where equation (34) employs conventional units. Definition (33) simplifies expression (31) for $\theta_{\rm src}$ in Box 2, leading to (32).

?

Objection 1. *Whoa! The caption to Figure 12 claims that the farthest source galaxy lies 22 billion light years distant. How can this be? Isn't the Universe less than 14 billion years old?*

!

Early in this book we emphasized that global coordinate separations have no dependable relation to measured quantities. Recall Section 2.7 and the

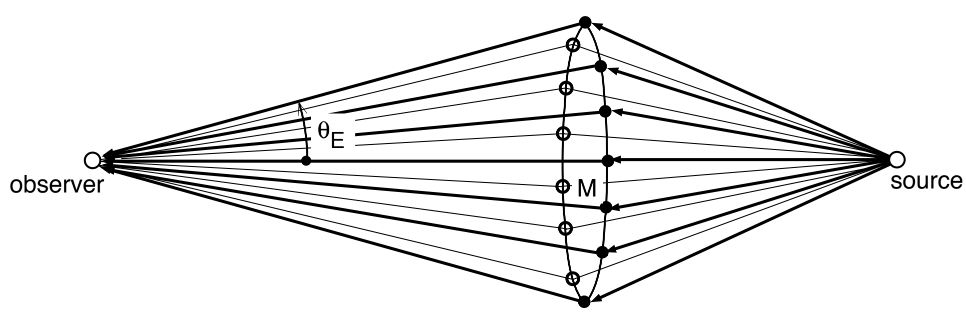


FIGURE 11 When the source, lens, and observer line up, then the deflection angle is the same on all sides of the lens, leading to the *Einstein ring*, whose observation angle $\theta_{\rm E}$ is given by equations (33) and (34). Note that fat rays deflect from one edge of the lens (solid dots) and narrow rays from the other side (little open circles). Not to scale.

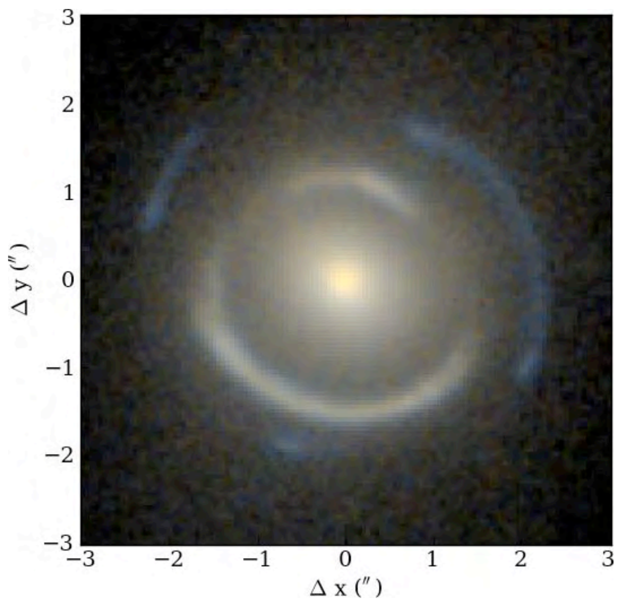


FIGURE 12 Two concentric Einstein rings that arise from two distant galaxy sources directly behind a foreground massive galaxy lens labeled SDSS J0946+1006. The horizontal and vertical scales are marked in units of 1" = one arcsecond. Redshifts of light from the three galaxies lead to the following estimates of their model distances from Earth: lens galaxy, 3.0 billion light years; nearest source galaxy, 7.4 billion light years; farthest source galaxy, 22 billion light years.

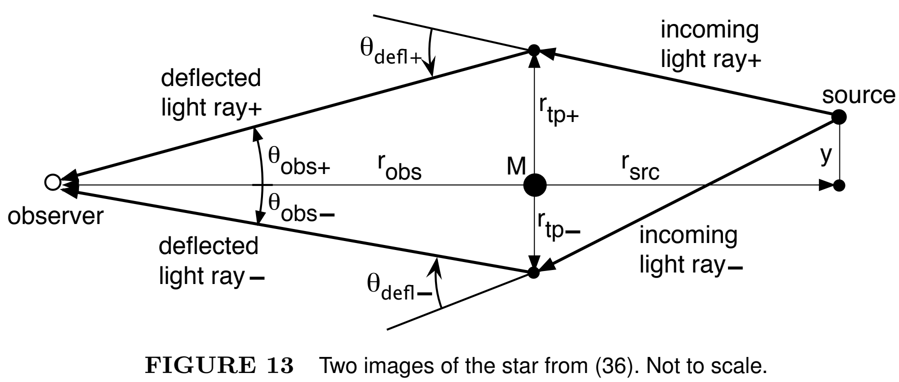


FIGURE 13 Two images of the star from (36). Not to scale.

First Strong Advice for this Entire Book (Section 5.6): "To be safe, it is best to assume that global coordinates never have any measurable meaning. Use global coordinates only with the metric in hand to convert a mapmaker's fantasy into a surveyor's reality." Chapter 15 shows how a cosmic metric translates coordinate differences into observable quantities like redshift—and also into estimated "distances" that depend on a model, such as those quoted in the captions of Figures 12 and 14.

QUERY 3. Use Einstein ring angle to measure the mass of a lensing galaxy.

A. Use the horizontal or vertical axis label in Figure 12 to make yourself a ruler in units of arcseconds. With this ruler measure the average angular radii of the two Einstein rings. From these average radii and the source and lens model distances given in the caption, calculate two independent estimates of the mass of the lensing galaxy in units of the mass of our Sun. Do the results agree to one significant figure?

B. Why does the lens equation (18) work for these Einstein rings, despite the warnings at the end of Box 1 and in the final paragraph of Section 13.5 that Einstein's deflection equation works only for a spherically symmetric lens? Does the image of the lensing structure in the center of Figure 12 give you a hint?

C. Figure 12 demonstrates that the Einstein ring angle θ_E created by a galaxy lens is observable with modern technology, as Zwicky predicted in 1937 (Section 13.1). In contrast, the Einstein ring angle for a star lens is too small to observe, as Einstein predicted in 1936. To demonstrate this, calculate the Einstein ring angle lensed by a star with the mass of our Sun located at the center of our galaxy 26 000 light years distant, with the source twice as far away. Order of magnitude result: $\theta_E \sim 10^{-3}$ arcsecond. Give your answer to one significant digit. [My answer: 7.2×10^{-4}.] Would you expect to see this image as an Einstein ring, given the resolution of Figure 12?

Equation (32) leads to the following quadratic equation in θ_{obs}:

$$\theta_{obs}^2 - \theta_{src}\theta_{obs} - \theta_E^2 = 0 \tag{35}$$

Two images

This equation has two solutions which correspond to two images of the source in the x, y plane of Figure 13:

$$\theta_{obs\pm} = \frac{\theta_{src}}{2} \pm \frac{1}{2}\left(\theta_{src}^2 + 4\theta_E^2\right)^{1/2} \tag{36}$$

QUERY 4. Use double images to measure lensing galaxy mass. Estimate of the mass of the gravitational lens in Figure 14, as follows:

A. Measure the angular separation $|\theta_{obs+} - \theta_{obs-}|$ of the two images in arcseconds.

B. Measure the angular separation $|\theta_{obs-}|$ of the lensing galaxy and the quasar image closest to it in arcseconds.

C. Use equation (36) to determine the separate values of θ_{src} and θ_E.

D. From your value of θ_E and model distances given in the caption of Figure 14, deduce the mass of the lensing galaxy in units of solar mass to one significant figure. Compare this result with the mass of the lensing galaxy in the system SDSS J0946+1006 of Figure 12 that you calculated in Query 3.
Note: The mass found here is only approximate, because the lens is not spherically symmetric. More complex modeling finds a lens mass $(3.9 \pm 1.2) \times 10^{14} M_{Sun}$ (Kundić et al in the references).

Comment 2. Local time delay between images
Light from the separate images in Figure 14 travel along different paths in global coordinates between source and observer, as shown in Figure 13. If the intensity of the source changes, that change reaches the observer with different local time delays in the two images. Kundić et al (see the references) measured the local time delay between the images in Figure 14 and found it to be 417 ± 3 days: more than one Earth-year. This difference in locally-measured time delay seems large, but is a tiny fraction of the total lapse of global t along either path from that distant source to us.

Gravitational lensing detects dark matter.

Gravitational lensing by galaxies provides some of the strongest evidence for the existence and importance of **dark matter**. Galaxy masses obtained by gravitational lensing are much larger than the combined mass of all the visible stars in the measured galaxies. Most of the *matter* in the Universe is not atoms but a mysterious form called dark matter. Chapter 15 discusses the cosmological implications of this result. That chapter also examines, in

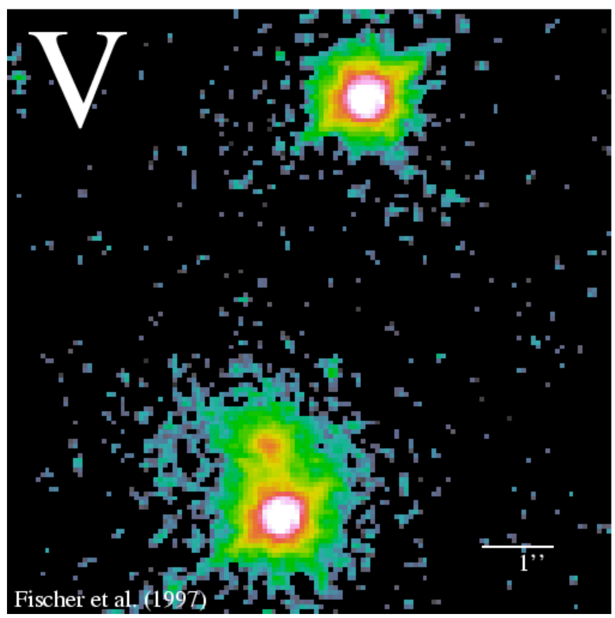


FIGURE 14 Double image from microwave observations of the distant quasar imaged by a foreground lensing structure—Figure 13 and equation (36). This is the first gravitationally lensed object, called QSO 0957+561A/B, observed in 1979 by Walsh, Carswell, and Weymann. The small patch above the lower image is the lensing galaxy. Redshifts of light from the distant quasar and the lensing galaxy lead to the following estimates of their model distances from Earth: lens galaxy, 4.6 billion light years; quasar source, 14.0 billion light years. The two images of the distant quasar are not collinear with that of the lens, which demonstrates that the lens is not spherically symmetric, as our analysis assumes.

addition to matter, the presence and importance of the mass-equivalent of light, neutrinos, and a larger and still more mysterious contribution called **dark energy**.

13.7 ■ MICROLENSING

The image brightens, then dims again.

Microlensing: when images cannot be resolved

Suppose that our detector—telescope, microwave dish, X-ray imaging satellite, or some other—cannot resolve the separate images of a distant star caused by an intermediate gravitational lens. Einstein warned us about this in 1936 (Section 13.1). In this case we see only one image of the source. Nevertheless, the intermediate lens directs more light into our telescope than would otherwise arrive from the distant source. We call this increase of light **microlensing**. How can we use that increased amount of light to learn about the lensing structure that lies between the source and us? We begin with a necessary set of definitions.

Instead, detect increased brightness of the single image.

DEFINITION 1. Solid angle

Solid angle

To measure star patterns in the night sky—whether detected by visible light, microwaves, infrared, ultraviolet, X-rays, or gamma rays—astronomers record the *angle* between any given pair of images. *For astronomers, angle is the only dependable geometric measure of the heavens.* The cross-hatched region of a distant source in Figure 15 has a length and width both measured in angle. We call the resulting measure **solid angle**. Solid angle is angular area, measured in square arcseconds or square radians (square radians has the technical name **steradians**). In the following we derive the *ratio* of solid angles, defined as *magnification*.

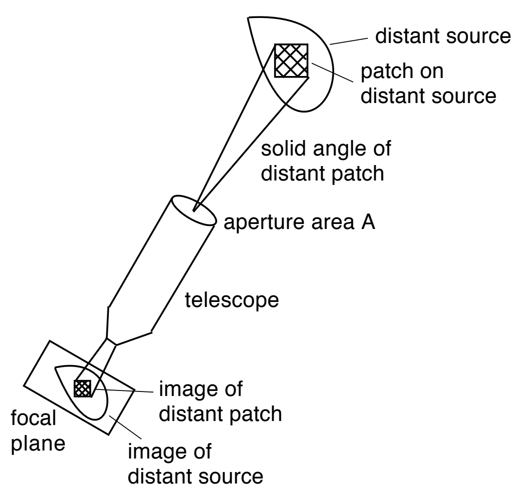


FIGURE 15 Figure for Definition 2 of Intensity, Flux, and Magnification.

DEFINITION 2. Intensity, Flux, and Magnification

Figure 15 helps to define intensity, flux, and magnification. We impose two conditions on these definitions: (1) The definitions must describe light from the source and not the instrument we use to measure it. (2) The distant object being observed is not a point source, so we can speak of a patch of solid angle on that source.

Intensity

A camera attached to a telescope of aperture area A (Figure 15) displays the image of a patch with a given solid angle on the sky, say the portion of a distant galaxy. The intensity I of the light is defined as:

$$I \equiv \frac{\text{total energy of light from patch recorded by camera}}{\text{local time} \times \text{aperture } A \times \text{solid angle of patch in the sky}} \quad (37)$$

Flux

The flux F of the source is the total energy striking the camera plane from the entire source per unit local time and per aperture area A :

$$\text{Flux} \equiv F = \int_{\text{over source}} (\text{Intensity}) d(\text{solid angle of source}) \quad (38)$$

Magnification

Now place a gravitational lens between source and detector. The image in the focal plane will be changed in size and also distorted. However, its intensity is not changed. *Example for a conventional lens:* Hold a magnifying lens over a newspaper. The lens directs more light into your eye; the flux increases. However, the area of the newspaper image on your retina increases by the same ratio. *Result:* The newspaper does not look brighter; you see its intensity as the same. However, a larger image with the same intensity means more flux, more energy, in the same ratio as solid angles, namely magnification. So magnification is equal to the ratio of fluxes, even when your detector cannot resolve the larger image. The result for a magnifying lens is also the result for a gravitational lens: the flux increases in proportion to the magnification.

$$\text{Magnification} \equiv \frac{(\text{solid angle with gravitational lens})}{(\text{solid angle without gravitational lens})} \quad (39)$$

$$= \frac{F(\text{with gravitational lens})}{F(\text{without gravitational lens})} \quad (40)$$

Figure 16 and Box 3 derive the magnification of a point gravitational lens.

Use microlensing to detect and study invisible lensing structure.

Astronomers use microlensing to detect the presence and estimate the mass of an intermediate lensing object, for example a star that is too dim for us to see directly.

?

Objection 2. *Wait a minute! When we see a distant source, it is just a source like any other. How can we tell whether or not the flux from this*

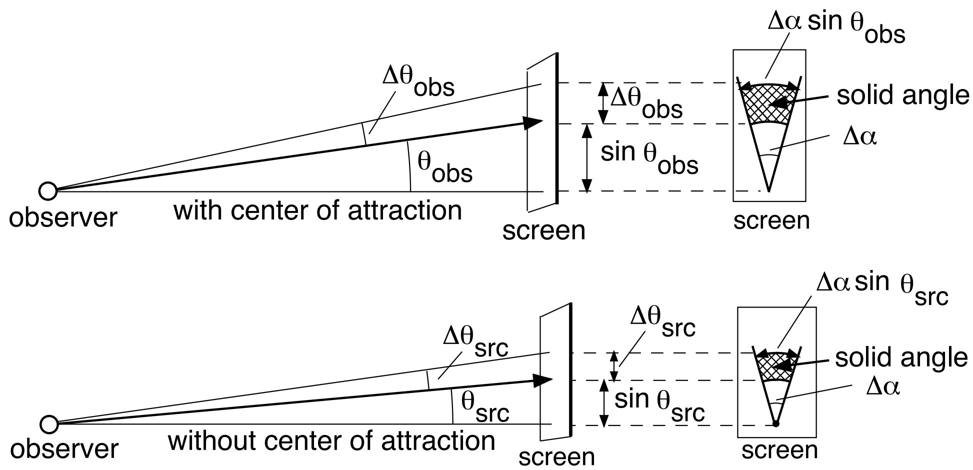


FIGURE 16 Magnification of the image of an extended source by a center of attraction acting as a gravitational lens. A patch on the source is the cross-hatched solid angle at the right of the lower panel; the corresponding patch on the image is the cross-hatched solid angle at the right of the upper panel. The magnification is the ratio of the upper to the lower cross-hatched solid angles. Box 3 employs this figure to derive the magnification of a gravitational lens.

Box 3. Image Magnification

Figure 16 shows cross-hatched solid angles whose ratio defines the magnification of an extended source by a gravitational lens. Magnification is defined as the ratio of the cross-hatched solid angle patch in the upper panel (with the center of attraction present) to the cross-hatched solid angle patch in the lower panel (with no center of attraction present).

To find this ratio, pick a wedge of small angle $\Delta\alpha$, the same for both panels. The radius of the cross-hatched solid angle for each wedge is proportional to $\sin\theta_{\rm obs}$ in the upper panel and $\sin\theta_{\rm src}$ in the lower panel. The angular spread in each case is $\Delta\alpha$ times the sine factor.

Then the magnification, equal to the ratio of solid angles with and without the center of attraction, becomes:

$$\text{Mag} = \left|\frac{\sin\theta_{\rm obs}\Delta\theta_{\rm obs}}{\sin\theta_{\rm src}\Delta\theta_{\rm src}}\right| \approx \left|\frac{\theta_{\rm obs}d\theta_{\rm obs}}{\theta_{\rm src}d\theta_{\rm src}}\right| \qquad (41)$$

where we add absolute magnitude signs to ensure that the ratio of solid angles is positive. In the last step of (41) we assume that observation angles are small, so that $\sin\theta \approx \theta$ and $\Delta\theta \approx d\theta$. Now into (41) substitute $d\theta_{\rm src}$ from the differential of both sides of (32), with $\theta_{\rm E}$ a constant:

$$d\theta_{\rm src} = d\theta_{\rm obs} + \frac{\theta_{\rm E}^2}{\theta_{\rm obs}^2}d\theta_{\rm obs} \qquad (42)$$

Equation (41) becomes

$$\begin{aligned}\text{Mag} &= \left|\frac{\theta_{\rm obs}d\theta_{\rm obs}}{\left(\theta_{\rm obs} - \dfrac{\theta_{\rm E}^2}{\theta_{\rm obs}}\right)\left(1 + \dfrac{\theta_{\rm E}^2}{\theta_{\rm obs}^2}\right)d\theta_{\rm obs}}\right| \\ &= \left|\left(1 - \frac{\theta_{\rm E}^2}{\theta_{\rm obs}^2}\right)^{-1}\left(1 + \frac{\theta_{\rm E}^2}{\theta_{\rm obs}^2}\right)^{-1}\right| \qquad (43)\end{aligned}$$

So finally,

$$\text{Mag} = \left|1 - \frac{\theta_{\rm E}^4}{\theta_{\rm obs}^4}\right|^{-1} \qquad (44)$$

source has already been increased by an intermediate gravitational lens? For example, it might just be a brighter star.

Good point. For a static image—one that does not change as we watch it—we cannot tell whether or not an intermediate lens has already changed the flux. However, if the source and lensing object move with respect to one another—which is the usual case—then the total flux changes with local time, growing to a maximum as source and gravitational lens line up with one another, then decreasing as this alignment passes. Figure 17 displays a theoretical family of such curves and Figure 18 shows the result of an observation.

Figure 13 shows that the observer receives two images of the source. Even though we cannot currently resolve these two images in microlensing, the total flux received is proportional to the summed magnification of both images:

$$\mathrm{Mag_{total}} = \mathrm{Mag}\left(\theta_{\mathrm{obs}+}\right) + \mathrm{Mag}\left(\theta_{\mathrm{obs}-}\right) \equiv \mathrm{Mag}_{+} + \mathrm{Mag}_{-} \qquad (45)$$

Total magnification equals increased flux.

where (36) gives $\theta_{\mathrm{obs}\pm}$.

Now we descend into an algebra orgy: Divide both sides of (36) by θ_{E} and substitute $q \equiv \theta_{\mathrm{src}}/\theta_{\mathrm{E}}$. Insert the results into (44). The expression for the separate magnifications M_{+} and M_{-} of the two images become:

$$\mathrm{Mag}_{\pm} = \left| \frac{1 + 2q^2 + \dfrac{q^4}{2} \pm \dfrac{q}{2}\left(q^2+2\right)\left(q^2+4\right)^{1/2}}{2q^2 + \dfrac{q^4}{2} \pm \dfrac{q}{2}\left(q^2+2\right)\left(q^2+4\right)^{1/2}} \right| \qquad \text{where} \quad q = \frac{\theta_{\mathrm{src}}}{\theta_{\mathrm{E}}} \quad (46)$$

Substitute this result into (45) to find the expression for total magnification.

$$\mathrm{Mag_{total}} = \frac{q^2+2}{q\left(q^2+4\right)^{1/2}} \qquad \text{where} \quad q = \frac{\theta_{\mathrm{src}}}{\theta_{\mathrm{E}}} \qquad (47)$$

Comment 3. Variation of $\mathrm{Mag_{total}}$ with q

It may not be obvious that smaller q results in larger total magnification. Convince yourself of this by taking the derivative of (47) with respect to q or by plotting its right hand side.

The maximum magnification—the maximum brightness of the microlensed background source—occurs for the minimum value of q:

$$\mathrm{Mag_{total,max}} = \frac{q_{\mathrm{min}}^2+2}{q_{\mathrm{min}}\left(q_{\mathrm{min}}^2+4\right)^{1/2}} \qquad \text{where} \quad q_{\mathrm{min}} = \frac{\theta_{\mathrm{src,min}}}{\theta_{\mathrm{E}}} \qquad (48)$$

What does the observer see as a the source passes behind the lens? To answer this question, give the source angle a time derivative in equation (47):

$$q = \frac{\dot{\theta}_{\mathrm{src}}}{\theta_{\mathrm{E}}}\left(t - t_0\right) \qquad (49)$$

The symbol $\dot{\theta}_{\mathrm{src}}$ is the angular velocity of the moving source seen by the observer—called its **proper motion** by astronomers—and t_0 is the observed

Proper motion

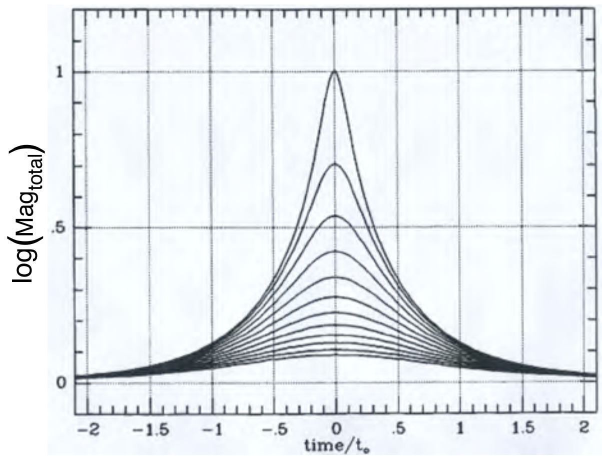


FIGURE 17 Log of total magnification $\text{Mag}_{\text{total}}$ (vertical axis) due to microlensing as the source moves past the deflecting mass, with t_0 a normalizing local time of minimum separation. Different curves, from top to bottom, are for the 12 values $q_{\min} = 0.1, 0.2 \ldots 1.1, 1.2$ in equation (47), respectively. From a paper by Paczynski, see the references.

time at which the source is at the minimum angular separation $\theta_{\text{src,min}}$. Substitute (49) into the definition of q in (47) to predict the local time dependence of the apparent brightness of the star. Figure 17 shows the resulting predicted set of light curves for different values of $q_{\min} = \theta_{\text{src,min}}/\theta_{\text{E}}$.

“Achromatic” gravitational lens

Comment 4. Gravitational lenses are achromatic.
Equations of motion for light around black holes are exactly the same for light of every wavelength. Technical term for any lens with this property: **achromatic**. Gravitational lenses often distort images terribly, but they do not change the color of the source, even when “color” refers to microwaves or gamma rays. The achromatic nature of a gravitational lens can be important when an observer tries to distinguish between increased light from a source due to microlensing and increased light due to the source itself changing brightness. A star, for example, can increase its light output as a result of a variety of internal processes, which most often changes its spectrum in some way. In contrast, the increased flux of light from the star due to microlensing does not change the spectrum of that light. Therefore any observed change in flux of a source without change in its spectrum is one piece of evidence that the source is being microlensed.

Predicted microlensing curve observed

Figure 18 shows a microlensing curve for an event labeled OGLE 2005-BLG-390. The shape of the observed magnification curve closely follows the predicted curves of Figure 17 when converted to a linear scale.

What can we learn from an observation such as that reported in Figure 18? In Query 5 you explore two examples.

QUERY 5. Results from Figure 18

A. From the value of the magnification in Figure 18, find the value of $q_{\min}$.

B. Measure the observed time between half-maximum magnifications in Figure 18; call this $2(t_{1/2} - t_0)$. The horizontal axis in Figure 17 expresses this observed time as a function of $\dot{\theta}(t - t_0)/\theta_{\text{src,min}}$. (Careful: this is a semi-log plot.) From these two results calculate the value of the quantity $\dot{\theta}/\theta_{\text{src,min}}$.

Exoplanet detected

The tiny spike on the right side of the curve in Figure 18—magnified in the inset labeled "planetary deviation"—shows another major use of microlensing: to detect a planet orbiting the lensing object. The term for a planet around a star other than our Sun is **extra-solar planet** or **exoplanet**.

How exoplanet detection is possible.

The presence of a short-duration, high-magnification achromatic spike in a long microlensing event is evidence for an exoplanet, which causes additional deflection and magnification of one of the two images. This additional magnification results from the small value of q in equation (48) and has much shorter duration $\theta_{\text{src,min}}/\dot{\theta}_{\text{src}}$ than that due to the primary lens star because the minimum separation between the exoplanet and light ray is very small, as shown in the middle panel of Figure 19.

Comment 5. Shape of the exoplanet curve

The faint background curve in the "Planetary deviation" inset of Figure 18 has the shape similar to curves in Figure 17 predicted for a static point-mass planet.

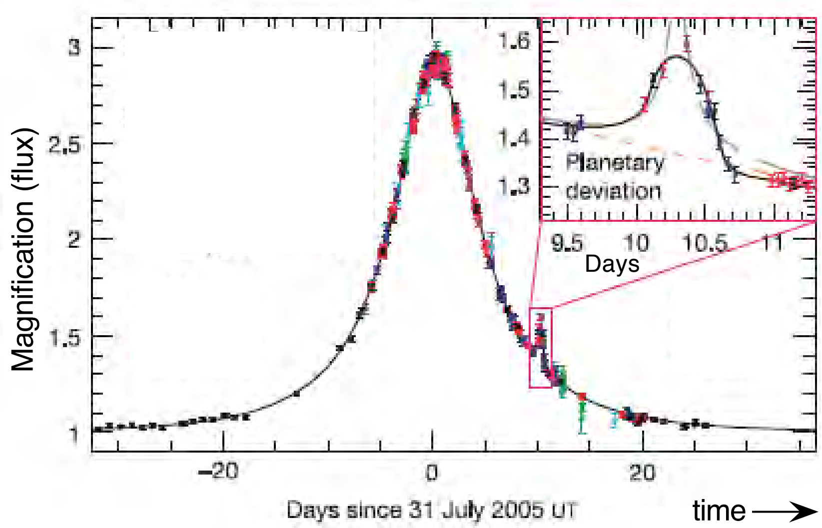


FIGURE 18 Microlensing image with the code name OGLE 2005-BLG-390. The lens is a dwarf star, a small relatively cool star of approximately 0.2 solar mass. Observed time along the horizontal axis shows that the variation of intensity can take place over days or weeks. The abbreviation UT in the horizontal axis label means "universal time," which allows astronomical measurements to be coordinated, whatever the local time zone of the observer. The inset labeled "planetary deviation" detects a planet of approximately 5.5 Earth masses orbiting the lens star.

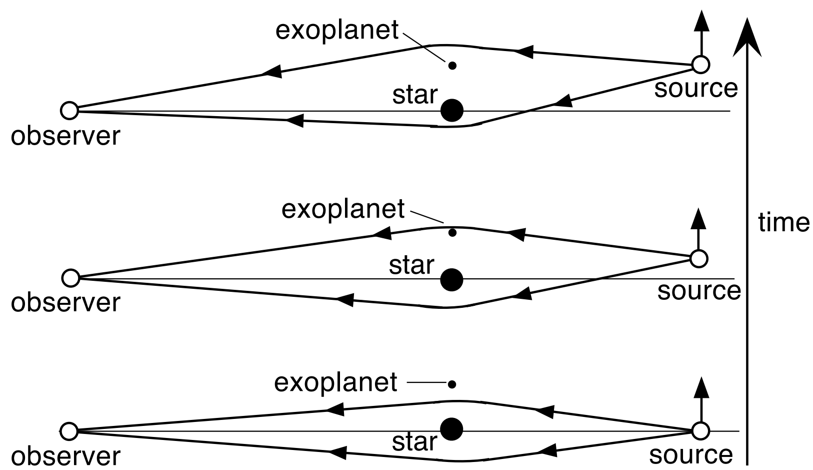


FIGURE 19 Schematic diagram of the passage of the source behind the lensing star with an exoplanet that leads to the small spike in the local time-dependent flux diagram of Figure 18. Observer time increases from bottom to top. The bottom panel displays the alignment at local time $t - t_0 = 0$ (Figure 17), when the source is directly behind the lens, which results in maximum flux from the source at the observer. The middle panel shows the alignment that leads to the maximum of the little spike in Figure 18. Figure not to scale.

> *Question:* Why does the *shape* of exoplanet-induced magnification curve differ from this prediction (as hinted by the phrase "planetary deviation")? *Answer:* Because the planet moves slightly around its mother star during the microlensing event, so $\dot{\theta}$ is not constant.

Analysis of the exoplanet spike on a microlensing flux curve is just one of several methods used to detect exoplanets; we do not describe other methods here.

13.8 ■ REFERENCES

Initial quote: Ilse Rosenthal-Schneider, in *Some Strangeness in the Proportion*, Edited by Harry Woolf, 1980, Addison-Wesley, Reading MA, page 523. The German original plus correction from Walter Isaacson, *Einstein: His Life and Universe*, 2007, Simon and Schuster, page 600, Note 22. Concerning the original, a native German speaker remarks, "This is really quite endearing, because you clearly hear his southern German dialect through the choice of words."

Some dates in Section 13.1 Timeline from a chronology in Abraham Pais, *Subtle is the Lord: The Science and Life of Albert Einstein*, 1982, Clarendon Press, pages 520-530.

Mrs. Freundlich's tablecloth: *Einstein, the Life and Times* by Ronald W. Clark, New York, Harper Perennial, 2007, page 222.

"Lights All Askew in the Heavens" *New York Times* headline, November 10, 1919.

J. G. von Soldner calculates an incorrect deflection (0.84 arcsecond) using Newtonian mechanics: *Astronomisches Jahrbuch für das Jahr 1804* (Späthen, Berlin, 1801), page 161.

Einstein calculates an incorrect deflection (0.83 arcsecond) effectively equal to that of Soldner using his incomplete general relativity, A. Einstein, *Annalen der Physik,* 1911, Volume 340, Issue 19, pages 898-906.

Einstein calculates the substantially correct deflection(1.7 arcseconds). German source: *Sitzungberichte der Preussischen Akademie der Wissenschaften zu Berlin,* Volume 11, pages 831-839 (1915). English translation by Brian Doyle in *A Source Book in Astronomy and Astrophysics, 1900-1975,* pages 820-825. Deflection result on page 823. 4.

Einstein demonstrates the possibility, suggested by Mandl, that a star can act as a gravitational lens for light from a star behind it. A. Einstein, *Science,* 1936, Volume 84, pages 506-507.

Zwicky proposes that a galaxy can form visible images, can be used to measure the total mass of the lensing galaxy and also to magnify distant structures: F. W. Zwicky, *Physical Review,* 1937, Volume 51, page 290.

Zwicky says that gravitational lensing by galaxies is not only possible but, given the angular density of galaxies, "becomes practically a *certainty.*" F. W. Zwicky, *Physical Review,* 1937, Volume 51, page 679.

Joshua Winn: Section 13.2 treatment of Newtonian starlight deflection adapted from his lecture to an MIT class.

Double Einstein ring: Gavazzi, Treu et al *Astrophysical Journal,* Volume 677, pages 1046-1059, April 20, 2008.

Image of 0957-5671: Walsh, Carswell, and Weyman, *Nature*, Volume 279, May 31, 1979, pages 381-384.

Time delay between two images in 0957-561: Kundić et al *Astrophysical Journal.* Volume 482, pages 75-82, June 10, 1997.

Magnification curves of Figure 17 from Bohdan Paczynski, *The Astrophysical Journal*, Volume 304, May 1, 1986, pages 1-5.

Microlensing curve of Figure 18 adapted from Beaulieu, Bennett, Fouqué et al, *Nature*, Volume 439, pages 437-440, January 26, 2006.

Download file name: Ch13GravitationalMirages160510v1.pdf

Chapter 14. Expanding Universe

- *What does "expansion of the Universe"* mean *and how can I observe it?*
- *What does the Universe expand* from*? What does it expand* into*?*
- *How can a metric describe the Universe as a whole?*
- *You assume "for simplicity" that the Universe is uniform, but a glance at the night sky shows your assumption to be false!*
- *How many different kinds of uniform curvature are possible for the Universe as a whole?*
- *How do galaxies move as the Universe expands?*
- *How do we measure the distance to a remote galaxy?*
- *How far away "now" is the most distant galaxy that we see "now"?*
- *And what does "now" mean, anyway? Even special relativity shouts, "Simultaneity is relative!"*

Download file name: Ch14ExpandUniverse170331v1.pdf

CHAPTER

14 Expanding Universe

Edmund Bertschinger & Edwin F. Taylor *

Nothing expands the mind like the expanding universe.

—Richard Dawkins

14.1 ■ DESCRIBING THE UNIVERSE AS A WHOLE

Finding words that correctly describe the unbounded

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

One-sentence description of our Universe

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

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

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

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

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

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

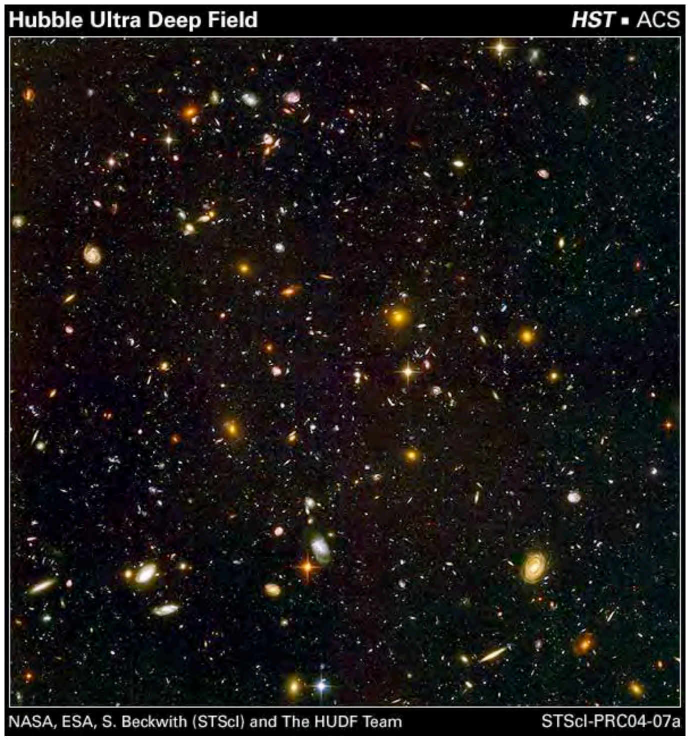


FIGURE 1 "Ultra deep field" image from the Hubble Space Telescope, named after astronomer Edwin P. Hubble. Every dot and every smear in this image is a galaxy, with the exception of a few nearby stars in our local galaxy. (Can you distinguish these exceptions?)

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

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

Brief history of the Universe

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

Box 1. Is this the only Universe?

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

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

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

How do we know?

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

14.2 ■ SPACE METRICS FOR A STATIC UNIVERSE

Describing a uniform space

Space metric for uniform space curvature

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

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

Flat, closed, and open spaces

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

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

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

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

Flat plane, sphere, and saddle

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

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

Traveling in flat space

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

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

This leads to the metric for flat space:

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

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

Variable χ automatically satisfies limits.

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

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

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

The differential dr is

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

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

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

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

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

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

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

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

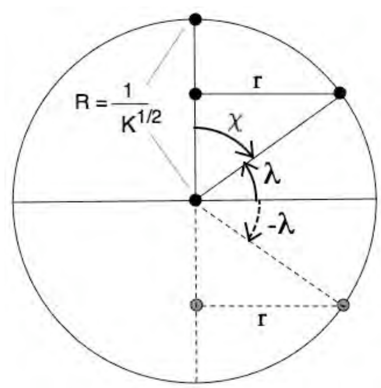


FIGURE 2 Relation between latitude λ and colatitude χ to determine the north-south coordinate on the sphere with $R = 1/K^{1/2}$ in Euclidean space. Latitude λ ranges over the values $-\pi/2 \leq \lambda \leq +\pi/2$, whereas colatitude χ ranges over $0 \leq \chi \leq \pi$. Equation (11) gives the relation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

Describing closed space

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

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

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

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

Comment 1. Scale factor R?

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

Describing open space

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

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

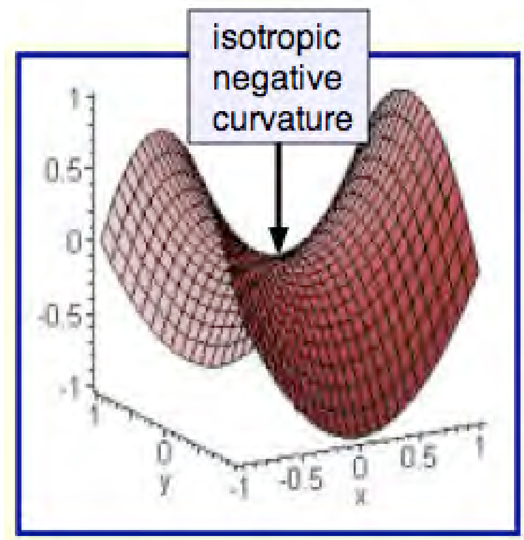


FIGURE 3 The saddle shape has intrinsic negative curvature. Only in the neighborhood of a single (central) point, however, is the negative curvature the same in all directions. Elsewhere on the surface the curvature is negative but varies from place to place and is different in different directions.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

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

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

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

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

and the corresponding space metric is

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

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

Box 2. What does the Universe expand into?

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

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

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

14.3 ■ ROBERTSON-WALKER GLOBAL METRIC

A Universe that expands

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

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

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

Robertson-Walker metric

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

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

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

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

Comoving coordinates

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

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

Box 3. Is a static, uniform Universe possible?

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

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

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

YOU ARE AT THE "CENTER OF THE UNIVERSE."

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

Global t on wristwatch of comoving observer

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

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

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

14.4 ■ REDSHIFT

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

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

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

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

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

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

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

Emit and detect two light flashes.

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

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

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

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

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

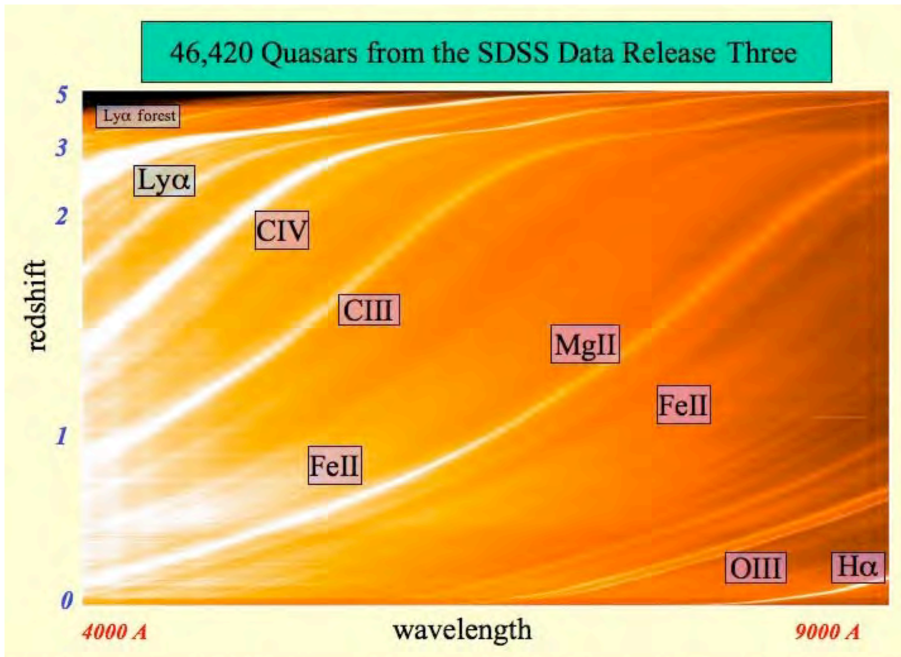
$$\frac{\Delta t_0}{R(t_0)} \approx \frac{\Delta t_{\text{emit}}}{R(t_{\text{emit}})} \qquad (\text{light}) \qquad (25)$$

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

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

Redshift z

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



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

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

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

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

Stretch factor: $1+z$

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

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

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

Cosmological redshift

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

Redshift a Doppler shift?

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

Redshift deduced from laboratory spectra

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

Astronomers use z for $t_{\rm emit}$.

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

14.5 ■ HOW DO GALAXIES MOVE?

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

Transverse galaxy motion is difficult to detect.

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

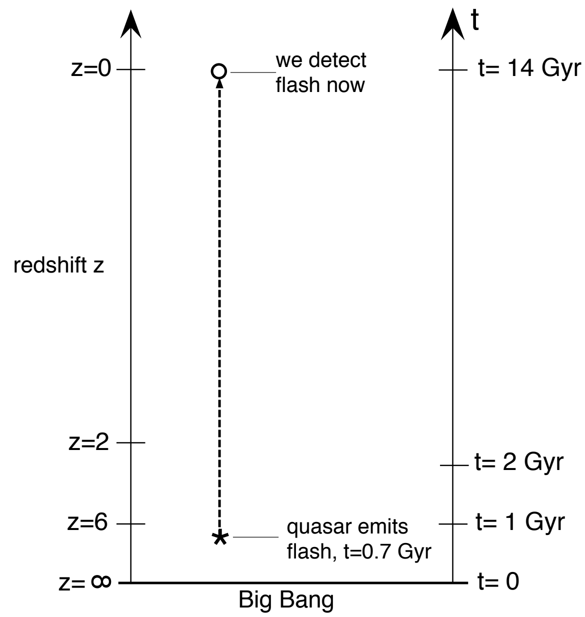


FIGURE 5 Schematic diagram comparing redshift z with cosmic t, in units of Gigayears (10^9 years). Calibration of the scale at the right of the figure depends on the t-development of the Universe, through $R(t)$, based on our current model. Astronomers use redshift as a proxy for t, both because it is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

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

Limit attention to radial motion.

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

Galaxy motion from Principle of Maximal Aging

For radial motion, the metric (17) becomes:

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

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

Seek a conserved quantity.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

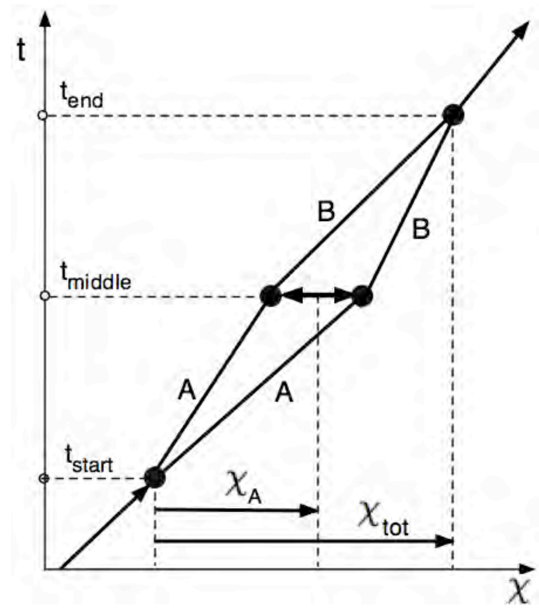


FIGURE 6 Greatly magnified picture of alternative worldlines across incremental segments A and B used in the derivation of the constant of motion (38). We vary the position $\chi_{\rm A}$ of the middle event between segments A and B and demand that the total wristwatch time across both segments be maximum.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

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

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

and

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

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

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

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

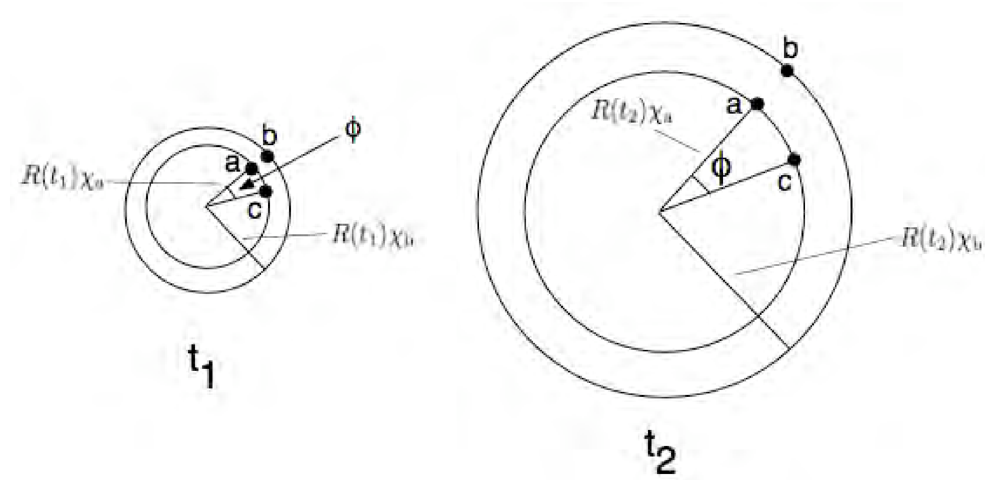


FIGURE 7 One possible radial motion for a galaxy is to remain at rest in the comoving coordinate χ and ϕ and ride outward, following $R(t)$, as the Universe expands. This figure shows the result for a flat Universe. All separations increase by the same ratio, so every observer can analyze galaxy motion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

or

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

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

Constant $Q_{\rm r}$ for radial motion only

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

Constant of motion for galaxy or light

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

Two possible radial motions

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

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

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

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

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

Constant Q_ϕ for *any* motion

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

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

$$\text{(constant for } any \text{ free motion)}$$

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

14.6 ■ MEASURING DISTANCE

Extending a ruler from one lonely outpost.

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

Problems with our observations

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

Box 5. Edwin P. Hubble

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

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

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

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

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

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

Determine "distance" with a "standard candle."

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

Cepheid variables: standard candles

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

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

Hubble's "island universes" = our galaxies.

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

Modern standard candle: Type Ia supernova

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

For astronomers, M and m are magnitudes.

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

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

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

Hubble Diagram

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

Expansion speeding up

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

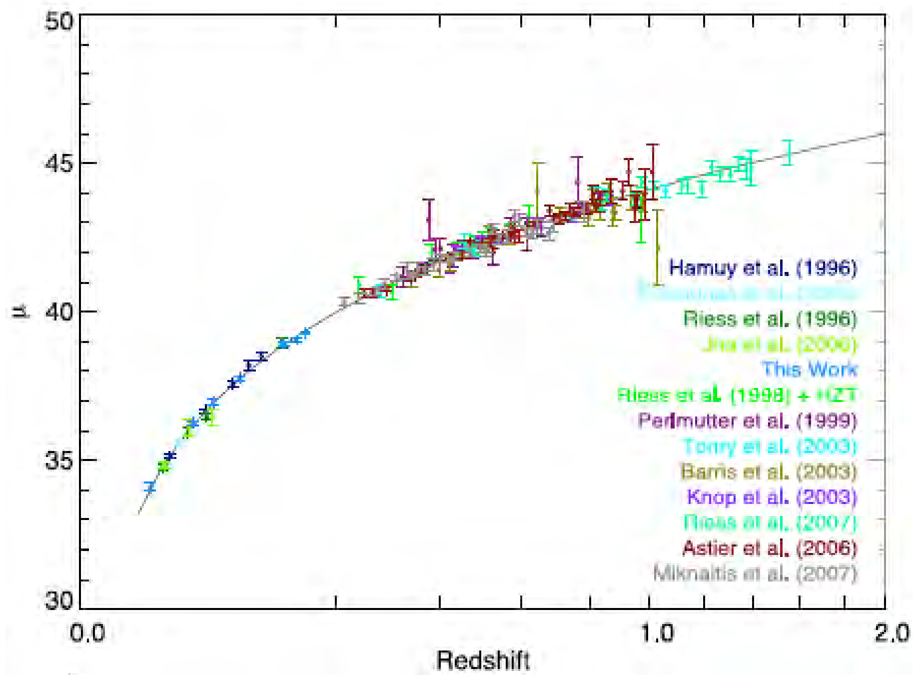


FIGURE 9 Effective magnitude of Type Ia supernovae as a function of their redshift z. The vertical axis is $\mu = m - M$, the difference between apparent magnitude and intrinsic magnitude.

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

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

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

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

Standard rulers

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

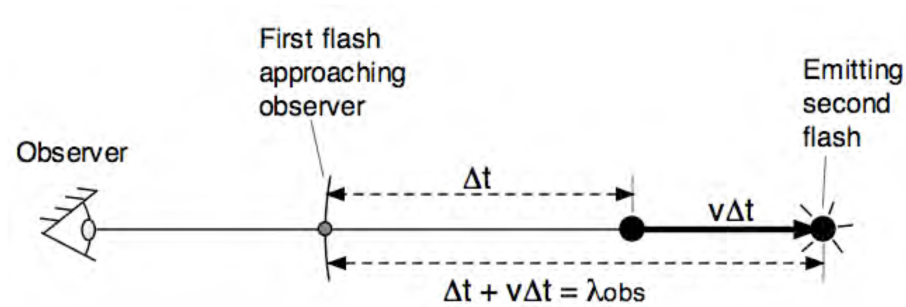


FIGURE 10 Doppler effect observed in a single inertial frame of special relativity, used by Hubble to analyze the speed of receding nearby galaxies.

14.7 ■ LAWS OF RECESSION

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

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

Velocity vs. Redshift

Hubble used special relativity Doppler shift.

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

Hubble uses special relativity Doppler shift

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

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

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

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

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

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

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

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

$$v = z \qquad (v \ll 1) \tag{48}$$

Doppler OK for small z

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

Measuring Distance with a "Standard Candle"

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

Hubble's Law of Recession

Hubble's law of recession

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

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

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

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

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

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

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

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

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

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

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

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

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

Hubble constant H_0

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

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

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

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

Robertson-Walker Law of Recession

Recession at great distance and great speed

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

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

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

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

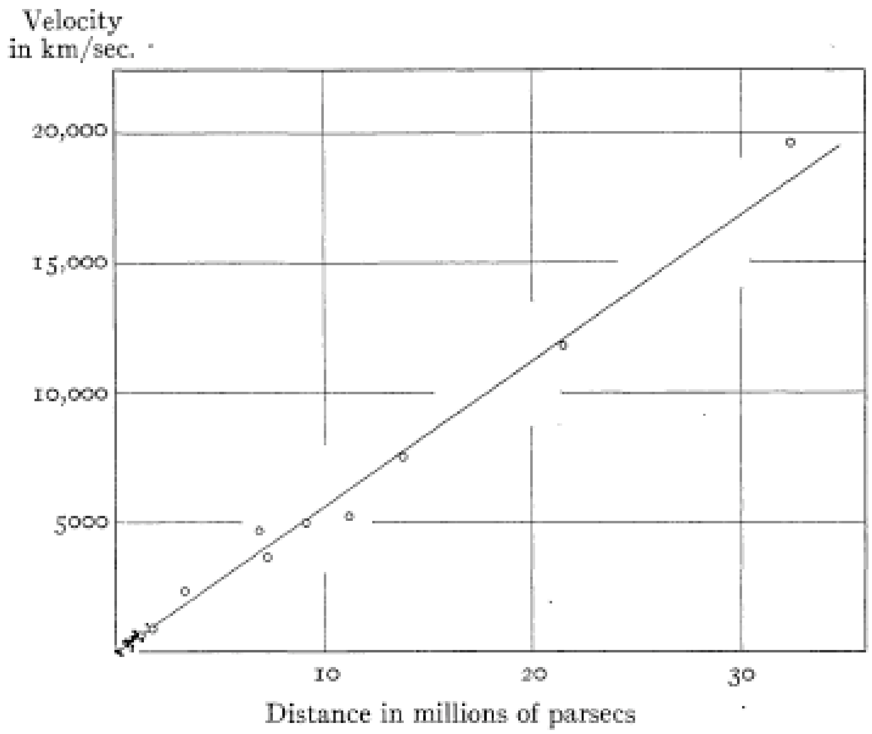


FIGURE 11 A plot of recession velocity as a function of distance by Hubble and Milton Humason (1931). Open circles represent averages of groups of galaxies; solid dots near the origin show individual galaxies from an earlier paper by Hubble. A parsec equals 3.26 light-years, so the most distant group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

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

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

The recession speed at t is expressed using elementary calculus:

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

Hubble parameter

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

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

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

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

We need radial function $R(t)$.

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

14.8 ■ EXERCISES

1. Tangential Momentum

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

2. Energy not a Constant of Motion

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

3. Transverse Motion

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

5. Hubble's Error

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

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

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

14.9 ■ REFERENCES

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

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

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

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

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

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

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

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

Download file name: Ch14ExpandUniverse170331v1.pdf

Chapter 15. Cosmology

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

Download file name: Ch15Cosmology170510v1.pdf

CHAPTER 15

Cosmology

Edmund Bertschinger & Edwin F. Taylor *

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

—Robert Frost, "Fire and Ice"

15.1 ■ CURRENT COSMOLOGY

Summary of current cosmology.

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

Our golden age of cosmology

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

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

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

Box 1. Fantasy: Present at the Creation

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

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

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

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

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

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

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

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

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

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

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

Dark matter and dark energy

Major goals of current astrophysics research are (1) to find more accurate values of quantities that make up the Universe as a whole, (2) to explore the nature of dark matter, which evidently accounts for about 23 percent of the mass-energy in the Universe, and (3) to explore the nature of dark energy, which makes up about 73 percent. Everything we are made of and can see and touch accounts for only four percent of the mass of the Universe. This consists

Study constituents of the Universe

of protons and neutrons in the form of atoms and their associated electrons—called **baryonic matter** because its nuclei are made of protons and neutrons, which are called baryons.

Einstein's general relativity fail?

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

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

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

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

Answer with scale factor $a(t)$.

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

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

Friedmann equation

FRW cosmology

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

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

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

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

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

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

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

Comment 1. An aside on units

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

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

(Friedmann equation, conventional units)

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

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

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

Einstein links geometry with energy.

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

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

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

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

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

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

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

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

QUERY 1. Value of the critical density now on Earth

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

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

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

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

Find t-variation of density components.

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

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

Matter, radiation, and dark energy.

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

Universe composed of matter, radiation, and dark energy.

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

Matter

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

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

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

Stars and gas: mostly protons & neutrons.

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

What we see: 4% of Universe.

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

Neutrino mass is negligible.

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

Dark matter is currently estimated to account for approximately 23 percent of the mass-energy of the Universe. What is dark matter? And how do

Dark matter holds galaxies together.

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

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

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

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

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

Radiation

Radiation: mass much less than energy.

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

Recombination: Universe becomes transparent.

At the present stage of the Universe, radiation is a whisper, but it used to be a shout. Shortly after the Big Bang, radiation contributed the dominant fraction of the mass-energy density of the Universe. In the hot ionized plasma of the early Universe, radiation and matter were tightly coupled: photons continually scattered from free electrons, so photons could not move in straight lines and escape. About 300 000 years after the Big Bang, however, the Universe cooled to a temperature of about 3 000 K, at which electrons combined with protons to create hydrogen gas (with some helium and a trace amount of lithium). This period is called **recombination**, even though the stable electron-nucleus combination was taking place for the first time. At recombination, the Universe became transparent to radiation, and photons were essentially decoupled from matter, free to stream across the Universe unimpeded. The cosmic microwave background radiation that we observe in all directions is a view of that early transition from opaque to transparent, with later expansion lowering our observed temperature to 2.725 degrees Kelvin. It is remarkable that the low-energy photons we detect as background radiation between the stars have been streaming freely for billions of years, not interacting with anything until they enter our detectors.

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

The number of photons emitted by all the stars in the history of the Universe is tiny compared with the number of photons created in the hot Big Bang. In the early Universe these photons were continually being emitted, absorbed, and scattered, but the *number* of photons remains approximately constant as the Universe expands. Therefore the *number of photons per unit volume* varies inversely as the scale factor cubed, or as $a^{-3}(t)$, just as the number of matter particles do. But there is an additional effect for photons. The equation $E = hf = hc/\lambda$ connects the energy E of a photon to the frequency f and wavelength λ of the corresponding electromagnetic wave. The symbol h stands for **Planck's constant**, with the value $h = 6.63 \times 10^{-34}$ kilogram-meter^2/second in conventional units. As this wave propagates through an expanding space, its wavelength increases in proportion to $a(t)$.

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

This increased wavelength is observed as the redshift of light from distant galaxies. An increasing wavelength implies a *decrease* in the energy of each photon, an energy that varies as $a^{-1}(t)$. This leads to an extra (inverse) power of $a(t)$ compared with that for matter in equation (9) because of the drop in energy of each photon as the Universe expands. Let $\rho_{\text{rad},\,0}$ represent the energy density of radiation at t_0, the present age of the Universe. Then we predict that the radiation density obeys the equation

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

Dark Energy

Dark energy composition is unknown.

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

QUERY 2. Energy density of radiation

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

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

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

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

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

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

Dark energy = vacuum energy?

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

Einstein's cosmological constant

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

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

Assume dark energy density remains constant.

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

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

?

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

!

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

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

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

t-variation of total density.

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

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

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

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

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

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

Fractional densities Ω

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

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

$$\Omega_{\mathrm{mat},\,0} \equiv \frac{\rho_{\mathrm{mat},\,0}}{\rho_{\mathrm{crit},\,0}} = 0.27 \pm 0.03 \tag{16}$$

$$\Omega_{\Lambda,\,0} \equiv \frac{\rho_\Lambda}{\rho_{\mathrm{crit},\,0}} = 0.73 \pm 0.03 \tag{17}$$

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

$$\Omega_{\mathrm{rad},\,0} \equiv \frac{\rho_{\mathrm{rad},\,0}}{\rho_{\mathrm{crit},\,0}} \approx 8.4 \times 10^{-5} \tag{18}$$

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

We live between matter domination and vacuum energy domination.

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

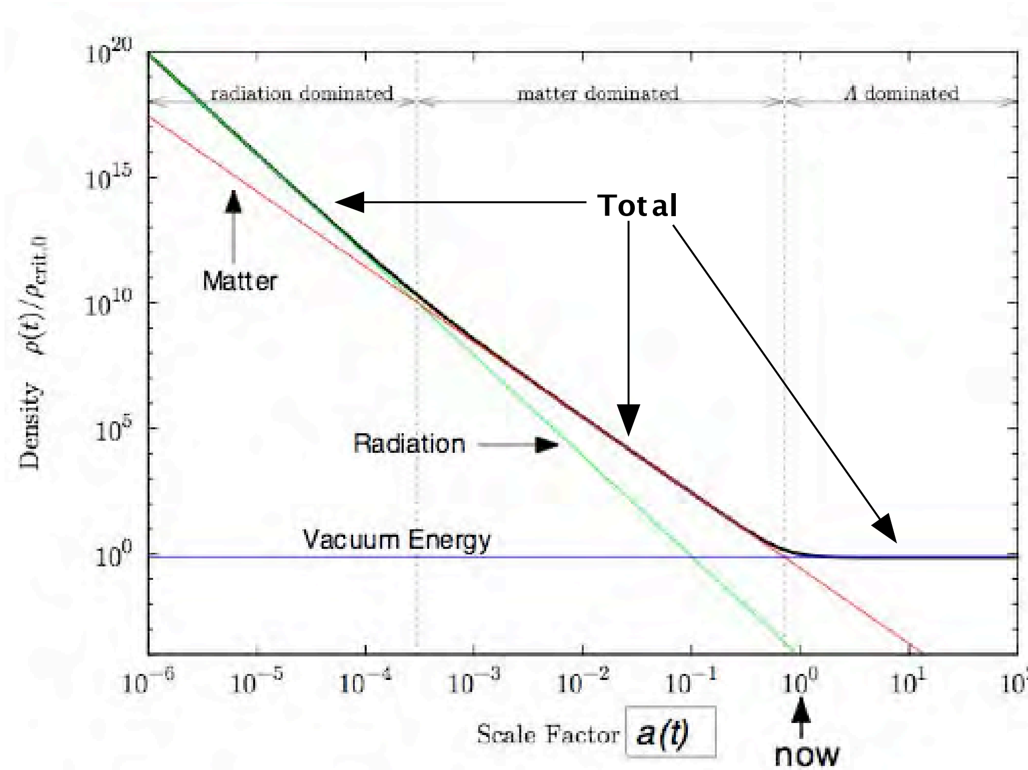


FIGURE 1 Total mass-energy density of the Universe (heavy line) in units of the present critical value as a function of the expansion scale factor. The vertical dashed lines denote transitions between the radiation-dominated early phase, the matter-dominated middle era, and the vacuum-energy-dominated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

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

QUERY 3. Contributions to the Density

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

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

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

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

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

?

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

!

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

15.4 ■ UNIVERSES WITH DIFFERENT CURVATURES

Effective potential for the Universe

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

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

where, remember, a dot over a symbol means its derivative with respect to t.

t-development of the Universe

Re-express equation (19) in terms of the components of Ω_{tot} defined in equations (8), (16), (17), and (18):

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Assumptions

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

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

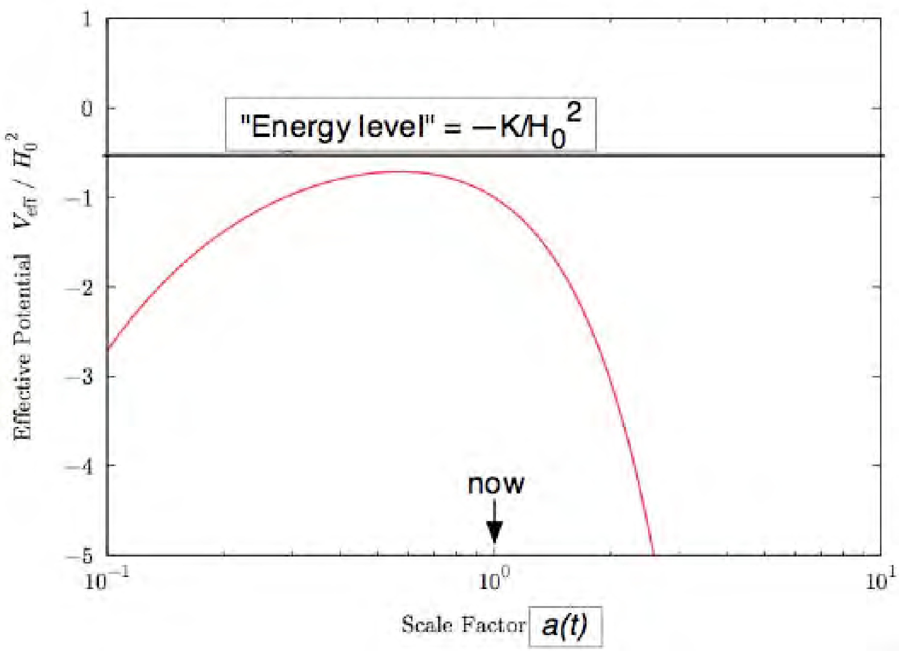


FIGURE 2 Effective potential governing the evolution of $a(t)$ according to equations (24) and (25). The "energy level" is set by $V_{\text{eff}}/H_0^2 = -K/H_0^2$. The figure shows an example of a closed Universe that expands endlessly. Our Universe has $K = 0$ to a good approximation and will apparently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

"Effective potential" for the Universe

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

QUERY 4. The Friedmann-Robertson-Walker Universe

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

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

15.5 ■ SOLVING FOR THE SCALE FACTOR

Integrating $\dot{a}(t)$

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

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

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

Integrate da/dt.

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

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

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

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

Present value of Hubble constant

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

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

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

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

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

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

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

QUERY 6. Various kinds of Universes

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

A. Assume the Universe contains only matter and that $\Omega_{\text{mat},0} = 1$. Find an expression for $a(t)$ and the corresponding value of $H_0 t_0$.
B. Assume the Universe contains only radiation and that $\Omega_{\text{rad},0} = 1$. Find an expression for $a(t)$ and the corresponding value of $H_0 t_0$.
C. Assume that the Universe contains only dark energy and that $\Omega_{\Lambda,0} = 1$. Find an expression for $a(t)$.
D. *Optional.* Discuss the validity of your results for parts A, B, and C for $t < t_0$ and in particular for $t = 0$.

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

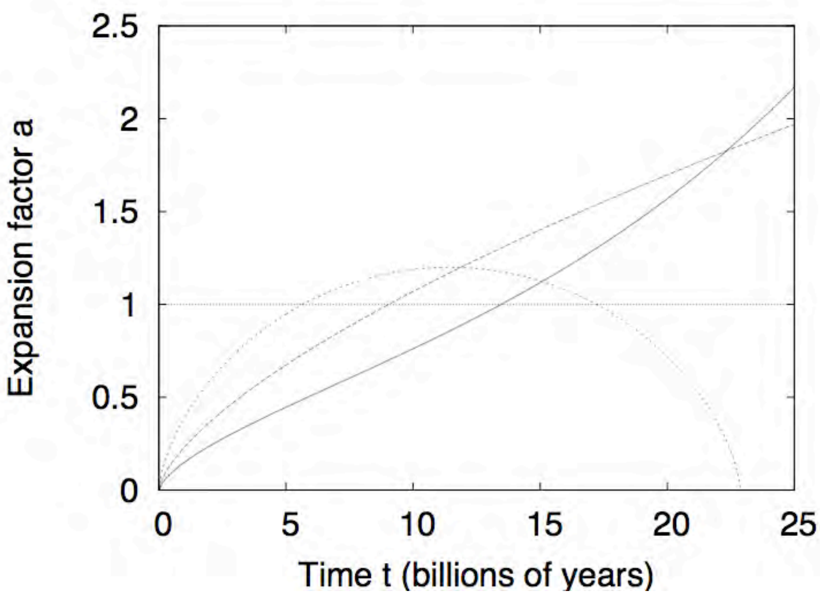


FIGURE 3 Expansion scale factor $a(t)$ versus t for three different models. The solid curve is the favored model with $\Omega_{\mathrm{mat},0} = 0.27$ and $\Omega_{\Lambda,0} = 0.73$. The two dotted curves show alternative models with no dark energy, $\Omega_{\mathrm{mat},0} = 1$ and $\Omega_{\mathrm{rad},0}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

15.6 ■ LOOK-BACK DISTANCE AS A FUNCTION OF REDSHIFT

Where are earlier emitters now?

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

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

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

?

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

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

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

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

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

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

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

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

where z' is a dummy variable of integration.

Look-back d_0 vs z

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

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

Negative pressure pushes!

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

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

Matter-dominated era: expansion slows down.

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

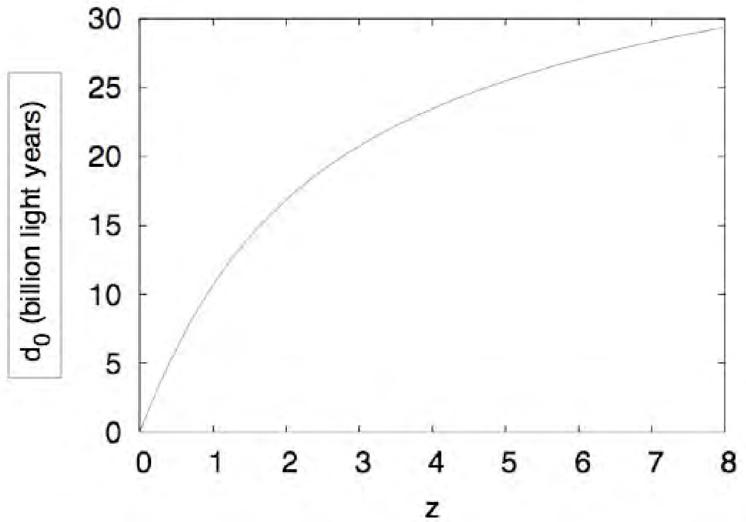


FIGURE 4 The present-day "look-back distance" d_0 to objects at redshift z. As explained in Box 4 in Section 14.4, in an expanding Universe an object that we see now (at its earlier position) is at present much farther away from us in light-years than the age of the Universe in years.

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

Thermodynamics

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

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

Expanding gas cools.

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

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

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

Positive pressure slows expansion.

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

QUERY 8. Acceleration of the Scale Factor

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

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

B. Rewrite the Friedmann equation (2) as

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

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

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

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

Negative pressure speeds up expansion.

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

QUERY 9. Pressure from Different Sources

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

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

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

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

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

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

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

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

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

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

Negative pressure? OK. Negative mass? No.

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

History of changes in expansion

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

MATTER: positive mass and zero pressure

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

RADIATION: positive energy density and positive pressure.

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

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

QUERY 10. Einstein's Static Universe

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

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

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

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

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

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

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

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

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

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

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

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

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

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

15.8 ■ CONTENTS OF THE UNIVERSE II: OBSERVATIONS

Galaxy rotation and cosmic background radiation

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

Galaxy Rotation: Evidence for Dark Matter

Evidence for dark matter.

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

Galaxy rotation curve

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

Surface luminosity density

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

$$\Sigma(R) = \Sigma_0 \exp(-R/h) \qquad (46)$$

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

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

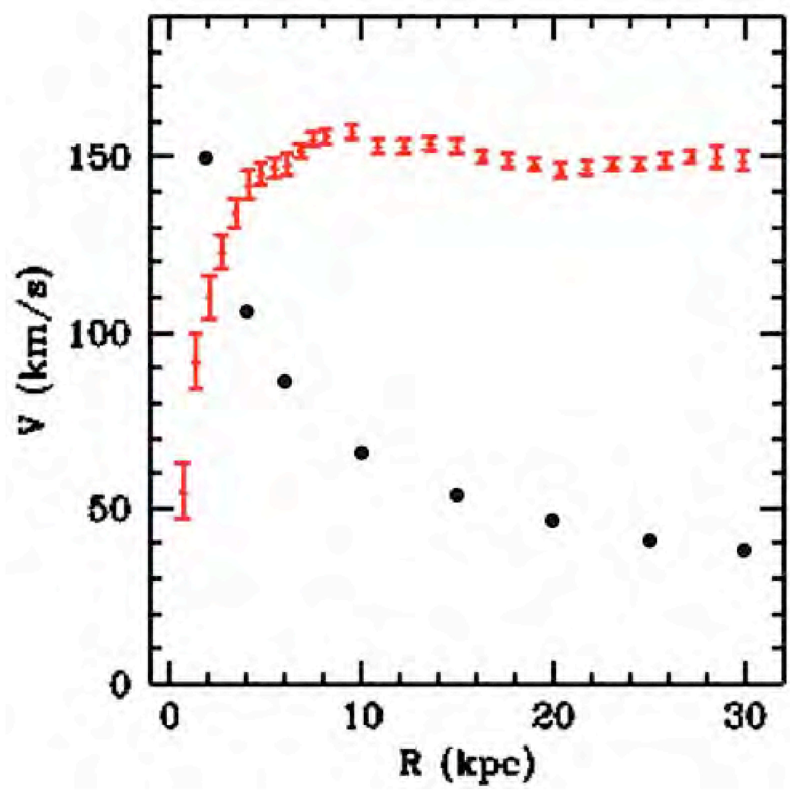


FIGURE 5 Upper plot: Rotation curve for spiral galaxy NGC 3198, from Begeman 1989, *Astronomy and Astrophysics*, 223, 47. Filled dots: Points showing the shape of a rotation curve if the attractive mass were concentrated at the center, for example in our solar system. The vertical position of the filled-dot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

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

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

Mass *vs* lumnosity

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

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

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

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

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

QUERY 11. Dark Matter from a Rotation Curve

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

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

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

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

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

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

Cosmic Microwave Background Radiation

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

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

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

Why blackbody spectrum?

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

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

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

Redshift at recombination.

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

Our earliest view of the Universe.

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

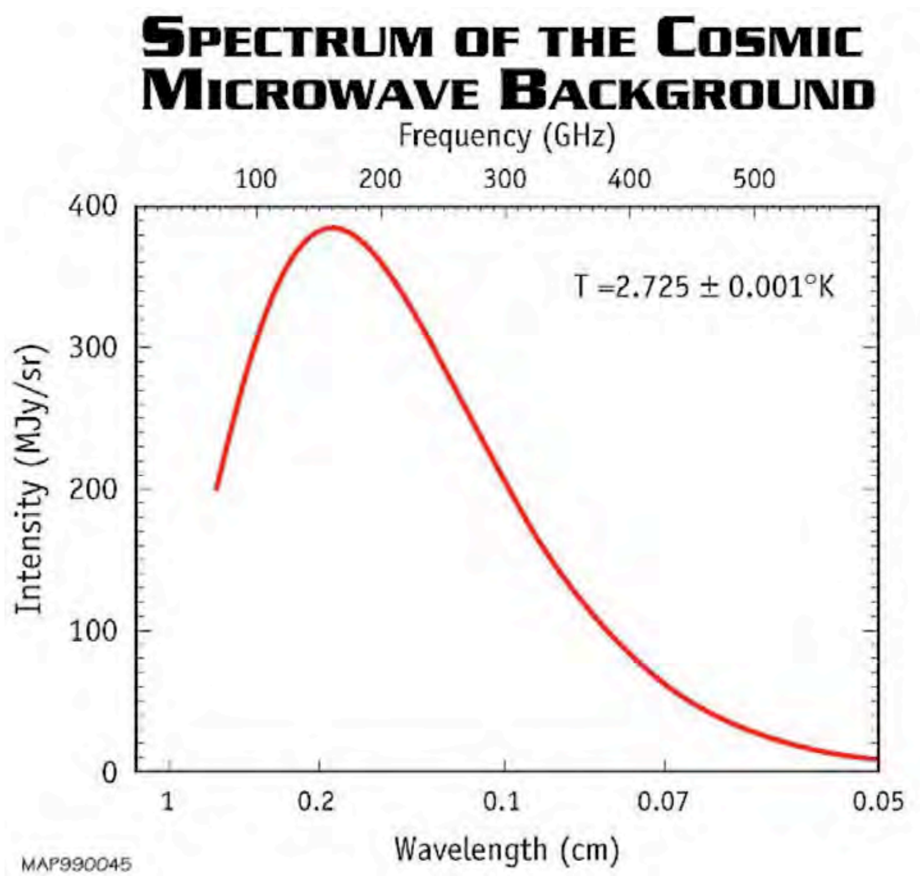


FIGURE 6 Spectrum of the cosmic microwave background radiation measured in the 1990s by the FIRAS instrument aboard the COBE satellite. (From the WMAP website.)

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

?

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

!

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

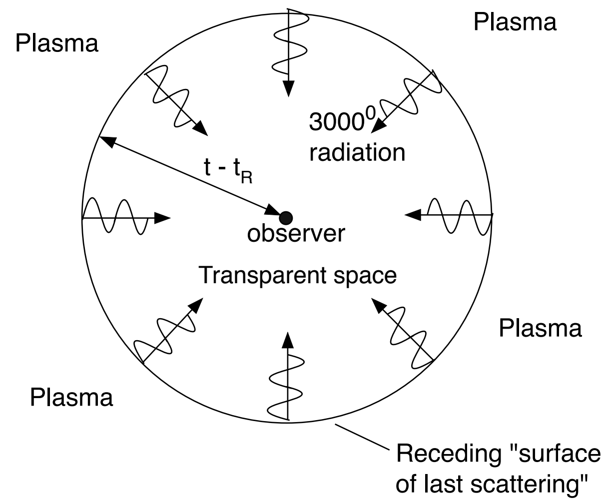


FIGURE 7 Observer's view of a non-expanding model Universe at $t - t_{\rm R}$, where $t_{\rm R}$ is the t-value, at which entire Universe becomes transparent to radiation. Looking outward, the observer cannot receive a signal from the entire Universe, but will see radiation released earlier from receding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

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

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

COBE satellite

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

WMAP satellite

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

Map of fluctuations: fingerprint of early Universe

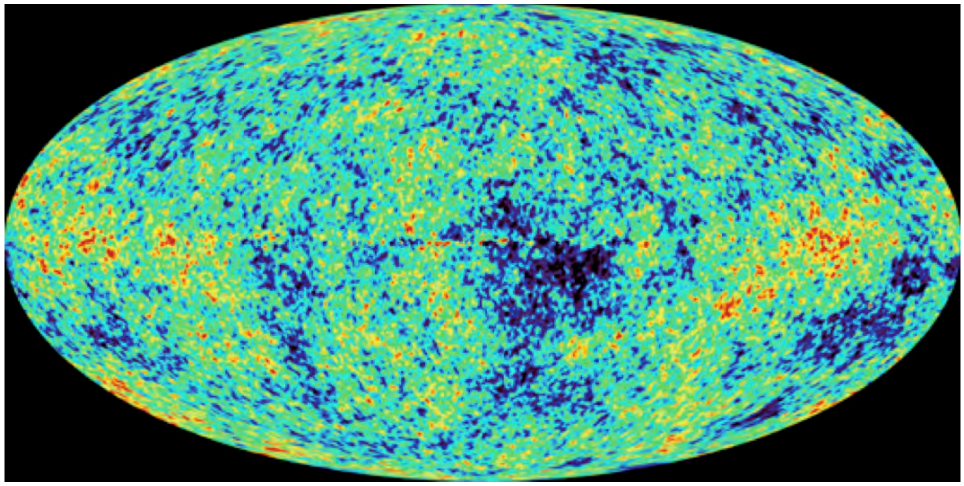


FIGURE 8 An all-sky map of the cosmic microwave background radiation at high contrast made by the WMAP satellite, with radiation from the nearby milky way stars removed. The oval is a projection of the entire sky onto the page. The colors in the original are "false colors" that indicate the temperature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

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

Fluctuations due to sound waves.

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

Sound waves: "standard ruler."

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

15.9 ■ EXPANSION HISTORY FROM STANDARD CANDLES

Finding t from redshift z

To find t, measure z.

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

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

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

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

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

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

which implies

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

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

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

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

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

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

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

then reintegrating:

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

Alternative model universes

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

15.10 ■ THE UNIVERSE NOW: THE OMEGA DIAGRAM

Squeeze the Universe model from all sides.

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

Squeezing the parameters

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

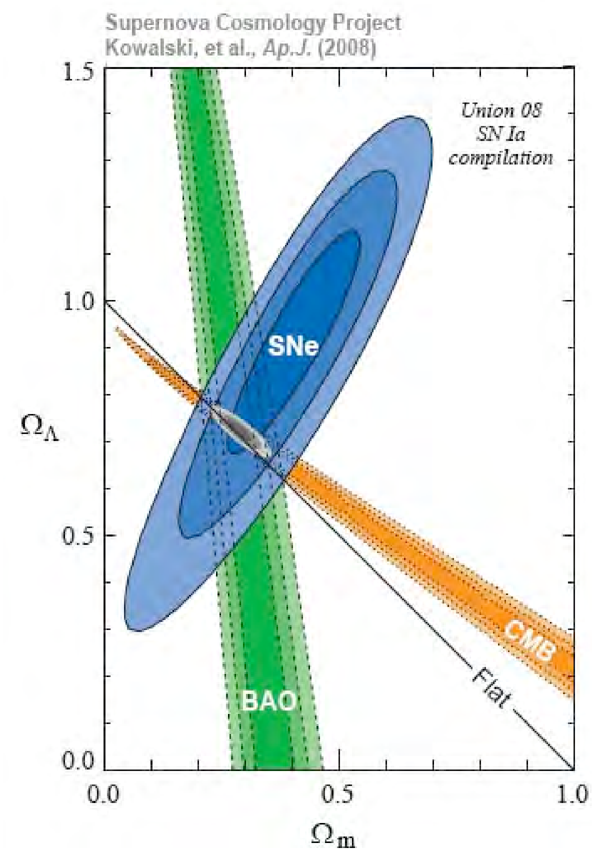


FIGURE 9 The Omega Diagram. Parameters Ω_m and Ω_Λ are called $\Omega_{mat,0}$ and $\Omega_{\Lambda,0}$ in this chapter. Relative amounts of matter and vacuum energy in the universe at present corresponds to the relatively tiny region of intersection of three sets of measurements: Type Ia supernovas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

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

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

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

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

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

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

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

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

15.11 ■ FIRE OR ICE?

You predict the fate of the Universe.

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

Fire or ice? **You** predict.

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

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

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

Download file name: Ch15Cosmology170510v1.pdf

Chapter 16. Gravitational Waves

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

Download file name: Ch16GravitationalWaves171018v1.pdf

CHAPTER 16

Gravitational Waves

Edmund Bertschinger & Edwin F. Taylor *

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

—Albert Einstein

16.1 ■ THE PREDICTION AND DISCOVERY OF GRAVITATIONAL WAVES

Gravitational wave: a tidal acceleration that propagates through spacetime.

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

Newton: Gravity propagates instantaneously.

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

Einstein: No signal propagates faster than light.

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

Compare gravitational waves to electromagnetic waves.

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

Gravitational waves propagate tidal accelerations.

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

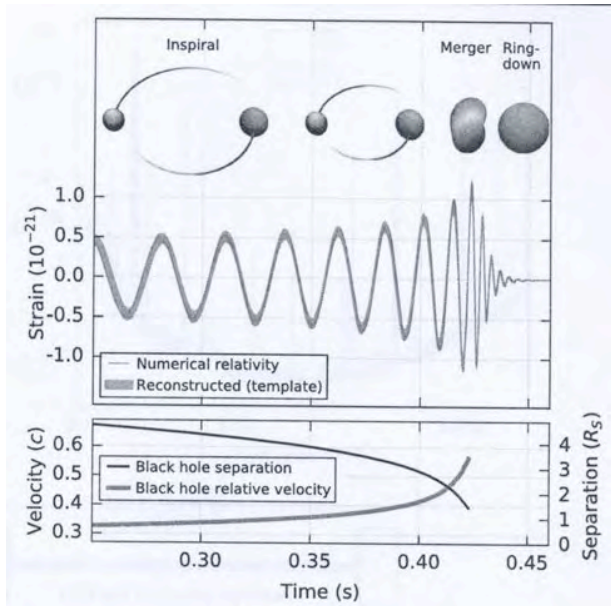


FIGURE 1 Predicted "chirp" of the gravitational wave as two black holes in a binary system merge. Frequency and amplitude increase, followed by a "ring down" due to oscillation of the merged black hole. The present chapter explains details of this figure.

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

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

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

Gravitational wave on Earth: An extremely small traveling tidal effect

Current gravitational wave detectors on Earth are interferometers in which light reflects back and forth between "free" test masses (mirrors) positioned at the ends of two perpendicular vacuum chambers. A passing gravitational wave changes the relative number of wavelengths along each leg, with a resulting change in interference between the two returning waves. The "free" test masses are hung from wires that are in turn supported on elaborate shock-absorbers to minimize the vibrations from passing trucks and even ocean waves crashing

Gravitational wave detectors are interferometers.

on a distant shore. The pendulum-like motions of these test masses are free enough to permit measurement of their change in separation due to tidal effects of a passing gravitational wave, caused by some remote gigantic distant event such as the coalescence of two black holes modeled in Figure **??**.

?

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

!

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

Comment 1. Why not "gravity wave"?

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

16.2 ■ GRAVITATIONAL WAVE METRIC

Tiny but significant departure from the inertial metric

Gravitational wave metric

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

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

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

$$\begin{aligned} 0 &\approx \Delta t^2_{\text{LIGO}} - [(1+h)^{1/2}\Delta x_{\text{LIGO}}]^2 - [(1-h)^{1/2}\Delta y_{\text{LIGO}}]^2 - \Delta z^2_{\text{LIGO}} \qquad (2)\\ &\approx \Delta t^2_{\text{LIGO}} - [(1+h/2)\Delta x_{\text{LIGO}}]^2 - [(1-h/2)\Delta y_{\text{LIGO}}]^2 - \Delta z^2_{\text{LIGO}} \qquad (h \ll 1) \end{aligned}$$

$h/2 =$ gravitational wave strain

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

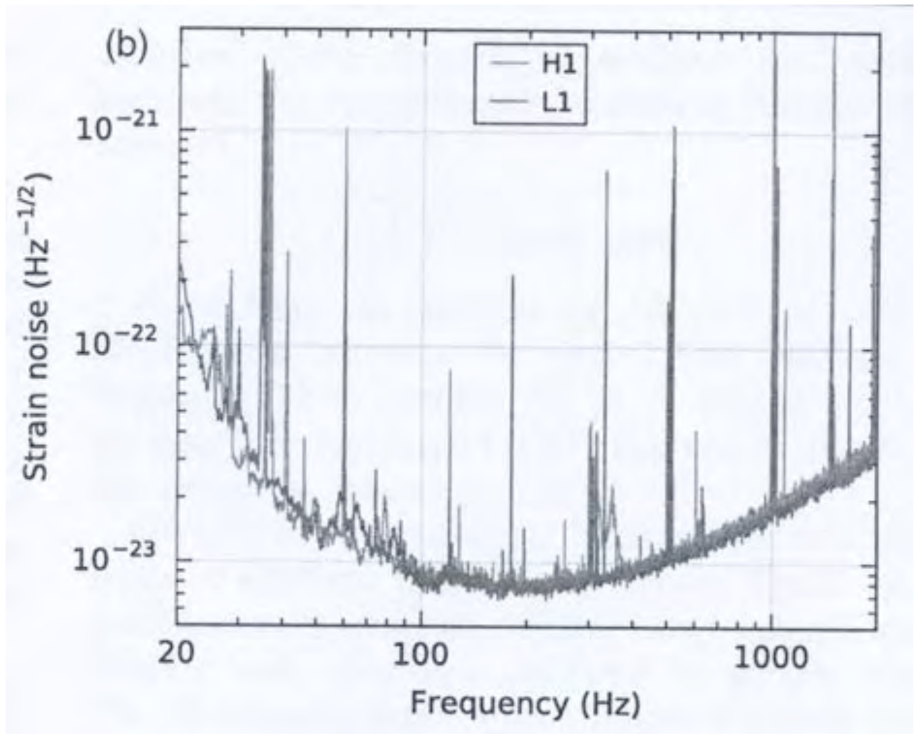


FIGURE 2 Strain noise of LIGO detectors at Hanford, Washington state (curve H1) and at Livingston, Louisiana state (curve L1) at the first detection of a gravitational wave on 14 September, 2015. On the vertical axis $h = 10^{-23}$, for example, means a fractional change in separation of 10^{-23} between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

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

?

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

!

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

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

LIGO gravity wave detector

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

Various kinds of noise

LIGO sensitivity

?

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

!

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

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

Einstein's equations become a wave equation.

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

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

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

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

Assume gravity wave moves in $+z$ direction.

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

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

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

LIGO sensitive 75 to 500 hertz

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

QUERY 1. Uniform h?

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

Analogy: draw global map coordinates on rubber sheet.

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

Gravitational wave distorts rubber sheet.

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

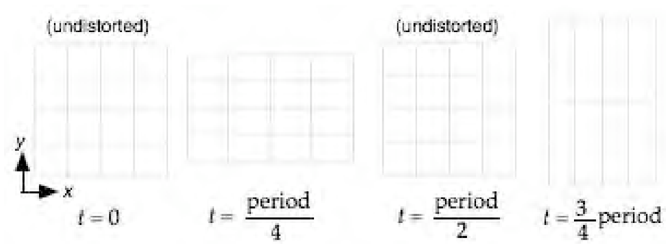


FIGURE 3 Change in shape (greatly exaggerated!) of the map coordinate grid at the same x, y location at four sequential t-values as a periodic gravitational wave passes through in the z-direction (perpendicular to the page). NOTE carefully: The x-axis is stretched while the y-axis is compressed and vice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

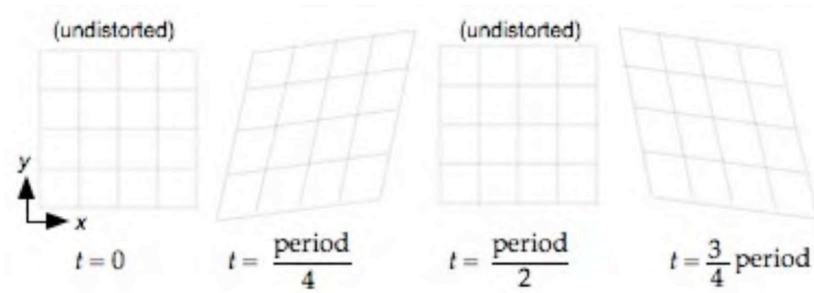


FIGURE 4 Effects of a periodic gravitational wave with polarization "orthogonal" to that of Figure 3 on the map grid in the xy plane. Note that the axes of compression and expansion are at 45 degrees from the x and y axes. All grids stay in the xy plane as they distort. As in Figure 3, the areas of the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

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

Map t read on clocks glued to the rubber sheet.

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

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

16.3 ■ SOURCES OF GRAVITATIONAL WAVES

Many sources; only one type leads to a clear prediction

No linear "antenna" for gravitational waves

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

Binary system emits gravity waves . . .

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

. . . whose amplitude is predictable.

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

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

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

From other sources: hard to predict.

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

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

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

16.4 ■ MOTION OF LIGHT IN MAP COORDINATES

Light reflected back and forth between mirrored test masses

LIGO is an interferometer.

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

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

Motion of light in map coordinates.

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

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

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

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

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

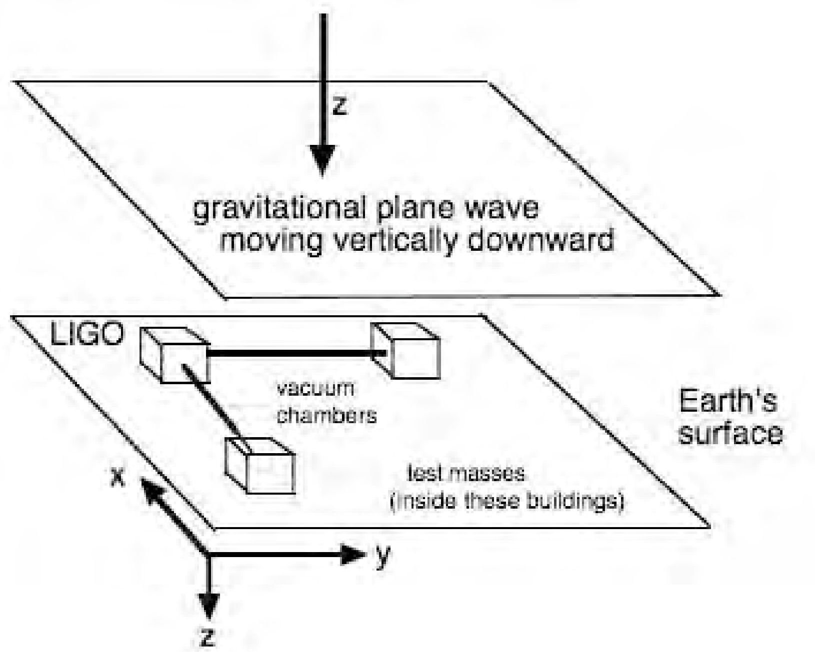


FIGURE 5 Perspective drawing of the relative orientation of legs of the LIGO interferometer lying in the x and y directions on the surface of Earth and the z-direction of the incident gravitational wave descending vertically. [Illustrator: Rotate lower plate and contents CCW 90 degrees, so corner box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

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

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

Apply this approximation to (7) to obtain

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

Gravitational wave modifies map speed of light.

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

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

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

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

16.5 ■ ZERO MOTION OF LIGO TEST MASSES IN MAP COORDINATES

"Obey the Principle of Maximal Aging!"

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

How does the test mass move?

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

Idealized case: Linear jogs out and back.

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

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

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

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

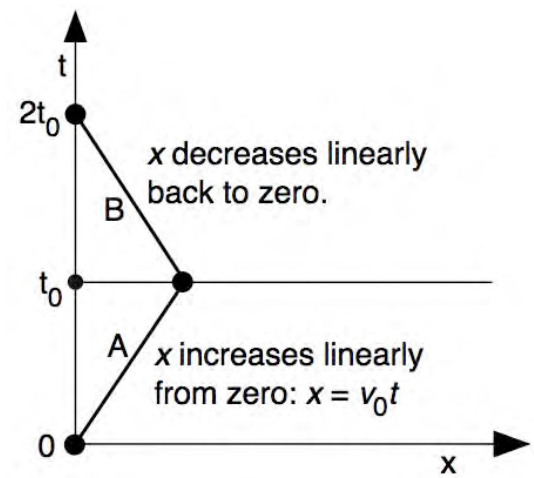


FIGURE 6 Trial worldline for a test mass; incremental departure from vertical line of a particle at rest. Segments A and B are very short.

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

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

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

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

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

so that

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

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

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

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

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

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

QUERY 4. Disproof of relativity? (optional)
"Aha!" exclaims Kristin Burgess. "Now I can disprove relativity once and for all. If the test mass *moves*, a passing gravitational wave can cause a kink in the worldline of the test mass as observed in the local inertial Earth frame. No kink appears in its worldline if the test mass is at rest. But if a worldline has a kink in it as observed in one inertial frame, it will have a kink in it as observed in all overlapping relatively-moving inertial frames. An observer in any such frame can detect this kink. So the *absence* of a kink tells me *and every other inertial observer* that the test mass is 'at rest'? We have found a way to determine absolute rest using a local experiment. Goodbye relativity!" Is Kristin right? (A detailed answer is beyond the scope of this book, but you can use some relevant generalizations drawn from what we already know to think about this paradox. As an analogy from flat-spacetime electromagnetism, think of a charged particle at rest in a purely magnetic field: The particle experiences no magnetic force. In contrast, when the same charged particle moves in the same frame, it may experience a magnetic force for some directions of motion.)

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

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

16.6 ■ DETECTION OF A GRAVITATIONAL WAVE BY LIGO

Make measurement in the local Earth frame.

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

Earth frame tied to LIGO slab

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

$$\Delta\tau^2 \approx \Delta t_{\text{Earth}}^2 - \Delta x_{\text{Earth}}^2 - \Delta y_{\text{Earth}}^2 - \Delta z_{\text{Earth}}^2 \tag{16}$$

Compare this with the approximate version of (1):

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

Earth frame coordinate differences

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

Compare (16) with (17) to yield:

$$\Delta t_{\text{Earth}} = \Delta t \tag{18}$$

$$\Delta x_{\text{Earth}} = (1+h)^{1/2}\Delta x \approx (1+\frac{h}{2})\Delta x \qquad h \ll 1 \tag{19}$$

$$\Delta y_{\text{Earth}} = (1-h)^{1/2}\Delta y \approx (1-\frac{h}{2})\Delta y \qquad h \ll 1 \tag{20}$$

$$\Delta z_{\text{Earth}} = \Delta z \tag{21}$$

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

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

Test masses move in Earth coordinates.

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

Light speed $= 1$ in local Earth frame.

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

Different Earth times along different legs

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

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

Time difference after N round trips.

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

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

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

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

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

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

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

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

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

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

QUERY 9. LIGO fails to detect a gravitational wave?

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

16.7 ■ BINARY SYSTEM AS A SOURCE OF GRAVITATIONAL WAVES

"Newtonian" source of gravitational waves

Unequal masses, each in circular orbit

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

Energy of the system.

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

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

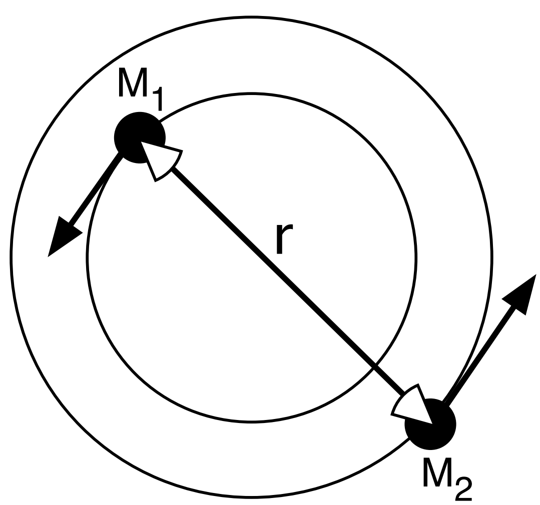


FIGURE 7 A binary system with each object in a circular path.

Rate of energy loss . . .

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

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

. . . derived from Einstein's equations.

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

QUERY 10. Energy and rate of energy loss

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

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

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

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

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

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

QUERY 11. Rate of change of radius

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

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

16.8 ■ GRAVITATIONAL WAVE AT EARTH DUE TO DISTANT BINARY SYSTEM

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

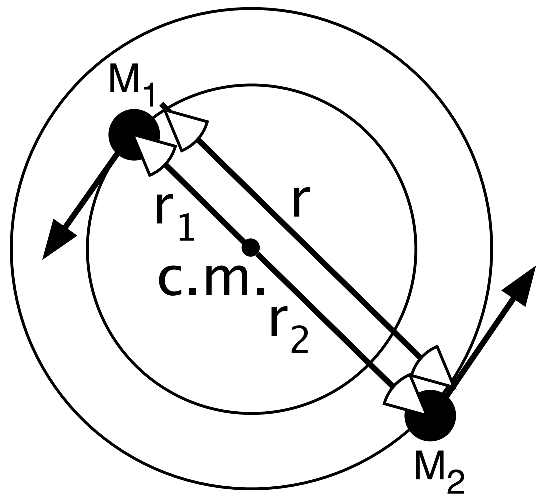


FIGURE 8 Figure 7 augmented to show the center of mass (c.m.) and orbital r-values of individual masses in the binary system.

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

Gravitational waveform

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

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

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

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

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

(b) they move at nonrelativistic speeds.

Additional assumptions are:

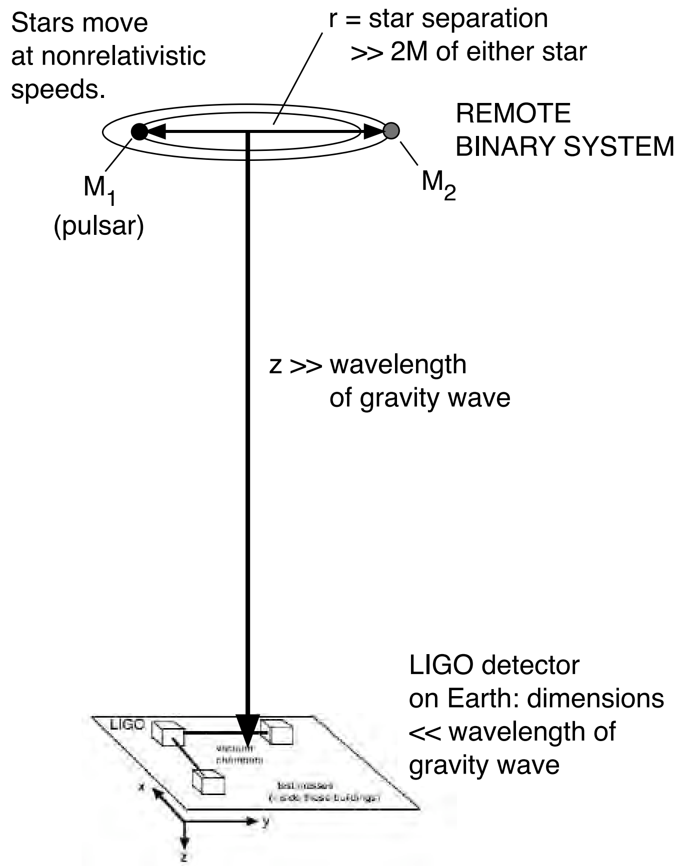


FIGURE 9 Schematic diagram, *not to scale*, showing notation and relative magnitudes for equation (29). The binary system and the LIGO detector lie in parallel planes.[Illustrator: See note in caption to Figure 5.]

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

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

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

. . . for one case

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

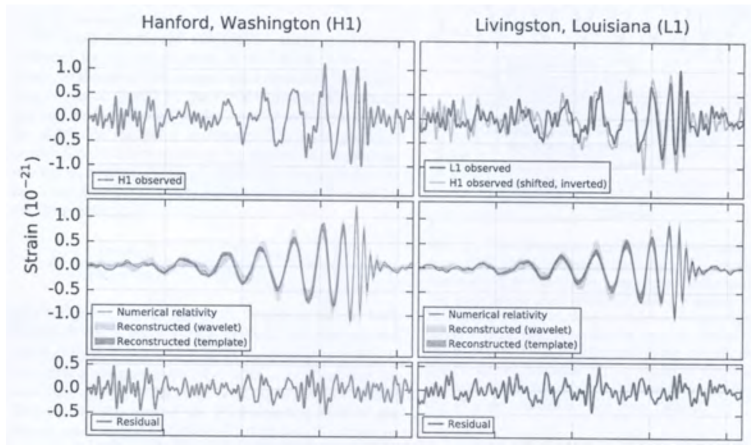


FIGURE 10 Detected "chirps" of the gravitational wave at two locations. The top row shows detected waveforms (superposed in the right-hand panel). The second row shows the cleaned-up image (again superposed). The bottom row displays "residuals," the noise deducted from images in the first row.

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

Detection requirements

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

QUERY 12. Amplitude of gravitational wave at Earth

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

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

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

QUERY 13. Frequency of emitted gravitational waves

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

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

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

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

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

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

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

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

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

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

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

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

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

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

"Chirp" at coalescence

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

16.9 ■ RESULTS FROM GRAVITATIONAL WAVE DETECTION; FUTURE PLANS

Unexpected details

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

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

How did observations lead to these results?

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

What are plans for future gravitational wave detections?

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

16.10 ■ REFERENCES

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

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

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

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

Download file name: Ch16GravitationalWaves171018v1.pdf

Chapter 17. Spinning Black Hole

- *What's the difference between a spinning and a non-spinning black hole?*
- *How does one spinning black hole differ from another spinning black hole?*
- *How fast can a black hole spin?*
- *Does the spin of a black hole keep me from falling to the singularity?*
- *If I can fall to the singularity, will that fall take longer than my lifetime?*
- *What local inertial frames are useful near a spinning black hole?*

Download file name: Ch17SpinningBlackHole171227v2.pdf

CHAPTER 17

Spinning Black Hole

Edmund Bertschinger & Edwin F. Taylor *

Black holes are macroscopic [large-scale] objects with masses varying from a few solar masses to billions of solar masses. When stationary and isolated, they are all, every single one of them, described exactly by the Doran solution. This is the only instance we have of an exact description of a macroscopic object. The only elements in the construction of black holes are our basic concepts of space and time. They are thus the most perfect macroscopic objects in the universe. They are the simplest objects as well.

—Subrahmanyan ("Chandra") Chandrasekhar [edited]

17.1 ■ THE AMAZING SPINNING BLACK HOLE

Add spin, multiply consequences

This and the following chapters describe the spinning black hole, which displays spectacular effects that outstrip most science fiction:

Spectacular physical effects

Some Physical Effects Near the Spinning Black Hole

1. There is a region outside the event horizon in which no rocket—no matter how powerful—can keep a spaceship stationary in our chosen global coordinates.
2. There is a region inside the event horizon in which a spaceship does *not* inevitably move toward the center, but can be repelled away from it (Chapter 18).
3. Stable orbits that do not cross the event horizon reach smaller r than do stable orbits for a non-spinning black hole. This result leads to dramatic general relativistic effects on the so-called **accretion disk** that circles around the spinning black hole (Chapter 18).

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

4. Unstable circular orbits exist in a region inside the event horizon and close to the singularity of the spinning black hole (Chapter 18).
5. Visual effects for the traveler near a spinning black hole are even wilder than those near the non-spinning black hole (Chapter 20).
6. The spinning black hole is an immense energy source, waiting to be tapped by an advanced civilization (Chapter 19).
7. The singularity of a spinning black hole is a ring through which a spaceship might pass undamaged (Chapter 21).
8. The spinning black hole may provide a gateway to other Universes (Chapter 21).

The present chapter sets the stage to describe these physical effects.

Why every black hole spins.

We expect every black hole to spin. Why? Because a group of stars or cloud of dust almost inevitably has *some* net spin angular momentum. When this system collapses to form a black hole, the spin rate increases in the same way that a spinning ice skater with arms extended rotates faster as she draws her arms inward. The skinnier the skater, the faster her final spin for a given initial angular momentum. The spinning black hole is the "skinniest possible astronomical skater." For this reason we expect (and have observational evidence) that black holes spin at a ferocious rate.

Apply the same toolkit to analyze the spinning black hole.

Comment 1. Have we wasted our time?
Since in Nature black holes spin, have we wasted our time studying the non-spinning black hole in the previous chapters of this book? Not at all! First, for most purposes the metric for the non-spinning black hole describes spacetime outside slowly rotating stars and planets such as Earth well enough so that we can use this metric to make predictions that are verified by observation. Second, we can generalize many of our non-spinning black hole tools to analyze the astonishing structure of the spinning black hole. Third, our analysis of the spinning black hole follows the same sequence as our analysis of the non-spinning black hole. Fourth, we can use our non-spinning black hole results as a limiting case to check predictions for the spinning black hole. Fifth—and most important—by now we have extensive experience using the power of the global metric plus the Principle of Maximal Aging to predict results of measurements and observations carried out near the spinning black hole.

Just two numbers: mass and spin

Spin parameter a

An isolated, uncharged spinning black hole is completely specified by just two quantities: its mass and its spin angular momentum. To avoid confusion between the rotational angular momentum of the spinning black hole (with mass M) and the orbital angular momentum of a stone (with mass m) around the black hole, we use the symbol J for the angular momentum of the spinning black hole and write J/M for this angular momentum per unit mass. The ratio J/M appears so often in the analysis that we define the lower-case italic a, called the **spin parameter**, which also has the unit of meters:

$$a \equiv \frac{J}{M} \qquad \text{(black hole spin parameter, unit of meters)} \qquad (1)$$

Note that the black hole spin parameter a has nothing to do with $a(t)$, the scale factor of the Universe defined in Section 15.2. We have run out of letters!

Think of an isolated star that collapses into a black hole while keeping its angular momentum constant. Its rotation rate will increase enormously. Look at the spinning black hole from either one side or the other. There is always a side for which the spin will be counterclockwise. We *choose* both J and a to be positive quantities for that counterclockwise spin direction. Now, the smallest value of J and a is zero. What is the largest possible value of each? In Query 5 you show that the ranges fit the following inequalities:

$$0 \leq J \leq M^2 \qquad \text{(range of spin angular momentum } J\text{, units of meters}^2\text{)} \quad (2)$$

$$0 \leq a \leq M \qquad \text{(range of spin parameter } a\text{, units of meters)} \qquad (3)$$

17.2 ■ THE DORAN GLOBAL METRIC

Eighty-five years after Einstein's equations!

Karl Schwarzschild derived his global metric for the non-spinning black hole less than a month after Einstein published his field equations. In contrast, not until 1963—forty-eight years later—did Roy P. Kerr publish a paper with a title that begins, "Gravitational Field of a Spinning Mass . . .". Brandon Carter and others showed that Kerr's metric describes not just a spinning mass but a spinning black hole. Only in the year 2000—eighty-five years after Einstein derived his equations—did Chris Doran express Kerr's results in the global metric that we use to analyze the spinning black hole. As usual, we restrict global coordinates and their metric to a slice through the center of the black hole. The non-spinning black hole is spherically symmetric, so this slice through the center can have any orientation. For the spinning black hole, however, we choose the slice in the symmetry plane of the equator, perpendicular to the axis of rotation. In one of many tetrad forms—the sum and difference of squares (Section 7.6)—the **Doran metric** reads:

Doran global metric

$$d\tau^2 = dT^2 - \left[\left(\frac{r^2}{r^2+a^2}\right)^{1/2} dr + \left(\frac{2M}{r}\right)^{1/2}(dT - a d\Phi)\right]^2 - \left(r^2+a^2\right) d\Phi^2 \qquad (4)$$

$$-\infty < T < \infty, \quad 0 < r < \infty, \quad 0 \leq \Phi < 2\pi \qquad \text{(Doran, equatorial plane)}$$

In Query 1 you multiply out (4) to obtain the Doran metric in expanded form:

$$d\tau^2 = \left(1 - \frac{2M}{r}\right) dT^2 - 2\left(\frac{2Mr}{r^2+a^2}\right)^{1/2} dT dr + 2a\left(\frac{2M}{r}\right) dT d\Phi \tag{5}$$
$$+ 2a\left(\frac{2Mr}{r^2+a^2}\right)^{1/2} dr d\Phi - \left(\frac{r^2}{r^2+a^2}\right) dr^2 - R^2 d\Phi^2$$
$$-\infty < T < \infty, \quad 0 < r < \infty, \quad 0 \le \Phi < 2\pi \qquad \text{(Doran, equatorial plane)}$$

The expanded Doran metric (5) contains every possible cross term—sorry! It also contains a new expression R, a function of both r and a that we call the **reduced circumference**:

Define R

$$R^2 \equiv r^2 + a^2 + \frac{2Ma^2}{r} \qquad (R = \text{reduced circumference}) \tag{6}$$

QUERY 1. Doran metric reduces to global rain metric for non-spinning black hole.

A. Let $a \to 0$ in the expanded Doran metric (5) for the spinning black hole and compare the result with the global rain metric for the non-spinning black hole, equation (32) in Section 7.5.

B. Now *demand* that the two global metrics of Item A be identical. Show that the result is that $d\Phi \to d\phi$ when $a \to 0$.

Figure 1 plots the reduced circumference R as a function of r for sample values of the spin parameter a. As $r \to \infty$ all curves converge asymptotically toward the curve for $a = 0$, the non-spinning black hole. Why do we call R the reduced circumference? Let $dr = dT = 0$. Then global metric (5) reduces to

$$d\tau^2 = -d\sigma^2 = -R^2 d\Phi^2 \qquad \text{(Doran: } dr = dT = 0\text{)} \tag{7}$$

or $\sigma = 2\pi R$ for a complete circle at fixed r around the spinning black hole. This justifies calling R the *reduced circumference.*

?

Objection 1. *Why not use (6) to eliminate r from metrics (4) and (5) and use R exclusively?*

!

Because R violates the rule that global coordinates must label each event uniquely (Section 5.8). Figure 1 shows that for every value of R greater than its minimum there correspond two different values of r.

?

Objection 2. *Why in the world are there two values of r for each value of the reduced circumference? Geometry does not allow this!*

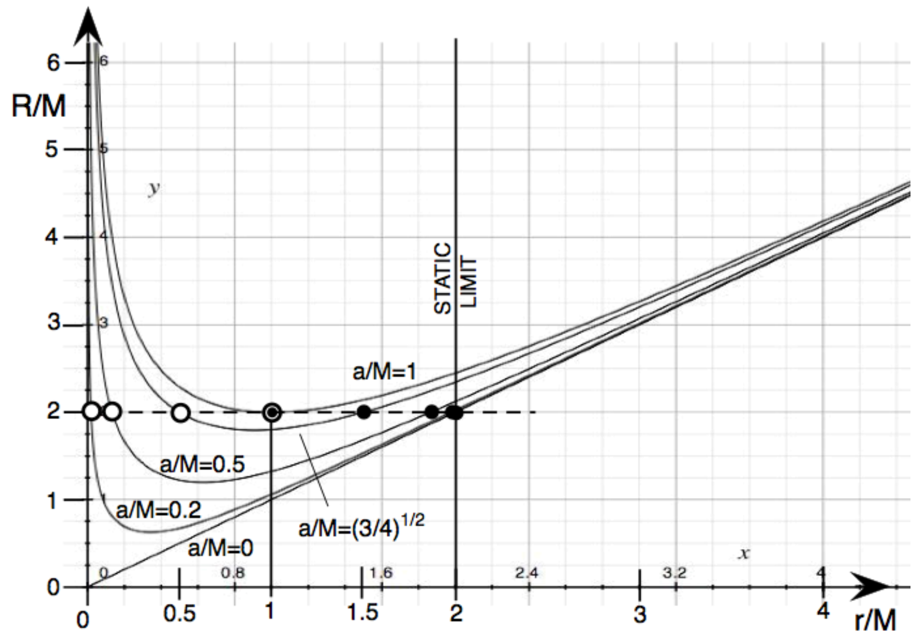


FIGURE 1 Plot of reduced circumference R vs. r for several values of the spin parameter a. Location of the static limit $r_S/M = 2$, equation (9), does not depend on spin. Section 17.3 and Figure 2 describe the significance of little filled and open circles along the dashed horizontal line $R/M = 2$.

Ah! You mean that *Euclidean geometry* does not allow this. Inside the static limit, especially, spacetime is radically distorted; Euclidean flat-space geometry simply does not apply there.

QUERY 2. Limiting cases of the Doran metric

A. Show that as $r \to \infty$ the Doran metric (4) becomes the metric for flat spacetime.

B. Write down the Doran metric (5) for the maximum-spin black hole ($a/M = 1$) and the expression for R_{max} in this case.

Comment 2. You do the math (if you wish).
At this point in the book some derivations become so algebraically complicated that we omit them, while leaving a skimpy trail to guide you if you choose to carry out these derivations yourself. Instead, we focus on results and predictions: What locations near the spinning black hole can we explore and still return home unharmed? What do we see and feel on the way? Which predictions can we verify now, and which must we leave to our descendants? Dive into the complications; enjoy the payoffs!

17.3 ■ A STONE'S THROW

Where you can go; how you can move

Now apply the Doran metric to two adjacent events that lie along the worldline of a stone. What commands does spacetime give to the stone through the metric? We examine two cases.

THE STONE AT REST IN DORAN COORDINATES

Where can the stone stand still in Doran coordinates?

The simplest possible motion of a stone is no motion at all: to stand still in global space coordinates. Where can the stone stand still? Expressed more carefully, can two adjacent events along the stone's worldline have $dr = d\Phi = 0$? To find out, put these conditions into the Doran metric:

$$d\tau^2 = \left(1 - \frac{2M}{r}\right) dT^2 \qquad (dr = d\Phi = 0) \tag{8}$$

Static Limit at $r_S = 2M$

Wristwatch time must be real along the worldline of a stone, so both sides of (8) must be positive. This tells us that the stone cannot remain at rest in Doran global coordinates when $r < 2M$. Does this place the event horizon of the spinning black hole at $r = 2M$? No. In what follows we discover that, for the spinning black hole, the event horizon lies inside $r = 2M$. For the minute, simply ask what equation (8) does say: Inside $r = 2M$ the stone *must* move in either r or Φ or both; the stone cannot remain static in Doran coordinates. Therefore we give this value of r the label **static limit** with the subscript S. Equation (8) shows that the static limit has the same value $r_S = 2M$ for all values of the spin parameter a:

$$r_S = 2M \qquad (r\text{-coordinate of } \textbf{static limit} \text{ for all } a) \tag{9}$$

THE STONE WITH $dr = 0$ IN DORAN COORDINATES

Now loosen restrictions on the stone. Where can the stone remain at fixed r-value but move in Φ? To find out, set $dr = 0$ in the global metric (5) for two adjacent events along the stone's worldline:

$$d\tau^2 = \left(1 - \frac{2M}{r}\right) dT^2 + 2\left(\frac{2Ma}{r}\right) dT d\Phi - R^2 d\Phi^2 \qquad (dr = 0) \tag{10}$$

We want a global metric in tetrad form—with no cross-term. Rewrite equation (10) as the sum and difference of squares on the right side. There are only two global coordinates in (10), so construct a linear combination of the form $dX = d\Phi - \omega dT$ and choose the function ω to eliminate the cross term in the metric. Substitute $d\Phi = dX + \omega dT$ into (10) and rearrange the result to obtain:

$$d\tau^2 = \left(1 - \frac{2M}{r} + \frac{4Ma\omega}{r} - \omega^2 R^2\right) + 2\left(\frac{2Ma}{r} - \omega R^2\right) dX dT - R^2 dX^2 \tag{11}$$

To eliminate the cross term, choose the function $\omega(r)$ to be

$$\omega(r) \equiv \frac{2Ma}{rR^2} \qquad \text{omega function} \tag{12}$$

With this choice of $\omega(r)$, the global metric for constant-r motion takes the tetrad form:

$$d\tau^2 = \left[1 - \frac{2M}{r} + \frac{4M^2a^2}{r^2R^2}\right] dT^2 - R^2\left[d\Phi - \omega dT\right]^2 \qquad (dr = 0) \tag{13}$$

Simplify the coefficient of dT^2 as follows:

$$\begin{aligned} 1 - \frac{2M}{r} + \frac{4M^2a^2}{r^2R^2} &\equiv \frac{\left(1 - \frac{2M}{r}\right)R^2 + \frac{4M^2a^2}{r^2}}{R^2} \\ &= \frac{\left(1 - \frac{2M}{r}\right)\left(r^2 + a^2 + \frac{2Ma^2}{r}\right) + \frac{4M^2a^2}{r^2}}{R^2} \\ &= \frac{r^2 + a^2 - 2Mr - \cancel{\frac{2Ma^2}{r}} + \cancel{\frac{2Ma^2}{r}} - \cancel{\frac{4M^2a^2}{r^2}} + \cancel{\frac{4M^2a^2}{r^2}}}{R^2} \\ &= \frac{r^2 - 2Mr + a^2}{R^2} = \left(\frac{rH}{R}\right)^2 \end{aligned} \tag{14}$$

Define: **Horizon function** H.

where we define the **horizon function** $H(r)$ from the last line of equation (14):

$$H^2(r) \equiv \frac{r^2 - 2Mr + a^2}{r^2} = \frac{(r - r_{\rm EH})(r - r_{\rm CH})}{r^2} \qquad (H \equiv \text{horizon function}) \tag{15}$$

Note that when $a \rightarrow 0$, then $H^2(r) \rightarrow (1 - 2M/r)$; so we can think of the common expression $(1 - 2M/r)$ for the non-spinning black hole to be a special case of $H^2(r)$.

Comment 3. Horizon function H is different from Hubble parameter. The horizon function H defined in (15) has nothing to do with the Hubble parameter H defined in Chapter 15. There are only so many letters in any alphabet; in this case we recycle the symbol H.

Use the new horizon function H to give the Doran metric (13) with $dr = 0$ the simple form:

$$d\tau^2 = \left(\frac{rH}{R}\right)^2 dT^2 - R^2\left[d\Phi - \omega(r)dT\right]^2 \qquad (dr = 0) \tag{16}$$

The roots of the numerator in expression (15) for H^2 introduce two special values of the r-coordinate, which we call the **event horizon** and the **Cauchy horizon**:

$$\frac{r_{\rm EH}}{M} \equiv 1 + \left(1 - \frac{a^2}{M^2}\right)^{1/2} \qquad \text{(event horizon)} \qquad (17)$$

$$\frac{r_{\rm CH}}{M} \equiv 1 - \left(1 - \frac{a^2}{M^2}\right)^{1/2} \qquad \text{(Cauchy horizon)} \qquad (18)$$

Comment 4. Augustin-Louis Cauchy
Mathematician Augustin-Louis Cauchy (1789 to 1852) derived results over the entire range of then-current mathematics and mathematical physics. Cauchy did not discover black holes or their horizons, but his work on differential equations is relevant to the properties of horizons.

Meaning of an horizon

How do we justify calling these special r-coordinates *horizons*? What do we mean by an horizon for the black hole? Look closely at the right side of equation (16). The second term is always negative unless $d\Phi = \omega dT$. Let's assume this equality, because it gives us the greatest possible latitude to have a worldline with $d\tau^2 > 0$ and $dr = 0$. The resulting equation tells us immediately that such a worldline is possible if and only if $(rH/R)^2 > 0$ or $H^2 > 0$. If this is not so, that is if $H^2 < 0$, then a stone *must* move in the r-coordinate. Why? Because if it does not move, that is if $dr/d\tau = 0$, then $d\tau^2 < 0$, which is forbidden along the worldline of a stone. (It will also move in the Φ-coordinate, because we just assumed that $d\Phi/dT = \omega$.) See Figure 2.

Question: How to define an event horizon?

How do we find an event horizon? A full definition of an event horizon involves examining the propagation of light, which we describe in Chapter 20. However a simplified (and in this case valid) definition can use the orbits of stones.

Answer: r-surface on one side of which nothing can remain at constant r.

We ask whether a stone can remain at constant r. The event horizon is the boundary where the answer changes from "Yes" to "No". For the *non-spinning* black hole, nothing can remain at constant r between $r = 2M$ and the singularity, so we label $r = 2M$ the event horizon. The *spinning* black hole is more complicated: Nothing can remain at constant r where $H^2 < 0$, which is the case between the upper event horizon and the lower Cauchy horizon. At r values between the Cauchy horizon and the singularity, amazingly, a stone can again remain at constant r-value. How can a free stone do this? One way is to travel in a circular orbit. Chapter 18 describes circular orbits of a stone, including circular orbits at r-values inside the Cauchy horizon and down almost to $r = 0$!

QUERY 3. Verify horizon equations
Solve the quadratic equation $r^2 - 2Mr + a^2 = 0$ from the numerator of equation (15). Show the roots are $r_{\rm EH}$ and $r_{\rm CH}$ in equations (17) and (18).

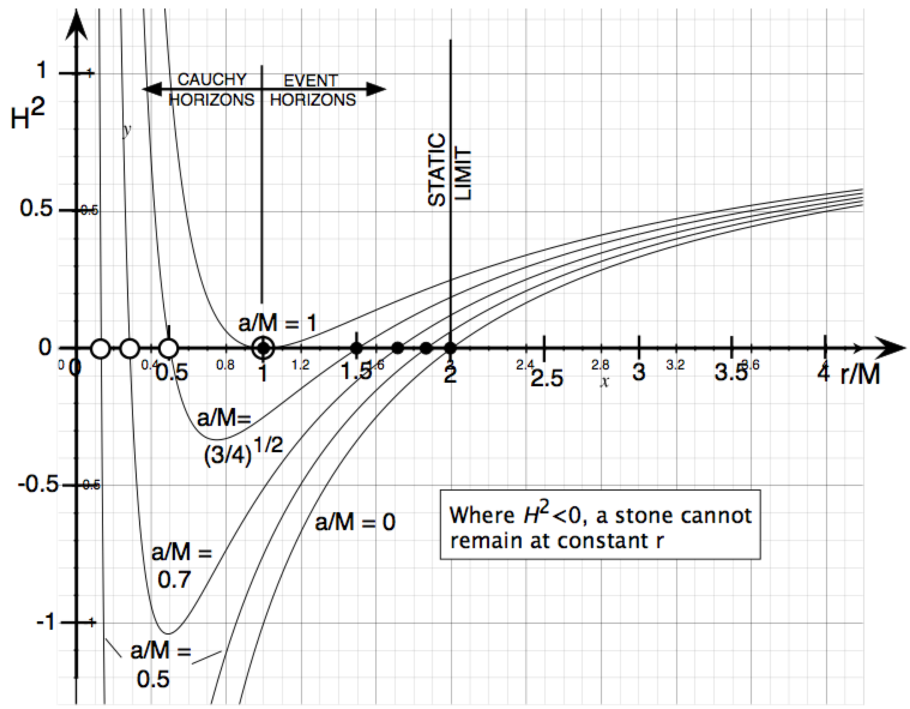


FIGURE 2 Plot of the function H^2 vs. r for selected values of a. Equation (16) says that when $d\Phi/dT = \omega(r)$, adjacent events along a stone's worldline are timelike—and that worldline is possible—only when $H^2 > 0$ in this plot. Little filled circles locate the event horizon for a given value of a, and little open circles locate the corresponding Cauchy horizons. For $a/M = 1$ these two horizons coincide at $r/M = 1$. Review similar symbols in Figure 1.

Sequence of horizons and static limit

Figure 3 plots r-values of event and Cauchy horizons for different spin parameters a. Equations (17) and (18) plus (9) lead to the following inequalities, also displayed in the figure:

$$0 \leq r_{\text{CH}} \leq M \leq r_{\text{EH}} \leq r_{\text{S}} = 2M \qquad (19)$$

QUERY 4. All horizons have reduced circumference $R = 2M$.
Substitute $r/M = 1 \pm (1 - a^2/M^2)^{1/2}$ from (17) and (18) into equation (6) for R^2 and verify that all horizons have reduced circumference $R = 2M$, as shown in Figure 1.

Prepare for local inertial frames

We can use any global metric expressed in tetrad form (Section 7.6) to define a local inertial frame. The next three sections prepare the way for us to

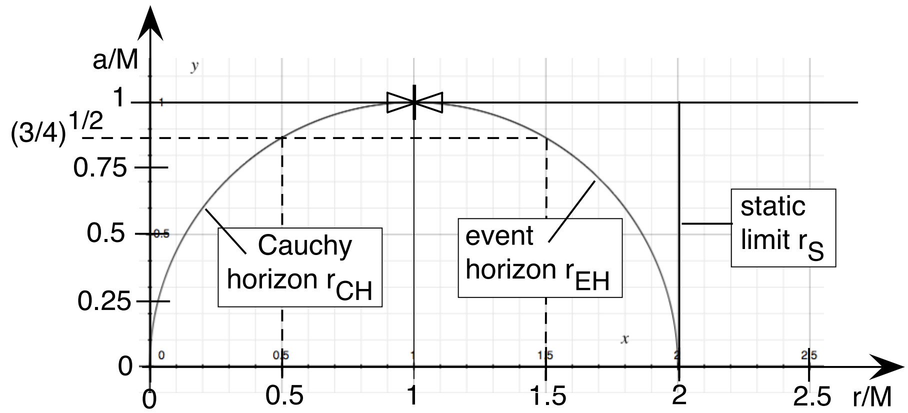


FIGURE 3 The r-values of the Cauchy and event horizons for different values of spin parameter a. Dashed lines are for $a/M = (3/4)^{1/2}$, for which $r_{\rm EH}/M = 1.5$ and $r_{\rm CH}/M = 0.5$. The static limit $r_{\rm S}/M = 2$ is independent of a. As the spin parameter a increases from zero, the event horizon drops from $r_{\rm EH}/M = 2$ to $r_{\rm EH}/M = 1$, while the Cauchy horizon emerges from the singularity and rises to the same final $r_{\rm CH}/M = 1$.

construct three useful local inertial frames from which to make measurements and observations near the spinning black hole.

QUERY 5. Horizons do not exist if $a > M$.

A. Show that if $a > M$, then $H^2(r) > 0$ everywhere.

B. Show that in this case, and for any given r, a stone can remain at that r while having $d\tau^2 > 0$ along its worldline.

C. Show that in this case a stone can move inward and outward from any r, while having $d\tau^2 > 0$.

D. Explain why this means that there is no event horizon.

Your analysis in this Query justifies the upper limit for a in relation (3).

We now describe the motion of a stone in the equatorial plane of the spinning black hole. For this we need global coordinate expressions for the stone's map energy and map angular momentum. Derivations of these expressions are closely similar to earlier derivations of similar quantities in Chapters 6, 8, and 9, so we relegate them to appendices in Sections 17.9 and 17.10. Here are the results:

$$\frac{E}{m} = \left(1 - \frac{2M}{r}\right)\frac{dT}{d\tau} - \left(\frac{2Mr}{r^2 + a^2}\right)^{1/2}\frac{dr}{d\tau} + \frac{2Ma}{r}\frac{d\Phi}{d\tau} \tag{20}$$

$$\frac{L}{m} = R^2 \frac{d\Phi}{d\tau} - \frac{2Ma}{r}\frac{dT}{d\tau} - a\left(\frac{2Mr}{r^2+a^2}\right)^{1/2}\frac{dr}{d\tau} \qquad (21)$$

QUERY 6. Map energy and map angular momentum for the non-spinning black hole. For $a \to 0$, show that (20) reduces to equation (35) in Section 7.5 for E/m and (21) reduces to equation (10) in Section 8.2 for L/m for a stone near a non-spinning black hole.

17.4 ■ THE RAINDROP

A simple case that gives deep insight

Major equations in this chapter look complicated. In contrast, John Wheeler insisted that "everything important is utterly simple" (Appendix I. Wheeler's Rules). We now examine an important case, the raindrop, and find that its equations of motion are indeed utterly simple.

Definition of the **raindrop**

The raindrop, remember, is a free stone that drops from initial rest starting at very large r. "Initial rest" means that $dr/d\tau \to 0$ and $d\Phi/d\tau \to 0$ as $r \to \infty$. In addition, equation (8) says that $dT \to d\tau$ as $r \to \infty$, and from (20) and (21), the raindrop's map energy and map angular momentum become:

$$\frac{E}{m} = 1 \qquad \text{and} \qquad \frac{L}{m} = 0 \qquad \text{(raindrop)} \qquad (22)$$

Doran: Make raindrop equations simple.

In Query 2 you showed that in the limit $a \to 0$, the Doran metric for the spinning black hole reduces to the global rain metric for the non-spinning black hole. Exercise 2 in Section 7.10 analyzed the raindrop for the non-spinning black hole in global rain coordinates and found that $d\phi/d\tau = 0$ along its worldline. Chris Doran *chose* global coordinates Φ and T so that the raindrop worldline lies along constant Φ—that is $d\Phi/d\tau = 0$ along the raindrop worldline—and the raindrop wristwatch ticks at the same rate that global T passes—that is, $dT/d\tau = 1$ along the raindrop worldline. For the raindrop, then, equations (20), (21), and (22) lead to:

$$\frac{E}{m} = 1 = \left(1 - \frac{2M}{r}\right) - \left(\frac{2Mr}{r^2+a^2}\right)^{1/2}\frac{dr}{d\tau} \qquad \text{(raindrop)} \qquad (23)$$

$$\frac{L}{m} = 0 = -\frac{2Ma}{r} - a\left(\frac{2Mr}{r^2+a^2}\right)^{1/2}\frac{dr}{d\tau} \qquad \text{(raindrop)} \qquad (24)$$

You can solve either one of these equations to find the same expression for $dr/d\tau$:

$$\frac{dr}{d\tau} = -\left(\frac{2M}{r}\right)^{1/2}\left(\frac{r^2+a^2}{r^2}\right)^{1/2} \qquad \text{(raindrop)} \qquad (25)$$

Raindrop equations of motion

With Chris Doran's raindrop-related choice of global coordinates, the equations of motion for the raindrop become:

$$\frac{dr}{d\tau} = -\left(\frac{2M}{r}\right)^{1/2}\left(\frac{r^2+a^2}{r^2}\right)^{1/2} \qquad \text{(raindrop)} \qquad (26)$$

$$\frac{dT}{d\tau} = 1 \qquad \text{(raindrop)} \qquad (27)$$

$$\frac{d\Phi}{d\tau} = 0 \qquad \text{(raindrop)} \qquad (28)$$

Raindrop wristwatch time from r_0 to r

How much time does it take, on the raindrop's wristwatch, to fall from an initial global coordinate r_0 to a lower value r? (*Slogan:* "How many ticks of a raindrop clock if a raindrop could tick tock?") To answer this question, integrate equation (26):

$$\tau[r_0 \rightarrow r] = \left(\frac{1}{2M}\right)^{1/2}\int_r^{r_0}\left(\frac{r^{*2}}{r^{*2}+a^2}\right)^{1/2} r^{*1/2}dr^* \qquad \text{(raindrop)} \quad (29)$$

where r^* is a variable of integration. The right side of this equation does not have a closed-form solution, so we integrate it numerically. Figure 4 plots some results and compares these curves with one curve for $a = 0$ in Section 7.5.

QUERY 7. Arrive sooner at the singulariy From a quick examination of equation (29), show that as you ride a raindrop into a spinning black hole,

A. your wristwatch time to fall from a given r to the singularity is less than for a non-spinning black hole, and

B. your wristwatch time to fall from a higher r_0 to a lower r when both are far from the black hole is the same as for a non-spinning black hole.

From (26) through (28), it follows immediately that the "global coordinate displacement" of the raindrop has the components:

$$\frac{dr}{dT} \equiv \frac{dr}{d\tau}\frac{d\tau}{dT} = -\left(\frac{2M}{r}\right)^{1/2}\left(\frac{r^2+a^2}{r^2}\right)^{1/2} \qquad \text{(raindrop)} \qquad (30)$$

$$\frac{d\Phi}{dT} \equiv \frac{d\Phi}{d\tau}\frac{d\tau}{dT} = 0 \qquad \text{(raindrop)} \qquad (31)$$

Comment 5. Goodbye "radial"
Does the raindrop follow a "radial" path down to the singularity of a spinning black hole? No. The word "radial" no longer describes motion near the spinning black hole.

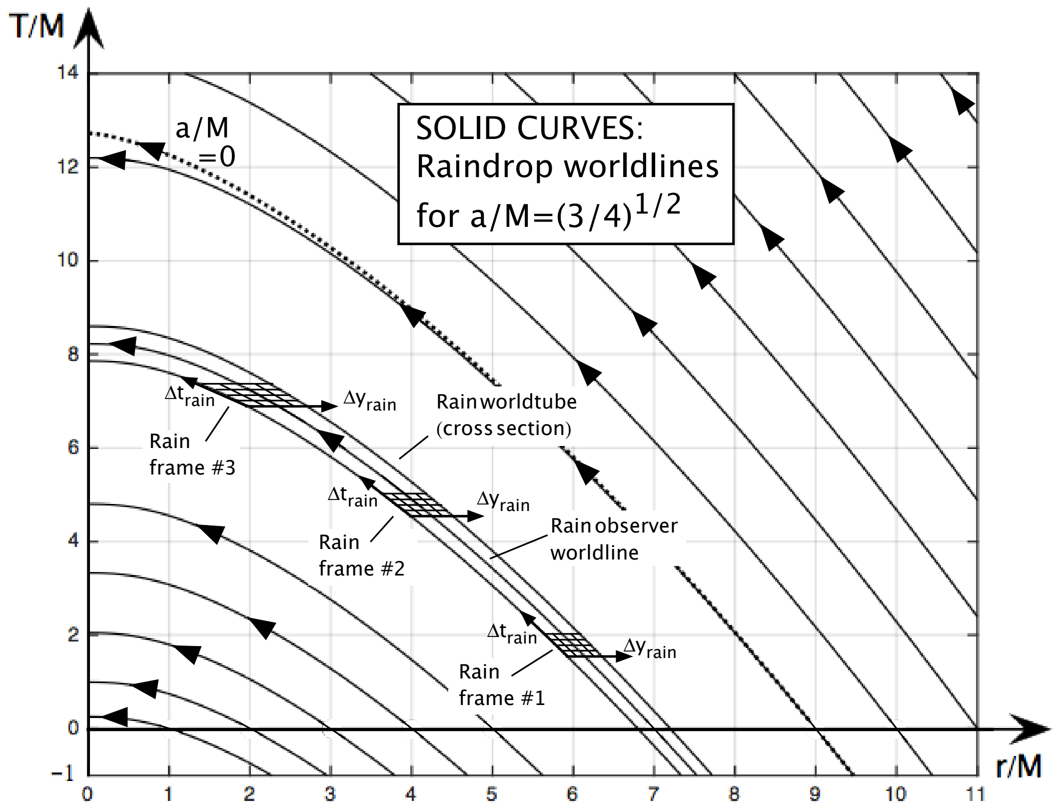


FIGURE 4 Solid curves: raindrop worldlines for a black hole with spin $a/M = (3/4)^{1/2}$, the numerical solution of equation (29), plotted on an $[r, T]$ slice. All these worldlines have the same shape and are simply displaced vertically with respect to one another. Note that these worldlines are continuous through the event and Cauchy horizons at $r_{\mathrm{EH}}/M = 1.5$ and $r_{\mathrm{CH}}/M = 0.5$. Around one of these worldlines we construct, in cross section, a worldtube that bounds local rain frames through which that rain observer passes. For local rain frame coordinates, see Section 17.7. Dotted curve for comparison: raindrop worldline for non-spinning black hole ($a/M = 0$); compare Figure 3, Section 7.5 for $a/M = 0$.

For the non-spinning black hole, we can still hang on to the intuitive term "radial," because the symmetry of that black hole demands that a raindrop—with zero map angular momentum—can veer neither clockwise nor counterclockwise as it descends.

Not so for the spinning black hole, which breaks the clockwise-counterclockwise symmetry. A stone with $dr/dT = d\Phi/dT = 0$ FINISH THIS COMMENT

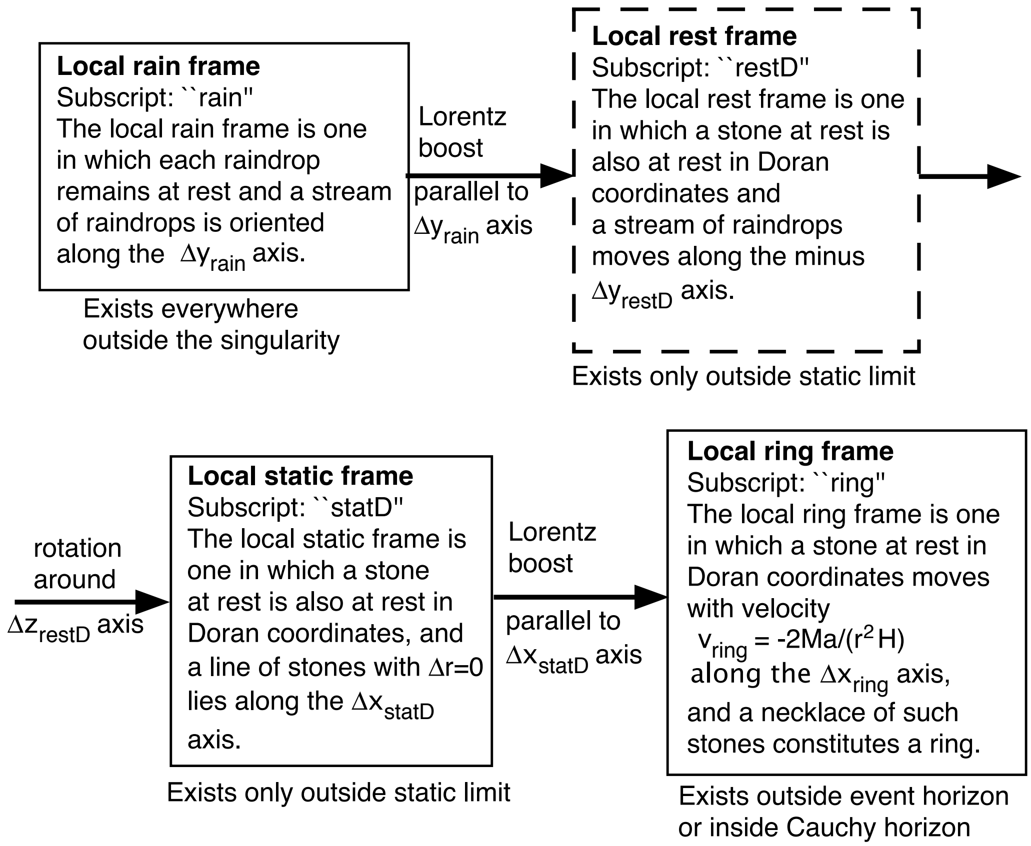


FIGURE 5 Definitions of several local inertial frames from which we choose to make measurements and observations near the spinning black hole. The so-called "local rest frame" (upper right box) serves mainly to connect the local rain frame to the local static frame, hence the dashed lines around the box that describes it.

17.5 ■ THE LOCAL RAIN FRAME

Take relaxed measurements as we fall

Choose local inertial frames for our measurements.

Thus far this chapter has introduced the Doran global metric and a few of its consequences for the motion of a free stone. As usual, our goal is to report measurements and observations made in local inertial frames; we now derive several of these from the Doran metric.

Figure 5 gives summary definitions of the local inertial frames we choose near the spinning black hole: local inertial rain, rest, static, and ring frames, described in this section and the following three sections. You will show that when $a \to 0$, the local rest, static, and ring frames all become the local shell frame (Section 5.7); and the local rain frame simply becomes the local rain frame for the non-spinning black hole (Section 7.5).

Comment 6. Generalized Lorentz transformation

The Lorentz transformations defined in Section 1.10 were limited to Lorentz boosts along the common $\Delta x_{\mathrm{frame}}$ axes of laboratory and rocket frames. In general, Lorentz boosts can take place along any direction in either frame. One way to do this is first to rotate the initial frame, then Lorentz-boost it to the desired final frame. Thus the general definition of **Lorentz transformation** also includes simple rotation of one frame with respect to the other. Look at labels on the arrows in Figure 5. Each of these labels describes a Lorentz transformation.

Initially Figure 5 may seem strange and perplexing; this section and the next three sections describe each of these frames in more detail.

Local rain frame from equation (4)

The right side of Doran metric (4) is in tetrad form—the sum and difference of squares (introduced in Section 7.6). Therefore its approximate form gives us *some* local inertial frame coordinates expressed in Doran global coordinates. Which particular local inertial frame? We will find that it earns the name **local inertial rain frame**; so the coordinates for the local rain frame in terms of Doran coordinates are:

Local rain frame coordinates

$$\Delta t_{\mathrm{rain}} \equiv \Delta T \tag{32}$$

$$\Delta y_{\mathrm{rain}} \equiv \left[\left(\frac{\bar{r}^2}{\bar{r}^2+a^2}\right)^{1/2}\Delta r - \left(\frac{2M}{\bar{r}}\right)^{1/2} a\Delta\Phi\right] + \left(\frac{2M}{\bar{r}}\right)^{1/2}\Delta T \tag{33}$$

$$\Delta x_{\mathrm{rain}} \equiv \left(\bar{r}^2+a^2\right)^{1/2}\Delta\Phi \tag{34}$$

The expression in square brackets in equation (33) appears also in equations for some later local inertial frames. Figure 5 contains a definition of the local rain frame.

Local rain frame: valid everywhere.

Expressions on the right sides of (32) through (34) are all real outside $r = 0$, so the local inertial rain frame exists everywhere outside the singularity. These three equations plus the approximate form of (4) guarantee that the local rain frame metric has the usual form:

$$\Delta\tau^2 \approx \Delta t_{\mathrm{rain}}^2 - \Delta y_{\mathrm{rain}}^2 - \Delta x_{\mathrm{rain}}^2 \tag{35}$$

Comment 7. The rain tetrad

Equations (32) through (34) express local rain coordinates in Doran coordinates when the global metric is in *tetrad* form. Notice that two of the components, Δt_{rain} and Δx_{rain}, depend on a single global coordinate difference, while Δy_{rain} depends on all three: ΔT, Δr, and $\Delta\Phi$. This result, due to black hole spin, generalizes the rain tetrad for a non-spinning black hole, where Δy_{rain} depends on two coordinate differences—equation (43) in Section 7.5.

QUERY 8. Compare rain frame coordinates for spinning and non-spinning black holes.

Compare local rain coordinate expressions (32) through (34) with those for the non-spinning black hole in Box 4 of Section 7.5. Under what assumption or assumptions do the spinning black hole expressions reduce to those for the non-spinning black hole when $a \rightarrow 0$?

The worldtube projected on the $[r, T]$ slice in Figure 4 embraces rain frames through which the rain observer passes. The time axis of a local inertial frame is always tangent to the worldline of a stone at rest in that frame. The raindrop is at rest in the local rain frame; therefore the Δt_{rain} axis is tangent to the raindrop worldline in Figure 4. What is the direction of the Δy_{rain} axis on the $[r, T]$ slice? The Δy_{rain} axis is a line along which $\Delta t_{\text{rain}} = \Delta x_{\text{rain}} = 0$. With these conditions, equation (33) tells us that the Δy_{rain} axis lies along the global Δr direction, as shown in Figure 4.

?

Objection 3. *Figure 4 is all wrong! Equation (32) clearly says that $\Delta t_{\text{rain}} = \Delta T$, so the Δt_{rain} axis must point along the vertical T/M axis in Figure 4.* More: *Equation (33) says that Δy_{rain} has contributions from all three global coordinates, so cannot point along the horizontal r/M axis in the figure.*

!

You *are* observant! To answer your objection, start with the Δy_{rain} axis: Note, first, that Figure 4 displays an $[r, T]$ slice. On that slice $\Delta\Phi = 0$. Second, for events simultaneous in the rain frame, $\Delta t_{\text{rain}} = 0$ so $\Delta T = 0$ from (32). That leaves the Δy_{rain} axis pointing along the r-direction, from (33). Now for the Δt_{rain} axis: By definition, raindrops lie at rest in the local rain frame. Setting $\Delta y_{\text{rain}} = \Delta x_{\text{rain}} = 0$ in (33) and (34) yields the worldline equation (30)—in its approximate form—so the local Δt_{rain} axis must lie along the raindrop worldline.

Equations (32) through (34) relate local measurement to global coordinates. An example is the velocity of a stone. Equations (32) through (34) lead to the following relation between global coordinate expressions $dr/dT, d\Phi/dT$ and the stone's velocity measured in the local rain frame:

Stone's velocity in local rain frame

$$v_{\text{rain,y}} \equiv \lim_{\Delta t_{\text{rain}} \to 0} \frac{\Delta y_{\text{rain}}}{\Delta t_{\text{rain}}} = \left(\frac{r^2}{r^2+a^2}\right)^{1/2} \frac{dr}{dT} + \left(\frac{2M}{r}\right)^{1/2} \left(1 - a\frac{d\Phi}{dT}\right) \quad (36)$$

$$v_{\text{rain,x}} \equiv \lim_{\Delta t_{\text{rain}} \to 0} \frac{\Delta x_{\text{rain}}}{\Delta t_{\text{rain}}} = \left(r^2 + a^2\right)^{1/2} \frac{d\Phi}{dT} \quad (37)$$

In the limit-taking process the local frame shrinks to a point (event) in spacetime, which removes the superscript bars that show average values.

Raindrop velocity in local rain frame

Now let the stone be a raindrop and verify its velocity components in the local rain frame. To do this, substitute for the raindrop from (30) and (31) into (36) and (37):

$$v_{\text{rain,y}} = v_{\text{rain,x}} = 0 \qquad \text{(raindrop)} \qquad (38)$$

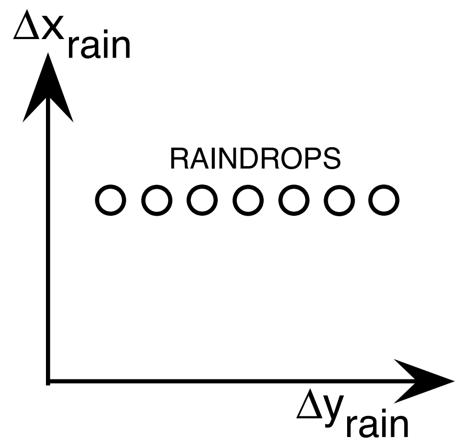


FIGURE 6 A snapshot ($\Delta t_{\text{rain}} = 0$) shows a line of raindrops, which are at rest in each local rain frame (Figure 4). Equations (36), (37), and (38) show that in Doran coordinates these raindrops have identical Φ and T but different r.

which shows that the raindrop is at rest in the local inertial rain frame. This justifies the name *rain frame*.

A line of raindrops

But the raindrop has more to tell us about the local rain frame. Consider a line of raindrops, for example a sequence of drops from a faucet, all with the same value of Φ but released in sequence so that a snapshot ($\Delta t_{\text{rain}} = 0$) shows the raindrops at slightly different r-values. Then equations (33) and (34) tell us that this line of raindrops (with $\Delta T = \Delta\Phi = 0$ but with slightly different values of Δr) all have the same Δx_{rain} but different values of Δy_{rain}. Therefore raindrops of equal Φ lie at rest in the rain frame and a line of raindrops lies parallel to the Δy_{rain} axis (Figure 6).

17.6 ■ THE LOCAL REST FRAME

At rest in Doran global coordinates

We want more choices for measurement than just a suicide raindrop trip to the singularity. For example, it is convenient to have a local frame in which a stone at rest has constant r.

Frame stands still in Doran coordinates

To find such constant-r frames, start with the rain frame, then apply a Lorentz boost in the Δy_{rain} direction so that a stone with $dr/dT = 0$ and $d\Phi/dT = 0$ has zero velocity in the new frame. Label this the **local inertial rest frame**, with the subscript "restD" to remind us that it is at rest in Doran global coordinates. The required Lorentz boost between rain and rest frames has the form of equation (40) in Section 1.10:

$$\Delta t_{\rm restD} = \gamma_{\rm rel}\left(\Delta t_{\rm rain} - v_{\rm rel}\Delta y_{\rm rain}\right) \tag{39}$$

$$\Delta y_{\rm restD} = \gamma_{\rm rel}\left(\Delta y_{\rm rain} - v_{\rm rel}\Delta t_{\rm rain}\right) \tag{40}$$

$$\Delta x_{\rm restD} = \Delta x_{\rm rain} \tag{41}$$

What is the value of $v_{\rm rel}$, the relative speed between the rest and rain frame? We want a stone with $\Delta r = \Delta\Phi = 0$ to have zero velocity in the new frame, that is $\Delta y_{\rm restD} = \Delta x_{\rm restD} = 0$. Now from (41) and (34) we already have $\Delta x_{\rm restD} = \Delta x_{\rm rain} = 0$ for a stone with $\Delta\Phi = 0$, and from equations (32) and (33):

$$\Delta y_{\rm rain} - v_{\rm rel}\Delta t_{\rm rain} = \left(\frac{\bar{r}^2}{\bar{r}^2 + a^2}\right)^{1/2}\Delta r - \left(\frac{2M}{\bar{r}}\right)^{1/2} a\Delta\Phi + \left(\frac{2M}{\bar{r}}\right)^{1/2}\Delta T - v_{\rm rel}\Delta T \tag{42}$$

$v_{\rm rel}$ between rest and rain frames

We want this expression to be zero when $\Delta r = \Delta\Phi = 0$. This will be the case if the last two terms on the right side of (42) cancel. That is, we need a Lorentz boost such that:

$$v_{\rm rel} = \left(\frac{2M}{\bar{r}}\right)^{1/2} \qquad \text{so} \qquad \gamma_{\rm rel} = \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2} \tag{43}$$

Local rest frame coordinates

Now substitute equations (43) and (32) through (34) into (39) through (41) to obtain local rest frame coordinates in global Doran coordinates:

$$\Delta t_{\rm restD} = \left(1 - \frac{2M}{\bar{r}}\right)^{1/2}\Delta T - \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2}\left(\frac{2M}{\bar{r}}\right)^{1/2}\left[\left(\frac{\bar{r}^2}{\bar{r}^2 + a^2}\right)^{1/2}\Delta r - \left(\frac{2M}{\bar{r}}\right)^{1/2} a\Delta\Phi\right] \tag{44}$$

$$\Delta y_{\rm restD} = \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2}\left[\left(\frac{\bar{r}^2}{\bar{r}^2 + a^2}\right)^{1/2}\Delta r - \left(\frac{2M}{\bar{r}}\right)^{1/2} a\Delta\Phi\right] \tag{45}$$

$$\Delta x_{\rm restD} = \left(\bar{r}^2 + a^2\right)^{1/2}\Delta\Phi \tag{46}$$

The two square-bracket expressions are the same as the one in (33). Figure 5 contains a definition of the local rest frame.

Equations (44) and (45) show that the local inertial rest frame exists only outside the static limit, because these local coordinates are imaginary for $r < 2M$. This result reinforces the interpretation of the static limit defined in Section 17.3.

Stone's velocity in local rest frame.

From equations (44) through (46) we derive expressions for the stone's velocity in the local inertial rest frame:

$$v_{\mathrm{restD,y}} \equiv \lim_{\Delta t_{\mathrm{restD}}\to 0} \frac{\Delta y_{\mathrm{restD}}}{\Delta t_{\mathrm{restD}}} \tag{47}$$

$$= \frac{\left(\dfrac{r^2}{r^2+a^2}\right)^{1/2}\dfrac{dr}{dT} - \left(\dfrac{2M}{r}\right)^{1/2} a\dfrac{d\Phi}{dT}}{\left(1-\dfrac{2M}{r}\right) - \left(\dfrac{2M}{r}\right)^{1/2}\left[\left(\dfrac{r^2}{r^2+a^2}\right)^{1/2}\dfrac{dr}{dT} - \left(\dfrac{2M}{r}\right)^{1/2} a\dfrac{d\Phi}{dT}\right]}$$

$$v_{\mathrm{restD,x}} \equiv \lim_{\Delta t_{\mathrm{restD}}\to 0} \frac{\Delta x_{\mathrm{restD}}}{\Delta t_{\mathrm{restD}}} \tag{48}$$

$$= \frac{\left(1-\dfrac{2M}{r}\right)^{1/2}\left(r^2+a^2\right)^{1/2}\dfrac{d\Phi}{dT}}{\left(1-\dfrac{2M}{r}\right) - \left(\dfrac{2M}{r}\right)^{1/2}\left[\left(\dfrac{r^2}{r^2+a^2}\right)^{1/2}\dfrac{dr}{dT} - \left(\dfrac{2M}{r}\right)^{1/2} a\dfrac{d\Phi}{dT}\right]}$$

In the limit-taking process the local frame shrinks to a point (event) in spacetime, which removes the superscript bars that specify average values.

The right sides of these equations are a mess, but the computer does not care and translates between global coordinate velocities and velocities in the local rest frame. For example, to find the speed of the raindrop in the local rest frame, substitute into these equations from (30) and (31). The result is:

$$v_{\mathrm{restD,y}} = -\left(\frac{2M}{r}\right)^{1/2} = -v_{\mathrm{rel}} \qquad \text{(raindrop)} \tag{49}$$

$$v_{\mathrm{restD,x}} = 0 \qquad \text{(raindrop)} \tag{50}$$

The last step in (49) is from (43); since a raindrop is at rest in the rain frame and we Lorentz boost $+v_{\mathrm{rel}}$ in the Δy_{rain} direction, therefore the raindrop must have velocity $-v_{\mathrm{rel}}$ in the new frame.

Stone at rest in Doran coordinates is at rest in local rest frame.

Now check that we are consistent: To verify that a stone at rest in Doran coordinates is indeed at rest in the local rest frame, substitute $dr/dT = d\Phi/dT = 0$ into (47) and (48) to obtain

$$v_{\mathrm{restD,y}} = v_{\mathrm{restD,x}} = 0 \qquad (\text{stone: } dr/dT = d\Phi/dT = 0) \tag{51}$$

The stone at rest in global Doran coordinates is also at rest in the local rest frame.

QUERY 9. Local rest frame coordinates when $a \to 0$ Show that when $a \to 0$ for the non-spinning black hole, equations (44) through (46) recover expressions for the local shell frame in global rain coordinates, Box 2 in Section 7.4.

17.7 ■ THE LOCAL STATIC FRAME

Lining up with the string of stones in a necklace.

Figure 6 shows a sequence of raindrops at rest in the local rain frame and lined up along the Δy_{rain} axis. The Lorentz boost from rain to rest frame takes place along the same Δy_{rain}, so the line of raindrops also lies along the Δy_{restD} axis, as shown in Figure 7. But in this local frame they are moving in the global inward direction shown in that figure.

For the non-spinning black hole we made observations from local shell frames outside the event horizon. On the symmetry slice through the center of a non-spinning black hole, each shell is a ring. The spinning black hole permits shell-rings only outside the static limit (see the exercises). More useful for the spinning black hole is a set of concentric rings that rotate with respect to global Doran coordinates. Think of each ring as composed of a necklace of stones at a given value of r that move in the Φ direction, as shown in Figure 7.

Rotating rings for $a > 0$ replace shell-rings for $a = 0$.

?

Objection 4. *In Figure 7 your Φ and r axes are not perpendicular. This violates the Pythagorean Theorem. It's illegal!*

!

Pythagoras was aware of what was later called Euclidean geometry in flat space, in which, for orthogonal coordinates,

$$\Delta s^2 = A\Delta r^2 + B\Delta\Phi^2 \qquad \text{(Phythagoras)} \qquad (52)$$

for some positive constants A and B. In contrast, you can show from (5) that, for $\Delta T = 0$,

$$\Delta s^2 = A\Delta r^2 + B\Delta\Phi^2 + C\Delta r\Delta\Phi \qquad \text{(Doran space)} \qquad (53)$$

that is, there is a cross term in the metric that signals non-orthogonality.

For every local inertial frame, we *demand* that spatial coordinates be orthogonal, so that

$$\Delta s^2 = \Delta x^2_{\text{frame}} + \Delta y^2_{\text{frame}} \qquad \text{(every local inertial frame)} \qquad (54)$$

Hence we force the Pythagorean Theorem to apply for space coordinates of every local inertial frame. It *need not* apply to global coordinates; Figure 7 is an example.

In the present section we start toward the rotating ring by finding a local inertial frame at fixed Doran global coordinates but with its local x-coordinate axis lying along the Φ-direction. We call it the **local static frame**, (subscript: "statD"). The local static frame is rotated with respect to the local rest frame (Figure 7).

local static frame

The rotation formulas between local rest and local static frames are:

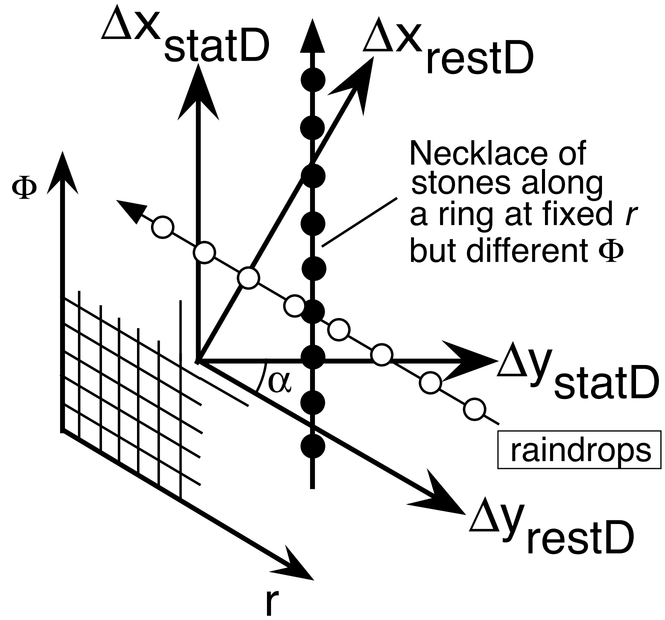


FIGURE 7 Three coordinate systems—local static and local rest plus global r-Φ—plotted on a single flat patch at a fixed global coordinate T. The line of raindrops lies along the global r-direction and moves in the negative r-direction. The necklace of stones around the spinning black hole forms a ring that lies along the global Φ-direction; stones in the necklace move in the positive Φ-direction. The relation between the local rest and static frames is a simple rotation through the angle α—equations (55) through (57). *Important:* This is a two-dimensional figure, not a perspective figure.

$$\Delta t_{\text{statD}} = \Delta t_{\text{restD}} \tag{55}$$

$$\Delta y_{\text{statD}} = \Delta y_{\text{restD}} \cos\alpha + \Delta x_{\text{restD}} \sin\alpha \tag{56}$$

$$\Delta x_{\text{statD}} = \Delta x_{\text{restD}} \cos\alpha - \Delta y_{\text{restD}} \sin\alpha \tag{57}$$

We choose the angle α so that Δy_{statD} has no terms that contain $\Delta\Phi$. In other words, orient the rotated frame so that a *ring* of stones with the same r but with different Φ-values all have $\Delta y_{\text{statD}} = 0$; the ring lies locally parallel to the Δx_{statD} axis. Equations (56), (45), and (46) yield:

$$\begin{aligned}\Delta y_{\text{statD}} = \left(1 - \frac{2M}{r}\right)^{-1/2} \left[\left(\frac{\bar{r}^2}{\bar{r}^2 + a^2}\right)^{1/2} \Delta r - \left(\frac{2M}{\bar{r}}\right)^{1/2} a\Delta\Phi\right] \cos\alpha \\ + \left(\bar{r}^2 + a^2\right)^{1/2} \Delta\Phi \sin\alpha\end{aligned} \tag{58}$$

Rearrange this equation to combine coefficients of $\Delta\Phi$:

$$\Delta y_{\text{statD}} = \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2} \left(\frac{\bar{r}^2}{\bar{r}^2 + a^2}\right)^{1/2} \Delta r \cos\alpha \qquad (59)$$
$$- \left[\left(\frac{2M}{\bar{r}}\right)^{1/2} \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2} a\cos\alpha - \left(\bar{r}^2 + a^2\right)^{1/2} \sin\alpha\right] \Delta\Phi$$

To eliminate $\Delta\Phi$ from the second line of equation (59), set the contents of the square bracket equal to zero. This determines angle α:

$$\frac{\sin\alpha}{\cos\alpha} \equiv \tan\alpha = \left(\frac{2M}{\bar{r}}\right)^{1/2} \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2} \left(\frac{a^2}{\bar{r}^2 + a^2}\right)^{1/2} \qquad (60)$$

In Query 10 you verify the following expressions for $\sin\alpha$ and $\cos\alpha$:

$$\sin\alpha = \left(\frac{2M}{\bar{r}}\right)^{1/2} \frac{a}{\bar{r}\bar{H}} \qquad (61)$$

$$\cos\alpha = \left(1 - \frac{2M}{\bar{r}}\right)^{1/2} \frac{\left(\bar{r}^2 + a^2\right)^{1/2}}{\bar{r}\bar{H}} \qquad (62)$$

The angle α should be written $\alpha(r)$ to remind us that it is a function of the r-coordinate, but we will not bother with this more complicated notation.

QUERY 10. Check expressions for $\sin\alpha$ and $\cos\alpha$.

A. Divide corresponding sides of (61) and (62) to check that the result gives $\tan\alpha$ in (60).

B. Confirm that $\sin^2\alpha + \cos^2\alpha = 1$.

C. Show that when $r \rightarrow \infty$, then $\alpha \rightarrow 0$.

D. Show that when $r \rightarrow 2M^+$ (that is, when $r \rightarrow 2M$ while $r > 2M$), then $\alpha \rightarrow \pi/2$.

E. Show that α is undefined for $r < 2M$. *Prediction:* The static frame exists only outside the static limit.

Local static frame coordinates

When we substitute (61) and (62) into (59), the second line on the right side of this equation goes to zero and the first line yields the simple expression for Δy_{statD} in (64). For rotation, $\Delta t_{\text{restD}} = \Delta t_{\text{statD}}$. Then substitution into (57) finds Δx_{statD}, which completes the coordinates of the static frame in global Doran coordinates:

$$\Delta t_{\rm statD} \equiv \left(1-\frac{2M}{\bar r}\right)^{1/2}\Delta T \tag{63}$$
$$- \left(1-\frac{2M}{\bar r}\right)^{-1/2}\left(\frac{2M}{\bar r}\right)^{1/2}\left[\left(\frac{\bar r^2}{\bar r^2+a^2}\right)^{1/2}\Delta r - \left(\frac{2M}{\bar r}\right)^{1/2} a\Delta\Phi\right]$$
$$\Delta y_{\rm statD} \equiv \frac{\Delta r}{\bar H} \tag{64}$$
$$\Delta x_{\rm statD} \equiv -\left(1-\frac{2M}{\bar r}\right)^{-1/2}\left[\left(\frac{2M}{\bar r}\right)^{1/2}\left(\frac{\bar r^2}{\bar r^2+a^2}\right)^{1/2}\frac{a}{\bar r\bar H}\Delta r - \bar r\bar H\Delta\Phi\right] \tag{65}$$

These equations show that, like the local rest frame, the local static frame exists only outside the static limit. Figure 5 contains a summary definition of the local static frame.

Stone's velocity in local static frame.

Now we derive expressions for the stone's velocity in the local inertial static frame:

$$v_{\rm statD,y} \equiv \lim_{\Delta t_{\rm statD}\to 0}\frac{\Delta y_{\rm statD}}{\Delta t_{\rm statD}} \tag{66}$$
$$= \frac{H^{-1}\left(1-\dfrac{2M}{r}\right)^{1/2}\dfrac{dr}{dT}}{\left(1-\dfrac{2M}{r}\right)-\left(\dfrac{2M}{r}\right)^{1/2}\left[\left(\dfrac{r^2}{r^2+a^2}\right)^{1/2}\dfrac{dr}{dT}-\left(\dfrac{2M}{r}\right)^{1/2}a\dfrac{d\Phi}{dT}\right]}$$
$$v_{\rm statD,x} \equiv \lim_{\Delta t_{\rm statD}\to 0}\frac{\Delta x_{\rm statD}}{\Delta t_{\rm statD}} \tag{67}$$
$$= \frac{(rH)^{-1}\left[r^2H^2\dfrac{d\Phi}{dT}-\left(\dfrac{2M}{r}\right)^{1/2}\left(\dfrac{r^2}{r^2+a^2}\right)^{1/2}a\dfrac{dr}{dT}\right]}{\left(1-\dfrac{2M}{r}\right)-\left(\dfrac{2M}{r}\right)^{1/2}\left[\left(\dfrac{r^2}{r^2+a^2}\right)^{1/2}\dfrac{dr}{dT}-\left(\dfrac{2M}{r}\right)^{1/2}a\dfrac{d\Phi}{dT}\right]}$$

In the limit-taking process the local frame shrinks to a point (event) in spacetime, which removes the superscript bars that show average values.

The right sides of these equations are a mess, but the computer does not care and translates between global coordinate velocities and velocities in the local static frame. For example, for the static frame components of a raindrop's velocity use equations (30) and (31):

$$v_{\mathrm{statD,y}} = -H^{-1}\left(\frac{2M}{r}\right)^{1/2}\left(1-\frac{2M}{r}\right)^{1/2}\left(\frac{r^2+a^2}{r^2}\right)^{1/2} \tag{68}$$

$$= -\left(\frac{2M}{r}\right)^{1/2}\cos\alpha \qquad \text{(raindrop)}$$

$$v_{\mathrm{statD,x}} = H^{-1}\left(\frac{2M}{r}\right)\frac{a}{r} \tag{69}$$

$$= \left(\frac{2M}{r}\right)^{1/2}\sin\alpha \qquad \text{(raindrop)}$$

Figure 7 shows us that the raindrop moves inward at an angle α with respect to the $\Delta y_{\mathrm{statD}}$ axis, in agreement with equations (68) and (69).

QUERY 11. Raindrop in the local static frame

A. Show that the speed of the raindrop in the static frame is $(2M/r)^{1/2}$.

B. Show that at large r, the raindrop moves slowly in the local static frame and in the direction $\alpha \to 0$ in that frame.

C. Show that as $r \to 2M^+$, the raindrop moves sideways at angle $\alpha \to \pi/2$ with respect to the $\Delta y_{\mathrm{statD}}$ axis at a speed approaching light speed in that frame.

Stone at rest in Doran coordinates is at rest in local static frame.

Finally, a consistency check: We verify that a stone at rest in Doran coordinates is indeed at rest in the local static frame. For this, substitute $dr/dT = d\Phi/dT = 0$ into (66) and (67) to obtain

$$v_{\mathrm{statD,y}} = v_{\mathrm{statD,x}} = 0 \qquad \text{(stone: } dr/dT = d\Phi/dT = 0) \tag{70}$$

QUERY 12. Local static frame coordinates when $a \to 0$ Show that when $a \to 0$ for the non-spinning black hole, equations (63) through (65) recover expressions for the local shell frame in global rain coordinates, Box 2 in Section 7.4. Compare the results of Query 9: when $a \to 0$, both rest frames and static frames become shell frames!

?

Objection 5. *Why are the line of raindrops and the string of necklace stones not perpendicular in Figure 7? You cannot tell me this is due to the non-measurability of global coordinates; These are real objects!*

!

Right you are: in a local frame the line of raindrops and the string of necklace stones are not perpendicular, regardless of the global

> coordinates that we use. The reason is subtle, but can be understood in analogy to raindrops that fall on Earth. Let a horizontal wind blow each raindrop sideways, so the line of raindrops deviates from the vertical. The spin of the black hole has a similar effect, a phenomenon sometimes called **dragging of inertial frames**. How big is the effect? Angle α in Figure 7 measures the size of this effect. In Query 10 you showed that far from the spinning black hole, $r \to \infty$, the angle $\alpha \to 0$. In contrast, as $r \to 2M^+$ the angle $\alpha \to \pi/2$ and the raindrop speed approaches that of light. At the static limit the "spinning black hole winds" are so great that raindrops are blown horizontal at the speed of light. Hurricanes on Earth are gentle beasts compared to the spinning black hole!

Dragging of inertial frames

17.8 ■ THE LOCAL RING FRAME

Relax on a ring that circles around the black hole.

The local static frame derived in Section 17.7 exists only outside the static limit. But we know from Section 17.3 that a stone can exist with no r motion all the way down to the event horizon if it has some tangential motion.

Necklace of stones becomes a ring.

We give the name **ring** to a necklace of stones, all at the same r, that have $dr/dT = 0$ with $d\Phi/dT = \omega(r)$; then we seek a corresponding set of **local inertial ring frames** that exist down to the event horizon. Each local inertial ring frame is at rest on the ring. We will discover, to our surprise, that the ring—and local ring frames—can exist also between the Cauchy horizon and the singularity.

To find a local inertial ring frame in which the necklace of stones are at rest, we perform a Lorentz boost in the $\Delta x_{\rm statD}$ direction.

$$\Delta t_{\rm ring} = \gamma_{\rm rel}\left(\Delta t_{\rm statD} - v_{\rm rel}\Delta x_{\rm statD}\right) \tag{71}$$

$$\Delta y_{\rm ring} = \Delta y_{\rm statD} \tag{72}$$

$$\Delta x_{\rm ring} = \gamma_{\rm rel}\left(\Delta x_{\rm statD} - v_{\rm rel}\Delta t_{\rm statD}\right) \tag{73}$$

Values of $v_{\rm rel}$ and $\gamma_{\rm rel}$ in these equations are *not* the same as the corresponding values in equations (39) and (40).

How do we find the value of $v_{\rm rel}$? We choose $v_{\rm rel}$ to fulfill our demand that $\Delta x_{\rm ring} = 0$ in (73) when $\Delta r = 0$ and $\Delta\Phi = \bar{\omega}(r)\Delta T$, where equation (12) defines $\omega(r)$. In Query 13 you show that this demand leads to:

$$v_{\rm rel} = \frac{2Ma}{\bar{r}^2\bar{H}} \qquad \text{(ring frame speed in stat frame)} \tag{74}$$

from which

$$\gamma_{\rm rel} \equiv \left(1 - v_{\rm rel}^2\right)^{-1/2} = \frac{\bar{r}\bar{H}}{\bar{R}}\left(1 - \frac{2M}{\bar{r}}\right)^{-1/2} \tag{75}$$

QUERY 13. Find $v_{\rm rel}$

A. Demand that $\Delta x_{\rm ring} = 0$ in equation (73) when $\Delta r = 0$ and $\Delta\Phi = \bar{\omega}\Delta T$. Show that this yields

$$v_{\rm rel} = \frac{\bar{r}\bar{H}\bar{\omega}}{1 - \dfrac{2M}{\bar{r}} + \dfrac{2M}{\bar{r}}a\bar{\omega}} \tag{76}$$

B. Substitute for $\bar{\omega}$ from (12) into (76) and manipulate the result to verify (74).

Local ring frame coordinates

Now we can complete Lorentz boost equations (71) through (73) using equations (63) through (65) plus equations (74) and (75). *Result:* coordinates of the local ring frame in global coordinates:

$$\Delta t_{\rm ring} \equiv \frac{\bar{r}\bar{H}}{\bar{R}}\Delta T - \frac{\bar{\beta}}{\bar{H}}\Delta r \tag{77}$$

$$\Delta y_{\rm ring} \equiv \frac{\Delta r}{\bar{H}} \tag{78}$$

$$\Delta x_{\rm ring} \equiv \bar{R}\left(\Delta\Phi - \bar{\omega}\Delta T\right) - \frac{\bar{\omega}\bar{r}}{\bar{\beta}}\Delta r \tag{79}$$

Definition of β

where

$$\beta \equiv \left(\frac{2M}{r}\right)^{1/2}\left(\frac{r^2+a^2}{R^2}\right)^{1/2} \tag{80}$$

The average $\bar{\beta}$ is the same expression with $r \to \bar{r}$ and $R \to \bar{R}$.

The unitless symbol β stands for a bundle of constants and global coordinates similar (but not equal) to dr/dT for a raindrop in equation (30). Box 1 summarizes useful functions defined in this chapter.

Ring frames valid for $r > r_{\rm EH}$ and $0 < r < r_{\rm CH}$

Equations (77) through (79) tell us that the local ring frame can exist wherever H is real, which from (15) is down to the event horizon. The function H is imaginary between the two horizons, so ring frames cannot exist there. Inside the Cauchy horizon, however, H is real again. This astonishing result predicts that local ring frames can exist between the Cauchy horizon and the singularity. *Question:* How can this possibly be? *Answer:* Close to the singularity of a spinning black hole our intuition fails. Recall our paraphrase of *Wheeler's radical conservatism*, Comment 1 in Section 7.1: Follow what the equations tell us, no matter how strange the results. Then develop a new intuition!

Figure 5 contains a definition of the local ring frame.

QUERY 14. Local ring frame coordinates when $a \to 0$ Show that when $a \to 0$ for the non-spinning black hole, equations (77) through (79) recover expressions for the local shell frame in global rain coordinates, Box 2 in Section 7.4.

Box 1. Useful Relations for the Spinning Black Hole

Many derivations manipulate these expressions.

Static limit from Section 17.3:

$$r_S = 2M \tag{81}$$

Reduced circumference from Section 17.2:

$$R^2 \equiv r^2 + a^2 + \frac{2Ma^2}{r} \tag{82}$$

Horizon function from Section 17.3:

$$H^2 \equiv \frac{1}{r^2}\left(r^2 - 2Mr + a^2\right) \tag{83}$$

$$= \frac{1}{r^2}\left(r - r_{EH}\right)\left(r - r_{CH}\right) \tag{84}$$

where r_{EH} and r_{CH} are r-values of the event and Cauchy horizons, respectively, from Section 17.3.

$$\frac{r_{EH}}{M} \equiv 1 + \left(1 - \frac{a^2}{M^2}\right)^{1/2} \quad \text{(event horizon)} \tag{85}$$

$$\frac{r_{CH}}{M} \equiv 1 - \left(1 - \frac{a^2}{M^2}\right)^{1/2} \quad \text{(Cauchy horizon)} \tag{86}$$

Ring omega from Section 17.3:

$$\omega \equiv \frac{2Ma}{rR^2} \tag{87}$$

An equivalence from Section 17.3:

$$1 - \frac{2M}{r} + R^2\omega^2 = \left(\frac{rH}{R}\right)^2 \tag{88}$$

Definition of α from Section 17.7:

$$\alpha \equiv \arcsin\left[\left(\frac{2M}{r}\right)^{1/2}\frac{a}{rH}\right] \tag{89}$$

$(0 \leq \alpha \leq \pi/2)$, namely $(r \geq 2M)$

Definition of β from Section 17.8:

$$\beta \equiv \left(\frac{2M}{r}\right)^{1/2}\left(\frac{r^2 + a^2}{R^2}\right)^{1/2} \tag{90}$$

Stone velocity in local ring frame

Now suppose that a stone moves in the local ring frame. Equations (77) through (79) lead to the following relation between components of global coordinate velocities dr/dT and $d\Phi/dT$ and components of the stone's velocity measured in the local ring frame:

$$v_{\text{ring,y}} \equiv \lim_{\Delta t_{\text{ring}} \to 0} \frac{\Delta y_{\text{ring}}}{\Delta t_{\text{ring}}} = \frac{\dfrac{dr}{dT}}{\dfrac{rH^2}{R} - \beta\dfrac{dr}{dT}} \tag{91}$$

$$v_{\text{ring,x}} \equiv \lim_{\Delta t_{\text{ring}} \to 0} \frac{\Delta x_{\text{ring}}}{\Delta t_{\text{ring}}} = \frac{R\left(\dfrac{d\Phi}{dT} - \omega\right) - \dfrac{\omega r}{\beta}\dfrac{dr}{dT}}{\dfrac{rH}{R} - \dfrac{\beta}{H}\dfrac{dr}{dT}} \tag{92}$$

In the limit-taking process the local frame shrinks to a point (event) in spacetime, which removes the superscript bars that show average values.

Stone at rest in Doran coordinates moves in local ring coordinates.

Suppose that a stone remains at rest in Doran coordinates. What is its velocity in the local ring frame? Recall from Section 7.3 that at or inside the static limit a stone cannot be at rest in Doran coordinates, so we require that $r \geq 2M$. But what goes wrong with observations at and inside the static limit? The trouble is different for different r-values there. Substitute $dr/dT = d\Phi/dT = 0$ into (91) and (92) to obtain

$$v_{\mathrm{ring,y}} = 0 \qquad \text{(stone at rest in Doran coordinates, } r \geq 2M) \tag{93}$$

$$v_{\mathrm{ring,x}} = -\frac{2Ma}{r^2 H} \qquad \text{(ditto)} \tag{94}$$

QUERY 15. Velocity in ring frame of stone at rest in Doran coordinates
Analyze equation (94) with the following steps:

A. For $r = 2M$, show that $v_{\mathrm{ring,x}} = -1$, the speed of light.
B. For $r_{\mathrm{EH}} < r < 2M$, show that $v_{\mathrm{ring,x}} < -1$, greater than light speed.
C. For $r_{\mathrm{CH}} < r < r_{\mathrm{EH}}$ show that no ring frame exists and $v_{\mathrm{ring,x}}$ is imaginary.
D. For $r < r_{\mathrm{CH}}$, show that $v_{\mathrm{ring,x}} < -1$, greater than light speed.

QUERY 16. Velocity of necklace stones in static frame With a symmetry argument, show that the velocity of the necklace stones measured in the static frame has the same y component as (93) but the negative of the x-component in (94).

Now let us find the velocity of the raindrop in the local ring frame. Into equations (91) and (92) substitute dr/dT from (30) and $d\Phi/dT = 0$ from (31).

QUERY 17. Denominator of (91). Show that for the raindrop, the denominator of the right side of (91) becomes R/r.

The result of Query 17 plus (30) and (90) lead to an expression for $v_{\mathrm{ring,y}}$:

$$v_{\mathrm{ring,y}} = -\left(\frac{2M}{r}\right)^{1/2}\left(\frac{r^2 + a^2}{R^2}\right)^{1/2} = -\beta \qquad \text{(raindrop)} \tag{95}$$

QUERY 18. Numerator of (92). Show that for the raindrop, the numerator of the right side of (92) is equal to zero.

Query 18 shows that:

$$v_{\mathrm{ring,x}} = 0 \qquad \text{(raindrop)} \tag{96}$$

Raindrop falls vertically in ring frame.

Surprising result: Every raindrop falls vertically through every local ring frame. Compare this result with parts B and C in Query 11; in the local static frame, raindrops move sideways. The local ring frame compensates for this

TABLE 17.1 Measured velocity of raindrop in several local inertial frames

Frame	Valid Region	$v_{\text{frame,y}}$	$v_{\text{frame,x}}$
Rain	Everywhere, $r > 0$	0	0
Rest	$r > r_S$	$-(2M/r)^{1/2}$	0
Static	$r > r_S$	$-(2M/r)^{1/2}\cos\alpha$	$+(2M/r)^{1/2}\sin\alpha$
Ring	$r \leq r_{\text{CH}}$ & $r \geq r_{\text{EH}}$	$-\beta$	0

sideways motion with a Lorentz boost, so raindrops fall vertically through the ring frame.

Table 1 summarizes the velocity components of the raindrop in the four local inertial frames we have set up.

Comment 8. Goodbye local rest frame.
We can construct an infinite number of local inertial frames at any point (event) in spacetime. From this infinite number, we choose a few frames that are useful for our purpose of making observations near a spinning black hole. The local rest frame (subscript: restD) helped to get us from the rain frame to the local static frame (subscript: statD), but has little further usefulness. Therefore we do not include the local rest frame in the exercises of this chapter or in later chapters about the spinning black hole.

In Query 19 you predict results of some measurements that observers can make in the local rain, static, and ring frames.

QUERY 19. Observations from local frames.

A. A stone is at rest in the local rain frame. What are the components of its velocity in the local static frame and in the local ring frame? What is its (scalar) speed in each of these frames?

B. A stone is at rest in the local static frame. What are the components of its velocity in the local rain frame and in the local ring frame? What is its (scalar) speed in each of these frames?

C. A stone is at rest in the local ring frame. What are the components of its velocity in the local rain frame and in the local static frame? What is its (scalar) speed in each of these frames?

D. Think of a static *ray* of stones, that is a set of stones with different r values but the same Φ values. Is this ray vertical in the local ring frame (with $\Delta x_{\text{ring}} = 0$ but $\Delta y_{\text{ring}} \neq 0$)? Is this ray vertical in the local rain frame (with $\Delta x_{\text{rain}} = 0$ but $\Delta y_{\text{rain}} \neq 0$)? Is it vertical in the local static frame (with $\Delta x_{\text{statD}} = 0$ but $\Delta y_{\text{statD}} \neq 0$)?

17.9 ■ APPENDIX A: MAP ENERGY OF A STONE IN DORAN COORDINATES

Derived using the Principle of Maximal Aging

We now show that the free stone has two global constants of motion: map energy and map angular momentum, just as the stone has as it moves around the non-spinning black hole. Happily we already have a well-honed routine for finding these constants of motion, most recently for the non-spinning black hole in Sections 6.2 and 8.2.

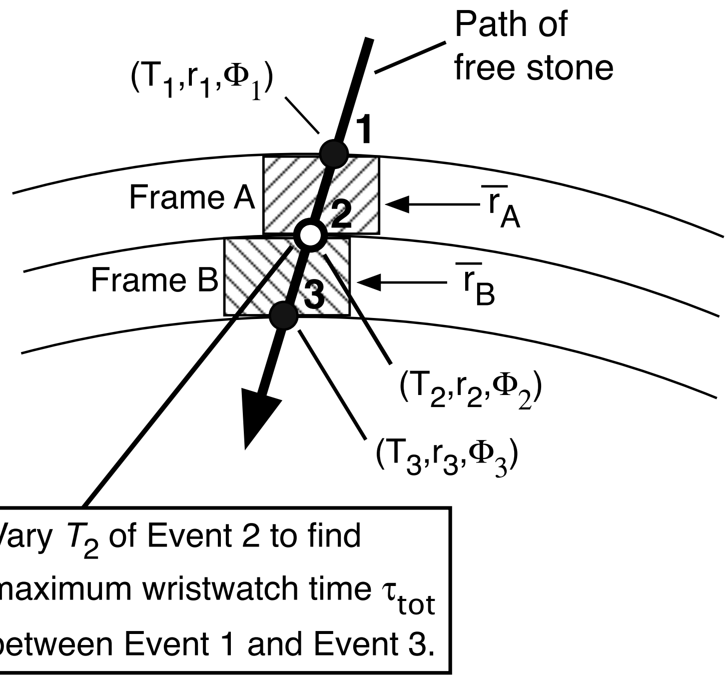


FIGURE 8 Use the Doran metric plus the Principle of Maximal Aging to derive the expression for map energy. Adaptation of Figure 3 in Section 6.2. Why does this arrow point at an angle, rather than vertically downward? See Objection 6.

Derive E and L using the Principle of Maximal Aging.

As usual, to derive map energy and map angular momentum we apply the Principle of Maximal Aging to the motion of the stone across two adjacent local inertial frames. This section adapts the procedure carried out for a non-spinning black hole in Section 6.2.

PREVIEW OF MAP ENERGY DERIVATION (Figure 8)

1. The stone enters the above local inertial Frame A at Event 1 with map coordinates (T_1, r_1, Φ_1).
2. The stone moves straight across the above inertial Frame A in time lapse τ_A measured on its wristwatch.
3. The stone crosses from the above inertial Frame A to the below inertial Frame B at Event 2 with map coordinates (T_2, r_2, Φ_2).
4. The stone moves straight across the below inertial Frame B in time lapse τ_B measured on its wristwatch.
5. The stone exits the below inertial frame at Event 3 with map coordinates (T_3, r_3, Φ_3) .
6. Use the Principle of Maximal Aging to define map energy of the stone: Vary *only* the value of T_2 at the boundary between above and below frames to maximize the total wristwatch time τ_{tot} across both frames.

The total wristwatch time τ_{tot} across both local frames is the sum of wristwatch times across the above and below frames:

$$\tau_{\text{tot}} \equiv \tau_{\text{A}} + \tau_{\text{B}} \tag{97}$$

To find the path of maximal aging, set to zero the derivative of τ_{tot} with respect to T_2:

$$\frac{d\tau_{\text{tot}}}{dT_2} = \frac{d\tau_{\text{A}}}{dT_2} + \frac{d\tau_{\text{B}}}{dT_2} = 0 \tag{98}$$

or

$$\frac{d\tau_{\text{A}}}{dT_2} = -\frac{d\tau_{\text{B}}}{dT_2} \tag{99}$$

Write approximate versions of metric (5) for the above and below patches; spell out only those terms that contain T. In the following, ZZ means "terms that do not contain T."

$$\begin{aligned}\tau_{\text{A}} \approx \Bigg[&\left(1-\frac{2M}{\bar{r}_{\text{A}}}\right)(T_2-T_1)^2 - 2\left(\frac{2M\bar{r}_{\text{A}}}{\bar{r}_{\text{A}}^2+a^2}\right)^{1/2}(T_2-T_1)(r_2-r_1) \\ &+2\left(\frac{2Ma}{\bar{r}_{\text{A}}}\right)(T_2-T_1)(\Phi_2-\Phi_1) + ZZ\Bigg]^{1/2}\end{aligned} \tag{100}$$

$$\begin{aligned}\tau_{\text{B}} \approx \Bigg[&\left(1-\frac{2M}{\bar{r}_{\text{B}}}\right)(T_3-T_2)^2 - 2\left(\frac{2M\bar{r}_{\text{B}}}{\bar{r}_{\text{B}}^2+a^2}\right)^{1/2}(T_3-T_2)(r_3-r_2) \\ &+2\left(\frac{2Ma}{\bar{r}_{\text{B}}}\right)(T_3-T_2)(\Phi_3-\Phi_2) + ZZ\Bigg]^{1/2}\end{aligned} \tag{101}$$

All coordinates are fixed except T_2. When we take the derivative of these two expressions with respect to T_2, the resulting denominators are simply τ_{A} and τ_{B}, respectively:

$$\frac{d\tau_{\text{A}}}{dT_2} \approx \frac{\left(1-\frac{2M}{\bar{r}_{\text{A}}}\right)(T_2-T_1) - \left(\frac{2M\bar{r}_{\text{A}}}{\bar{r}_{\text{A}}^2+a^2}\right)^{1/2}(r_2-r_1) + \left(\frac{2Ma}{\bar{r}_{\text{A}}}\right)(\Phi_2-\Phi_1)}{\tau_{\text{A}}} \tag{102}$$

$$\frac{d\tau_{\text{B}}}{dT_2} \approx -\frac{\left(1-\frac{2M}{\bar{r}_{\text{B}}}\right)(T_3-T_2) - \left(\frac{2M\bar{r}_{\text{B}}}{\bar{r}_{\text{B}}^2+a^2}\right)^{1/2}(r_3-r_2) + \left(\frac{2Ma}{\bar{r}_{\text{B}}}\right)(\Phi_3-\Phi_2)}{\tau_{\text{B}}} \tag{103}$$

Note the initial minus sign on the right side of the second equation.

Now substitute these two equations into (99). The minus signs cancel to yield expressions of similar form on both sides of the equation. *Result:* The expression on the left side of (99) depends only on $\bar{r}_{\rm A}$ plus differences in the global coordinates across that local inertial frame. The expression on the right side of (99) depends only on $\bar{r}_{\rm B}$ plus corresponding differences in the global coordinates across that frame. In other words, we have found an expression in global coordinates that has the same form and the same value in two adjacent frames; it is a **map constant of the motion** (Comment 6, Section 1.11). We call this expression **map energy:** E/m. Shrink the differences to differentials (Comment 4, Section 1.7). Map energy becomes:

Map energy in Doran coordinates

$$\boxed{\frac{E}{m} = \left(1 - \frac{2M}{r}\right)\frac{dT}{d\tau} - \left(\frac{2Mr}{r^2+a^2}\right)^{1/2}\frac{dr}{d\tau} + \frac{2Ma}{r}\frac{d\Phi}{d\tau}} \qquad (104)$$

QUERY 20. Cleanup questions for map energy of a stone.

A. Why do we give the name E/m to the expression on the right side of (20)? Verify that for $r \gg 2M$, that is in flat spacetime, this expression reduces to $E/m = dt/d\tau$, the special relativity expression for energy—equation (23) in Section 1.7.

B. Show that for the non-spinning black hole equation (20) for E/m reduces to equation (35) in Section 7.5.

Map energy

The map energy E of a free stone on the left side of (20) is a constant of motion whose numerical value is independent of the global coordinate system. The *form* of the right side, however, looks different when expressed in different global coordinate systems.

?

Objection 6. *In your derivation of map energy for the non-spinning black hole in Section 6.2, the arrow pointed vertically downward. Why does the arrow in Figure 8 in the present chapter point in another direction?*

!

A perceptive question! The term *ZZ* in both equations (100) and (101) represents "terms that do not contain T." Now look at the fourth term on the right side of global metric (5). This term does not contain dT, but it does contain $d\Phi$, so this term would be eliminated if the arrow in Figure 8 pointed vertically downward (for which $d\Phi = 0$). With this error, equation (20) for map energy would be incomplete; it would not contain the term that ends with $d\Phi/d\tau$. You can show that this complication does not exist in the earlier derivation of map energy for the non-spinning black hole (Section 6.2).

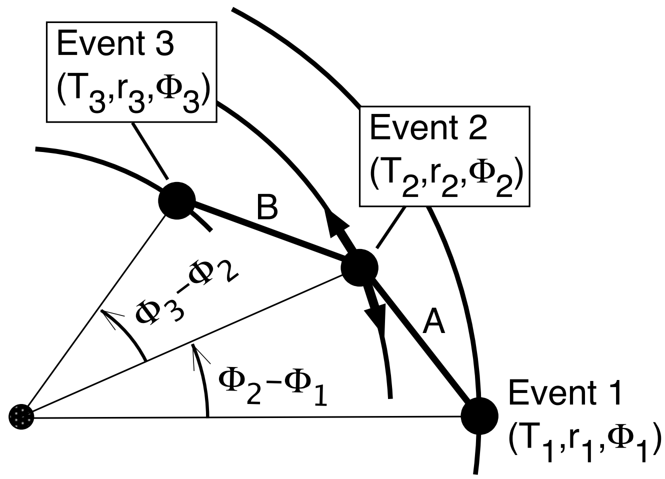


FIGURE 9 Use the Principle of Maximal Aging to derive the expression for map angular momentum in Doran coordinates. Vary Φ_2 of Event 2 to find the Φ-coordinate that leads to maximum τ_{tot} along worldline segments A and B between Events 1 and 3. Adaptation of Figure 2 in Section 8.2.

17.10 ■ APPENDIX B: MAP ANGULAR MOMENTUM OF A STONE IN DORAN COORDINATES

Again, use the Principle of Maximal Aging

To derive the expression for map angular momentum in Doran coordinates, our overall strategy closely follows that of the derivation of E/m in Section 17.9, with the notation shown in Figure 9. Run your finger down the Summary of Map Energy Derivation in Section 17.9 to preview the parallel derivation here.

In this case let the adjacent local inertial frames straddle the straight segments A and B in Figure 9. Write approximate versions of metric (5); spell out only those terms that contain Φ. In the following equations, YY stands for "terms that do not contain Φ."

$$\tau_{\text{A}} \approx \left[2\left(\frac{2Ma}{\bar{r}_{\text{A}}}\right)(T_2 - T_1)(\Phi_2 - \Phi_1) + 2a\left(\frac{2M\bar{r}_{\text{A}}}{\bar{r}_{\text{A}}^2 + a^2}\right)^{1/2}(r_2 - r_1)(\Phi_2 - \Phi_1) - \bar{R}_{\text{A}}^2(\Phi_2 - \Phi_1)^2 + YY \right]^{1/2} \qquad (105)$$

$$\tau_{\text{B}} \approx \left[2\left(\frac{2Ma}{\bar{r}_{\text{B}}}\right)(T_3 - T_2)(\Phi_3 - \Phi_2) + 2a\left(\frac{2M\bar{r}_{\text{B}}}{\bar{r}_{\text{B}}^2 + a^2}\right)^{1/2}(r_3 - r_2)(\Phi_3 - \Phi_2) - \bar{R}_{\text{B}}^2(\Phi_3 - \Phi_2)^2 + YY \right]^{1/2} \qquad (106)$$

All event coordinates are fixed except for Φ_2. To apply the Principle of Maximal Aging, take the derivatives of both these expressions with respect to Φ_2 and set the resulting sum equal to zero:

$$\frac{d\tau_{\text{tot}}}{d\Phi_2} = \frac{d\tau_{\text{A}}}{d\Phi_2} + \frac{d\tau_{\text{B}}}{d\Phi_2} = 0 \tag{107}$$

or

$$\frac{d\tau_{\text{A}}}{d\Phi_2} = -\frac{d\tau_{\text{B}}}{d\Phi_2} \tag{108}$$

Take these derivatives with respect to Φ_2 of each expression in (105) and (106). The resulting two equations have τ_{A} and τ_{B} in the denominator, respectively:

$$\frac{d\tau_{\text{A}}}{d\Phi_2} \approx \frac{\left(\dfrac{2Ma}{\bar{r}_{\text{A}}}\right)(T_2 - T_1) + a\left(\dfrac{2M\bar{r}_{\text{A}}}{\bar{r}_{\text{A}}^2 + a^2}\right)^{1/2}(r_2 - r_1) - \bar{R}_{\text{A}}^2(\Phi_2 - \Phi_1)}{\tau_{\text{A}}} \tag{109}$$

$$\frac{d\tau_{\text{B}}}{d\Phi_2} \approx -\frac{\left(\dfrac{2Ma}{\bar{r}_{\text{B}}}\right)(T_3 - T_2) + a\left(\dfrac{2M\bar{r}_{\text{B}}}{\bar{r}_{\text{B}}^2 + a^2}\right)^{1/2}(r_3 - r_2) - \bar{R}_{\text{B}}^2(\Phi_3 - \Phi_2)}{\tau_{\text{B}}} \tag{110}$$

Note the initial minus sign on the right side of the second equation.

Now substitute these two equations into (108). The minus signs cancel, yielding expressions of similar form on both sides of the equation. *Result:* The left side of (108) depends only on $\bar{r}_{\text{A}}$ plus differences in the global coordinates across that frame. The right side of (108) depends only on $\bar{r}_{\text{B}}$ plus corresponding differences in the global coordinates across that frame. In other words, we have found an expression in global coordinates that—in this approximation—has the same form and the same value in two adjacent frames. Shrink to differentials and the expression becomes exact. It is another constant of motion, which we call **map angular momentum**:

Map angular momentum in Doran coordinates

$$\frac{L}{m} = R^2\frac{d\Phi}{d\tau} - \frac{2Ma}{r}\frac{dT}{d\tau} - a\left(\frac{2Mr}{r^2 + a^2}\right)^{1/2}\frac{dr}{d\tau} \tag{111}$$

Comment 9. The sign of L/m: our choice
Notice that the right side of (21) is the negative of what we would expect, given its derivation from (109) and (110). The sign of L/m is arbitrary, our choice because either way L/m is constant for a free stone. We choose the minus sign so that when r becomes large, L/m is positive when the tangential component of motion is in the positive (counterclockwise) Φ direction. Recall the discussion after equation (1).

Map angular momentum

The map angular momentum L/m of a free stone, on the left side of (21), is a constant of motion whose numerical value is independent of the global coordinate system. The *form* of the right side, however, will look different when expressed in different global coordinate systems.

QUERY 21. Cleanup questions for map angular momentum of a stone.
Why do we give the name L/m to the expression on the right side of (21)? Verify that *either* for $r \gg 2M$ (far from the spinning black hole) *or* for $a \to 0$ (the non-spinning black hole) this expression reduces to $L/m = r^2 d\phi/d\tau$, the expression for the non-spinning black hole—equation (10) in Section 8.2.

17.11 ■ PROJECT: BOYER-LINDQUIST GLOBAL COORDINATES

Metric in Boyer-Lindquist coordinates

In 1963 Roy Kerr published his paper that first contained a global metric for the spinning black hole. In 1967 R. H. Boyer and R. W. Lindquist published a global metric that simplifies the form of Kerr's original metric. Here it is, expressed in so-called **Boyer-Lindquist global coordinates**. As usual, for simplicity we restrict global coordinates and their metric to a slice through the equatorial plane of the black hole, perpendicular to its axis of rotation.

$$d\tau^2 = \left(1 - \frac{2M}{r}\right) dt^2 + \frac{4Ma}{r} dt d\phi - \frac{dr^2}{H^2} - R^2 d\phi^2 \quad \text{(Boyer-Lindquist...} \quad (112)$$
$$-\infty < t < \infty, \quad 0 < r < \infty, \quad 0 \leq \phi < 2\pi \qquad \text{...on the equatorial slice)}$$

Box 2 defines H^2 and R^2. Global ϕ has the same meaning as it does in the global rain metric for the non-spinning black hole, equation (32) in Section 7.5.

Comment 10. Why not use Boyer-Lindquist coordinates?
The Boyer-Lindquist metric (112) has only one cross term instead of all possible cross terms in the Doran metric (5). Why does this chapter use and develop the consequences of this complicated Doran metric? The first term on the right of (112) tells why: this term goes to zero as $r \to 2M^+$. As a result, Boyer-Lindquist map time t increases without limit along the worldline of a descending stone as it approaches $r = 2M$. This is the same inconvenience we found in the Schwarzschild metric for the non-spinning black hole. To avoid this problem, in Chapter 7 we converted from Schwarzschild coordinates to global rain coordinates. We could have carried out the same sequence in the present chapter: begin with the Boyer-Lindquist metric, then convert to the Doran metric. But this conversion is an algebraic mess (with the simple result given in the following exercise). Instead, we chose to start immediately with the Doran metric and to relegate investigation of the Boyer-Lindquist metric to these exercises.

BL-1. Conversion from Doran coordinates to Boyer-Lindquist global coordinates

Substitute the following expressions into the Doran global metric and simplify the results to show that the outcome is the Boyer-Lindquist metric (112):

$$dT = dt + \frac{R\beta}{rH^2}dr \tag{113}$$

$$d\Phi = d\phi + \frac{\omega R}{rH^2\beta}dr \tag{114}$$

BL-2. Limiting cases of the Boyer-Lindquist metric

A. Show that for zero spin angular momentum ($a = 0$), the Boyer-Lindquist metric (112) reduces to the Schwarzschild metric, equation (6) in Section 3.1.

B. Show that the Boyer-Lindquist metric for a maximum-spin black hole ($a = M$) takes the form

$$d\tau^2 = \left(1 - \frac{2M}{r}\right)dt^2 + \frac{4M^2}{r}dtd\phi - \frac{dr^2}{H^2_{\text{max}}} - R^2_{\text{max}}d\phi^2 \quad (a = M) \tag{115}$$

BL-3. Tetrad form of the Boyer-Lindquist metric

To put the Boyer-Lindquist metric into a tetrad form, eliminate the $dtd\phi$ cross term by completing the square: Add and subtract a function $G(r)d\phi^2$ to terms on the right side of the metric, then define $G(r)$ to eliminate the cross term.

Show that the resulting tetrad form of the Boyer-Lindquist metric is:

$$d\tau^2 = \left(1 - \frac{2M}{r}\right)^{-1}\left[\left(1 - \frac{2M}{r}\right)dt + \frac{2Ma}{r}d\phi\right]^2$$
$$- \frac{dr^2}{H^2} - \left(1 - \frac{2M}{r}\right)^{-1}\left[R^2\left(1 - \frac{2M}{r}\right) + \frac{4M^2a^2}{r^2}\right]d\phi^2 \qquad \text{(Boyer-Lindquist)} \tag{116}$$

BL-4. Local shell frame in Boyer-Lindquist coordinates

A. Adapt equation (14) to simplify the coefficient of $d\phi^2$ in (116).

B. Use the results of Item A and exercise 2 to derive the following local shell coordinates in Boyer-Lindquist coordinates.

$$\Delta t_{\text{shell}} \equiv \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2}\left[\left(1 - \frac{2M}{\bar{r}}\right)\Delta t + \frac{2Ma}{\bar{r}}\Delta\phi\right] \tag{117}$$

$$\Delta y_{\text{shell}} \equiv \frac{\Delta r}{\bar{H}} \qquad \text{(Boyer-Lindquist)} \tag{118}$$

$$\Delta x_{\text{shell}} \equiv \left(1 - \frac{2M}{\bar{r}}\right)^{-1/2}\bar{r}\bar{H}\Delta\phi \tag{119}$$

C. How do we know that equations (117) through (119) define a local *shell* frame and not, for example, a local ring frame or rain frame?

E. Show that as $a \to 0$ equations (117) through (119) recover shell frame expressions in global rain coordinates (Section 7.5).

Comment 11. Shell frame in Doran coordinates.
You can use conversion equations (113) and (114) to express local shell coordinates in Doran global coordinates. Like equations (117) and (119), the resulting equations show that shell frames exist only outside the static limit.

BL-5. Local ring frame in Boyer-Lindquist coordinates

A. Show that the following tetrad form reduces to the Boyer-Lindquist metric (112):

$$d\tau^2 = \left(\frac{rH}{R}\right)^2 dt^2 - \frac{dr^2}{H^2} - R^2\left[d\phi - \omega(r)dt\right]^2 \quad \text{(Boyer-Lindquist)} \quad (120)$$

where Box 1 defines $\omega(r) \equiv 2Ma/(rR^2)$.

B. Individual terms in (120) allow us to define the local ring frame:

$$\Delta t_{\text{ring}} \equiv \frac{\bar{r}\bar{H}}{\bar{R}}\Delta t \qquad \text{(Boyer-Lindquist)} \quad (121)$$

$$\Delta y_{\text{ring}} \equiv \frac{\Delta r}{\bar{H}} \quad (122)$$

$$\Delta x_{\text{ring}} \equiv \bar{R}\,(\Delta\phi - \bar{\omega}\Delta t) \quad (123)$$

C. Use transformations (113) and (114) to show that Boyer-Lindquist ring equations (121) through (123) imply Doran ring equations (77) through (79).

D. What is the measurable relative velocity, call it v_{ring}, between local ring coordinates and local shell coordinates?

E. Show that as $a \to 0$ equations (121) through (123) recover shell frame expressions in global rain coordinates (Section 7.5).

BL-6. Local rain frame in Boyer-Lindquist coordinates

A. Substitute the Δ forms of equations (113) and (114) into equations (32) through (34) to obtain the following expressions for local rain coordinates in Boyer-Lindquist coordinates:

$$\Delta t_{\text{rain}} = \Delta t + \beta\frac{\bar{R}}{\bar{r}\bar{H}^2}\Delta r \quad (124)$$

$$\Delta y_{\text{rain}} = \frac{\bar{R}}{\bar{r}\bar{H}^2}\Delta r + \beta\Delta t \quad (125)$$

$$\Delta x_{\text{rain}} = \Delta x_{\text{ring}} = \bar{R}\,(\Delta\phi - \bar{\omega}\Delta t) \quad (126)$$

B. Use these equations to write the Boyer-Lindquist metric in tetrad form.

BL-7. Not "at rest" in both global coordinates

Show that a stone at rest in Boyer-Lindquist global coordinates ($dr = d\phi = 0$) is not at rest in Doran global coordinates; in particular, $d\Phi \neq 0$ for that stone.

BL-8. Boyer-Lindquist metric for $M = 0$.

Show that when the mass of the spinning black hole gets smaller and smaller, $M \to 0$ in (112), but the angular momentum parameter a keeps a constant value, then the Boyer-Lindquist metric becomes equal to the Doran metric under the same limits, as examined in Exercies 3.

17.12 ■ EXERCISES

1. Our Sun as a black hole

Suppose that our Sun collapses into a spinning black hole without blowing off any mass. What is the value of its spin parameter a/M? The magnitude of the Sun's angular momentum is approximately:

$$J_{\text{Sun}} \approx 1.63 \times 10^{41} \quad \text{kilogram meters}^2/\text{second} \tag{127}$$

A. Use equation (10) in Section 3.2 to convert kilograms to meters. The result to one significant digit is $J = 1 \times 10^{14}$ meters^3/second. Derive the answer to three significant digits. [My answer: 1.21×10^{14} meters^3/second]

B. Divide your answer to Item A by c to find the angular momentum of the Sun in units of meters^2.

C. Divide the result of Item B by the square of the mass of our Sun in meters (inside the front cover) to show that $a_{\text{Sun}}/M_{\text{Sun}} = 0.185$.

2. Ring frame time for one rotation

How does someone riding in the ring frame know that she is revolving around the spinning black hole? She can tell because the same pattern of stars overhead repeats sequentially, separated by ring frame time we can call Δt_{ring1}. Derive an expression for Δt_{ring1} using the following outline or some other method:

A. The observer is stationary in the ring frame. Show that this means that $\Delta r = 0$ and $\Delta\Phi = \bar{\omega}\Delta T$.

B. Show from equation (77) and results of Item A that, for one rotation, that is for $\Delta\Phi = 2\pi$:

$$\Delta t_{\text{ring1}} = \frac{\bar{r}\bar{H}}{\bar{R}}\Delta T = \frac{2\pi(\bar{r}\bar{H})}{\bar{R}\bar{\omega}} \qquad \text{(in meters)} \tag{128}$$

C. Substitute for the various factors in (128) to obtain

$$\Delta t_{\text{ring1}} = \frac{\pi \bar{R}\bar{r}}{Ma}\left(\bar{r} - r_{\text{EH}}\right)^{1/2}\left(\bar{r} - r_{\text{CH}}\right)^{1/2} \quad \text{(meters)} \qquad (129)$$

$$= \frac{\pi M}{a^*} R^* r^* \left[\left(r^* - r^*_{\text{EH}}\right)\left(r^* - r^*_{\text{CH}}\right)\right]^{1/2} \quad \text{(meters)} \qquad (130)$$

Equation (130) uses unitless variables, for example $r^* \equiv r/M$, and for simplicity we have deleted the average value bar over the symbols.

D. For a spinning black hole of mass $M = 10M_{\text{Sun}}$ and spin $a^* = a/M = (3/4)^{1/2}$, find the ring rotation times for one rotation at ring r-values given in items (b) through (f) in the following list. Express your results in both meters and seconds.

(a) Show that $\pi M/a^* = 5.369 \times 10^4$ meters.
(b) $r^* = 10^3$
(c) $r^* = 10$
(d) $r^* = 3$
(e) $r^* = 1.51$
(f) $r^* = 0.25$

Notice that each of these short times is measured in the local inertial ring frame.

E. For the spinning black hole in Item D, what is the value of Δt_{ring1} for a ring at the radius of Mercury around our Sun? Use Mercury orbit values in Chapter 10. Compare this value of Δt_{ring1} for our spinning black hole with the orbital period of Mercury around our Sun.

F. Equation (130) tells us that, for a given value of a^*, the ring frame time for one rotation of the ring is proportional to the mass M of the black hole. As a result, you can immediately write down the corresponding times Δt_{ring1} for Item D around the spinning black hole at the center of our galaxy whose mass $M = 4 \times 10^6 M_{\text{Sun}}$. Assume that the (unknown) value of its spin parameter $a^* = (3/4)^{1/2}$.

3. Distance between rings measured by a rain observer

A rain observer measures the distance between two adjacent concentric rings around a spinning black hole. The two rings are separated by dr in Doran r-coordinate. The rain observer their distance in two distinct ways:

[1] As she travels past the two rings, she measures, on her wristwatch, the time d? it takes her to get from the outer ring to the inner ring. She knows her speed vrel relative to the two adjacent rings. She then calculates the distance between the two adjacent rings from these two numbers.

[2] During her short travel through the two adjacent rings she is in a local inertial rain frame. She considers two events along the yrain axis in this frame: one takes place on the inner ring, the other on the outer ring, and they

simultaneous as measured in her local inertial rain frame. She then determines the distances between the rings as the separation of yrain-coordinates between these two events.

A. Write an expression for distance ds between the two adjacent rings, according to her first measurement technique? [Hint: Use (26) through (28) and (43).]

B. What is the distance ds between the two adjacent rings, according to her second measurement technique? [Hint: Use (32) through (34).] Show that the two techniques give the same result for the distance between the two rings as measured by a rain observer.

C. Take the limit of ds as $a \to 0$, and compare the result with Box 5 in Chapter 7 which suggested that for a non-spinning black hole the distance between two adjacent shells as measured by a rain observer is $ds = dr$, where dr is the incremental difference in Schwarzschild r-coordinate between the two shells.?

4. Raindrop speed measured in local inertial ring frame

Use (95) and your favorite plotting program to plot the speed of a raindrop measured in a local inertial ring frame, as a function of the Doran r-coodinate of that ring frame, for each of the following black hole spin parameters:

- (a) $a/M = 0$ (non-spinning black hole). Compare this plot with Figure 2 in Chapter 6.
- (b) $a/M = (3/2)^{1/2}$
- (c) $a/M = 1$ (maximally spinning black hole)

Show that wherever a local inertial ring frame can be constructed, the speed of the raindrop measured in that frame does not exceed the speed of light. At what r-values does the measured speed of the raindrop reach the speed of light?

5. Relative orientation of local ring frame and local rest frame axes

Table 1 shows that the velocity of a raindrop measured in the local ring frame points along the Δy_{rain} axis. Table 1 also tells us that the velocity of the same raindrop measured in the local rest frame points along the Δy_{rest} axis. Does this mean that the spatial axes in the local ring frame have the same orientation as the spatial axes in the local rest frame? Isn?t this in contradiction with Figure 7, which implies that the orientation of the spatial axes in the local ring frame matches the orientation of spatial axes in the local static frame?

6. Stone released from rest on a local ring frame

Release a stone from rest in a local ring frame at Doran coordinate r_0. Derive an expression for the velocity v_{ring} of the stone measured in a local ring frame as a function of the Doran r-coordinate of that ring frame ($r < r_0$). Show that in the limit in which the stone drops from rest far away ($r_0 \to \infty$), the expression for the velocity of the stone reduces to expression (95) for a raindrop.

7. Stone hurled inward from a local ring frame far away

Hurl a stone inward with velocity components $v_{\text{ring},x} = 0$ and $v_{\text{ring},y} = -v_{\text{far}}$ from a local inertial ring frame far away from a spinning black hole.

A. Derive an expression for the velocity components of the stone measured in a local ring frame as a function of the Doran r-coordinate of that ring frame.

B. Show that in the limit in which the stone drops *from rest* in a ring frame far away ($v_{\text{far}} \to 0$), the expression for the velocity of the stone reduces to expression (95) for a raindrop.

8. Tetrad form of the Doran global metric

A. From equations (77) through (79), write down the corresponding tetrad form of the Doran global metric.

B. Multiply out the resulting global metric to verify that the result is Doran metric (5).

9. Doran metric for $M \to 0$

Let the mass of the spinning black hole get smaller and smaller, $M \to 0$, while the angular momentum parameter a retains a a constant value. Then metric (5) becomes:

$$d\tau^2 = dT^2 - \frac{r^2}{r^2 + a^2} dr^2 - \left(r^2 + a^2\right) d\Phi^2 \qquad (M = 0) \qquad (131)$$

Does metric (131) represent flat spacetime? To find out we show a coordinate transformation that reduces (131) to an inertial metric in flat spacetime. Let

$$\rho \equiv \left(r^2 + a^2\right)^{1/2} \qquad (132)$$

The last term in metric (131) becomes $\rho^2 d\Phi^2$ and ρ is the reduced circumference.

A. Take the differential of both sides of (132) and substitute the result for the second term on the right side of (131). Show that the outcome is the metric

$$d\tau^2 = dt^2 - d\rho^2 - \rho^2 d\Phi^2 \qquad (M = 0) \qquad (133)$$

The global metric (131) has been transformed to the globally flat form (133). This is *not* the metric of a local frame; it is a *global metric*—but with a strange exclusion, discussed in the following Items.

B. Does the spatial part of the metric (133) describe the Euclidean plane? To describe Euclidean space, that spatial part of the metric

$$ds^2 = d\rho^2 + \rho^2 d\Phi^2 \qquad \text{(Euclid)} \qquad (134)$$

must, by definition, be valid for the full range of ρ, the radial coordinate in equation (134), namely $0 \leq \rho < \infty$. But this is not so: Definition (132) tells us that $\rho = a$, when $r = 0$. So global metric (131) is undefined for $0 < \rho < a$. Can we "do science"—that is, carry out measurements—in the region $0 < \rho < a$?

C. Is $\rho = 0$ *actually* a point or a ring? What is the meaning of the word *actually* when we describe spacetime with (arbitrary!) map coordinates.

D. Does the Doran metric for $M \to 0$ but $a > 0$ reduce to the flat spacetime metric of special relativity? Show that the answer is no, that the black hole spin remains imprinted on spacetime like the Cheshire cat's grin after its body—the mass—fades away.

10. Free stone vs. powered spaceship vs. light

Review Section 17.3, A stone's throw. Which formulas in that section describe only a free stone? Which formulas apply generally to any object with nonzero mass (free stone, powered spaceship, etc.)? Which formulas apply to light also? [*Hint:* The metric describes nearby events along the worldline of any object: free stone, powered spaceship, or light ray. The Principle of Maximal Aging is valid only for objects that move freely.]

11. Toy model of a pulsar

A **pulsar** is a spinning neutron star that emits electromagnetic radiation in a narrow beam. We observe the pulsar only if the beam sweeps across Earth. Box 5 in Section 3.3 tells us that "General relativity significantly affects the structure and oscillations of the neutron star." In particular, the neutron star has a maximum spin rate related to a_{max} for a black hole—equation (3). Let the neutron star have the mass of our Sun with the surface at $R = 10$ kilometers. Use Newtonian mechanics to make a so-called *toy model* of a pulsar—that is, a rough first approximation to the behavior of a non-Newtonian system. The pulsar PSR J1748-2446, located in the globular cluster called Terzan 5, rotates at 716 hertz $\equiv$ 716 revolutions per second. Set the neutron star's angular momentum to that of a uniform sphere rotating at that rate and call the result "our pulsar." Then the angular momentum, as a function of the so-called **moment of inertia** I_{sphere} and spin rate ω radians per second is:

$$J \equiv I_{\text{sphere}}\omega = \left(\frac{2M}{5}M_{\text{kg}}R^2\right)\omega \qquad \text{(Newton, conventional units)} \quad (135)$$

Our pulsar spins once in Newton universal time $t = 1.40$ millisecond. Use numerical tables inside the front cover to answer the following questions:

A. What is the value of our pulsar's angular momentum in conventional units?

B. Express the our pulsar's angular momentum in meters^2.

C. Find the value of $J/(Ma_{\text{max}}) = J/M^2$ for our pulsar, where M is in meters.

D. Suppose that our pulsar collapses to a black hole. Explain why it would have to blow off some of its mass to complete the process.

12. Spinning baseball a naked singularity?

A standard baseball has a mass $M = 0.145$ kilogram and radius $r_{\text{b}} = 0.0364$ meter. The Newtonian expression for the spin angular momentum of a sphere of uniform density is, in conventional units

$$J_{\text{conv}} = I_{\text{conv}}\omega = \frac{2}{5}M_{\text{kg}}r_{\text{b}}^2\omega = \frac{4\pi M_{\text{kg}}r_{\text{b}}^2}{5}f \qquad \text{(Newton)} \qquad (136)$$

where ω is the rotation rate in radians per second. The last step makes the substitution $\omega = 2\pi f$, where f is the frequency in rotations per second. We want to find the value of the angular momentum parameter $a = J/M$ in meters. Begin by dividing both sides of (136) by the baseball's mass M_{kg}:

$$\frac{J_{\text{conv}}}{M_{\text{kg}}} = \frac{4\pi r_{\text{b}}^2}{5}f \qquad \text{(Newton: conventional units)} \qquad (137)$$

The units of the right side of (137) are meters^2/second. Convert to meters by dividing through by c, the speed of light, to obtain an expression for a:

$$a \equiv \frac{J}{M} = \frac{4\pi r_{\text{b}}^2}{5c}f \qquad \text{(Newton: units of meters)} \qquad (138)$$

A. Insert numerical values to show the result in the unit meter:

$$a = 1.1 \times 10^{-11}\ \text{second} \times f \qquad \text{(Newton: units of meters)} \quad (139)$$

B. We want to know if a is greater than the mass of the baseball. What is the mass M of the baseball in meters? [My answer: 1.1×10^{-28} meter.]

C. Suppose that a pitched or batted baseball spins at 4 rotations per second. What is the value of a for this flying ball? [My answer: 4.4×10^{-11} meter.] Does this numerical value violate the limits on the spin angular momentum parameter a for a spinning black hole? [My answer: And how!]

QUESTION: Is this baseball a naked singularity?
ANSWER: No, because the Doran metric is valid only in curved *empty* space; it does not apply inside a baseball. ("Outside of a dog, a book is man's best friend. Inside of a dog it's too dark to read." –Groucho Marx)

D. What is the value of r/M at the surface of the baseball, that is, what is the value of r_b/M? Calculate the resulting value of H^2 at the surface of the baseball. What is the value of R^2/M^2 at this surface?

E. Divide Doran metric (5) through by M^2 to make it unitless. At the surface of the baseball, determine how much each term in the resulting metric differs from the corresponding term for flat spacetime:

$$\left(\frac{d\tau}{M}\right)^2 = \left(\frac{dT}{M}\right)^2 - \left(\frac{dr}{M}\right)^2 - \left(\frac{r}{M}\right)^2 d\Phi^2 \qquad \text{(flat spacetime)} \quad (140)$$

F. Will the gravitational effects of the baseball's spin be noticeable to the fielder who catches the spinning ball?

G. Use equation (12) and the values of M and a calculated in Items B and C to calculate the $\omega_{\text{framedragging}}$ function that expresses the "frame dragging effect" of this baseball at its surface. How many orders of magnitude is this greater or less than ω_{rotation}, the angular speed of the spinning baseball.

13. Spinning electron a naked singularity?

The electron is a quantum particle; Einstein's classical (non-quantum) general relativity cannot predict results of experiments with the electron. Ignore these limitations in this exercise; treat the electron as a classical particle.

The electron has mass $m_e = 9.12 \times 10^{-31}$ kilogram and spin angular momentum $J_e = \hbar/2$, where the value of "h-bar," $\hbar = 1.05 \times 10^{-34}$ kilogram-meter^2/second. Calculate the numerical value of the quantity a/m_e for the electron. If the electron is a point particle, then the Doran metric describes the electron all the way down to (but not including) $r = 0$. *Questions:* Is the electron a spinning black hole? Is the electron a naked singularity?

17.13 ■ REFERENCES

Initial quote: S. Chandrasekhar, *Truth and Beauty: Aesthetics and Motivations in Science,* University of Chicago Press, 1987, pages 153–154.

The story of Roy Kerr's discovery of the spinning black hole metric: *Cracking the Einstein Code: Relativity and the Birth of Black Hole Physics* by Fulvio Melia, with an Afterword by Roy Kerr, University of Chicago press, 2009.

The first paper: R. P. Kerr, "Gravitational Field of a Spinning Mass as an Example of Algebraically Special Metrics," *Physical Review Letters,* Volume 11, pages 237–238 (1963)

Introduction of the Boyer-Lindquist metric: Robert H. Boyer and Richard W. Lindquist, "Maximum Analytic Extension of the Kerr Metric," *Journal of Mathematical Physics,* Volume 8, Number 2, pages 265–281 (February 1967). See also Brandon Carter, "Global Structure of the Kerr Family of Gravitational Fields," *Physical Review,* Volume 174, Number 5, pages 1559–1571 (1968).

Introduction of the Doran global metric: Chris Doran, "New form of the Kerr solution," Physical Review D, Volume 61, 2000, pages 067503-1 to 067503-4.

For completeness, the Newman electrically charged black hole: E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence," Metric of a Rotating, Charged Mass," *Journal of Mathematical Physics;* Volume 6, Number 6, pages 918–919 (1965); also E. T. Newman and A. I. Janis, "Note on the Kerr Spinning-Particle Metric," *Journal of Mathematical Physics,* Volume 6, Number 6, pages 915–917 (1965).

Reference for naked singularities:
`http://www.theory.caltech.edu/people/preskill/nyt_bet_story.html`

Matthew W. Choptuik's original discussion of the naked singularity: *Physical Review Letters,* Volume 70, Number 1, 4 January 1993, pages 9-12. Review article: Carsten Gundlach and José M. Martín García, *Living Reviews in Relativity*, Volume 10, pages 5-57.

Download file name: Ch17SpinningBlackHole171227v2.pdf

Chapter 18. Circular Orbits around the Spinning Black Hole

- *How do circular orbits around the spinning black hole differ from those around the non-spinning black hole?*
- *Can a stone in orbit close to the black hole move in a direction opposite to the black hole spin?*
- *Can circular orbits exist inside the event horizon?*
- *Does black hole spin make orbiters go faster? slower?*
- *What happens to material that circles in the accretion disk of a spinning black hole?*
- *Are quasars associated in some way with spinning black holes? If so, how can these structures emit so much radiation?*

Download file name: Ch18CircularOrbitsSpinBH170905v3.pdf

CHAPTER 18 Circular Orbits around the Spinning Black Hole

Edmund Bertschinger & Edwin F. Taylor *

The Mevlevi Order, founded in 1273 by Jalal ad-Din Muhammad Rumi's followers, perform their "dance" and musical ceremony known as the Sama, which involves the whirling from which the order acquires its nickname, Whirling Dervish. The Sama represents a mystical journey of humanity's spiritual ascent. Turning towards the truth, the follower grows through love, deserts ego, finds the truth, and arrives at the "Perfect."

—Wikipedia, The Free Encyclopedia [edited]

18.1 ■ REPRISE: THE DORAN METRIC

Prepare for a trip into the spinning black hole

Prepare to fall into a spinning black hole.

"What's it like to fall into a black hole?" Our first twelve chapters developed answers to this question for the *non-spinning* black hole. We could not give details until Chapter 12, because we needed the background provided by earlier chapters. "What is it like to fall into a *spinning* black hole?" Again, we cannot give details until Chapter 21, because we need the background provided by Chapters 17 through 20.

This chapter: circular orbits

But we can say this now: Falling into the spinning black hole has many more possibilities—and is much more interesting—than falling into the non-spinning black hole. To reach this conclusion we study orbits of stones and light. The present chapter examines *circular orbits* of a stone around the spinning black hole.

Most circular orbits unstable

We find that around the spinning black hole, most of the circular orbits are unstable. An unpowered spaceship can perch temporarily in an unstable circular orbit on its way to a stable circular orbit (Section 18.8).

Blazing accretion disk: a sequence of stable circular orbits

In the *accretion disk* (Section 18.9), gas and dust slowly cascade down through a series of (semi-)stable circular orbits of decreasing r, each successive orbit with slightly smaller orbital energy. Electromagnetic radiation carries away the energy difference between orbits (Section 18.9). We can detect this

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

emitted energy at our location far from the black hole. Eventually however, no circular orbit exists for smaller r, and the accreted material spirals inward across the event horizon.

Doran global metric

To begin, recall the Doran global metric in the equatorial plane of the isolated spinning black hole—equations (4) and (5) in Section 17.2:

$$d\tau^2 = dT^2 - \left[\left(\frac{r^2}{r^2+a^2}\right)^{1/2} dr + \left(\frac{2M}{r}\right)^{1/2}(dT - ad\Phi)\right]^2 - \left(r^2+a^2\right)d\Phi^2 \qquad (1)$$

$$-\infty < T < \infty, \quad 0 < r < \infty, \quad 0 \leq \Phi < 2\pi \qquad \text{(Doran, equatorial plane)}$$

The black hole spin parameter $a \equiv J/M$, with J the angular momentum of the black hole (Section 17.2). The spin parameter a has the unit meter. In Query 1 of Section 17.2 you multiplied out (1) to obtain:

$$d\tau^2 = \left(1 - \frac{2M}{r}\right)dT^2 - 2\left(\frac{2Mr}{r^2+a^2}\right)^{1/2} dTdr + 2a\left(\frac{2M}{r}\right)dTd\Phi \qquad (2)$$

$$+ 2a\left(\frac{2Mr}{r^2+a^2}\right)^{1/2} drd\Phi - \left(\frac{r^2}{r^2+a^2}\right)dr^2 - R^2 d\Phi^2$$

$$-\infty < T < \infty, \quad 0 < r < \infty, \quad 0 \leq \Phi < 2\pi \quad \text{(Doran, equatorial plane)}$$

Equation (6) in Box 1 defines the symbol R.

Comment 1. Heavy algebra
This chapter requires a great deal of algebra to derive many of its equations, algebra that we mostly omit. *Question:* Would more advanced mathematics—for example tensors—make these derivations simpler? *Answer:* We don't think so, but you can try!

18.2 ■ EQUATIONS OF MOTION FOR A STONE; TWO EFFECTIVE POTENTIALS

Algebra orgies lead to powerful results.

Two equations in three unknowns

Our first task is to find equations of motion for a stone in Doran coordinates. Equation (103) for E/m in Section 17.9 and equation (110) for L/m in Section 17.10 give us two linear equations in the three unknowns $dT/d\tau$, $dr/d\tau$, and $d\Phi/d\tau$. Solve them to find $dT/d\tau$ and $d\Phi/d\tau$ as functions of E/m, L/m and $dr/d\tau$. The result is two equations of motion for the stone, both of them functions of the still-undetermined expression for $dr/d\tau$. Box 1, repeated from Section 17.8, provides expressions for H, ω, β, and R in the following equations:

Box 1. Useful Relations for the Spinning Black Hole

This box repeats Box 1 in Section 17.8.

Static limit from Section 17.3:

$$r_S = 2M \qquad (5)$$

Reduced circumference from Section 17.2:

$$R^2 \equiv r^2 + a^2 + \frac{2Ma^2}{r} \qquad (6)$$

Horizon function from Section 17.3:

$$H^2 \equiv \frac{1}{r^2}\left(r^2 - 2Mr + a^2\right) \qquad (7)$$

$$= \frac{1}{r^2}\left(r - r_{EH}\right)\left(r - r_{CH}\right) \qquad (8)$$

where r_{EH} and r_{CH} are r-values of the event and Cauchy horizons, respectively, from Section 17.3.

$$\frac{r_{EH}}{M} \equiv 1 + \left(1 - \frac{a^2}{M^2}\right)^{1/2} \quad \text{(event horizon)} \qquad (9)$$

$$\frac{r_{CH}}{M} \equiv 1 - \left(1 - \frac{a^2}{M^2}\right)^{1/2} \quad \text{(Cauchy horizon)} \qquad (10)$$

Ring omega from Section 17.3:

$$\omega \equiv \frac{2Ma}{rR^2} \qquad (11)$$

An equivalence from Section 17.3:

$$1 - \frac{2M}{r} + R^2\omega^2 = \left(\frac{rH}{R}\right)^2 \qquad (12)$$

Definition of α from Section 17.7:

$$\alpha \equiv \arcsin\left[\left(\frac{2M}{r}\right)^{1/2}\frac{a}{rH}\right] \quad (0 \le \alpha \le \pi/2) \qquad (13)$$

Definition of β from Section 17.8:

$$\beta \equiv \left(\frac{2M}{r}\right)^{1/2}\left(\frac{r^2 + a^2}{R^2}\right)^{1/2} \qquad (14)$$

Box 2 examines the values of some of these expressions at the event and Cauchy horizons.

$$\frac{dT}{d\tau} = \left(\frac{R}{rH}\right)^2\left(\frac{E - \omega L}{m}\right) + \frac{\beta R}{rH^2}\frac{dr}{d\tau} \quad \text{(equations of motion)} \qquad (3)$$

$$\frac{d\Phi}{d\tau} = \frac{1}{(rH)^2}\left[\left(1 - \frac{2M}{r}\right)\frac{L}{m} + \frac{2Ma}{r}\frac{E}{m} + a\left(\frac{2Mr}{r^2 + a^2}\right)^{1/2}\frac{dr}{d\tau}\right] \qquad (4)$$

Find $dr/d\tau$, the third equation of motion.

To find $dr/d\tau$ on the right sides of these equations, divide both sides of the Doran metric (1) by $d\tau^2$; into the result substitute $dT/d\tau$ and $d\Phi/d\tau$ from equations (3) and (4). Extensive algebra leads to the third equation of motion:

$$\frac{dr}{d\tau} = \pm\frac{R}{r}\left(\frac{E - V_L^+}{m}\right)^{1/2}\left(\frac{E - V_L^-}{m}\right)^{1/2} \quad \text{(stone)} \qquad (15)\text{q}$$

Here $V_L^{\pm}(r)$ are the **effective potentials** (two of them!) for the spinning black hole:

$$\frac{V_L^{\pm}(r)}{m} \equiv \omega\frac{L}{m} \pm \frac{rH}{R}\left(1 + \frac{L^2}{m^2R^2}\right)^{1/2} \quad \text{(stone, effective potentials)} \qquad (16)$$

TWO effective potentials

The $\pm$ sign in (16) chooses between the two effective potentials, while the $\pm$ sign in (15) tells us whether the stone moves to larger or smaller r. Note that the effective potentials are not real-valued (do not exist) at values of r that make the horizon function H imaginary; namely between the event and Cauchy horizons.

QUERY 1. Effective potentials at selected r-values

Show the following:

A. The two effective potential functions become equal, $V_L^+(r) = V_L^-(r)$, at both horizons and at $r = 0$.

B. As $r/M \to \infty$, the two effective potentials become, respectively, $V_L^+(r)/m \to +1$ and $V_L^-(r)/m \to -1$.

Objection 1. *Impossible! Item B in Query 1 says that the spinning black hole has an effective potential that extends outward to infinity. No black hole, spinning or non-spinning, can possibly be that powerful.*

There is no problem with $V_L^+(r)$: Item B in Query 1 simply reaffirms that a stone far from the black hole has $V_L^+(r) \to 1$, the special relativity result in flat spacetime. For the case of $V_L^-(r)$ far from the black hole, read on!

QUERY 2. Map angular momentum of a stone when $a \to 0$

Show that when $a \to 0$ then $d\Phi \to d\phi$ and $R \to r$, so the angular momentum equation (110) in Section 17.10 reduces to the expression for the non-spinning black hole (Section 8.2):

$$\frac{L}{m} = r^2 \frac{d\phi}{d\tau} \qquad \text{(non-spinning black hole)} \qquad (17)$$

QUERY 3. Expression for $dr/d\tau$ for the non-spinning black hole

A. Show that when $a \to 0$, equation (15) reduces to equation (19) in Section 8.3 for the non-spinning black hole:

$$\left(\frac{dr}{d\tau}\right)^2 = \left(\frac{E}{m}\right)^2 - \left(\frac{V_L}{m}\right)^2 = \left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{m^2 r^2}\right) \qquad \text{(non-spinning BH)} \quad (18)$$

B. Show that when $a \to 0$, then $V_L^{\pm}(r)$ reduces to the single effective potential for a non-spinning black hole in Section 8.4:

Box 2. At the Horizons

What happens to our constants and variables at the event and Cauchy horizons? Here's a summary. (You can derive these expressions as a Query or exercise.)

In the following, the subscript H stands for the value of that quantity at either the event horizon or the Cauchy horizon.

$$R(r) \to R_{\rm H} = 2M \quad \text{(Fig. 1, Section 17.2.)} \qquad (22)$$

$$H(r) \to H_{\rm H} = 0 \qquad (23)$$

$$\omega \to \omega_{\rm H} = \frac{a}{2Mr_{\rm H}} \qquad (24)$$

$$\frac{r^2+a^2}{2Mr} \to 1 \qquad (25)$$

$$\left(\frac{E-V_{\rm L}^+}{m}\right)^{1/2}\left(\frac{E-V_{\rm L}^-}{m}\right)^{1/2} \to \frac{E-\omega_{\rm H}L}{m} \qquad (26)$$

$$\beta = \left(\frac{2M}{r}\right)^{1/2}\left(\frac{r^2+a^2}{R^2}\right)^{1/2} \to \beta_{\rm H} = 1 \qquad (27)$$

$$\frac{V_{\rm L}(r)}{m} \equiv \left(1-\frac{2M}{r}\right)^{1/2}\left(1+\frac{L^2}{m^2r^2}\right)^{1/2} \qquad \text{(non-spinning black hole)} \qquad (19)$$

Equations of motion $dT/d\tau$ and $d\Phi/d\tau$

Use expressions (15) and (16) for $dr/d\tau$ to complete the equations of motion begun with (3) and (4), and rearrange the results to give the following expressions. These extensive derivations use several expressions in Box 1.

$$\frac{dT}{d\tau} = \left(\frac{R}{rH}\right)^2\left[\frac{E-\omega L}{m} \pm \beta\left(\frac{E-V_{\rm L}^+}{m}\right)^{1/2}\left(\frac{E-V_{\rm L}^-}{m}\right)^{1/2}\right] \qquad (20)$$

$$\frac{d\Phi}{d\tau} = \frac{L}{mR^2} + \frac{\sin^2\alpha}{a}\left[\frac{E-\omega L}{m} \pm \frac{1}{\beta}\left(\frac{E-V_{\rm L}^+}{m}\right)^{1/2}\left(\frac{E-V_{\rm L}^-}{m}\right)^{1/2}\right] \qquad (21)$$

In these equations, the plus sign in front of β or $1/\beta$ corresponds to an increasing r-value and the minus sign to a decreasing r-value.

18.3 ■ USING EFFECTIVE POTENTIALS

Where to go, where to stop, where to bounce, where to stay

Every equation of motion—(15), (20), and (21)—contains the following expression, which must be real if the stone can move, or even exist, with that map energy E:

$$\left(\frac{E-V_{\rm L}^+}{m}\right)^{1/2}\left(\frac{E-V_{\rm L}^-}{m}\right)^{1/2} \qquad \text{must be real.} \qquad (28)$$

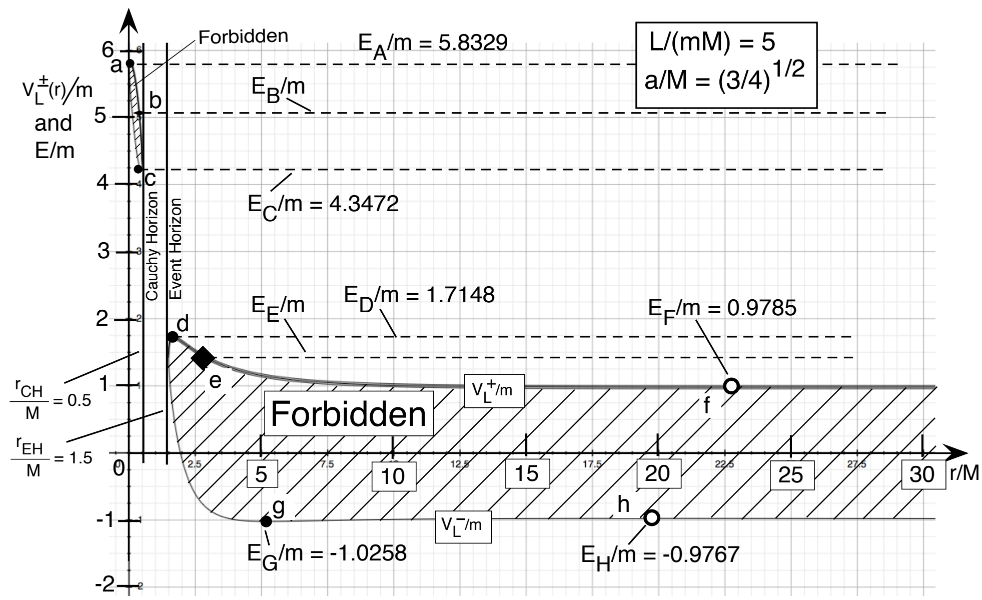


FIGURE 1 Effective potentials $V_L^+(r)$ and $V_L^-(r)$ for a stone with $L/m = 5M$ orbiting a spinning black hole with spin parameter $a/M = (3/4)^{1/2}$. Turning points (Definition 2) lie on the effective potential curves: a little filled circle at the r-value of an unstable circular orbit; a little open circle at the r-value of a stable circular orbit; a rotated little black square at a bounce point. Figure 2 shows a magnified view of effective potentials inside the Cauchy horizon.

Equations of motion must be real.

From (16), $V_L^+(r) > V_L^-(r)$ at every r-value where effective potentials exist. Expression (28) is real at these r-values when either $E > V_L^+(r)$ or $E < V_L^-(r)$. In contrast, expression (28) is imaginary in regions where map energy lies between the effective potentials, that is where $V_L^+(r) > E > V_L^-(r)$. The stone cannot move, or even exist, with map energy E in that region. We say that this is a *forbidden map energy region* (Definition 1).

Figures 1 and 2 plot the two effective potentials from (16) for given values $a/M = (3/4)^{1/2}$ and $L/(mM) = 5$, along with several values of the stone's map energy. These figures illustrate forbidden map energy regions, which we now define.

Definition: **Forbidden energy region**

DEFINITION 1. Forbidden map energy region

A **forbidden map energy region** (which we often call simply a *forbidden region*) is a region between the $V_L^-(r)$ and $V_L^+(r)$ effective potential curves on the $V_L^{\pm}(r)/m$ vs r/M plot. Why forbidden? Because if the map energy E/m of the stone did lie in this region, its equations of motion would be imaginary or complex.

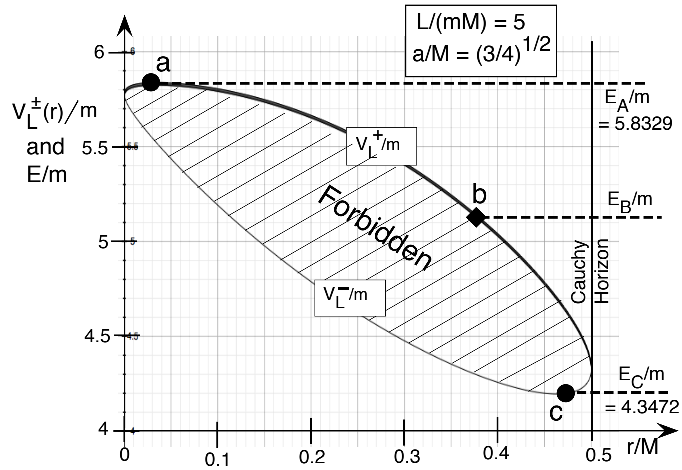


FIGURE 2 Magnified view of the pair of effective potentials in Figure 1 inside the Cauchy horizon. Little filled circles at points a and c show r-values for unstable circular orbits; the rotated filled square symbol locates a bounce point.

Figures 1 and 2 exhibit not only forbidden map energy regions but also what we call *turning points*, which we subdivide into *circle points* and *bounce points*. (Recall similar definitions in Section 8.4 for the non-spinning black hole.)

DEFINITION 2. Turning point, circle point, and bounce point

Definition: **Turning point**

A **turning point** is a point on the $V_L^{\pm}(r)/m$ vs r/M curve for which either $E = V_L^+$ or $E = V_L^-$. At a turning point $dr/d\tau = 0$—equation (15). Examples: points labeled a through h in Figure 1. We distinguish two kinds of turning points: circle point and bounce point.

Definition: **Circle point**

A **circle point** is a turning point at a maximum or minimum of the effective potential. At a circle point $dr/d\tau = 0$ and remains zero, at least temporarily, so a stone at a circle point is in an unstable or stable circular orbit. We plot a circle point as a little filled circle (at an unstable circular orbit) or a little open circle (at a stable circular orbit). See Definition 3. Examples: points labeled a, c, d, f, g, and h in Figure 1.

Definition: **Bounce point**

A **bounce point** is a turning point that is *not* at a maximum or minimum of the effective potential. At a bounce point, $dr/d\tau = 0$ for an instant but then reverses sign. We plot a bounce point as a little filled rotated square (a diamond). Examples: points b, and e in Figure 1 and point b in Figure 2.

Return to the circle point. There are two different kinds of circular orbits: *stable* and *unstable*.

DEFINITION 3. Stable and unstable circular orbits

Definition: **Stable circular orbit**

A stone occupies a **stable circular orbit** when it lies at a circle point in the $V_{\rm L}^{\pm}(r)/m$ *vs.* r/M diagram at which displacement either right or left, while keeping E/m constant puts it inside a forbidden map energy region. We plot a stable circular orbit location as a little open circle. Examples: points f and h in Figure 1.

Definition: **Unstable circular orbit**

The stone occupies an **unstable circular orbit** when it lies at a circle point in the $V_{\rm L}^{\pm}(r)/m$ *vs.* r/M diagram at which displacement either right or left, while keeping E/m constant does *not* put it inside a forbidden map energy region in that diagram. We often call an unstable circular orbit a **knife-edge** orbit to emphasize its instability. We plot an unstable circular orbit location as a little filled circle. Examples: points a, c, d, and g in Figure 1.

Table 18.1 expresses these definitions analytically. Table 18.2 lists details for turning points in Figures 1 and 2.

Objection 2. *Stop! Figure 1 shows circular orbits g and h at negative map energies; negative-energy orbits cannot exist. Everyone knows that energy must be a positive quantity. Circular orbits at points g and h in Figure 1 cannot exist!*

Beware of phrases such as "everyone knows." First, even in Newton's mechanics we can choose the zero of gravitational energy at any height in a gravitational field; then the potential energy of a stationary stone at any lower height becomes negative. Second, in general relativity the map energy is typically not measurable; it's a constant of motion that can be negative without physical consequence. Chapter 19 gives formulas for the energy of a free stone measured in a local inertial frame, which yields a positive frame energy even for a negative map energy.

Objection 3. *Phooey! Your whole analysis is a fantasy! Even Figures 1 and 2 describe structures inside the event horizon that no observer can possibly see or measure. Physical theory has to be "falsifiable:" it must be vulnerable to disproof by observation.*

In principle (or possibly in the future) we can observe and measure these results: Someone who rides a free stone inward across the event horizon can make measurements to verify results of this theory. Let an astronaut initially outside the event horizon have positive map energy above the forbidden map energy region. Chapter 21 describes a set of maneuvers inside the event horizon that brings this astronaut back out through the event horizon with negative map energy. Then she can report on her measurements during her earlier descent. More generally, a scientific theory often predicts what we will observe when new conditions or improved equipment become available.

TABLE 18.1 Classification of Circular Orbits using $V_L^{\pm}$

When $E = V_L^+$ and	When $E = V_L^-$ and
$dV_L^+/dr = 0$, then the orbit is	$dV_L^-/dr = 0$, then the orbit is
STABLE if $d^2V_L^+/dr^2 > 0$, but	**STABLE** if $d^2V_L^-/dr^2 < 0$, but
UNSTABLE if $d^2V_L^+/dr^2 < 0$.	**UNSTABLE** if $d^2V_L^-/dr^2 > 0$.

TABLE 18.2 Map Energies of Circular Orbits with $L/(mM) = 5$ and $a/M = (3/4)^{1/2}$ (Figures 1 and 2). Circle orbit Type numbers from equations (31)–(38).

Circular orbit letter: r/M-value	Type: E/m-value, unstable or stable
Point a: $r/M = 0.0341$	Type 1: $E_A/m = 5.8329$, unstable
Point c: $r/M = 0.4660$	Type 2: $E_C/m = 4.3472$, unstable
Point d: $r/M = 1.6963$	Type 1: $E_D/m = 1.7148$, unstable
Point f: $r/M = 22.744$	Type 1: $E_F/m = 0.9785$, STABLE
Point g: $r/M = 5.2469$	Type 3: $E_G/m = -1.0258$, unstable
Point h: $r/M = 19.7855$	Type 3: $E_H/m = -0.9767$, STABLE

QUERY 4. Application of Table 18.1

Which entries in Table 18.1 apply to circular orbits around the *non-spinning* black hole?

Comment 2. Two non-communicating regions

What goes on below the forbidden map energy region in Figure 1? This figure implies, and equations show, that this forbidden map energy region extends as far as $r \to \infty$. Apparently both stable and unstable circular orbits exist below the forbidden map energy region. We have verified that no stone can exist in the forbidden map energy region, and Chapter 20 demonstrates that light is similarly forbidden to travel directly between an upper and lower region. *Result:* two regions that cannot communicate directly with one another.

Map energy is negative below the forbidden map energy region, but that need not worry us: nobody observes or measures map energy. You can show that almost every (but not every) local inertial frame (defined in Chapter 17) that exists above the forbidden region can exist below the forbidden region. Indeed, for almost every (but not every) event that occurs at T, r, Φ above the forbidden map energy region an event can occur at T, r, Φ below this region.

Where are events that occur below the forbidden map energy region? Is there an entire separate Universe there, a Universe we cannot see from ours? Can we get to that Universe? Can we come back? Answers in Chapter 21!

18.4 ■ FOUR TYPES OF CIRCULAR ORBITS

How many circular orbits, and of what types?

Multiple circular orbits at the same r

The spinning black hole has (many!) more surprises for us. One of these is the existence of *multiple distinct circular orbits at the same r-value.* Figure 3 shows

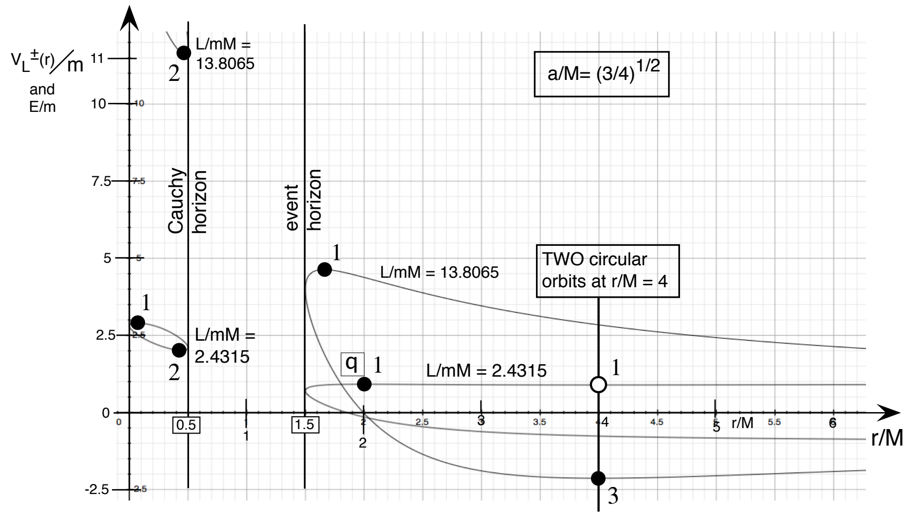


FIGURE 3 Two different effective potentials for a spinning black hole, each of which leads to a circular orbit at $r/M = 4$, one stable and the other unstable. Numbers 1 through 3 indicate circular orbit Types from equations (31) through (38). Figure 4 shows the possibility of *four* circular orbits at $r/M = 4$. The label q refers to the same orbit in Figure 13. In order to display all turning points clearly, we do not shade forbidden map energy regions in this plot.

two different effective potentials for a spinning black hole with $a/M = (3/4)^{1/2}$ that lead to two different circular orbits at $r/M = 4$. Note that these occur for two different (positive) values of the map angular momentum $L/(mM)$. Even more astonishing, Figure 4 shows a total of four circular orbits at $r/M = 4$, two for the pair of positive values of $L/(mM)$ in Figure 3 plus two more for the corresponding *negative* values of these map angular momenta.

QUERY 5. Number of circular orbits at given r: Newton and the non-spinning black hole

Both the non-spinning black hole and the spherically symmetric center of attraction of Newton's mechanics are spherically symmetric, which allows an unlimited number of differently oriented $[r, \Phi]$ slices through the centers of these objects on which circular orbits can exist. On a single one of these slices,

A. Newton: For what values of r do circular orbits exist?

B. Newton: How many distinct circular orbits exist at that r?

C. Newton: If your answer to Item B predicts more than one circular orbit, what determines the difference between circular orbits at that r-value?

D. Repeat Items A through C for the non-spinning black hole.

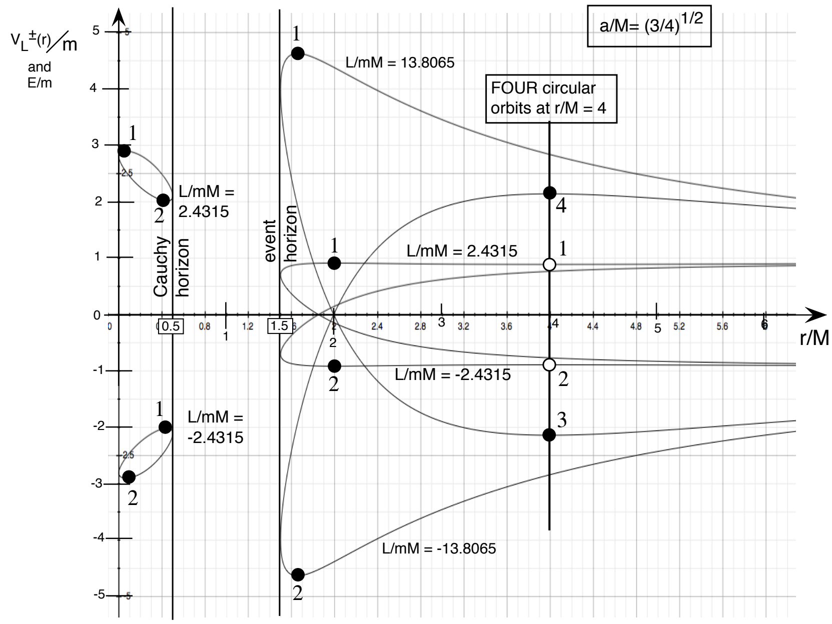


FIGURE 4 Four different effective potentials for a spinning black hole with $a/M = (3/4)^{1/2}$, all of which have circular orbits at $r/M = 4$, two stable and two unstable. This figure adds to Figure 3 effective potential curves for negative values of the stone's angular momentum. Effective potentials for $L/(mM) = \pm 13.8065$ inside the Cauchy horizon lie beyond the vertical range of this plot. The number on each circular orbit symbol gives its Type. We do not shade forbidden map energy regions, in order to display all turning points clearly.

No circular orbits between event horizon and Cauchy horizon

We want to derive general expressions for map energies and map angular momenta of circular orbits around a spinning black hole. Definition 2 tells us that a circular orbit occurs at r-values for which either $E = V_L^+(r)$ and $dV_L^+(r)/dr = 0$ or $E = V_L^-(r)$ and $dV_L^-(r)/dr = 0$. Between the event horizon and the Cauchy horizon the third equation of motion (15) is imaginary, so carries no physical meaning there. In addition, circular orbits near the horizons lie separated in r-value from the horizons, illustrated in Figures 1 and 2 (Query 6). Now we turn these qualitative observations into analytical and numerical results.

QUERY 6. Circular orbits avoid horizons and the singularity.
In this Query you show that the circular orbits do not exist at the singularity or at the two horizons.

A. Show that the slope of each effective potential function increases without limit ($dV_L^\pm/dr \to \infty$) at both horizons and at $r = 0$.

B. The slope of the effective potential is zero at the r-value of every circular orbit. Item A tells us that this slope is vertical at three r-values: both horizons and the singularity. The effective potentials are continuous at these three r-values and at the nearest circle points. Circular orbits are impossible at each horizon and at the singularity.

Generating equation for circular orbits

To find all r-values of circular orbits, set the derivatives of the two functions $V_{\mathrm{L}}^{+}(r)$ and $V_{\mathrm{L}}^{-}(r)$ equal to zero and from them derive an equation that contains all terms containing L. *Result:* an expression for the value of L for a circular orbit (if any) at that r-value. Equations (29) and (30) are the generating equations for circular orbits.

$$\pm AL = B\left(L^2 + m^2R^2\right)^{1/2} + \frac{C}{\left(L^2 + m^2R^2\right)^{1/2}} \qquad (29)$$

where the $\pm$ symbol matches that in the superscript of $V_{\mathrm{L}}^{\pm}$, and symbols A, B, C stand for the following functions of a and r:

$$A \equiv -\frac{d\omega}{dr}, \qquad B \equiv \frac{d}{dr}\left(\frac{rH}{R^2}\right), \qquad C \equiv m^2\left(r - \frac{Ma^2}{r^2}\right)\frac{rH}{R^2} \qquad (30)$$

QUERY 7. *Optional:* Derive the generating equation for circular orbits.
Carry out the derivation of equations (29) and (30).

QUERY 8. Pairs of solutions

A. Show that when $L = +L_1$ is a solution of (29) with $E = V_{\mathrm{L}}^{+}(r)$, then $L = -L_1$ is also a solution at the same r with $E = V_{\mathrm{L}}^{-}(r)$. Conclusion: *Circular orbits come in pairs.*
B. Identify all such pairs in Figure 4.
C. Show also that the orbits in a pair are either both stable or both unstable. Hint: Use a symmetry argument.

Four circular orbit types

Solve equation (29) for L/m as a function of r. Lots of algebra yields two solutions for L/m. For each of these solutions set $E = V_{\mathrm{L}}^{+}$ or $E = V_{\mathrm{L}}^{-}$ at this value of r. Result: four types of circular orbits described by the following equations. (Section 18.5 defines the labels on the right sides of these equations.)

<u>TYPE 1</u> for $E = V_{\mathrm{L}}^{+}(r)$

$$\left(\frac{L}{m}\right)_{\mathrm{Type\,1}} = \left(\frac{M}{r}\right)^{1/2} \frac{r^2 + a^2 - 2a(Mr)^{1/2}}{\left[r^2 - 3Mr + 2a(Mr)^{1/2}\right]^{1/2}} \quad \text{(forward, prograde)} \tag{31}$$

$$\left(\frac{E}{m}\right)_{\mathrm{Type\,1}} = \frac{V_{\mathrm{L}}^{+}(r)}{m} = \frac{r^2 - 2Mr + a(Mr)^{1/2}}{r\left[r^2 - 3Mr + 2a(Mr)^{1/2}\right]^{1/2}} \quad \text{(forward, prograde)} \tag{32}$$

<u>TYPE 2</u> for $E = V_{\mathrm{L}}^{-}(r)$

$$\left(\frac{L}{m}\right)_{\mathrm{Type\,2}} = -\left(\frac{L}{m}\right)_{\mathrm{Type\,1}} \quad \text{(backward, prograde)} \tag{33}$$

$$\left(\frac{E}{m}\right)_{\mathrm{Type\,2}} = \frac{V_{\mathrm{L}}^{-}(r)}{m} = -\left(\frac{E}{m}\right)_{\mathrm{Type\,1}} \quad \text{(backward, prograde)} \tag{34}$$

<u>TYPE 3</u> for $E = V_{\mathrm{L}}^{-}(r)$

$$\left(\frac{L}{m}\right)_{\mathrm{Type\,3}} = \left(\frac{M}{r}\right)^{1/2} \frac{r^2 + a^2 + 2a(Mr)^{1/2}}{\left[r^2 - 3Mr - 2a(Mr)^{1/2}\right]^{1/2}} \quad \text{(backward, retrograde)} \tag{35}$$

$$\left(\frac{E}{m}\right)_{\mathrm{Type\,3}} = \frac{V_{\mathrm{L}}^{-}(r)}{m} = -\frac{r^2 - 2Mr - a(Mr)^{1/2}}{r\left[r^2 - 3Mr - 2a(Mr)^{1/2}\right]^{1/2}} \quad \text{(backward, retrograde)} \tag{36}$$

<u>TYPE 4</u> for $E = V_{\mathrm{L}}^{+}(r)$

$$\left(\frac{L}{m}\right)_{\mathrm{Type\,4}} = -\left(\frac{L}{m}\right)_{\mathrm{Type\,3}} \quad \text{(forward, retrograde)} \tag{37}$$

$$\left(\frac{E}{m}\right)_{\mathrm{Type\,4}} = \frac{V_{\mathrm{L}}^{+}(r)}{m} = -\left(\frac{E}{m}\right)_{\mathrm{Type\,3}} \quad \text{(forward, retrograde)} \tag{38}$$

QUERY 9. Pairs of map energies and map angular momenta
Show that Figure 4 illustrates the results of Query 8. As a result, show that Type 1 implies the existence of Type 2 and also that Type 3 implies the existence of Type 4.

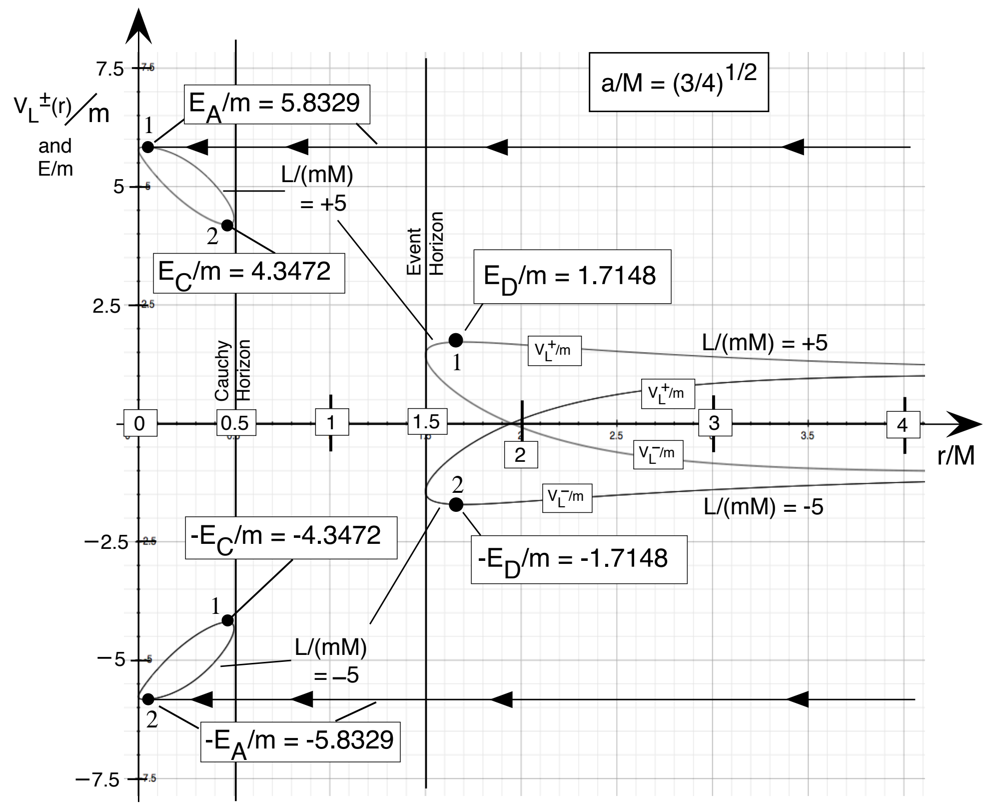


FIGURE 5 Extension of Figure 1 to positive and negative values of angular momentum: $L/(mM) = \pm 5$ to show the relation between Types 1 and 2 circular orbits. Reverse the sign of L to reverse the sign of E at the same r-value (Query 8). A stone of map energy E_{A} and $L/(mM) = +5$ (horizontal line at the top of the plot) goes into a Type 1 circular orbit, which is distinct from the Type 2 circular orbit with $E = -E_{\mathrm{A}}$ at the same r (bottom of the plot). Similarly for other circular orbits at the same r-values but of different types.

QUERY 10. Other pairs of solutions

A. Show that when we change ω to $-\omega$ in (16), then $V_{\mathrm{L}}^{+}(r)$ becomes $-V_{\mathrm{L}}^{-}(r)$ and $V_{\mathrm{L}}^{-}(r)$ becomes $-V_{\mathrm{L}}^{+}(r)$.

B. From Item A and equation (11), show that when we change a to $-a$ in (31) and (32)—that is, when the black hole spins in the opposite sense—then a circular orbit of Type 1 becomes a circular orbit of Type 3 at the same r-value.

C. Likewise, show that when we change a to $-a$ in (33) and (34), then a Type 2 circular orbit becomes a Type 4 circular orbit at the same r-value.

Comment 3. Convenient to define four types of circular orbits
Queries 8 through 10 show that reversing the sign of the orbital angular momentum of a stone and/or the spin parameter of the black hole yield new circular orbits. Result: We can derive from Type 1 the other three types of circular orbits for a given absolute value of the black hole spin parameter $|a/M|$. It is informative, however, to consider each of the four types separately.

How many circular orbits at a given r?

How many circular orbits exist at r for the spinning black hole with a given value of a/M? To answer this question, look at equations (31) through (38). Map energy and map angular momentum of the stone must be real, so orbits exist only at r-values where functions inside the square roots in the denominators of these equations are positive:

$$r^2 - 3Mr + 2a(Mr)^{1/2} > 0 \quad \text{(where orbits exist for Types 1 and 2)} \quad (39)$$

$$r^2 - 3Mr - 2a(Mr)^{1/2} > 0 \quad \text{(where orbits exist for Types 3 and 4)} \quad (40)$$

How many circular orbits at various values of r?

From these inequalities we can sort out the r-locations at which different circular orbit types exist. As $r \to \infty$, both inequalities (39) and (40) are satisfied, so all four types of circular orbits exist far from the black hole. At some intermediate values of r (but outside the event horizon) inequality (39) is satisfied, but inequality (40) is not satisfied, so only prograde orbits exist at those r-values. Only prograde orbits exist inside the Cauchy horizon, as in Figure 2 (Table 1). Finally, a region exists in which even $r^2 - 3Mr + 2a(Mr)^{1/2} < 0$, so no circular orbits can exist in that region. Each of these conditions depends on the value of the black hole's spin parameter a/M. Figure 6 plots these results for different values of a/M.

Comment 4. Orbits of light
The r-values where equations (39) and (40) become equalities are places where the denominators vanish in equations (31) through (38). Multiply both sides of each of these equations through by m, the mass of the orbiting stone. Then circular orbits can exist with the corresponding values of E and L if, and only if, $m \to 0$. Therefore, these are r-values for circular orbits of *light* (Figure 6). Chapter 20 explores orbits of light in greater generality.

Which of these circular orbits are stable? Figures such as 2 and 4 preview the result that *all* circular orbits inside the Cauchy horizon are unstable. Sections 18.6 through 18.8 pursue the stability question after we investigate further the differences among Types 1 through 4 circular orbits outside the event horizon. In this process we will finally define *prograde vs. retrograde* circular orbits and *forward vs. backward* circular orbits.

18.5 ■ MAP $dT/d\tau$ AND MAP $d\Phi/dT$ FOR CIRCULAR ORBITS

Add Doran Φ and T to the specification of circular orbits.

$dT/d\tau$ and $d\Phi/d\tau$ for circular orbits

Look again at equations of motion (20) and (21). The final term on the right side of each of these equations equals zero for the special case of a circular orbit, for which either $E = V_{\rm L}^{+}$ or $E = V_{\rm L}^{-}$:

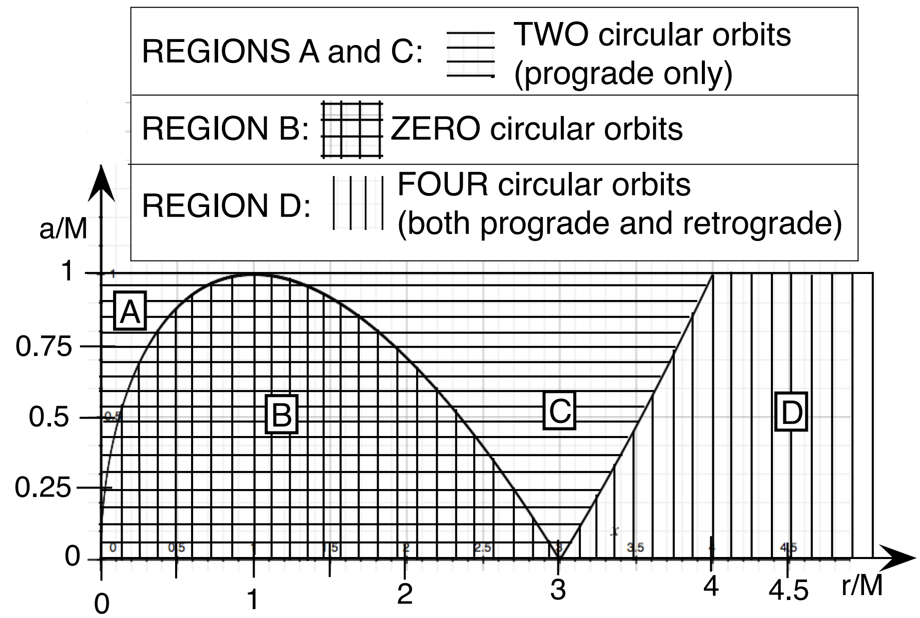


FIGURE 6 This figure uses inequalities (39) and (40) to answer the question, "How many circular orbits of a stone exist at a given r for different values of the spin parameter a/M?" In Region B, zero circular orbits exist. In Regions A and C, only Type 1 and Type 2 (prograde) circular orbits exist. In Region D, all four types of circular orbits exist. Circular orbits along the curves that divide regions are photon orbits (Comment 4).

$$\frac{dT}{d\tau} = \left(\frac{R}{rH}\right)^2 \left(\frac{E - \omega L}{m}\right) \qquad \text{(circular orbit)} \qquad (41)$$

$$\frac{d\Phi}{d\tau} = \frac{L}{mR^2} + \frac{\sin^2\alpha}{a}\left(\frac{E - \omega L}{m}\right) \qquad \text{(circular orbit)} \qquad (42)$$

Four types of $dT/d\tau$ and $d\Phi/d\tau$

Now plug values of L/m and E/m from (31) through (38) into equations (41) and (42). This leads to expressions for $dT/d\tau$ and $d\Phi/d\tau$ for the four types of circular orbits in in Section 18.4:

TYPE 1 for $E = V_{\mathrm{L}}^{+}$

$$\left(\frac{dT}{d\tau}\right)_{\mathrm{Type\,1}} = \frac{r + a\,(M/r)^{1/2}}{\left[r^2 - 3Mr + 2a(Mr)^{1/2}\right]^{1/2}} \qquad \text{(forward, prograde)} \qquad (43)$$

$$\left(\frac{d\Phi}{dT}\right)_{\mathrm{Type\,1}} = \frac{(Mr)^{1/2}}{r^2 + a\,(Mr)^{1/2}} \qquad \text{(forward, prograde)} \qquad (44)$$

TYPE 2 for $E = V_{\mathrm{L}}^{-}$

$$\left(\frac{dT}{d\tau}\right)_{\mathrm{Type\,2}} = -\left(\frac{dT}{d\tau}\right)_{\mathrm{Type\,1}} \qquad \text{(backward, prograde)} \qquad (45)$$

$$\left(\frac{d\Phi}{dT}\right)_{\mathrm{Type\,2}} = +\left(\frac{d\Phi}{dT}\right)_{\mathrm{Type\,1}} \qquad \text{(backward, prograde)} \qquad (46)$$

TYPE 3 for $E = V_{\mathrm{L}}^{-}$

$$\left(\frac{dT}{d\tau}\right)_{\mathrm{Type\,3}} = \frac{-r + a(M/r)^{1/2}}{\left[r^2 - 3Mr - 2a(Mr)^{1/2}\right]^{1/2}} \qquad \text{(backward, retrograde)} \qquad (47)$$

$$\left(\frac{d\Phi}{dT}\right)_{\mathrm{Type\,3}} = \frac{(Mr)^{1/2}}{-r^2 + a\,(Mr)^{1/2}} \qquad \text{(backward, retrograde)} \qquad (48)$$

TYPE 4 for $E = V_{\mathrm{L}}^{+}$

$$\left(\frac{dT}{d\tau}\right)_{\mathrm{Type\,4}} = -\left(\frac{dT}{d\tau}\right)_{\mathrm{Type\,3}} \qquad \text{(forward, retrograde)} \qquad (49)$$

$$\left(\frac{d\Phi}{dT}\right)_{\mathrm{Type\,4}} = +\left(\frac{d\Phi}{dT}\right)_{\mathrm{Type\,3}} \qquad \text{(forward, retrograde)} \qquad (50)$$

Note: Equations for $d\Phi/dT$, with dT in the denominator, are not typographical errors: We choose to solve for $d\Phi/dT$, not for $d\Phi/d\tau$, for two reasons: *Minor reason*: Equations for $d\Phi/dT$ are simpler than equations for $d\Phi/d\tau$. *Major reason*: This choice simplifies the categories. Type 1 and 2 circular orbits (labeled prograde) always have $d\Phi/dT > 0$, while Type 3 and 4 circular orbits (labeled retrograde) always have $d\Phi/dT < 0$.

QUERY 11. Plus or minus? Signs of important expressions

A. From the requirement that $r^2 - 3Mr - 2a(Mr)^{1/2} > 0$ for Types 3 and 4 circular orbits, show that $-r^2 + a(Mr)^{1/2} < 0$.

B. As a result, show that for Types 3 and 4 circular orbits, we have $d\Phi/dT < 0$ for all values of r.

C. Show that $(dT/d\tau)_3 < 0$ for all values of r, so $(dT/d\tau)_4 > 0$ for all values of r.

This analysis leads to definitions of *prograde* and *retrograde* orbits.

Prograde and **retrograde** orbits

DEFINITION 4. Prograde and retrograde orbits

We divide circular orbits into two classes, **prograde** and **retrograde**. In a prograde orbit the stone "revolves in the direction that the black hole rotates" in global Doran coordinates so that $d\Phi/dT > 0$, while in a retrograde orbit the stone revolves in the opposite direction, $d\Phi/dT < 0$. Note that the condition $d\Phi/dT = 0$ for the raindrop worldline (Section 17.7) marks the separation between prograde and retrograde orbits. As shown in Figure 6, retrograde orbits exist only outside the event horizon, while prograde orbits exist inside the Cauchy horizon as well as outside the event horizon.

?

Objection 4. *Your definitions of prograde and retrograde orbits are nothing but manipulations of Doran map coordinates Φ and T. You keep saying that we cannot observe map coordinates directly. Worse: Except for wristwatch time τ, this chapter uses **only** map coordinates. Your messy results tell us nothing about what we can **see** and **measure** as we move near a spinning black hole. Stop wasting our time!*

!

Nice objection! We use global constants of motion to discover possible motions of a stone. For example, we now know how many circular orbits—zero, two, or four—can exist at each r-value around a black hole with given spin parameter a. This significant achievement says nothing whatsoever about what you will *see* as you ride an unpowered rocket ship in any circular orbit. Such predictions require analysis of orbits of light near the spinning black hole. Hang on: Visual results arrive in future chapters!

Forward or backward orbits from sign of $dT/d\tau$

The other pair of labels attached to circular orbits, *forward* or *backward*, derive from the sign of $dT/d\tau$. We have chosen the stone's wristwatch time τ to increase—to make $d\tau$ positive—as the stone proceeds along its worldline (Comment 7, Section 1.11). So the sign of dT determines the sign of $dT/d\tau$. If $dT/d\tau > 0$, then T also runs forward along the worldline of that stone. In contrast if $dT/d\tau < 0$ then T runs backward along that worldline. This leads to definitions of forward and backward orbits.

Definition: **forward** and **backward** orbits

DEFINITION 5. Forward and backward orbits

Along a **forward orbit**, $dT/d\tau > 0$, so both T and τ increase as the stone proceeds along its worldline. Along a **backward orbit**, $dT/d\tau < 0$, so τ increases and T decreases as the stone proceeds along its worldline.

The concept of a global rain T (or global Schwarzschild t for a non-spinning black hole) that runs backward along a stone's worldline is

TABLE 18.3 Signs of circular orbit quantities

Type	$E =$	E/m	L/m	$dT/d\tau$	$d\Phi/dT$	$d\Phi/d\tau$
1	V_{L}^{+}	$\pm$	$\pm$	$+$	$+$	$+$
2	V_{L}^{-}	$\mp$	$\mp$	$-$	$+$	$-$
3	V_{L}^{-}	$-$	$+$	$-$	$-$	$+$
4	V_{L}^{+}	$+$	$-$	$+$	$-$	$-$

Types 1 and 2 for L/m and E/m: Upper sign for orbits outside the event horizon, either sign for orbits inside the Cauchy horizon. Type 3 and 4 orbits exist only outside the event horizon.

nothing new. Figure 8 in Section 3.7 displayed the worldline of Stone B inside the event horizon along whose worldline Schwarzschild global t runs backward. No contradiction results; nobody measures these global coordinate differences.

Global T can run either forward or backward along a worldline.

For the spinning black hole there are two new results: *First new result:* The orbits that run forward and backward in T come in pairs: if one exists, the other exists at the same r, with opposite signs of E/m and L/m—equations (32) through (37). *Second new result:* For a spinning black hole, global T can run backward along a stone's worldline even outside the event horizon, indeed, all the way out: $r/M \to \infty$.

Comment 5. ALWAYS forward? ALWAYS backward?
Are orbits with $E = V_{\mathrm{L}}^{+}(r)$ *always* forward? Are orbits with $E = V_{\mathrm{L}}^{-}(r)$ *always* backward? Yes to both questions—at least for circular orbits. These results follow from (31) through (38) and (43) through (50). Can you fill in the argument?

QUERY 12. Orbit pairs

A. Show that the signs of $dT/d\tau$ and $d\Phi/dT$ in Table 18.3 agree with Definitions 4 and 5.
B. Show that for each pair of circular orbits in Query 8, one orbit is forward, the other is backward.
C. Show that for eqch pair of circular orbits in Query 10, one orbit is prograde, the other is retrograde.

QUERY 13. More signs of important expressions

A. Use Table 18.3 and the signs of L/m and E/m to verify the assignment of Types to the 6 circular orbits listed in Table 18.2.
B. Verify the signs of L/m and E/m in Table 18.3. Hint: To show that both signs are possible for Types 1 and 2, examine Point c in Table 18.2.
C. Verify the signs of $dT/d\tau$ and $d\Phi/d\tau$ in Table 3 using equations (43) to (50) and Query 11.

QUERY 14. Elapsed ΔT and $\Delta\tau$ for one circular orbit

A. Define one complete circular orbit to have $\Delta\Phi = 2\pi$. Use equations (44), (46), (48), and (50) to find the following expression for ΔT, the advance of Doran global T-coordinate, during one circular orbit:

$$\Delta T(\text{one orbit}) = \pm 2\pi \left[\frac{\pm r^2 + a(Mr)^{1/2}}{(Mr)^{1/2}}\right] = \pm 2\pi M\left[\pm\left(\frac{r}{M}\right)^{3/2} + \frac{a}{M}\right] \qquad (51)$$

The $\pm$ sign outside the square brackets comes from $\pm\Delta T$ for forward and backward orbits and the $\pm$ sign inside the square bracket for prograde and retrograde orbits.

B. Next, define one complete circular orbit to have $\Delta\Phi = +2\pi$ if $d\Phi/d\tau > 0$ but $\Delta\Phi = -2\pi$ if $d\Phi/d\tau < 0$. Then use all eight equations (43) through (50) to find the elapsed wristwatch time $\Delta\tau$:

$$\Delta\tau(\text{one orbit}) = 2\pi r\left[\frac{r^2 - 3Mr \pm 2a(Mr)^{1/2}}{Mr}\right]^{1/2} = 2\pi r\left[\frac{r}{M} - 3 \pm \frac{2a}{(Mr)^{1/2}}\right]^{1/2} \qquad (52)$$

with the plus for prograde and the minus for retrograde orbits.

C. Show that for the non-spinning black hole, equation (52) reduces to equation (37) in Section 8.5. What happens to the $\pm$ sign in (52) in this reduction?

D. Answer one of the initial questions on the first page of this chapter: *Does black hole spin make orbits go faster? slower?* Pay special attention to the meaning(s?) of the word "go" in that question.

In the following three sections we examine which circular orbits are stable and which are unstable: Section 18.6 for Newton's circular orbits; Section 18.7 for circular orbits around the non-spinning black hole; Section 18.8 for circular orbits around the spinning black hole.

Why do we care about *stable* circular orbits? Why are they important? Stable circular orbits are important to us for two primary reasons:

WHY ARE STABLE CIRCULAR ORBITS IMPORTANT?

1. A stone perched at the peak of the effective potential does not stay there long, so you do not observe unstable circular orbits in Nature. In contrast, the accretion disk around the spinning black hole (Section 18.9) consists of a series of nested stable circular orbits which a stone occupies in sequence as it radiates away its loss of orbital energy.
2. When we carry out an exploration program of the spinning black hole (Chapter 19), we can temporarily perch our unpowered spaceship in an unstable circular orbit on our way to somewhere else. "Somewhere else" is often a stable circular orbit, from which we can make relaxed observations without worry about falling off the effective potential maximum.

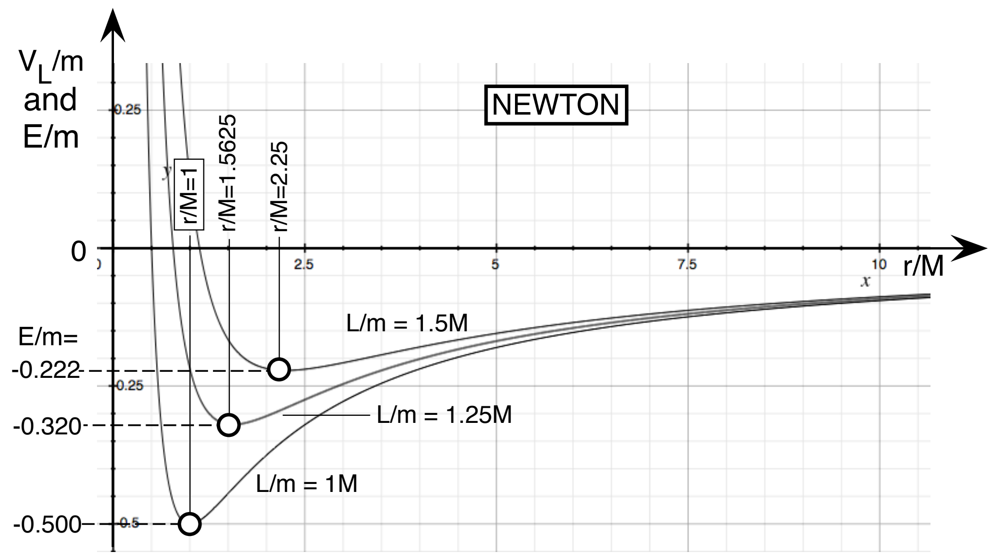


FIGURE 7 Examples of effective potentials and the radii of Newton's stable circular orbits around a point mass. A stable orbit (little open circle) exists at the minimum of each effective potential curve. The area under each effective potential is a forbidden map energy region for the stone with that angular momentum.

18.6 ■ STABILITY OF NEWTON'S CIRCULAR ORBITS

Angular momentum makes the world go 'round.

Newton: Total energy

Begin the analysis of Newton's circular orbits with his expression for the total energy (kinetic plus potential) of a stone in a central gravitational field:

$$E = \frac{1}{2}mv^2 - \frac{mM}{r} \qquad \text{(Newton, conservation of energy)} \qquad (53)$$

Newton's force law $F = ma$ demands that in a circular orbit the inward gravitational force $-mM/r^2$ equals mass m times the inward acceleration $-v^2/r$:

$$-\frac{mM}{r^2} = -\frac{mv^2}{r} \quad \text{so} \quad r^2v^2 = Mr \quad \text{(Newton force law, circular orbit)} \qquad (54)$$

Newton: Angular momentum of a circular orbit

Newton defines the angular momentum of a stone in a circular orbit as its radius r times its tangential linear momentum mv:

$$L \equiv mrv \qquad \text{(Newton, circular orbit)} \qquad (55)$$

so that from (54):

$$L = m(Mr)^{1/2} \qquad \text{(Newton, circular orbit)} \qquad (56)$$

Figure 7 suggests that total orbital energy decreases with decreasing radius of the stable circular orbit. To check this, find an expression for v^2 from (55) and substitute the result into (53), thereby defining the effective potential:

$$V_{\rm L}(r) \equiv \frac{L^2}{2mr^2} - \frac{mM}{r} \qquad \text{(Newton, effective potential)} \qquad (57)$$

Newton: Total energy of a circular orbit

Now substitute for L from (56) and rearrange the result to yield the energy of a stone in a circular orbit as a function of the radius of that orbit:

$$E = V_{\rm L}(r) = \frac{1}{2}\frac{mM}{r} - \frac{mM}{r} = -\frac{1}{2}\frac{mM}{r} \qquad \text{(Newton, circular orbit energy)} \qquad (58)$$

Newton's conclusion: Every circular orbit is stable, all the way down to $r = 0$.

Figure 7 and our accompanying algebraic analysis tell us that Newton's effective potential has only one zero-slope point, and that one point is at a minimum. Definition 3 then tells us that *in Newton's mechanics EVERY circular orbit is stable.* More: Newton's circular orbits are stable all the way down to $r = 0$, or until the stone strikes the surface of a spherically symmetric center of attraction.

Add a little friction.

Now suppose that a stone in a circular orbit encounters a little friction—perhaps from dust or a rarified atmosphere. This friction converts some orbital energy into heat, electromagnetic radiation, or other forms of energy. Where does this converted energy come from? For Newton the only source is the orbital energy of the stone. We analyze the result with a simple model: Assume that this loss of energy per orbit is minuscule, so the stone's orbit remains circular, but its radius changes slightly. How can we track changes in energy, angular momentum, and radius of the orbit during this process? Begin to answer these questions by differentiating both sides of (58):

$$\frac{dE}{dr} = +\frac{1}{2}\frac{mM}{r^2} \qquad \text{(sequence of Newton's circular orbits)} \qquad (59)$$

Similarly, differentiate both sides of (56):

$$\frac{dL}{dr} = +\frac{m}{2}\left(\frac{M}{r}\right)^{1/2} \qquad \text{(sequence of Newton's circular orbits)} \qquad (60)$$

Circular orbit E, L, and r all decrease.

Figure 7 shows what equations (59) and (60) tell us, namely that when the energy of the circular orbit decreases, the angular momentum also decreases, as does the radius of the orbit.

Equations (59) and (60) imply that energy and angular momentum can *change.* How can this be?

External force: friction

The stone's energy and angular momentum are constant for free-fall motion, but they change if an external force is applied to the stone, whether this force arises from a rocket or from friction in an accretion disk. For a circular orbit, r, E, and L are all related. As E and L change, the radius of the circular orbit changes. To see how, think of an incremental change ΔE in energy. Equation (59) then implies that r changes by the amount

$$\Delta r \approx \left(\frac{dE}{dr}\right)^{-1} \Delta E \qquad \text{(Newton AND Einstein circular orbits)} \qquad (61)$$

We can adapt (60) to express the same change in radius between stable orbits of different angular momentum:

$$\Delta r \approx \left(\frac{dL}{dr}\right)^{-1} \Delta L \qquad \text{(Newton AND Einstein circular orbits)} \qquad (62)$$

To summarize: For Newton's circular orbits, a small amount of friction decreases the energy E and the angular momentum L of the orbiting stone and causes it to move to smaller radii through a sequence of *stable* circular orbits. Why stable? Because *all* Newton's circular orbits are stable; every circular orbit nests at a minimum of an effective potential (Figure 7).

?

Objection 5. *Whoa! In this section you use the terms "radius," "energy," and "angular momentum" without modifiers. But you keep saying that these terms have no measurable meaning. Instead, you force us to use modifiers such as "map energy," "map angular momentum," "shell frame energy," and so forth. Why did you use single-word terms that you label forbidden? Follow your own rules!*

!

These distinctions—important in general relativity—do not exist in Newton's mechanics. When carefully used, everyday terms are perfectly accurate for Newton. So we have just enjoyed a short vacation from our terminology rules for general relativity. Sorry, our little vacation is now over!

Comment 6. Wide application of Definitions 1 and 3

Our analysis of Newton's circular orbits uses Definition 1 (forbidden region) and Definition 3 (stable and unstable orbits). The energy region under each of Newton's effective potential curves in Figure 7 is forbidden to the stone (Definition 1), because in that region the stone's kinetic energy would be negative. The stable circular orbit (Definitition 3) nestles at the minimum of the effective potential. These same definitions have wide usefulness: They apply to circular orbits around the non-spinning black hole (Section 18.7) and around the spinning black hole (Section 18.8).

QUERY 15. Time for one orbit according to Newton

A. From Newton's equation for orbit speed in (54) and the circumference of a circle $= 2\pi r$ in flat spacetime, show that for Newton the elapsed time for one circular orbit is:

$$\frac{\Delta t}{M}(\text{one orbit}) = \frac{\Delta \tau}{M}(\text{one orbit}) = 2\pi \left(\frac{r}{M}\right)^{3/2} \qquad \text{(Newton)} \qquad (63)$$

B. Show that equations (51) and (52) both reduce to Newton's result (63) when $r/M \to \infty$.

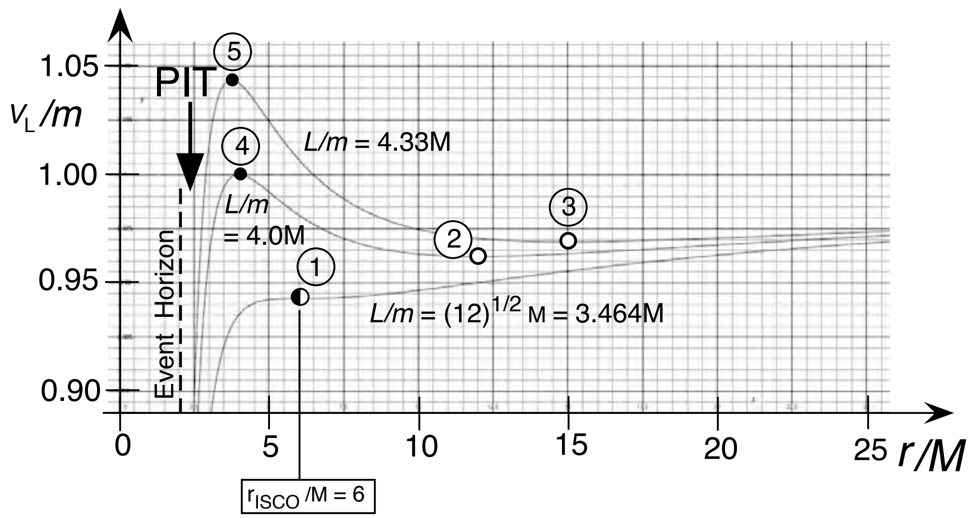


FIGURE 8 Effective potentials for the non-spinning black hole (repeat of Figure 4 in Section 8.4). The area under each curve is the forbidden map energy region for a stone with that value of map angular momentum. Little filled circles locate unstable circular orbits, little open circles locate stable circular orbits, and the little half-filled circle locates a "half-stable" $r_{\rm ISCO}$ circular orbit, one that is "stable to the right and unstable to the left." A small amount of friction moves stable orbits downward and to the left along the sequence of circled numbers $3 \rightarrow 2 \rightarrow 1$ until $r = r_{\rm ISCO}$, after which the stone spirals inward across the event horizon.

18.7 ■ STABILITY OF CIRCULAR ORBITS: NON-SPINNING BLACK HOLE

Add unstable circular orbits to stable circular orbits.

Non-spinning black hole: Stable circular orbits exist for $r > 6M$.

Next analyze the stability of circular orbits around the non-spinning black hole. Figure 8 replots the effective potential for several values of L from Figure 4 in Section 8.4. In Newton's case, Figure 7, all curves have one minimum, the location of a stable circular orbit. But for the spinning black hole, Figure 8, the effective potential to the left of each minimum is radically different. In particular, Figure 8 exhibits the famous PIT in the potential of the non-spinning black hole. Unstable orbits exist at maxima of the effective potential between this pit and the stable-orbit r-values, provided that $L/(mM) > (12)^{1/2}$. Points 4 and 5 are examples of this maximum. Unstable circular orbits are the new contribution of the non-spinning black hole.

To analyze circular orbits for the non-spinning black hole, let $a/M \rightarrow 0$ in equations (31) for E/m and (32) for L/m. Results:

$$\frac{E}{m} = \frac{r^2 - 2Mr}{r\left(r^2 - 3Mr\right)^{1/2}} \qquad \text{(circular orbits, non-spinning black hole)} \qquad (64)$$

$$\frac{L}{m} = \left(\frac{M}{r}\right)^{1/2} \frac{r^2}{\left(r^2 - 3Mr\right)^{1/2}} \ \text{(circular orbits, non-spinning black hole)} \qquad (65)$$

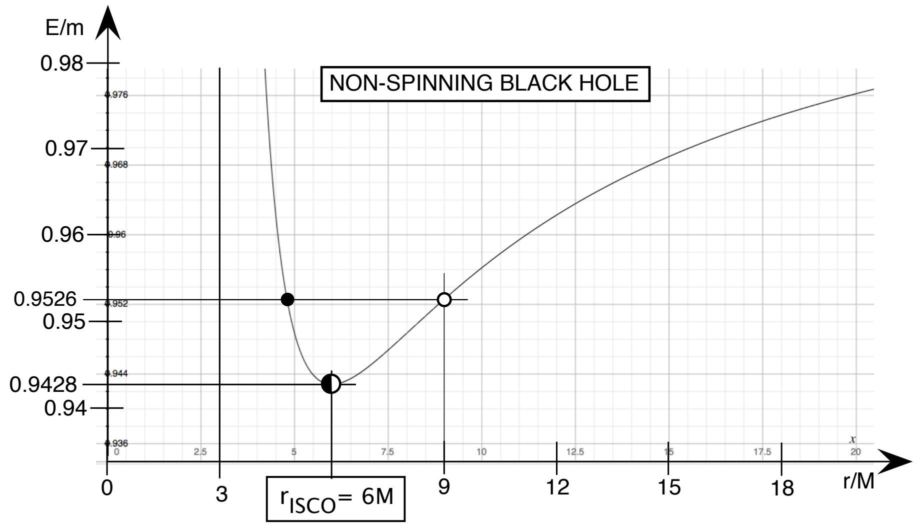


FIGURE 9 Plot of equation (64) for circular orbits around the non-spinning black hole. *Every point* on this curve represents the map energy of a circular orbit. The curve has a minimum $(E/m)_{\rm min} = (8/9)^{1/2} = 0.9428$ at $r_{\rm ISCO} = 6M$ (little half-filled circle). A horizontal line above this minimum at, say, $E/m = 0.9526$ fixes the r-value of an unstable circular orbit (little filled circle) and also the r-value of a stable circular orbit (little open circle).

These correspond to equations (58) and (56) in Newton's case.

QUERY 16. Circular orbits in Newton's limit

Check (64) and (65) in Newton's limit $r/M \to \infty$, that is $M/r \to 0$.

A. Does (65) reduce to (56)?

B. Does (64) reduce to (58). Before doing the algebra, guess the answer by comparing the vertical scales of Figures 7 and 8 and the number that E/m approaches as $r/M \to \infty$.

C. Interpret the physical difference between Newton's circular orbit energy (58) and the Newtonian limit of circular orbit energy (64).

We want to trace the result of a little friction on these orbits. To follow an analysis similar to that for Newton's circular orbits in Section 18.6, take derivatives of both sides of (64) and (65) in Query 17.

QUERY 17. Non-spinning black hole: dE/dr and dL/dr for a sequence of circular orbits.

A. Differentiate (64) and (65) to obtain, for a non-spinning black hole:

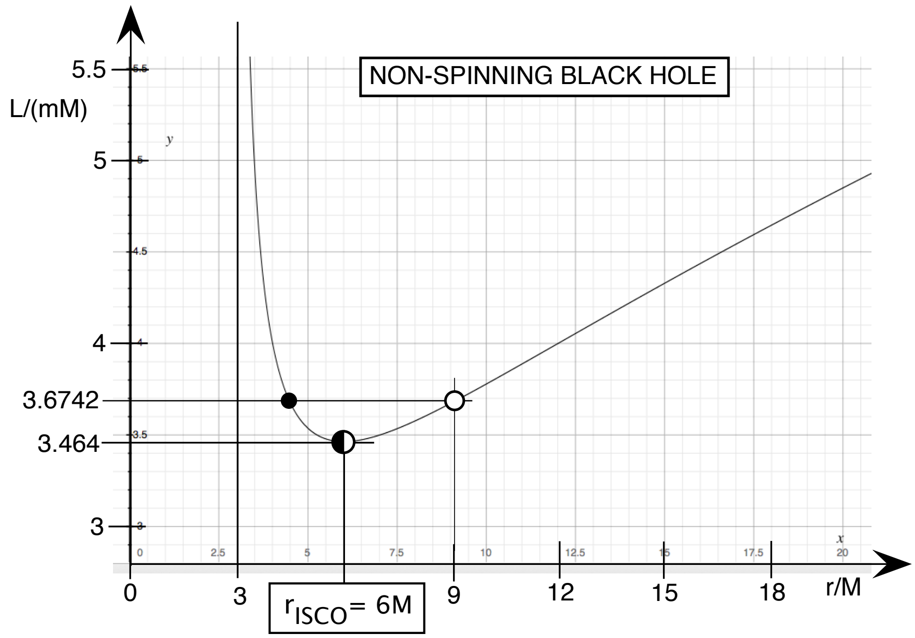


FIGURE 10 Plot of equation (65) for circular orbits around the non-spinning black hole. *Every point* on this curve represents the map angular momentum of a circular orbit. This curve has a minimum $[L/(mM)]_{\min} = (12)^{1/2} = 3.464$ at $r_{\rm ISCO} = 6M$ (little half-filled circle). A horizontal line above this minimum at, say, $L/(mM) = 3.6742$ fixes the r-value of an unstable circular orbit (little filled circle) and a stable circular orbit (little open circle).

$$\frac{dE}{dr} = \frac{mM(r-6M)}{2r^3(r-3M)^{3/2}} \qquad \text{(sequence of circular orbits)} \qquad (66)$$

$$\frac{dL}{dr} = \frac{mM^{1/2}(r-6M)}{2(r-3M)^{3/2}} \qquad \text{(sequence of circular orbits)} \qquad (67)$$

B. Show that when $r \gg M$, these reduce to Newton's results (59) and (60).

C. Show how Figure 8 reflects the result that the right sides of both equations (66) and (67) reverse sign at $r = 6M$.

Figure 9 plots E/m vs r/M from equation (64), while Figure 10 plots L/m vs r/M from equation (65). These figures show what equations (66) and (67) describe: E and L have minima at $r = 6M$ for circular orbits around a non-spinning black hole and both have positive slopes, $dL/dr > 0$ and $dE/dr > 0$, for $r > 6M$. From Figure 8, orbits in this range of r-values are stable because r-displacement in either direction at constant E moves the circle point into a forbidden map energy region (Definition 2).

Comment 7. Not another kind of effective potential
Figure 9 looks like an effective potential for the non-spinning black hole, but it is not. Instead, it tells us the r-values of circular orbits for *all possible* values of E/m.

Add friction: Shrinking orbits for non-spinning black hole unstable for $3M < r \leq 6M$.

Now trace the consequences of a little friction for circular orbits around the non-spinning black hole. Start with a stone in a circular orbit at $r > 6M$ in Figures 9 and 10. Friction causes the orbit to lose both angular momentum and energy. Because $dL/dr > 0$ and $dE/dr > 0$ for $r > 6M$, therefore both L and E decrease when r decreases: the orbit shrinks, as confirmed by equations (61) and (62).

What happens after the orbit r-value reaches $r = 6M$, where $L/(mM) = (12)^{1/2} = 3.4641$ and $E/m = (8/9)^{1/2} = 0.9428$? Answer: Friction continues to drain angular momentum and energy. But $dL/dr = 0$ and $dE/dr = 0$ for circular orbits at $r = 6M$, so the stone can no longer change L and E by changing its orbital r-value: *No circular orbits exist for* $L/(mM) < (12)^{1/2}$ *and* $E/m < (8/9)^{1/2}$. Equations (61) and (62) bear this out: Δr is undefined at $r = 6M$.

To determine what happens next, see circled number 1 in Figure 8: Displacement to the left does *not* move the circle point into a forbidden map energy region. Instead, it leads to a continual decrease of r. Result: The stone spirals inward across the event horizon.

For stable circular orbit: $dE/dr > 0$ and $dL/dr > 0$

As long as $dE/dr > 0$ and $dL/dr > 0$ along a sequence of circular orbits, the orbits are stable. Query 17 shows that dE/dr and dL/dr both change sign at $r = 6M$, which marks the transition to unstable circular orbits. Comparing Figures 8 through 10, we see that circular orbits are unstable at r-values where $dE/dr < 0$ and $dL/dr < 0$.

The smallest r-value of a stable circular orbit is called r_{ISCO}. The subscript ISCO stands for **Innermost Stable Circular Orbit,** defined in Section 8.5.

Recall that the ISCO is both stable and unstable: Increasing the r-value at the same energy puts the stone into a forbidden map energy region, but decreasing the r-value does not; the orbit is stable to increasing r, but unstable to decreasing r. We can call the r_{ISCO} orbit a **half-stable circular orbit.**

?

Objection 6. *Wrong again! You tell us that "Map quantities L and E are not measured quantities." So how can you say that friction causes them to decrease? Only physical quantities like velocity and energy in a local frame have a measurable meaning. So you talk nonsense when you say that friction causes (unmeasurable!) map quantities L and E to decrease!*

!

Guilty as charged! Values of L and E are not directly measurable. However, we can use global coordinates to predict, for example, the energy E_{shell} of the stone measured in a local inertial shell frame. For a non-spinning black hole, the result shows that when the tangential velocity measured, for example, in the local shell frame decreases, then L decreases, and *vice versa*. And when local E_{shell} decreases, then map E also decreases, and *vice versa*. Map angular momentum and map energy serve as "proxies" for measurable quantities and both do decrease as claimed. Chapter 19 carries out this analysis for a spinning black hole using the ring frame.

18.8 ■ STABILITY OF CIRCULAR ORBITS: SPINNING BLACK HOLE

Find four types of stable and unstable circular orbits.

Four types of circular orbits for spinning black hole.

How many stable and unstable circular orbits exist around the spinning black hole? We follow an analysis similar to the one for the non-spinning black hole (Section 18.7). But there is a complication: The spinning black hole has four types of circular orbits, introduced in Section 18.4. The symmetry among these four types allows us to concentrate on the two types with positive map energy outside the event horizon, Type 1 and Type 4. (The other two types are related to these by sign changes, described in Query 9.) Figure 11 plots effective potentials that show locations of two Type 1 circular orbits. Compare this plot with Figure 8 for the non-spinning black hole. Figure 12 plots effective potentials that show locations of two Type 4 circular orbits.

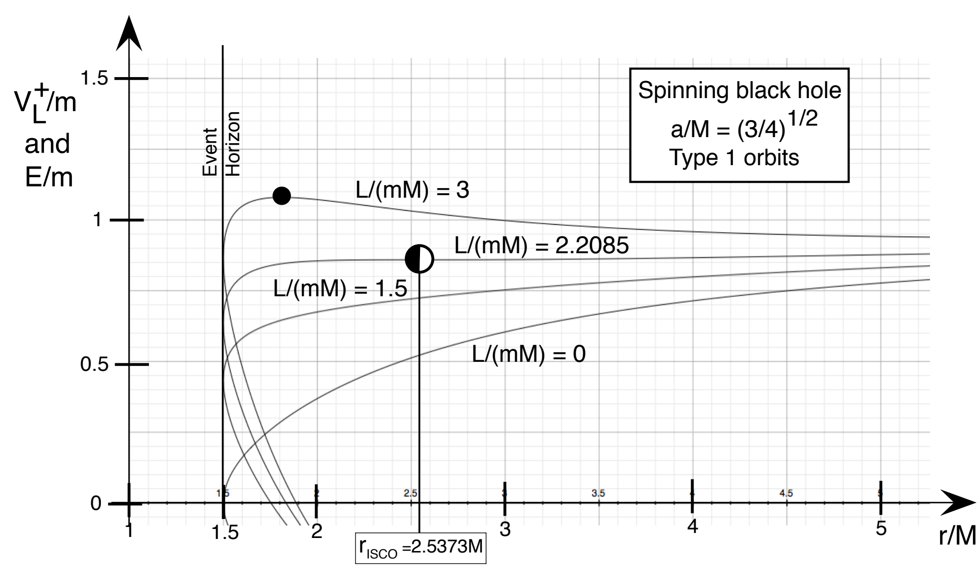


FIGURE 11 Magnified view of the effective potential $V_L^+(r)$ near the event horizon for several values of $L/(mM)$, showing r-values of two Type 1 (prograde) circular orbits from (32). Compare with Figure 8. In this plot the forbidden map energy region exists below and to the right of each curve for every value of map angular momentum, including zero. The horizontal axis begins at $r/M = 1$ to hide the distraction of unstable circular orbits inside the Cauchy horizon.

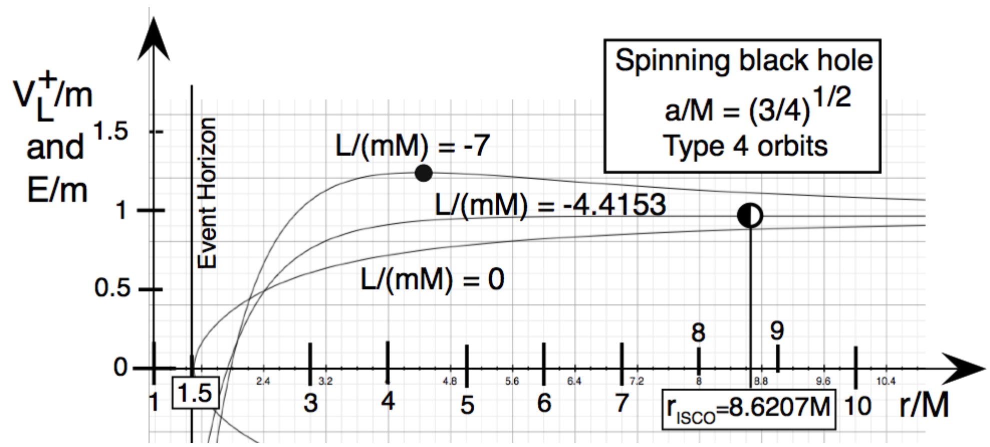


FIGURE 12 Magnified view of the effective potential $V_{\mathrm{L}}^{+}(r)$ near the event horizon for several values of $L/(mM)$, showing r-values of two Type 4 (retrograde) circular orbits from (37).

> **Comment 8. Always a forbidden map energy region for spinning black hole**
> Figures 11 and 12 show that for $a/M = (3/4)^{1/2}$ the forbidden map energy region exists for every value of the stone's angular momentum, including zero. This result is general: For *every* spinning black hole and for *every* value of the stone's angular momentum in orbit around it, *every pair* of effective potentials $V_{\mathrm{L}}^{-}(r)$ and $V_{\mathrm{L}}^{+}(r)$ embrace a forbidden map energy region.

Figure 13 plots E/m vs. r/M from Type 1 (prograde) and Type 4 (retrograde) orbits for $a/M = (3/4)^{1/2}$ from equations (32) and (38). Figure 14 shows corresponding plots of $L/(mM)$ vs. r/M from equations (31) and (37). Sample horizontal lines show pairs of unstable and stable orbits at the same map energy or map angular momentum.

Adding friction shrinks stable orbits for spinning black hole

To see where and why circular orbits become unstable, start with the stone in a stable prograde (Type 1) circular orbit at large r. Now introduce a little friction that decreases the stone's energy E. Figures 13 and 14 show positive derivatives $dL/dr > 0$ and $dE/dr > 0$ for stable Type 1 orbits at large r. Then equations (61) and (62) tell us that the r-value of the orbit shrinks.

Orbits stable down to r_{ISCO}

The condition for stability of Type 1 orbits is $dE/dr > 0$ and $dL/dr > 0$ from equations (31) and (32), or equivalently $dV_{\mathrm{L}}^{+}/dr = 0$ and $d^2V_{\mathrm{L}}^{+}/dr^2 > 0$ (Table 18.1). Either way gives, after lots of algebra, the inequality:

$$r^2 - 6Mr + 8a(Mr)^{1/2} - 3a^2 > 0 \qquad \text{(stable orbits, Types 1 and 2)} \qquad (68)$$

Although we derived equation (68) for Type 1, it is also valid for Type 2 ($E = V_{\mathrm{L}}^{-}$ and, outside the event horizon, $L < 0$). You can see this from Figure 4 and Query 8. Both stable and unstable circular orbits come in pairs.

The left hand side of equation (69) vanishes at only one r-value and is negative for smaller r-values. The r-value of the innermost stable circular orbit is therefore given by the solution of this equation:

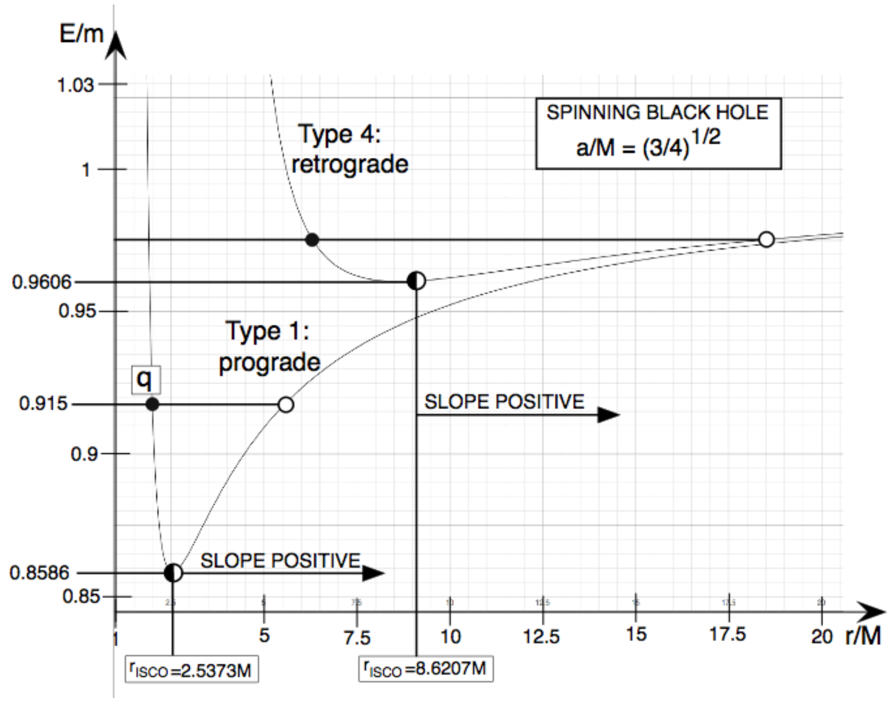


FIGURE 13 Map energy vs r for circular orbits outside the event horizon of the spinning black hole with $a/M = (3/4)^{1/2}$, from equation (32) for Type 1 and equation (38) for Type 4, showing $r_{\rm ISCO}$ at the minima and one example of unstable and stable orbits for each type. The prograde circular orbit labeled q at $r = 1.95$ and energy $E = 0.915M$ is the orbit labeled q in Figure 3; the figure above proves that orbit q in Figure 3 is unstable.

$$r_{\rm ISCO}^2 - 6Mr_{\rm ISCO} + 8a\left(Mr_{\rm ISCO}\right)^{1/2} - 3a^2 = 0 \qquad \text{(prograde orbits)} \qquad (69)$$

Stable circular orbits exist only for $r > r_{\rm ISCO}$.

Minimum $|L|$ and $|E|$ at $r_{\rm ISCO}$

For a stone in a Type 1 or 2 (prograde) circular orbit at $r_{\rm ISCO}$, further decrease of $|L|$ or $|E|$ can no longer result in a circular orbit, because $|L|$ and $|E|$ have already reached their minimum values for circular orbits, shown in Figures 13 and 14. To determine what happens next, look at the little half-filled circle in Figure 11: Displacement of the stone to the left does *not* move it into a forbidden map energy region. Instead, it leads to a continual decrease of r. Result: The stone spirals inward, then crosses the event horizon!

$r_{\rm ISCO}$ for retrograde orbits.

Next turn attention to retrograde orbits, Types 3 and 4. It is simplest to start with Type 4, $E = V_{\rm L}^{+} > 0$ and $L < 0$ (Table 18.3). Then stability for Type 3 follows as a "mirror image," as was the case for prograde circular orbits. At large r for Type 4, $dE/dr > 0$ (Figure 13), while $dL/dr < 0$ (Figure

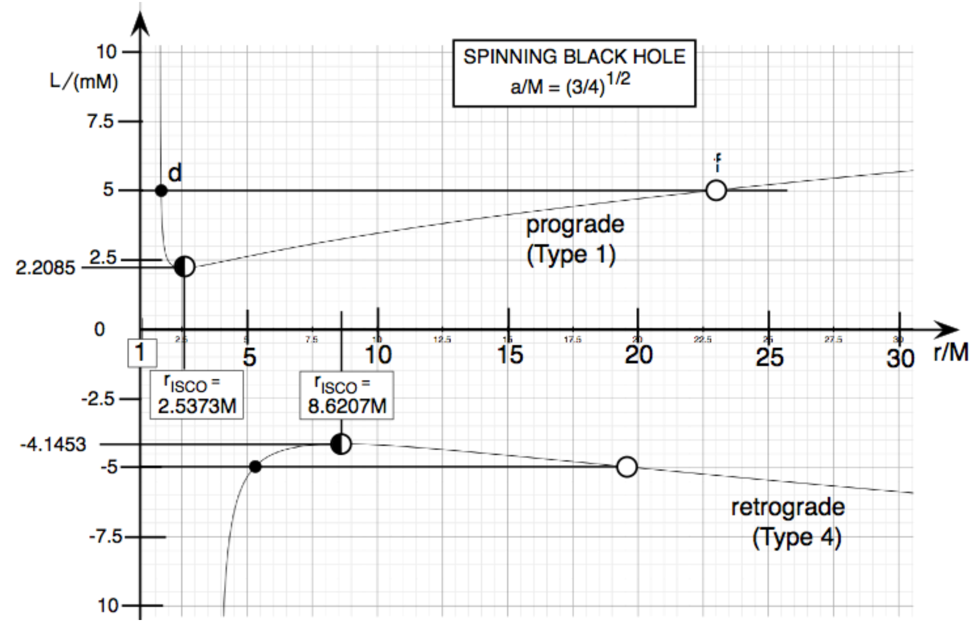


FIGURE 14 Map angular momentum vs r for circular orbits outside the event horizon of the spinning black hole with $a/M = (3/4)^{1/2}$, from equation (31) for Type 1 and equation (37) for Type 4, showing $r_{\rm ISCO}$ at the minima and one example of unstable and stable orbits for each type. Points d and f along the horizontal line at $L/(mM) = +5$ have the same labels in Figure 1 and Table 2.

14). Whether L is positive or negative, a little friction decreases $|L|$. Thus the condition for stability is that there exists a circular orbit of slightly smaller r and slightly smaller $|L|$; this condition requires that $d|L|/dr > 0$ and therefore $dL/dr < 0$ when $L < 0$ in Figure 14.

Comment 9. Signs of dE/dr and dL/dr for stable orbits
When $E < 0$, as in Type 3, the condition on E for stability becomes $d|E|/dr > 0$. For both signs of E, the stability condition is $d|E|/dr > 0$, similar to the condition $d|L|/dr > 0$ for stability. The reason for this is that a little friction decreases both $|L|$ and $|E|$ regardless of the signs of L and E, and for orbits to exist with smaller $|L|$ and $|E|$, the graphs of $|L(r)|$ and $|E(r)|$ must have positive slope with respect to r.

The stability condition for Type 4 circular orbits is $dE/dr > 0$ and $dL/dr < 0$ from equations (37) and (38), or equivalently $dV_{\rm L}^{+}/dr = 0$ and $d^2V_{\rm L}^{+}/dr^2 > 0$ (Table 18.1). Either way yields, after lots of algebra, the inequality:

$$r^2 - 6Mr - 8a(Mr)^{1/2} - 3a^2 > 0 \qquad \text{(stable orbits, Types 3 and 4)} \qquad (70)$$

Although we derived equation (70) for Type 4 orbits, it is also valid for Type 3. At $r = r_{\text{ISCO}}$, equation (70) becomes an equality.

$$r_{\text{ISCO}}^2 - 6Mr_{\text{ISCO}} - 8a\left(Mr_{\text{ISCO}}\right)^{1/2} - 3a^2 = 0 \qquad \text{(retrograde orbits)} \qquad (71)$$

As in Query 10, this result follows from the prograde case merely by changing the sign of a. We can solve the two equations (69) and (71) to find two expressions for $a(r_{\text{ISCO}})$.

$$a(r_{\text{ISCO}}) = \pm\frac{1}{3}\left(Mr_{\text{ISCO}}\right)^{1/2}\left[4 - \left(3\frac{r_{\text{ISCO}}}{M} - 2\right)^{1/2}\right] \qquad (72)$$

The plus sign in this equation describes prograde orbits and the minus sign describes retrograde orbits.

Limits on value of r_{ISCO}

Black holes exist only for $0 \leq a/M \leq 1$. Equation (72) then limits prograde and retrograde orbits to to the following values of r_{ISCO}:

$$M \leq r_{\text{ISCO}} \leq 6M \qquad (0 \leq a/M \leq 1,\ \text{prograde}) \qquad (73)$$
$$6M \leq r_{\text{ISCO}} \leq 9M \qquad (0 \leq a/M \leq 1,\ \text{retrograde}) \qquad (74)$$

Values of r_{ISCO}

The curves in Figure 15 plot a as a function of r_{ISCO} from equation (72). Bardeen, Press, and Teukolsky solved (72) to give r_{ISCO} as a function of a, a combination of three equations (see the references):

$$\frac{r_{\text{ISCO}}}{M} = 3 + Z_2 \mp \left[(3 - Z_1)(3 + Z_1 + 2Z_2)\right]^{1/2} \qquad (75)$$

minus sign for prograde, plus sign for retrograde, and

$$Z_2 \equiv \left(3a^2/M^2 + Z_1^2\right)^{1/2} \qquad (76)$$
$$Z_1 \equiv 1 + \left(1 - a^2/M^2\right)^{1/3}\left[(1 + a/M)^{1/3} + (1 - a/M)^{1/3}\right] \qquad (77)$$

QUERY 18. Values of r_{ISCO} for $a/M = (3/4)^{1/2}$. (*Optional*)
Use equations (75) through (77) to verify the following values of r_{ISCO} for a spinning black hole with $a/M = (3/4)^{1/2}$:

$$r_{\text{ISCO}}/M = 2.537331951 \qquad \text{for prograde orbit} \qquad (78)$$
$$r_{\text{ISCO}}/M = 8.620665097 \qquad \text{for retrograde orbit}$$

Summary: For circular orbits around a spinning black hole, a small amount of friction decreases the absolute values of map energy and map

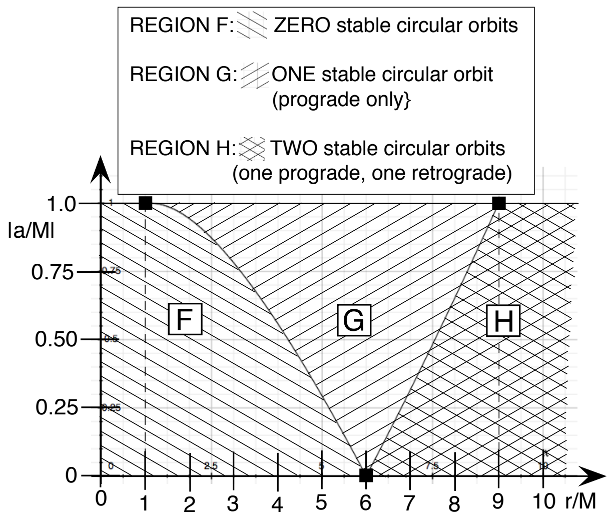


FIGURE 15 How many *stable* circular orbits exist at a given r for different values of the spin parameter a/M? This figure uses inequalities (68) and (70) to answer that question. The regions are separated by curves for $r_{\rm ISCO}$ from equations (69) and (71). In Region F there are *zero* stable circular orbits; in Region G there is *one* stable prograde circular orbit; in Region H there are *two* stable circular orbits, one prograde and one retrograde. Compare this figure with Figure 6 for *all* circular orbits.

Summary: Sequence of stable circular orbits

angular momentum, $|E|$ and $|L|$, which causes the stone to occupy a sequence of stable circular orbits with decreasing r—until both $|E|$ and $|L|$ reach their minima at $r = r_{\rm ISCO}$. Increasing black hole spin moves the ISCO inward from $r_{\rm ISCO} = 6M$ to $r_{\rm ISCO} = M$ for prograde orbits and outward from $r_{\rm ISCO} = 6M$ to $r_{\rm ISCO} = 9M$ for retrograde orbits (Figure 15). These results have profound consequences for the accretion disk around the spinning black hole, which we explore in Section 18.10.

18.9 ■ TIMING CIRCULAR ORBITS FROM A LARGE r

On whose watch?

We are (thank goodness!) far from a spinning black hole. *Surprise:* We can nevertheless hold a stopwatch on each circular orbit in the sequence of circular orbits as a stone works its way inward through the accretion disk (Section 18.10). In practice we might observe a blob of incandescent matter as it moves in each circular orbit. This section provides the background for such an observation.

Flash-emitting orbiter

Replace the circulating blob with an astronaut in a circular orbit who emits a flash of light as she completes each orbit. Equation (52) tells us the lapse $\Delta\tau_{\rm emit}$ on her wristwatch for one orbit, where we have added the

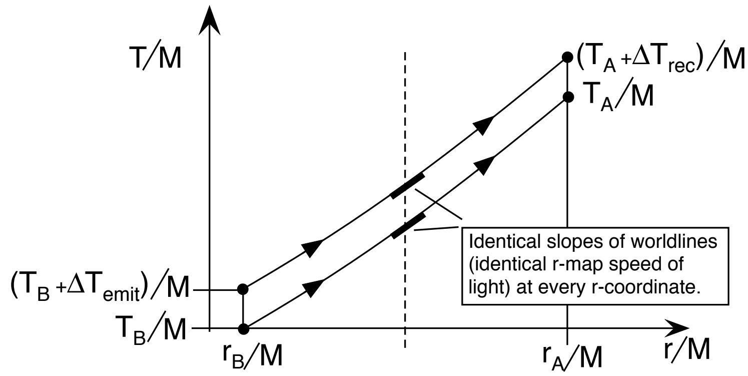


FIGURE 16 Schematic plot in Doran global coordinates of worldlines of two flashes emitted by the Below emitter at the beginning and end of one circular orbit and received by a distant Above observer. The lapse ΔT_{rec} between receptions is equal to the lapse ΔT_{emit} between emissions. Similar plot for the Global Positioning System: Figure 2 in Section 4.2.

subscript "emit" for clarity in what follows. How does the orbiter *know* that she has completed one orbit? The pattern of stars she sees overhead repeats as she returns to the same r, Φ. We have not yet predicted this star pattern, which depends on the observer's orbit and the worldline of light from each distant star to the observer. Still, we know that this visual pattern repeats, so the observer can emit a flash at each repetition.

Equation (51) tells us the Doran coordinate lapse ΔT_{emit} between flash emissions by the orbiter. A distant observer at rest in Doran coordinates $(dr/dT = d\Phi/dT = 0)$ receives two sequential flashes emitted by the orbiter and records his wristwatch time lapse $\Delta\tau_{\text{rec}}$ between these two receptions.

Timing these flashes from far away

At the location of this stationary distant observer the Doran metric reduces to $d\tau^2 = dT^2$. Therefore, the distant observer measures a time lapse $\Delta\tau_{\text{rec}} = \Delta T_{\text{rec}}$ between flashes, where ΔT_{rec} is the Doran coordinate lapse between the receptions of sequential light flashes.

How is ΔT_{rec} related to ΔT_{emit}? Light rays travel along curves $r(T)$ in global coordinates. Let one light flash be emitted at $r = r_{\text{A}}$ and $T = T_{\text{A}}$ and a second one from the same r-value at $T = T_{\text{A}} + \Delta T_{\text{emit}}$ (Figure 16). When are these two flashes received by a distant observer stationary in Doran coordinates?

We cannot answer this question without integrating the equation of motion of light, but we can answer a simpler question: What is the difference between two global T-values of reception by a distant observer? That is, how is ΔT_{rec} related to ΔT_{emit}?

Figure 16 shows that at every value of r the curves $r(T)$—or equivalently $T(r)$—have the same slope for two sequential light pulses emitted from the same global location. Therefore these curves are vertically displaced by the same offset in Doran T at every r-value. As a result, $\Delta T_{\rm rec} = \Delta T_{\rm emit}$.

This analysis leads to the prediction that the wristwatch time $\Delta\tau_{\rm far}$ for one orbit measured by a distant observer at rest in Doran coordinates is equal to the lapse ΔT for one orbit given by equation (51). This answers the question, "What is the wristwatch time lapse $\Delta\tau_{\rm far}$ for one circular orbit measured by a distant observer?"

QUERY 19. Careful with wristwatch times!
Show that the wristwatch time $\Delta\tau_{\rm far}$ between reception of flashes for the distant observer is NOT equal to the wristwatch time $\Delta\tau_{\rm emit}$ between emission of flashes for the orbiter.

Pulse emitter: black hole GR1915

Figure 17 shows X-ray pulses emitted by the spinning black hole labeled GR1915, with about 14 times the mass of the Sun located near the plane of the galaxy about 40 light-years from us. A companion star feeds a pulse of material to the accretion disk of GR1915. This pulse of matter heats to high temperature and emits radiation whose pressure temporarily prevents more matter from entering the accretion disk from the companion. After the accreted material drops into the black hole, a new blob enters the accretion disk from the companion. The resulting "heartbeat" of X-rays are about 50 seconds apart.

18.10 ■ THE ACCRETION DISK

Circling toward doom

QUASAR: Emission as material circles inward through accretion disk.

Section 8.6 constructed the toy model of an accretion disk around a non-spinning black hole, but we have not observed a non-spinning black hole, much less one with an accretion disk. We do observe energetic radiation from quasars, each of which appears to be a spinning black hole surrounded by an accretion disk that emits this radiation. What creates this radiation? Interactions within the accretion disk are complex and defy simple analysis, but here is the basic idea: The accretion disk consists of dust and particles in orbit. This material changes energy as it moves inward through a sequence of circular orbits. The change in energy heats up the accretion disk, with consequent emission of radiation.

Assume that material in the accretion disk passes in sequence through a series of circular orbits. Initial circular orbits are at large r-values; their final circular orbit is at $r_{\rm ISCO}$, after which the material spirals inward through the event horizon. We cannot see radiation emitted after stones and dust pass through the event horizon. Now for some details.

Stone in distant circular orbit has $E/m = 1$ and $L/(mM) = 0$.

Start with a stone far from the black hole, a stone that moves so slowly in its circular orbit that it is effectively at rest in Doran global coordinates, with initial map energy $E/m = 1$ and initial map angular momentum $L/(mM) = 0$. Consider this stone to be in a forward, prograde Type 1 circular orbit.

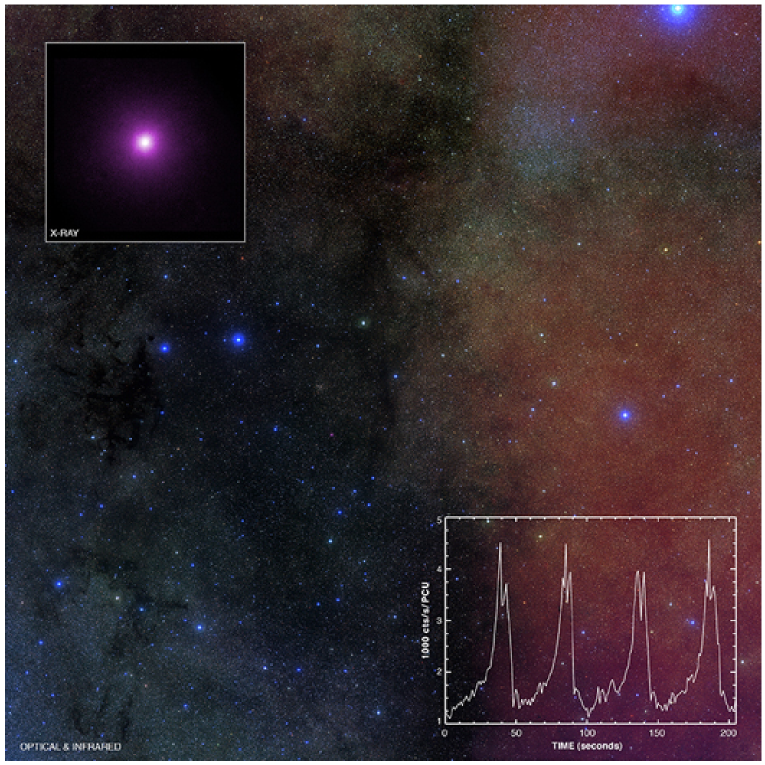


FIGURE 17 Upper left corner: the spinning black hole GR1915-105 fed by material from a companion star (not visible). Lower right corner: the “heartbeat” of emitted X-rays.

For very large r, $R \to r$ and the Doran metric (2) becomes:

$$d\tau^2 \to dT^2 - dr^2 - r^2 d\Phi^2 \qquad \text{(for } r \to \infty) \qquad (79)$$

This is the metric of flat spacetime in which we can define local shell coordinates: $\Delta t_{\text{shell}} \equiv \Delta T$, $\Delta y_{\text{shell}} \equiv \Delta r$, and $\Delta x_{\text{shell}} \equiv \bar{r}\Delta\Phi$. A stone at rest in this local frame must have $(E/m)_{\text{shell}} = 1 = E/m$, where E/m is the map energy. *Summary: Far from the black hole the directly measurable shell energy* $(E/m)_{\text{shell}}$ *of a stone is equal to its Doran map energy* E/m.

Next the stone loses map energy as it passes gradually inward through a series of circular orbits of decreasing r until it reaches the innermost stable circular orbit at r_{ISCO}. How much map energy does the stone lose during this process? Assume the material emits its change in map energy in the form of radiation. What total radiated energy do we detect far from the black hole?

What is the map energy of the stone in the ISCO orbit just before it drops across the event horizon?

When $a/M = (3/4)^{1/2}$, equations (75) through (76) tell us that $r_{\rm ISCO} = 2.5373M$ so that from equation (32) $E/m = 0.8586$. Hence the radiated energy is $\Delta E = (1 - 0.8586)m = 0.1414m$.

In contrast, when $a/M = 1$, then equations (75) through (76) tell us that $r_{\rm ISCO} = M$ so that from equation (32) $E/m = 0$. Hence the radiated energy is $\Delta E = m$. The entire rest energy of the stone is emitted as radiation. No wonder the quasar shines so brightly!

QUERY 20. More typical emission of radiation

A more typical upper value of a/M for a spinning black hole is 0.85. Use Figure 15 to estimate numerical value of $r_{\rm ISCO}$ for $a/M = 0.85$. Optional: Use equations (75) through (77) to calculate the numerical value of $r_{\rm ISCO}$ to four decimal digits in this case.

QUERY 21. Power output of a quasar

A distant quasar swallows $m = 10M_{\rm Sun}$ = ten times the mass of our Sun every Earth-year. Recall that watts equals joules/second and, from special relativity,
$\Delta E[\text{joules}] = \Delta m[\text{kilograms}]c^2[\text{meters}^2/\text{second}^2]$. Assume that this quasar has $a/M = 0.85$.

A. How many watts of radiation does this quasar emit, according to our model?
B. Our Sun emits radiation at the rate of approximately 4×10^{26} watts. The quasar is how many times as bright as our Sun?
C. Compare your answer in Item B to the total radiation output of a galaxy of approximately 10^{11} Sun-like stars.

QUERY 22. How long does a quasar shine?

We see most quasars with large red shifts of their light, which means they were formed not long after the Big Bang, about 14×10^9 years ago. A typical quasar is powered by a black hole of mass less than 10^9 solar masses. Explain, from the results of Query 21, what this says about the lifetime during which the typical quasar shines.

18.11 ■ CHAPTER SUMMARY

Key ideas of the chapter

Two effective potentials

The spinning black hole has not one but *two* effective potentials, which depend on the stone's angular momentum and the spin parameter of the black hole. Circular orbits of a stone are possible at maxima and minima of these effective potentials, which (for different values of the stone's map angular momentum) can occur at most r-values outside the event horizon and inside the Cauchy horizon.

Forbidden energy region

Each pair of effective potentials encloses a **forbidden map energy region**. A stone cannot have its map energy in a forbidden map energy region.

Prograde and **retrograde** orbits

We divide circular orbits into two classes, **prograde** and **retrograde**. In a prograde orbit the stone "revolves in the direction that the black hole rotates" in global Doran coordinates, $d\Phi/dT > 0$, while in a retrograde orbit the stone revolves in the opposite direction, $d\Phi/dT < 0$.

Stable circular orbits and the *innermost* stable circular orbit

Most circular orbits around the spinning black hole are unstable; a few are stable. To analyze orbital stability, we trace the effects of a little friction, which slowly decreases orbital r (leaving the orbit effectively circular), while it also decreases values of $|L|$ and $|E|$. The r-value of the **innermost stable circular orbit**, labeled r_{ISCO}, occurs when values of $|L|$ and $|E|$ for a circular orbit reach their minima. When the circular orbit of a stone reaches r_{ISCO}, further loss of energy to friction leads the stone to spiral inward through the event horizon.

Accretion disk

In Nature a spinning black hole is surrounded by an *accretion disk* that consists of material circulating in stable prograde circular orbits in the equatorial plane. (Why prograde? Because a stone circulating in a prograde ISCO has a much smaller map energy than a stone in a retrograde ISCO; see Figure 13.) Orbiting dust and particles emit energy in the form of electromagnetic radiation as they descend gradually through circular orbits of decreasing r. A distant stationary observer measures this emitted radiation to have energy equal to the map energy E/m. We continue to observe this radiation as material spirals down from the minimum stable ISCO orbit, but not after the material crosses the event horizon.

18.12 ■ EXERCISES

0. SOLVED EXERCISE. Add angular momentum to a maximum-spin black hole?

Suppose that the spinning black hole has maximum spin: $a/M = 1$. Can you increase this (maximum!) spin by sending into the black hole a stone with positive angular momentum? Try a specific example:

Figure 18 plots the effective potential for a black hole with maximal spin $a/M = 1$ and incoming stones with angular momentum $L/(mM) = 5$ and three different map energies, including $E_{\text{C}}/M = 6$, above the energy of the forbidden map energy regions. When it falls into the black hole, can this highest-energy stone increase the black hole spin beyond its maximum value? Answer this question using the following steps.

$$\frac{E}{m} = 6 \qquad \text{and} \qquad \frac{L}{mM} = 5 \tag{80}$$

A. When this stone enters the black hole, it changes the black hole's mass according to equation (28) in Section 6.5 and increases the black hole's

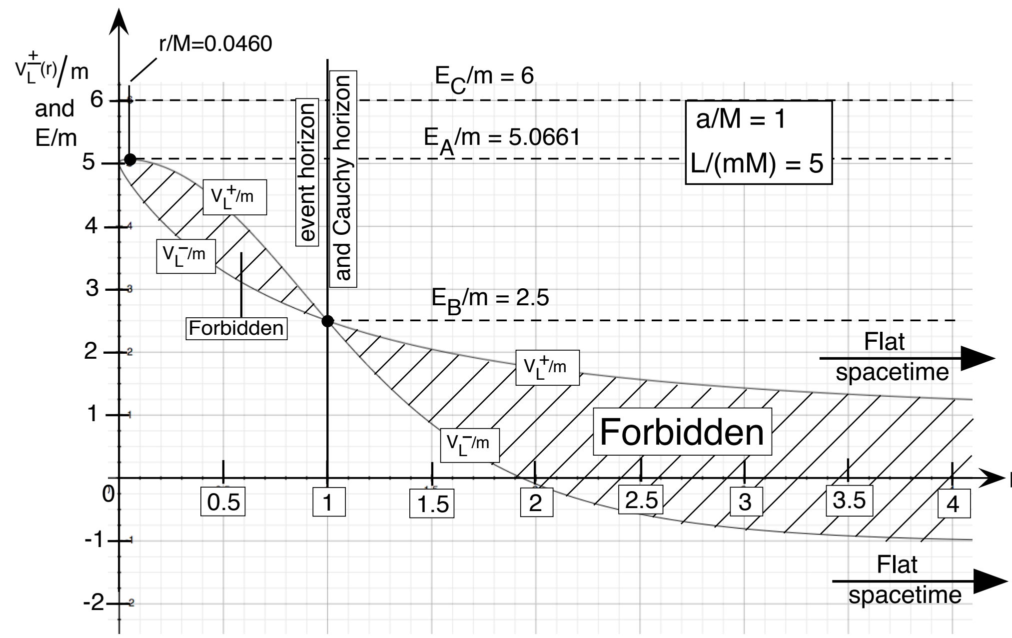


FIGURE 18 Effective potentials $V_{\mathrm{L}}^{+}(r)$ and $V_{\mathrm{L}}^{-}(r)$ for a stone with $L/m = 5M$ in orbit around a spinning black hole with maximum spin parameter $a/M = 1$. There are two stable circular orbits at larger r/M than the maximum in this diagram, one prograde, one retrograde. Two of the dashed lines show map energies E_{A}/m and E_{B}/m of two stones that take up unstable circular orbits. Can a third stone, with $E_{\mathrm{C}}/m = 6$ and angular momentum $L/(mM) = 6$ fall into this black hole and increase its angular momentum above the maximum?

angular momentum beyond the old maximum in equation (2) in Section 17.1:

$$M_{\text{new}} = M + E_{\text{stone}} \qquad \text{and} \qquad J_{\text{new}} = M^2 + L_{\text{stone}} \tag{81}$$

B. Then equation (1) in Section 17.1 tells us that

$$\frac{a_{\text{new}}}{M_{\text{new}}} = \frac{J_{\text{new}}}{M_{\text{new}}^2} = \frac{M^2 + L_{\text{stone}}}{(M + E_{\text{stone}})^2} = \frac{1 + L_{\text{stone}}/M^2}{(1 + E_{\text{stone}}/M)^2} \tag{82}$$

C. Now L_{stone} and E_{stone} are properties of the incoming stone, which has mass $m \ll M$, therefore $L_{\text{stone}} \ll M^2$ and $E_{\text{stone}} \ll M$, so we can approximate (82) with the formula inside the front cover:

$$\frac{a_{\rm new}}{M_{\rm new}} \approx (1 + L_{\rm stone}/M^2)(1 - 2E_{\rm stone}/M) \tag{83}$$

$$\approx 1 + \frac{L_{\rm stone}}{M^2} - \frac{2E_{\rm stone}}{M} \tag{84}$$

$$\approx 1 + \left(\frac{L_{\rm stone}}{mM}\right)\left(\frac{m}{M}\right) - 2\left(\frac{E_{\rm stone}}{m}\right)\left(\frac{m}{M}\right)$$

$$\approx 1 + \frac{m}{M}(5 - 2 \times 6) = 1 - 7\frac{m}{M} \tag{85}$$

The step from (83) to (84) neglects the product of two small quantities. The final expression (85) is (slightly) smaller than the initial (maximum) spin parameter $a/M = 1$.

For this example, the incoming stone does not increase the spin parameter of the black hole. Why not? Because it increases the mass of the black hole, which changes the value of its maximum spin.

1. *Optional:* Repeat exercise 0 with GRorbits

Use interactive GRorbits software to plot the analysis of Exercise 0

A. Plot the case described in Exercise 0 with your choice of numerical values for $m \ll M$ and $M = 10M_{\rm Sun}$.

B. Repeat Item A for $M = 10^7 M_{\rm Sun}$. Describe how your results differ from those in Item A?

C. Report what you have learned in this exercise that supplements or reinforces results in Exercise 0.

2. Fast orbits!

Write a computer program to fill in Tables 18.4 and 18.5 for a spinning black hole with $a/M = (3/4)^{1/2}$. Write "None" in entries for which circular orbits do not exist. Section 18.10 shows that a distant observer records a wristwatch time equal to map ΔT for one circular orbit. In the table, "progr." means "prograde" and "retrogr." means "retrograde".

A. For a "small" black hole with mass $M = 10M_{\rm Sun}$ fill in entries in Table 18.4.

B. For a "large" black hole with mass $M = 4 \times 10^6 M_{\rm Sun}$ (the approximate mass of the spinning black hole at the center of our galaxy), fill in the entries in Table 18.5.

TABLE 18.4 "Small" black hole: "TIMES" for one orbit, in SECONDS

$M = 10M_{\text{Sun}}$	$r/M =$ 0.2	$r/M =$ 2	$r/M =$ 6	$r/M =$ 10	$r/M =$ 20
Newton time					
Nonspin $\Delta\tau$					
Spin progr. $\Delta\tau$					
Spin retrogr. $\Delta\tau$					
Nonspin ΔT					
Spin progr. ΔT					
Spin retrogr. ΔT					

NOTE: Spinning black hole has $a/M = (3/4)^{1/2}$. Equation (52) for τ and (51) for T.

TABLE 18.5 "Large" black hole: "TIMES" for one orbit, in DAYS

$M = 4 \times 10^6 M_{\text{Sun}}$	$r/M =$ 0.2	$r/M =$ 2	$r/M =$ 6	$r/M =$ 10	$r/M =$ 20
Newton time					
Nonspin $\Delta\tau$					
Spin progr. $\Delta\tau$					
Spin retrogr. $\Delta\tau$					
Nonspin ΔT					
Spin progr. ΔT					
Spin retrogr. ΔT					

NOTE: Spinning black hole has $a/M = (3/4)^{1/2}$. Equation (52) for τ and (51) for T.

3. Can a stone exist in a region where the effective potential is not real-valued?

In Section 18.2 we found from equation (16) that the effective potentials are not real-valued (do not exist) at r-values for which the horizon function H is imaginary, namely between r_{C} and r_{E}. This seems to imply that the equation of motion (15) for $dr/d\tau$ is complex-valued, so the stone cannot move or even exist between the horizons. Demonstrate conclusively that the stone can exist and move between the two horizons.

4. Forbidden map energy region for non-spinning black hole?

Review the effective potential diagrams for the *non-spinning* black hole in Chapter 8 Circular Orbits and answer the following questions without doing any calculation.

A. Show that a forbidden map energy region exists for the non-spinning black hole.

B. Does this forbidden map energy region extend all the way to flat spacetime, $r \to \infty$?

C. What is the experimental (observational) consequence—if any—of the forbidden map energy region near the non-spinning black hole for an observer far away where spacetime is flat?

D. *Optional:* Take the limit of equation (16) as $a/M \to 0$ and $L/(mM) \to 0$. Plot the resulting effective potential curve for a stone moving radially near a non-spinning black hole.

5. Forward time travel using a knife edge circular orbit of a spinning black hole.

Review Exercise 7 in Chapter 8. The Space Administration is now accepting proposals for forward time travel that use a forward prograde knife-edge circular orbit around a spinning black hole with $a/M = (3/2)^{1/2}$. They consider a satellite with a non-relativistic velocity far from the black hole so that $E/m \approx 1$. While still far from the black hole, the spaceship captain uses small rocket thrusts to achieve the value of map angular momentum L required so that $V_L^+/m = E/m = 1$ on the peak of the $V_L^+(r)/m$ curve.

A. Substitute the condition that $V_L^+/m = 1$ at the peak of the $V_L^+(r)/m$ curve into equation (32). Solve the resulting equation for r.

B. Substitute the solution of Item A into (31) to find the factor $d\tau/dT$ for the spaceship in this knife-edge orbit. What speed in flat spacetime gives the same time-stretch ratio?

C. Compare $d\tau/dT$ in Item B with the time-stretch ratio for the non-spinning black hole (Exercise 7, Item B in Chapter 8).

6. Effect of friction when starting from an unstable circular orbit

Section 18.7 analyzes the motion of a stone that starts from a *stable* circular orbit at $r > 6M$ around a non-spinning black hole, and loses map energy and angular momentum through friction (see Figures 9 and 10). Use Figures 9 and 10 to answer the following question: What happens if a stone is in an *unstable* circular orbit at $r < 6M$, then loses map energy and map angular momentum in small steps through friction?

7. How many stable circular orbits are there for the non-spinning black hole?

Figure 15 shows that at $a/M = 0$ regions F, G and H meet in a single point at $r/M = 6$. Are there ZERO, ONE or TWO stable circular orbits there?

8. Circular orbits inside the Cauchy horizon

Figures 11 through 14 all plot the horizontal r-axes for $r/M > 1$ in order to avoid complications with the spacetime region between the singularity and the Cauchy horizon. Yet Figure 15 plots the horizontal axis all the way down to the singularity at $r/M = 0$. Use Figures 1 and 2 to explain why the region $0 < r/M < 1$ in Figure 15 is correct.

9. Stone map energy and map angular momentum at the ISCO for $a = M$

Equation (69) shows that for the maximum-spin black hole, $r_{\rm ISCO} = M$. For these values of a and $r_{\rm ISCO}$, equations (31) and (32) give indeterminate values $(L/(mM))_{\rm Type1} = 0/0$ and $(E/m)_{\rm Type1} = 0/0$.

To find the numerical values of L/m and E/m for this orbit, we need to set $r/M = 1 + \epsilon$ and take the limit of equations (31) and (32) as $\epsilon \rightarrow 0$. The answer, to one significant digit, is $L/(mM) = 1.2$ and $E/m = 0.6$.

A. Find numerical values for $L/(mM)$ and E/m to three significant digits. [Warning: our familiar approximation inside the front cover does not work everywhere in this case. Under the square root in the denominator of the right side of (31) and (32) you need to include the second (quadratic) term in the expansion, so that:
$(r/M)^{1/2} = (1+\epsilon)^{1/2} \approx 1 - \epsilon/2 - \epsilon^2/8$

B. *Optional*. Plot $V_{\rm L}^{+}(r)/m$ vs. r/M for the value of $L/(mM)$ you calculated in Item A. Check that the minimum of the effective potential occurs at $r/M = 1$ at the value of E/m you obtained in Item A.

10. Two light cone diagrams for the maximally spinning black hole ($a = M$)

A. Review Sections 3.6 through 3.9 in Chapter 3 for the meaning of *spacetime slice, light cone diagram*, and *embedding diagram*. Use the technique outlined there to construct a light cone diagram, similar to Figure 8 of Chapter 3, on the $[r, T]$ slice of a spinning black hole with $a/M = 1$.

B. Construct a light cone diagram on the $[\Phi, T]$ slice of a spinning black hole with $a/M = 1$.

C. Answer the following questions for both light cones in Items A and B: Why cannot a stone or spaceship remain static in Doran coordinates for $r < 2M$? How can a stone or spaceship still escape to infinity from $r = 2M$? Does the rotation of the black hole drag a stone or spaceship at $r = 2M$ inevitably along the direction in which the black hole spins? Is your answer to this third question coordinate-free?

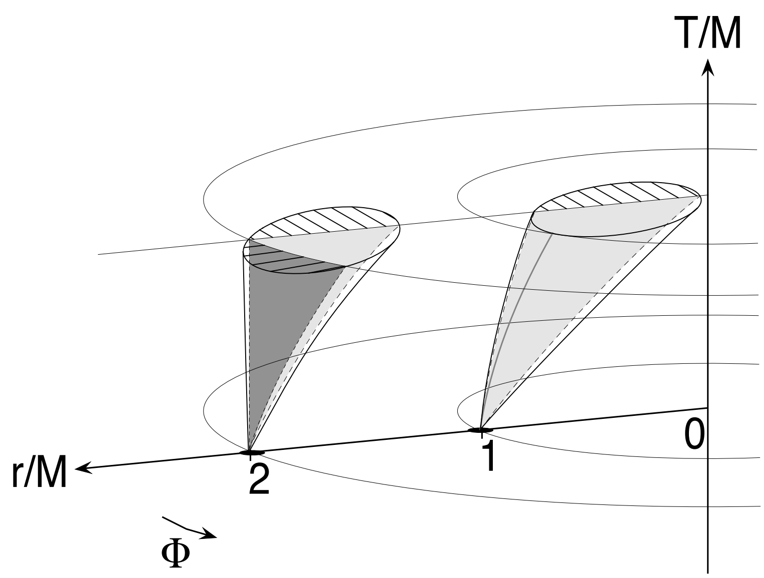


FIGURE 19 A three-dimensional Doran coordinate r, Φ, T plot of two light cones near the maximally-spinning black hole $a/M = 1$.

11. ***Difficult!*** Three-dimensional light cone diagram for the maximally-spin black hole

Figure 19 shows a three-dimensional Doran coordinate plot of two light cones for the maximally-spinning black hole. Discuss the following characteristics of these light cone plots/ plot.

A. Both light cones start on the r/M axis. Why are they both deflected inward in the r direction? Are they deflected in the Φ direction? Why or why not?

B. Why is the light cone that starts at $r/M = 1$ deflected more in the r direction than the light cone that starts at $r/M = 2$?

C. What is the physical difference between that part of the area at the top of the $r/M = 2$ light cone whose lines lie in the r direction and the part of that area whose lines lie in the Φ direction? Why is there no corresponding area of the $r/M = 1$ light cone lined in the r direction?

D. Does either light cone tell you that a circular orbit of a stone is possible at that value of r/M? If not, why not? If so, what does it say about that circular orbit?

E. Answer Item C in exercise 10 for the two lightcones of Figure 19.

12. Light cone diagrams for a spinning black hole with $a/M = (3/2)^{1/2}$

Refer to your answers for Items A through C of exercise 10. The present exercise asks to you apply a similar analysis to a black hole with $a/M = (3/2)^{1/2}$.

A. Repeat Item A of exercise 10 for $a/M = (3/2)^{1/2}$.

B. Ditto for Item B of exercise 10.

C. In Section 17.8 we found from equations (77) through (79) the surprising result that local ring frames can exist between the Cauchy horizon and the singularity. Use the 3D light cone diagram of Item C to show how once a stone crosses the Cauchy horizon, in principle—*that is, without any mathematical analysis of particular orbits*—the stone is *not necessarily* dragged further towards smaller r-values and into the singularity, but can remain in circular orbits.

D. Knowing what you know from the present chapter, how many different circular orbits can there be for a free stone inside the Cauchy horizon? Why is your answer to this Item D different from your answer to Item C?

13. Limiting values of constants and variables at the horizons

Derive expressions (22) through (27) in Box 2.

14. Stable circular orbits at $r/M = 9$ for maximum-spin black hole

Equations (68) and (70) tell us that stable orbits come in pairs (prograde Types 1 and 2 always occur together, and retrograde Types 3 and 4 also always occur together). Figure 15 shows that for a maximum-spin black hole, $r/M = 9$ is on the boundary between region G (where one prograde pair of stable circular orbits exist) and region H (where two pairs of stable circular orbits exist—one prograde, one retrograde).

This argument implies that $r/M = 9$ is the innermost stable circular orbit (ISCO) for retrograde (Types 3 and 4) orbits, but just an ordinary stable circular orbit for prograde (Types 1 and 2) orbits.

Use equations (31) through (38) for L/m and E/m and equation (16) for $V_{\mathrm{L}}^{\pm}/m$ to verify the conclusion in the preceding paragraph.

15. Orbiting in the direction of rotation of the black hole

Out of the four types of circular orbits discussed in this chapter, in which type(s) does the stone actually orbit in the direction that the black hole rotates? Does this question have a coordinate-free meaning?

16. Circle points for the maximum-spin black hole

Table 2 shows the r/M and E/m values of circular orbits for a black hole with $a/M = (3/2)^{1/2}$ and a stone with a map angular momentum $L/(mM) = 5$. How were these numerical values calculated? Construct a similar table for stone moving with the same map angular momentum around a spinning black hole with $a/M = 1$. Display the effective potentials $V_{\mathrm{L}}^{\pm}(r)$ for this case in a plot similar to Figure 1.

17. Possible orbits and their orbit parameters for a given a/M and r/M

Use equations (31) through (38) and equations (43) through (50) to find all possible types of circular orbits and their values of $L/(mM)$, E/m, $dT/d\tau$, and $d\Phi/dT$, for black hole spin $a/M = (3/2)^{1/2}$ at the following r-coordinates.

(a) $r/M = 22.76$. Check your result in Figure 1.
(b) $r/M = 19.87$. Check your result in Figure 1.
(c) $r/M = 4$. Check your result in Figure 3.
(d) $r/M = 0.4475$. Check your result in Figure 3.

Download file name: Ch18CircularOrbitsSpinBH170905v3.pdf

Chapter 19. Orbiting the Spinning Black Hole

- *As my spaceship approaches the spinning black hole, how do I insert it into an initial circular orbit?*
- *Which of the four Types of circular orbits at a given r do I choose?*
- *How can I transfer from one circular orbit to a closer one?*
- *Can I put a probe into a circular orbit inside the Cauchy horizon?*
- *Can I harness the black hole spin to "throw" stones (or photons) out to a great distance?*
- *At what r-value do tides in a circular orbit become lethal?*

Download file name: Ch19OrbitingTheSpinningBH171115v2.pdf

CHAPTER 19 Orbiting the Spinning Black Hole

Edmund Bertschinger & Edwin F. Taylor *

Einstein was not only skeptical, he was actively hostile, to the idea of black holes. He thought the black hole solution was a blemish to be removed from the theory by a better mathematical formulation, not a consequence to be tested by observation. He never expressed the slightest enthusiasm for black holes, either as a concept or a physical possibility.

—Freeman Dyson

19.1 ■ EXPLORE THE SPINNING BLACK HOLE

The sequence of orbits in our exploration plan

Observe the black hole from circular orbits.

Chapter 18 described circular orbits of a free stone around a spinning black hole. The present chapter shows how the captain of an approaching spaceship can insert her ship into an initial circular orbit at arbitrarily-chosen $r = 20M$, then transfer to circular orbits of progressively smaller r-value to provide closer looks at the black hole.

The exploration program for the spinning black hole is similar to that for the non-spinning black hole (Chapter 9) but in some ways strikingly different. In particular, the spinning black hole may be monitored from unstable circular orbits *inside* the Cauchy horizon (Step 3 in the following exploration program).

EXPLORATION PROGRAM FOR THE SPINNING BLACK HOLE [$a/M = (3/4)^{1/2}$]

Exploration program

Step 1. Insert the approaching spaceship into an initial stable circular orbit at $r = 20M$.

Step 2. Transfer an observation probe from this initial circular orbit to the innermost stable circular orbit (ISCO) at $r_{\rm ISCO} = 2.5373M$.

Step 3. Transfer the probe from $r_{\rm ISCO}$ into either of two unstable circular orbits *inside the Cauchy horizon.*

Step 4. Tip the probe off the unstable circular orbit so that it spirals into the singularity.

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

Box 1. Useful Relations for the Spinning Black Hole

This box repeats Box 1 in Section 17.8.

Static limit from Section 17.3:

$$r_{\mathrm{S}} = 2M \qquad (1)$$

Reduced circumference from Section 17.2:

$$R^2 \equiv r^2 + a^2 + \frac{2Ma^2}{r} \qquad (2)$$

Horizon function from Section 17.3:

$$H^2 \equiv \frac{1}{r^2}\left(r^2 - 2Mr + a^2\right) \qquad (3)$$

$$= \frac{1}{r^2}\left(r - r_{\mathrm{EH}}\right)\left(r - r_{\mathrm{CH}}\right) \qquad (4)$$

where r_{EH} and r_{CH} are r-values of the event and Cauchy horizons, respectively, from Section 17.3.

$$\frac{r_{\mathrm{EH}}}{M} \equiv 1 + \left(1 - \frac{a^2}{M^2}\right)^{1/2} \quad \text{(event horizon)} \qquad (5)$$

$$\frac{r_{\mathrm{CH}}}{M} \equiv 1 - \left(1 - \frac{a^2}{M^2}\right)^{1/2} \quad \text{(Cauchy horizon)} \qquad (6)$$

Ring omega from Section 17.3:

$$\omega \equiv \frac{2Ma}{rR^2} \qquad (7)$$

An equivalence from Section 17.3:

$$1 - \frac{2M}{r} + R^2\omega^2 = \left(\frac{rH}{R}\right)^2 \qquad (8)$$

Definition of α from Section 17.7:

$$\alpha \equiv \arcsin\left[\left(\frac{2M}{r}\right)^{1/2}\frac{a}{rH}\right] \quad (0 \le \alpha \le \pi/2) \qquad (9)$$

Definition of β from Section 17.8:

$$\beta \equiv \left(\frac{2M}{r}\right)^{1/2}\left(\frac{r^2 + a^2}{R^2}\right)^{1/2} \qquad (10)$$

Compare with the non-spinning black hole.

This chapter does not contain queries that ask you to "Compare these results with those for a non-spinning black hole." Nevertheless, we recommend that you do so automatically: Run your finger down the text of Chapter 9 as you read Chapter 19. The similarities are as fascinating as the differences!

Box 1 reminds us of useful relations for the spinning black hole, taken from earlier chapters. Box 2 clarifies what it means to plot the orbits of a stone.

REVIEW FROM CHAPTER 18: KINDS OF MOTION

Classify the motion of a stone by how its Doran global coordinates change during that motion. Section 18.5 defined prograde/retrograde motion and also forward/backward motion as follows:

Kinds of motion

- **Prograde motion** has $d\Phi/d\tau > 0$.
- **Retrograde motion** has $d\Phi/d\tau < 0$.
- **Forward motion** has $dT/d\tau > 0$.
- **Backward motion** has $dT/d\tau < 0$.

Recall that the raindrop (released from rest far from the black hole) falls with $d\Phi/d\tau = 0$ (Section 17.4). Raindrop motion provides the dividing line between prograde and retrograde motion.

Wristwatch time τ runs forward along the worldline of a stone. In backward motion ($dT/d\tau < 0$), map T runs backward along the stone's worldline—a reminder that map coordinate T is *not* measured time.

Sections 18.4 and 18.5 described four Types of circular orbits:

REVIEW: FOUR TYPES OF CIRCULAR ORBITS

Types of circular orbits

- **Type 1 Circular:** $E/m > 0$, $L/m > 0$, forward, prograde, with $E/m = V_{\rm L}^{+}$
- **Type 2 Circular:** $E/m < 0$, $L/m < 0$, backward, prograde, with $E/m = V_{\rm L}^{-}$
- **Type 3 Circular:** $E/m < 0$, $L/m > 0$, backward, retrograde, with $E/m = V_{\rm L}^{+}$
- **Type 4 Circular:** $E/m > 0$, $L/m < 0$, forward, retrograde, with $E/m = V_{\rm L}^{-}$

Note: In Type 1 and 2 orbits, the signs of E/m and L/m apply outside the event horizon. Inside the Cauchy horizon the signs may be different. (Table 3, Section 18.5).

In addition to circular orbits, the present chapter studies a series of transfer orbits that take us from one circular orbit to another.

Comment 1. Follow the Figures
This chapter continues, even increases, the heavy use of algebra, but it has a simple central theme: how to insert a spaceship into an outer circular orbit, then how to transfer from this outer circular orbit to inner circular orbits. Pay attention to the figures, which illustrate and summarize these transitions.

19.2 ■ INSERT APPROACHING SPACESHIP INTO AN INITIAL CIRCULAR ORBIT

Approach from far away and enter an initial circular orbit.

Insert incoming spaceship into initial circular orbit.

A spaceship approaches the spinning black hole from a great distance. The captain chooses $r = 20M$ for her initial circular orbit, near enough to the spinning black hole to begin observations. How does she manage this insertion? Analyze the following method: While still far from the black hole, the captain uses speed- and direction-changing rocket thrusts to put the spaceship into an unpowered orbit whose minimum r-value matches that of the desired initial circular orbit (Figure 1). At that minimum, when the spaceship moves tangentially for an instant, the captain fires a tangential rocket to slow down the spaceship to the speed in a stable circular orbit at that r-value.

Comment 2. Both unpowered spaceship and unpowered probe = stone
In the present chapter, our spaceship or probe sometimes blasts its rockets, sometimes remains unpowered. The unpowered spaceship or probe moves as a free stone moves. It is important not to confuse powered and unpowered motions of "spaceship" or "probe."

Insertion orbit

What values of map E and L lead a distant incoming unpowered spaceship later to move tangentially for an instant at the chosen $r = 20M$ (Figure 1)? To find out, manipulate equations (15) and (16) in Section 18.2 and introduce the condition $dr/d\tau = 0$ (tangential motion), so that $E = V_{\rm L}^{\pm}(r)$. The result is:

Box 2. How do we plot orbits of a stone?

This chapter asks and answers two questions about a stone's orbit: **Question 1:** How do we calculate the orbit of a stone? **Question 2:** How do we plot that orbit? Question 2 is a central subject of this chapter.

What is the precise definition of the stone's orbit? Technically **An orbit is a parameterized worldline expressed in global coordinates**. Huh, what does that mean? Here's an example: Go for a walk, during which you glance occasionally at your wristwatch. Later you announce, "When I arrived at the corner of Main and Pleasant Streets, my wristwatch read seven." Wristwatch time is—literally!—the *parameter* by which you report on your walk. Of course, the wristwatch time between two locations depends on the path you choose between them. If you go by way of Lester Street (for example), your wristwatch will record a longer time of, say, ten minutes.

Every orbit plotted in this book is a parameterized worldline expressed in global coordinates. In Doran coordinates, for example, three functions $T(\tau)$, $r(\tau)$, and $\Phi(\tau)$ give a full description of the stone's orbit, parameterized by its wristwatch time τ. On our two-dimensional page we plot the orbit on a two-dimensional slice, typically the $[r, \Phi]$ slice.

Restate the two questions that began this box:
Question 1: How do we obtain these orbit functions?
Question 2: How do we plot these orbit functions?
Answer Question 2 first.

Question 2: For reasons discussed in Section 19.5, we translate r and Φ into Cartesian-like global coordinates $X \equiv (r^2+a^2)^{1/2}\cos\Phi$ and $Y \equiv (r^2+a^2)^{1/2}\sin\Phi$—equations (40) and (41). In these global coordinates the black hole singularity $r = 0$ is a ring with $(X^2 + Y^2)^{1/2} = a$. The X and Y coordinates of an orbit are plotted as if they were Cartesian (Figure 10). (Indeed, behind the scenes we plot every orbit in this book using similar Cartesian-like coordinates, including those plotted by the software GRorbits.)

Back to Question 1, how to obtain functions $r(\tau)$ and $\Phi(\tau)$, is easy to answer in principle but more difficult in practice. In principle, we simply integrate equations of motion for $dr/d\tau$ and $d\Phi/d\tau$—equations (4) and (15) in Section 18.2. In practice, the $\pm$ signs in these equations make them difficult to solve. Instead, our plotting programs—including GRorbits—use different equations of motion defined for $\tau(T), r(T), \Phi(T)$, with global T-coordinate as the parameter (Section 20.1). These equations do not contain $\pm$ signs and are valid provided $d\tau/dT \neq 0$. *Full disclosure:* We do not display these equations, which are based on so-called "Hamiltonian methods."

?

Objection 1. *Fraud! You just admitted that neither the calculation of orbits nor their plots in this book use the equations of motion you give us. Stop lying to us.*

!

You're right—and wrong. A wise manager lays out the general strategy to reach a goal, shows an in-principle path to that goal, then delegates to others completion of the project. Shall we take a side trip into "Hamiltonian methods" (whatever that means) to calculate orbits from equations of motion that do not have plus-or-minus signs? We choose not to. Instead we plunge ahead with our story of diving into the spinning black hole.

$$\frac{E}{m} = \omega\frac{L}{m} \pm \frac{rH}{R}\left(1 + \frac{L^2}{m^2R^2}\right)^{1/2} \qquad \text{(tangential)} \tag{11}$$

Here the $\pm$ on the right side is the same as the superscript on $V_L^{\pm}(r)$. Write (11) as a quadratic equation in L/m:

$$\left[\omega^2 - \left(\frac{rH}{R^2}\right)^2\right]\left(\frac{L}{m}\right)^2 - 2\omega\frac{E}{m}\left(\frac{L}{m}\right) + \left[\left(\frac{E}{m}\right)^2 - \left(\frac{rH}{R}\right)^2\right] = 0 \tag{12}$$

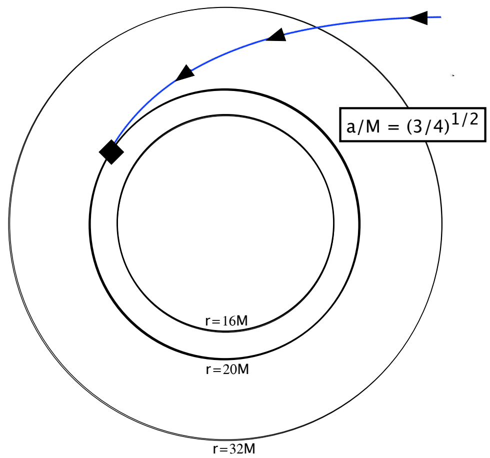


FIGURE 1 An insertion orbit with instantaneous tangential motion at $r = 20M$. At that instant the spaceship fires a tangential rocket burst that reduces the local ring velocity to that for a Type 1 circular orbit there (Figure 2).

This quadratic equation is in the standard form:

$$A\left(\frac{L}{m}\right)^2 - 2B\left(\frac{L}{m}\right) + C = 0 \qquad (13)$$

with the standard solution:

$$\frac{L}{m} = \frac{B \pm (B^2 - AC)^{1/2}}{A} \qquad (14)$$

Use (8) to simplify coefficient A:

$$A \equiv \omega^2 - \left(\frac{rH}{R^2}\right)^2 = -\frac{1}{R^2}\left(1 - \frac{2M}{r}\right) \qquad (15)$$

Show that:

$$B^2 - AC = \left(\frac{rH}{R^2}\right)^2 \left[\left(\frac{E}{m}\right)^2 - \left(1 - \frac{2M}{r}\right)\right] \qquad (16)$$

With these substitutions, (14) yields the solution:

$$\boxed{\frac{L}{m} = \frac{-\omega R^2\left(\dfrac{E}{m}\right) \pm rH\left[\left(\dfrac{E}{m}\right)^2 - \left(1 - \dfrac{2M}{r}\right)\right]^{1/2}}{1 - \dfrac{2M}{r}} \qquad \text{(tangential)} \qquad (17)}$$

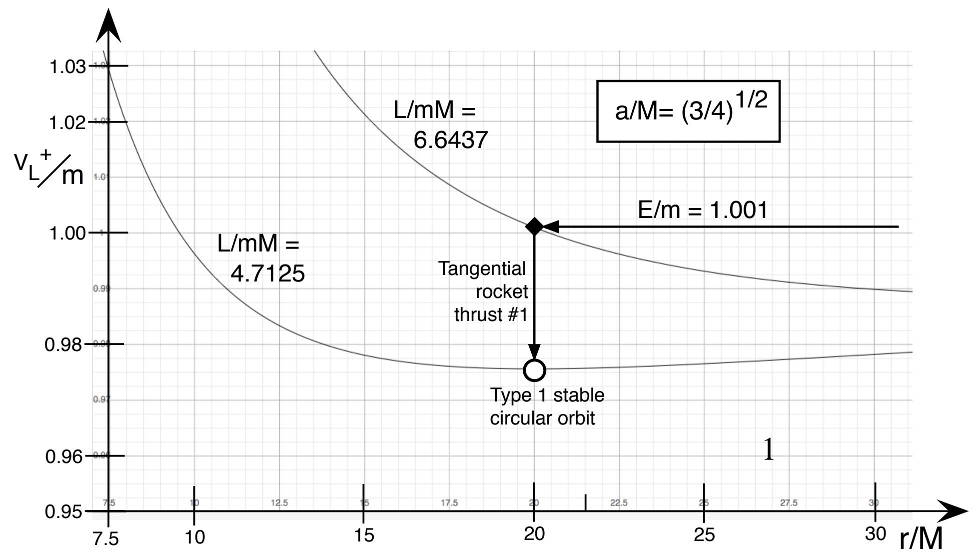


FIGURE 2 At the instant when the incoming spaceship moves tangentially at the turning point $r = 20M$ (Figure 1), it fires tangential rocket thrust #1 to change its map energy and map angular momentum to those for a Type 1 stable circular orbit at that r.

You choose the value of r; then equation (17) gives you the value of L/m for which the free stone moves tangentially at this r. This equation is valid at all turning points and everywhere along a circular orbit.

We want to place the incoming spaceship into a circular orbit at $r = 20M$. But Section 18.4 tells us that there are *four* Types of circular orbits at every $r > r_{\rm ISCO}$. Which of these four circular orbit Types do we choose for our incoming spaceship?

Choose Type 1 at $r = 20M$.

We choose the map energy of a stone to be positive, while map angular momentum can be either positive or negative. This limits circular orbits to either Type 1 or Type 4. Figure 4 in Section 18.4 shows the Type 1 circular orbit at $r = 4M$ to be stable; similarly for the Type 1 orbit at $r = 20M$. In contrast, the Type 4 circular orbit is unstable—too dangerous for our astronauts. Therefore we choose the Type 1 (stable) circular orbit.

Comment 3. Turning point symbols, a reminder
Figures in this chapter use *turning point* symbols from Definition 2 and Figure 1 in Section 18.3: The little open circle lies at the r-value of a *stable* circular orbit. The little filled circle lies at the r-value of an *unstable* circular orbit. The little half-filled circle lies at the r value of the half-stable innermost stable circular orbit, ISCO. Finally, the little filled diamond lies at a *bounce point*, where an incoming free stone "bounces" off the effective potential, reversing its r-component of motion.

Insertion orbit tangential at $r = 20M$.

We want the insertion orbit to be tangential at the instant when the unpowered spaceship reaches $r = 20M$. What map values E and L of the distant spaceship lead to its later tangential motion at $r = 20M$? We

TABLE 19.1 Numerical values at $r = 20M$ and $r = r_{\text{ISCO}}$ for $a/M = (3/4)^{1/2}$

Values of	at $r = 20M$	at $r_{\text{ISCO}} = 2.537\ 331\ 95M$
R^2	$400.825\ 000M^2$	$7.779\ 225\ 58M^2$
R	$20.020\ 614\ 4M$	$2.789\ 126\ 31M$
$2Ma/r$	$0.086\ 602\ 540\ 4M$	$0.682\ 626\ 807M$
$\omega = 2Ma/(rR^2)$	$2.160\ 607\ 26 \times 10^{-4}M^{-1}$	$0.087\ 749\ 969\ 5M^{-1}$
rH	$18.993\ 419\ 9M$	$1.453\ 750\ 16M$
rH/R	0.948 693 158	0.521 220 626
$1 - (2M/r)$	0.9	0.211 770 458
$(L/m)_{\text{insert}}$	$6.643\ 724\ 95M$	———
$(E/m)_{\text{insert}}$	1.001	———
$v_{x,\text{ring,insert}}$	0.314 955 478	———
$(L/m)_{\text{Type 1}}$	$4.712\ 495\ 61M$	$2.208\ 530\ 40M$
$(E/m)_{\text{Type 1}}$	0.975 638 130	0.858 636 605
$v_{x,\text{ring,Type 1}}$	0.229 120 545	0.620 784 509
$(L/m)_{\text{transfer}}$	$2.678\ 687\ 02M$	$2.678\ 687\ 02M$
$(E/m)_{\text{transfer}}$	0.957 725 762	0.957 725 762
$v_{x,\text{ring,transfer}}$	0.132 614 709	0.692 683 307

arbitrarily choose incoming spaceship map energy $E/m = 1.001$, as we did in Section 9.2. With this choice, equation (17) yields the value of $(L/m)_{\text{insert}}$ for the insertion orbit at $r = 20M$. Add this value to Table 19.1.

Subscripts in Table 19.1

DEFINITION 1. Subscripts in Table 19.1

Here are definitions of the subscripts in Table 19.1. All of them describe the motion of a free stone or an unpowered spaceship or probe.

insert: Quantities for a stone approaching from a great distance that leads it to move tangentially at the given r.

Type: Quantities for a stone in a circular orbit of that Type at the given r (Section 18.4).

transfer: Quantities for a stone in a transfer orbit between tangential motion at both of the given values of r.

ring: Value of the quantity measured in the local inertial ring frame at that r.

Comment 4. Significant digits

In this chapter we analyze several unstable (knife-edge) circular orbits. Interactive software such as GRorbits requires accurate inputs to display the orbit of an unpowered probe that stays in an unstable circular orbit for more than one revolution. To avoid clutter, we relegate to tables most numbers that have many significant digits.

Insert into circular orbit

The insertion maneuver shown in Figures 1 and 2 brings the unpowered spaceship to instantaneous tangential motion at $r = 20M$. Before it can move outward again, a tangential rocket thrust slows it down to the orbital speed of a stable Type 1 circular orbit at that r-value. What change in tangential velocity must this rocket thrust provide? To answer this question, we must choose a local inertial frame in which to measure tangential velocities. Sections 17.5 through 17.8 describe *four* different local inertial frames. Which one should we choose? Figure 5 in Section 17.5 shows that of our four local inertial frames, only the ring frame exists both outside the event horizon and inside the Cauchy horizon—locations where circular orbits also exist. Therefore we choose to measure the tangential velocity in the local ring frame.

The *ring frame* is the local rest frame of a ring rider who circles the black hole with map angular speed:

$$\frac{d\Phi}{dT} = \omega \equiv \frac{2Ma}{rR^2} \tag{18}$$

where Box 1 defines both ω and R^2. As with all local inertial frames, we define the ring frame so that local coordinate increments satisfy the flat spacetime metric,

$$\Delta\tau^2 \approx \Delta t_{\text{ring}}^2 - \Delta x_{\text{ring}}^2 - \Delta y_{\text{ring}}^2 \tag{19}$$

Doran metric

where each local coordinate difference equals a linear combination of global coordinate increments appearing in the global metric. The approximate Doran metric becomes:

$$\Delta\tau^2 \approx \left(1 - \frac{2M}{\bar{r}}\right)\Delta T^2 - 2\left(\frac{2M\bar{r}}{\bar{r}^2 + a^2}\right)^{1/2}\Delta T\,\Delta r + \frac{4Ma}{\bar{r}}\Delta T\,\Delta\Phi$$
$$-\frac{\bar{r}^2\,\Delta r^2}{\bar{r}^2 + a^2} + 2a\left(\frac{2M\bar{r}}{\bar{r}^2 + a^2}\right)^{1/2}\Delta r\,\Delta\Phi - \bar{R}^2\Delta\Phi^2\ . \tag{20}$$

Ring frame coordinates

We define ring frame coordinates by equations (77) to (80) of Section 17.8:

$$\Delta t_{\text{ring}} = \frac{\bar{r}H(\bar{r})}{R(\bar{r})}\Delta T - \frac{\beta(\bar{r})}{H(\bar{r})}\Delta r \tag{21}$$

$$\Delta y_{\text{ring}} = \frac{\Delta r}{H(\bar{r})} \tag{22}$$

$$\Delta x_{\text{ring}} = R(\bar{r})\left[\Delta\Phi - \omega(\bar{r})\Delta T\right] - \frac{\bar{r}\omega(\bar{r})}{\beta(\bar{r})}\Delta r \tag{23}$$

where Box 1 defines β. You can substitute equations (21) through (23) into (19) to verify that the result matches (20).

To complete the insertion of the incoming spaceship, we need to find the value of the rocket thrust required to put the ship into the Type 1 circular orbit at $r = 20M$. Appendix B has the general results. Here we use equation (94) for tangential motion.

$$v_{x,\text{ring}} = \frac{p_{x,\text{ring}}}{E_{\text{ring}}} = \frac{rH}{R^2}\left(\frac{L}{E - \omega L}\right) \tag{24}$$

TABLE 19.2 Rocket Thrusts in Instantaneous Initial Rest Frames (IIRF)

Thrust	at $r =$	$\Delta v_{x,\text{IIRF}}$	Description	$m_{\text{final}}/m_{\text{initial}}$
#1	$20M$	$\Delta v_{x,\text{IIRF1}} = -0.092\ 510\ 766\ 2$	into circular orbit	0.9113976
#2	$20M$	$\Delta v_{x,\text{IIRF2}} = -0.099\ 530\ 031\ 6$	into transfer orbit	0.9049635
#3	$2.5373M$	$\Delta v_{x,\text{IIRF3}} = -0.126\ 139\ 806$	into ISCO	0.8808964
#4	$2.5373M$	$\Delta v_{x,\text{IIRF4}} = -0.545\ 847\ 072$	into transfer to r_1	0.5420231
#5	$2.5373M$	$\Delta v_{x,\text{IIRF5}} = -0.402\ 281\ 976$	into transfer to r_2	0.4743450

NOTE: A first probe uses thrusts #2, #3, and #4 to carry it from the spaceship in orbit at $r = 20M$ to orbit r_1 inside the Cauchy horizon. A second probe uses thrusts #2, #3, and #5 to carry it from the spaceship to orbit r_2 inside the Cauchy horizon.

What insertion velocity change?

What "change in velocity" must the spaceship rocket thrust provide in order to convert its "insertion velocity" at $r = 20M$ to its "circular orbit velocity" at that r-value? Quotation marks in the preceding sentence warn us that values of *velocity* and *velocity change* depend on the local inertial frame from which we measure them. We measure velocities $v_{x,\text{ring,insert}}$ and $v_{x,\text{ring,Type 1}}$ with respect to the local inertial ring frame. But what does the spaceship captain care about the ring frame? Indeed, from her point of view a stone at rest in the ring frame can be lethal! All she cares about are answers to questions like, "Do I have enough rocket fuel left to escape from this black hole?" The answer to this question depends only on the change in velocity in the spaceship's initial rest frame. In Chapter 9 we labeled the inertial frame in which the spaceship is initially at rest the **Instantaneous Initial Rest Frame (IIRF)** (Definition 2, Section 9.2). The present chapter describes five different IIRF velocity changes. Table 19.2 lists these velocity changes with the number 1 through 5 added to each subscript.

Instantaneous initial rest frame IIRF

A special relativity equation for addition of velocities—equation (54) of Section 1.13—allows us to use the two ring-frame velocities $v_{x,\text{ring,Type 1}}$ and $v_{x,\text{ring,insert}}$ to calculate the required rocket-thrust velocity change $\Delta v_{x,\text{IIRF1}}$:

$$\Delta v_{x,\text{IIRF1}} = \frac{v_{x,\text{ring,Type 1}} - v_{x,\text{ring,insert}}}{1 - v_{x,\text{ring,Type 1}} v_{x,\text{ring,insert}}} \tag{25}$$
$$= -0.092\ 510\ 766\ 2 \qquad \text{(place in circular orbit at } r = 20M\text{)}$$

shown in Figure 2. Enter the numerical result in Table 19.2. This is the rocket-thrust velocity change ($-27\ 734$ kilometers/second) that places the incoming spaceship in the Type 1 circular orbit at $r = 20M$.

QUERY 1. Why use special relativity here?

Examine equation (25). Why do we assign the special relativity roles of v_{rel}, $v_{x,\text{lab}}$, and $v_{x,\text{rocket}}$ from equation (54) of Chapter 1 to $v_{x,\text{ring,insert}}$, $v_{x,\text{ring,Type1}}$, and $\Delta v_{x,\text{IIRF1}}$ in equation (25)?

Every change in spaceship (or probe) velocity $\Delta v_{x,\text{frame}}$ with respect to a local inertial frame requires a rocket burn. Every rocket burn uses fuel that

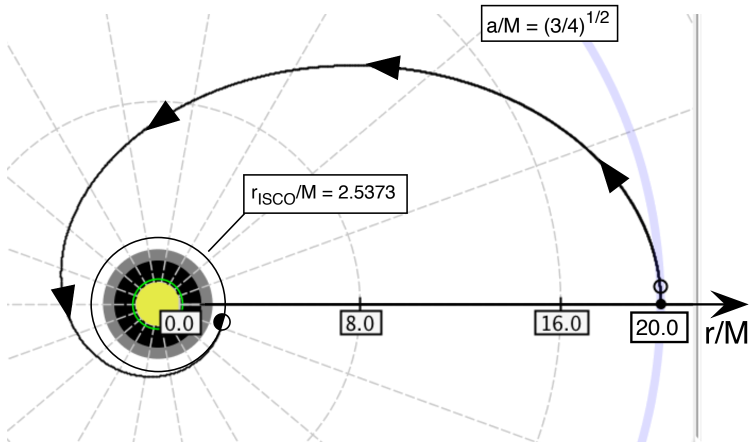


FIGURE 3 Transfer orbit in which the unpowered probe coasts from tangential motion at $r_{\rm A} = 20M$ to tangential motion at $r_{\rm B} = r_{\rm ISCO}$ and $\Phi_{\rm insert} = 350^{\rm o}$. Figure 4 indicates the required (single) tangential rocket thrust #2 to put the probe into this transfer orbit.

Use the photon rocket

changes the net mass of the spaceship or probe itself from initial mass $m_{\rm initial}$ to final mass $m_{\rm final}$. Query 2 recalls our analysis of the most efficient rocket, the so-called *photon rocket*, that combines matter and anti-matter and directs the resulting radiation out the back of the spaceship or probe. The final column of Table 19.2 lists the spaceship or probe mass ratio $m_{\rm final}/m_{\rm initial}$ for each burn described in that table.

QUERY 2. Mass ratios for transfer between circular orbits at $r = 20M$ and $r_{\rm ISCO}$. Suppose our probe uses a photon rocket defined in Exercise 2 of Section 9.8, with the resulting mass ratio:

$$\frac{m_{\rm final}}{m_{\rm initial}} = \left[\gamma + \left(\gamma^2 - 1\right)^{1/2}\right]^{-1} \qquad \text{(photon rocket)} \qquad (26)$$

where $\gamma = [1 - (\Delta v_{x,\rm IIRF})^2]^{-1/2}$ with $\Delta v_{x,\rm IIRF}$ from the third column in Table 19.2. Verify all entries in the right hand column of Table 19.2.

19.3 ■ TRANSFER FROM THE INITIAL CIRCULAR ORBIT TO ISCO, THE INNERMOST STABLE CIRCULAR ORBIT

Balanced near the abyss

The spaceship completes observations in the stable Type 1 circular orbit at $r = 20M$. The captain wants to make further observations from a smaller circular orbit. To take the entire spaceship down to this smaller orbit requires a large amount of rocket fuel. Instead, the captain launches a small probe to the inner orbit to radio observations back to the mother ship.

Transfer to circular orbit at $r = r_{\rm ISCO}$.

What r-value shall we choose for the inner circular orbit? Be bold! Take the probe all the way down to the Innermost (prograde) Stable Circular Orbit at $r_{\rm ISCO} = 2.5373M$ for the black hole with $a/M = (3/4)^{1/2}$.

Comment 5. ISCO as a limiting case
The ISCO is hazardous because it's a "half stable" circular orbit that may lead to a death spiral inward through the event horizon. In practice the inner circular orbit r-value needs to be slightly greater than $r_{\rm ISCO}$ to make it fully stable. In what follows we ignore this necessary small r-adjustment.

Figures 3 and 4 illustrate the following two-step transfer process.

ORBIT TRANSFER STEPS

Step 1: A tangential rocket thrust

Step 2: A second tangential rocket thrust

Table 19.1 shows L and E values of our initial circular orbit at $r = 20M$. To carry out Step 1, we need to find two global quantities and one local quantity: map E and L of the transfer orbit plus rocket thrust #2 to put the probe at $r = 20M$ into this transfer orbit. Calculate the global quantities E and L first.

STEP 1A: CALCULATE $(E/m)_{\rm transfer}$ AND $(L/m)_{\rm transfer}$ OF THE TRANSFER ORBIT.

Transfer from $r = 20M$ to $r_{\rm ISCO}$

Call the outer r-value of the transfer orbit $r_{\rm A}$ for Above and the inner r-value $r_{\rm B}$ for Below. At these **turning points** $E = V_{\rm L}^{\pm}$. From equation (15) in Section 18.2 for $V_{\rm L}^{+}(r)$:

$$\left(\frac{E}{m}\right)_{\rm transfer} = \frac{V_{\rm L}^{+}(r_{\rm A})}{m} = \frac{V_{\rm L}^{+}(r_{\rm B})}{m} \qquad \text{(at turning points)} \qquad (27)$$

We use the $V_{\rm L}^{+}$ effective potential because the transfer orbit takes us from one Type 1 orbit at $r_{\rm A}$ to another Type 1 orbit at $r_{\rm B}$. Substitute for $V_{\rm L}^{+}$ from equation (16) in Section 18.2:

$$\left(\frac{E}{m}\right)_{\rm transfer} = \omega_{\rm A}\left(\frac{L}{m}\right)_{\rm transfer} + \frac{r_{\rm A}H_{\rm A}}{R_{\rm A}}\left[1 + \frac{1}{R_{\rm A}^2}\left(\frac{L}{m}\right)^2_{\rm transfer}\right]^{1/2} \qquad (28)$$

$$= \omega_{\rm B}\left(\frac{L}{m}\right)_{\rm transfer} + \frac{r_{\rm B}H_{\rm B}}{R_{\rm B}}\left[1 + \frac{1}{R_{\rm B}^2}\left(\frac{L}{m}\right)^2_{\rm transfer}\right]^{1/2} \qquad (29)$$

Our task is to find the value of $(L/m)_{\rm transfer}$ that makes the right side of (28) equal to the right side of (29). When this is accomplished, (28) yields the value of $(E/m)_{\rm transfer}$.

Find L/m of transfer orbit.

The Section 19.3 analysis for $r_{\rm A} = 20M$ gives us values of the coefficients on the right side of (28), already entered in the middle column of Table 19.1.

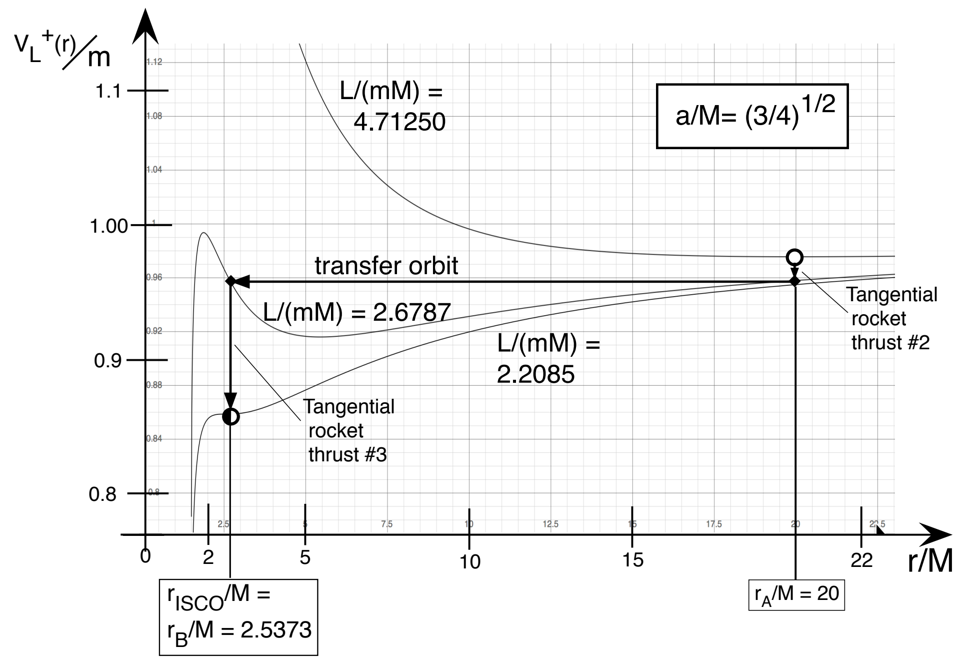


FIGURE 4 Rocket thrusts and resulting effective potential changes for transfer orbit between the stable Type 1 circular orbit at $r_{\mathrm{A}} = 20M$ and the half-stable Type 1 circular orbit at $r_{\mathrm{ISCO}} = r_{\mathrm{B}} = 2.5373M$ (Figure 3).

Now calculate coefficients on the right side of (29) using $r_{\mathrm{B}} = r_{\mathrm{ISCO}}$ and enter results in the right column of Table 19.1.

To find the value of $(L/m)_{\mathrm{transfer}}$, equate the right sides of (28) and (29). The result is a fourth order equation in $(L/m)_{\mathrm{transfer}}$, which has no straightforward algebraic solution. So we use a numerical software algorithm to find the value of $(L/m)_{\mathrm{transfer}}$ that makes equal the right sides of (28) and (29). Substitute the resulting value of $(L/m)_{\mathrm{transfer}}$ into equation (28) to find the value of $(E/m)_{\mathrm{transfer}}$ on the left side. Enter resulting values of $(L/m)_{\mathrm{transfer}}$ and $(E/m)_{\mathrm{transfer}}$ in the right-hand column of Table 19.1. Now use equation (94) to calculate values of $v_{x,\mathrm{ring,transfer}}$ at $r = 20M$ and at r_{ISCO}; enter them in Table 19.1.

STEP 1B: CALCULATE THE ROCKET THRUST VELOCITY CHANGE TO PUT THE PROBE INTO THE TRANSFER ORBIT.

IIRF2 transfer velocity change

What change in velocity must the rocket thrust provide to put the probe into the transfer orbit from $r = 20M$ to r_{ISCO}? This is our second tangential thrust to be given in an instantaneous initial rest frame IIRF, this time with the number 2 added to the subscript. From Table 19.1 and equation (54) of Section 1.13:

$$\Delta v_{x,\mathrm{IIRF2}} = \frac{v_{x,\mathrm{ring,transfer}} - v_{x,\mathrm{ring,Type\,1}}}{1 - v_{x,\mathrm{ring,Type\,1}} v_{x,\mathrm{ring,transfer}}} \quad \text{(into transfer orbit \ldots} \tag{30}$$
$$= -0.099\ 530\ 031\ 6 \qquad \text{from } r = 20M \text{ to } r_{\mathrm{ISCO}})$$

shown as tangential rocket thrust #2 in Figure 4. Enter the numerical value in Table 19.2. This rocket thrust ring velocity change (−29 838 kilometers/second) inserts the probe from the circular orbit at $r = 20M$ into the transfer orbit that takes it down to instantaneous tangential motion at r_{ISCO}.

?

Objection 2. *In Figure 3 when the probe reaches the little half-black circle, will it automatically go into the circular orbit at r_{ISCO}?*

!

No, its map angular momentum is too high. Look at Figure 4. If there is no insertion rocket thrust, the probe will simply move back and forth along the "transfer orbit" line between r_{ISCO} and $r = 20M$. Step 2 describes the rocket-thrust insertion into ISCO.

STEP 2: ROCKET THRUST TO INSERT PROBE INTO ISCO

Insert into r_{ISCO} orbit.

The probe that follows the transfer orbit from $r = 20M$ arrives for an instant at global coordinates $r = r_{\mathrm{ISCO}}$ and some value of Φ different from zero (Figure 3). At that instant it has tangential velocity $v_{x,\mathrm{ring,transfer}}$ measured in local ring coordinates, which is too high for a circular orbit at r_{ISCO}. Equation (94) gives us this tangential ring velocity, calculated from selected values in the right column of Table 19.1. Enter the result in the lower right hand position in this table.

Now we want to change this tangential transfer velocity to the velocity $v_{x,\mathrm{ring,Type\,1}}$ of the circular orbit at r_{ISCO}. Use equation (94) and enter the result in Table 19.1.

IIRF3 insertion velocity change

Again we must calculate the change in velocity the rocket thrust must provide to put the probe into the circular orbit at r_{ISCO}. We measure this third tangential change—call it $\Delta v_{x,\mathrm{IIRF3}}$ with the number 3 added to the subscript—with respect to the probe's instantaneous initial rest frame. From Table 19.1 and equation (54) of Section 1.13:

$$\Delta v_{x,\mathrm{IIRF3}} = \frac{v_{x,\mathrm{ring,Type\,1}} - v_{x,\mathrm{ring,transfer}}}{1 - v_{x,\mathrm{ring,Type\,1}} v_{x,\mathrm{ring,transfer}}} \tag{31}$$
$$= -0.126\ 139\ 806 \qquad \text{(inserts into circular orbit at } r_{\mathrm{ISCO}})$$

shown as tangential rocket thrust #3 in Figure 4. Enter the numerical result in Table 19.2. This velocity reduction (−37 815 kilometers/second) installs the probe into the Type 1 innermost stable circular orbit at r_{ISCO}.

Rocket mass ratios

Figure 4 shows that the transfer between $r = 20M$ and $r_{\rm ISCO} = 2.5373M$ requires two rocket thrusts, #2 and #3, with values listed in Table 19.2, each with its mass ratio given in the last column of that table. Thrust #2 results in mass ratio $(m_{\rm final}/m_{\rm initial})_{\#2}$. The final probe mass of thrust #2 becomes the initial probe mass of thrust #3 in the mass ratio $(m_{\rm final}/m_{\rm initial})_{\#3}$. After both thrusts take place, the net result is that the probe arrives at $r_{\rm ISCO}$ with the net mass ratio equal to the product of the two mass ratios in the right hand column of Table 19.2:

$$\left(\frac{m_{\rm final}}{m_{\rm initial}}\right)_{\#2}\left(\frac{m_{\rm final}}{m_{\rm initial}}\right)_{\#3} = 0.9049635 \times 0.8808964 = 0.7971791 \qquad (32)$$

This completes our analysis of the transfer between the initial circular orbit at $r = 20M$ and the ISCO at $r_{\rm ISCO} = 2.5373M$.

19.4 ■ ROCKET THRUSTS TO TRANSFER FROM ISCO TO CIRCULAR ORBITS INSIDE THE CAUCHY HORIZON

Teetering next to the singularity

Orbits inside the Cauchy horizon!

The probe carries out observations at $r_{\rm ISCO}$. What's next? The captain examines two alternatives: observations from one of two *unstable* circular orbits inside the Cauchy horizon. We analyze both of theses alternatives.

?

Objection 3. *Either choice is stupid! Nothing comes back from inside the event horizon, not even a radio signal. So you cannot receive a report of what happens there. You are wasting resources to place the probe in any orbit inside the event horizon.*

!

Hamlet cautions us: "There are more things in heaven and earth, Horatio, than are dreamt of in your philosophy." Chapter 21 contains surprises about what rocket-blast maneuvers inside the event horizon can accomplish. In the meantime we can still predict what the diver inside the the Cauchy horizon experiences, as we did in Section 7.8 for the (doomed!) diver inside the event horizon of the non-spinning black hole, even though neither diver can report these observations to us on the outside.

This is the first of two sections on the probe transfer from the ISCO to orbits inside the Cauchy horizon. The present section derives rocket thrusts for transfers, summarized in Table 19.2. The following Section 19.5 plots the transfer orbits themselves. Why a separate section on these orbit plots? Because close to the singularity spacetime curvature is so large, and coordinates become so stretched, that plotting any orbit requires great care.

Start with a strategic overview: To install the probe into a *stable* circular orbit (Sections 19.2 and 19.3) requires a final rocket thrust to drop the probe's map energy into the minimum of the effective potential at that r (Figures 2 and 4). In contrast, we need no such final rocket thrust to install a probe into

No final insertion rocket thrust

TABLE 19.3 Circular orbits at $r_{\rm ISCO}$, r_1, r_2 and some transfer orbits between them

Circular orbits	$r_{\rm ISCO} =$ 2.537 331 95M Type 1 outside $r_{\rm EH}$	$r_1 =$ 0.170 763 678M Type 1 inside $r_{\rm CH}$	$r_2 =$ 0.353 627 974M Type 2 inside $r_{\rm CH}$
L/m	2.208 530 40M	0.318 183 046M	0.849 088 850M
E/m	0.858 636 605	0.552 521 8506	0.619 345 540
R	2.789 126 311	3.092 447 193	2.262 034 177
ω	0.087 749 969 5 M^{-1}	1.060 621 78 M^{-1}	0.957 228 652 M^{-1}
rH/R^2	0.186 875 948M^{-1}	0.069 175 194 1M^{-1}	0.080 055 930 0M^{-1}
$v_{x,\rm ring,circle}$	0.620 784 511	0.102 350 039	-0.351 423 150
Transfer orbits	From $r_{\rm ISCO}$	to r_1	to r_2
L/m	—	0.318 183 046M	0.849 088 850M
E/m	—	0.552 521 851	0.619 345 540
$v_{x,\rm ring,transfer}$	0.113 344 665	0.102 350 039	—
$v_{x,\rm ring,transfer}$	0.291 232 033	—	-0.351 423 150

an *unstable* circular orbit such as those inside the Cauchy horizon. Why not? Because the transfer orbit is already at this maximum or minimum; the probe simply coasts onto that maximum or minimum (Figures 5 and 6). So we need only a single rocket thrust at $r_{\rm ISCO}$ to change map energy and map angular momentum to that of a circular orbit inside the Cauchy horizon. Now the details.

Transfer from $r_{\rm ISCO}$ to r_1

Transfer from $r_{\rm ISCO}$ to r_1: As a first alternative, transfer the probe from the $r_{\rm ISCO}$ orbit to the Type 1 unstable circular orbit at r_1 inside the Cauchy horizon (Figure 5). To do this, use a tangential rocket thrust that slows the probe so that it enters the transfer orbit in which it coasts directly into the unstable circular orbit at r_1.

Find L and E for transfer

How do we find values of L and E for this coasting orbit? Look again at equations (28) and (29). On the right side of (28), we know the value of $r_{\rm A}$ (the r-value of the ISCO), but we do not know the value of $(L/m)_{\rm transfer}$. On the right side of (29), we do not know values of either $r_{\rm B}$ or $(L/m)_{\rm transfer}$. Thus (29) has two unknowns, namely $(L/m)_{\rm transfer}$ and $r_{\rm B} = r_1$. However, we can find a second equation for these two unknowns, because we know that the circular orbit at $r_{\rm B}$ is Type 1, for which equation (31) in Section 18.4 takes the form

$$\left(\frac{L}{m}\right)_{\rm Type\,1} = \left(\frac{M}{r_{\rm B}}\right)^{1/2} \frac{r_{\rm B}^2 + a^2 - 2a(Mr_{\rm B})^{1/2}}{\left[r_{\rm B}^2 - 3Mr_{\rm B} + 2a(Mr_{\rm B})^{1/2}\right]^{1/2}} \;\text{(circular orbit)} \qquad (33)$$

Substitute this expression for (L/m) into equations (28) and (29), then equate the right sides of these two equations. The result is a (complicated!) equation

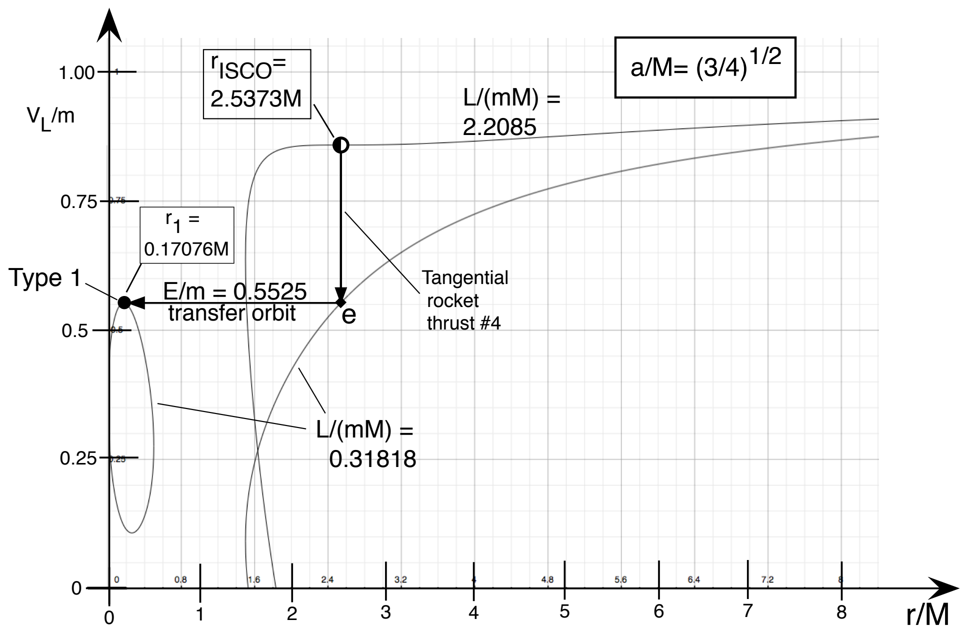


FIGURE 5 Tangential rocket thrust followed by coasting transfer orbit between ISCO (half-stable) prograde circular orbit and the Type 1 unstable circular at $r_1 = 0.17076M$, the *maximum* of the effective potential inside the Cauchy horizon.

in the single unknown $r_{\rm B}$. Again use a numerical software algorithm to find the value of $r_{\rm B}$ and enter the result in the third column of Table 19.3.

QUERY 3. Identical table entries

Look at the two right-hand columns in Table 19.3, the ones labeled r_1 and r_2. Why are so many entries for circular orbits inside the Cauchy horizon the same as the corresponding entries for the transfer orbits?

Numerical values in Table 19.3 allow us to calculate the tangential $v_{x,\rm ring,transfer}$ in (94) for the transfer orbit that starts at $r_{\rm ISCO}$ and ends at r_1 (Figure 5). The result is $v_{x,\rm ring,transfer}$ =0.113 344 264 at $r_{\rm ISCO}$, entered in Table 19.3.

IIRF4 transfer velocity change

Once again we must calculate the change in velocity the rocket thrust provides to put the probe into the transfer orbit at $r_{\rm ISCO}$. Measure this change—call it $\Delta v_{x,\rm IIRF4}$, with the number 4 added to the subscript—with respect to the instantaneous initial rest frame. From Tables 1 and 3 plus equation (54) of Section 1.13:

$$\begin{aligned}\Delta v_{x,\mathrm{IIRF4}} &= \frac{v_{x,\mathrm{ring,transfer}} - v_{x,\mathrm{ring,Type\,1}}}{1 - v_{x,\mathrm{ring,Type\,1}} v_{x,\mathrm{ring,transfer}}} \quad &\text{(into transfer orbit} \\ &= -0.545\ 847\ 072 &\text{from } r_{\mathrm{ISCO}} \text{ to } r_1) \end{aligned} \tag{34}$$

shown in Figure 5. Enter the numerical value in Table 19.2. This change in rocket velocity (−163 641 kilometers/second) puts the probe into a transfer orbit between r_{ISCO} and r_1. Figure 5 shows that the probe then coasts into the unstable circular orbit at r_1 without the need for an insertion rocket thrust.

?

Objection 4. *Unbelievable! Are you really going to demand that a human-built rocket engine change the velocity of a probe by $\Delta v = 0.545\ 847$—more than half the speed of light? Get real!*

!

Even today we use multi-stage rockets to achieve large velocity changes. Still, mass ratios to achieve a speed reduction $c/2$—and even more the overall mass ratios in Items A and B of Query 4—require the resources of an **advanced civilization**, defined as one that can achieve any technical goal not forbidden by the laws of physics. Photon rocket technology may be in our future!

Transfer from r_{ISCO} to r_2

<u>Transfer from r_{ISCO} to r_2</u>: As a second alternative, the spaceship captain transfers the probe from the ISCO to the Type 2 unstable circular orbit at r_2, a minimum of the effective potential inside the Cauchy horizon. Figure 6 shows this maneuver. A tangential rocket thrust drops the map angular momentum of the probe to $L/m = 0.849\ 088\ 850M$. Then the probe coasts inward to the minimum of the effective potential at r_2 inside the Cauchy horizon, no insertion rocket thrust required.

IIRF5 transfer velocity change

The change in velocity the rocket thrust provides puts the probe into the transfer orbit at r_{ISCO}. We measure this change—call it $\Delta v_{x,\mathrm{IIRF5}}$, with the number 5 added to the subscript—with respect to the instantaneous initial rest frame. From Tables 1 and 3 plus equation (54) of Section 1.13:

$$\begin{aligned}\Delta v_{x,\mathrm{IIRF5}} &= \frac{v_{x,\mathrm{ring,transfer}} - v_{x,\mathrm{ring,Type\,1}}}{1 - v_{x,\mathrm{ring,Type\,1}} v_{x,\mathrm{ring,transfer}}} \quad &\text{(into transfer orbit} \\ &= -0.402\ 281\ 976 &\text{from } r_{\mathrm{ISCO}} \text{ to } r_2) \end{aligned} \tag{35}$$

labeled "Tangential rocket thrust #5" in Figure 6. Enter the numerical result in Table 19.2. This change in velocity (−120 601 kilometers/second) puts the probe into a transfer orbit toward the unstable Type 2 circular orbit at r_2, shown in Figure 6. When the probe arrives there, it already has the map energy and map angular momentum of that unstable circular orbit, so does not require an insertion rocket thrust.

Recall our overall strategy: Thrust #1 takes the entire spaceship into the stable circular orbit at $r = 20M$. The spaceship then launches two separate

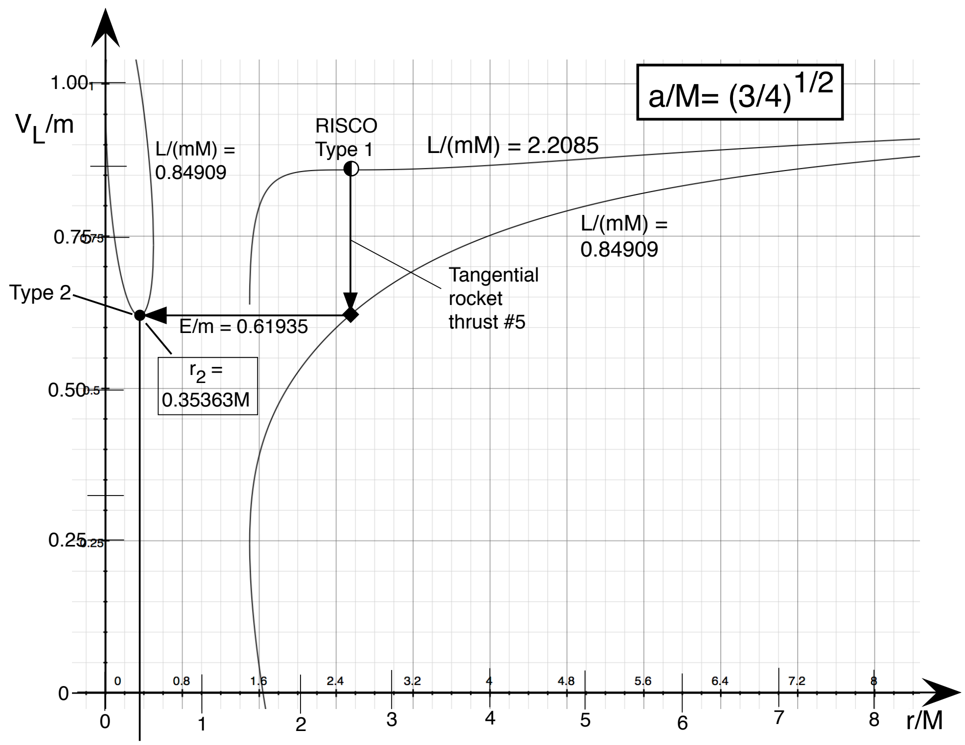


FIGURE 6 Transfer from Type 1 $r_{\rm ISCO}$ circular orbit to Type 2 unstable circular orbit at r_2, the *minimum* of the effective potential inside the Cauchy horizon.

probes. The first probe uses the sequence of thrusts #2, #3, and #4 to enter the unstable circular orbit at r_1 inside the Cauchy horizon. The second probe uses the sequence of thrusts #2, #3, and #5 to enter the unstable circular orbit at r_2 inside the Cauchy horizon.

QUERY 4. Net mass ratios for transfer between circular orbit $r_{\rm ISCO}$ and circular orbits inside the Cauchy horizon.

A. Analyze the entire sequence of thrusts #2, #3, and #4 that carry the first probe from the spaceship to the unstable circular orbit at r_1 inside the Cauchy horizon. What is the net mass ratio for this sequence of thrusts. [My answer: 0.4320895]

B. Next analyze the sequence of thrusts #2, #3, and #5 that carry the second probe from the spaceship to the unstable circular orbit at r_2 inside the Cauchy horizon. What is the net mass ratio for this sequence of thrusts. [My answer: 0.3781379].

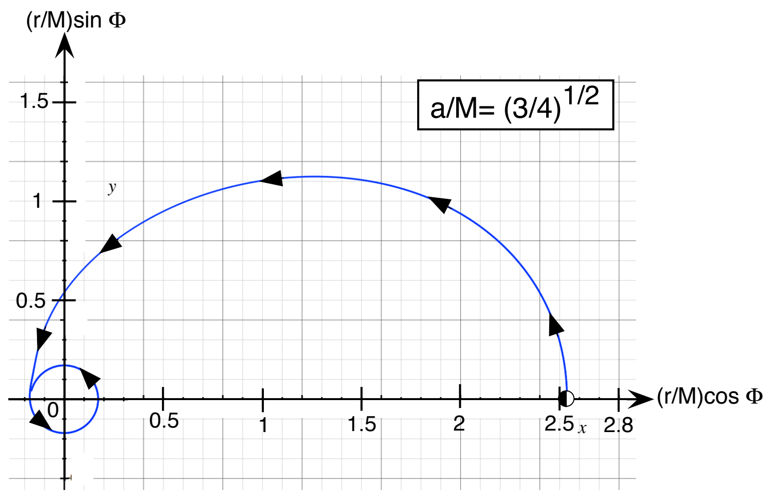


FIGURE 7 First plot of the transfer orbit between the circular ISCO and the circular orbit at $r_1 = 0.17076M$ inside the Cauchy horizon (Figure 5). This plot of $(r/M)\sin\Phi$ vs. $(r/M)\cos\Phi$ is the one we usually call an "orbit." This plot is totally correct, but near the singularity it misrepresents the geometry of spacetime.

19.5 ■ PLOTTING TRANSFER ORBITS FROM ISCO TO CIRCULAR ORBITS INSIDE THE CAUCHY HORIZON

One transfer, one failure

TWO circular orbits inside the Cauchy horizon

This section plots transfer orbits from the innermost stable circular orbit at r_{ISCO} to two different unstable circular orbits inside the Cauchy horizon: one at r_1, the maximum of an effective potential, the other at r_2, the minimum of another effective potential. For $a/M = (3/4)^{1/2}$, the circular orbit at $r_1 = 0.1707M$ lies very close to the singularity. Spacetime there is so radically warped that no global coordinate system—even Doran coordinates—gives us a picture that conforms to our everyday intuition. In what follows we do the best we can to find orbit plots that inform our intuition about this strange world.

Figure 7 shows a first orbit plot of the transfer from r_{ISCO} to r_1. This plot seems straightforward, with the singularity at $r \to 0$ as expected. But a closer look reveals that this first plot fails to correctly represent spacetime near the singularity.

To see this, look again at the Doran global metric, equation (4) in Section 17.2 when $dT = 0$, that is, on an $[r, \Phi]$ slice. Then the squared differential of measured distance $d\sigma^2$ expressed in Doran coordinates becomes:

$$d\sigma^2 = \left[\left(\frac{r^2}{r^2+a^2}\right)^{1/2} dr - a\left(\frac{2M}{r}\right)^{1/2} d\Phi\right]^2 + \left(r^2+a^2\right) d\Phi^2 \qquad (36)$$

$$0 < r < \infty, \quad 0 \le \Phi < 2\pi, \qquad dT = 0, \text{ on an } [r, \Phi] \text{ slice}$$

What happens to $d\sigma$—the differential of a measurable quantity—as $r \to 0$? The final $d\Phi^2$ term on the right side behaves reasonably; it goes to $a^2 d\Phi^2$ as $r \to 0$. In contrast, the first $d\Phi$ term blows up as $r \to 0$.

Singularity not a point.

However a little rearrangement simplifies this metric and allows us to predict a measurable result. Expand metric (36) and collect terms.

$$d\sigma^2 = \frac{r^2}{r^2+a^2}dr^2 - 2a\left(\frac{2Mr}{r^2+a^2}\right)^{1/2} dr d\Phi + \left(r^2 + a^2 + \frac{2Ma^2}{r}\right) d\Phi^2 \quad (37)$$

$$d\sigma^2 = \frac{r^2}{r^2+a^2}dr^2 - 2a\left(\frac{2Mr}{r^2+a^2}\right)^{1/2} dr d\Phi + R^2 d\Phi^2 \quad (38)$$

$$0 < r < \infty, \quad 0 \leq \Phi < 2\pi, \qquad dT = 0, \text{ on an } [r, \Phi] \text{ slice}$$

The step between (37) and (38) applies the definition of R^2 in Box 1.

Now, let r become very small and see what the singularity looks like. The first two terms in global metrics (37) and (38) become negligibly small and the third terms become:

$$d\sigma^2 \to R^2 d\Phi^2 \to a^2\left(1 + \frac{2M}{r}\right) d\Phi^2 \qquad (r \ll a \leq M) \quad (39)$$

Singularity has the topology of a circle.

As the value of r continues to decrease, the coefficient of $d\Phi^2$ *increases.* Two locations with the same small r-value but different Φ lie along a *circular* arc of length $R\Delta\Phi$. And σ, remember, is a measurable quantity. *The singularity of a spinning black hole has the topology of a circle, not a point!* In the limit of small r, we call the circular topology a **ring singularity.**

Now ask: Is there a way to plot transfer orbits so that the measurable result in (39) becomes apparent? Yes: Use R as the separation from the origin. Figure 8 shows such a plot. As we now expect from Figure 7, the probe starts moving inward but its trajectory soon deflects outward because R^2 increases as $r/(2M)$ decreases. R begins at $R = 2.7891M$ and ends at $R = 3.0924M$. Yet Figure 5 clearly shows that during this transfer the probe moves steadily inward from $r = 2.5373M$ to $r = 0.1708M$. A paradox!

To resolve this paradox, note that R is double-valued (Figure 1 in Section 17.2), and that as $r \to 0$, $R \to \infty$. *Conclusion:* Using R to plot the orbit creates a bigger problem than it solves.

Try plotting the same orbit in global map coordinates r and Φ, as in Figure 9. In this plot the global map angle Φ increases from zero at r_{ISCO} to a value that increases without limit at r_1 as the probe continues to circle there. This plot is correct but tells us nothing that we do not already know from Figure 7. And it is ugly!

New global coordinates: X and Y

So far we have failed to discover how to plot the transfer orbit between r_{ISCO} and r_1 in such a way that it correctly displays the singularity as a circle, while preserving inward motion. To accomplish this, we choose a new radial global coordinate that does not blow up as $r \to 0$, but correctly plots a circle there. This radial coordinate is $(r^2 + a^2)^{1/2}$, shown in Figure 10. The global Cartesian coordinates become:

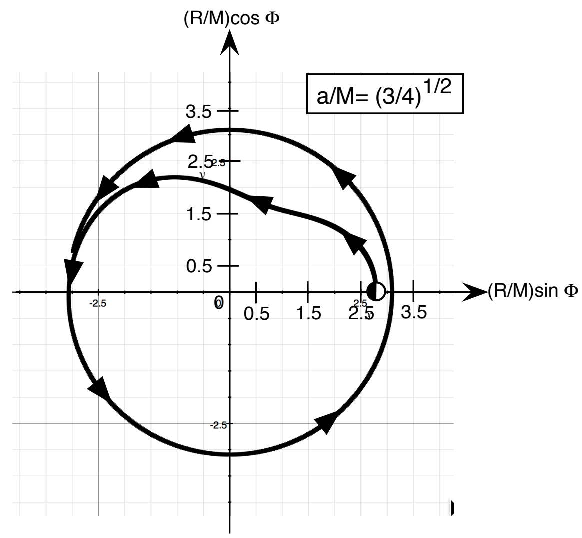


FIGURE 8 Second orbit plot of the transfer between ISCO and the circular orbit at $r_1 = 0.17076M$ inside the Cauchy horizon (Figure 5). This plot of $(R/M)\sin\Phi$ vs. $(R/M)\cos\Phi$ shows a strange coordinate behavior: The probe moves inward toward $r = 0$, yet arrives in a circular orbit of larger R than it started. See entries for R in Table 19.3.

$$X \equiv (r^2 + a^2)^{1/2} \cos\Phi \qquad \text{(global coordinates on } [r, \Phi] \text{ slice)} \qquad (40)$$

$$Y \equiv (r^2 + a^2)^{1/2} \sin\Phi \qquad (41)$$

$$X^2 + Y^2 > a^2, \quad 0 < r < \infty, \quad 0 \le \Phi < 2\pi \qquad (42)$$

Do global coordinates (40) and (41) correctly describe spacetime around a spinning black hole? They do, because they satisfy the conditions for a *good coordinate system* (Section 5.9). As we shall see, X and Y are good coordinates for much, but not all, of spacetime.

Figure 10 plots the transfer orbit in global X, Y coordinates.

GRorbits software uses $[X, Y]$ coordinates.

This book often employs the interactive software program GRorbits, which provides plots for many of our figures. Now it can be told: GRorbits makes its plots using X and Y coordinates.

?

Objection 5. *What's inside the blank disk at the center of Figure 10? The range of coordinates given in expressions (42) does not include the inside of this disk. Where can I find this inside region?*

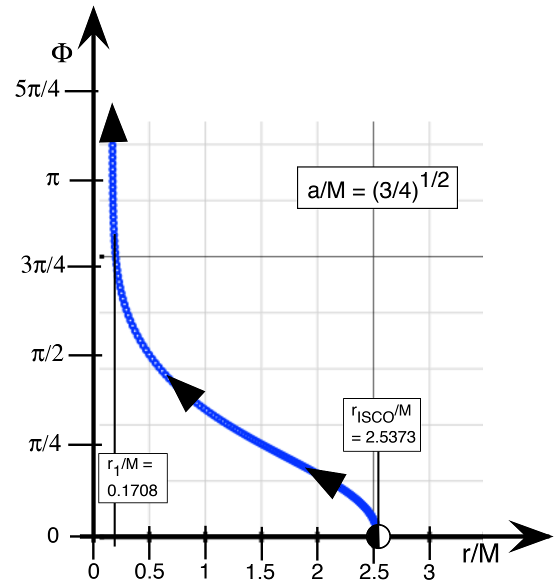


FIGURE 9 Second "orbit plot" of the transfer between ISCO and the circular orbit at $r_1 = 0.17076M$ inside the Cauchy horizon (Figure 5). This plot of Φ vs. r is not one we usually call an "orbit," but is perfectly valid as such.

!

There is no region inside the disk in the equatorial plane of the spinning black hole. Equations (40) through (42) show that points inside the ring at $r = 0$ have imaginary r-values, which is impossible.

Comment 6. Ring?

Except for gravitational waves (Chapter 16), almost all global metrics and global orbits in this book are restricted to the $[r, \Phi]$ slice. Therefore we can say nothing about the topology of any three-dimensional surface—perhaps a sphere or a cylinder—that might intersect the X, Y surface as our ring. However, advanced treatments show that the singularity of the Doran metric is confined to the $[r, \Phi]$ slice. It's a ring, not a sphere or cylinder. Moreover, we can access the central disk by traveling out of the equatorial plane to pass over or under the sigularity, as described in Chapter 21.

?

Objection 6. *The plot in which Figure: 7, 8, 9, or 10, is the correct one for the transfer orbit between circular orbits at r_{ISCO} and r_1?*

!

Every one of these orbits is equally valid and correct. Every one is distorted, because the geometry of spacetime near the spinning black hole is radically different from the flat space of our everyday lives. We select plots such as those in Figures 7 and 10 that display those features of the

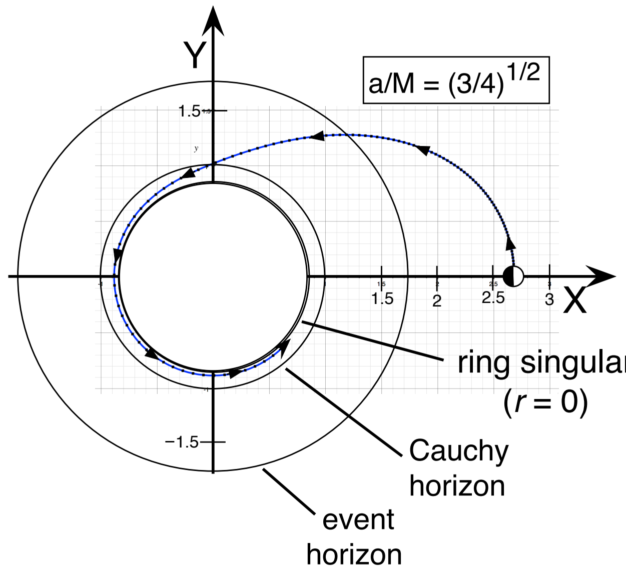


FIGURE 10 Fourth "orbit plot" of the transfer between ISCO and the circular orbit at $r_1 = 0.17076M$ inside the Cauchy horizon (Figure 5). This plot uses global coordinates X, Y.

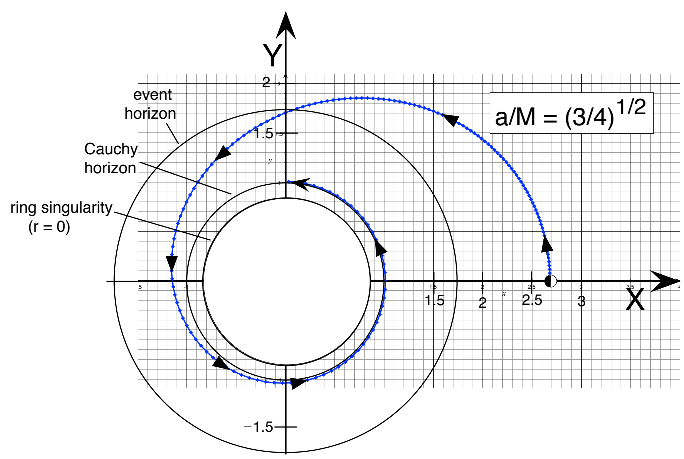


FIGURE 11 Unsuccessful attempt to plot the X, Y transfer orbit from r_{ISCO} to the minimum of the effective potential at $r_2 = 0.353\ 6M$ (Figure 6). The descending orbit shown here gets stuck at the Cauchy horizon and does not make it in to r_2. *Reason:* global Doran coordinates are not good everywhere. Chapter 21 presents new global coordinates that solve this problem.

geometry near a spinning black hole most likely to help us develop a useful intuition and ability to make predictions.

Hangup at the Cauchy horizon

So much for the transfer between $r_{\rm ISCO}$ to $r_2 = 0.17076M$. To complete our exploration inside the black hole, we also want to transfer from $r_{\rm ISCO}$ to $r_2 = 0.35363M$ at an effective potential minimum (Figure 6). This should be easier because $r_2 > r_1$ and spacetime is less warped at r_2, right? Figure 11 shows an attempt to make this plot. Oops! In this plot the probe does not move inward past the Cauchy horizon at $r_{\rm CH} = 0.5M$, shown in the figure at $X_{\rm CH} = (r_{\rm CH}^2 + a^2)^{1/2} = (1/4 + 3/4)^{1/2})M = M$.

Our history of *bad* global coordinates

Question: Why—in Figure 11—does the probe not pass inward through the Cauchy horizon? *Beginning of an answer:* We have run into this kind of problem before. Recall that the raindrop did not cross the event horizon of the non-spinning black hole when described in Schwarzschild global coordinates (Section 6.4). *Reason:* The Schwarzschild global t-coordinate is bad at the event horizon. *Solution:* Change to global rain coordinates (Section 7.5), whose T-coordinate ushers the raindrop inward through the event horizon to its doom. For the spinning black hole we started with Doran coordinates, chosen because they are good across the event horizon. But that is not enough to ensure that they are always good across the Cauchy horizon. What other coordinates are available?

Comment 7. Boyer-Lindquist t-coordinate bad at the event horizon
The exercises of Chapter 17 introduce the Boyer-Lindquist global coordinates for the spinning black hole, whose global metric is simpler than the Doran global metric. However, the Boyer-Lindquist t-coordinate is bad at the event horizon, where it increases without limit along the worldline of the raindrop.

Even Doran coordinates *bad.*

Doran global coordinates smoothly conduct our raindrop inward across the event horizon of the spinning black hole and all the way to the circular orbit at r_1, but fail to allow penetration of the Cauchy horizon on the way to the different circular orbit at r_2. Why are these two results different? Doran coordinates are okay for transfer to the circular orbit at r_1, but—it turns out—both T and Φ are bad at the Cauchy horizon for transfer to the circular orbit at r_2 (though they are good for transfer to r_1!). Figure 11 displays the problem with Φ: As $r \to r_{\rm CH}$, then $\Phi \to \infty$. To cross the Cauchy horizon we sometimes need different global coordinates. Chapter 21 explains why and shows us where new global coordinates can take us.

? **Objection 7.** *Is Nature fundamentally "bad"', or are you incompetent?*

! Nature is not bad. Mathematicians have proved that, for many curved spaces, it is impossible to cover the entire space with a single global coordinate system that is free of singularities. For these curved spaces there is no completely "good" coordinate system. The simplest example is a sphere: Earth's latitude and longitude coordinates are singular at the poles (Section 2.3), even though, for a non-spinning sphere, the poles are no different from any other points on the sphere. General relativity is difficult not because the mathematics is hard, but because we have to unlearn so many everyday assumptions that are false when applied to curved spacetime. One of these everyday false assumptions is the existence of a single global coordinate system that works everywhere.

Dispose of the probe.

To complete the Exploration Program for the Spinning Black Hole (Section 19.1), tip the probe off either unstable circular orbit inside the Cauchy horizon, so that it spirals into the singularity. Good job!

19.6 ■ ORBITING SUMMARY

Orbit descriptions

1. Effective potential plots (Figures 2, 4, 5, and 6) show us what orbits exist and help us to plot the transfer and circular orbits of an exploration program.
2. We must plot orbits using global coordinates, even though it is difficult to plot orbits in a way that is faithful to the twisted topology near the spinning black hole.
3. Sometimes one global coordinate system is not enough to cover the entire trajectory. It can take us only to the edge of a map; to go beyond that map, we need new global coordinates and a new map (Sections 2.5 and 7.5).
4. Doran global coordinates are effective across the event horizon but not necessarily through the Cauchy horizon. Also, Doran coordinates require help to show the topology of spacetime near the singularity, where a more revealing plot uses $(r^2 + a^2)^{1/2}$ rather than r or R.
5. *Preview:* Chapter 21 shows that the reason why Doran coordinates sometimes fail at the Cauchy horizon is that there are actually *two different* Cauchy horizons at the same $r_{\text{CH}} = M - (M^2 - a^2)^{1/2}$, called the *Cauchy horizon* and the *Cauchy anti-horizon.*

19.7 ■ THE PENROSE PROCESS MILKS ENERGY FROM THE SPINNING BLACK HOLE

Harness the black hole spin to hurl a stone outward.

The spinning black hole has an obvious motion that distinguishes it from the non-spinning black hole: *it spins!* Everywhere in physics, motion implies energy. Can we extract black hole spin energy for use? We know that an observer measures and extracts energy only in a local inertial frame. Can we find a local inertial frame in which the the black hole spin affects the measured energy of a stone, thus making it available for use? Roger Penrose found a way to harness the black hole spin as a local frame energy, then to send this energy to a distant observer. The present section examines what has come to be known as the **Penrose process**.

Three Penrose processes: energy conserved

Here are three physical processes in which energy does not appear to be conserved, but it is. We shall find that each process is an example of the Penrose process.

First process: A spaceship crosses inward through the static limit ($r_S = 2M$) with map energy $E/m < 1$, a value less than the minimum escape energy $E/m = 1$. Even if its rockets are not powerful enough to increase E/m above the value one, a clever ejection of ballast allows it to escape.

Second process: A distant observer launches a stone toward a black hole. Over the course of a few weeks, the observer records outgoing photons followed by a high-speed outgoing stone. He measures the combined energy of the photons and outgoing stone to exceed that of the original stone.

Third process: A uranium atom with $E/m < 1$ radioactively decays while located between the static limit and the event horizon. A distant observer measures a thorium nucleus pass outward with greater total energy than the mass of the initial uranium atom.

In all three processes, an energetic body whizzes past a distant observer. To compensate for this emitted energy, the black hole swallows a second body (ballast, photons, or decay fragments) with map energy $E < 0$. In each of these Penrose processes the black hole mass decreases, along with its spin parameter a.

Third process: uranium nucleus decays

Begin with the third process, the spontaneous decay of a uranium nucleus into an alpha particle plus a thorium nucleus. Label this process as $b \rightarrow c + d$:

$$^{235}\text{U} \rightarrow {}^{4}\text{He} + {}^{231}\text{Th} \tag{43}$$

$$b \rightarrow \quad c \quad + \quad d \qquad\qquad \text{(labels)} \tag{44}$$

Conservation of energy-momentum in ring frame

This reaction conserves the total energy and total momentum observed in every local inertial frame. We choose the ring frame. The ring frame observer verifies the following conservation statements:

$$E_{\text{ring},b} = E_{\text{ring},c} + E_{\text{ring},d} \tag{45}$$

$$p_{x,\text{ring},b} = p_{x,\text{ring},c} + p_{x,\text{ring},d} \tag{46}$$

$$p_{y,\text{ring},b} = p_{y,\text{ring},c} + p_{y,\text{ring},d} \tag{47}$$

We do not assume that the initial uranium nucleus (label b) is at rest in the ring frame; in general $E_{\text{ring},b} > m_b$ with non-zero linear momentum components $p_{x,\text{ring},b} = v_{x,\text{ring},b}E_{\text{ring},b}$ and $p_{y,\text{ring},b} = v_{y,\text{ring},b}E_{\text{ring},b}$, and similar equations apply for each of the two daughter fragments.

Equations (45) through (47), combined with equations (96) and (97) in Appendix B imply that

$$E_b = E_c + E_d \qquad\qquad \text{(map quantities)} \tag{48}$$

$$L_b = L_c + L_d \tag{49}$$

Surprise conservation of map quantities

Surprise! Even though map energy and map angular momentum are not directly measured, they are conserved in the sense that when a uranium nucleus splits in two, the total map E and total map angular momentum L are each unchanged. This remarkable fact shows how map angular momentum and map energy can act as (conserved!) proxies for measurable quantities, as Section 18.7 anticipated.

More map energy out than in

To milk energy from the spinning black hole, a successful Penrose process requires that $E_\mathrm{d} > E_\mathrm{b}$; therefore $E_\mathrm{c} < 0$ in equation (48). We shall see that this is possible only for $r < r_\mathrm{S} = 2M$, and then only if $L_\mathrm{c} < 0$ (retrograde motion). Particle d recoils with increased map energy and map angular momentum: $E_d > E_b$ and $L_d > L_b$. This is surprising, because a spontaneous decay that takes place in the ring frame always *removes* energy:

$$E_d = E_b - E_c > E_b \tag{50}$$

$$E_{\mathrm{ring},d} = E_{\mathrm{ring},b} - E_{\mathrm{ring,c}} < E_{\mathrm{ring},b} \tag{51}$$

The map energy increases while the ring frame energy decreases!

Difference milked from spin

The Penrose process takes advantage of the fact—shown in equation (91)—that ring frame energy is proportional to $E - \omega L$, not map energy E alone. Consequently, even if E increases, E_ring can decrease if L increases: $L_b - L_c = L_d$ must be sufficiently positive. The process works only if $\omega > 0$. Spacetime curvature "makes a contribution" to ring frame energy through the negative spin factor $-\omega$ in equation (91). When map angular momentum is also negative, $L < 0$, then spacetime curvature increases the ring frame energy: $E - \omega L > E$. The stone draws from spacetime curvature through the term $-\omega L$ to create a daughter nucleus (thorium) that escapes with more map energy than the initial nucleus (uranium) had when it arrived.

?

Objection 8. *How can a stone "draw from spacetime curvature" to increase its map energy? Never before have we equated curvature with energy.*

!

Curvature is *not* energy, just as map energy is not measured energy. Map energy depends on the metric, and therefore on spacetime curvature, even though measured energy is independent of the metric. Measurements are local, curvature is global. But global affects local!

Stone decays into light flash plus recoiling stone

Next look at the second process, in which a stone of mass m_b with (r, L_b, E_b) emits a light flash c with ring frame momentum components $(p_{x,\mathrm{ring},c}, p_{y,\mathrm{ring},c})$. The ring-frame energy of the light flash is $E_\mathrm{ring,c} = (p^2_{x,\mathrm{ring},c} + p^2_{y,\mathrm{ring},c})^{1/2}$. We want to find the mass, map angular momentum, and map energy of the stone—labeled d—that recoils from its backward emission of light. Note that $m_d < m_b$ because, in the rest frame of b, the light flash carries away energy. To determine the trajectory of stone d following the emission, we need both m_d and E_d because the motion depends on E_d/m_d and not on E_d alone.

Let $\phi_{\mathrm{ring},c}$ be the angle of photon momentum in the ring frame, defined so that

$$p_{x,\text{ring},c} = E_{\text{ring},c}\cos\phi_{\text{ring},c} \qquad \text{(light)} \tag{52}$$

$$p_{y,\text{ring},c} = E_{\text{ring},c}\sin\phi_{\text{ring},c} \qquad \text{(light)} \tag{53}$$

(For tangential retrograde motion, $\phi_{\text{ring},c} = \pi$.) Equations (48), (96), and (97) lead to the following equations (54) and (55). Equation (56) for m_d derives from equations (52) through (55) when substituted into the special relativity equations $m_\text{d}^2 = E_{\text{ring},d}^2 - p_{\text{ring},d}^2$ and $0 = E_{\text{ring},c}^2 - p_{\text{ring},c}^2$.

$$L_d = L_b - L_c = L_b - RE_{\text{ring},c}\cos\phi_{\text{ring},c} \tag{54}$$

$$E_d = E_b - E_c = E_b - E_{\text{ring},c}\left(\frac{rH}{R} + \omega R\cos\phi_{\text{ring},c}\right) \tag{55}$$

$$m_d = \left[m_b^2 + 2E_{\text{ring},c}\left(-E_{\text{ring},b} + p_{x,\text{ring},b}\cos\phi_{\text{ring},c} + p_{y,\text{ring},b}\sin\phi_{\text{ring},c}\right)\right]^{1/2} \tag{56}$$

Substitute equations (54) and (55) into equations (91) and (92) to give:

$$E_{\text{ring},b} = \frac{R}{rH}(E_b - \omega L_b) \tag{57}$$

$$p_{x,\text{ring},b} = \frac{L_b}{R} \tag{58}$$

$$p_{y,\text{ring},b} = \pm\left(E_{\text{ring},b}^2 - m_b^2 - p_{x,\text{ring},b}^2\right)^{1/2} \tag{59}$$

We have written (L_d, E_d, m_d) in terms of $(r, m_b, L_b, E_b, E_{\text{ring},c}, \phi_{\text{ring},c})$ and can now determine under what conditions $E_\text{d} > E_\text{b}$.

Simplest case

The simplest case to analyze is for stone b to be in a tangential prograde circular orbit, $p_{x,\text{ring},b} > 0$ and $p_{y,\text{ring},b} = 0$. In this case, E_d is maximized and m_d is minimized when $\phi_{\text{ring},c} = \pi$, that is when the light flash is emitted tangentially in the reverse (retrograde) direction.

?

Objection 9. *You said the stone was launched toward the black hole. That's not a circular orbit!*

!

The stone can be deflected into a circular orbit as it approaches the black hole, for example by encountering an accretion disk. The stone slowly loses energy to friction in the disk. After spiralling inward, it will be conveniently in a nearly circular prograde orbit inside the static limit from which the Penrose process can begin.

Fraction q of stone's mass becomes photon

We choose initial conditions $(r_b, L_b/m_b, E_b/m_b)$. The use of L_b/m_b and E_b/m_b as parameters instead of L_b and E_b generalizes the results to a stone of any mass m_b. All of the unknowns in equations (54) and (55) are now fixed except $E_{\text{ring},c}/m_b$, which we rewrite for the case in which the fraction q of the mass m_b is emitted as a photon:

$$\frac{E_{\text{ring},c}}{m_b} = \frac{E_{\text{IIRF},c}}{m_b}\left(\frac{1 - v_{x,\text{ring},b}}{1 + v_{x,\text{ring},b}}\right)^{1/2} \equiv q\left(\frac{1 - v_{x,\text{ring},b}}{1 + v_{x,\text{ring},b}}\right)^{1/2} \tag{60}$$

where $E_{\text{IIRF},c}$ is the energy of photon c in the initial rest frame (the rest frame of b), and we have used the Doppler formula of special relativity, equation (48) in Section 1.13. The ratio q on the right side of (60) can also be expressed using (56) with $\phi_{\text{ring},c} = \pi$ in terms of the final/initial mass ratio of the stone:

$$m_d^2 = m_b^2 - 2E_{\text{ring},c}\left(E_{\text{ring},b} + p_{x,\text{ring},b}\right) \tag{61}$$

$$1 - \frac{m_d^2}{m_b^2} = 2\frac{E_{\text{ring},c}}{m_b}\frac{E_{\text{ring},b}}{m_b}\left(1 + \frac{p_{x,\text{ring},b}}{E_{\text{ring},b}}\right)$$

$$\frac{1}{2}\left(1 - \frac{m_d^2}{m_b^2}\right) = \frac{E_{\text{ring},c}}{m_b}\frac{E_{\text{ring},b}}{m_b}\left(1 + v_{x,\text{ring},b}\right) \tag{62}$$

From special relativity:

$$\frac{E_{\text{ring},b}}{m_b} = \left(1 - v_{\text{ring},b}^2\right)^{-1/2} = \left(1 - v_{\text{ring},b}\right)^{-1/2}\left(1 + v_{\text{ring},b}\right)^{-1/2} \tag{63}$$

Substitute from (63) with $v_{y,\text{ring},b} = 0$ into (62) and use (60):

$$\frac{1}{2}\left(1 - \frac{m_d^2}{m_b^2}\right) = \frac{E_{\text{ring},c}}{m_b}\left(\frac{1 + v_{x,\text{ring},b}}{1 - v_{x,\text{ring},b}}\right)^{1/2} \tag{64}$$

$$= q\left(\frac{1 - v_{x,\text{ring},b}}{1 + v_{x,\text{ring},b}}\right)^{1/2}\left(\frac{1 + v_{x,\text{ring},b}}{1 - v_{x,\text{ring},b}}\right)^{1/2} = q$$

so that finally the fraction q of stone b's mass that is carried away by photon c is given by the expression:

$$q \equiv \frac{E_{\text{IIRF},c}}{m_b} = \frac{1}{2}\left(1 - \frac{m_d^2}{m_b^2}\right) \tag{65}$$

Assume that stone b is unable to escape the black hole without help, $E_b/m_b < 1$. This will be the case, for example, if the stone spirals inward in an accretion disk: the sequence of circular orbits in an accretion disk have $E/m < 1$ (Section 18.10). Can the stone escape by emitting a photon?

Answer this by evaluating E_d/m_d using (55) and (60):

$$\frac{m_d}{m_b}\left(\frac{E_d}{m_d}\right) = \frac{E_b}{m_b} - q\left(\frac{1 - v_{x,\text{ring},b}}{1 + v_{x,\text{ring},b}}\right)^{1/2}\left(\frac{rH}{R} - \omega R\right) \tag{66}$$

a similar calculation using (54) gives

$$\frac{m_d}{m_b}\left(\frac{L_d}{m_d}\right) = \frac{L_b}{m_b} + qR\left(\frac{1 - v_{x,\text{ring},b}}{1 + v_{x,\text{ring},b}}\right)^{1/2} \tag{67}$$

How large can E_d/m_d be? First, in order for the stone's map energy to increase, $E_d > E_b$, the final factor in (66) must be negative:

$$\omega R > \frac{rH}{R} \quad \text{(first condition for Penrose process)} \tag{68}$$

Definition of **ergoregion**

Equation (68) is equivalent to $r < r_S \equiv 2M$. The region $r_{EH} < r < r_S$ is called the **ergoregion**.

QUERY 5. Where is a Penrose process possible?

A. Starting from (68), show that (70) implies $r < r_S \equiv 2M$. This explains the origin of the term **ergoregion** for $r_{EH} < r < r_S$: inside the ergoregion it is possible to extract energy from a spinning black hole.

B. Show that for $r > r_S$, all particles (both stones and photons) have $E > 0$. Thus, a stone with $E < 0$ is trapped inside the ergoregion. [Hint: use (97).]

C. Show that for $r < r_S$, a retrograde photon ($\phi_{\text{ring}} = \pi$) necessarily has $E < 0$, a prograde photon ($\phi_{\text{ring}} = 0$) necessarily has $E > 0$, and a photon moving in other directions may have $E > 0$ or $E < 0$.

D. Show that as $r \to r_{EH}$, a photon with even a slight backwards direction, $\phi_{\text{ring}} = \frac{\pi}{2} + \epsilon$, has $E < 0$ and is therefore trapped.

Every Penrose process relies on the existence of particles with negative map energy. When is this possible? From (97), particle c (stone or photon) has negative map energy when

$$-v_{x,\text{ring},c} > \frac{rH}{\omega R^2} \quad \text{(second condition for Penrose process)} \tag{69}$$

Conditions for the Penrose process

The two conditions can be combined into one equation:

$$\frac{rH}{\omega R^2} < -v_{x,\text{ring},c} \leq 1 \quad \text{(Conditions for Penrose process)} \tag{70}$$

For a photon moving tangentially backward, $v_{x,\text{ring},c} = -1$ so that

$$E_c = -E_{\text{ring},c}\left(\omega R - \frac{rH}{R}\right) < 0 \quad (\text{for } r < r_S) \tag{71}$$

The emission of a negative map energy photon inside the ergoregion increases the map energy of a stone but does not guarantee that the stone will escape, which requires $E_d/m_d > 1$. Equation (66) shows that this ratio depends on several quantities: the E/m of the original stone b, the velocity of the stone in the ring frame, and the ratio of final to original mass m_d/m_b or equivalently the fraction q of the stone's original mass that is converted to retrograde-moving photons. For a stone in a given orbit, q is the only quantity that we can vary. The Penrose process is most efficient when q is maximized.

Maximal energy: annihilation

Equation (65) shows that largest possible value is $q = \frac{1}{2}$. This limit corresponds to $m_d/m_b = 0$, i.e., the stone loses all its mass! If the stone is half matter and half anti-matter, their annihilation can extract the largest possible energy from the spinning black hole. Half the mass goes into the energy of a photon emitted in the prograde direction, and half to a photon emitted in the retrograde direction. The escaping photon ("stone d") has energy

$$E_d = E_b + \frac{1}{2} m_b \left(\frac{1 - v_{x,\mathrm{ring},b}}{1 + v_{x,\mathrm{ring},b}} \right)^{1/2} \left(\omega R - \frac{rH}{R} \right) \qquad (72)$$

The energy extracted depends on the motion of the stone before it annihilates into photons. Amazingly, the map energy is largest when the stone is moving retrograde, $v_{x,\mathrm{ring},b} < 0$.

?

Objection 10. *This is crazy! How can going backwards increase the stone's map energy?*

!

You're right, this is wild, but it's true! As your intuition suggests, the map energy of stone b is *decreased* by backward motion, as shown by equation (97) with $v_{x,\mathrm{ring},b} < 0$. However, the stone's final map energy E_d also depends on the map energy of photon c. The measured energy of photon c depends on the motion of the emitter, stone b. From the Doppler formula (60), when $v_{x,\mathrm{ring},b}$ decreases, $E_{\mathrm{ring},c}$ increases and is positive: increasing stone b's velocity in the backward direction increases the energy of a photon emitted in that direction.

Now comes the real wildness: inside the ergoregion, the map energy E_c has the opposite sign to $E_{\mathrm{ring},c}$! Increasing $E_{\mathrm{ring},c}$ makes E_c more negative. The result? $E_d = E_b - E_c$ increases when $v_{x,\mathrm{ring},b}$ decreases.

Saving a crippled spaceship

Converting all one's mass to photons is a steep price to pay to escape from a black hole. Consider the first process described in the beginning of this section and ask what is the minimum energy fraction q that will allow a crippled spaceship to escape from inside the ergoregion without using rockets aside from a single thrust of a photon rocket. We seek the most frugal solution, which retains as much mass as possible.

Previous sections showed that it is very costly to transfer to circular orbits inside the Cauchy horizon (e.g., Table 19.2). Take instead the innermost stable circular orbit, the ISCO, to be the one from which we seek to return home. Taking advantage of the Penrose process requires $r_{\mathrm{ISCO}} < r_{\mathrm{S}}$. Exercise 1 below shows that this condition gives

$$\frac{a}{M} > \frac{2}{3}\sqrt{2} = 0.94281 \quad (r_{\mathrm{ISCO}} < r_{\mathrm{S}}) \qquad (73)$$

We do not know whether real black holes have such high spins (though some astronomers think so, e.g. Risalti et al., Nature, 494, 449, 2013;

doi:10.1038/nature11938). (In 1974, Kip Thorne set a theoretical limit of $a/M < 0.998$: ApJ 191, 507, 1974).

Fast spinning black hole!

As an example, take $a/M = 0.96$, for which $r_{\text{EH}} = 1.2M$. At the ISCO,

$$r_{\text{ISCO}} = 1.84\ 300\ 573\, M\ , \quad \frac{L_b}{m_b} = 1.83\ 102\ 239\, M\ ,$$
$$\frac{E_b}{m_b} = 0.798\ 919\ 307\ , \quad v_{b,x,\text{ring}} = 0.621\ 811\ 282\ . \tag{74}$$

QUERY 6. ISCO for a rapidly spinning black hole

Confirm the entries in equation (74) using equations (31), (32), and (75)–(77) of Chapter 18 and (94) below.

Given these parameters, find the minimum q for an escape orbit, by setting $E_d/m_d = 1$. Solve (66) using a numerical method and substitute into (67) to find

$$q = 0.173\ 658\ 866\ , \quad \frac{L_d}{m_d} = 2.50\ 581\ 328\, M\ . \tag{75}$$

After this photon rocket thrust, the spaceship has tangential velocity given by (24), which evaluates to

$$v_{x,\text{ring},d} = 0.735\ 812\ 177 \tag{76}$$

As in previous sections, we calculate the velocity change provided by this rocket thrust to put the spaceship into an escape orbit from r_{ISCO}. The velocity change in the instantaneous initial rest frame follows from equation (54) of Section 1.13:

$$\Delta v_{x,\text{IIRF}b} = \frac{v_{x,\text{ring},d} - v_{x,\text{ring},b}}{1 - v_{x,\text{ring},b} v_{x,\text{ring},d}} \quad \text{(from the ISCO . . .} \tag{77}$$
$$= 0.210\ 153\ 964 \qquad \text{into an escape orbit)} \tag{78}$$

Compared with the velocity changes required to transfer from the ISCO to orbits inside the Cauchy horizon of a more slowly spinning black hole (Thrusts 4 and 5 in Table 19.2), this is economical!

Figure 12 shows the effective potentials of the spaceship (stone b) before and after the rocket thrust. Compare with thrust #3 in Figure 4, which inserted the spaceship into orbit at the ISCO. Figure 12 shows the opposite: ejection from the ISCO.

The final to initial mass ratio follows from equation (26) or (65):

$$\frac{m_d}{m_b} = (1 - 2q)^{1/2} = 0.807\ 887\ 533 \tag{79}$$

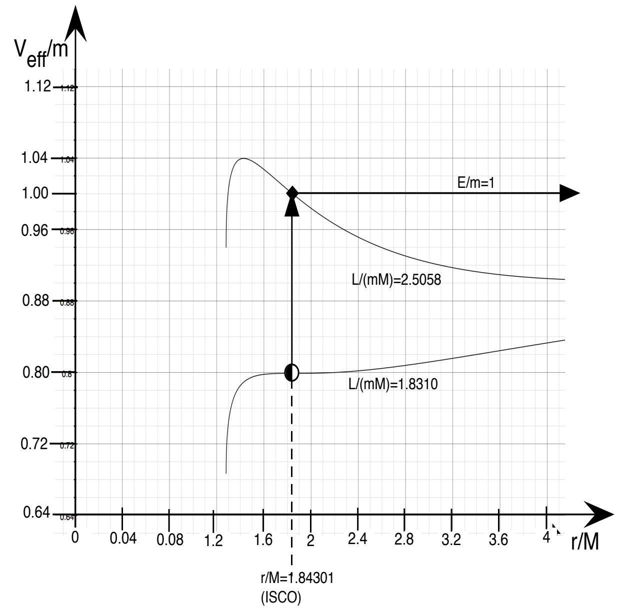


FIGURE 12 Effective potentials for a spinning black hole with $a/M = 0.96$ and two choices of the map angular momentum. The energy expended in the rocket thrust is less than the difference in map energy, providing a potential power source.

Extracting energy from the spinning black hole

This result seems almost mundane (it is comparable to thrust #3 in Table 19.2) until we compare it with E_b/m_b in equation (74), which is smaller than m_d/m_b. Although the difference is small, it reveals a crucial opportunity: the spinning black hole is an energy source!

To see this, recall that the map energy is the energy at infinity. As a stone spirals inward in an accretion disk, the photons emitted can escape to infinity, where their total energy is the difference in map energy, or $1 - (E_b/m_b) = 20.1\%$ of the original rest mass m_b. In principle, that energy is available to do work at infinity. Then, in order to escape back to infinity, a thrust must be applied that reduces m_b by a factor $1 - (m_d/m_b) = 19.2\%$. The energy difference is 0.9% of the rest mass, vastly more than the energy released by fission of uranium into thorium, and more even than is liberated by fusion of hydrogen into helium, 0.7%.

?

Objection 11. *Your numbers don't add up! You said that $q = 17.4\%$ of the initial mass goes into rocket thrust, and 80.8% is left. You're missing 1.8%!*

!

'The missing piece is the change in kinetic energy, $(\gamma - 1)m_d$, where γ is calculated using $\Delta v_{x,\mathrm{IIRF}b}$.

Justify claim: "Immense source of energy."

We can now justify the statement in Section 17.1, "The spinning black hole is an immense energy source, waiting to be tapped by an advanced civilization." Suppose you drop a stone from rest far from the black hole. Initially, $E/m = 1$ (a raindrop) and the stone enters an accretion disk. It looses map energy as it spirals inward (Section 18.8) emitting this map energy as photons. Recall that the energy of photons received at infinity is just the map energy lost, described in Sections 8.6 and 18.10 for accretion disks. By the time the stone reaches the ISCO, it has radiated a fraction $1 - E_b/m_b = 0.20108$ of its mass, as measured at infinity. In order to return to infinity, the stone's mass must decrease by a fraction $1 - m_d/m_b = 0.19211$. Thus the radiation received at infinity more than makes up for the loss of mass by the stone. We've extracted energy from the spinning black hole! This is possible because of the negative map energy of the retrograde photons emitted inside the static limit.

Comment 8. Many cycles; large extracted energy
The difference between E_b/m_b and m_d/m_b seems small, but a stone or spaceship can be reused to extract lots of energy over many cycles. Some of the energy radiated to infinity can be used to replenish the stone for another trip to the black hole. Note that the amount extracted is larger when the black hole spin is greater or the rocket thrust is applied closer to the event horizon.

?

Objection 12. *Great! We have an endless supply of energy; a perpetual motion machine! We just toss stones into the spinning black hole and program them to emit powerful laser pulses when they are inside the static limit.*

!

Sorry, this is a false hope. Here's the hitch: The photons with negative map energy fall into the black hole, where they decrease the black hole mass. We do not prove it here, but the gravitational effect of negative map energy is to decrease the gravitational field far from the black hole, exactly as if the black hole mass decreases. (Back to Newtonian physics for slow motion far from the black hole!) Still, the spinning black hole is a promising energy source for an advanced civilization.

Summary of the Penrose process for $a/M = 0.96$

Summary: Penrose process

1. Initially a stone with $E_b = m_b$ drops from rest at a great distance.
2. The stone enters an accretion disk, where it radiates 20.1% of its mass as it descends to the ISCO. The radiation—"quasar light"—travels to a great distance.
3. At the ISCO, the stone emits a photon rocket thrust; the surviving piece has mass $m_d = 0.808 m_b$
4. The surviving piece with $E_d = m_d$ escapes to a great distance, where it comes to rest and can be furbished or replaced.
5. The stone's mass decreased by 19.2%, but more than this was received at a great distance as quasar light.

The Penrose process is reminiscent of Hawking radiation (Box 5 in Section 6.6), whereby energy is also extracted from a black hole. For Hawking radiation, however, the stone that falls into the black hole is a virtual stone, a temporary entity living on time borrowed from the Heisenberg Uncertainty Principle. In contrast the Penrose process photons with negative map energy are real, not virtual. In addition, non-spinning and spinning black holes both emit Hawking radiation, while the Penrose process works only for the spinning black hole.

Penrose process compared with Hawking radiation

19.8 ■ APPENDIX A: KILLER TIDES NEAR THE SPINNING BLACK HOLE

How close is a safe orbit?

In the Appendix of Chapter 9 we saw how local inertial frames are "spaghettified" by tidal accelerations when they move near a non-spinning black hole. Equations (38) to (40) and (46) to (48) in that chapter gave the expressions for the components of the tidal acceleration Δg_{local} for local inertial frames that move along the Schwarzschild r-direction and along the Schwarzschild ϕ-direction, respectively.

In the present Appendix A we list similar expression for the *spinning* black hole. We give all the equations in *Boyer-Lindquist coordinates.* (See the Project: Boyer-Lindquist Global Coordinates at the end of Chapter 17.) In the local inertial frames the $x-$, $y-$, and $z-$ directions are along the global $\phi-$direction, $r-$direction, and perpendicular to the spinning black hole's equator, respectively.

TIDES IN THE LOCAL RING FRAME

Expressions for the the tidal accelerations around the spinning black hole are messy. Fortunately, in the equatorial plane the equations reduce to a fairly simple form. For the local ring frame:

$$\Delta g_{\text{local,y}} \approx \frac{M}{\bar{r}^3}\frac{2+Z}{1-Z}\Delta y_{\text{local}} \tag{80}$$

$$\Delta g_{\text{local,x}} \approx -\frac{M}{\bar{r}^3}\Delta x_{\text{local}} \tag{81}$$

$$\Delta g_{\text{local,z}} \approx -\frac{M}{\bar{r}^3}\frac{1+2Z}{1-Z}\Delta z_{\text{local}} \tag{82}$$

where

$$Z \equiv \frac{a^2\bar{H}\bar{r}}{\left(\bar{r}^2+a^2\right)^2} \tag{83}$$

The value of the dimensionless quantity Z always lies between 0 and 0.043 (ref: Bardeen, Press, and Teukoilsky, 1972), so the deviations for the expressions from the Schwarzschild case are small.

As an exercise, check that for $a = 0$, the three equations above reduce to Schwarzschild expressions (38) through (40) in Chapter 9.

As another exercise, check that

$$\frac{\Delta g_{\mathrm{local},x}}{\Delta x_{\mathrm{local}}} + \frac{\Delta g_{\mathrm{local},y}}{\Delta y_{\mathrm{local}}} + \frac{\Delta g_{\mathrm{local},z}}{\Delta z_{\mathrm{local}}} \approx 0 \qquad (84)$$

and compare the result with equation (45) in Chapter 9.

TIDES IN THE LOCAL ORBITER FRAME

The orbiter frame moves with speed v in the x-direction relative to the ring frame. The tidal acceleration components for the local orbiter frame are:

$$\Delta g_{\mathrm{local,y}} \approx \frac{M}{\bar{r}^3}\frac{2+Z}{1-Z}\Delta y_{\mathrm{local}} - 3\frac{M}{\bar{r}^3}\frac{\bar{H}a(\bar{r}^2+a^2)}{\bar{r}\bar{R}^2}\frac{v}{(1-v^2)^{1/2}}\Delta x_{\mathrm{local}} \qquad (85)$$

$$\Delta g_{\mathrm{local,x}} \approx -\frac{M}{\bar{r}^3(1-Z)}\left(1 - Z\frac{1+2v^2}{1-v^2}\right)\Delta x_{\mathrm{local}} \qquad (86)$$

$$-3\frac{M}{\bar{r}^3}\frac{\bar{H}a(\bar{r}^2+a^2)}{\bar{r}\bar{R}^2}\frac{v}{(1-v^2)^{1/2}}\Delta y_{\mathrm{local}}$$

$$\Delta g_{\mathrm{local,z}} \approx -\frac{M}{\bar{r}^3(1-Z)}\left(1 + Z\frac{2+v^2}{1-v^2}\right)\Delta z_{\mathrm{local}} \qquad (87)$$

Definition:
Shear terms

Note the second term on the right side of (85). It tells us that the y-component of the tidal acceleration depends on the x-coordinate too, not only the y-coordinate. Similarly, the second term on the right side of (86) tells us that the x-component of the tidal acceleration depends on the y-coordinate too, not only the x-coordinate. We call these two terms **shear** terms.

TIDES IN THE LOCAL RAIN FRAME

The tidal acceleration components in the local rain frame are:

$$\Delta g_{\mathrm{local,y}} \approx \frac{M}{\bar{r}^3}\frac{2+Z}{1-Z}\Delta y_{\mathrm{local}} - 3\frac{M}{\bar{r}^3}\frac{a(\bar{r}^2+a^2)^{3/2}}{\bar{r}^2\bar{R}^2}\frac{2M}{\bar{r}}\Delta x_{\mathrm{local}} \qquad (88)$$

$$\Delta g_{\mathrm{local,x}} \approx -\frac{M}{\bar{r}^3}\left(1 - \frac{3a^2(\bar{r}^2+a^2)}{\bar{r}^2\bar{R}^2}\left(\frac{2M}{\bar{r}}\right)^2\right)\Delta x_{\mathrm{local}} \qquad (89)$$

$$-3\frac{M}{\bar{r}^3}\frac{a(\bar{r}^2+a^2)^{3/2}}{\bar{r}^2\bar{R}^2}\frac{2M}{\bar{r}}\Delta y_{\mathrm{local}}$$

$$\Delta g_{\mathrm{local,z}} \approx -\frac{M}{\bar{r}^3}\left(1 + \frac{3a^2}{\bar{r}^2}\right)\Delta z_{\mathrm{local}} \qquad (90)$$

Again, note the presence of shear terms in the y-component and the x-component of the tidal acceleration: the second term on the right side of (88) and the second term on the right side of (89), respectively. A raindrop is not simply stretched in its local y-direction and compressed in its local x- and z-directions, but feels a sideways tension (shear) in its own rest frame too.

19.9 ■ APPENDIX B: RING FRAME ENERGY AND MOMENTUM

Measured energy and momentum

This appendix derives the map energy E and map angular momentum L of a stone from its ring frame energy E_{ring} and components of momentum $p_{x,\text{ring}}$ and $p_{y,\text{ring}}$ at a given r. The result is valid for any motion of the stone for which $H^2 > 0$, that is, everywhere except between the horizons. Start with equation (21):

$$\frac{E_{\text{ring}}}{m} \equiv \lim_{\Delta\tau\to 0}\frac{\Delta t_{\text{ring}}}{\Delta\tau} = \frac{rH}{R}\frac{dT}{d\tau} - \frac{\beta}{H}\frac{dr}{d\tau} = \frac{R}{rH}\left(\frac{E-\omega L}{m}\right) \tag{91}$$

The last step uses equation (111) of Section 17.10. Next, apply similar limits to equations (22) and (23) to obtain momentum components in the local ring frame:

$$\frac{p_{x,\text{ring}}}{m} \equiv \lim_{\Delta\tau\to 0}\frac{\Delta x_{\text{ring}}}{\Delta\tau} = R\left(\frac{d\Phi}{d\tau} - \omega\frac{dT}{d\tau}\right) - \frac{r\omega}{\beta}\frac{dr}{d\tau} = \frac{L}{mR} \tag{92}$$

$$\frac{p_{y,\text{ring}}}{m} \equiv \lim_{\Delta\tau\to 0}\frac{\Delta y_{\text{ring}}}{\Delta\tau} = \frac{1}{H}\frac{dr}{d\tau} \tag{93}$$

The velocity components in the local ring frame follow from these equations:

$$v_{x,\text{ring}} = \frac{p_{x,\text{ring}}}{E_{\text{ring}}} = \frac{rH}{R^2}\left(\frac{L}{E-\omega L}\right) \tag{94}$$

$$v_{y,\text{ring}} = \frac{p_{y,\text{ring}}}{E_{\text{ring}}} = \frac{r}{R}\left(\frac{m}{E-\omega L}\right)\frac{dr}{d\tau} \tag{95}$$

Expressions for map L and E.

Solve equations (91) and (92) for the map constants of motion in terms of the locally-measured ring energy and ring x-momentum:

$$L = Rp_{x,\text{ring}} \qquad \text{(not between horizons)} \tag{96}$$

$$E = \left(\frac{rH}{R}\right)E_{\text{ring}} + \omega R p_{x,\text{ring}} = E_{\text{ring}}\left(\frac{rH}{R} + \omega R v_{x,\text{ring}}\right) \tag{97}$$

Section 19.7 uses these two equations in the description of the Penrose process.

19.10 ■ EXERCISES

1. When the ISCO lies at the static limit

The innermost stable circular orbit (ISCO) for the non-spinning black hole lies at $r = 6M$. For the non-spinning black hole there is no distinction between prograde and retrograde orbits. For the maximum spinning black hole ($a/M = 1$), the prograde ISCO drops to $r_{\text{ISCO1}} = M$, while the retrograde orbit rises to $r_{\text{ISCO2}} = 9M$. Figure 15 in Section 18.9 plots r_{ISCO} as a function of a/M for both prograde and retrograde circular orbits.

A. What is the intermediate value of a/M at which the prograde ISCO lies at the same r-value as the static limit, $r_{\text{S}} = 2$? Use equations (75)—(77) of Section 18.8 to show that this intermediate value is $a/M = 0.94281$.

B. Verify that the numerical value of a/M in Item A is equal to $2^{3/2}/3$.

C. What is the r value of the retrograde ISCO for the value of a/M in Item A?

2. Choose incoming spaceship energy E/m for exploration program

Figure 2 shows that our explorers choose $E/m = 1.001$ for their initial energy as they start their journey from far away towards the spinning black hole. Justify this choice for the incoming value of E/m. Why should they not choose a value of E/m much larger than this? a value of E/m much closer to 1 than this? Are your reasons fundamental to general relativity theory or practical for particular spaceships and black holes?

3. Can a transfer orbit violate Kepler's second law?

Examine the second-to-last row of Table 19.3. For the transfer orbit between the circular orbit at $r_{\text{ISCO}} \approx 2.537M$ and the circular orbit at $r_1 \approx 0.170M$, the value of $v_{x,\text{ring,transfer}}$ appears to contradict Kepler's second law: The freely-moving probe appears to move faster at the larger r-coordinate than at the smaller r-coordinate. Explain how this is possible.

4. What kind of motion is raindrop motion?

Section 19.1 reviewed definitions of prograde/retrograde motions and forward/backward motions. Does *raindrop* motion provide the dividing line between forward and backward motion? between prograde and retrograde motion? Summarize your answers in a clear definition of *raindrop motion.*

5. "Size" of the ring singularity

How large is the ring singularity at r = 0?

A. Is the size of the ring singularity zero, as Figure 7 in Section 19.5 seems to show?

B. Does the *radial* size of the ring singularity equal the value of the spin-parameter a, as Figure 10 and equation (42) seem to imply?

C. Is the the ring singularity infinitely large, as Figure 8 and equation (39) seem to say? Show that from equation (39):

$$\text{circumference} = \lim_{r\to 0} 2\pi R = \lim_{r\to 0} 2\pi a \left(1 + \frac{2M}{r}\right)^{1/2} = \infty \qquad (98)$$

D. If the ring singularity is indeed infinitely large, as Item C implies, does this mean that the ring singularity extends to infinity and embraces the entire Universe? If so, why the limit $r \to 0$ in equation (98)?

E. From results of Items A through D, explain why quotes embrace the word "Size" in the title of this exercise.

6. Spacetime trajectory or spatial trajectory of the transfer orbit?

Figures 7, 9 and 10 are distorted maps, visual representations of transfer orbit, similar to the way that every flat map necessarily gives a distorted view of an arbitrary airplane route on Earth's spherical surface. But do these at least correctly depict the *spatial* trajectory of the transfer orbit? To answer this question look at the coordinates on the axes of these figures to check whether those coordinates are spacelike or timelike.

19.11 ■ REFERENCES

The original reference for the Penrose process is R. Penrose and R. M. Floyd, "Extraction of Rotational Energy from a Black Hole," Nature Physical Science. Volume 229, pages 177-179 (1971).
J. M. Bardeen, W. H. Press, and S. A. Teukolsky, "Rotating Black Holes: Locally Nonrotating Frames, Energy Extraction, and Scalar Synchrotron Radiation," The Astrophysical Journal, Volume 178, pages 347-369 (1972)

Download file name: Ch19OrbitingSpinningBH17115v2.pdf

Chapter 20. Orbits of Light around the Spinning Black Hole

- *What variety of paths does light follow around a spinning black hole?*
- *Can a spinning black hole reverse the direction of a light beam?*
- *Can a light beam go into orbit around a spinning black hole? If so, how many different orbits are available to it?*
- *What does a distant spinning black hole look like? How can I distinguish it visually from a non-spinning black hole?*

Download file name: Ch20OrbitsOfLightAroundSpinningBH170906v1.pdf

CHAPTER 20 Orbits of Light around the Spinning Black Hole

Edmund Bertschinger & Edwin F. Taylor *

Light does strange things near a spinning black hole; it misleads you about the locations and shapes of things, so you must grope along as if you are in a haunted house. Seeing is definitely not believing!

—The authors

20.1 ■ INTRODUCTION: THE PURPOSES OF LOOKING

Who cares what we see?

New questions about the spinning black hole

Chapter 11 described orbits of light around the non-spinning black hole. That earlier chapter focussed on the question, "What is the visual size of the black hole seen by a raindrop diver falling from a great distance?" The same question about the spinning black hole is of little interest today. Instead, we ask the questions:

- How can we identify a distant spinning black hole in the heavens?
- How can we measure its mass M and spin parameter a?

To answer these questions, the present chapter does use a method similar to that of these earlier chapters:

Stone's mass goes to zero.

- Start with a stone of mass m.
- Let the stone's mass go to zero.

We begin this program with a review of the stone's equations of motion in global Doran coordinates.

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

20.2 ■ DORAN GLOBAL EQUATIONS OF MOTION FOR THE STONE

Goal: Get rid of wristwatch time.

Motion of a stone

This chapter analyzes the motion of light in the equatorial plane of the spinning black hole. We derive equations of motion for light by extending equations of motion for a free stone. Begin this process with equations (15), (21), and (20) and (16) for motion of a stone in Section 18.2: . Write them in the form:

$$\frac{dr}{d\tau} = \pm\frac{R}{mr}(E - V_{\rm L}^{+})^{1/2}(E - V_{\rm L}^{-})^{1/2} \qquad \text{(stone)} \qquad (1)$$

$$\frac{d\Phi}{d\tau} = \frac{L}{mR^2} + \frac{\sin^2\alpha}{ma}\left[E - \omega L \pm \frac{1}{\beta}(E - V_{\rm L}^{+})^{1/2}(E - V_{\rm L}^{-})^{1/2}\right] \qquad \text{(stone)} \qquad (2)$$

$$\frac{dT}{d\tau} = \left(\frac{R}{rH}\right)^2 \frac{1}{m}\left[E - \omega L \pm \beta(E - V_{\rm L}^{+})^{1/2}(E - V_{\rm L}^{-})^{1/2}\right] \qquad \text{(stone)} \qquad (3)$$

$$V_{\rm L}^{\pm}(r) \equiv \omega L \pm \frac{rH}{R}\left(m^2 + \frac{L^2}{R^2}\right)^{1/2} \qquad \text{(stone)} \qquad (4)$$

Box 1 in Section 18.2 defines α, β, and ω.

Photon does not age.

To apply these equations to light, we must overcome a fundamental problem: The differential aging $d\tau$ of a light flash along its worldline is automatically zero (Section 1.4), so these equations of motion for the stone have no meaning for the light flash. To give them meaning, we eliminate $d\tau$ from these equations of motion, recasting them without $d\tau$. Use the following equations:

$$\frac{dr}{dT} = \left(\frac{dr}{d\tau}\right)\left(\frac{d\tau}{dT}\right) \qquad \text{(stone)} \qquad (5)$$

$$\frac{d\Phi}{dT} = \left(\frac{d\Phi}{d\tau}\right)\left(\frac{d\tau}{dT}\right) \qquad \text{(stone)} \qquad (6)$$

Carry out this combination on equations (1) through (3). *Result:* equations for dr/dT and $d\Phi/dT$. Note that in this process, m disappears from the coefficients—but remains in the expression for $V_{\rm L}^{\pm}(r)$ in equation (4).

To convert these equations to describe light, we take the limit of the resulting equation for dr/dT and $d\Phi/dT$ as $m \rightarrow 0$. Carry this out first on expressions that contain $V_{\rm L}^{+}$ and $V_{\rm L}^{-}$ defined in equation :

$$\left[\left(E - V_{\rm L}^{+}\right)\left(E - V_{\rm L}^{-}\right)\right]^{1/2} \qquad (7)$$

$$= \left[\left\{E - \omega L - \frac{rH}{R}\left(m^2 + \frac{L^2}{R^2}\right)^{1/2}\right\}\left\{E - \omega L + \frac{rH}{R}\left(m^2 + \frac{L^2}{R^2}\right)^{1/2}\right\}\right]^{1/2}$$

$$= \left[(E - \omega L)^2 - \left(\frac{rH}{R}\right)^2\left(m^2 + \frac{L^2}{R^2}\right)\right]^{1/2} \qquad \text{(stone)} \qquad (8)$$

Let $m \to 0$.

To convert this expression to light, take the limit as $m \to 0$:

$$\lim_{m\to 0}\left[\left(E - V_{\mathrm{L}}^{+}\right)\left(E - V_{\mathrm{L}}^{-}\right)\right]^{1/2} = \lim_{m\to 0}\left[(E-\omega L)^2 - \left(\frac{rH}{R}\right)^2\left(m^2 + \frac{L^2}{R^2}\right)\right]^{1/2} \quad (9)$$

$$= E\left[\left(1 - \omega\frac{L}{E}\right)^2 - \left(\frac{rH}{R}\right)^2\left(\frac{L}{RE}\right)^2\right]^{1/2} \quad (10)$$

$$= E\left[(1-\omega b)^2 - \frac{b^2}{R^2}\left(\frac{rH}{R}\right)^2\right]^{1/2} \quad (11)$$

$$\equiv E \times F_{\mathrm{spin}}(a,b,r) \qquad \text{(light)} \quad (12)$$

Impact parameter b for light

Equation (11) defines a new constant b, while (12) defines a new function $F_{\mathrm{spin}}(a,b,r)$:

$$b \equiv \frac{L}{E} \qquad \text{(impact parameter for light)} \quad (13)$$

$$F_{\mathrm{spin}}(a,b,r) \equiv \left[(1-\omega b)^2 - \frac{b^2}{R^2}\left(\frac{rH}{R}\right)^2\right]^{1/2} \qquad \text{(light)} \quad (14)$$

QUERY 1. Motion of light around the non-spinning black hole Show that when $a \to 0$, then $F_{\mathrm{spin}}(a,b,r) \to F(b,r)$, defined in equation (16), Section 11.3.

Only L/E matters.

The spinning black hole shares an important simplification with the non-spinning black hole (Section 11.2): the motion of light does not depend on L or E separately, but only on their ratio, the impact parameter $b \equiv L/E$.

Comment 1. Both L and b can be positive or negative.
The angular momentum L of a stone can be positive (prograde orbit) or negative (retrograde orbit). Equation (13) shows the same to be true of impact parameter b. For the non-spinning black hole, orbits in the two directions are simply mirror images of one another. In contrast, for the spinning black hole counterclockwise and clockwise orbits of a stone are quite different, as Chapters 18 and 19 show. Prograde and retrograde orbits of light are also different from one another, as Section 20.3 will show.

Equations of motion for light

These results allow us to write down the equations of motion for light in the equatorial plane of the spinning black hole:

$$\frac{dr}{dT} = \pm\left(\frac{rH^2}{R}\right)\frac{F_{\rm spin}}{1-\omega b \pm \beta F_{\rm spin}} \qquad \text{(light)} \qquad (15)$$

$$\frac{d\Phi}{dT} = \left(\frac{rH}{R}\right)^2 \frac{\dfrac{b}{R^2} + \dfrac{\sin^2\alpha}{a}\left[1-\omega b \pm \dfrac{1}{\beta}F_{\rm spin}\right]}{1-\omega b \pm \beta F_{\rm spin}} \qquad \text{(light)} \qquad (16)$$

$$\frac{dr}{d\Phi} = \pm\frac{R}{r}\frac{F_{\rm spin}}{\dfrac{b}{R^2} + \dfrac{\sin^2\alpha}{a}\left[1-\omega b \pm \dfrac{1}{\beta}F_{\rm spin}\right]} \qquad \text{(light)} \qquad (17)$$

QUERY 2. And when $a \to 0$?

In Query 1 you showed that as $a \to 0$, $F_{\rm spin} \to F$ for the non-spinning black hole. Apply the same limit to expressions in equations(15) through (17). Box 1 in Section 18.2 may be useful.

As $a \to 0$:

A. $R \to r$
B. $H^2 \to (1 - 2M/r)$
C. $\omega \to 0$
D. $\beta \to (2M/r)^{1/2}$
E. $\sin^2\alpha/a \to 0$

With these changes, show that equations (15) through (17) for the motion of light around the spinning black hole reduce to equations (17) through (19) in (Section 11.3) for the non-spinning black hole.

20.3 ■ EFFECTIVE POTENTIAL FOR LIGHT

Orbits at a glance

Derive effective potential for light.

Now we derive the effective potential for light, which allows us to predict at a glance the r-motion of a light flash that moves in the equatorial plane of the spinning black hole. This derivation follows a procedure similar to that for the non-spinning black hole in Section 11.3. Run your finger down that earlier derivation to compare the following derivation for the spinning black hole.

To start, multiply both sides equation (15) by an expression that leaves only $\pm F_{\rm spin}$ on the right side. Then multiply both sides by M/r and square both sides of the resulting equation. The result has the form:

$$A^2(a,b,r)\left(\frac{dr}{dT}\right)^2 = \left(\frac{M}{b}\right)^2 F^2_{\rm spin}(a,b,r) \qquad \text{(light)} \qquad (18)$$

$$= \left(\frac{M}{b}\right)^2 - 2M\omega\left(\frac{M}{b}\right) + M^2\omega^2 - \frac{M^2}{R^2}\left(\frac{rH}{R}\right)^2 \qquad (19)$$

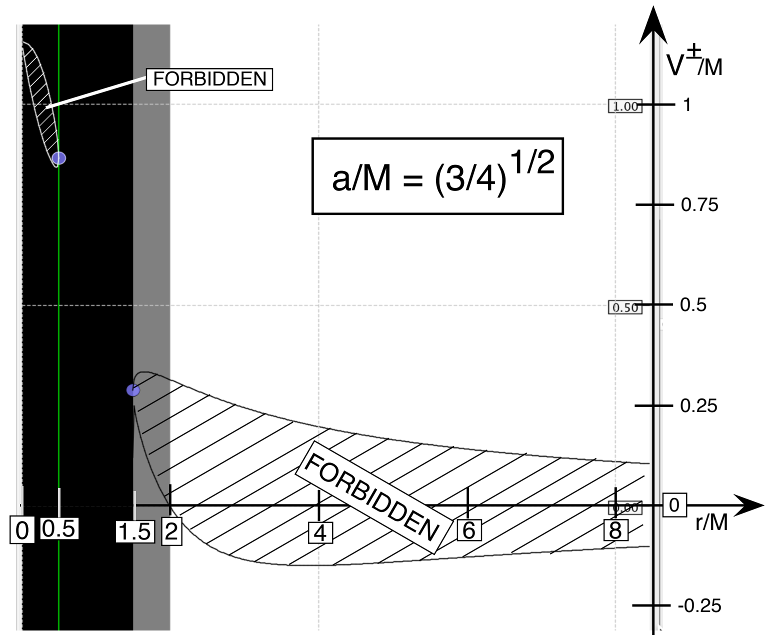


FIGURE 1 Effective potential for light with $a/M = (3/4)^{1/2}$. The gray region extends from the static limit at $r_{\text{S}} = 2M$ down to the event horizon at $r_{\text{EH}} = 1.5M$. The Cauchy horizon is at $r_{\text{CH}} = 0.5M$. There are two forbidden regions for light, one outside the event horizon and one inside the Cauchy horizon. Equation (22) shows that inside a forbidden region dr/dT is imaginary .

where

$$A(a,b,r) \equiv \left(\frac{M}{b}\right)\left(\frac{rH^2}{R}\right)\left[1 - \omega b \pm \beta F_{\text{spin}}(a,b,r)\right] \qquad (20)$$

The right side of (19) is quadratic in M/b, so factor it using the quadratic equation:

$$\frac{M}{b} = M\omega \pm \left[\frac{M^2}{R^2}\left(\frac{rH}{R}\right)^2\right]^{1/2} \qquad (21)$$

Substitute this expression for M/b into the second term on the right side of (19) and collect terms, with the result:

$$\boxed{A^2(a,b,r)\left(\frac{dr}{dT}\right)^2 = \left(\frac{M}{b}\right)^2 - \left[\frac{V^{\pm}(a,r)}{M}\right]^2 \qquad (22)}$$

where

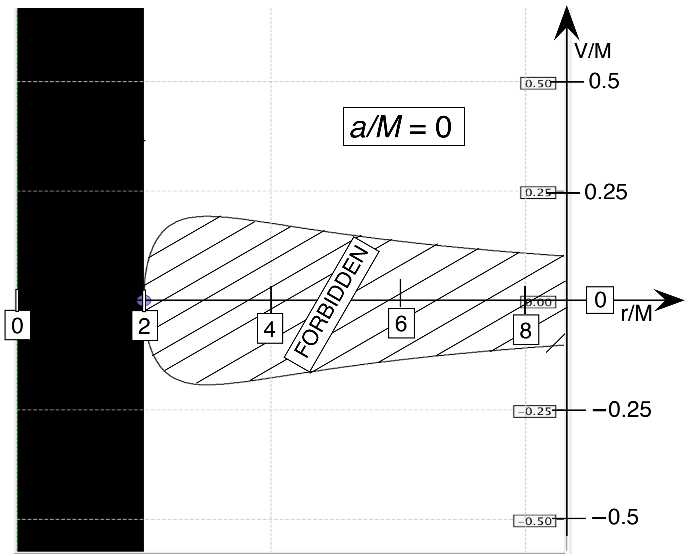


FIGURE 2 For comparison, effective potential for light for the non-spinning black hole ($a = 0$), to the same scale as Figure 1.

$$\left[\frac{V^{\pm}(a,r)}{M}\right]^2 \equiv \frac{M^2}{R^2}\left(\frac{rH}{R}\right)^2 \pm \frac{2M^2\omega}{R}\left(\frac{rH}{R}\right) - M^2\omega^2 \tag{23}$$

The superscript $\pm$ on the left side of (23) is the same as the $\pm$ on the right side.

Track r-motion of light.

Equation (22) tracks the r-motion of a light flash in the equatorial plane: The first term on the right side is a function of b but not a function of a or r, while the second term—the square of the effective potential—is a function of a and r, but not a function of b. Equation (22) then permits us to plot on the same diagram the effective potential function (23)—Figure 1—and a horizontal line that represents any given value of $(M/b)^2$ in (22).

Forbidden region for light

Equation (22) tells us why the region between V^- and V^+ in Figure 1 is forbidden: If $(M/b)^2$ is less than $(V^+/M)^2$ but greater than $(V^-/M)^2$, then dr/dT is imaginary, which is indeed forbidden. Figure 2 reminds us of the corresponding effective potential for the non-spinning black hole.

QUERY 3. Effective potential for the non-spinning black hole

Show that when $a \to 0$, equations (22) and (23) reduce to equations (25) and (26) in Section 11.3, shown in Figure 2.

QUERY 4. Mass vs. massless
Compare Figure 1 for light with the corresponding Figure 1 in Section 18.3 for a stone, both for $a/M = (3/4)^{1/2}$. Is the following statement true or false: Outside the event horizon, the forbidden region for light is entirely enclosed in the forbidden region for the stone.

Where light goes

Global radial motion dr/dT must be real. Therefore the right side of (22) must be positive. *Result:* The light flash that moves along a horizontal line at $(M/b)^2$ in Figure 1, for example, cannot enter either forbidden region between the two effective potential curves. A light flash from far away that reaches one of these curves either reverses the sign of its r-motion or holds its r-value at the curve's maximum or minimum.

20.4 ■ EXERCISE

1. Does a spinning back hole appear smaller or larger than a non-spinning black hole?

Think of two black holes. The first is a non-spinning black hole of a particular given mass M. The second is a spinning black hole of the *same* mass M. Can a distant observer tell them apart?

Specifically, (1) How large does a black hole look to an Earth observer? (2) Can the Earth observer determine whether or not the black hole is spinning or non-spinning? (3) Since most black holes in Nature spin, can the Earth observer determine *how fast* the spinning black hole rotates—that is, what is its value of a/M? When you finish this exercise, you should be able to answer these three questions clearly and definitively.

Non-spinning black hole: Review Figures 3 and 12 in Chapter 11. They show that r-coordinate of the knife-edge circular orbit for light around a non-spinning black hole $r = 3M$ corresponds to the maximum of the effective potential curve $V(r)$, and that this knife-edge orbit determines the visual size of the non-spinning black hole through the critical impact parameter b_{critical}: the visual diameter of the non-spinning black hole is $2b_{\mathrm{critical}}$.

Spinning black hole: The visual size of the spinning black hole in the equatorial plane is determined by two critical impact parameters b^+_{critical} and b^-_{critical} that belong to the prograde and retrograde knife-edge orbits of light, respectively, as shown in Figure 3 for the special case $a = (3/4)^{1/2}M$.

A. Figure 1 in this chapter plots the effective potentials $V^{\pm}(r)$ for a light beam that moves near a black hole with $a/M = (3/4)^{1/2}$. From Figure 1, show that

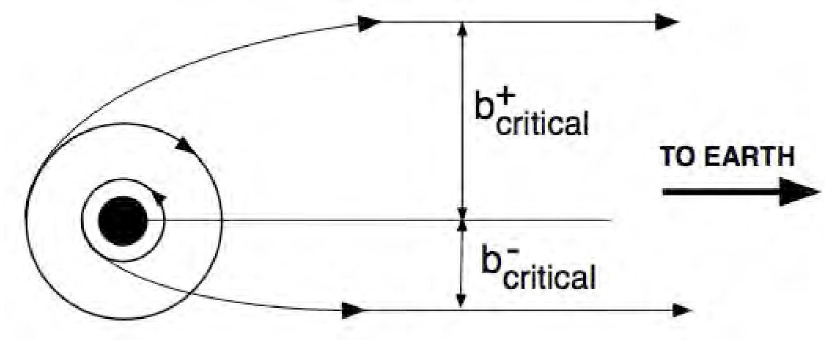


FIGURE 3 Figure for Exercise 1. Schematic diagram showing the visual size of a spinning black hole with $a/M = (3/4)^{1/2} = 0.866$.

$$b^+ = \frac{1}{V^+_{\text{max}}} = \frac{1}{V^+(r^+_{\text{knife edge}})} \quad \text{prograde knife-edge orbit} \tag{24}$$

$$b^- = \frac{1}{V^-_{\text{max}}} = \frac{1}{V^-(r^-_{\text{knife edge}})} \quad \text{retrograde knife-edge orbit} \tag{25}$$

B. Derive the following expressions for $b^{\pm}_{\text{critical}}$ and $r^{\pm}_{\text{knife edge}}$:

$$b^{\pm}_{\text{critical}} = \left(r^{\pm}_{\text{knife edge}} + 3M\right)\left(\frac{r^{\pm}_{\text{knife edge}}}{4M}\right)^{1/2} \tag{26}$$

$$r^{+}_{\text{knife edge}} = 4M\cos^2\Psi^{\pm} \text{ where } \Psi = \frac{1}{3}\arccos\left(\mp\frac{a}{M}\right) \tag{27}$$

C. Evaluate equations (26) and (27) for $a = 0$ and show that both results agree with equation (28) in Chapter 11.

D. Find the values of b^+ and b^- for a spinning black hole with (a) $a/M = (3/4)^{1/2}$ and (b) $a/M = 1$ (maximum-spin black hole).

E. *Optional:* Plot the visual size $(b^+_{\text{critical}} + b^-_{\text{critical}})$ of the spinning black hole as a function of its spin parameter a/M.

F. Answer the question posed in the title of this exercise. Include a sentence that starts, “This depends on . . . .”.

Download file name: Ch20OrbitsOfLightAroundSpinningBH170906v1.pdf

Chapter 21. Inside the Spinning Black Hole

- *Why can't I escape from inside the event horizon of the* non-spinning *black hole?*
- *How can I escape from inside the spinning black hole?*
- *When I do emerge from inside the spinning black hole, where am I?*
- *After I emerge from inside the spinning black hole, can I return to Earth?*
- *What limits does my finite wristwatch lifetime place on my personal exploration of spacetimes?*
- *What limits are there on spacetimes that a group in a rocket can visit?*

Download file name: Ch21TravelThroughTheSpinningBH170831v1.pdf

CHAPTER 21

Inside the Spinning Black Hole

Edmund Bertschinger & Edwin F. Taylor *

The non-spinning black hole is like the spinning black hole, but with its gate to other Universes closed. For the spinning black hole, the gate is ajar.

—Luc Longtin

21.1 ■ ESCAPE FROM THE BLACK HOLE

Exit our Universe; appear in a "remote" Universe!

Travel to another Universe . . .

. . . on a one-way ticket!

Chapters 18 through 20 examined orbits of stones and light around the spinning black hole. We study orbits to answer the question, "Where *do* we go near the spinning black hole?" The present chapter shifts from orbits to topology—the connectedness of spacetime. Topology answers the question, "Where *can* we go near the spinning black hole?" *Astonishing result:* We can travel from our Universe to other Universes. These other Universes are "remote" from ours in the sense that from them we can no longer communicate with an observer in our original Universe, nor can an observer in our original Universe communicate with us. *Worse:* Once we leave our Universe, we cannot return to it. Sigh!

Begin with effective potential.

Figure 1 previews this chapter by examining the r-motion of a free-fall stone—or observer—in the effective potential of the spinning black hole. Free stones with different map energies have different fates as they approach the spinning black hole from far away. Two stones with map energies $(E/m)_2$ and $(E/m)_3$, for example, enter unstable circular orbits. In contrast, the stone with map energy $(E/m)_4$ reaches a turning point where its map energy equals the effective potential, then it reflects outward again into distant flat spacetime. *Question:* What happens to a stone with map energy $(E/m)_1$? Two question marks label its intersection with the forbidden region inside the Cauchy horizon. Does the stone reflect from this forbidden region? Does it move outward again through the Cauchy and event horizons? Does it emerge into our Universe? into some other Universe? The present chapter marshalls general relativity to answer these questions.

*Draft of Second Edition of *Exploring Black Holes: Introduction to General Relativity*

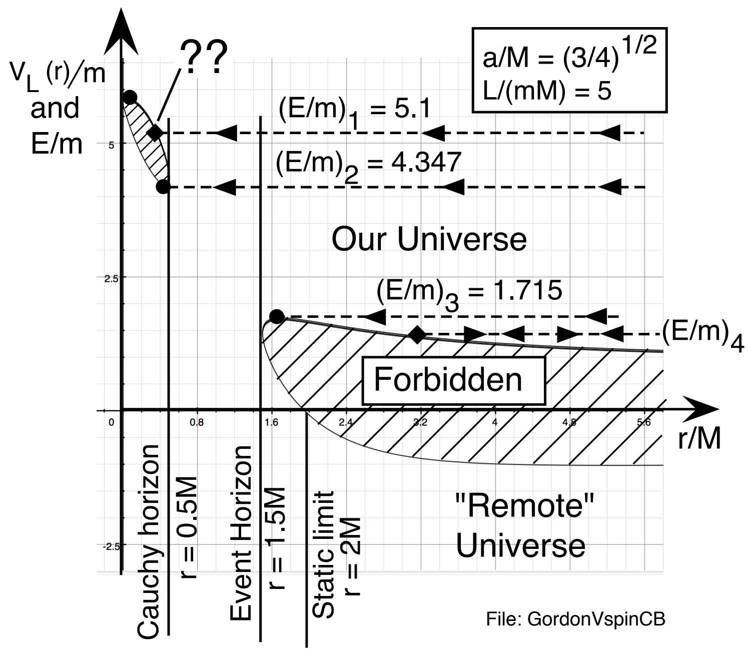


FIGURE 1 Effective potential for a stone with $L/(mM) = 5$ near a spinning black hole with $a/M = (3/4)^{1/2}$. What happens at the intersection of the horizontal line $(E/m)_1$ with the forbidden region inside the Cauchy horizon? (Adapted from Figure 5 in Section 18.4.)

Carter-Penrose diagram

The idea of traveling from our Universe to another Universe is not new. In 1964 Roger Penrose devised, and in 1966 Brandon Carter improved, what we now call the **Carter-Penrose diagram** for spacetime, a navigational tool for finding one's way across Universes. This diagram will be the subject of the following sections.

21.2 ■ THE CARTER-PENROSE DIAGRAM FOR FLAT SPACETIME

Begin around the edges, then fill in.

Global metric flat spacetime

As usual, we develop our skills gradually, first with flat spacetime, then with the non-spinning black hole, and finally with the spinning black hole. Here is a global metric on an $[x, t]$ slice in flat spacetime:

$$d\tau^2 = dt^2 - dx^2 \qquad \text{(global metric, flat spacetime)} \qquad (1)$$

$$-\infty < t < \infty, \quad -\infty < x < \infty \qquad (2)$$

The following transformation from $[t, x]$ to $[v, u]$ corrals the infinities in (2) onto a single flat page:

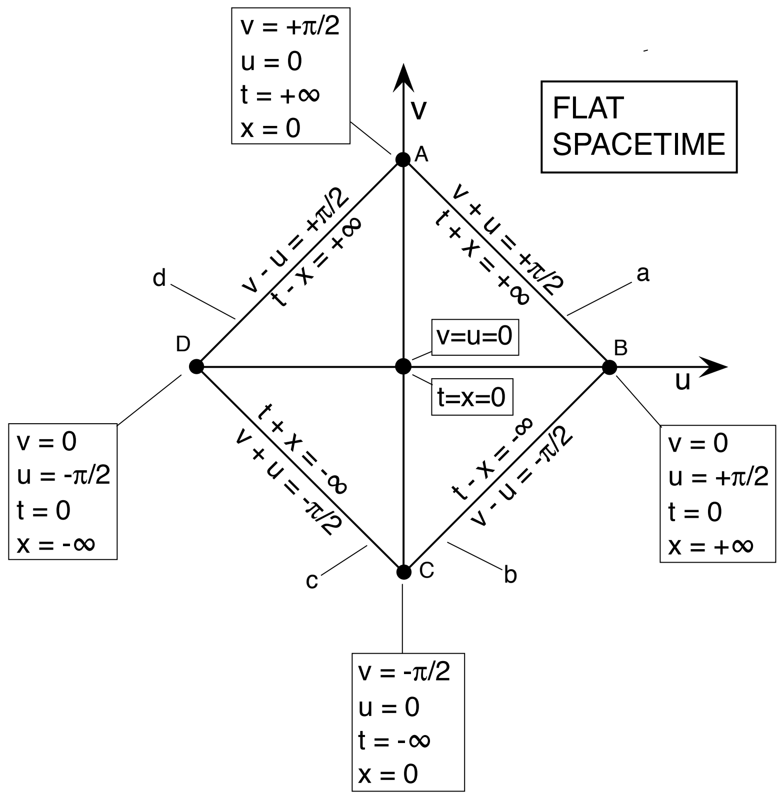


FIGURE 2 Points and lines on the boundaries in the Carter-Penrose diagram for flat spacetime.

$$t = \frac{1}{2}\left[\tan(u+v) - \tan(u-v)\right] \quad \text{(global coordinates, flat spacetime)} \tag{3}$$

$$x = \frac{1}{2}\left[\tan(u+v) + \tan(u-v)\right] \tag{4}$$

$$-\pi/2 < v < +\pi/2, \quad -\pi/2 < u < +\pi/2 \tag{5}$$

QUERY 1. Coordinate ranges
Show that transformations (3) and (4) convert the coordinate ranges of t and x in (2) into the coordinate ranges of v and u in (5). In other words, the Carter-Penrose diagram brings map coordinate infinities onto a finite diagram.

Carter-Penrose diagram

Figure 2 shows the result of this transformation, which we call the **Carter-Penrose diagram**. It plots positive infinite t at point A, negative infinite t at point C, distant positive x at point B, and distant negative x at

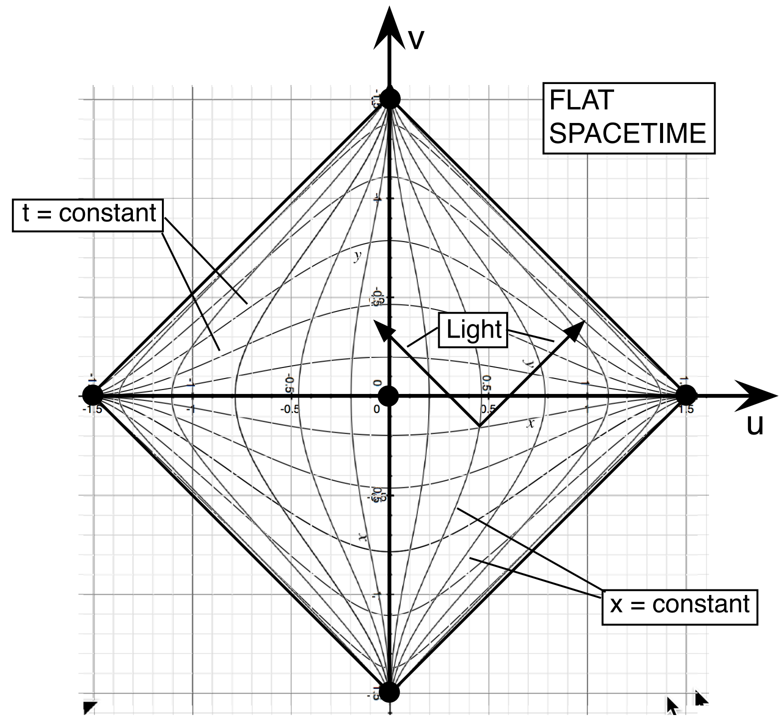


FIGURE 3 The Carter-Penrose diagram that fills in coordinates of Figure 2 on the $[x, t]$ slice of flat spacetime. These curves plot v vs. u from the inverse of equations (3) through (5). These particular conformal coordinates preserve the $\pm 45^o$ angles for worldlines of light.

point D. In Query 2 you use equations (3) through (5) to verify map coordinate values in this figure.

QUERY 2. Points and boundaries in the Carter-Penrose diagram

Use equations (3) and (4) to verify the following statements about points A through D and boundaries a through d in Figure 2:

A. Show that when $u = 0$ then $x = 0$, and when $v = 0$ then $t = 0$.

B. Verify the boxed values of t and x at points A through D.

C. Verify the values of $v + u$ along the two lines labeled a and c.

D. Verify the values of $v - u$ along the two lines labeled b and d..

E. Verify the values of $t + x$ along the two lines labeled a and c.

F. Verify the values of $t - x$ along the two lines labeled b and d.

The Carter-Penrose diagram is a **conformal diagram** that brings global coordinate infinities onto the page. A conformal diagram is simply an ordinary spacetime diagram for a metric on which we have performed a particularly

Conformal diagram

clever coordinate transformation. This particular coordinate transformation preserves the causal structure of spacetime defined by the light cone.

To find the global metric on the $[u, v]$ slice for flat spacetime, take differentials of (3) and (4) and rearrange the results:

$$dx = \frac{1}{2}\left[\frac{du + dv}{\cos^2(u+v)} + \frac{du - dv}{\cos^2(u-v)}\right] \tag{6}$$

$$dt = \frac{1}{2}\left[\frac{du + dv}{\cos^2(u+v)} - \frac{du - dv}{\cos^2(u-v)}\right] \tag{7}$$

$$-\pi/2 < v < +\pi/2, \quad -\pi/2 < u < +\pi/2 \tag{8}$$

Global metric in u, v coordinates

Substitute dx and dt from (6) and (7) into global metric (1) and collect terms. Considerable manipulation leads to the global metric on the $[u, v]$ slice:

$$d\tau^2 = \frac{dv^2 - du^2}{\cos^2(u+v)\cos^2(u-v)} \tag{9}$$

$$-\pi/2 < v < +\pi/2 \qquad -\pi/2 < u < +\pi/2 \tag{10}$$

Equation (9) has the same form as equation (1) except it is multiplied by $[\cos^2(u+v)\cos^2(u-v)]^{-1}$, called the **conformal factor**. Indeed, equations (9) and (10) are examples of a **conformal transformation:**

Definiton: **Conformal transformation**

DEFINITION 1. Conformal transformation

A conformal transformation has two properties:

- It transforms global coordinates.
- The new global metric that results has the same form as the old global metric, multiplied by the *conformal factor.*

Conformal factor

The transformation (3) through (5) has both of these properties. In particular, the resulting metric (9) has the same form (a simple difference of squares) as (1), multiplied by the conformal factor $[\cos^2(u+v)\cos^2(u-v)]^{-1}$.

Infinities on the $[x, t]$ slice correspond to finite (non-infinite) values on the $[u, v]$ slice, due to the conformal factor in (9), which goes to $x + t = \pm\infty$ or $x - t = \pm\infty$ when $u + v = \pm\pi/2$ or $u - v = \pm\pi/2$, as shown around the boundaries of Figure 2.

Worldlines of light at $\pm 45°$

For the motion of light, set $d\tau = 0$ in (9). Then the numerator $dv^2 - du^2 = 0$ on the right side ensures that $dv = \pm du$, so the worldline of light remains at $\pm 45°$ on the $[u, v]$ slice. Therefore a light cone on the $[u, v]$ slice has the same orientation as on the $[x, t]$ slice. We deliberately choose conformal coordinates to make this the case.

QUERY 3. Standing still; limits on worldlines

A. Show that when $dx = 0$ in (6), then $du = -du$, which means that $du = 0$. *Result:* The stone with a vertical worldline on the $[x, t]$ slice has a vertical worldline on the $[u, v]$ slice.

B. Show that in Figure 2 the worldline of every stone lies inside the light cone $\pm 45°$.

QUERY 4. You cannot "reach infinity."

Show that as $x \rightarrow \pm\infty$ the global equation of motion dx/dt for a stone takes the form $dx/dt \rightarrow \pm 0$. Therefore a stone cannot reach that limit, any more than it (or you!) can reach infinity.

? **Objection 1.** *Are these predictions real? They sound like science fiction to me!*

! We do not use the word "real" in this book; see the Glossary. These predictions can in principle be validated by future observations carried out by our distant descendents. In that sense they are scientific. They also satisfy *Wheeler's radical conservativism:* "Follow what the equations tell us, no matter how strange the results, then develop a new intuition."

21.3 ■ TOPOLOGY OF THE NON-SPINNING BLACK HOLE

The one-way worldline

We move on from flat spacetime to spacetime around the non-spinning black hole. Equations (17) and (18) of Section 8.4 connect the global r-motion of a stone to the effective potential $V_L(r)$:

$$\left(\frac{dr}{d\tau}\right)^2 = \left(\frac{E}{m}\right)^2 - \left(\frac{V_L(r)}{m}\right)^2 \qquad (11)$$

A remote Universe

Because all terms in this equation are squared, the effective potential $V_L(r)$ and the map energy E/m can be either positive or negative, as shown in Figure 4. our Universe lies above the forbidden region. Below the forbidden region lies a second, "remote" Universe.

Meaning of "forbidden"

What does "forbidden" mean? Equation (11) tells us that global r-motion $dr/d\tau$ becomes imaginary when $(E/m)^2$ is smaller than $(V_L/m)^2$. In other words, neither stone nor observer can exist inside the forbidden region.

The forbidden region prevents the direct passage from our Universe to this remote Universe. To do so we would have to move inward through the event horizon with positive map energy, then use rocket blasts to re-emerge below

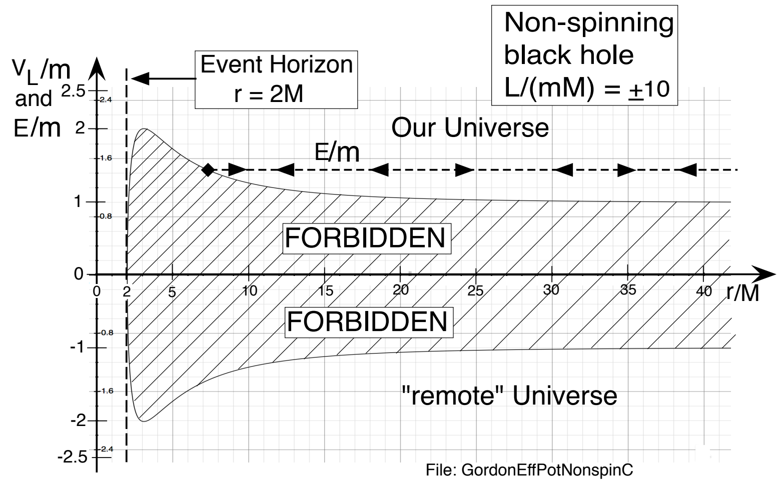


FIGURE 4 Effective potential for the non-spinning black hole, copy of Figure 5 of Section 8.4.

Door to remote Universe is closed.

the forbidden region with negative map energy. But inside the event horizon motion to smaller r is inevitable. *Result:* For the non-spinning black hole the door to to the remote Universe is closed.

Figure 5 displays the double-ended funnel-topology of the non-spinning black hole. The upper and lower flat surfaces represent flat spacetime in our Universe and in the remote Universe, respectively. The pinched connection in

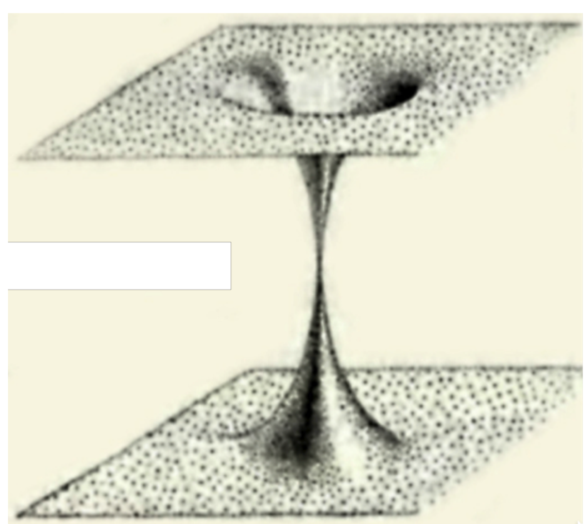


FIGURE 5 Topology of the non-spinning black hole that supplements Figure 4. The upper flat surface represents our Universe. It is connected to a remote Universe (lower flay surface) by the impassable *Einstein-Rosen bridge.*

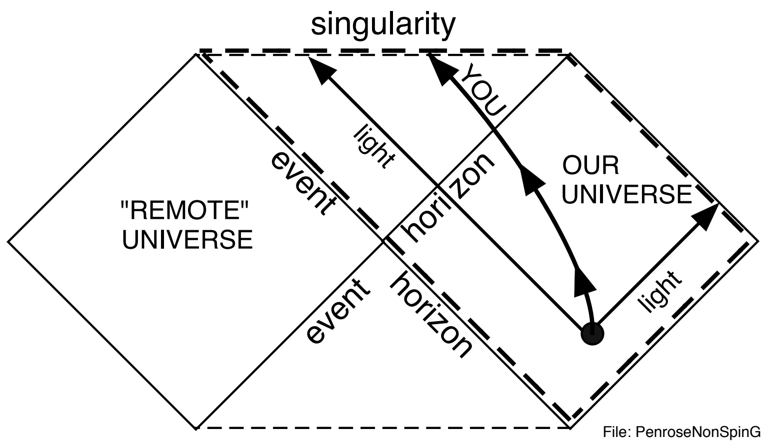


FIGURE 6 Carter-Penrose diagram for the non-spinning black hole, which has two event horizons. Heavy dashed lines enclose spacetime spanned by the Schwarzschild Metric, which has access to only one of these event horizons. From our Universe a stone, light flash, or observer cannot reach the "Remote" Universe in Figures 6 and 1 by crossing the second event horizon.

Einstein-Rosen bridge unpassable

the center, called the **Einstein-Rosen Bridge,** is too narrow for a stone or light flash to pass between the two Universes.

Now turn attention to the Carter-Penrose diagram for the non-spinning black hole, displayed in Figure 6. This two-dimensional diagram suppresses the ϕ-coordinate, leaving t and r global coordinates. The Schwarzschild metric, equation (5) in Section 3.1, becomes:

$$d\tau^2 = \left(1 - \frac{2M}{r}\right) dt^2 - \left(1 - \frac{2M}{r}\right)^{-1} dr^2 \qquad (12)$$

$$-\infty < t < +\infty, \qquad 0 < r < \infty \qquad (13)$$

Two event horizons

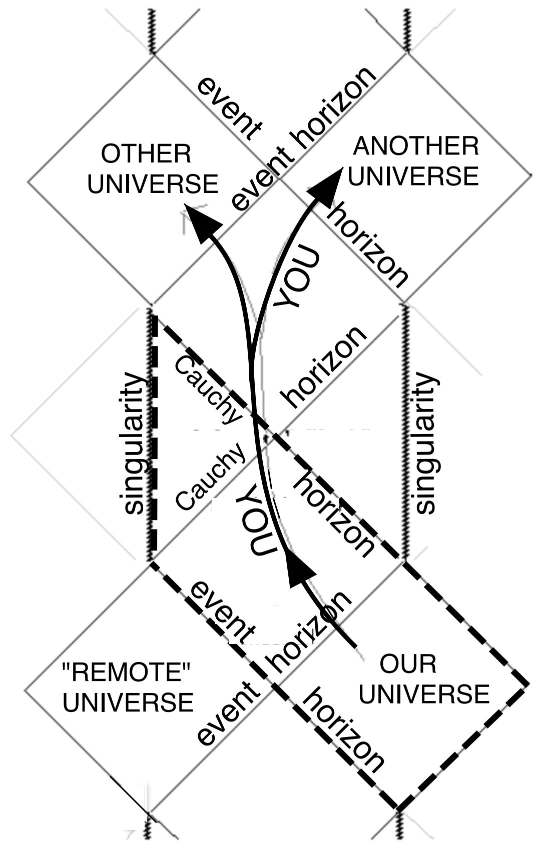
In this Carter-Penrose diagram an inward-moving stone or light flash crosses the event horizon, then moves inevitably to the singularity represented by the spacelike horizontal line. Topologically there is a second event horizon that is not available to this stone or light flash, because their worldlines are corralled within the upward-opening light cones.

21.4 ■ TOPOLOGY OF THE SPINNING BLACK HOLE

No two-way worldline!

Reflect outward from inside Cauchy horizon?

Figure 1 displays the effective potential for a stone with map angular momentum $L/(mM) = 5$ near the spinning black hole with $a/M = (3/4)^{1/2}$. The striking new feature of this effective potential is the added forbidden region *inside* the Cauchy horizon. This added forbidden region raises the



File: PenroseSpinJ

FIGURE 7 Carter-Penrose diagram of the spinning black hole that answers questions posed in the caption to Figure 1. The heavy dashed line shows the boundaries of Doran global coordinates, which enclose one event horizon and one Cauchy horizon. With calibrated rocket blasts, you can choose to enter either the Other Universe or Another Universe at the top of the diagram. The upward orientation of your worldline shows that you cannot return to our Universe once you leave it—according to general relativity.

possibility that the stone with, say, $(E/m)_1 = 5.1$ can reflect from this forbidden region and move back outward into a distant region of flat spacetime.

Worldline moves between Universes.

Figures 7 and 8 present the topology of such a spinning black hole. You, the observer who travels along the worldline in Figure 7, start in our Universe, pass inward through the event horizon and the Cauchy horizon, reflect from the forbidden region inside the Cauchy horizon, and emerge from a second Cauchy horizon. Then, with the use of rockets, you can choose which event horizon to cross into one of two alternative Universes at the top of this diagram.

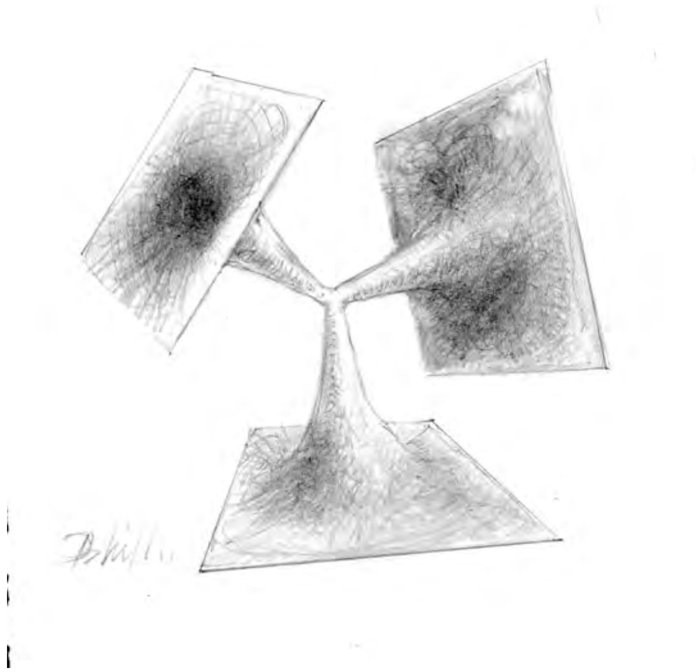


FIGURE 8 Topology of spacetime around the spinning black hole. In this case the central Einstein-Rosen bridge is wide enough for a traveler to pass through on her one-way trip to another Universe. Indeed, she may use rocket thrusts to choose between two alternative Universes. This figure supplements Figures 1 and 7.

To construct Figure 7 suppress the Φ-coordinate of the Doran metric, equation (4) in Section 17.2. The result:

$$d\tau^2 = dT^2 - \left[\left(\frac{r^2}{r^2+a^2}\right)^{1/2} dr + \left(\frac{2M}{r}\right)^{1/2} dT\right]^2 \quad \text{(Doran, } d\Phi = 0) \quad (14)$$

$$-\infty < T < \infty, \quad 0 < r < \infty$$

?

Objection 2. *Why are the lines labeled "singularity" in Figure 7 vertical, while the line labeled "singularity" in Figure 6 is horizontal?*

!

These diagrams show *topology*: where you *can* go, and where you *cannot* go. Categories "vertical" and "horizontal" in such a diagram carry no prediction for observation. Each case shows that you cannot climb out of the singularity.

Emerge into another Universe

The heavy dashed line in Figure 7 outlines the spacetime region included in Doran global coordinates. Notice that this included region is only part of available spacetime. Compare the worldline in Figure 7 with the horizontal line $(E/m)_1$ in Figure 1. This comparison shows that the reflected observer does not re-emerges into our Universe, but into one of the alternative Universes at the top of Figure 7. *Conclusion:* For the spinning black hole, the gate between alternative Universes is ajar (initial quote of this chapter). But your worldline in Figure 7 moves relentless upward; you cannot return to the Universe you have left. You can't go home again!

?

Objection 3. *How can I tell that I have reached the limits of a map, but not the limits of spacetime, when there is uncharted territory ahead.*

!

Geodesic incompleteness

If you reach the boundary of a global coordinate system in finite wristwatch time (so that $dT/d\tau \to \infty$) or some other singularity arises, and if you have not reached a singularity, then you have just demonstrated that the original coordinates are incomplete and need to be extended. This limited feature of global coordinates is called **geodesic incompleteness:** There exist (portions of) geodesics (or other curves) that reach the edge of your coordinate range and continue beyond your present global coordinates—unless you introduce new global coordinates that cover the new range.

?

Objection 4. *How many different Universes are there?*

!

In principle the Carter-Penrose diagram of Figure 7 extends indefinitely both upward and downward, embracing an unlimited number of Universes.

Triple funnel for spinning black hole

Figure 8 displays the topology through which you pass along the worldline of Figure 7. You enter the funnel from Our Universe, then use rockets to choose the Universe into which you emerge. Your worldline in Figure 7 shows that you cannot re-enter that funnel in order to return to our Universe. Your trip between Universes is a one-way street!

?

Objection 5. *So WHERE are these other Universes? Show them to me!*

!

Spacetimes multiply inside the event horizon of the spinning black hole. What does this mean for regions far from the spinning black hole? *Big surprise:* An observer can use rockets to maneuver inside the event horizon of the spinning black hole in order to choose the remote Universe into which she emerges. *Example:* Figure 7 shows that the "bouncing" traveler with $(E/m)_1$ in Figure 1 can emerge into either one of two alternative Universes shown in Figure 8. *Conclusion:* Neither of these alternative Universes "exist" in our Universe in the everyday sense—but you can travel there!

21.5 ■ EXERCISES

SUGGESTED EXERCISES, PLEASE!

21.6 ■ REFERENCES

Penrose, Roger, “Conformal treatment of infinity,” in *Relativity, Groups and Topology*, ed. C.M. de Witt and B.de Witt, Les Houches Summer Schools, 1963 Gordon and Breach, New York and London, pages 565-584.

Carter, Brandon (1966). ”Complete Analytic Extension of the Symmetry Axis of Kerr’s Solution of Einstein’s Equations,” Physical Review, 1966, Volume 141, Number 4, pages 1242–1247. Bibcode:1966PhRv..141.1242C. doi:10.1103/PhysRev.141.1242.

Kruskal, M.D, “Maximal Extension of Schwarzschild Metric,” Physical Review, Volume 119, Number 1, September 1, 1960, pages 1743-1745.

Figure 5 adapted from From Kip Thorne, *Black Holes and Time Warps: Einstein’s Outrageous Legacy,* New York, W. W. Norton, 1994, page 487.

Figure 8 drawn by Thomas Dahill.

Useful Carter-Penrose diagrams reference: William J. Kaufmann, III, *Black Holes and Warped Spacetime*, W.H. Freeman Co., San Francisco, 1979.

Download file name: Ch21TravelThroughTheSpinningBH170831v1.pdf

CHAPTER 0

Appendix A: Wheeler's Rules of Writing

If you haven't found something strange during the day, it hasn't been much of a day.

—John Archibald Wheeler

Wheeler's rules of writing

(These rules were assembled over several years by Edwin F. Taylor, one of many collaborators with John Archibald Wheeler (JAW). JAW has read these and generally approved them, but he has not edited them.)

Motivate! Motivate! Motivate! The text should read like a detective story, keeping the reader on the edge of her chair, gasping for the next handout. Every sentence quotable! Book design must contribute to the rich, headlong plunge.

Simplify! Simplify! Simplify! JAW: ''Everything important is, at bottom, utterly simple.'' Einstein: ''I want to know His [God's] thoughts, the rest are details.''

The power and generality of the singular, the specific, the committed: Avoid plurals. ''those designing Earth satellites'' becomes ''anyone designing an earth satellite.'' Use *the* rather than *a*: ''Center of the black hole,'' not ''center of a black hole.'' No ''if,'' no ''suppose;'' instead, use ''when.''

The power of the present: Avoid past tense unless talking about history. Avoid unnecessary future tense.

The power of the active: Avoid passives.

The dullness of simply being: Suppress the use of the verb ''to be.''

JAW: ''Whenever I have an 'is' in a sentence, I know there is something wrong with that sentence.''

''...is not an harmonic oscillator'' becomes ''...does not rate as an harmonic oscillator.''

''He is happy'' becomes ''He beams happiness.''

''Schwarzschild spacetime geometry is distinguished from all other conceivable geometries...'' becomes ''Schwarzschild spacetime geometry distinguishes itself from all other conceivable geometries... .''

Avoid the subjunctive (''We would like to express the metric as...'') except in cases in which you are presenting something with which you do not agree (''Some would conclude incorrectly that...'').

Avoid ''ing'' words. ''before escaping or plunging'' becomes ''before it escapes or plunges''; ''The Earth is rotating'' becomes ''The Earth rotates.''

Put the key word first or early in the sentence or at the end of the sentence, not in the middle.

Rhetorical rule of threes. Use three descriptions to establish a triangle that spans the idea being presented: ''proper time, interval, wristwatch time'' or ''Schwarzschild radial coordinate, *r*, reduced circumference.'' It is also a reminder of the different descriptors of the same thing.

Use ''we'' to include the student, rather than ''you,'' which is not so friendly: ''As we plunge into a black hole... .''

However... .

Use infinitive construction: ''To find'', ''to learn'', ''to determine'' rather than ''we do so and so'' or ''let us do so and so,'' which is condescending because the author is going to do it anyway.

''We use the Principle of Relativity to derive the invariance of the interval.'' or ''Let us use the Principle of Relativity to derive the invariance of the interval.'' or ''Use the Principle of Relativity to derive the invariance of the interval.'' all become ''To derive the invariance of the interval, use the Principle of Relativity.''

Use commands to stir the blood—but sparingly.

''Find...''

''Determine...''

''Reckon...'' (rather than ''compute'' or ''calculate,'' which seem technical)

''Plunge...''

But not so much as to seem bossy.

Appeal to experiment or logic—not to the professions. Do not invoke ''scientists'' to enforce a point.

John A. Wheeler
Department of Physics
Princeton University
Princeton, New Jersey 08544

(#WheelersRulesOnly)

John Archibald *Wheeler's Rules of Writing* influence every page of this book. We struggle to follow his commands "Motivate!" "Simpify!" "Use infinitives" and the rest. Sometimes we find it awkward or impossible to do so. Other times we intentionally ignore him: For example, Wheeler's word "reckon" seems to us old-fashioned.

Wheeler did not include in his published Rules of Writing some central hallmarks of his composition. Chief among these is **self-descriptive terminology**; the most important of these is the name **black hole**. Wheeler did not create the term *black hole*, but adopted it immediately when someone (now unknown) used it in a question at one of Wheeler's lectures (initial quote, Chapter 3). Similarly, we call a reading on your wristwatch *wristwatch time*. The standard term is *proper time*, but that is not self-descriptive. (What could *improper time* possibly mean?) *Wristwatch time* is our effort to duplicate the gorgeous German noun "die Eigenzeit," literally "one's own time."

The self-descriptive term *shell* refers to an imaginary spherical latticework concentric to a black hole and a local inertial frame at rest on that shell (Sections 5.7 and 7.4). *Rain* describes a stone or local inertial frame that falls from initial rest at a great distance (Section 7.6); *hail* a stone or local inertial frame flung radially inward from a great distance; and *drip* a stone or local inertial frame dropped from rest off a shell (all summarized in Section 9.7).

Wheeler uses the same unit for distance and time measured in a local inertial frame (Section 1.2), either: (1) meters of distance and meters of light-travel time, or (2) years of time and light-years of distance. He measures energy and momentum in the common unit mass, so that E/m and $\boldsymbol{p}/m$ have no units (Section 1.10).

Other rules of writing we developed ourselves: We use no abbreviations whatsoever in this book, except in subscripts and in an occasional equation label. We always spell out the words *second* and *meter* because abbreviations are ambiguous. (Does m mean mass or meters? Does s refer to seconds or distance?) Our goal is to eliminate what we call *hiccups*: moments when the reader must pause to recall the meaning of a term.

We have developed notation rules of our own: *Be consistent and avoid redundant notation!* Always write subscripts in the order (component, frame, particle); for example, $E_{\text{ring},b}$ as the ring-frame energy of particle b, or $p_{x,\text{ring},b}$. Finicky? Absolutely! Avoid hiccups at all cost.

Are we similarly limited in everyday conversation with friends and colleagues? Of course not. Everyday life is full of gorgeous ambiguity. But ambiguity does not belong in our textbook. We believe that to be slightly boring is much better than to be unclear!

We talk constantly about components but do not use vectors. Most derivatives are total; only twice in the book do we use partial derivatives. For this introductory text, we simplify the metric and the resulting analysis by describing spacetime and motion on a *slice*, a spatial symmetry plane through the center of a non-spinning black hole (Section 3.6) or in the equatorial plane of the spinning black hole (Section 17.2).

Wheeler exclaimed, "I *hate* footnotes!" So there are none in this book: Full reference to every quote is given in the References section at the end of each chapter.

Throughout the book we employ the *radical conservatism* of John Archibald Wheeler: *Follow what the equations tell us, no matter how strange the results, then develop a new intuition!*

REFERENCE

John Archibald Wheeler, "Wheeler's rules of writing," *American Journal of Physics*, Volume 67, Number 11, November 1999, page 945.

Download file name: VAppendixAWheelersRules170511v1.pdf

GLOSSARY, Exploring Black Holes

<u>MANY TERMS</u> ARE ALSO DEFINED INSIDE THE BACK COVER

WORDS <u>NOT</u> USED IN THIS BOOK, EXCEPT IN QUOTATIONS, MATHEMATICAL EXPRESSIONS, OR NEWTON'S ANALYSIS. (SEVERAL TERMS ARE MENTIONED ONCE IN ORDER DISMISS THEM.)

absolute, absolutely, actual, actually, amazing, angular momentum (with no modifier), beam (used for global motion of light only), component (never for global quantities; instead use "r-motion" or "phi-motion"), correct, distance (with no modifier), endless, energy (with no modifier), established, eternal, eternally, everyone knows that, exact, exactly, fact, in fact, forever, geodesic (instead: worldline of a free particle), Hertz, imaginary, impossible (better: forbidden by the current laws of physics), incredible, incredibly, infinite, infinity, infinitely, Minkowski spacetime (instead, say flat spacetime), never, obvious, obviously, permanent, proof, proper time, radius, radial (OK for the non-spinning black hole), real, reality, rest mass, scalar, Schwarzschild radius, space (with no modifier), tensor, time (with no modifier), trajectory, true, truth, vector, why (asking for purpose or intention).

We are not creating a new dogma here. "Why" and "fact" and "truth" are not forbidden in everyday discussion, for example in analysis of human motivation. But our careful definition of words in this physics textbook helps us to enforce the discipline of predicting observation and measurement and the creation, verification, and application of established theory.

WORDS USED IN THIS BOOK.
Each entry is a pointer to a description or to the full definition in context.

aberration, Section 12.5

advanced civilization, Section 19.4

black hole, Box 4 in Section 6.6

bounce orbit, Section 11.4

bounce point, non-spinning black hole, Section 8.4

circle point, non-spinning black hole, Section 8.4

circlar orbit point for light, Section 11.5

conformal transformation, Section 21.2

constant of motion, Section 1.10

critical impact parameter, Section 11.3

curvature of spacetime, Section 1.10

effective potential for a stone, Section 10.4, Section 11.4

ACKNOWLEDGMENTS for the Second Edition of *Exploring Black Holes*

Except for Edwin Taylor's USA income tax deductions, no one has made a single dollar on the second edition of *Exploring Black Holes*, now posted online for free download. All participated out of professional intellectual delight in developing the general relativity concepts of John Archibald Wheeler.

Nandor Bokor helped to resurrect a stalled project. His deep understanding, love of the subject, and thirst for clarity informs every section of every chapter. Nandor instantly grasps each argument and moves seamlessly toward its optimum expression. He organizes and archives our dropsite versions. Generations of his students at the Budapest University of Technology and Economics have tried out endless revisions.

Anyone who reads more than one paragraph of this book is deeply in debt to Luc Longtin. From the smallest typo, through physics errors, to recasting entire sections, Luc's comments and recommendations on literally hundreds of revisions shape what you read today. Along with Nandor Bokor, Luc serves as enforcer of our gradually-developing understanding of the radical difference between (arbitrary) global coordinates and the local inertial frames in which we make every measurement and observation and analyze it using special relativity. These two reviewers have helped to exterminate prohibited words that lack sufficiently precise definition for a textbook --- see the Glossary.

William Ingham and students in his classes have laid us low and helped us up more times that we can count. In one of his classes, Michael Trop's advice motivated the huge question mark and exclamation point that label each objection/response, while Laura Simone's questions led to construction of the Personal Planetarium in Chapter 12.

Chris Gould of North Carolina State University, among others, also relayed difficulties that he and more than one generation of his students found in our treatments.

The Physics Librarian Tan-Qui Tran at the Massachussetts Institute of Technology (MIT) located original references beyond number, so that giants in the field are (often!) correctly acknowledged.

John Rogosich of Techsetters, Inc. formatted our chapters without compensation from us, from MIT, or from any publisher.

Peter Brown forced us to rethink just about everything more often than we wanted to.

For many years, students in many classes in various colleges and universities read through sequential versions of this text, shared with us their detailed difficulties, and gave us advice. See "Only the Student Knows," *American Journal of Physics,* Volume 60, Number 3, March 1992, pages 201-202.

Ruth Chabay advised us to forget hard-copy book publishing and distribute our work from this free-download dropsite.

More individuals than we can recall or reward helped us toward this current second edition of *Exploring Black Holes*. Rejoice in them as you read.

---The authors

GENERAL RELATIVITY BRIEFING

THE FOLLOWING TWO PAGES WILL BE FACING PAGES INSIDE THE BACK COVER

Please goto next page

Download file name: YInsideBackCover170421v4.pdf

GENERAL RELATIVITY BRIEFING

(Inside back cover, left and right sides)

EVENT

An occurrence, such as your birth, that we can locate at a point in space at an instant of time. Events are the elemental nails on which physics hangs. An event exists independent of any method we use to locate it.

SPACETIME

The arena in which events occur. Newton thought that time is universal, the same for all observers. Einstein realized that different observers typically measure different values of time between a pair of events, along with different values of their spatial separation. Einstein combined space and time to provide a spacetime measure of separation between events on which all observers agree. [See INTERVAL and METRIC.]

SPACETIME REGION

A volume of space during a period of time, for example the vicinity of a black hole during one Earth-year.

SPECIAL RELATIVITY

A theory of motion in a spacetime without gravity or curvature.

GRAVITY

In Newtonian physics, a universal *force* arising from mass. For Newton, a force may be real (such as gravity) or fictitious (such as centrifugal force or Coriolis force) and we can correctly analyze gravity in a single *global* inertial reference frame. [See INERTIAL FRAME.] In general relativity, however, gravity is *always* a fictitious force which we can eliminate anywhere by changing to a local frame that is in free fall (a different free-fall frame for each event).

CURVATURE

A measure of those properties of spacetime which prevent us from having a global inertial frame. Sources of curvature include mass-energy and pressure.

GENERAL RELATIVITY

A theory of curved spacetime and motion.

SINGULARITY

A spacetime location at which curvature is so extreme that general relativity fails.

PATCH

A spacetime region purposely limited in size and duration so that curvature does not noticeably affect a given measurement or observation. A patch can have uniform gravity (which generates no curvature). We can approximate curved spacetime to any desired accuracy with a set of patches, in the same way that the curved surface of the space shuttle was covered with many small flat tiles.

STONE

A particle whose location at each instant is described by an event and whose mass warps spacetime too little to be measured. A stone has nonzero mass and moves slower than light with respect to every local inertial frame.

WORLDLINE

The path of a stone through spacetime. We can mark on a worldline a chain of intermediate events, such as ticks of the stone's wristwatch. The **wristwatch time** is the reading on the stone's wristwatch at any event along its worldline.

GLOBAL COORDINATE SYSTEM

Any system that assigns unique numbers ("coordinates") to every event in order to locate it in an extended spacetime region. General relativity frees us to use (almost) any global coordinate system in a given spacetime region. In this book we usually analyze events that occur on a single spatial plane, for which three coordinates, such as (t, x, y) or (t, r, ϕ), suffice to locate an event.

FRAME

A local coordinate system of our choice installed on a patch. This local coordinate system is limited to that single patch. We may construct a frame at any event except on a singularity such as that inside a black hole.

INERTIAL or FREE-FALL FRAME

A local coordinate system in which special relativity is valid, typically Cartesian spatial coordinates and synchronized clocks. A free stone moves with constant velocity as measured in coordinates of an inertial frame. In general relativity every inertial frame is local; spacetime curvature precludes a global inertial frame.

ROOM

A room is a physical enclosure of fixed spatial dimensions in which we make measurements and observations over an extended period of time measured in that room.

OBSERVER

An observer is a person or machine that moves through spacetime making measurements, each measurement limited to a local inertial frame. Thus an observer moves through a series of local inertial frames.

WORLDTUBE

A worldtube is a bundle of worldlines of objects at rest in a room and worldlines of the structural components of that room. Think of a worldtube as sheathing the worldline of an observer at work in the room.

INTERVAL

Any one of three possible alternative directly-measured **spacetime separations** between any pair of adjacent events. If a free stone can travel directly from one event to the other, the time lapse on that stone's wristwatch is called the **wristwatch time** or **timelike interval**. If only a light flash is fast enough to travel directly from one event to the other, we say there is a **null interval** or **lightlike interval** between them. If nothing is fast enough to move between the two events, then the separation is a **spacelike interval**. In this case there exists an inertial frame such that the two adjacent events are simultaneous in that frame. The distance between the events measured along a ruler at rest in that frame is called the **ruler distance**.

GLOBAL METRIC

A function that satisfies Einstein's field equations, whose inputs are global coordinates and global coordinate differentials (such as dt, dr, $d\phi$) between an adjacent pair of events and whose output is the interval between the events. The global metric (plus the topology of the region) completely determines local spacetime and gravitational effects within the global region in which it applies. The global metric combines with the Principle of Maximal Aging [see entry] to predict the global motion of stones and massless particles in curved spacetime.

FRAME METRIC or LOCAL METRIC

A metric expressed in local coordinates on a spacetime patch and valid only on that patch. A frame metric is *always* distinguished from a global metric by the notations: (1) increment Δ instead of differential d, (2) an approximately equal sign $\approx$, (3) a one-word subscript label on every local coordinate increment (such as Δt_{shell}) that identifies the local frame. In contrast, a global differential such as dt *never* carries a one-word subscript.

INVARIANT

A physical quantity that has the same value, whether measured or calculated, with respect to any possible frame or global coordinate system. Examples include the interval between two infinitesimally close events, the wristwatch time along a worldline, the mass of a stone, and the mass of a center of gravitational attraction.

PRINCIPLE OF MAXIMAL AGING

A free stone follows the worldline of maximal wristwatch time (maximal aging) across any two adjoining local frames. We can completely tile spacetime with adjoining frames, so the stone knows how to move everywhere in a global spacetime region, such as that near a black hole or near Earth.

CONSTANT OF MOTION

A quantity that describes the motion of a free stone or massless particle and whose value does not change with time in the given global coordinate system. Constants of motion arise when all coefficients in the metric are independent of one or more global coordinates. Map energy and map angular momentum are each a constant of motion for a free stone near a black hole.

METRIC ON A CURVED SURFACE *VS.* METRIC IN CURVED SPACETIME

ON A STATIC CURVED SURFACE IN SPACE:

First, measure directly the DISTANCE between nearby points using a ruler or tape measure—or timed round trip of laser or radar pulse. This direct measurement requires no coordinates. Second, set up a SURFACE MAP, which is just a rule that assigns unique coordinates to each point on the surface. Aside from requirements of uniqueness and smoothness, map coordinates are totally arbitrary. Third, ask for the relation between the directly-measured distance between two given nearby points and their coordinate differences. The result is a local MAP SCALE between the two points. Finally, generalize, seeking a GLOBAL METRIC: a function whose inputs are coordinates and coordinate differences between a pair of adjacent points anywhere on the surface and whose output is the squared distance between them.

The output of the metric is an INVARIANT: the directly-measured distance between two nearby points has the same value no matter what map coordinates we choose.

Almost everywhere on our surface we can treat as FLAT a sufficiently small surface patch. Choose local coordinates to turn this patch into a FRAME. The resulting LOCAL METRIC tells us the invariant distance between any pair of points in that frame. Demand that *every* measurement on the entire curved surface be made in such a local frame.

The frame metric derives from the global metric. The reverse is not true; the frame metric tells us *nothing* about the global metric from which it is derived.

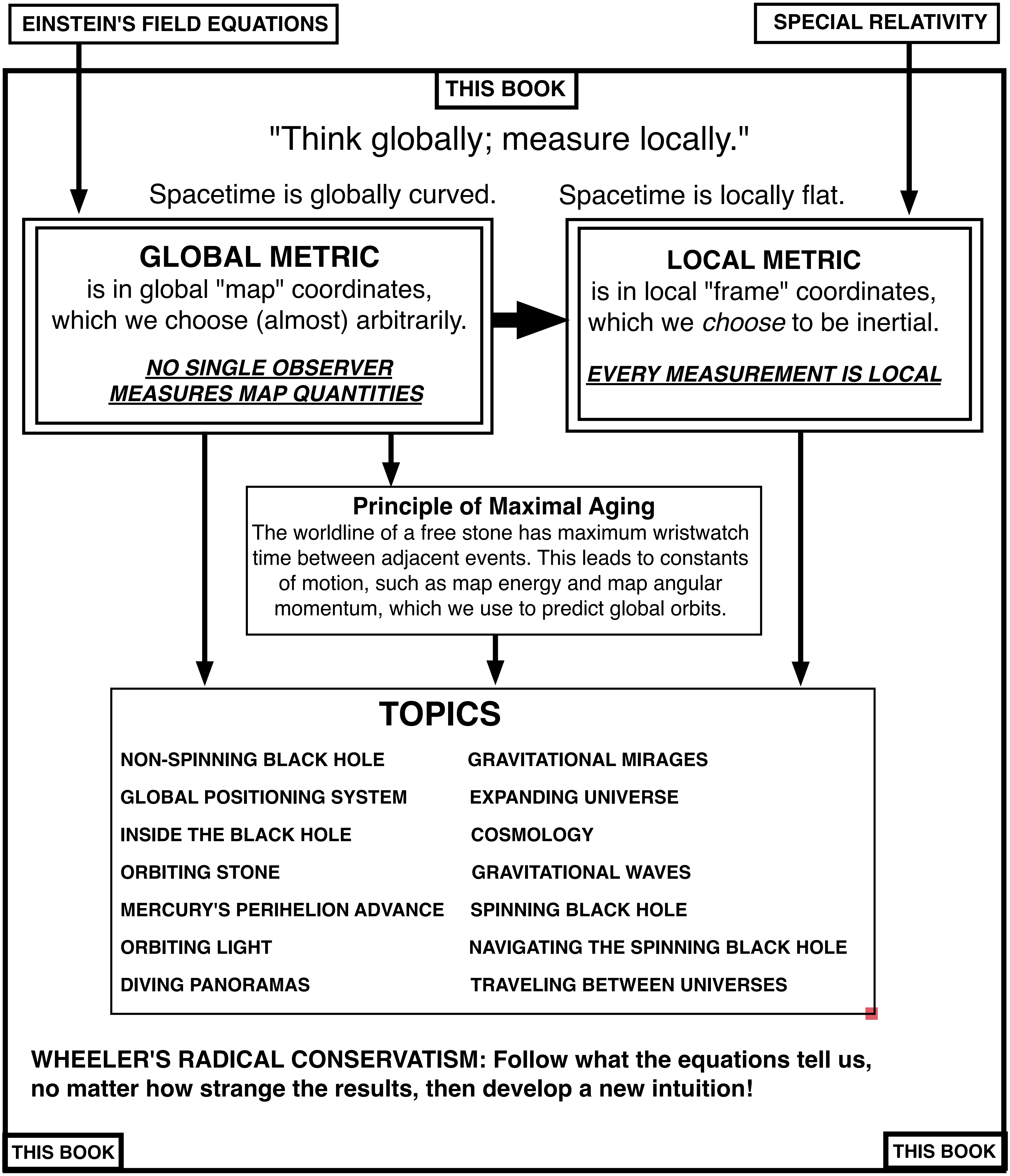
IN A REGION OF CURVED SPACETIME:

First, measure directly the SPACETIME SEPARATION (*timelike*, *lightlike*, or *spacelike* interval) between nearby events. This direct measurement requires no coordinates. Second, set up a SPACETIME MAP, which is just a rule that assigns unique coordinates to each event in the spacetime region. Aside from some simple requirements of uniqueness and smoothness, map coordinates are totally arbitrary. Third, ask for the relation between the directly-measured spacetime separation between two given nearby events and their coordinate differences. The result is a local MAP SCALE between the two events. Finally generalize, seeking a GLOBAL METRIC: a function that satisfies Einstein's field equations whose inputs are coordinates and coordinate differences between a pair of adjacent events and whose output is the squared spacetime separation between them.

The output of the metric is an INVARIANT: the directly-measured spacetime interval between two nearby events has the same value no matter what map coordinates we choose.

Almost everywhere in our spacetime region we treat as FLAT a sufficiently small spacetime patch. Choose local coordinates to turn this patch into a FRAME. The resulting LOCAL METRIC tells us the invariant interval between any pair of events in that frame. Demand that *every* measurement in the global spacetime region be made in such a local frame.

The frame metric derives from the global metric. The reverse is not true; the frame metric tells us *nothing* about the global metric from which it is derived.



KEY IDEAS: Just three words summarize this book: <u>spacetime</u>, <u>motion</u>, <u>measurement</u>!
The global metric--with arbitrary global coordinates--describes <u>spacetime</u>.
The Principle of Maximal Aging describes free <u>motion</u>.
C*hoose* to report every <u>measurement</u> with respect to a local inertial frame.