# Astronomy in the Third Millennium
# Special and General Relativity

Norman Gray

February 2002

[Version: 1.4, 2002/01/27 23:16:45]

This pair of lectures is concerned with Special and General Relativity, and the rôle that General Relativity (GR) will have in the astronomy of the next century. Some bits of it are, I'm aware, rather hard going. What I want to concentrate on, however, are not any of the many details, but on the ideas underpinning these details, most specifically the idea of taking a geometrical approach to general relativity.

# 1   Special Relativity

Special Relativity really *is* special. It fundamentally challenges our view of the world, and yet the mathematical knowledge required to explain it is negligible, in comparison to other areas of physics. I am going to present a *full* account of special relativity (SR), up to length contraction and time dilation. This will include a couple of equations, but they will include nothing beyond multiplication and addition; furthermore, there is nothing of great intrinsic importance in those equations, and you can safely skip over them. The point I want to emphasise is that, although the material in this block is rather demanding, the axioms of SR and the logic that springs from them are intelligible without any hugely sophisticated mathematics.

I'm going to start off by setting the scene and describing the important terms we're going to use. Then I'm going to describe the two axioms of SR – the only actual *physics* I'm going to introduce in this block – and describe our common-sense view of the world in these new terms. Then I'm going to come to the central part of the block, and describe how the two axioms of SR inevitably upset our notions of time, space and simultaneity.

## 1.1   The first postulate: the Relativity Principle

A form of the Principle of Relativity, was described very clearly by Galileo:

> SALVATIUS: Shut yourself up with some friend in the main cabin below decks on some large ship, and have with you there some flies, butterflies, and other small flying animals. Have a large bowl of water with some fish in it; hang up a bottle that empties drop by drop into a wide vessel beneath it. With the ship standing still, observe carefully how the little animals fly with equal speed to all sides of the cabin. The fish swim indifferently in all direction; the drops fall into the vessel beneath; and, in throwing something to your friend, you need throw it no more strongly in one direction than another, the distances being equal; jumping with your feet together, you pass equal spaces in every direction. When you have observed all these things carefully (though there is no doubt that when the ship is standing still everything must happen in this way), have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that. You will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still. In jumping, you will pass on the floor the same spaces as before, nor will you make larger jumps toward the stern than toward the prow even though the ship is moving quite rapidly, despite the fact that during the time that you are in the air the floor under you will be going in a direction oposite to your jump. In throwing something to your companion, you will need no more force to get it to him whether he is in the direction of the bow or the stern, with yourself situated opposite. The droplets will fall as before into the vessel beneath without dropping toward the stern, although while the drops are in the air the ship runs many spans. The fish in their water will swim toward the front of their bowl with no more effort than toward the back, and will go with equal ease to bait placed anywhere around the edges of the bowl. Finally the butterflies and flies will continue their flights

indifferently toward every side, nor will it ever happen that they are concentrated toward the stern, as if tired out from keeping up with the course of the ship, from which they will have been separated during long intervals by keeping themselves in the air. *Galileo Galilei, Dialogue Concerning the Two Chief World Systems, quoted in [Taylor and Wheeler, 1992], §3.1*

This is a very clear account of the Relativity Principle (RP). Another way of putting it is that 'you can't tell if you're moving' – there's no experiment you can do which would allow you to distinguish between a moving and a stationary reference frame (a term we will return to in Sect. 1.3.1). This argument immediately makes redundant the whole idea of a universal stationary frame, since being in such a putative frame cannot have any observational consequence.

The RP as quoted here implicitly refers only to mechanics. However, given that you need mechanical components to do any electomagnetic experiment, and given that all mechanical objects are held together by (atomic) electromagnetic forces, it would seem unavoidable that it must apply to electromagnetism as well. See also Einstein's remarks quoted below.

From the GT, one can show that, with certain obvious (but, as we shall discover, wrong) assumptions about the nature of space and time, one could derive the (apparently also rather obvious) *Galilean transformation* (GT). This relates the positions and times of events as measured by, for example, someone on a roadside, to the positions and times as measured by someone in a moving car, and immediately leads to the conclusion that, for example, if you throw a ball at 10mph from a car moving at 60mph, someone standing on the roadside will see it moving at (10+60)mph.

If you take the RP as true, then it follows that any putative law of mechanics which *does* appear to allow you to distinguish between reference frames cannot in fact be a law of physics.

Everything, therefore, seems to be rosy.

Everything, in fact, *was* rosy, until the end of the nineteenth century. Around then, physicists were investigating *Maxwell's Equations*, one of the highpoints of nineteenth-century

physics, which unified all of the phenomena of electricity and magnetism into a single formalism of tremendous insightful power and overwhelmingly successful application – Maxwell's Equations worked.

However, it appeared that Maxwell's Equations had their simplest form only in a frame which was *not moving* – the fact that the equations of electromagnetism were not invariant under the GT appeared to indicate that, whenever you watched an electromagnetic experiment (such as an ammeter, or a microwave oven) in a moving frame, it should work differently from that same experiment in a stationary frame. Specifically, it suggests that there actually exists such an absolutely stationary frame, which is otherwise rendered unnecessary by the RP: saying 'you can't tell if you're moving' is another way of saying 'there is no (experimentally accessible) standard of absolute rest'.

Another, linked, problem was that of the aether. Since light is an electromagnetic wave, it seems obvious that, like water waves or sound waves, there must be something that light is a wave *in*. This 'light medium' was named the aether, and had the apparently contradictory properties of being both very rigid (so that it could sustain the very high frequencies of light) and very tenuous (so that objects such as planets could move through it freely). The aether would be a clear candidate for the frame of absolute rest.

The Earth moves around the sun in its orbit, with a constantly changing velocity. It followed, therefore, that there was some point in its orbit at which it had a maximum, and another point at which it had its minimum, velocity with respect to the putative aether. Although this velocity is rather slow compared to the speed of light, it should have been possible to *measure* the change in the velocity of the Earth with respect to the aether or the absolute rest frame. There was therefore a series of experiments in the late nineteenth and early twentieth centuries which attempted to measure this relative velocity: the Michelson-Morley aether-wind experiment attempted to measure the different light-travel times for beams directed along and across the flow of the aether; the Fizeau experiment and Lodge's experiments attempted to detect the extent to which the aether could be dragged along by fast-moving objects on Earth. All of them failed: no-one was able to detect the Earth's movement through the aether, or

the movement relative to the absolute rest frame, which the Galilean Transformation and the apparently necessary properties of an apparently necessary aether demanded.

> The aether drift experiments are discussed in most relativity textbooks. For an interesting sociological and historical take on the Michelson-Morley experiments, and the context in which they were interpreted, see also [Collins and Pinch, 1993, ch. 2].

So there was a conflict between the predictions of Maxwell's equations plus the GT plus the Relativity Principle, and what was observed in experiment. At this stage there were a number of options.

(i) Perhaps Maxwell's equations were wrong – perhaps light didn't behave as a powerfully simplifying theory, and a huge number of successful experiments, suggested it should.

(ii) Perhaps the Relativity Principle was wrong, although one possible result of the repeated failures to measure the Earth's motion through the aether could, I imagine, have been the weakening of the idea of an absolute rest-frame.

(iii) Perhaps the GT was wrong, though this transformation seems so obvious, and so bound up with our other preconceptions that it would be difficult to see *how* it could be wrong.

(iv) Perhaps there was some further physics at work. There were suggestions that an experimental apparatus, or even the Earth itself, might be able to drag the aether along with it enough to wipe out any detection from the Michelson-Morley experiment. Lorentz managed to find a transformation – now known as the Lorentz Transformation – under which Maxwell's equations *are* invariant, but was then left with the problem of explaining why those equations were apparently uniquely subject to a different transformation law from everything else. Attempts to explain the latter transformation included suggestions that objects might inexplicably change their lengths, and time be distorted, when moving head-on into the aether; but this in turn would require further explanation (see Lorentz's papers in [Lorentz et al., 1952] for an account of the latter attempts). It would have been clear that something was very wrong.

The resolution was that (i) Maxwell's equations are right, (ii) the Relativity Principle is right, (iii) the Galilean Transformation is inadequate and (iv) Special Relativity is the new physics to come out of this.

Einstein explains this as clearly as anyone. In his 1905 paper which introduced SR ('*On the Electrodynamics of Moving Bodies*'), he opens with a paragraph commenting that only *relative* motion is important in Maxwell's equations, and then goes on to say, with magisterial finality:

> Examples of this sort, together with the unsuccessful attempts to discover any motion of the Earth relatively to the 'light medium', suggest that the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest. They suggest rather that, as has already been shown to the first order of small quantities, the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good. We will raise this conjecture (the purport of which will hereafter be called the 'Principle of Relativity') to the status of a postulate…[Einstein, 1905]

In other words, he is saying that Galileo's Relativity Principle, which had really only been a statement about mechanical experiments, now applied to *all* of physics.

We can recast the Relativity Principle, the first postulate of Special Relativity, as follows:

> *The Principle of Relativity*: All inertial frames are equivalent for the performance of *all* physical experiments.

There is no physical (or chemical or biological or sociological or musical) experiment I can do which will have a different result when I'm moving from when I'm stationary. There is therefore no need for even the idea of a standard of absolute rest.

## 1.2 The second postulate: the constancy of the speed of light

The passage I quoted above, from Einstein's 1905 paper, goes on to say:

> We will raise this conjecture (the purport of which will hereafter be called the 'Principle of Relativity') to the status of a postulate, and also introduce another postulate, which is only apparently irreconcilable with the former, namely, that light is always propagated in empty space with a definite velocity $c$ which is independent of the state of motion of the emitting body.

This is the second postulate of Special Relativity.

This doesn't seem particularly remarkable at first reading; after all, we *know* that 'the speed of light' is one of Nature's fundamental constants, at $c = 299792458\,\mathrm{ms}^{-1}$. The sting is in the final remark, 'independent of the state of motion of the emitting body'. At first thought, there would seem to be three things that 'the speed of light' could mean:

1. The speed relative to the emitter (like a projectile);

2. The speed relative to the transmitting medium (like water or sound); or

3. The speed relative to the detector.

Option 2 is ruled out by the first postulate: if this were true then the frame in which light had this special value would be picked out as special; the RP also incidentally excludes the notion of the aether. Option 1 also turns out not to be the case. The GT says that velocities add, so that a projectile emitted from a moving object has not only the speed it was fired with, but also the speed of the emitter. This is not so for light, or indeed any object moving at a significant fraction of the speed of light, and this can be amply verified by observing the light coming from Jupiter's moons, or binary stars, or other astrophysical objects moving at great speeds.

No, option 3 is the case, so that, no matter what sort of experiment you are doing, whether you are directly observing the travel-time of a flash of light, or doing some interferometric

experiment, the speed of light relative to your apparatus will always have the same numerical value. This is perfectly independent of how fast you are moving: it is independent of whichever inertial frame you are in, so that another observer, measuring the *same* flash of light from their moving laboratory, will measure the speed of light relative to *their* detectors to have exactly the same value.

There is no real way of *justifying* this postulate: it is simply a truth of our universe, and we can do nothing more than simply *demonstrate* its truth through experiment.

## 1.3  Simultaneity

This section is where we investigate the consequences of the axioms I've described above.

### 1.3.1  Measuring lengths and times: simultaneity

Below, we will talk repeatedly of 'reference frames', and 'inertial reference frames' (in fact, we mentioned them in passing in Sect. 1.1). The term 'reference frame' can be explained rather informally as a 'point of view', or a set of reference points. Thus a room might be regarded as a reference frame, and a person walking through it might represent another. A train station is a reference frame, and two trains going through it, at different speeds, accelerating or decelerating, represent two others. Some frames are special, in that they are not accelerating; these are called 'inertial reference frames'. In all but earthquake zones, train stations are generally inertial frames; the trains in them represent inertial frames when they are standing at the platform, are demoted to mere refererence frames while they are accelerating from the station or turning a bend, and get the 'inertial' status back whenever they are moving at a constant speed in a straight line. For subtleties and quibbles, see Sect. A.1.

How do we measure times? In SR, we repeatedly wish to talk about the time at which an event happens. If the event happens in front of our nose, we can just look at our own watch. This is an important point. One of the things we can hold onto in the rest of this course, is that if two events at the same spatial position happen at the same time, they are simultaneous for everybody. This is why a time measurement by a local observer is always reliable. Observers at that same spatial position but moving in different frames may produce different numbers, but their measurement is, by definition, the measurement of the time of that event in that frame.

> We must take into account that all our judgments in which time plays a part are always judgments of *simultaneous events.* If, for instance, I say, "That train arrives here at 7 o'clock," I mean something like this: "The pointing of the small hand of my watch to 7 and the arrival of the train are simultaneous events." [Einstein, 1905]

If the event happens some distance away, however, or if we want to know what time was measured by someone in a moving frame, things are not so simple (as most of the rest of this lecture course makes clear). SR defines a clear procedure for measuring times. When we talk about the time of an event, we *always* mean the time of the event *as measured by a local observer*, that is, an observer at the same spatial position as the event (which is rather unfortunate if the event in question is an explosion of some type – but what are friends for?), whether they are stationary or moving. Crucially, *all these observers have clocks which are synchronised with all others in their frame.*

How should we measure the length of a moving object? The obvious methods – some variant of laying a ruler along the object to be measured and examining the marks on the ruler – make too many assumptions about how the world is. To measure something accurately this way, you have to correct for the time it takes for light to travel from the ruler and the (moving) object to you. You obviously have to measure both ends of the object simultaneously: does that 'simultaneously' depend on where you're standing relative to the object? The obvious approaches to this problem beg too many questions to which SR provides surprising answers.

The way we measure lengths and times in SR is therefore as follows. We station observers at strategic points in the reference frames of interest (in principle, we have an infinite array of observers scattered throughout space). We can know these observers' coordinates in one frame or another. The observers make measurements of events that happen at their location, and afterwards compare notes and draw conclusions. For example, you would measure the 'length of a rod' by subtracting the coordinates of the two observers who observed opposite ends of the rod at a prearranged time.

This approach relies on three things.

1. It requires a specific procedure for synchronising clocks. This isn't too difficult to define, within the context of the two axioms mentioned in sections 1.1 and 1.2.

2. It assumes that there is no ambiguity about two events at the same position and time being regarded as simultaneous. This has to be true: the fact that two cars crash –

because they were in the same position at the same time, and so are attempting to occupy the same space simultaneously – cannot depend in any way on your point of view.

3. We assume that moving clocks measure the passage of time accurately. We do not assume anything about accelerating clocks, because in SR we largely avoid discussion of acceleration, but we do assume that there is nothing magical about motion which causes clocks to go wrong. This is known as the *clock hypothesis*, and is related to the Equivalence Principle which is one of the axioms of SR.

### 1.3.2  Train-spotting

Imagine standing in the centre of a train carriage, with suitably agile friends at either end: Fred (at the Front) and Barbara (at the Back). At a prearranged time, on your carefully synchronised watches, you fire off a flashbulb and your friends note down the time showing on their watches when the flash reaches them. Comparing notes afterwards, you all find that it took some time for the flash to travel from the middle of the carriage to the end, and that your friends have noted down the same arrival time on their watches, time '3', say (not seconds!)[1]. In other words, Fred's and Barbara's watches both reading '3', are *simultaneous* events.

But if this train is moving through a station as all this goes on, and you look from the platform into the carriage, what would you see from this point of view? You would see the light from the flash move both forward towards Fred and backwards towards Barbara, but remember that you would *not* see the light moving forwards faster than the speed of light – its speed would not be enhanced by the motion of the train – nor would you see the light moving backwards at less than $c$. Since the back of the train is rushing towards where the light was emitted, the flash would naturally get to Barbara first. If, standing on the platform, you

---

[1] These watches are obviously not calibrated in seconds. For a flash of light to travel 10m, would take roughly $3 \times 10^{-8}$ s – three hundredths of a millisecond.
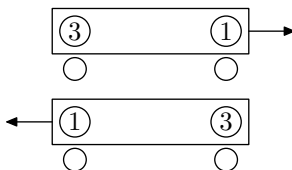
Figure 1: Trains passing in the night.

were to take a photograph at this point, you would get something like the upper part of Fig. 1. Barbara's watch reads '3'. But at this point, the light moving towards Fred cannot yet have caught up with him: since the light reaches Fred when *his* watch reads '3', his watch must still be reading something less than that, '1', say. In other words, Barbara's watch reading '3' and Fred's watch reading '1' are simultaneous events in *your* inertial frame on the platform.

What is going on here? Are these events simultaneous or not? What this tells us is that our notion of simultaneity is rather naïve, and that we have to be very careful exactly what we mean when we talk of events as being simultaneous. The only case where two events are quite unambiguously simultaneous is if they take place at exactly the same point in space. That's why we could say without hesitation that the light reached Barbara when her watch read '3', because that's what she had seen and noted down.

### 1.3.3 Length contraction and time dilation

We're not finished with the trains, yet. Imagine now we're standing on the platform and see now two trains go past. We've cunningly arranged the speeds, timetable and flashbulbs so that we can get the photograph in Fig. 1, where the light has reached both rear observers and neither front one. Now pause a moment, and take another photograph when the two rear observers are beside each other, this time getting Fig. 2.

After all this hectic fun is over, everyone calms down, ambles together, and compares
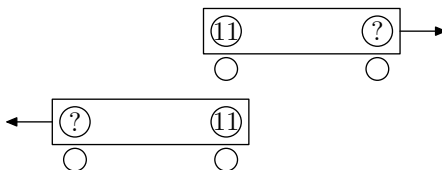
Figure 2: The trains, a little later.

notes. Barbara (standing at the back of the top carriage) could remark "I saw the front of the other carriage pass us when my clock was reading '3' " (this is perfectly correct, as you can confirm by looking at the 'photograph' in Fig. 1). At which Fred would say "But the back of the carriage passed me at time '1' – it must have been well past me at time '3' ". From this they, and we, can quite correctly remark that the carriage they observed moving past them was measured to be shorter than their own. They have measured the length of a moving carriage, and found that it is shorter than a similar carriage (their own) which they can measure at leisure. This is *length contraction*.

Fred then says "I looked through the window at the clock at the back of the other carriage, and I noticed that it was reading '3', when mine was reading '1' – it was two seconds fast". Barbara says "Well, I saw that *same* clock a bit later [in Fig. 2], and it was reading '11', just like mine – it wasn't fast at all." They know that their own clocks were synchronised throughout the encounter (they can make sure that their clocks are synchronised at some point, and they know that they both go at the same rate), so they can only conclude (correctly) that the clock they both saw was going more slowly than theirs were. Time in the other carriage is passing more slowly than in their own, the same phenomenon as we saw with the light clock, which led up to the mathematical expression of this in Eqn. (2) above.

The extraordinary thing is that Barbara and Fred's counterparts in the other carriage would come to precisely the same conclusions. Because this setup is perfectly symmetrical, they

would measure Barbara and Fred's clocks to be moving slowly, and their carriage to be shorter. There is no sense in which one of the carriages is *absolutely* shorter than the other.

## 1.4   Time dilation

Imagine you're on a train which happens to be charging through a station at some significant fraction of the speed of light. You throw a ball into the air and catch it again: how would you describe this? You'd say that the ball started off in your left hand, followed a parabolic path (like anything thrown), landed in your right hand, and that it took one second (say) to do it all. Now imagine you're on the platform watching this go on: how would you describe it now? You'd say that it started off at the start of the platform, landed a good way down the platform, and took one second to do it. These two perceptions agree that the ball follows a parabolic path (different parabolae, yes, but parabolae nontheless – this is the Equivalence Principle at work), but they disagree on how far the ball travelled in flight. That disagreement is easily explained: from the point of view of the observer on the platform, the ball was travelling very quickly, since it had the train's speed as well as its own, so of *course* it covered more ground before it landed again.

   None of this is mysterious. I've made it mysterious by the elaborate way in which I've described it. But I've described it that way to pull this perfectly normal situation into line with the next step, the *light clock*.

### 1.4.1   The light clock

The light clock (see Fig. 3) is an idealised timekeeper, in which a flash of light leaves a bulb, bounces off a mirror, and returns. If the mirror and the flashbulb are a distance $L$ apart, so that the round trip is a distance of $2L$ and I, standing by the light clock, time the round trip as $t'$ seconds then, since distance is equal to speed times time, and since the speed of light is the constant $c$, it must be true that

$$2L = ct' \tag{1}$$

(note that $t'$ here is the time on my watch, standing by, and moving with, the clock).
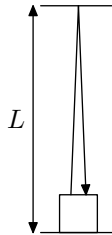
Figure 3: The light clock

Now imagine the light clock sitting on the train going through the station, as you watch it from the platform. The clock is in motion, at a speed $v$, while the light crosses the clock, so that when the light bounces off the mirror, the clock is further down the platform that it was when the light started, and by the time it gets back to the bulb, the whole thing is further down the platform still (see Fig. 4). If, standing on the platform, you time all this and find that it takes a time $t$ for the light to make the round trip, then (remembering again that distance is speed times time) the clock will have moved a distance $vt$ down the platform.

How far has the light travelled? We know it travelled at a speed $c$ (the second axiom tells us that, no matter where or how the light was travelling – in a clock sitting beside us on the platform, or a clock zooming though the station in front of us – we would *always* measure its speed to be the same number $c$), and you timed its round trip at $t$ seconds (this is a number on your watch, standing on the platform). Using no mathematics more sophisticated than Pythagoras' theorem, it turns out that we must have

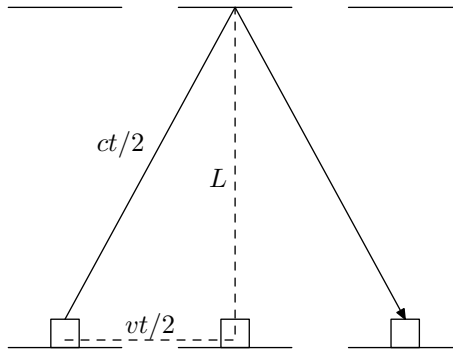$$t' = t\sqrt{1 - \frac{v^2}{c^2}} \tag{2}$$

(see Sect. A.2 for some details).

Figure 4: The light clock, as it's perceived when it's moving past you.

Now, the important thing about this equation is that it involves $t'$, the time for the clock to 'tick' as measured by the person standing next to it on the train, and it involves $t$, the time as measured by the person on the platform, and *they are not the same.*

How can this possibly be? Why is this different from the perfectly reasonable behaviour of the ball thrown down the carriage? The difference is that when you watched the ball from the platform, you saw it move with the speed it was given plus the speed of the train – in other words, the person on the platform and the person on the train had a perfectly reasonable disagreement about the speed of the ball, which resulted in them agreeing on the time the ball was in flight. However, both of them *agree* on the speed of the light in the light clock, as the second axiom says they must. Something has to give, and the result is that the two observers disagree on how long the light takes for a circuit.

So at least one of the clocks is broken? They're both in perfect working order. They only work properly when they're stationary? No, the Equivalence Principle tells us that there's no sense in which either of them is 'more stationary' than the other, so that the clocks work in

exactly the same way whether they're moving or not. No…

> *Both clocks are perfectly accurately measuring the passage of time. Time is flowing differently for the two observers.*

A useful image I find is to imagine a clock like some design of ship's log, which is some type of propellar arrangement towed behind a ship which records how much water the ship has moved through. Analogously, a clock is a device which records how much time the clock has been dragged through.

## 2  General Relativity and the geometry of spacetime

Hold a metre-stick in front of you. How long is it? A metre, of course. Now hold it at an angle: how long is it now? It's still a metre long, but it doesn't look like it – because it's at an angle, we see it as being shorter. Our intuitive knowledge of the geometry of three-dimensional space tells us that the thing really is the same length as it always was, but because it's been turned in space, its projection onto our field of view is less, and so this observed shortening is merely an artefact of the geometry of the situation.

Special Relativity tells us that we do not in fact live in the three-dimensional world we thought we did. We live in a four-dimensional world, albeit one with an unfamiliar geometry, and time is one of the directions. That isn't just wordplay: movement is something with a temporal as well as a spatial element, and when we move, we 'rotate' in this four-dimensional space. When others see us move, they experience the analogue of the foreshortening of the ruler and they naturally see our time, and our spatial extent, change.

## 2.1  The geometry of moving frames

We can make some sense of the behaviour we have seen by looking at the analogous situation where a line of observers are examining a pole at some angle to them (see Fig. 5). The observers can measure the 'length', $l$, of the pole by measuring the distance between the two observers who can see one end of the pole directly in front of them; and they can measure the 'depth', $d$, of the pole by having the same two observers measure the distance to the end of the pole that they can see. Both these measurements $l$ and $d$ have *some* physical significance, but as long as the observers insist on using $l$ and $d$ as their measurement of 'the length of the pole', then they will get different answers for the pole's length depending on where they are standing when they make the measurements. This is an undesirable situation, which is only resolved when we make use of our knowledge of geometry to construct the invariant length, $s$, of the pole, where Pythagoras tells us that

$$s^2 = l^2 + d^2. \tag{3}$$

No matter how the two observers are arranged, it is a geometrical fact that the paired values of $l$ and $d$ that they measure will lead to the *same* value for the invariant length $s$.

Now consider measuring the separation of two events such as a banger going off here at this time, and one going off there at that time. If we insist on using the distance between the bangers, $l$, or the time between the explosions, $t$, as our measure of the 'separation' of the two events, then we will measure different values for the 'separation' depending on how fast we are moving relative to the events (or, equivalently, how fast they are moving relative to us). It is only our knowledge of Special Relativity that allows us to resolve the conflict and calculate the *invariant interval*, $s$, separating the events, which is related to the time and distance separating them by

$$s^2 = (ct)^2 - l^2. \tag{4}$$

No matter how fast we are moving relative to the two bangs, it is a fact that the values of $l$ and $t$ that we measure will lead to the *same* value for the invariant interval $s$.

The similarity between these two situations is more than a coincidence – we can very profitably approach SR in geometrical terms. In this view, space and time are not fundamentally different things. We do not move in time through the three-dimensional world of our intuitive experience: instead we exist in a four-dimensional world in which time and space are merely different 'directions'. This four-dimensional world is called *spacetime*.

Geometry in spacetime is not the same as the Euclidean geometry we are familiar with. In Euclidean geometry, we have familiar theorems like Pythagoras' theorem, Eqn. (3), the familiar statement that "the square on the hypotenuse is equal to the sum of the squares on the other two sides". In this new four-dimensional world, the analogue of Pythagoras' theorem, and the invariant we can cling to in our dizzy new circumstances, is Eqn. (4). It is clear that this is similar to Pythagoras' theorem; it is equally clear that the difference is hugely significant.

This is the first hint of the fundamental role that geometry has in our understanding of the nature of spacetime.
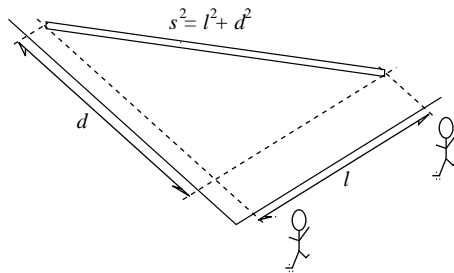


Figure 5: Observers make different measurements of the 'length' of the pole.

## 2.2   The geometry of a rotating frame

Our consideration of the mechanics of motion at high speed – Special Relativity – has been confined to high *constant* speeds, unaffected by gravity. When we generalise from this to motion near large masses, or in accelerating frames, our first question has to be 'what can we still hold on to?' I'll try to answer that by considering a particular kind of accelerated motion: rotation.

It may not be immediately clear to you that rotation at a constant speed involves any acceleration at all. To stay on a roundabout, however, you definitely need to hold on – the roundabout is obviously exerting some pull on you which, in another context, you would naturally associate with an acceleration.

Newton's first law says that in the absence of acceleration, objects move with a constant velocity. In physics, the term 'velocity' includes both speed and direction, so this law is formally stating what you would expect to see, watching a puck move across an ice rink: the puck moves at a constant speed, in a straight line. The only thing that disrupts this motion is if some other agency, such as a hockey stick, or some roughness on the ice, or a draught from the side, accelerates it away from this natural motion.

So, if you were to loosen your grip on the roundabout, you would resume your natural motion, and fly from the roundabout in a straight, tangential, line. Your grip on the roundabout is accelerating you away from the motion you would have in its absence: motion at constant velocity.

### 2.2.1   Co-moving inertial frames

Suppose you are holding on to the outer rim of a turning roundabout, just as someone runs straight past it. If they have chosen their speed to match yours, then there will be a short time when you and they are moving at the same speed, in the same direction. From the point of view of someone perched on the centre of the roundabout, therefore, there will be no difference between rulers and watches held by you and by the runner (see Fig. 6). They can therefore

use SR to discuss the motion of the runner, and apply the results to you, temporarily moving at the same speed. The inertial frame attached to the runner is called a *co-moving* inertial frame. Taking over our results from SR, the person perched at the centre of the roundabout will measure your ruler (and the runner's) to be shorter than their own, and will measure your watch to be showing less time passing than their own (do remember that we are implicitly assuming that the runner and the roundabout are moving at relativistic speeds here).

We have come to a remarkable conclusion. All the rulers throughout an *inertial* frame are the same length, and all the clocks move at the same rate, but if we imagine a frame turning with the roundabout – an *accelerated* frame – we see that rulers and clocks have systematically different lengths and rates at different points. This has an interesting consequence….



Figure 6: Running past a roundabout.

### 2.2.2  Geometry is distorted

If you draw out a circle on the ground, measure its diameter, and measure its circumference, perhaps by laying a ruler round the edge, you will find that the ratio of the circumference of the circle to the diameter will be $\pi = 3.1415\ldots$. What if you were to do the same thing on the roundabout?

Use your ruler to mark out a circle on the roundabout with a radius of, say, five ruler-lengths. Now measure its circumference by repeatedly laying the ruler along the circle you've marked. Someone standing by the roundabout, watching all this going on, will agree about the diameter of the circle, but will see these circumferential rulers whip past them at speed, and so will measure them as being shorter than a similar ruler of their own. Naturally, therefore,
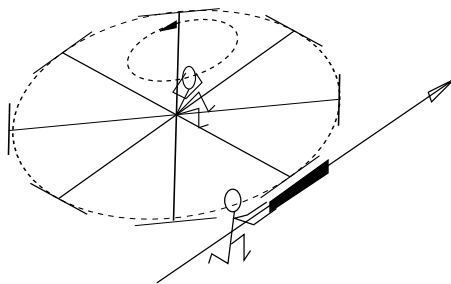
it will take *more* rulers to mark out the circumference than it would if the roundabout were stationary. In other words, the ratio of the measured circumference to the diameter – the number of ruler lengths it takes to circumnavigate the circle, divided by the number it takes to cross it – will be *greater than* $\pi$. Geometry on the roundabout does not obey Euclid's axioms. This is not an illusion – the frame attached to the rotating roundabout is a different shape from one that isn't accelerating.

⚛ This is a good place to emphasise that a foot-rule remains a foot-rule, even when the space it's sitting in expands. If all the atoms that make up the rule were free particles, sitting quietly in space-time, then when that space expanded, the atoms would remain in position with respect to the space, and expand with it. However, the atoms are *not* free particles, and are instead held together by forces (largely electromagnetic) which mean that the particles remain in position with respect to each other. Analogously, imagine a group of ants sitting on a balloon which is being inflated: as the balloon expands, the ants will move further and further away from each other; but if they sat in a circle and held hands (how picturesque!), they would be dragged over the surface of the expanding balloon, with the result that the circle would remain its original size.

## **2.3 The equivalence principle, GR, and curvature**

The equivalence principle in GR is just as central as the corresponding principle in SR. However, it is difficult to discuss the equivalence principle in detail without becoming enmeshed in subtleties which, though they are very important, easily distract from the main line of the argument, which aims to show how the framework on which SR is built is incompatible with the existence of gravitating matter.

### **2.3.1 The nature of the problem**

The argument is as follows:

1. Briefly, the equivalence principle is: in a small laboratory, physical systems in a gravitational field behave the same way as in an accelerated frame.

2. The behaviour of a free particle at loose in an accelerated frame is obviously independent of the composition of the particle. The equivalence principle tells us that the same *must* be true for a free particle in a gravitational field.

3. Any frame attached to a freely falling particle behaves like an inertial frame, and we can *define* an inertial frame as being one attached to a particle in free-fall.

4. Thus two inertial frames (such as those attached to a particle falling down a liftshaft on earth, and one cruising through outer space, or two free-falling lifts on opposite sides of the earth) can be *accelerating* relative to each other.

In point 1, the limitation to a small laboratory is important. In a laboratory which is large enough, in time or in space, you will always be able to tell whether your laboratory is accelerating, or is sitting near a planet. But the difference between these phenomena is a 'large' scale effect, and *locally*, the two effects are equivalent.

Point 2 is known as the *universality of free fall*, and translates into the profound observation that all masses, of whatever material, fall in a gravitational field with *exactly* the same acceleration. 'Free fall' here is just code for 'movement under the influence of gravity alone', and refers to a satellite in orbit, a plummeting skydiver (before she has reached terminal velocity), and a thrown ball (like the skydiver, ignoring air-resistance). Galileo was the first to emphasise this (and was why he was reputedly imperilling passers-by by flinging cannonballs from the leaning tower of Pisa), and it is an extraordinary, and completly inexplicable fact. That is, the gravitational mass of a particle, which tells you how much it is attracted to the Earth, and the inertial mass, which tells you how susceptible it is to being pushed, are exactly the same. This direct consequence of the equivalence principle has been elaborately verified.

Free fall means motion under the influence of gravitation alone. Since gravitation affects all matter equally (the universality of free fall), the particles in a plummetting lift will not be accelerated relative to each other. Things will stay where they are put, so that the lift acts as an inertial frame, hence point 3.

Point 4 is the problem. In SR, *all* inertial frames move with constant velocity relative to each other. They cannot accelerate. Here is a blunt contradiction with SR, created by the presence of gravitating matter.

## 2.3.2  The nature of the solution

It is important to be clear exactly where the problem is.

When we are dealing purely with local inertial frames, for example the frame attached to a rocket cruising through space, or the frame attached to a lift plunging towards destruction, we can use Special Relativity to discuss motion in those frames. It is an indirect consequence of the equivalence principle that, for a particle in free fall, the geometry of spacetime is *locally* that of SR (which, if you fancy a bit of name-dropping, is called Minkowski space).

That word 'locally' is the important one. We have seen that Special Relativity (or, more precisely, the notion of globally applicable inertial frames) fails in the presence of matter

because that matter causes inertial frames attached to freely falling particles to accelerate relative to one another. Although SR correctly describes the local spacetime of a freely falling frame, it is the job of GR to tie these local inertial frames together and so to tell you how to move from one to the other.

In other words, we already know the essential local properties of spacetime, namely that described by Special Relativity. General Relativity describes the *global* properties, with acceleration and matter included.

### 2.3.3  The curved geometry of space

It should not come as any surprise, after the various mentions I've made of geometry, that GR produces a solution to the problem in *geometrical* language, using the notion of curvature.

You are probably more familiar with non-Euclidean geometry than you think! Draw a circle and a triangle on a piece of paper. If you measure the circumference and diameter of the circle, they will have a ratio of $\pi$, as everyone knows. If you look at the internal angles of the triangle, they will add up to 180°, as Euclid says they should.

Now do the same thing on a globe: use a piece of string to draw a circle with one pole as its centre, and draw a triangle with one vertex at one pole and the other two at the equator. If you compare the circumference of your circle with its diameter, you will find that the ratio is *less than* $\pi$, and if you add up the internal angles of your triangle, they will come to *more than* 180°. The sphere is intrinsically curved, and the geometry of shapes drawn on its surface does not obey Euclid's axioms: this is a non-Euclidean surface.

How do you draw a triangle on the surface of a sphere, if a triangle is three points connected by straight lines? In curved geometries, just as in Euclidean geometry, a straight line is the shortest distance between two points, so you can get your straight line by tying a piece of string to the two points you want to connect, and simply pulling it taut.

When we talk of curved spacetime, this is what we mean, except that the curvature is a feature of four-dimensional spacetime of three spatial dimensions plus time, rather than the

two-dimensional surface of the sphere.

The principal application of these ideas is in Cosmology.

### 2.3.4  Towards cosmology: the curved geometry of spacetime

We can get an idea of what it means to say that a 'space' is 'curved' by looking at curvature on familiar two-dimensional surfaces, where we have some intuitive notions to guide us. We will be guided by geometry.

On a plane, as we know, Euclid's geometry is valid. This means that, for example, we can find parallel lines (straight lines which never meet), the internal angles on a triangle add up to 180°, and the ratio of the circumference of a circle to its diameter is $\pi$. This is naturally just as true for a group of ants on a piece of paper as it is for us. Confined to the (two-dimensional) surface of the paper, they could make measurements of the circles and triangles we draw, and reach the same conclusions. A sheet of paper is obviously *flat*: we can extend the notion of flatness, and say that a *space* is flat if Euclid's geometry is valid, and triangles drawn in space do what they should. The space we live in is, to an excellent approximation, flat.

What would our ants see if they were to make their measurements on the surface of a sphere, say (remember that they cannot see the sphere from above, or outside, so making measurements of the curvature of the surface they live in is the only way they can find out about their world)? They could draw large circles or triangles on their sphere, and measure the ratio of the circumference to the diameter, or add up the internal angles, and they would indeed find out that they were living on a curved surface. With a little mathematics, they could develop an quantitative measure of the 'amount of curvature' on their sphere, and might well come up with a number that corresponded to what we, looking at the sphere in our three-dimensional world, would call the radius of the sphere.

The important point is that the ants have a notion of curvature that does not depend on the sphere being embedded within an three-dimensional world. This will be very puzzling to the ants – 'But where does the surface curved *into*?'; 'Into the *third dimension*!'; 'The what?!'. In

the same way, we can talk about curvature in three-dimensional space: if we were to measure the surface area and diameter of a ball in a curved 3D space, we would find that the ratio would, again, not be $\pi$, we would conclude that the space is curved, and we would be puzzled by questions 'where does the space curve *into*?' These notions of curvature also work in four dimensions, and in the more general geometrical spaces of special and general relativity, even though our intuition gives out.

The surface of a sphere is an example of a space with a constant curvature. Not all spaces or surfaces have constant curvature – think of the surface of an egg, for example – and the four-dimensional spacetime we live in is like that. How do we detect this curvature?

The curvature of spacetime affects us because it governs how matter moves. When matter is in free-fall, it moves along a *geodesic*, which is a 'straight line' in spacetime. In a spacetime such as that of SR, this 'straight line' (for which I will now use the term 'geodesic') in spacetime would appear to us as a straight line in space, but a geodesic in a *curved* spacetime, appears as a curved line in space. We will come back to this in a moment.

⚛ How do you determine a straight line through space? A straightforward means, which is a lot more fundamental than it seems at first sight, is to stretch a piece of string between the two endpoints: the taut string is the shortest distance between the two points, and thus a straight line. That this is so is an axiom of Euclid's geometry and may be taken to be a definition of what 'straight line' means in that context. Within the different geometry of four-dimensional spacetime, and fundamentally linked to the minus-sign in Eqn. (4), the precisely corresponding definition is that a 'straight line' geodesic is the *longest* distance between two endpoints, but the apparently important jump from shortest to longest distance is only an accident of the mathematical details of the geometry.

What makes a curved spacetime curved? Matter. In the presence of (suitably large amounts of) matter, spacetime alters its shape, and becomes curved. For us to casually notice this alteration in shape, we need planet-sized amounts of matter, but in principle all matter, from an atom upwards and down, minutely alters the spacetime around it. Matter in free fall – a ball thrown in the air, perhaps – follows a geodesic through that curved spacetime, and
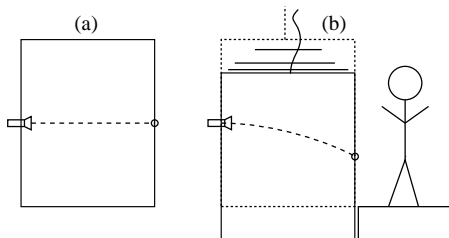
Figure 7: (a) A beam shone across an inertial lift cabin. (b) The same, measured from a non-inertial frame.

we perceive this as a curved passage through space, and an uneven passage through time. In other words, we see the path of the ball as a graceful (*curved*) parabola, and we see the ball accelerate towards the earth. We are *seeing* a geodesic drawn through the curved spacetime in which we live.

In Fig. 7a, we see a freely-falling lift cabin: in the context of GR, this means that the lift cabin constitutes an inertial frame (remember that this means that the lift could be plummeting towards the ground, or it could be floating in outer space). A flash from a torch on the left hand wall, travels across the cabin in a straight line and hits the opposite wall at a marked spot. In Fig. 7b, we consider what would happen if this took place in a gravitational field. If the lift cable were cut at the instant the flash left the left-hand wall, then the cabin would revert to being an inertial frame, and we know from Fig. 7a that the flash would travel across the cabin and strike the marked spot. Since the lift is falling, as perceived (or measured) by the horrified observer who was about to step into the lift, we can see that the light would take a curved path, again as measured by that observer. There is nothing about being in the lift cabin which *affects* the path of the light (it doesn't drag it along in any sense): this construction

merely allows us to work out what the path of the light *would be*, irrespective of any nonsense about falling lifts and suddenly nervous lift passengers. That is, we have elegantly deduced that a beam of light would curve whenever it is in a gravitational field, and this is an effect which was famously confirmed by Eddington's observation of the bending of starlight when its path to us grazes the surface of the Sun. Even more directly than with the thrown ball, massless light's bent path is a very direct illustration of the curvature of the spacetime the light is moving straight through.

The amazing thing here is that gravity, in the sense of Newton's mysterious action at a distance, has to some extent 'disappeared'. In GR, there is *nothing* called 'gravity' – particles do not 'know' that a large mass is near them, they simply move as straight as they can through a curved spacetime, the curvature of which was created by the presence of that large mass. There is a slogan: 'Space tells matter how to move; matter tells space how to curve'. This is bizarre; it is also very elegant and very beautiful.

Newton, by the way, was never happy with the implications of his model of gravitation. There is nothing in his inverse-square law which corresponds to the *propagation* of gravitational force: if an arrangement of masses was changed, the effects would be felt throughout the universe *instantly*. This was in straightforward contradiction to the Cartesian belief that mechanical effects must proceed from their causes by impacts – local interactions. As a consequence of this, Newton believed his model to be incomplete, but neither he, nor anyone else for the next 250 years, managed anything better.

The picture of reality that this gives us is of a curved stage, on which all the rest of physics takes place. The stage is nonetheless affected by the actors on it, and the two are linked by (and only by) the mass of the actors.

As a final remark, it is important to separate what is mere mathematics in what I have said, and what is physics. The remarks about the geometry of curved spaces, and about the existence and properties of geodesics are mathematics: there is nothing debatable about them and, for me at least, nothing fundamentally interesting. The statements that free-falling matter follows geodesics, and that matter causes a difference in the spacetime in which it sits, are

fundamentally different. These are *physical* remarks – statements about the universe which we could readily imagine to be false, in a way that we could not imagine '2 + 2 = 4' to be false. That they are not in fact false, hugely enriches our picture of the universe.

# 3  Cosmology: The Schwartzschild solution and black holes

The big question now is, *how* is curvature related to spacetime? After some trial and error, Einstein produced what has become known as the Einstein Field Equation(s), which I might as well quote in full:

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = -\kappa T_{\mu\nu}$$

(there are *many* mathematical details hidden behind this deceptively simple line; but *The Waste Land* isn't an easy read either). The terms on the left are related to the curvature of a spacetime, and the term on the right gives the distribution of matter there (actually energy, which is the same thing, in relativity). This is by no means easy to solve, or even to understand, but solutions have been found in certain simple situations.

One of the first people to produce a solution to the Einstein field equations was Karl Schwartzschild, in 1916. This describes the spacetime round a symmetric, non-rotating, mass at rest in an otherwise empty universe. This (relatively) simple solution describes mathematically various effects I have hinted at before: if you were sitting close to the mass, an observer far away would see your watch move more slowly than his; free particles placed around the mass will accelerate in towards it, relative to the observer; and so on.

One of the interesting things about the Schwartzschild solution is that it makes clear the extent to which even *light* is affected by matter. Since light is massless, you might expect that it would be unaffected by gravity. However, light is in free fall as well, if it's moving only under the influence of gravity, and it follows a geodesic just as matter does, so if it's moving through a curved spacetime then its path will be curved. Near the Sun, this is a tiny effect, but it *is* observable.

This solution becomes interesting and exotic when you manage to squeeze all the mass in the problem into a tiny volume. If you were to compress the entire mass of the Sun into a sphere 6 km across, or an entire galaxy into a sphere the size of the solar system, you would

create a *black hole*. As you go closer and closer to the centre of a black hole, spacetime becomes more and more distorted until, when you cross the *event horizon*, it is so distorted that there is no possible route back out of the horizon. Nothing that falls past the horizon, including light, can ever get out again, and has no option but to move to smaller and smaller radii from the centre, until they eventually disappear from the universe.

What would you see if you were falling into a black hole? You would notice nothing particularly odd, as your watch would carry on working as before, and so on. Even as you crossed the event horizon, you would notice nothing odd – there isn't any 'surface' there, for example – but once you had crossed the horizon you would be doomed, as there is nothing you could do to push yourself back past it again. Just about the only thing you would notice would be the 'tidal' forces due to the difference in 'gravity' between your head and your feet, say. If the black hole were a rather small one, made from our Sun, for example, the tidal forces near the event horizon would be sufficiently strong to pull you apart. If all the matter in our galaxy were compressed into a volume the size of our solar system, however, that would make a black hole with an event horizon large enough that you could fall across it without noticing anything amiss.

What would be seen by an observer watching you from a safe distance? As you fell in, they would see time, as measured on your watch, slow down; they would see you fade away as the light from you was shifted into red (losing energy as it climbs out of the gravitational well). But they would never actually see you cross the event horizon: as you got closer and closer to the horizon, they would see your watch – measuring the passage of time for you – get slower and slower, relative to theirs and, from their point of view, they would never see the time when you actually crossed the horizon.

## 3.1   The Robertson-Walker solution and expanding universes

The Robertson-Walker (RW) spacetime is another of the solutions to Einstein's equations. Whereas the spacetime in the Schwartzschild solution is unchanging, the RW spacetime evolves in time. Specifically, one of the possible solutions is for it to *expand* as time goes on. This, or one of its variants, is the spacetime most usually thought to reflect the large scale structure of our own universe, whilst the Schwartzschild solution is the one most often used used to model the spacetime round isolated bodies.

If the universe is expanding, it must have expanded *from* something. This 'something', commonly called the Big Bang, must have been a hugely dense collection of energy, expanding at a huge rate. Whether it was created like that, or whether it bounced back from a 'Big Crunch' – a collapsing universe – is, really, still an open question.

Our universe is pretty certainly expanding still. We can tell this because the light from distant galaxies is *redshifted*, meaning that it is made more red by those galaxies' recession from us in much the same way that a sound is moved to lower pitch by its source moving from us. We can tell how quickly galaxies are receding from us by measuring that redshift, and we can make estimates of how far away they are. The two measurements are related, in that a galaxy twice as far from us as another will be measured to be receding from us twice as quickly. This effect was first noted by *Edwin Hubble* at the beginning of this century, and the parameter that relates the velocity and distance is known as the *Hubble constant* (a misnomer, as it may not be constant in fact). It is of crucial importance to astronomy, as it has a very direct connection with the rate of expansion in the RW spacetime, and thus with the age of our universe. It is also, unfortunately, particularly difficult to measure, and current estimates can disagree by a factor of two.

## 3.2  The Big Bang

The Robertson-Walker solution allows us to 'look back in time'.

After establishing the properties of the universe now, we can unwind the evolution of the Robertson-Walker universe, and try to pin down the properties the universe must have had at various earlier times. Using physics that we understand fairly well, we can account for the evolution of the universe back to times only $10^{-43}$ s after the creation of the universe. At the point when the Hubble expansion took over (around $t = 10^{-32}$ s) the universe was only 0.1 m in size. Since energy is conserved (one of the physical principles that survives even into this extreme situation), all the energy that is in the present huge universe will have been concentrated into a tiny volume, which consequently had a *huge* temperature.

As the universe expanded, it cooled down, and as it did so, it became cool enough for first protons and neutrons, then atomic nuclei, and finally atoms to exist (before that, the universe was filled with a gas of radiation and subnuclear particles). At around this time, the matter and the radiation in the universe stopped being as closely bound together as they had been before, and they continued their evolution separately. The matter in the universe started to clump together to form the heirarchy of galaxies, galaxy clusters, and superclusters, that we see today (the details of this 'structure formation' form one of the most important areas of modern cosmology). The radiation simply carried on expanding with the rest of the universe, and it is this radiation that we observe today, as the Cosmic Microwave Background.

The microwave background presents a great problem, however. No matter which direction we look in, the microwave background is the same – it is *homogeneous* – to a very high degree of accuracy. For this to be the case after all the expansion it has undergone, it must have been almost unbelievably homogeneous before. Related to this, a good deal of evidence points to our universe having just the right amount of mass to make its curvature, on the very largest scale, almost zero – the universe is flat. Again, this is so intrinsically unlikely that it calls for an explanation.

The currently believed explanation is that, before the Hubble expansion got under way, some small part of the tiny universe that was there suffered a period of *inflation*, during

which it experienced a huge increase in size (something like 50 orders of magnitude), and the homogeneity of the universe at present is explained as a consequence of the (easily explained) homogeneity of a small part of a tiny universe.

This has been a particularly rapid account of the Big Bang, which I've included as an example of how GR can be applied. For further details, I'll refer you to other popular accounts of the subject.

# A    Some of the details

When discussing relativity, one has to tread a fine line between a presentation which is chummy but inaccurate, and a presentation which is pedantically accurate but about as accessible as a conveyancing document.

I hope I have trodden a middle line, and produced an account which is accurate, but made accessible by carefully skimming subtleties. These subtleties are more likely to bite you, however, the more deeply you think about the topic. That bite can turn septic, and send you off into delirium.

This is why I have added this section. It contains rather more detailed discussions of some knotty points which crop up in the earlier part of these notes. It is not meant to be immediately intelligible, and I will not refer to it in the evening lectures, but I hope it will provide a resource when you are puzzling over the obscurities in this or some other account of relativity.

## A.1  Inertial frames

Above, I used a station as an example of an inertial frame.

To be strictly correct, the station is *not* an inertial frame, as long as the force of gravity is present. This is not an issue until we start to discuss general relativity, however, and as long as we confine ourselves to motion on a flat surface, the platform is equivalent to an inertial frame.

Similarly, we should be careful when talking about throwing balls or juggling (as I do repeatedly) within an inertial frame (and to be correct should confine ourselves to discussion of clearly manifest forces such as springs or rockets). However, as long as we are talking about SR rather than GR; as long as all the relevant motion (of inertial frames) is horizontal; and as long as no-one throws the ball further than a hundred miles or so (!), denying ourselves any mention of projectile motion would achieve nothing beyond removing a vivid and natural example to focus on. If you really want to, you can remove gravity from the examples by imagining the events take place in a free-falling space capsule, with some suitably baroque arrangement of downward jets of air or rocket packs, to supply the forces.

I don't want to make a great big deal about inertial frames, but if you're still a little puzzled by them, the following remarks might help.

We need to understand first what a *reference frame* is, then what is special about an *inertial* (reference) frame, and finally what is different about the way that special and general relativity treat the notion of inertial frames.

A *reference frame* is simply a way of assigning a position to events. The scheme that is possibly most familiar to you is that of map references: every point on the earth can be specified by a latitude and a longitude. For example, the centre of Glasgow is roughly $55°52'$ north of the equator, and $4°18'$ east of the Greenwich meridian. The lines of latitude and longitude constitute a reference frame centred on, and fixed to, the centre of the earth. You could specify any point in the universe using those coordinates.

Similarly, the distance posts that lie along railway tracks give the distance to that point from the last station (or is it the last signal box?). This is another reference frame – you

could specify any point by giving its distance along the track from Queen Street station, say. Reference frames need not be fixed to a stationary body, though. A train driver most naturally sees the world in terms of distances in front of the train. An approaching station can quite legitimately be said to be moving through the driver's reference frame.

You can generate an indefinite number of reference frames, fixed to various things moving in various ways. However, we can pick out some frames as special, namely those frames which are *not accelerating*, in a fairly obvious sense. In SR we can complete the definition by saying that the movement is horizontal, and we can ignore gravity. We have to do a little more work to make the definition precise, but this captures the essence of the notion as far as SR is concerned.

## A.2 The light clock – calculating time dilation

In Sect. 1.4.1, I described how time as measured by an observer with the light clock was related to time as measured by an observer on the platform. Here, I want to walk through this brief calculation. If the following makes no sense to you, that doesn't matter. The maths really does not add anything. The point of adding this section is merely to demonstrate that the maths is far from sophisticated, and to satisfy the curiosity of anyone who is comfortable with the maths (I'm aware that me saying "this involved only simple maths" is like someone saying to me "but it only involves *very* simple Greek").

We know the light travelled at a speed $c$ (the second axiom tells us that, no matter where or how the light was travelling – in a clock sitting beside us on the platform, or a clock zooming though the station in front of us – we would *always* measure its speed to be the same number $c$), and we timed its round trip at $t$ seconds, so the light beam must have travelled a distance $ct$ in the time that the clock itself travelled a distance $vt$. But from Fig. 4 we can see that this distance is the distance from the emitter to the mirror and back to the detector, so that the distance from the emitter to the reflection is $vt/2$; using Pythagoras' theorem, we can write

$$\left(\tfrac{1}{2}ct\right)^2 = L^2 + \left(\tfrac{1}{2}vt\right)^2.$$

From Eqn. (1), we know that $2L = ct'$, or $L = ct'/2$, so

$$\left(\tfrac{1}{2}ct\right)^2 = \left(\tfrac{1}{2}ct'\right)^2 + \left(\tfrac{1}{2}vt\right)^2.$$

We can rearrange and simplify this to find

$$c^2t'^2 = c^2t^2 - v^2t^2.$$

And we can divide both sides of this by $c^2$ and rearrange to find

$$t' = t\sqrt{1 - \frac{v^2}{c^2}}. \tag{5}$$

# B   Booklist

[Schwartz and McGuinness, 1999] is a lighthearted, but nonetheless very thorough, description of Special Relativity. It's very accessible, but not very detailed. [Gamow, 1965] is a whimsical account of the phenomena of relativity and quantum mechanics, which is entertaining, even if it is a little old-fashioned. Beyond these two, there seems a shortage of books which make serious popular (ie, non-mathematical) attempts to introduce special relativity.

[Davies, 1984] is a roller-coaster ride through some of the more exotic parts of particle physics and cosmology. It's hard going in places, but rewarding and accurate.

[Atkins, 1992] is a provocative and slightly eccentric book, which takes an exhilirating gallop round the fundamentals of the natural sciences, to leave you gasping at the conclusion that the universe has that of necessity in it, that it could slide unbidden and uncreated into existence. It doesn't set out to talk much about relativity, but it includes a number of insights into the topics we've covered here, as throwaway remarks.

Einstein's own popular account of relativity [Einstein, 2001] is very readable, though it's naturally a little old-fashioned in places. You can see the influence of this book, and its examples, in many later SR textbooks.

*The Principle of Relativity* [Lorentz et al., 1952] is a collection of (translations of) original papers on the Special and General theories, including Einstein's paper of 1905 [Einstein, 1905], but also some earlier papers by Lorentz suggesting interpretations of the Michelson-Morley experiment. The first few sections, at least, of the 1905 paper are worth reading as a current introduction to SR.

Finally, [Taylor and Wheeler, 1992] is a more advanced book. It's an undergraduate textbook, and so it doesn't pull any punches when it comes to the maths. The level of maths required is, however, relatively modest, so if you can cope with that and are prepared for some hard thought, you will benefit from the book's immense insight.

The Open University course *Space, Time and Cosmology* [Open University, 1998] is an advanced undergraduate course which covers much the same ground as this block. I drew on

its predecessor [Open University, 1979] in preparing these notes.

# References

[Atkins, 1992] Atkins, P. W. (1992). *Creation Revisited*. Penguin.

[Collins and Pinch, 1993] Collins, H. and Pinch, T. (1993). *The Golem: What Everyone Should Know about Science*. Cambridge University Press.

[Davies, 1984] Davies, P. (1984). *Superforce: the Search for a Grand Unified Theory of Nature*. Unwin.

[Einstein, 1905] Einstein, A. (1905). Zur Elektrodynamik bewegter Körper (on the electrodynamics of moving bodies). *Annalen der Physik*, 17:891. Reprinted in [Lorentz et al., 1952].

[Einstein, 2001] Einstein, A. (2001). *Relativity: The Special and the General Theory*. Routledge Classics.

[Gamow, 1965] Gamow, G. (1965). *Mr Tompkins in Paperback*. Cambridge University Press.

[Lorentz et al., 1952] Lorentz, H. A., Einstein, A., Minkowski, H., and Weyl, H. (1952). *The Principle of Relativity*. Dover.

[Open University, 1979] Open University (1979). *S354: Understanding Space and Time*. Open University Press. Replaced by [Open University, 1998].

[Open University, 1998] Open University (1998). *S357: Space, Time and Cosmology*. Open University Press.

[Schwartz and McGuinness, 1999] Schwartz, J. and McGuinness, M. (1999). *Introducing Einstein*. Icon Books. Used to be known as 'Einstein for Beginners'.

[Taylor and Wheeler, 1992] Taylor, E. F. and Wheeler, J. A. (1992). *Spacetime Physics*. W H Freeman, 2nd edition.