1
Introduction

Aims: you should:

1.1. understand the importance of events within Special Relativity, and the distinction between events and their coordinates in a particular frame; and
1.2. appreciate why we have to define very carefully the process of measuring distances and times, and how we go about this.

1.1 The Basic Ideas

Relativity is simple. Essentially the only *new physics* which will be introduced in this text boils down to two axioms:

1. All inertial reference frames are equivalent for the performance of all physical experiments;
2. The speed of light has the same constant value when measured in any inertial frame.

The work of understanding relativity consists of (i) appreciating what these two axioms really mean, (ii) examining their direct consequences, and (iii) thus discovering the way that we have to adjust the physics we already know. Each of these steps presents challenges.

So when I say that 'relativity is simple', I do not mean that it is easy, simply that this list of axioms is a short one. We will discover that these axioms are more subtle than may at first appear, and that they lead to conclusions which go against our usual intuitions. It is here that the difficulties arise: the maths isn't particularly hard, but we have to put a lot of effort into understanding ideas we thought were already clear, and try to think precisely about processes we thought were intuitive: what do we mean when we talk about 'the length of a stick'?

Our first step is to understand the axioms, and we'll start on that in the next chapter. What does it mean to talk about a 'transformation between frames', and why should the speed of light have such a significant place in this story?

Chapters 3 and 4 are about the reasonably direct consequences of the axioms. It's in this pair of chapters that we discover the most surprising features of SR – length contraction and time dilation – first qualitatively then quantitatively. This pair of chapters is where the most profound conceptual challenges are.

The various strands here are tied together by the main calculational tool of SR, the 'Lorentz transformation', which we derive and study in Chapter 5. This chapter is quite a long one, and detailed, so that it would be fairly easy to get lost in it. However it is really only Section 5.1 that has the new material, and the rest of the chapter is, again, an exploration of the consequences. Those consequences are often surprising, and will frequently, I think, cause you to re-read and re-think Chapters 3 or 4.

In Chapters 6 and 7 we look outwards to the rest of physics, and learn how the principles of SR oblige us to recast familiar ideas such as velocity, acceleration, frequency, momentum, mass, and energy.

Finally, in Appendix A, we look at how we can apply these ideas more generally, and discover Einstein's theory of gravity: General Relativity (GR). We can't dive too far beneath the surface here, without more advanced mathematics, but our understanding of SR will allow us to at least see the connection to the main structures of GR, and how they link to gravity.

Prior to all that, however, there are a couple of bits of terminology that it's useful to clarify in this first chapter, to avoid breaking the flow of the argument in Chapter 2. In particular:

Events These are things like a light-flash, or a bang, which happen at a particular place and time.

Reference frames These are the coordinate systems that we use to describe where and when events happen. We are concerned, in SR, with the *multiple* reference frames that may be relevant in our analysis of the physical world, and how the measurements in these frames relate to each other. We need to pay particular attention to the special case of the 'inertial frame'.

Measuring lengths and times We have to be quite careful about how we measure distances, in space or time. The obvious ways of doing this contain ambiguities which can easily lead to confusion.

I expand on each of these points below. The following sections are rather short (and possibly a little dry), but we will use the ideas in them again and again and again.

It's probably a good idea to re-read these sections repeatedly as you work through the rest of the text. It possibly follows that these remarks might not make perfect sense first time; they may even at first seem absurdly over-precise, since they are making distinctions which may appear unnecessary until you have understood some of the rest of the material.

1.2 Events

An *event* in SR is something that happens at a particular place, at a particular instant of time. The standard examples of events are a flashbulb going off, or a banger or firecracker exploding, or two things colliding.

There is nothing *relative* about an event: if two cars crash and metal is bent, there is no 'point of view' from which the crash did not happen. Although this may seem at this point to be too obvious to be worth stating, we will discover that more things than we may expect are relative to our 'point of view', and we will use events as our way of navigating through the puzzles this produces.

We will soon discover that, although we can all agree that a particular event did happen, we might well have different answers to the questions 'where?' and 'when?' These are questions that we can answer using a *reference frame*. [Exercise 1.1]

1.3 Inertial Reference Frames

We need to understand first what a *reference frame* is, and then what is special about an *inertial* (reference) frame.

A *reference frame* is simply a method of assigning a position, as a set of numbers, to events. Whenever you have a coordinate system, you have a reference frame, and I will use the two terms almost interchangeably. The coordinate systems that spring first to mind are possibly the (x, y, z) or (r, θ, ϕ) of physics problems. Reference frames need not be fixed to a stationary body: 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 – speeding up and slowing down – in the driver's reference frame.

In mechanics problems, we are used to thinking of time as a free variable: often, the point of a physics problem is to work out how the position of a thrown ball, for example, varies in time – what is $x(t)$? When thinking of events, however, it can be useful to think of them as being located using *four* coordinate numbers, (t, x, y, z). We will come back to this point of view in Chapter 4.

You can generate any number of reference frames, associated with various things moving in various ways. In the context of SR, however, we can pick out some frames as special, namely those frames which are *not accelerating*.

Imagine placing a ball at rest on a table: you'd expect it to stay in place. Similarly, if you roll a ball across a table, you'd expect it to move in a straight line. This is merely the expression of Newton's first law: 'bodies move in straight lines at constant velocity, unless acted on by an external force'. In what circumstance – that is, in which frames – will this *not* be true?

Suppose you're sitting in a train which is accelerating out of a station.[1] A ball placed on a table in front of you will start to roll towards the rear of the train, rather than staying put in the way that Newton's first law seems to say it should. This observation makes perfect sense from the point of view of someone on the station platform, who sees the ball as stationary, and the train being pulled from under it. But in a reference frame attached to the train, where 'position' is perhaps measured as the distance, $x(t)$, from the rear of the train, this position will change without any force acting on the ball. We refer to the station as an 'inertial frame', and the accelerating train carriage, with respect to which Newton's law appears not to hold, as non-inertial. Similarly, if you are perched on a spinning children's roundabout, and toss a ball to someone on the opposite side, it veers off to one side (interpreting this as either 'it appears to veer off to one side, from your point of view' or, more formally, 'it will be measured to veer off, as observed by someone using the rotating reference frame which is fixed to the roundabout'). This motion, again, is immediately intelligible from the point of view of someone standing watching all this go on, who sees the ball go exactly where it should, but the catcher rotate out of the way. The playpark

[1] We're going to hear an awful lot more about this train. Although I will occasionally vary the examples by talking about rockets or boats, a train going past a station platform presents such an immediate picture of two reference frames, in constant relative motion, that it will be hard to avoid. I see no reason for wanton innovation with respect to this particular aspect of the subject.

is an inertial frame, the spinning roundabout is not. In both cases, you can tell whether you're the one in the non-inertial frame: in the first case you feel yourself pushed back into the train seat, and in the second case, it's only your grip on the roundabout that stops you flying off, pulled towards the outside by centrifugal 'force'.

Acceleration and force are intimately connected with the notion of inertial frames – an inertial frame is one which isn't accelerated in any way. From that, you would be correct to conclude that once the train has stopped accelerating, and is speeding smoothly on its way, it becomes an inertial frame again; if you closed your eyes, you wouldn't be able to tell if you were on a moving train or at rest in the station. Anything you can do whilst standing on a station platform (such as juggling, perhaps), you can also do whilst racing through that station on a train, even though, to the person watching the performance from the platform, the balls you're juggling with are moving at a hundred kilometres an hour, or so.

What we have concluded here is that, although different observers may reasonably ascribe different coordinates to events, and different speeds, there is no ambiguity about who is accelerating or not. If you are on a train picking up speed as it leaves a station, you can feel the pressure of the seat on your back, and be under no illusion that you are not moving, and there is no point of view from which the drink on the table in front of you does not look likely to spill. We will have more to say about this point in Section 2.1.

Newton's second law is more quantitative, since it relates the amount of force applied to an object, the amount it is accelerated, and the body's inertial mass, through the well-known relation $F = ma$.

We can therefore define an 'inertial frame' as follows:

> *Definition of Inertial Frames:* An *inertial reference frame* is a reference frame, with respect to which Newton's first law holds.

All this being said: don't over-think this. An inertial frame is one that isn't accelerating.

Note that, in the context of SR, inertial frames are infinite in extent; also, since all inertial frames move with constant velocity, it follows that no pair of such frames mutually accelerate. [Exercise 1.2]

1.3.1 Further Remarks on Frames ⚠

Of course, there is rather more to it than that. This definition suffices for Special Relativity, but once we consider *General Relativity* (GR) we have both the need, and the mathematical tools, for a more fundamental definition.

In brief, in GR the definition of an inertial frame is one which is in free fall, meaning one which is moving freely in a gravitational field, or freely floating, unaccelerated in interstellar space. As in SR, this is a frame in which Newton's first law still holds. This definition is locally consistent with the definition in SR, but allows us to start to discuss inertial frames which are mutually accelerating (in the specific sense that the second derivative of the separation is non-zero), such as two free-falling objects on opposite sides of the Earth. I discuss this at greater length in Appendix A.

If we acknowledge the existence of GR then, to be nit-pickingly precise, I shouldn't really talk of train carriages and station platforms as inertial frames. Firstly, there are tiny corrections due to the fact that we are on the curved surface of a rotating planet; we can ignore these in the huge majority of cases. Secondly, we should be careful when talking about throwing balls or juggling within an inertial frame, since, because of the presence of the force of gravity, a frame sitting on Earth is not inertial according to GR's stricter definition. 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 kilometres 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 taking place not in train carriages going through stations, but in space capsules flying past asteroids, with some suitably baroque arrangement of air jets or rockets, to supply the forces when necessary.

Also, this section is one of the few places in this text where I mention 'acceleration' (another is coming up in Section 1.6, when I talk of the 'clock hypothesis'). This is not because SR 'cannot deal with acceleration' as you might see mistakenly claimed, but simply because the novel and counter-intuitive relativistic effects, that we are going to discover in the chapters to come, are not a result of accelerating frames or observers, but instead of non-accelerating (that is, inertial) observers in mutual motion at large relative velocity. Similarly, when we (implicitly) talk of motion under gravity – such as when we talk of throwing a ball – it is perfectly reasonable for us to do so using a newtonian theory of gravity ($F = mg$) rather than anything distractingly exotic.

The mass which is the constant of proportionality in Newton's second law is what defines *inertial mass* – it indicates, roughly speaking, how resistant an object is to being accelerated by an applied force. This is distinct from the *gravitational mass* of an object, which describes how much gravitational field the object generates – in Newton's law of gravitation, $F = GMm/r^2$, both masses are gravitational masses. It turns out, however, that whatever the composition or construction of an object, its gravitational and inertial masses, though logically completely distinct, are always measured to be equal. This fact is more surprising than it may at first appear; we will examine some of its consequences in Appendix A.

1.4 Simultaneity: Measuring Times

How do we measure times? In SR, we repeatedly wish to talk about the times at which events happen, and more particularly the time *intervals* between events. In Chapter 3, we will discover that two observers in relative motion, who observe a pair of events, will not only ascribe different time coordinates to those events (unsusprisingly, because they are in different coordinate systems), but may also disagree about which event happens first, or whether they are simultaneous. We must therefore be careful just what we mean by ‘the time of an event’.

One of the things we can hold onto in the rest of this text is that, if two events at the same spatial position happen at the same time, they are simultaneous for everybody. Einstein made this particularly clear, when he talked about what it means to assign a ‘time’ to an event:

> 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, however, the event happens some distance away (answering a question such as ‘what time did the train pass the next signal box?’), or if we want to know what time was measured by someone in a moving frame (answering, for example, ‘what was the time on the train-driver’s watch as the train passed the signal box?’), things are not so simple, as most of the rest of this text makes clear. Special Relativity is very clear about what we mean by ‘the time of an event’: when we talk about the time of an event, we *always* mean the time of the event *as measured on a clock carried 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?), who is stationary with respect to the frame they represent. We will typically imagine more than one observer at an event; indeed we imagine one local observer per frame of interest, stationary in that frame, and responsible for reporting the space and time coordinates of the event as

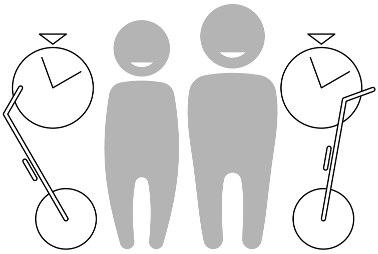

Figure 1.1 Our observers, equipped with their clocks and surveyors' wheels.

measured in that observer's frame.[2]

What we never do is have an observer in one location measure and report the time of an event at another location. To do so, we'd have to concern ourselves with a number of complications, most prominently correcting for the time of flight of the light from the event to our observer's eye, which depends on the distance between them, and so on. It would be possible to be clever and correct for these, but we simply avoid doing so by exclusively using local observers.

We suppose that a frame has a plentiful supply of observers, and that we can position them, and their clocks, wherever we need to make an observation (we're going to take 'observation' and 'measurement' to be synonyms here). Or you can imagine observers positioned *everywhere* in a frame, ready to take note of any events which happen next to them. Our observers – see Figure 1.1 – are both lazy and very short-sighted: once they have established their location they never move, and they *will only ever observe events which happen right in front of them*. When an event happens, they note the time on their clock, ready to report it along with their measured position when required.

Since the observers never move, the frame they are standing in is special to them, it is the *rest frame* of the observer. At the risk of belabouring the point, if we have two observers, one on a train and one on a station platform, then the rest frame for one observer is the frame attached to the station (within which neither the station nor the observer are moving), and the rest frame for the other is attached to the train (within which the train is not moving, but the station is).

The observers in a particular frame have one further property of impor-

[2] Note that it makes no sense to talk of being 'stationary with respect to an event' or to talk of 'the rest frame of an event': since an event is instantaneous, it cannot be said to be moving in any frame. This also means that there is no observer who is 'special' with respect to an event.

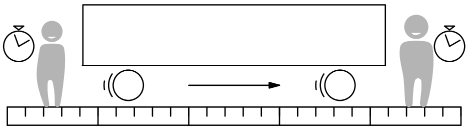

Figure 1.2 Measuring the length of a train, using a ruler painted on the edge of the station platform. The observers are standing on the platform. You should imagine observers all along the platform; I've shown only the two who happen to be next to the ends of the train at the observation time.

tance: as well as being stationary in the frame, we presume that they have arranged things so that the clocks they carry are, and remain, mutually synchronised. I have a little more to say about this in Section 2.2.1, but I mention it here only to reassure you that this is not on the list of unexpectedly complicated things.

Although an event might have a number of adjacent observers, in different frames, this doesn't imply that all observers report the *same* time. The observers' watches may differ for trivial reasons – perhaps their watches are set to different time zones, or designed to run at different rates. Or, less trivially, they may be running at different rates for relativistic reasons that we'll come to later. We assume that, despite these complications, all of the clocks tick out a time – produce a *number* – which is linearly related to the passage of time. [Exercise 1.3]

1.5 Simultaneity: Measuring Lengths

How should we measure the length of a moving object such as the train carriage of Figure 1.2? The obvious method is some variant of laying out a metre stick or measuring tape, and looking at how the markings line up with the moving carriage. But doing so raises some awkward questions. I said above that we want to avoid making any non-local measurements, and make all observations with co-located observers. How do we go about that in this situation? You obviously have to measure both ends of the object simultaneously: does that 'simultaneously' depend on where you're standing relative to the object? There are questions here which are only apparently straightforward, and to which SR provides surprising answers.

The way we measure lengths and times in SR is therefore as follows. We position observers at strategic points in the reference frames of interest. We can know these observers' coordinates in one frame or another (I find it useful to imagine the x-axis painted on the platform edge). The observers make records of events which happen at their location, and afterwards compare notes and draw conclusions. Specifically, you would measure the ‘length of a train’ by subtracting the coordinates of the two observers who observed opposite ends of the train at a pre-arranged time (remember that we are assuming that all of the observers in a particular frame have pre-synchronised clocks). We return to this in Chapter 3.

If these two observers are inside the train carriage, stationary at opposite ends, then they get the value you would intuit: this is merely a complicated way of measuring the length they could obtain by stretching a tape-measure between them. For the case of the observers on the platform measuring the moving train, this does initially seem a complicated way of organising things, but it has the virtue of being unambiguous about when things happen and where.

To summarise, this approach relies on three things.

1. It requires a specific procedure for synchronising clocks, which is described briefly in Section 2.2.1.
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 – that they are ‘good clocks’ in the specific sense of the ‘clock hypothesis’ (see Section 1.6).

The second point in this list does need saying, possibly surprisingly: we will later discover that the simultaneity of two events not at the same position *does* depend on your point of view.

1.6 The Clock Hypothesis

The *clock hypothesis* is the assumption that there is nothing intrinsic to motion, or to acceleration, which stops a clock being a reliable measure of the passage of time.

The first (rather trivial) part of this is the presumption that, whatever the technology we're using for our clocks, it's appropriate for the situation, so that we're not, for example, using a pendulum clock in space or on board ship, or a clock which will snap if we accelerate it. A very concrete example of a ‘good clock’ is an atomic clock, which depends on fundamental physics