Cosmology and Relativity

AA3053

Available as a module for the *BSc in Astronomy*.

Sample Notes

These sample pages from the Course Notes for the *Relativity* section of the module *Cosmology and Relativity* have been selected to give an indication of the level and approach of the course. They are not designed to be read as a whole, but are intended to give you a flavour of the syllabus, style, diagrams, images, equations, mathematical content and presentation. They are a subset of the navigable on-line version of the learning materials.

All the Course Notes and other learning materials are made available via the course website.

- All sections of notes will be available in modest colour and basic navigation in pdf format suitable for downloading and printing at home. It is anticipated that most students will prefer to use the notes in the colour pdf files on the CDRom

July 2008

1.4 Consequences of the Lorentz transformations.

The new transformations lead directly to consequences without parallel in our everyday experience and which appear to be at variance with "common sense". So startling are the predictions that some notable intellects could never accept them. The author of one University text book on Special Relativity later came to the view that the theory must be incorrect, and the preface of the second edition of his book contained the somewhat bizarre warning that the theory described herein must now be regarded as wrong! The same author pursued an energetic but somewhat fruitless correspondence in the scientific journal *Nature* and in the National press, arguing that, in accepting special relativity, physicists had made a dreadful and possibly dangerous blunder.

These particular objections were rooted in a refusal to accept the relativity of time, as required by the Lorentz time transformation. However, this transformation was forced upon us by the need to make sense of experimental results and in science, it is experiment rather than philosophy that is the final arbiter. Though surprising, many of the consequences of special relativity have been confirmed by experiment, and there are no experimental results that contradict the theory.

Some of the most startling consequences of these transformations were described in Section 8.3.4 of EMU. The discussion that follows will consider some of the same phenomena in a little more detail.

1.4.1 Causality: the speed of light as a speed limit.

If there existed two inertial frames of relative speed greater than c it is clear that the Lorentz factor $\gamma = (1 - v^2/c^2)^{-1/2}$ would be an imaginary number and the coordinates in S$'$ of a real point in S would also be imaginary. This is clearly impossible and demonstrates that, for consistency with postulate 2, the relative speed of any two inertial frames must be less than the speed of light.

There are also other senses in which c represents a speed limit. In Newtonian physics where time is absolute, it is transparently obvious that cause must precede effect in all inertial frames. In special relativity, the requirement that cause must follow effect must be imposed. If event A causes event B as observed in S, then A must occur first in *all* inertial frames. Failure to impose this restriction would raise the possibility of influencing events that occurred in the past, leading to all manner of logical problems!

Suppose A and B occur on the x axis. Since A occurs first, $\Delta t = t_B - t_A$ is positive. Causality requires that Δt should be positive in all inertial frames. In one such frame, S$'$, Δt can be found from the Lorentz transformations:

$$\Delta t' = \gamma(\Delta t - v\Delta x/c^2)$$

The requirement that $\Delta t' > 0$ translates into the condition

$$u = \frac{\Delta x}{\Delta t} < \frac{c^2}{v} \tag{27}$$

where u is clearly the speed of a signal which travelled from event A arriving at event B and so can be regarded as the signal which mediated cause and effect. If condition 27 is to be met for *all* inertial frames, i.e. for all values of v from zero to speeds approaching c, then the most stringent condition (for v approaching c) requires

$$u < c \tag{28}$$

It follows that the need to preserve causality forbids the possibility of signals propagating at speeds in excess of the speed of light.

Problem 2: (i) The phase velocity of light for some types of wave (e.g. for radio waves in a plasma) exceeds the speed of light. Can such waves lead to violations of causality? (ii) A beam from a distant light house illuminates a spot on some cliffs. If the light spins fast enough, the spot can move at speeds greater than c. Can this lead to violations of causality?

Problem 3: Two events are separated by Δx, Δy, Δz and Δt in S and $\Delta x'$, $\Delta y'$, $\Delta z'$ and $\Delta t'$ in S′. Show that $(c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 = (c\Delta t')^2 - (\Delta x')^2 - (\Delta y')^2 - (\Delta z')^2$.

Problem 3 shows that the quantity $(\Delta s)^2 = (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2$ is an *invariant* quantity. It is known as the *interval* between two events. Pairs of events for which $(\Delta s)^2 > 0$ can be connected by a signal propagating at speed $u = [(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2]^{1/2}/\Delta t < c$ and can therefore be regarded as cause and effect, i.e. they are *causally related*. The interval between them is classed as a *time-like* interval and by the invariance of the interval it is time-like in all inertial frames. Pairs of events for which $(\Delta s)^2 < 0$ are too widely separated to be connected by a causally permitted signal (i.e. one propagating at speed below c) and therefore cannot be regarded as cause and effect. The interval separating such events is described as *space-like*.

At this point it is interesting to consider the astronomical phenomenon of **superluminal motion**. At the centre of many active galaxies (galaxies that emit mainly non-stellar radiation) are highly collimated beams of energetic particles and magnetic field. These emit synchrotron radiation which is seen chiefly at radio frequencies. Using large arrays of radio telescopes the motion of bright patches in these jets can be seen and their angular speed on the sky can be measured. By obtaining the distance of the source from its redshift, the angular speed can be translated into a linear speed and, remarkably, these often turn out to be in excess of the speed of light. When this phenomenon was first observed some commentators suggested that the sources must be closer than their redshifts imply, bringing the linear velocity (which is proportional to their distance) below c. In that case the redshifts would have to be non-cosmological. However it was soon realised that this phenomenon can be explained (retaining cosmological redshifts) if the jets are aligned at a small angle to the line of sight and their speed is very close to (but still less than) c. To see how this is possible, consider a source B ejected at time $t = 0$ from the nucleus A of a galaxy at angle θ to a remote observer's line of sight (as shown in Fig. 3).

At time $t = 0$ the source B emerges from the nucleus A, and emits a ray of light toward the observer as it does so. After time t, B has travelled a distance vt and has advanced a distance $vt\cos\theta$ toward the observer. In the same time the ray emitted at $t = 0$ has travelled distance ct toward the observer, so the source is effectively a distance $ct - vt\cos\theta$ behind the ray. The time interval between the moments when the observer sees the original ray emitted by B at $t = 0$ and a ray emitted at time t is therefore $\Delta t = (ct - vt\cos\theta)/c$. In the same time the observer sees the source moving a projected distance $\Delta y = vt\sin\theta$. The apparent speed is therefore given by

$$\begin{aligned} v_{app} &= \frac{vt\sin\theta}{t(1 - v\cos\theta/c)} \\ &= \frac{v\sin\theta}{1 - v\cos\theta/c} \end{aligned} \tag{29}$$

The maximum value of this apparent velocity occurs for $\sin\theta = 1/\gamma$ (or equivalently $\cos\theta = v/c$) for which $v_{app} = \gamma v$. Hence if the speed of the source is close to c, the apparent speed can be as high as γ times the speed of light! A graph showing the variation of v_{app} with angle θ is shown for $\gamma = 5$ on the right of Fig. 3.

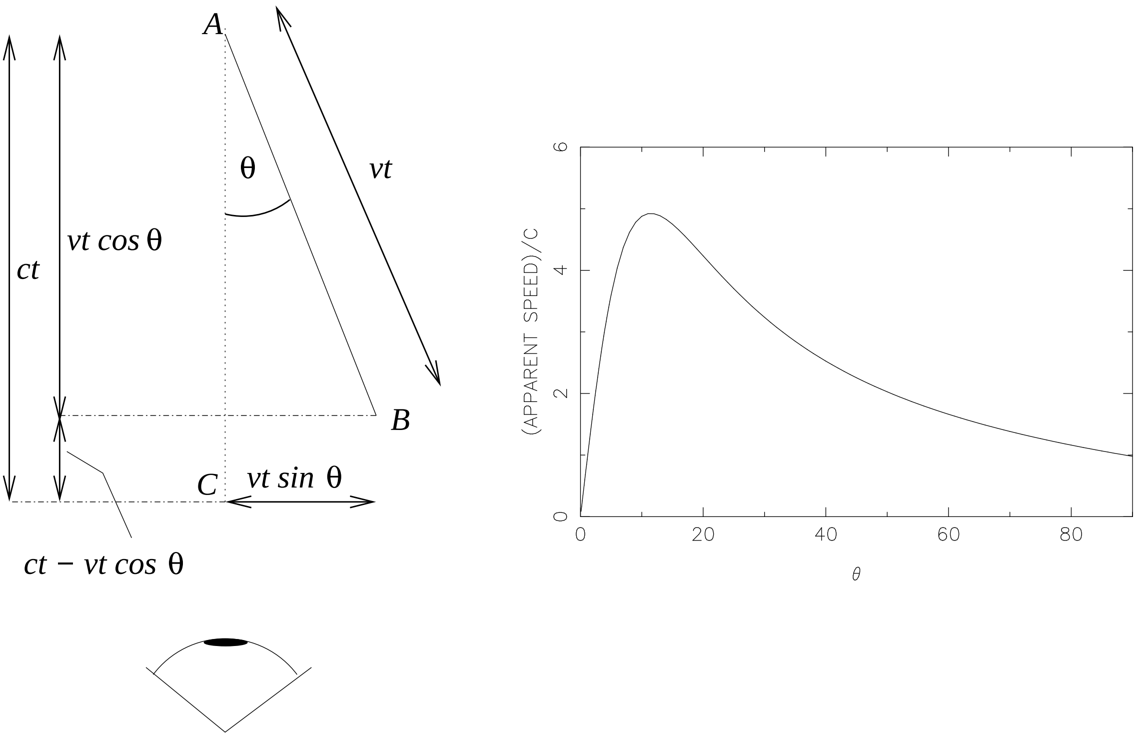

Figure 3: Left: Figure illustrating the explanation of superluminal motion. Source B leaves the nucleus A of a galaxy at $t = 0$ moving with speed v at angle θ to the line of sight of a remote observer. At time t later, the source is only a distance $ct - vt\cos\theta$ behind the radiation emitted at $t = 0$. Right: v_{app} is plotted as a function of θ (in degrees) for $\gamma = 5$.

Note that the since the entire calculation is done in one frame of reference, the Lorentz transformations have not been used. From the point of view of relativity, the phenomenon is of interest because if the speed v of the source B exceeded c, equation 29 shows that the apparent speed could be negative, i.e. the source would be seen to move inward. The fact that the observed motion is always outward confirms the view that in nature, real signals preserve causality by obeying the speed limit c.

There are lots of animated sequences illustrating superluminal motion on Matt Lister's website: www.physics.purdue.edu/astro/MOJAVE/movies.html. You can download these to your computer and watch them using standard software. The sequence for the quasar 0923+392 is particularly good: there are several clear ejections from the galactic nucleus (on the left of the image) that cause brightening of the extended structure.

1.4.2 Time Dilation

The Lorentz transformations relate the space and time coordinates of events as recorded in different frames of reference. In order to understand just what these coordinates represent it is useful to think about how they might be measured. We can imagine both S and S′ to be populated by many 'observers' at rest in their frames. All observers are equipped with clocks that are synchronised throughout each frame. This synchronisation could be accomplished by a light flash emitted from the origin at $t = t' = 0$. Knowing the speed of light and its own coordinates, each observer could set its clock to read the time $(x^2 + y^2 + z^2)^{1/2}/c$ (or the equivalent in S′) at the instant the flash of light is seen. In this way, the coordinates of events can be recorded in a manner which eliminates the effect of time delays arising from propagation of light (or other signals) from the position of an event to the position at which it is recorded.

> **Problem 4**: Think of another means of sychronising clocks throughout an inertial frame.

Consider two events which occur at the same position in S′ with coordinates (x', y', z', t'_1) and (x', y', z', t'_2).

Using the Lorentz transformation equation 22, the time coordinates of these events in S are

$$t_1 = \gamma(t'_1 + vx'/c^2)$$
$$t_2 = \gamma(t'_2 + vx'/c^2)$$

Differencing these two equations gives the result

$$\Delta t = t_2 - t_1 = \gamma(t'_2 - t'_1) = \gamma \Delta t'$$

or

$$\Delta t' = \frac{\Delta t}{\gamma} \tag{30}$$

This equation shows that *the time interval between two events is shortest in the inertial frame in which they occur at the same position* (assuming that such a frame exists). In other frames the time interval is greater by a factor of γ.

This effect, known as *time dilation*, has been verified experimentally. One manifestation of this effect is the survival of cosmic ray muons created in the upper atmosphere to reach the Earth's surface. Muons are electron-like particles, but more massive. In the laboratory they have been shown to decay with half-life about $1\,\mu$s, whereas the muons take about $6.5\,\mu$s to reach the Earth's surface during which all but a fraction $(0.5)^{6.5} \simeq 0.01$ should decay. Yet a combination of high and low altitude observations shows that most of the muons survive to reach the Earth. The explanation is that the muons travel at speed very close to the speed of light with $\gamma \simeq 10$, and in the rest frame of the muons, the time elapsed is therefore shorter by a factor of 10, i.e. less than one half-life. For more details of this fascinating experiment see French, pages 97-99.

The time dilation effect has also been measured directly using highly accurate maser clocks flown on jet aircraft. As these aircraft move at $\simeq 150\,\mathrm{ms}^{-1}$ the Lorentz factor is approximately $1 + (1.25 \times 10^{-13})$, and so the difference in reading between the airborne and stationary clocks at the end of the flight is extremely small, but measurable.

A graphic illustration of the effects of time dilation is provided by the case of two twins, one of whom (A) remains stationary, whilst the other (B) goes on a return journey at a speed approaching the speed of light.

Suppose that B travels in a straight line with a Lorentz factor of 10 for a time of 4 years as measured on his own clocks. In the rest frame of A the time interval between the beginning and end of the journey is (using the time dilation equation 30) 40 years. B then reverses his journey and travels at the same speed back to A. At the end of the journey another 4 years has elapsed in the frame of B while another 40 years has elapsed in the frame of A. During the entire trip, A has aged by 80 years whereas B has aged by only 8. (You might reasonably object that we have ignored the effects of the accelerations and decelerations experienced by B, however the effects of these on the times elapsed can be ignored as small effects provided they occupy only a small part of the journey time.)

This curious result, in which the twins, initially the same age, acquire radically different ages, is sometimes (incorrectly) referred to as the 'twins paradox'. It could be argued that from B's point of view it is A that makes the round-trip and therefore A that should age less rapidly.

2 Relativistic Mechanics

2.1 The need for a new momentum

Fundamental laws of physics must be invariant under the Lorentz transformations, i.e. they must apply equally in all inertial frames. The most fundamental law of mechanics is the law of momentum conservation that lies at the root of Newton's laws of motion. This section demonstrates that the Lorentz transformations require a new definition of momentum if the conservation of momentum is to be an invariant law, following the section headed "Two views of an inelastic collision", page 172, from Chapter 6 in French's book.

Consider the inelastic collision of two particles of equal mass m. In S, one mass is stationary, the other has velocity U; after they collide, the two masses adhere and move together with velocity u. The collision can also be viewed from the frame S′ in which the two masses move with equal but opposite velocities and therefore equal and opposite momenta. This frame of reference, in which the total momentum of the interacting particles is zero, is known as the 'Centre of Mass' (or sometimes 'Centre of Momentum') frame for a given system of particles. In this frame the post-collision mass must be at rest if momentum is conserved. The collision is shown in Fig. 1. The particle moving to the left as seen from S′ is at rest in S, so in S′, the

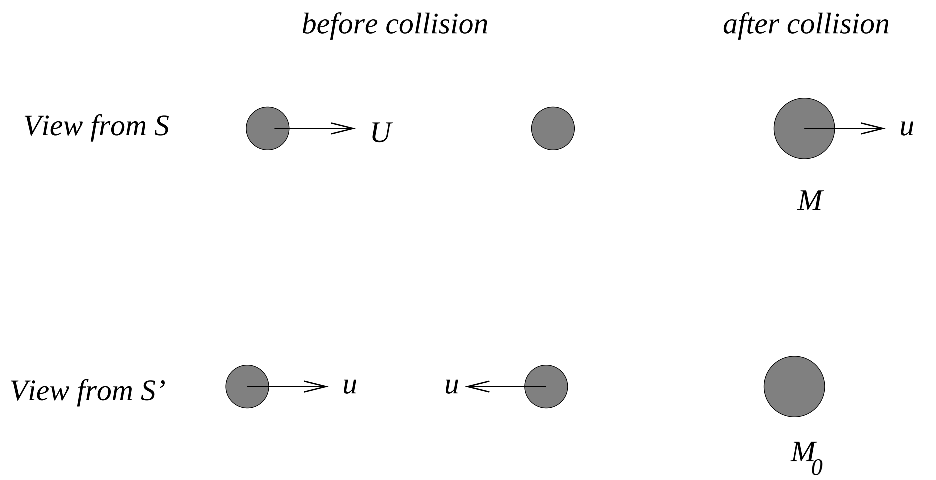

Figure 1: The inelastic collision of two particles of mass m. Left: before collision; Right: After collision. Top: the view from S; bottom the view from S' (the Centre of Mass frame).

frame S moves to the left at speed u. It follows that in S, the S′ frame is moving to the right at speed u. From the transformation for the x component of velocity (Section 1.5), the velocity U of the moving, pre-collision particle in S is

$$\begin{aligned} U &= \frac{u'_x + v}{1 + vu'_x/c^2} \\ &= \frac{u + u}{1 + u^2/c^2} \\ &= \frac{2u}{1 + u^2/c^2} \end{aligned} \qquad (1)$$

In Newtonian mechanics, mass is conserved in the collision so that the mass of the post collision particle is $2m$. In the centre of mass frame, S′, Newtonian momentum has been conserved by assumption that the post-collision particle is stationary. In S, pre-collision momentum is $mU = 2mu/(1 + u^2/c^2)$, but the post-collision momentum is $2mu$. This demonstrates that Newtonian momentum mv is only approximately conserved in S in the limit $u \ll c$. Therefore conservation of Newtonian momentum is, in general, not an invariant law and therefore cannot be regarded as a fundamental law of physics. (It only takes one counter-example to disprove the law!)

The minimal change in the definition of momentum required to reinstate the conservation law is to allow the mass of the particle to be a function of its velocity. Therefore in S, the mass of the moving particle becomes $m(U)$. The mass of a particle at rest is called the rest mass, usually indicated by a suffix of zero, so that $m(u = 0) = m_0$. In the problem considered here, the condition that the pre-collision masses are equal becomes a requirement that the *rest masses* are equal.

It is easy to see that, according to this new definition, momentum is conserved in S′: once again, the total momentum is zero before and after collision. In S, conservation of momentum gives

$$m(U)U \quad = \quad M(u)u \tag{2}$$

while conservation of mass gives

$$m(U) + m_0 = M(u). \tag{3}$$

The aim, here, is to determine how $m(U)$ varies with U, or equivalently obtain $m(U)/m_0$ as a function of U. Eliminating $M(u)$ between equations 2 and 3 gives

$$\frac{m(U)}{m_0} = \frac{u}{U - u} \tag{4}$$

so it remains to express the right hand side of equation 4 in terms of only U. The relationship between u and U is given by equation 1, which can be expressed as the quadratic equation

$$u^2 - (2c^2/U)u + c^2 = 0$$

which can be solved using the quadratic formula to give

$$\begin{aligned} u &= \frac{c^2}{U} \pm \left[\left(\frac{c^2}{U} \right)^2 - c^2 \right]^{1/2} \\ &= \frac{c^2}{U} \left[1 \pm (1 - U^2/c^2)^{1/2} \right]. \end{aligned} \tag{5}$$

The ambiguity in sign can be resolved by noting that, as u approaches zero, so U must approach zero as well. Choosing the positive sign in equation 5 clearly fails to yield $u \to 0$ as $U \to 0$. When the negative sign is chosen, the square root can be expanded as a Taylor series[1], to give $(1 - U^2/c^2)^{1/2} \simeq 1 - (U^2/2c^2)$ for $(U/c) \ll 1$. Then as $U \to 0$, equation 5 gives

$$u \to \frac{c^2}{U} \times \frac{U^2}{2c^2} = \frac{U}{2} \to 0$$

as required. Therefore

$$u \quad = \quad \frac{c^2}{U} \left[1 - (1 - U^2/c^2)^{1/2} \right] \tag{6}$$

and so

$$\begin{aligned} U - u &= \frac{c^2}{U} \left[\frac{U^2}{c^2} - 1 + (1 - U^2/c^2)^{1/2} \right] \\ &= \frac{c^2}{U} (1 - U^2/c^2)^{1/2} \left[1 - (1 - U^2/c^2)^{1/2} \right] \\ &= \frac{c^2}{U} \frac{1}{\gamma(U)} (1 - 1/\gamma(U)). \end{aligned} \tag{7}$$

[1]Taylor series were described in the Mathematics Section of AA1056, Energy, Matter and the Universe.

Then, from equation 4, 6 and 7

$$\begin{aligned} \frac{m(U)}{m_0} &= \frac{1-1/\gamma(U)}{(1-1/\gamma(U))/\gamma(U)} \\ &= \gamma(U). \end{aligned} \tag{8}$$

Hence the only possible conclusion that preserves the invariance of momentum conservation is that the mass of a moving body increases in proportion to its Lorentz factor. The same conclusion is reached by considering other types of interaction, eg. the elastic scattering of two particles, as described in the section of that title in Chapter 6 of French's book (page 169).

2.2 The energy of a relativistic body

In this section, the relationships between mass, momentum and energy are obtained by assuming the mass and momentum formulae derived in Section 2.1.

For a particle moving at velocity v, momentum $p = mv = \gamma m_0 v$, the force acting on it is (by Newton's second law) the rate of change of momentum, or $F = dp/dt$.

The rate of doing work on the particle is $vF = vdp/dt$ and so the work done in time dt is $dW = F(dp/dt)dt$. It follows that the total work done between times t_1 and t_2 is

$$W = \int_{t_1}^{t_2} v\frac{dp}{dt}dt$$

This can be re-expressed using the product rule for differentiation: $d(vp)/dt = v(dp/dt) + p(dv/dt)$ or $v(dp/dt) = d(vp)/dt - p(dv/dt)$. Hence

$$\begin{aligned} W &= \int_{t_1}^{t_2} \left(\frac{d}{dt}(pv) - p\frac{dv}{dt}\right) dt \\ &= \Delta(pv) - \int_{t_1}^{t_2} pdv \\ &= \Delta(pv) - \int_{t_1}^{t_2} \frac{m_0 v}{(1-v^2/c^2)^{1/2}} dv \end{aligned}$$

where $\Delta(pv)$ is the change in pv between t_1 and t_2. Substituting $x = v^2/c^2$, from which $2vdv = c^2dx$, the integral is easily obtained:

$$\begin{aligned} W &= \Delta(pv) - \int_{x_1}^{x_2} \frac{m_0c^2/2}{(1-x)^{1/2}} dx \\ &= \Delta(pv) - \frac{m_0c^2}{2}\left[\frac{(1-x)^{1/2}}{(1/2)\times(-1)}\right]_{x_1}^{x_2}. \end{aligned} \tag{9}$$

If the particle begins from $v = 0$ and finishes with $v = V$, then $\Delta(pv) = \gamma(V)m_0V^2$, $x_1 = 0$ and $x_2 = V^2/c^2$, from which (noting that $\gamma^2v^2/c^2 = \gamma^2 - 1$)

$$\begin{aligned} W &= \gamma(V)m_0V^2 + m_0c^2\left(\frac{1}{\gamma(V)} - 1\right) \\ &= m_0c^2\left(\frac{(\gamma(V))^2V^2/c^2 + 1}{\gamma(V)} - 1\right) \\ &= m_0c^2\left(\frac{(\gamma(V))^2 - 1 + 1}{\gamma(V)} - 1\right) \\ &= (\gamma(V) - 1)m_0c^2 \end{aligned} \tag{10}$$

The kinetic energy gained by the particle is equal to work done upon it, so it follows that the kinetic energy of the particle of rest mass m_0 and Lorentz factor γ is

$$K = (\gamma - 1)m_0c^2 \tag{11}$$

or, equivalently,

$$K = (m - m_0)c^2 \tag{12}$$

where $m = \gamma m_0$ is the relativistic mass. Equation 12 suggests a very close relationship between mass and energy. If the particle's kinetic energy increases by ΔK, its mass increases by $\Delta m = \Delta K/c^2$. This is a large effect, dramatically increasing the mass of high-energy particles, and is therefore easily testable. For example the radius of the path a particle of electric charge q orbiting in a magnetic field B is (see EMU section 6.6.2)

$$r = \frac{mv}{qB}$$

and therefore particles moving at speeds approaching c will orbit in much larger radii than expected from Newtonian mechanics due to the increase in mass. This behaviour has been rigorously tested in particle accelerators where experiments involving highly relativistic particles are now a matter of routine.

Problem 1. Show that, in the limit of velocities $v \ll c$, Equation 11 yields the well-known Newtonian result, $K = m_0v^2/2$.

The equivalence between mass and kinetic energy suggested by Equation 12 led Einstein to the hypothesis that there exists a wider, more complete equivalence between energy and mass in which all types of matter and energy are, in principle, interchangeable. This idea is encapsulated by the famous equation

$$E = mc^2. \tag{13}$$

This suggests, for example, that rest mass can be transformed into energy, as has been verified by the phenomena of nuclear fusion and nuclear fission. The energy associated with the rest mass, $E_0 = m_0c^2$ is the *rest mass energy*. The total energy of a free particle, E is therefore the sum of the kinetic and rest mass energies. Equation 13 also predicts that pure energy can be converted into matter, as occurs in the phenomenon of pair creation in which an electron-positron pair is created from a gamma-ray photons. Less spectacularly, it also predicts that transitions between energy levels of an atom result in (small) changes in the mass of the atom.

The relationship between energy and momentum can be obtained from Equation 13

$$\begin{aligned} E^2 &= (mc^2)^2 = \gamma^2(m_0c^2)^2 \\ &= \frac{(m_0c^2)^2}{1 - v^2/c^2} \\ &= \frac{(m_0v)^2c^2 + (m_0c^2)^2(1 - v^2/c^2)}{1 - v^2/c^2} \\ &= (\gamma m_0v)^2c^2 + (m_0c^2)^2 \end{aligned}$$

from which, since $p = \gamma m_0v$,

$$E^2 = p^2c^2 + m_0^2c^4 \tag{14}$$

3 Four Vectors

3.1 Space-time and momentum-energy four vectors

The similarity between the relativistic transformations for space and time (from section 1) and momentum and energy (section 2) suggests the possibility of inventing a standard set of transformations applicable to both (x, y, z, t) and (p_x, p_y, p_z, E), and to other similar groups of physical parameters. The four parameters in each group can be regarded as components of a four vector. For reasons that should become clear later, the fourth component of each vector is a pure imaginary number. The space-time four vector is therefore

$$\mathbf{R} = (x, y, z, ict) \tag{1}$$

and the momentum-energy four vector is

$$\mathbf{P} = (p_x, p_y, p_z, iE/c) \tag{2}$$

where $i = \sqrt{-1}$. The transformation of general four vectors $\mathbf{Q} = (q_1, q_2, q_3, q_4)$ between inertial frames S and S$'$ is given by

$$q'_1 = \gamma(q_1 + i\beta q_4) \tag{3}$$
$$q'_2 = q_2 \tag{4}$$
$$q'_3 = q_3 \tag{5}$$
$$q'_4 = \gamma(-i\beta q_1 + q_4) \tag{6}$$

where $\beta = v/c$. As usual the inverse transformations are obtained by switching primed and unprimed coordinates and reversing the sign of v (or β).

$$q_1 = \gamma(q'_1 - i\beta q'_4) \tag{7}$$
$$q_2 = q'_2 \tag{8}$$
$$q_3 = q'_3 \tag{9}$$
$$q_4 = \gamma(i\beta q'_1 + q'_4) \tag{10}$$

It is easy to show that these transformations yield the now-familiar transformations derived earlier in this course. For example, applying equation 3 to the first component of **P** yields

$$\begin{aligned} p'_x &= \gamma(p_x + i\beta \times iE/c) \\ &= \gamma(p_x - vE/c^2) \end{aligned}$$

which is the transformation for the x component of momentum given by equation 42 in Section 2. Similarly applying equation 6 to the fourth component of **P** yields

$$iE'/c = \gamma(-i\beta p_x + iE/c)$$

from which, multiplying throughout by $-ic$,

$$E' = \gamma(E - vp_x)$$

which is the transformation for energy given by equation 41 in Section 2. The correspondence between equations 4 and 5 above and equations 39 and 40 in section 2 is obvious.

Problem 1. Substitute the four components of **R** into equations 3 to 6 to recover the Lorentz transformations of space and time.

Just as the scalar product of two ordinary (three component) vectors is invariant in Galilean relativity (for example the scalar product of a vector with itself is just the length of the vector) the scalar product of two four vectors is invariant in special relativity.

To prove this, take any two four vectors $\mathbf{Q}$ and $\mathbf{U}$. Their scalar product is

$$\mathbf{Q}.\mathbf{U} = q_1u_1 + q_2u_2 + q_3u_3 + q_4u_4 \tag{11}$$

$$\mathbf{Q}'.\mathbf{U}' = q_1'u_1' + q_2'u_2' + q_3'u_3' + q_4'u_4' \tag{12}$$

Using equations 3 to 6 to express $\mathbf{Q}'.\mathbf{U}'$ in terms of $\mathbf{Q}$ and $\mathbf{U}$ yields:

$$q_1'u_1' = \gamma^2(q_1u_1 - \beta^2 q_4u_4 + i\beta q_1u_4 + i\beta q_4u_1) \tag{13}$$

$$q_2'u_2' = q_2u_2 \tag{14}$$

$$q_3'u_3' = q_3u_3 \tag{15}$$

$$q_4'u_4' = \gamma^2(-\beta^2 q_1u_1 + q_4u_4 - i\beta q_1u_4 - i\beta q_4u_1) \tag{16}$$

Adding first Equations 13 and 16:

$$q_1'u_1' + q_4'u_4' = \gamma^2(q_1u_1 + q_4u_4)(1-\beta^2) \tag{17}$$

$$= (q_1u_1 + q_4u_4) \tag{18}$$

which, combined with equations 14 and 15 yields

$$q_1'u_1' + q_2'u_2' + q_3'u_3' + q_4'u_4' = q_1u_1 + q_2u_2 + q_3u_3 + q_4u_4$$

or

$$\mathbf{Q}'.\mathbf{U}' = \mathbf{Q}.\mathbf{U} \tag{19}$$

This result proves that any two four vectors satisfying transformation equations 3 to 10 will have an invariant scalar product.

There are two obvious cases. First, $\mathbf{R}'.\mathbf{R}' = x^2 + y^2 + z^2 - c^2t^2 = -(\Delta s)^2$, where $(\Delta s)^2$ is the *interval* between between events (x, y, z, t) and $(0, 0, 0, 0)$, which was shown to be invariant in Section 1.4.1. Second, $\mathbf{P}'.\mathbf{P}' = (p^2 - E^2/c^2) = -(m_0c)^2$ from Section 2 equation 14, verifying that the rest mass m_0 is an invariant property of a particle.

Invariant parameters can be constructed easily from scalar products of four vectors and can greatly simplify the solution of some problems, since the invariant parameter in one inertial frame can be replaced by the same parameter in a frame (eg. the centre of mass frame) in which it takes a much simpler form.

As an example, the problem of inverse Compton scattering can be tackled using four vectors. The scattering process is shown in Fig. 1. $\mathbf{K}_1$ and $\mathbf{K}_2$ are the momentum-energy four vectors of the photon before and after scattering; $\mathbf{P}_1$ and $\mathbf{P}_2$ are the momentum-energy four vectors of the electron before and after scattering. Conservation of momentum-energy is therefore expressed by the four vector equation

$$\mathbf{K}_1 + \mathbf{P}_1 = \mathbf{K}_2 + \mathbf{P}_2 \tag{20}$$

The strategy here is to simplify equation 20 by generating scalar products which are invariant. However, first, it makes sense here to isolate the most complicated term, namely $\mathbf{P}_2$ which is the only term that has contributions from both momentum and rest mass energy. Hence, rearranging yields

$$\mathbf{K}_1 - \mathbf{K}_2 + \mathbf{P}_1 = \mathbf{P}_2 \tag{21}$$

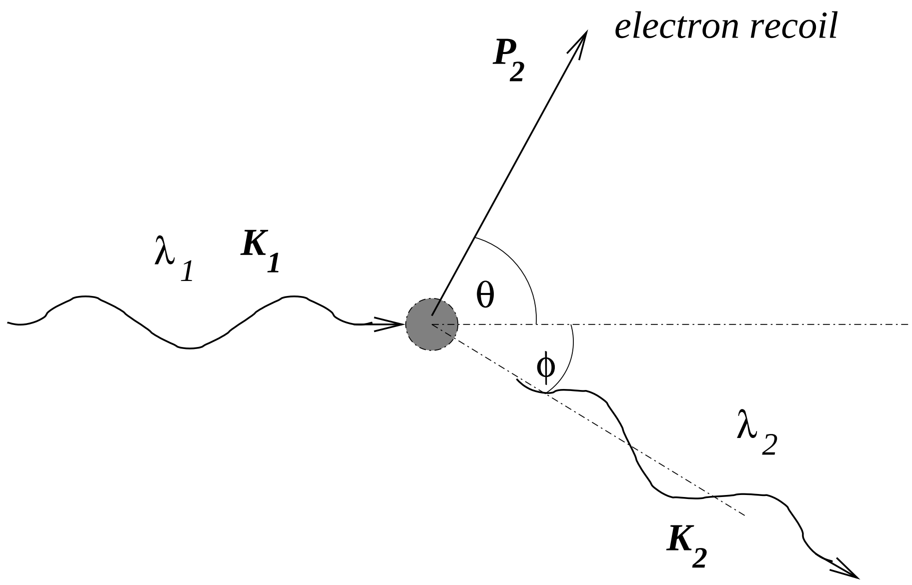

Figure 1: A photon of wavelength λ_1 is scattered by an electron. The electron recoils and the photons is redshifted to wavelength λ_2. $\mathbf{K}_1$ and $\mathbf{K}_2$ are the momentum-energy four vectors of the photon before and after scattering, and $\mathbf{P}_2$ is the momentum-energy four vector of the electron after scattering.

and then squaring both sides gives

$$\mathbf{K}_1.\mathbf{K}_1 + \mathbf{K}_2.\mathbf{K}_2 + \mathbf{P}_1.\mathbf{P}_1 - 2\mathbf{K}_1.\mathbf{K}_2 - 2\mathbf{K}_2.\mathbf{P}_1 + 2\mathbf{K}_1.\mathbf{P}_1 = \mathbf{P}_2.\mathbf{P}_2 \tag{22}$$

The square of a momentum-energy four vector is simply $(m_0c)^2$, which is zero for the photon and $(m_ec)^2$ for the electron. Therefore $\mathbf{K}_1.\mathbf{K}_1 = \mathbf{K}_2.\mathbf{K}_2 = 0$ and $\mathbf{P}_1.\mathbf{P}_1 = \mathbf{P}_2.\mathbf{P}_2 = (m_ec)^2$. Hence equation 22 reduces to

$$\mathbf{K}_1.\mathbf{K}_2 = \mathbf{K}_1.\mathbf{P}_1 - \mathbf{K}_2.\mathbf{P}_1 \tag{23}$$

The energy-momentum four vector is $(\mathbf{p}, iE/c)$ and the the four vectors $\mathbf{P}_1$, $\mathbf{K}_1$ and $\mathbf{K}_2$ are therefore given by

$$\mathbf{K}_1 = \left(\frac{h}{\lambda_1}\mathbf{u}, \frac{ih}{\lambda_1}\right) \tag{24}$$

$$\mathbf{K}_2 = \left(\frac{h}{\lambda_2}\mathbf{v}, \frac{ih}{\lambda_2}\right) \tag{25}$$

$$\mathbf{P}_1 = \left(\mathbf{p_1}, \frac{iE}{c}\right) = (0, im_ec) \tag{26}$$

where $\mathbf{u}$ and $\mathbf{v}$ are unit vectors parallel to the photon directions before and after scattering. Therefore

$$\mathbf{K}_1.\mathbf{K}_2 = \frac{h^2}{\lambda_1\lambda_2}(\cos\phi - 1) \tag{27}$$

$$\mathbf{K}_1.\mathbf{P}_1 = -\frac{hm_ec}{\lambda_1} \tag{28}$$

$$\mathbf{K}_2.\mathbf{P}_1 = -\frac{hm_ec}{\lambda_2} \tag{29}$$

(30)

where $\cos\phi = \mathbf{u}.\mathbf{v}$. Substitution in Equation 23 then yields

$$\frac{h^2}{\lambda_1\lambda_2}\cos\phi - \frac{h^2}{\lambda_1\lambda_2} = -\frac{h}{\lambda_1}m_e c + \frac{h}{\lambda_2}m_e c \tag{31}$$

from which

$$\lambda_2 - \lambda_1 = \frac{h}{m_e c}(1-\cos\phi) \tag{32}$$

as found in Section 2. The important point to notice here is how the use of invariant quantities allowed a rapid simplification of Equation 22.

As a second example of the use of four vectors in mechanics, consider the creation of pion at threshold in a proton-proton collision, another problem first solved in Section 2 (Problem 7). Recall that the threshold for particle creation occurs when the products of the reaction are at rest in the centre of mass frame. The reaction is shown in Fig 2. The momentum-energy

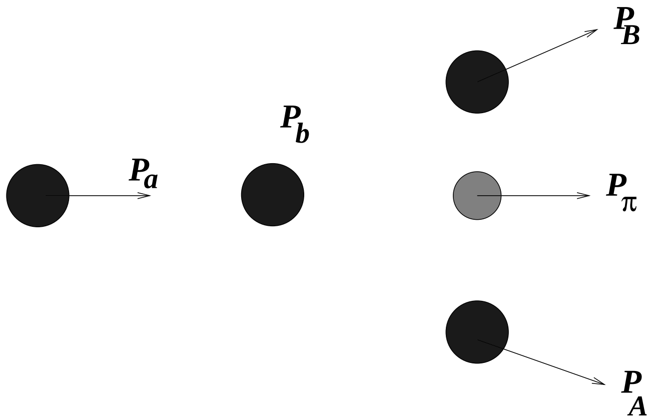

Figure 2: The collision between a moving proton and a stationary proton results in creation of a pion. The momentum-energy four vectors are $\mathbf{P}_a$ and $\mathbf{P}_b$ for the protons before collision, $\mathbf{P}_A$ and $\mathbf{P}_A$ for the protons after collision and $\mathbf{P}_\pi$ for the pion.

four vectors are $\mathbf{P}_a$ and $\mathbf{P}_b$ for the protons before collision, $\mathbf{P}_A$ and $\mathbf{P}_A$ for the protons after collision and $\mathbf{P}_\pi$ for the pion.

Conservation of momentum-energy can then be expressed through the four vector equation

$$\mathbf{P}_a + \mathbf{P}_b = \mathbf{P}_A + \mathbf{P}_B + \mathbf{P}_\pi \tag{33}$$

Squaring the equation (or forming the scalar product of each side with itself) gives

$$(\mathbf{P}_a + \mathbf{P}_b)^2 = (\mathbf{P}_A + \mathbf{P}_B + \mathbf{P}_\pi)^2 \tag{34}$$

Both sides of equation 34 are invariant quantities because they are scalar products of four vectors. They can therefore be calculated in any inertial frame of reference in sure and certain knowledge that the result will be the same in all other inertial frames. It is easiest to evaluate the left hand side in the laboratory frame, S, as illustrated in Fig. 2 while the right hand side can is most easily expressed in S′, the centre of mass frame, where (at threshold) the particles are at rest. Equation 34 can therefore be subtly recast as

$$(\mathbf{P}_a + \mathbf{P}_b)^2 = (\mathbf{P'}_A + \mathbf{P'}_B + \mathbf{P'}_\pi)^2 \tag{35}$$

4 A brief excursion into General Relativity.

4.1 The Equivalence Principle

Why is special relativity an incomplete theory?

- Special relativity is 'special' in the sense that it relates the values of physical quantities in different *inertial* frames, i.e. between the special set of frames, all of which are in uniform motion relative to the others, in which Newton's laws hold. It does not apply directly to frames that are accelerating with respect to inertial frames.

- Special Relativity leaves unanswered one very important question: suppose a set of reference frames S_1,S_2,S_3... are all stationary with respect to each other at some instant but are in different states of acceleration. What is it that selects *one* of these systems to be the inertial frame?

- As is explained later, inertial frames only exist *locally* in a gravitational field and so the Special Theory of Relativity is applicable to such fields only in a very restricted sense.

The complete theory of relativity applicable to gravitational fields is General Relativity. This section makes a brief excursion around the periphery of the General Theory, the inner workings of which are rather complex. In fact it is perhaps some measure of this complexity that some ten years elapsed between the publication of Einstein's papers setting out the Special and General Theories.

The starting point for relativity in accelerating frames and gravitational fields is the Equivalence Principle, which can be stated in two parts. First the Weak Equivalence Principle states that acceleration and gravitational pull are equivalent. Consider two laboratories as shown in Fig. 1 (a) and (b). The first is a static laboratory in a gravitational field with acceleration g. The second is a laboratory accelerating upward at a rate $a = |g|$. In each case the spring balance will be extended and gives the same reading. Other physical experiments performed in the two frames will also give the same result. Second, the Strong Equivalence Principle states that laboratories in free fall in a gravitational field are locally equivalent to laboratories in free space where the gravitational field is negligible. This is illustrated by Fig. 1 (c) and (d). The first shows a laboratory in a region of zero gravity. The second shows a laboratory in free fall in a gravitational field, i.e. accelerating a rate $a = g$. In both cases, the spring balance is unextended. The results of all other experiments in the two frames will also be locally equivalent.

The 'locality' of the equivalence implied above can be illustrated by considering neighbouring particles freely falling in the gravitational field of the Earth (or similar body). Each falls toward the centre and therefore their paths are not quite parallel, as shown in Fig. 2. If the particles are separated by a horizontal distance x near the Earth's surface, then the rate at which they drift together is $\dot{x} = R\dot{\theta} = vx/R$, where R is the Earth's radius and v is the free fall velocity. Thus, the free fall laboratory is only *locally* an inertial frame: over too large a range of distance, free particles start to move relative to each other without the application of a force, in violation of Newton's first law.

The Equivalence Principle provides a means to analyse physics problems in the absence of gravity by solving the problem in a reference frame in free fall. Although the result is usually required in a non-inertial frame (i.e. one not in free fall) the result will be the same near the origin of inertial and non-inertial frames that are momentarily at rest with respect to each other (i.e. they have the same velocity but different accelerations). To visualise this, imagine two children clinging to a rope hanging from a tree. At $t = 0$, one child lets go of the rope. At this instant, both are stationary, but one child is in free fall (and is therefore an inertial observer),

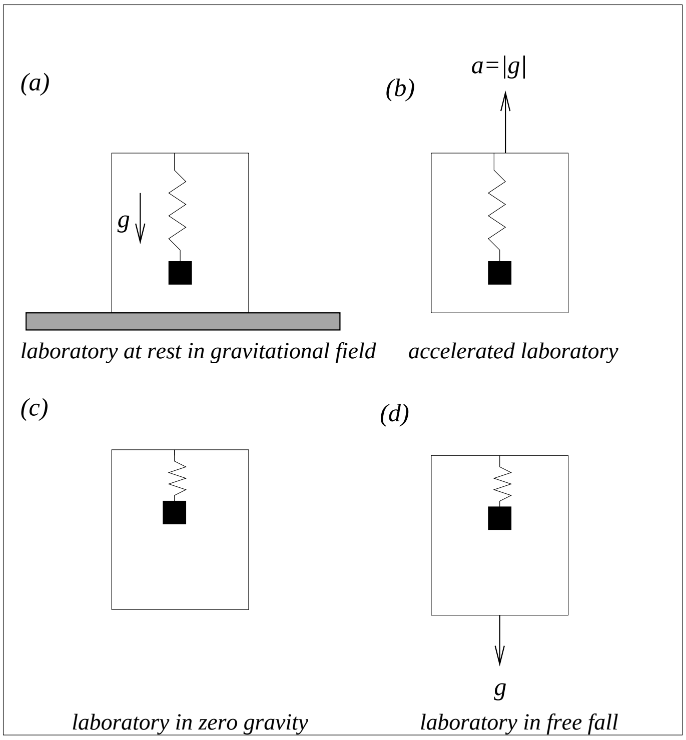

Figure 1: The Weak Equivalence Principle stipulates that the results of physics experiments in a gravitational field and in a laboratory accelerating at $a = |g|$ are locally indistinguishable (panels (a) and (b)). The Strong Equivalence Principle asserts that the results of physics experiments in a region of zero gravity and in a laboratory in free fall in a gravitational field are locally indistinguishable (panels (c) and (d)).

while the other is not. Nevertheless, at the instant $t = 0$, both will agree on coordinates, speeds, frequencies, and other measurements that do not depend on acceleration. In a gravitational field such a transformation is only valid in the neighbourhood of one point. Over an extended path, it must be repeated for all points along the path. This is done using the machinery of General Relativity.

The Equivalence Principle also provides a solution to the problem posed by the special status of inertial frames: what is it that makes Newton's laws apply relative to one set of coordinates rather than another which is accelerating relative to the first? The answer is that the inertial frames are the frames of reference in free fall in the local gravitational field. In fact, it is now clear that, since gravity pervades the entire Universe, inertial frames are an idealised concept and only exist over restricted distances, near the origins of coordinate systems in free fall in the local gravitational field.

4.2 Gravitational redshift and time dilation.

Consider a laboratory of height h released from rest into free fall, as shown in Fig. 3. A beam of light of wavelength λ is emitted from the floor and travels toward the roof. The Equivalence Principle requires that, to an observer at the roof in free fall with the laboratory, the wavelength must be unchanged. To observers at rest, however, the laboratory has increased in speed by $gt = gh/c$ in direction away from the recipient of the light. To such observers, the light must be redshifted by $\Delta\lambda = (v/c)\lambda = (gh/c^2)\lambda$. This corresponds to a 'gravitational' redshift

$$z_g = \frac{gh}{c^2}. \quad (1)$$

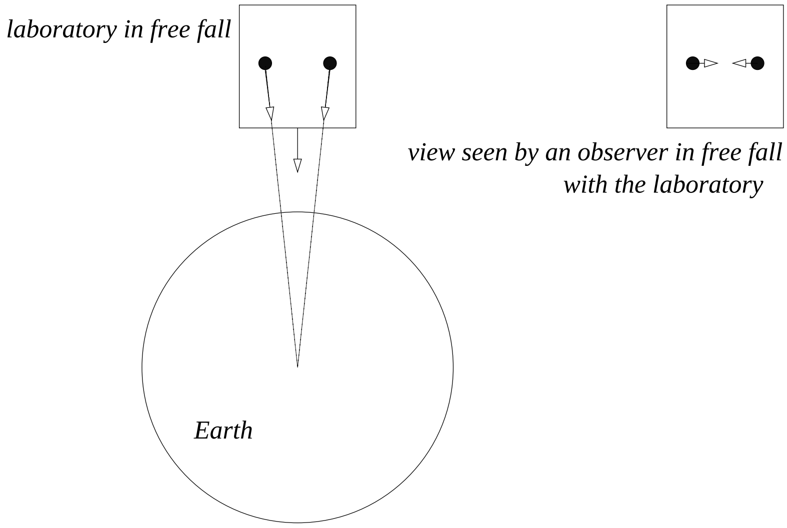

Figure 2: Left: A laboratory in free fall toward the centre of the Earth. Particles on either side of the laboratory also accelerated toward the centre of the Earth and so their directions of motion are not quite parallel. Right: As seen by an observer in free fall with the laboratory, the two particles approach each other at speed proportional to their separation.

(Note that the transformation above relates the value of λ in two frames, both in free fall (i.e. both inertial frames), one momentarily at rest and one moving with the laboratory. However a *non-inertial* observer (such as someone standing on a ladder holding a spectrometer) will record the same value for λ as the free-fall observer that is also at rest.) The quantity gh is the change

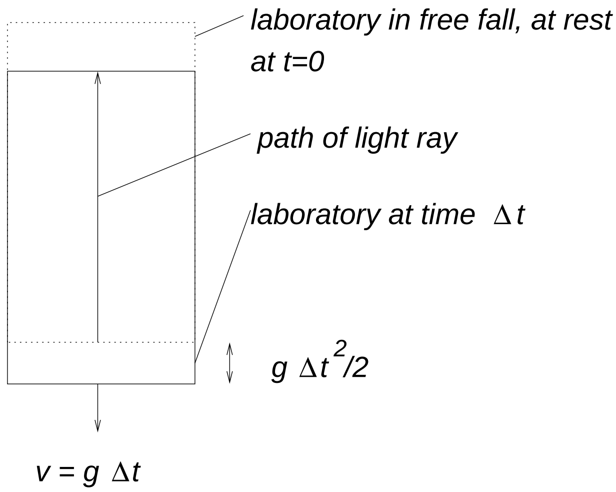

Figure 3: A laboratory in free fall is initially at rest. A light ray is emitted from the floor toward the roof. As the light ray reaches the roof, the laboratory has started to move downward.

in gravitational potential, $\Delta\phi$, so that the redshift could equally be written

$$z_g = \frac{\Delta\phi}{c^2} \tag{2}$$

This result has been tested very precisely using the Mossbauer effect, in which gamma rays of very narrow spectral width are emitted from nuclei in a crystal lattice. The locking of the nuclei in a lattice minimises their motion and reduces the Doppler shift they cause.

Problem 1. In the experiment of Pound and Rebka (1964), gamma rays were emitted from the top of a 22.4 m high tower. Find the predicted value of z_g and compare this with measured value of $(-2.57 \pm 0.26) \times 10^{-15}$.

Problem 2. The above argument shows that light climbing out of a gravitational field loses energy. Show that the energy lost by a photon of energy E in climbing distance h is the same as that lost by a mass $m = E/c^2$ in climbing the same distance.

An observer detecting light from a source of frequency ν located deeper within a gravitational potential will measure a lower frequency $\nu/(1+z_g)$. It follows that two events separated by N cycles of light at the source will be separated by the same number of cycles at the observer, and therefore by a time which is longer than that at the source by a factor $(1+z_g)$. In effect, the clock (in this case a source of light, but it could be any type of clock) runs slower the deeper within the gravitational field it lies. Hence, in the presence of a gravitational field due to a mass M, the gravitational potential at radius r is (see Tipler, Section 11-3)

$$\phi = \frac{GM}{r} \tag{3}$$

and so

$$\frac{\Delta t(r)}{\Delta t(\infty)} = \frac{1}{1+z_g} \simeq 1 - z_g = 1 - \frac{GM}{rc^2} \tag{4}$$

where $\Delta t(r)$ and $\Delta t(\infty)$ are the time intervals between two events as recorded on clocks at r and at an infinite distance from M, where the gravitational field is effectively zero. A Taylor expansion (see EMU Maths notes, Section 3) has been performed assuming $z_g \ll 1$.

It is possible to construct another story similar to the 'Twin's paradox' of Section 1. In this case the mobile twin B moves deep into a gravitational field. The stationary twin A monitors B's clocks and sees them ticking more slowly than his own throughout B's journey. As a result when B returns, a shorter time has elapsed on B's clocks than on A's and so B is younger. Again, there is no real paradox because there is no symmetry between A and B: A remains at a constant gravitational potential whereas B spends the entire journey deeper in the potential well.

It is interesting to note that these effects must be corrected in determining positions using GPS (the Global Positioning System). This system uses a network of satellites in orbits of radius 26,000 km each making approximately two orbits per day to locate the position of radio receivers to an accuracy of about 15 m. Each receiver listens to broadcasts from the GPS satellites giving precise measurements of time and position at the satellite. Using these, the receiver can determine its precise location. However, in comparing the clock readings at the Earth and the satellite their different clock rates must be taken into account. These differing clock rates arise from (i) the effect of the gravitational field predicted by General Relativity and (ii) the relativistic time dilation associated with Special Relativity.

The General Relativistic effect can be determined from the ratio of the clock rates given by equation 4:

$$\begin{aligned}
\frac{(\Delta t)_{Earth}}{(\Delta t)_{satellite}} &= \frac{1 - GM/c^2 r_{Earth}}{1 - GM/c^2 r_{satellite}} \\
&\sim (1 - GM/c^2 r_{Earth})(1 + GM/c^2 r_{satellite}) \\
&\sim 1 - (GM/c^2)(1/r_{Earth} - 1/r_{satellite}) \\
&\sim 1 - (GM/c^2)\frac{r_{satellite} - r_{Earth}}{r_{Earth} r_{satellite}}
\end{aligned}$$