Background

There is a well-known legend that a reporter once asked the great British Astrophysicist Sir Arthur Eddington “Is it true that only three people in the world understand Albert Einstein’s theory”, and Eddington, with becoming modesty, replied “Who’s the third?”. The mathematical abstruseness of General Relativity is all part of the mystique that surrounds Einstein. It seems a reasonable supposition that a theory so famously difficult and abstract cannot have any measurable and important effect on everyday life.

This is not so, which I suspect is a surprise to most people. General Relativity is used every day in the third most ancient use of mathematics (navigation). One of the effects of General Relativity — gravitational time dilation — has a major impact on the Global Positioning System (GPS), used in more or less all (commercial and military) air and sea transport.

The GPS network can locate you anywhere on the earth’s surface to within 100 metres and give your height to within about 150 metres (or rather less if you’re the American Defence forces: the signal is deliberately degraded for commercial users) — for a few hundred dollars, one can buy a hand held GPS receiver.

GPS works by communicating with at least 4 of 24 satellites, orbiting about 20 thousand kilometres above the surface of the earth, using time-tagged, coded microwave signals from each of the four satellites. In order to translate time to distance, you multiply by the speed of the microwave signal, which is the same speed as light: 3×10^8 metres a second. To get distances to 100 metres you need to measure time to 233×10^{-9} seconds (233 nanoseconds).

One of the consequences of Special Relativity the faster you move, the slower time passes. Now the GPS satellites orbit every 12 hours and we on the surface orbit every 24 hours, so the satellites are moving faster than we are.

On the other hand, one of the consequences of general relativity is that gravity changes the rate that time passes — if gravity is stronger, time passes more slowly. In fact, a black hole can be considered a place where gravity is so strong that time stands still. (There was an interesting episode of the TV programme Stargate SG-1 shown in Australia in late 1999 that had this effect down pat: the ‘stargate’ opened close to the black hole and while in the control room a few hours passed the Pentagon had days to think about what to do).

At over 20 thousand kilometres above the Earth, the clocks on the GPS satellites (four on each: two Caesium, two Rubidium) are going to be running just a little faster than clocks down here on the Earth, as the Earth’s gravity is weaker at the satellites.

The two relativistic effects — weaker gravity speeding up the satellite clock and faster motion slowing it down (relative to the surface) — only partly cancel with one another. I mean to go through the calculations so we can see what difference relativity makes.

In practice, the satellites are occasionally reset to match ground stations to within 1 microsecond of the reference time known as gps time. This also has the effect of cancelling out all the other errors (such as orbit perturbations, refraction of the microwave signal, terrestrial oblateness) that affect the clock rate.
Special Relativity

Special Relativity is for physics without gravity: mathematically, it is linear algebra in the affine space $\mathbb{R}^4$ with the indefinite inner product (in Cartesian-like coordinates)

$$\langle X, Y \rangle = c^2 X_0 Y_0 - X_1 Y_1 - X_2 Y_2 - X_3 Y_3.$$  

(Or equivalently, swap the positive and negative signs around.) Here $c$ is the speed of light. This form of the Minkowski inner product applies to those observers who are inertial: Newton’s first law holds in their coordinate system. We call $\mathbb{R}^4$ with this inner product Minkowski space-time $M^4$. Points in $M^4$ are called events: they are the fundamental objects in relativity.

There is an article on Special Relativity in the March 2000 American Mathematical Association Monthly that takes this viewpoint.

Physically meaningful quantities are those (and only those) formed from this geometry. Compare Euclidean case: “diagonals of a rhombus are perpendicular” is a theorem of Euclid and can be written totally in terms of the Euclidean inner product on $\mathbb{R}^2$; “projection of this line onto the $x$-axis in 2 units long” is not a geometrically invariant statement. The former is preserved under arbitrary orthogonal transformations, the latter is not.

Similarly (and this is the whole point of Special Relativity from a mathematician’s viewpoint), a statement like “this experiment takes 5 seconds” is not formed using just the geometry of Minkowski space — it is coordinate dependent — and so is not physically meaningful on its own. This statement is not preserved under the general transformations that preserve the Minkowski inner product.

The most important difference between this Minkowski inner product and the usual Euclidean one is that the “square norm” $\langle X, X \rangle$ of a non-zero vector can have any sign. The non-zero vector $X$ is said to be

- **timelike** if $\langle X, X \rangle$ is positive.
- **spacelike** if $\langle X, X \rangle$ is negative.
- **null** if $\langle X, X \rangle$ is zero.

The point of this terminology is that a particle with mass (e.g. a clock) moves on a curve whose tangent vector is timelike, and a light ray follows a curve whose tangent vector is null.

An **ideal clock** is a clock that is unaffected by acceleration: its rate depends only on its (instantaneous) speed. Such things do seem to exist, as tests of Special Relativity using particles moving in a circle at close to the speed of light (with accelerations of the order of $10^{19} g$: CERN 1968) bear out the theory’s predictions.

A clock sitting at $X_1 = X_2 = X_3 = 0$ reads time equal to $X_0$. For this reason we usually label the 0-coordinate $t$ and call it coordinate time. The other directions are space coordinates.
On of the most famous consequences of SR is that a moving clock runs slowly. In fact, a clock moving on a curve \( r(\lambda) \) measures a proper time \( \tau \) defined

\[
c^2 \left( \frac{d\tau}{d\lambda} \right)^2 = \left\langle \frac{dr}{d\lambda}, \frac{dr}{d\lambda} \right\rangle.
\] (1)

Note that the right hand side here is positive for a real clock. In the instantaneous rest frame when \( r = (t(\lambda), 0, 0, 0) \) we get \( \tau = t \). We must then use equation (1) in a moving frame, as (1) is geometrically invariant.

We can now see immediately why a moving clock runs slowly: if any spatial component of \( \frac{dr}{d\lambda} \) increases, \( \frac{d\tau}{d\lambda} \) decreases. This is time dilation: a clock runs fastest in its rest frame. It’s really just a velocity perspective effect (as is its dual, length contraction).

If two observers pass each other at the one event \( E \) then the inner product of their 4-velocities is the “beta factor” that gives relative time dilation.

A GPS satellite, for example, travels in a roughly circular orbit at about 3870 ms\(^{-1}\) — satellite clocks would be running slower than earth clocks. At that speed proper time is slower than coordinate time \( \tau \approx (1 - 8.33 \times 10^{-11})t \).

This doesn’t seem like much, but in the course of a day (86 400 seconds) this comes to 7200 nanoseconds, 30 times the required accuracy, or 2 km off target.

**General Relativity**

General relativity is Einstein’s theory of gravity. What does it say about time measurement?

Another of the legends of relativity is that one day Einstein spoke to a man who had survived falling off the roof of a building. Einstein asked the man if he had felt the force of gravity as he fell and was told ‘no’. This is supposed to have lead Einstein to the realisation that gravity is a fictitious force, in the same sense as centrifugal and coriolis forces are: they are all accidents of the coordinate system and are caused by inertia.

One can consider a body to have three different masses:

1. the mass that resists motion: the **inertial mass**. This is the \( m \) that appears in Newton’s second law \( \mathbf{F} = m\mathbf{a} \).
2. the mass that causes gravitational attraction: the **active gravitational mass**. This is the \( m \) that appears in the potential for the Newtonian gravitational field: \( \phi = -\frac{Gm}{r} \).
3. the mass that reacts to gravitation: the **passive gravitational mass**. This is the mass that appears in the force law \( \mathbf{F}_{\text{grav}} = -m \nabla \phi \).

Probably the best attested result in physics is that these three masses are all the same. Newton’s third law forces active and passive gravitational masses to be the same (with appropriate units). Passive gravitational and inertial masses were first proposed as the same (independent of composition) by Galileo who (reputedly) dropped two different masses from the leaning tower of Pisa. He observed the accelerations to be (roughly: neglect air resistance) the same, from which we get the equality of passive and inertial masses. In rather more subtle experiments Baron von Eötvös (ut’vush) in 1922 showed that the difference is less than 1 part in 10\(^9\), and more recent experiments by Braginski and co-workers in Moscow (1971) showed the difference to be less than 1 part in 10\(^{12}\) (lunar laser ranging experiments have shown that the moon and earth fall to the sun with the same acceleration to better
than 7 in $10^{13}$). The fact that these masses are the same leads us to suppose that there is something fundamental at work, rather than a coincidence.

If we assume that this equality is no accident, we are lead to the conclusion that gravity is an inertial force. This **Principle of Equivalence** can be stated in several ways, one being

*A frame linearly accelerated with respect to an inertial frame in SR is locally identical to a frame at rest in a gravitational field.*

In this statement, “locally” is a deliberately rubbery term. What local means depends on the variation of the gravitational field and just how accurately it can be measured: in a room the size of a lecture theatre, you’d need very delicate equipment to detect the variation on the direction of the Earth’s gravitational field between opposite sides (although the variation does exist). In mathematical terms, what the principle is saying is that the tangent space at an event is $\mathbb{M}^4$; after all, a smooth manifold is locally the same as its tangent space (locally, the Earth is $\mathbb{R}^2$).

The simplest way to tie the principle of equivalence to motion is to assume that rather than gravity being a field stuck “on top of” a special relativistic background, it in fact arises from space-time being not flat (as in SR) but curved. The equation of motion of a free particle in an inertial system in SR is just

$$\frac{d^2 X}{d\tau^2} = 0.$$  

If we change to a non-inertial (e.g. rotating) coordinate system, then this equation alters to include terms that are precisely the inertial forces (centrifugal, coriolis etc). The principle of equivalence requires that gravity should appear in exactly the same way, and the simplest generalisation is to assume that space-time is no longer flat but curved.

This changes the definition of proper time as the “inner product” on a curved manifold is a (position dependent) matrix $G$ (the **metric**). The metric which tells us how space-time is curved. So a body moving on a curve $r(\lambda)$ measures a proper time $\tau$ where

$$c^2 \left(\frac{d\tau}{d\lambda}\right)^2 = d\tau T G \frac{dr}{d\lambda}.$$  

Again, the right hand side is positive (that is $dr/d\lambda$ timelike) for a particle with mass.
Exact calculation

So what metric do we use the model space-time outside the earth? We assume that the earth is perfectly spherical, ignore rotation (whose relativistic effects are not noticeable at the accuracy of our data) and also ignore everything else in the Universe. This gives us the Schwarzschild metric, which in spherical coordinates \( \{t, r, \phi, \theta\} \) is

\[
\begin{pmatrix}
(1 - 2m/r)c^2 & 0 & 0 & 0 \\
0 & -(1 - 2m/r)^{-1} & 0 & 0 \\
0 & 0 & -r^2 & 0 \\
0 & 0 & 0 & -r^2 \sin^2 \phi
\end{pmatrix}
\]

where \( m = GM/c^2 \), for \( M \) the mass of the earth, \( G \) Newton’s Gravitational constant. We usually write this is the suggestive form

\[
ds^2 = \left(1 - \frac{2m}{r}\right) c^2 dt^2 - \left(1 - \frac{2m}{r}\right)^{-1} dr^2 - r^2 \left(d\phi^2 + \sin^2 \phi d\theta^2\right),
\]

where \( ds \) is an infinitesimal proper displacement as a function of infinitesimal displacements in the coordinates.

Karl Schwarzschild found this metric in 1915 within a few weeks of Einstein’s announcement of General Relativity, rather to Einstein’s surprise as he thought no-one would ever find an exact solution. Sadly, Schwarzschild (who was in the German army) was killed on the Russian front in 1916.

The apparent problem with the Schwarzschild metric at \( r = 2m \) will not concern us, as for the Earth \( m \approx 4.434 \times 10^{-3} \) metres, so \( r = 2m \) is well inside the earth’s surface. For more compact bodies, this surface does have an important role: it will be the event horizon for a black hole, the place where “time stands still”.

In order to calculate the difference in proper times between us on the ground and a GPS satellite in orbit, what we need is the ratio

\[
\frac{d\tau_{\text{ground}}}{d\tau_{\text{satellite}}} = \frac{d\tau_{E}}{d\tau_{s}}.
\]

Now since we are ignoring rotation, we may assume wlog that the satellite orbit is equatorial (\( \phi = \pi/2 \)), and we assume circular too of radius \( r_s \). So the path of the satellite is \( r_s(\lambda) = (t(\lambda), r_s, \pi/2, \theta(\lambda)) \). We’ll assume the receiver follows the path \( r_E(\lambda) = (t_E(\lambda), r_E, \pi/2, \theta_E(\lambda)) \), for \( r_E = 6371 \)km the mean radius of the earth.

Now the satellites (and we on the ground) move at constant angular velocity: we take 24 hours to orbit once, and the satellites take 12 hours (actually, it’s 12 hours and 5 minutes). Whose hours are these, because they’ll be different? Is that coordinate \( t \) hours, proper hours on the ground or proper hours as measured by the appropriate body?

In fact, it’s the proper time on the ground, but I’m going to assume that these hours are measured in coordinate time \( t \). To the accuracy we can manage with the data we have it makes no actual difference, and with coordinate time, the calculations are marginally simpler. This means that for the satellite \( d\theta/dt = 2\pi/(12\times86400) = \Omega_s \) and for the receiver \( d\theta/dt = 2\pi/(24\times86400) = \Omega_E \).
The only other number we now need is the orbit radius of the satellite. Happily, Kepler’s laws are still valid (at least formally) in this situation, so we know that \( r_s^3 \Omega_s^2 = GM \), and this gives us \( r_s = 26610 \text{km} \) (about 20 200 km above the surface).

So, using coordinate time as the orbit parameters we get

\[
\left( \frac{d\tau_E}{dt} \right)^2 = \left( 1 - \frac{2m}{r_E} \right) c^2 - r_E^2 \Omega_E^2,
\]

and

\[
\left( \frac{d\tau_s}{dt} \right)^2 = \left( 1 - \frac{2m}{r_s} \right) c^2 - r_s^2 \Omega_s^2.
\]

Thus the (constant) ratio giving us the difference in proper times is

\[
\frac{d\tau_E}{d\tau_s} = \sqrt{\frac{(1 - 2m/r_E)c^2 - r_E^2 \Omega_E^2}{(1 - 2m/r_s)c^2 - r_s^2 \Omega_s^2}} = 1 - 4.472 \times 10^{-10}.
\]

So we see that the earth clocks are running slower than the satellite clocks: the gravity effect is stronger than the special relativity effect. We can tease out the pure gravity effect by setting the \( \Omega \) terms to zero if we want.

However, we want to multiply the deficit \( 4.47 \times 10^{-10} \) by 86 400 and then by \( c = 2.99 \times 10^8 \) to get how far off course we’d be after a day. We do this and get 11.5 km.

Alternatively, we can ask how long it would take to be 500m off. We calculate

\[
\frac{500}{c \times \text{deficit}} \approx 3720 \text{s}
\]

— about an hour, which is too short a time to rely on re-setting by ground based clocks.

John Steele
13-April-2000