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

Local inertial frame available anywhere

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

*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

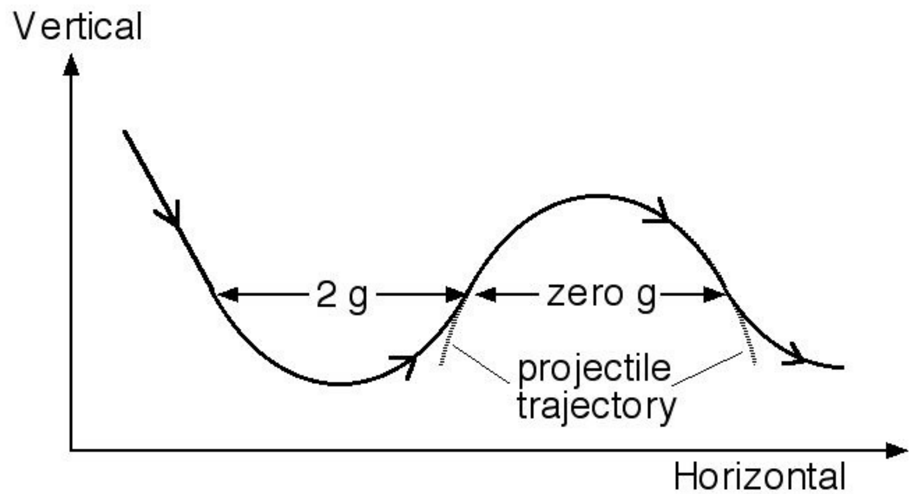

FIGURE 2 Trajectory followed by the Vomit Comet airplane above Earth's surface. Portions of the trajectory marked "2 g" and "zero g" are parabolas. During the zero-g segment, which lasts up to 30 seconds, the plane is guided to follow the trajectory of a free projectile in the absence of air resistance. 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

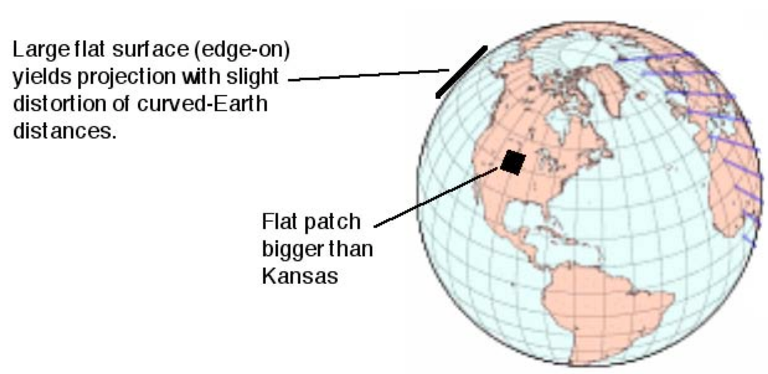

FIGURE 4 Small space patch and large flat plane tangent to Earth's surface. Projecting Earth's features onto the large flat plane can lead to distortion of those features on the resulting flat map. For precise mapmaking, the larger surface does not satisfy the requirements of a *space patch*.

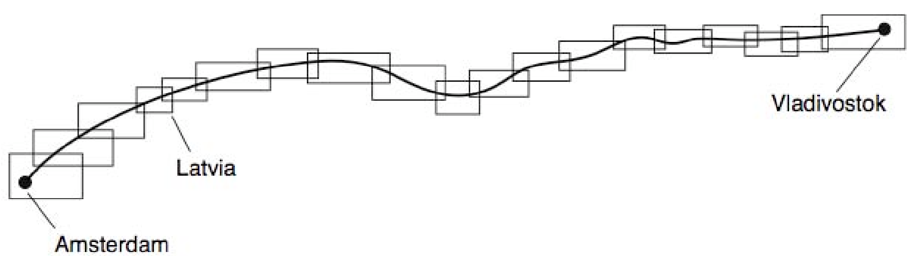

FIGURE 5 To reckon the total length of the path between Amsterdam and Vladivostok, sum the short separations across a series of small, overlapping, flat maps lined up along our chosen path. One of these small, flat maps covers all of Latvia. The smaller each map is—and the greater the total number of 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:

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

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.

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

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, $\Delta y_{\rm Latvia}$ aligns itself with a great circle that passes through the north geographic pole, while $\Delta x_{\rm Latvia}$ lies in the perpendicular east-west direction.

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

Geographic north and magnetic north yield same Δs.

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

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 $\Delta s_{\rm Latvia}$ 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.

Pythagorean Theorem valid on rotated flat maps.

2.3 ■ GLOBAL COORDINATE SYSTEM ON EARTH

Global space metric using 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.

Use latitude and longitude.

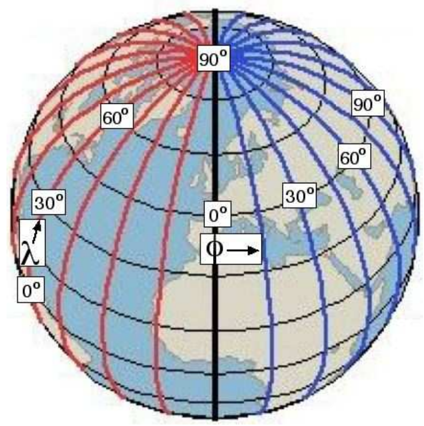

FIGURE 6 Conventional global coordinate system for Earth using angles of latitude λ and longitude ϕ.

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.

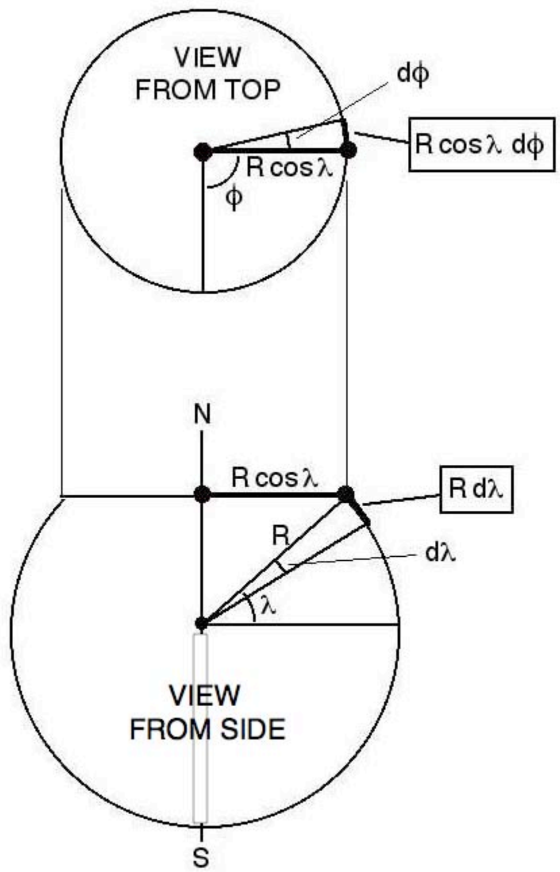

FIGURE 7 Derive the global space metric (3), as the sum of the squares of the north-south and east-west sides of a little box on Earth's surface. The north-south side of the little box is $Rd\lambda$, where R is the radius of Earth and $d\lambda$ is the differential change in latitude. The east-west 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:

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

. . . 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):

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

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.

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.

Integrate differential separation ds to calculate exact length of long 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.

Find shortest path

Does Earth care what global coordinate system we use to indicate positions on it? Not at all! An accident of history (and international politics)

Use any global coordinate system whatsoever.

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):

$$d\tau^2 = dt_{\text{lab}}^2 - ds_{\text{lab}}^2 = dt_{\text{rocket}}^2 - ds_{\text{rocket}}^2 \qquad \text{(flat spacetime)} \qquad (6)$$

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:

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

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.

DEFINITION 2. Spacetime patch

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:

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

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

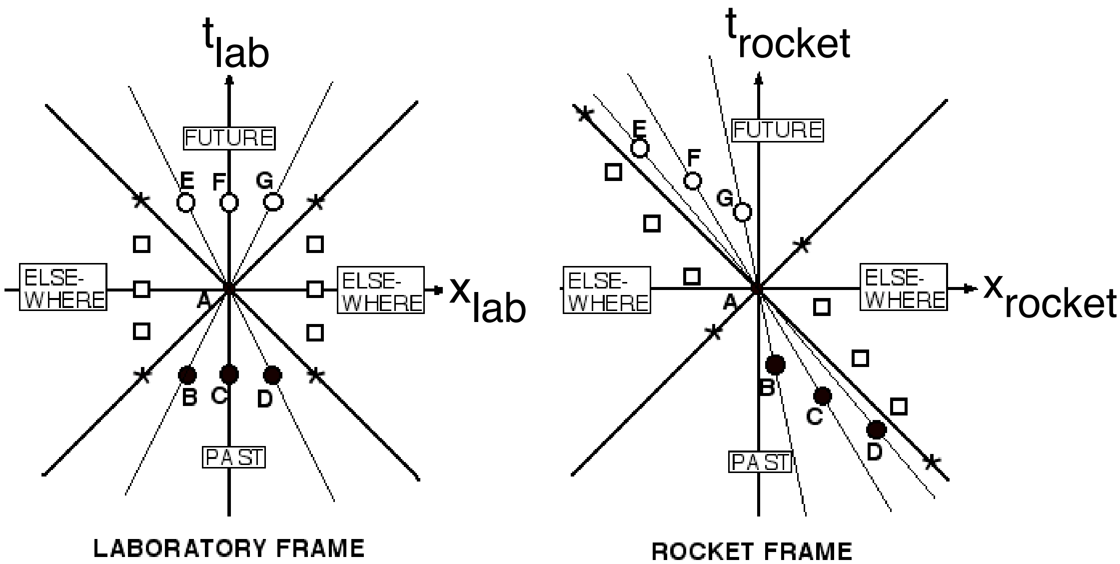

FIGURE 8 Preservation of cause and effect in special relativity. The laboratory spacetime diagram is on the left, an unpowered rocket spacetime diagram is on the right. Both diagrams plot a central Event A, and other events that may or may not cause A or be caused by A. Heavy diagonal lines are worldlines 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.

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.

Impossibility of cause and effect is also preserved.

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

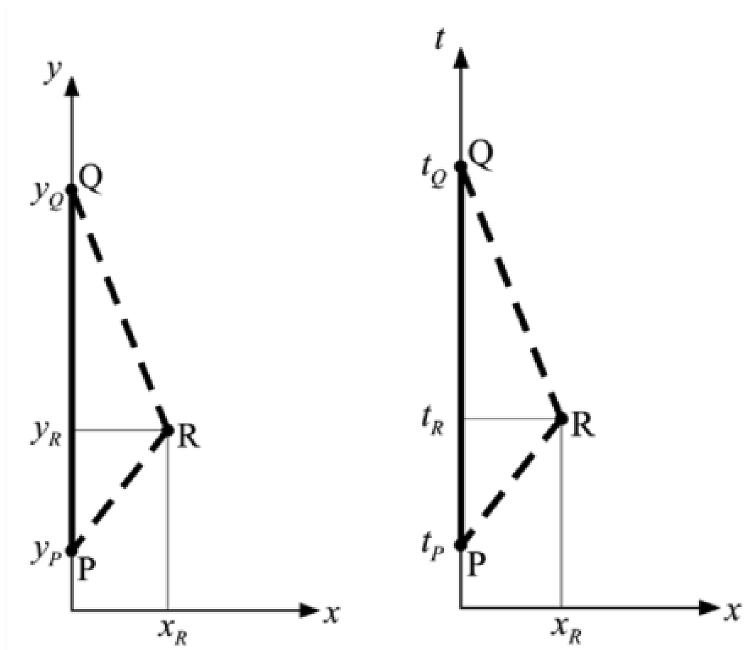

FIGURE 9 Left panel: Euclidean plane showing straight line PQ and broken line PRQ. Right panel: Spacetime diagram showing straight worldline PQ and broken worldline PRQ.

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.