Chapter 1

Introducing Engineers to Relativity

space, time and
gravitation,
briefly

One of the best introductions to his theory of relativity is found in Einstein's essays [Einstein], in which he stated:

"The theory of relativity is that physical theory which is based on a consistent physical interpretation of the concepts of motion, space and time. The name 'theory of relativity' is connected with the fact that motion from the point of view of possible experience always appears as the *relative* motion of one object with respect to another.

"Motion is never observable as 'motion with respect to space' or, as it has been expressed, as 'absolute motion.' The 'principle of relativity' in it's widest sense is contained in the statement: The totality of physical phenomena is of such a character that it gives no basis for the introduction of the concept of 'absolute motion'; or shorter but less precise: There is no absolute motion." Einstein continued with:

"The development of the theory of relativity proceeded in two steps, 'special theory of relativity' and 'general theory of relativity.' The latter presumes the validity of the former as a limiting case and is its consistent continuation."

Einstein then briefly described both the special and the general theories of relativity. It is worth to emphasize the fact which Einstein stated above—it is not two theories; general relativity completely includes special relativity as a limiting case where the gravitational field is negligible. Such a limiting case can happen in empty space, far away from massive objects and in the absence of acceleration. It can also happen closer to massive objects, inside a small laboratory that is in free-fall, where the gravitational field is practically not detectable inside the laboratory.

1.1 The special theory of relativity, briefly

Einstein was lead to his special theory of relativity by his believe that there is no way to detect absolute motion. This dictated that the measured speed of light must be the same in all inertial frames of reference.* Einstein called

*Inertial frames are uniformly moving coordinate systems, far away from gravitational or any other form of influence, where inertia is isotropic, meaning a given force will cause the same acceleration on identical masses in whatever direction the force is applied.

this the *"Light-principle"*, or L-principle for short.

Einstein realized that of if this principle does not hold, there must be a 'rest-frame' for light. This means that we could in principle set up an inertial frame in which light would not propagate in the forward direction at all (if the frame moves at the speed of light relative to the aether). Einstein reportedly contemplated if he would still be able to see his own face in a mirror if they were both at rest in such a frame.

We can extend this to say that radars as we know them would not work in such a moving inertial frame. Even at less extreme speeds, standard radars would report wrong distances, with errors that depend on direction of movement. More about radar measurements later.

Einstein realized that it is paradoxical to assume the same light ray can actually move with the same speed c (in an absolute Newtonian sense) relative to all inertial frames. This would require that light adapts it's "absolute speed" to the frame that measures it. He decided that either time intervals or distance measurements (or both) must change if measured by observers in different inertial frames that are in relative motion.

Exactly how these intervals must change, Einstein found in the transformation equations that Lorentz has developed before. These equations transforms time and distance measurements from one inertial frame to another precisely as required by the L-principle.

We will deal with the Lorentz transformation in a later chapter. For the purpose of introduction, it requires that for any two events, A and B that occurs in space and time, there is a quantity called the *spacetime interval* that remains unchanged, irrespective of in which inertial frame the components of the spacetime interval are measured.

The spacetime interval can be 'spacelike', 'lightlike' or 'timelike', as defined below:

$$\Delta s^2 = \begin{cases} (\Delta \mathrm{s}pace)^2 - (\Delta \mathrm{t}ime)^2 & \text{if } \Delta \mathrm{s}pace > \Delta \mathrm{t}ime \quad \text{(spacelike)}, \\ 0 & \text{if } \Delta \mathrm{s}pace = \Delta \mathrm{t}ime \quad \text{(lightlike)}, \\ (\Delta \mathrm{t}ime)^2 - (\Delta \mathrm{s}pace)^2 & \text{if } \Delta \mathrm{s}pace < \Delta \mathrm{t}ime \quad \text{(timelike)}, \end{cases}$$

in geometric units, where $c = 1$ so that $\Delta \mathrm{t}ime$ and $\Delta \mathrm{s}pace$ are expressed in the same units. Most engineers would probably prefer this to rather be expressed in the normal SI units of metres and seconds. It can be done by simply replacing all references to time by ct, thus converting seconds to metres. The spacetime interval will look then like this:

$$\Delta s^2 = \begin{cases} \Delta\mathbf{x}^2 - c^2\Delta t^2 & \text{if } \Delta\mathbf{x} > c\Delta t \quad \text{(spacelike)}, \\ 0 & \text{if } \Delta\mathbf{x} = c\Delta t \quad \text{(lightlike)}, \\ c^2\Delta t^2 - \Delta\mathbf{x}^2 & \text{if } \Delta\mathbf{x} < c\Delta t \quad \text{(timelike)}. \end{cases}$$

The author attempts to use SI units throughout, but here and there it is so much clearer if the constant c is not cluttering the equations that geometric units are being used. It is usually very clear which units are under consideration.

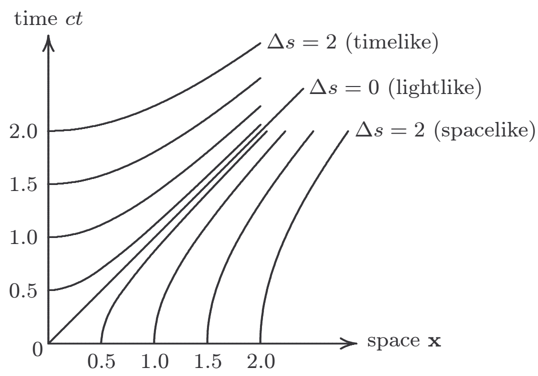

Figure 1.1: Spacetime intervals plotted for Δs ranging from 0 to 2 in 0.5 steps, both lightlike and timelike. For $\mathbf{x}$ large, they all approach the $\Delta s = 0$ (lightlike) line asymptotically. Note the precise symmetry of spacelike and timelike intervals around the lightlike interval.

Figure 1.1 illustrates the three types of interval on a standard spacetime diagram. A timelike interval is the only type where an observer, traveling slower than light, can be present at both events. This is so because there is enough time to cover the distance between the two events at a speed slower than that of light.

It then follows that a spacelike interval is the type where nothing, not even light, can be present at both events. In relativistic jargon, the two events are not causally connected, or stated more simply, the two events could not have influenced each other.

A lightlike interval is the borderline between the above two intervals and is only applicable to some types of waves and to massless particles—light, radio waves, gravitational waves, etc—things that move at the speed of light.

Spacelike intervals are normally denoted by Δs and timelike intervals with $\Delta\tau$, so that $\Delta s = -c\Delta\tau$.

The fact that the (spacetime) interval remains unchanged, irrespective of which inertial system measures it, may appear to be completely unremarkable. In Newton mechanics, where time intervals and space intervals remain unchanging when you change your inertial frame of reference, the interval will obviously remain unchanged.

Newton however, demands that the *measured speed of light* is different in different inertial reference frames. In order to conform to Einstein's *L-principle* (the invariance of the measured speed of light), either Δt*ime* or Δs*pace* or *both* must change if you switch between inertial reference frames that is moving relative to each other.

In order to satisfy the L-principle and leave the interval unchanging, both must change in a very specific way.

The speed connection If we take the timelike interval between two events A and B as

$$c\Delta\tau = \sqrt{c^2\Delta t^2 - \Delta s^2}$$

and factorize $c^2\Delta t^2$ out from the righthand side, we get

$$c\Delta\tau = \sqrt{1 - \dot{\mathbf{x}}^2}\Delta t, \quad (1.1)$$

where $\dot{\mathbf{x}} = \frac{\Delta s}{c\Delta t}$, which can be interpreted as the uniform speed that an observer must maintain (relative to the reference frame) to be present at both events.

The arrow AB in figure 1.2 represents an observer that leaves event A at time t_A and arrives at event B at time t_B, as measured in the coordinate system $\mathbf{x}, ct$. In the coordinate system of the moving observer $(\mathbf{x}', ct')$, the arrival time is at time t'_B, so that $\Delta t' = t'_B - t'_A = \Delta\tau$.

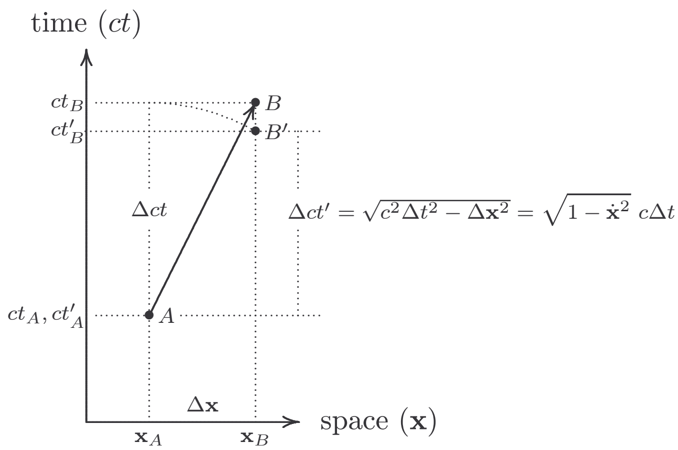

Figure 1.2: The arrow AB represents a uniformly moving observer that is present at both events A and B. In order to keep the interval Δs invariant, the moving observer must measure a time interval of $\Delta t' = \sqrt{1 - \dot{\mathbf{x}}^2}\Delta t$.

By definition, the 'moving' observer is stationary in an inertial frame that moves at a speed $\dot{\mathbf{x}}$ relative to the original reference frame. Since the observer is present at both events, the two events are separated in the observer's space by $\Delta \mathbf{x}' = 0$. The interval $\Delta\tau$ must be identical for both inertial frames, so

$$c\Delta\tau = \sqrt{c^2\Delta t^2 - \Delta\mathbf{x}^2} = \sqrt{c^2\Delta t'^2 - 0} = c\Delta t',$$

meaning the time difference between the two events must be measured by the moving observer as

$$\Delta t' = \Delta\tau = \sqrt{1 - \dot{\mathbf{x}}^2}\Delta t. \tag{1.2}$$

This statement counters the argument that is sometimes expressed, namely that special relativity is ambiguous in that either of the two observers can be considered as moving relative to the other one, so either could be considered as having a clock that is 'slow' when compared to the other's.

Here the situation is not symmetrical—one observer is present at both events and the other one is not. It is true that any one of the observers can be chosen as 'stationary' and the other one as moving relative to this 'stationary' frame of reference.

However, if the two (inertial) observers are moving relative to each other, *only one of them can be present at both events.** The time interval mea-

*Provided that the two events do not both happen at the place and moment where the two observers pass each other—then they will both measure $\Delta t = \Delta\mathbf{x} = 0$, which is not an interesting experiment.

sured by that observer is called the *propertime interval*, $\Delta\tau$. Propertime is an extremely important concept in relativity theory.

The above interpretation has nothing to do with the fact that distant observers will detect events with a time delay caused by the finite speed of light. By 'measure the time difference' we mean that the event times have been corrected for the time that light takes to travel from the event to the observer.

This does of course mean that the distance between the observer and the events must be known, which brings us to the way inertial observers will measure the distance between the two events.

Let Pam be the observer that moves between events A and B at a speed $\dot{\mathbf{x}}$ relative to Jim. Let event A happens as the two of them pass each other. There is no problem to understand how Pam measures the distance between the events. She is after all present at both and the distance in her inertial frame is zero.

How does Jim, who is *not* present at both events, measure the distance between events A and B? Let event B be a flash of light, generated by Pam after she has moved some distance away from Jim.

Equip Jim with a good radar with which he can constantly monitor Pam's distance as she moves away from him. Jim can read the distance ($\Delta\mathbf{x}$) of event B at the moment he observes the light flash, directly from his radar.

The fact that by the time Jim observes the flash from event B, Pam will be some distance past the position of the event, does not influence Jim's confidence in his measurement. The return signal of his radar and the flash of event B started out at precisely the same place and time and came to him at the same speed—the speed of light.

It is fairly obvious that Jim will measure a longer distance between the two events than Pam. Pam is the 'moving observer', who here measured the distance as zero!

However, Pam was stationary in her own inertial frame of reference, so for her Jim was the 'moving observer'. The only—and very significant—difference between Pam and Jim is that Pam was present at both events and Jim was not.

Inertial observers that are moving relative to Pam cannot also be present at both events.* They will all measure time and distance intervals between

*Remember, they are inertial observers, so they cannot turn around in any way.

the two events that are longer then those measured by Pam.

The reason for laboring the observation of the time and space intervals between events is this: events in empty space give us something 'tangible' to base comparisons between inertial frames on.

It does not say whose clock is running faster or slower than anybody else's. It does say unequivocally who will *measure* the shorter time and distance between two events—it the observer who is present at both events.

One of the classic 'tests' of special relativity's predictions is the case of the muon particles. They are created high in the earth's atmosphere by cosmic rays hitting oxygen atoms. The muons have such a short 'half-life'* that

*Half-life is a statistical parameter, meaning the time in which half of the particles (on average) would have decayed.

even if they travel at the speed of light,* virtually none of them could

*Which they don't, but they come quite close to the speed of light.

possibly make it to the surface of the Earth.

Yet they are routinely detected in laboratories on Earth in reasonable abundance. The secret lies in the fact that the muons are present at two events, C (their creation) and D (their detection on the ground), while we as observers on the ground are clearly not present at both events.

On Newtonian grounds, we predict that the time that the muons would take to reach the ground is much too long for any appreciable number to survive. Because of their high speed, the muons experience a time difference between events C and D that is much shorter than what Newton would have predicted, allowing a lot of them to make it to the ground.

So far, this introduction was an 'engineering-like' attempt to acquaint the reader with the all important spacetime interval. So before we go any further, it will be appropriate to give a brief overview of how the relativists view and express the spacetime interval.

1.2 The formal spacetime metric, briefly

Spacetime, in the absence of gravity, is expressed by the Minkowski metric with a *line element*

$$\begin{aligned} ds^2 &\equiv \eta_{\mu\nu}\, d\mathbf{x}^\mu d\mathbf{x}^\nu \qquad (\mu, \nu = 0, 1, 2, 3) \\ &= -c^2 dt^2 + dx^2 + dy^2 + dz^2, \end{aligned} \tag{1.3}$$

where $\eta_{\mu\nu}$ is the Minkowski metric tensor. The indices μ and ν indicate which component of 4-space is under consideration (i.e., t, x, y or z). In this notation, $\mathbf{x}^\mu$ does not mean $\mathbf{x}$ raised to the power μ, but rather that μ is an index that indicates how $\mathbf{x}$ must be summed.

This 'Einstein summation convention' sums over repeated indices, e.g., the μ and ν in $\eta_{\mu\nu}$ and $d\mathbf{x}^\mu d\mathbf{x}^\nu$, without the implicit summation sign being used. For example, if both μ and ν range from 0 to 1 only, the summation will result in

$$\eta_{\mu\nu} d\mathbf{x}^\mu d\mathbf{x}^\nu = \eta_{\mu\nu}(d\mathbf{x}^0 d\mathbf{x}^0 + d\mathbf{x}^0 d\mathbf{x}^1 + d\mathbf{x}^1 d\mathbf{x}^0 + d\mathbf{x}^1 d\mathbf{x}^1). \tag{1.4}$$

When μ and ν range from 0 to 3, the summation will naturally have all combinations up to $d\mathbf{x}^3 d\mathbf{x}^3$, i.e., 16 terms in all. Each $d\mathbf{x}^\mu d\mathbf{x}^\nu$ term is multiplied by the corresponding element of the metric tensor $\eta_{\mu\nu}$, which is best represented by a 4x4 matrix.

For Minkowski spacetime the $\eta_{\mu\nu}$ matrix is relatively simple, presented here as a "bordered matrix" for clarity:

$$(\eta_{\mu\nu}) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array},$$

where only the diagonal elements are non-zero and they are unitary, indicating 'flat' spacetime. It means that only terms with $\mu = \nu$ remains after multiplication and the coefficients are all unity, so that

$$\begin{aligned} \eta_{\mu\nu} d\mathbf{x}^\mu d\mathbf{x}^\nu &= -(d\mathbf{x}^0)^2 + (d\mathbf{x}^1)^2 + (d\mathbf{x}^2)^2 + (d\mathbf{x}^3)^2 \\ &= -c^2 dt^2 + dx^2 + dy^2 + dz^2. \end{aligned} \tag{1.5}$$

This may look like a very round-about way to achieve a simple result. And what is more, to rigorously prove that the metric tensor $\eta_{\mu\nu}$ has the form shown above, requires quite complex tensor analysis. We will skip that and accept the $\eta_{\mu\nu}$ matrix at face value. The complexity is the price paid for mathematical generality.

The $d\mathbf{x}^\mu d\mathbf{x}^\nu$ terms can represent many things, not just 4-space coordinates. Further, $ds^2 = \eta_{\mu\nu}\ d\mathbf{x}^\mu d\mathbf{x}^\nu$ is not restricted to 'flat' spacetime, as we will see later. Also, the formalism can handle virtually any coordinate system. We can replace the Cartesian coordinate system (x, y, z) with an equivalent spherical coordinate system (r, θ, ϕ), as shown in figure 1.3, where

$$\begin{aligned} x &= r\cos\theta\sin\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\phi. \end{aligned}$$

When dx, dy and dz are computed from the above and substituted into the Cartesian line element, the line element for spherical coordinates becomes

$$ds^2 = -c^2dt^2 + dr^2 + r^2d\phi^2 + r^2\sin^2\phi\ d\theta^2, \tag{1.6}$$

valid for 'flat' spacetime.

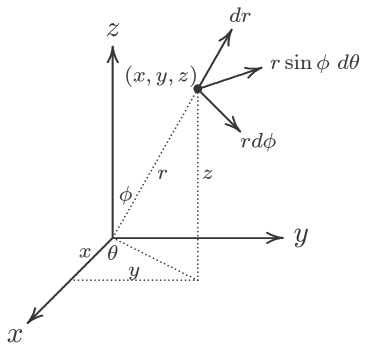

Figure 1.3: Small changes in r, θ and ϕ create an 'orthogonal coordinate system' dr, $rd\phi$ and $r\sin\phi\ d\theta$ at point x, y, z. Note that while dr signifies radial displacement, the other two directions signify transverse (or tangential) displacements relative to the origin.

This gives the elements of the $\eta_{\mu\nu}$ matrix for spherical coordinates as

$$(\eta_{\mu\nu}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2\sin^2\phi \end{pmatrix}.$$

The above spherical form of the metric is not important in flat spacetime, but is very convenient in the curved spacetime environment of general relativity, as will be discussed later in this chapter.

The spacetime line element ds corresponds to a spacelike interval Δs. We have seen before that a timelike interval, normally indicated by $\Delta\tau$, can be obtained from $c^2\Delta\tau^2 = -\Delta s^2$. So the timelike line element can be written as

$$c^2d\tau^2 = -ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2.$$

Some authors prefer to work solely with the timelike interval as the metric of spacetime, e.g., [Faber]. This is presumably because most of the intervals that we can observe and measure are timelike.

The Lorentz transformation. A Dutch physicist H.A. Lorentz is credited with a set of transformation equations that transformed space and time between inertial frames in relative motion. Historically, Lorentz was not the first person making the suggestion,* but he was the first to publish

*G.Fitzgerald, an Irish physicist, first postulated a contraction in the direction of movement. Relativistic length contraction is commonly known as *Lorentz/Fitzgerald contraction.*

the set of equations, now known as the *Lorentz transformation*

Lorentz did not discover special relativity though. He simply found a mathematical way to transform measurements made on objects moving through the aether (or absolute space) that made them conform to the null result of the aether-drift experiment of Michelson and Morley.

In essence, his equations transformed the space interval ($\Delta\mathbf{x}'$) and time interval ($\Delta t'$) as measured by a frame moving relative to the aether, to the 'absolute' space interval $\Delta\mathbf{x}$ and 'absolute' time interval Δt.

Lorentz had no physical theory for why his transformations seem to agree with experiments. He did postulate that length contraction might be a physical reality, but he could not explain why time had to transformed too, other than that it explained observations. The Lorentz transformation equations in SI units are:

$$\Delta\mathbf{x} = \frac{\Delta\mathbf{x}' + \dot{\mathbf{x}}c\Delta t'}{\sqrt{1-\dot{\mathbf{x}}^2}} \tag{1.7}$$

$$c\Delta t = \frac{c\Delta t' + \dot{\mathbf{x}}\Delta\mathbf{x}'}{\sqrt{1-\dot{\mathbf{x}}^2}} \tag{1.8}$$

where $\dot{\mathbf{x}} = \frac{d\mathbf{x}}{cdt}$, the speed relative to the aether. Einstein's contribution was that these equations can be used to transform time and space directly between any two inertial frames in relative motion and not just between an inertial frame and the absolute (aether) frame. In short, they conform to the principles of relativity.

The meaning of Einstein's interpretation of the Lorentz transformation is simply this: measure a space interval and a time interval in any inertial frame. Through the invariance of the spacetime interval, the transformation tells you what the value of the space and time intervals will be in any other inertial frame.

The 'absolute frame of reference', the aether, was not required at all. As we will see later, things are sometimes much simpler when we do not have to contend with the aether—especially in one-way Doppler measurements.

1.3 The general theory of relativity, briefly

Einstein's *general theory of relativity* is essentially a theory about the gravitational fields generated by massive objects. It is also about the dynamics of objects moving in such gravitational fields.

The objects can be massless particles like photons that always move at the speed of light relative to every inertial reference frame; or they can be massive objects that always move at speeds less than that of light relative to every inertial reference frame.

The gravitational field can be thought of as a ‘deformation’ of the fabric of spacetime caused by massive objects. ‘Test objects’ move ‘as straight as possible’ through this deformed spacetime.

All cases are locked up in Einstein’s field equations, *the Einstein equation* for short. With $c = 1$, $G = 1$,* it is given by

*This means that geometric units are being used for simplicity and clarity.

$$R_{\mu\nu} - \frac{1}{2}R\, g_{\mu\nu} = T_{\mu\nu} \qquad (\text{where } \mu, \nu = 0, 1, 2, 3), \tag{1.9}$$

where $R_{\mu\nu}$ is the *Ricci tensor*, R the *Ricci scalar*, $g_{\mu\nu}$ the generalized form of the metric tensor $\eta_{\mu\nu}$ and $T_{\mu\nu}$ the *energy-momentum tensor*. The Ricci tensor is a contraction of the *Riemann curvature tensor* and the Ricci scalar is the trace of the Ricci tensor.

So the equation tells us how the energy and momentum in space (the right hand side) cause the curvature of spacetime (the left hand side). The curvature of spacetime influences the movement of massive bodies through spacetime, thus changing the momenta. So there is ‘cross-talk’ between the left- and right hand sides, making the full equation very, very difficult to solve.

For any given situation there are up to 10 different $T_{\mu\nu}$ values to be established in terms of energy, distance and time—not 16, because the tensor is symmetrical, meaning $T_{01} = T_{10}$ etc. Various solutions to these equations presumably represent every possible equation of motion that exists in the macroscopic universe.

We will not dig into all that complexity, but rather attempt to provide some intuitive feel for certain specific solutions. The first, and perhaps best known exact solution to Einstein’s field equation was derived by Karl Schwarzschild in 1916, only months after Einstein published his general theory of relativity.

This solution provides the gravitational field outside of an isolated, spherically symmetrical, non-rotating mass, permanently at rest at the origin of a 3-d spherical coordinate system r, θ, ϕ (as in figure 1.3).

In such a case $R_{\mu\nu} = T_{\mu\nu} = 0$, but all the components making them up are not zero (the components simply sum to zero). The solution is then simpler, though not trivial at all. The spacetime metric of the gravitational field is obtained through the *metric tensor*, just like for flat spacetime, as

$$ds^2 \equiv g_{\mu\nu}\, d\mathbf{x}^\mu d\mathbf{x}^\nu,$$

or

$$c^2 d\tau^2 \equiv -g_{\mu\nu}\, d\mathbf{x}^\mu d\mathbf{x}^\nu,$$

with

$$(g_{\mu\nu}) = \begin{array}{c} \\ cdt \\ dr \\ d\phi \\ d\theta \end{array} \begin{array}{c} \begin{array}{cccc} cdt & dr & d\phi & d\theta \end{array} \\ \begin{pmatrix} g_{00} & 0 & 0 & 0 \\ 0 & g_{11} & 0 & 0 \\ 0 & 0 & g_{22} & 0 \\ 0 & 0 & 0 & g_{33} \end{pmatrix} \end{array},$$

so that the metric becomes

$$c^2 d\tau^2 = -(g_{00}\ c^2dt^2 + g_{11}\ dr^2 + g_{22}\ d\phi^2 + g_{33}\ d\theta^2). \tag{1.10}$$

By (tediously)* solving for the $T_{\mu\nu}$ of the energy-momentum tensor and

*Found in most general relativity texts. The details fall outside of the scope of this book.

casting them into $g_{\mu\nu}$ form, the following values are obtained for the non-zero elements of the metric tensor:

$$g_{00} = -(1 - \tfrac{2GM}{rc^2}) \tag{1.11}$$
$$g_{11} = (1 - \tfrac{2GM}{rc^2})^{-1} = -1/g_{00} \tag{1.12}$$
$$g_{22} = r^2 \tag{1.13}$$
$$g_{33} = r^2 \sin^2\phi. \tag{1.14}$$
$$\tag{1.15}$$

This gives the ‘Schwarzschild metric’ as

$$c^2 d\tau^2 = (1 - \tfrac{2GM}{rc^2})c^2dt^2 - \frac{dr^2}{1 - \frac{2GM}{rc^2}} - r^2 d\phi^2 - r^2 \sin^2\phi\ d\theta^2. \tag{1.16}$$

It is easy to see that when $rc^2 \gg 2GM$, i.e., far from the central mass, the metric reduces to the ‘flat’ Minkowski spacetime of special relativity.

Because of the ‘awkwardness’* of using, especially, $\sqrt{-g_{00}}$ in many places,

*Some texts, e.g. [Pathria] use the convention of labeling indices from 1 to 4 and making time the 4th coefficient g_{44}.

g_{00} will be replaced by $g_{tt} = -g_{00} = 1 - 2GM/(rc^2)$, meaning the ‘time-time’ coefficient of the Schwarzschild metric.

To make it clear that the usage is nonstandard, the other coefficient that is regularly used, g_{11}, will be relabeled g_{rr}, loosely meaning the ‘radial-radial’ coefficient. Now, if we express $d\tau$ in terms of dt, like we did for special relativity, we get

$$c^2 d\tau^2 = [g_{tt} - g_{rr}\frac{dr^2}{c^2dt^2} - \frac{r^2(d\phi^2 + \sin^2\phi\ d\theta^2)}{c^2dt^2}]\ c^2dt^2.$$

Since dr^2 is a radial spatial displacement squared and $r^2(d\phi^2 + \sin^2\phi\ d\theta^2)$ is a transverse spatial displacement squared, we can write

$$d\tau^2 = [g_{tt} - g_{rr}\frac{v_r^2}{c^2} - \frac{v_t^2}{c^2}]\ dt^2. \tag{1.17}$$

where v_r and v_t are the radial and transverse coordinate velocity components respectively.

This is a most illuminating expression of the spacetime metric. It tells us that, compared to coordinate time flow dt^2, propertime flow $d\tau^2$ is reduced by three terms inside the bracket.

The first is a static term: g_{tt}, which is always less than unity. The other two are velocity related terms: $g_{rr}v_r^2/c^2$ and v_t^2/c^2. We will first examine the static term in more detail.

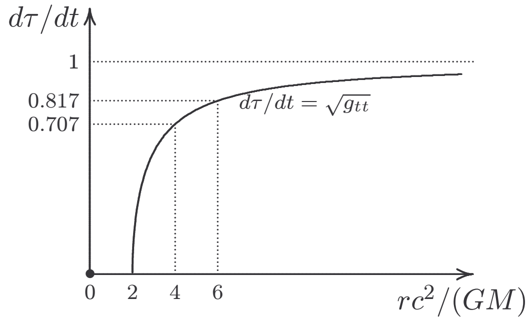

Figure 1.4: The gravitational time dilation factor (or propertime flow $d\tau/dt$), against coordinate radial distance r from mass M. The region $2\frac{GM}{rc^2} < r <= 6\frac{GM}{rc^2}$ has special significance (see text for details). Far from mass M, when $r \to \infty$, $d\tau/dt \to 1$.

As shown in figure 1.4, the 'rate of propertime flow' $d\tau/dt$ is zero at $r = 2GM/c^2$ and then increases rapidly until $r = 4GM/c^2$, where the slope of the curve is unity (45 degrees). After that, the curve starts to approach unity asymptotically.

We will later see that the *measured* static gravitational acceleration (i.e., the initial acceleration of an object kept stationary and then released so that it free-falls), is proportional to the slope of the curve $d\tau/dt$ against r, at least to a first approximation.

This suggests that the measured gravitational acceleration at $r = 2GM/c^2$ will approach infinity. This radial distance is called the Schwarzschild radius r_S and the spherical surface associated with r_S is called the event horizon of a static black hole. Because of the infinite acceleration, nothing, not even light, can escape from within the event horizon.

The velocity related terms are a bit more subtle. We have met the velocity time dilation factor of flat spacetime: $d\tau^2/dt^2 = 1 - v^2/c^2$, where $v^2 = v_r^2 + v_t^2$, the square of the vector sum of the radial and transverse components of velocity v.

In curved spacetime, the vector sum differs, because radial velocity is affected by the curvature of space. It is illuminating to write the Schwarzschild solution in the following form:

$$\frac{d\tau^2}{dt^2} = g_{tt}[1 - g_{rr}^2\frac{v_r^2}{c^2} - g_{rr}\frac{v_t^2}{c^2}]. \tag{1.18}$$

This can be viewed as the product of a gravity related time dilation factor

(g_{tt}) and a velocity related time dilation factor $(1 - v^2/c^2)$. One can guess the vector summation equation from $d\tau/dt$ above as:

$$v_{lo}^2 = g_{rr}^2 \frac{v_r^2}{c^2} + g_{rr} \frac{v_t^2}{c^2}, \tag{1.19}$$

where v_{lo} is the velocity as measured by the *local observer* and v_r, v_t are the coordinate (i.e. distant observer) velocities. A local observer must be inertial (free falling) and momentarily stationary in the reference frame at the time and place of the measurement. The timelike metric can then be simply written as

$$d\tau^2 = g_{tt}(1 - v_{lo}^2)\ dt^2, \tag{1.20}$$

relating proper time and coordinate time by the product of the gravitational time dilation (redshift) and a simple velocity time dilation.

Note that despite the local velocity being measured by a locally stationary observer, the equation gives the rate of the locally moving clock ($d\tau$) as a function of the rate of the coordinate clock (dt).

This shows very clearly that special relativity is a special case of general relativity. Special relativity rules when the gravitational field is weak or absent ($g_{tt} = 1$)—then it is only velocity time dilation that occurs.

From the above we have the very useful transformation formulae between local (v_{lo}) and coordinate (v_{co}) velocities in Schwarzschild spacetime (recall that $g_{rr} \geq 1$ and $g_{tt} = 1/g_{rr}$)

$$v_{r(lo)}^2 = g_{rr}^2\ v_{r(co)}^2, \tag{1.21}$$
$$v_{t(lo)}^2 = g_{rr}\ v_{t(co)}^2, \tag{1.22}$$
$$v_{r(co)}^2 = g_{tt}^2\ v_{r(lo)}^2, \tag{1.23}$$
$$v_{t(co)}^2 = g_{tt}\ v_{t(lo)}^2. \tag{1.24}$$

To make sense out of the velocity transformations, remember that someone with a slower clock (the local observer) will measure a shorter time and thus a higher speed than someone with a faster clock (the distant observer). This explains the transverse velocity transformation, but why the additional factor g_{rr} for radial velocities?

This is caused by the gradient of curved space. Near the source of gravity, distances appear to be 'compressed' in the radial direction, as viewed by the distant observer. Therefore, radial movement appears to the distant observer to slow down by a further factor g_{tt}, as shown in figure 1.5, where geometrized units have been used for clarity, i.e. $\bar{M} = GM/c^2$, or $G = 1,\ c = 1$.

The 'compression' of radial distances by a gravitational field can be viewed as caused partially by gravitational time dilation and partially by the gradient of curved space (actually, partially here means that both have an equal share in the outcome). The gradient of curved space at a distance r from a mass M is given by

$$\frac{dz}{dr} = \sqrt{\frac{2\bar{M}}{g_{tt}\ r}}, \tag{1.25}$$

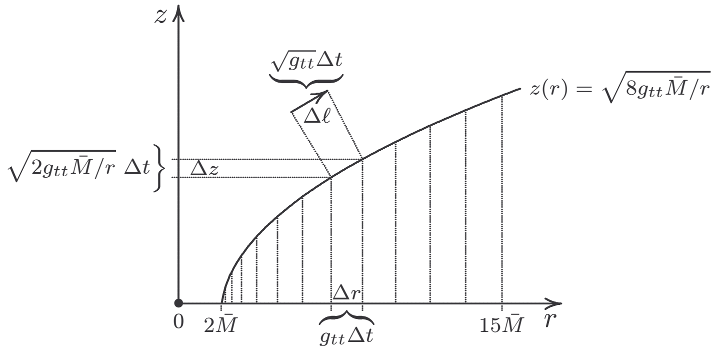

Figure 1.5: Proper radial distance increments ($\Delta\ell$) against coordinate radial distance increments (Δr), showing both gravitational time dilation (redshift) and space curvature. The segments $\Delta\ell$ represent the proper distance that light travel in time interval Δt, which become shorter closer to the origin due to gravitational time dilation. Then the gradient of the local space curve $z(r)$ causes the projection onto the coordinate radial axis to be 'compressed' further.

giving z as a function of r (after integration, with the expanded g_{tt})*

*An easy to follow derivation of $z(r)$ is given in [MTW], section 23.8.

$$z(r) = \sqrt{\frac{8g_{tt}\bar{M}}{r}} + \text{constant}, \tag{1.26}$$

as used in figure 1.5. This figure illustrates so much of the gravitational field around a static, spherically symmetric mass, that it warrants a closer look. Firstly, if space had no curvature, i.e., $z(r) =$ constant, there would still have been an apparent contraction in the radial direction, as observed by the distant observer, i.e.,

$$\Delta\ell = \sqrt{g_{tt}}\Delta t,$$

the distance that light propagates in coordinate time interval Δt.

If clocks slow down near the central mass, then so does the propagation of light, at least as viewed by the distant observer. A local observer cannot detect this 'slowing down' of light, because if your clock and your measuring rod* changes precisely in step, you cannot detect the 'slowing down'.

*All distance measurements are directly or indirectly based on the speed of light.

A distant observer can, in principle, detect the 'slowing down' of light by measuring the round trip travel time of light, beamed from a large distance to the mass and being reflected back from the surface of the mass.*

This is the effect of gravitational redshift alone. Because of the gradient of curved space, the effective movement of light in the radial direction is still 'slower' than that. From the gradient equation and since $dr^2 = d\ell^2 - dz^2$ (see figure 1.5), it is easy to show that

$$dr = \sqrt{g_{tt}}\ d\ell = g_{tt}\ dt, \tag{1.27}$$

*A black hole will not reflect light, but a neutron star will work nicely.

the projection of $d\ell$ onto the coordinate radial distance axis. Therefore, the radial 'contraction' factor of curved space is identical to the contraction factor of gravitational redshift, both being $\sqrt{g_{tt}}$.

This means that as far as distant observers are concerned, light moving precisely radially relative to a central mass 'slows down' to $g_{tt}c$.

Light moving (momentarily)* in a purely transverse direction relative to a

*There is a case where light can be in orbit around a black hole, but more about that later.

mass, is slowed down just by the gravitational redshift factor, to $\sqrt{g_{tt}}\ c$.

Since the full effect can only be measured indirectly, it is not called a 'slowing down of light', but rather a 'delay of light'.*

*It is called the *Shapiro delay*, after the man who first measured it accurately, as is discussed further in chapter 5.

1.4 Summary of this introduction to relativity

We have seen that in the gravity-free space of special relativity, there is a quantity called the spacetime interval that is invariant, i.e., it has the same value, no matter which inertial observer measures the components of the interval.

This lead us to the conclusion that an observer that is present at two spacetime events will always measure a time interval that is shorter than what any observer that is not present at both events will measure. This time interval (measured by the observer present at both events) is called the propertime between two events.

The more formal view of the spacelike and the timelike interval was then derived, using just a sprinkling of tensor algebra. We then moved on to a fairly loose discussion of Einstein's field equations and an even more loose derivation of the metric for the gravitational field was presented.

This lead us to the (disturbing?) realization that the speed of light varies in Schwarzschild coordinates with the direction of movement—at least as measured by a distant observer. This forced us to accept that distant observers will measure different velocities than what local observers will measure. There are however relatively simple transformation equations for velocities as measured locally and remotely.

Underlying to this is the fact that time and distance are measured differently by distant clocks and local clocks. In particular, distant clocks runs faster than local clocks and the difference becomes more noticeable if the local clock is moving relative to the coordinate system.

Now, to move forward, we will discuss a few topics in special relativity that are of fundamental importance—clock synchronization, energy, momentum and Doppler shift. They are the sort of things that many engineers use daily and may perhaps sometimes wonder how relativity influences their work.