Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 1
Introduction; Newtonian mechanics and relativity

1.1 Goal of this course

Our goal this semester is to understand Albert Einstein's theories of relativity, and what they tell us about the structure of the laws of physics. We will study the special theory of relativity in great detail for most of the semester. In the last several weeks of the term, we will briefly explore general relativity, focusing on situations in which what we learn from special relativity can be "upgraded" to the general case with relative ease.

More generally, our goal this semester is to think about how we formulate physics in such a way that a deep underlying principle is built into these laws. During the first few weeks of this course, we will motivate how on both theoretical and experimental grounds we came to understand that our universe respects a principle we call *Lorentz symmetry*. We will find that some of the laws of physics we learned previously exactly respect this symmetry, whereas others are approximations (albeit incredibly accurate approximations under the "everyday" conditions that one typically encounters in our day-to-day life, and even in cases where very precise measurements can be made). We will then develop a way of representing physical laws such that they automatically satisfy Lorentz symmetry. You can take this as a kind of "warm-up exercise" for including other symmetries that nature respects, and other principles that we may need to build into physics.

1.2 Newtonian physics and Galilean relativity

Let us begin by examining Newtonian physics. Newton's laws obey **a** law of relativity, though it is not the one that we usually think of when we discuss "relativity" in physics. Rather than Einstein's relativity, Newton's laws respect what we call *Galilean relativity*, following principles that were originally laid out by Galileo. We will begin by examining Galilean relativity in order to see a relativity principle in action, based on laws of physics that we thoroughly know and love ... and to see an interesting shortcoming we quickly find when we combine Galilean relativity with physics that we encounter early in our physics studies.

To begin this discussion, we need to define some terms:

- **Event**: An "event" is something that happens somewhere at some time. An event is essentially a particular point in both space and time. A key feature is that the event's reality is independent of how we label it, which can depend on our "reference frame."
- **Reference frame**: A system for labeling events in space and time. One can regard a reference frame as essentially a clock and a set of coordinate axes that are tied to a particular observer. For example, the professor, standing in front of the classroom uses a clock on the wall to define time, and takes the position of their feet as the origin. They imagine an x axis pointing from their feet toward the back of the classroom; a

y axis pointing from the professor's feet to their left, and a z axis pointing from the professor's feet up to the ceiling. (We have carefully defined the axes so that they form a *right-handed* coordinate systems. If you are unfamiliar with this concept, please ask one of the staff for a clarification.)

A student, sitting in the front row, sets up a similar reference frame. Also using the wall clock for time, the student likewise defines their position as the origin, imagines an x axis that points from their feet to the *front* of the classroom (they are facing in the opposite direction as the professor, so "forward" for them is opposite of "forward" for the professor), a y axis pointing from their feet to their left, and z axis pointing from their feet to the ceiling.

These two reference frames assign different labels to events, but both are perfectly valid provided they are used consistently.

⌈ASIDE: This is a good point to introduce some notation and definitions. Once we've introduced coordinate axes, it is very useful to define *unit vectors* which point along these axes. We will call the unit vectors along the x, y, and z axes $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ respectively. $\mathbf{e}_x$ is a dimensionless vector of magnitude 1 that is parallel to the x axis; likewise for the other two unit vectors. If more than 1 reference frame is being discussed, we will include some kind of label to distinguish them; e.g., $\mathbf{e}_x^P$ is the x unit vector for the professor's reference frame; $\mathbf{e}_x^S$ is the x unit vector for the student's reference frame.

In our typed-up notes, we will use boldface to denote vectors in 3-dimensional space. This does not work well in chalk, so we will instead use an undertilde (e.g., $\utilde{e}_x$) when we write such vectors on the chalkboard.⌋

- **Geometric object**: Something with properties that exist independent of the reference frame that we use to describe it.

An example of a geometric object is an event. Suppose that at 2:47 pm on 4 September 2024, a piece of chalks strikes the professor's forehead. All observers note this event. They may assign different labels to it — the professor calls it $t = $ 14:47 04-09-2024, $x = 0$, $y = 0$, $z = 1.8$ meters; the student calls it $t = $ 14:47 04-09-2024, $x = 3$ meters, $y = -1$ meter, $z = 1.8$ meters. Our labels differ, but they describe exactly the same thing. This very much like the way in which different languages use different sounds to describe the same thing: whether you call a bound collection of pages and words a book, un libro, ein Buch, or any of the many other words used by people around the world, you know what it is.

Another example of a geometric object is a vector. Consider a meter stick that points from a table at the front of the classroom. All observers agree[1] that it has a length of 1 meter, and all agree that is poking out of the table at some angle. However, depending on the reference frames being used, different observers will use different *representations* of that vector. This is of course fine as long as the different representations are used consistently in describing the physics of the system under study. Figure 1.2 illustrates this concept in a simple 2-dimensional example.

[1]The agreement among different observers about the length of this stick in this example won't hold up once we move beyond Galilean relativity! Hold that thought for now.

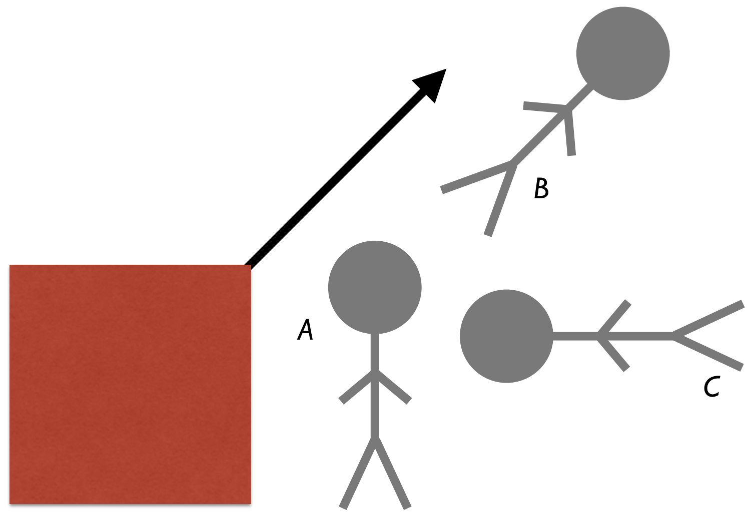


Figure 1: Three observers, three different inertial reference frames. (Imagine they are in a gravity-free environment so that each observer is inertial.) Each agrees that there is a large stick embedded in the table, making a 45° angle with its top and side. Observer A orients it along the direction $(\mathbf{e}_x^A + \mathbf{e}_z^A)/\sqrt{2}$; observer B orients it along $\mathbf{e}_z^B$; and observer C orients it along $(\mathbf{e}_x^C - \mathbf{e}_z^C)/\sqrt{2}$. These different representations are consistent, provided we correctly relate each observer's choice of coordinate axes to those of the other observers.

Much of relativity, whether it is Galileo's or Einstein's, is about making sure that we carefully and consistently describe things in different reference frames, and that we correctly relate the description of quantities according to one reference frame to those quantities according to another reference frame. Geometric objects are excellent tools for describing physics because objects like events, vectors, and tensors (which will be introduced and defined soon) have a meaning that transcends a particular reference frame's representation of that object. It is very important (and useful) to maintain a distinction in your mind between the object (e.g., a particular vector) and the representation of that object (e.g., the components of that vector according to some given frame of reference).

A particularly important reference frame is an *inertial reference frame*, or IRF; we will use this term enough that it is worth abbreviating. This is a frame of reference in which a body's momentum is constant if no forces act on it: It is an *unaccelerated* reference frame. Non-inertial reference frames certainly exist, and require us to introduce what are sometimes called "fictitious" or "non-inertial" forces[2], like the centrifugal force or the Coriolis force.

1.3 Galilean transformations

How do we relate quantities as described in one reference frame to those in another? We can deduce how to do this by thinking about the (hopefully, straightforward) connection between how the two frames relate the representations of geometric objects. Doing so, we build a mathematical "machine" for connecting quantities between two reference frames. In

[2] I prefer "non-inertial" to "fictitious," since "fictitious" sounds quite a bit like "fake." Anyone who has ever crashed a bike taking a turn too fast or hurt their neck on an amusement park ride can tell you that there is nothing at all fictitious about those forces if you happen to be in the non-inertial frame.

Newtonian physics, we call the resulting mathematical machine the *Galilean transformation*.

Let us say IRF C is used by the class to label objects and events; IRF P is used by the professor. The professor and the class are oriented in the same way, so that their x, y, and z axes all point in the same direction. However, the professor is walking across the classroom: the class sees the professor moving with constant speed v in the x direction. How do we relate these two IRFs?

Consider time first. Is there any difference in time according to the class and to the professor? In Newtonian physics, the answer is no: Both the class and the professor get their time from the wall clock, no matter whether they are in motion or not. So we have

$$t_P = t_C \ . \tag{1.1}$$

Their representations of space differ, however. An object at a fixed position in IRF C "falls back" along x in IRF P:

$$\begin{aligned} x_P &= x_C - vt_P \\ &= x_C - vt_C \ , &(1.2) \\ y_P &= y_C \ , &(1.3) \\ z_P &= z_C \ . &(1.4) \end{aligned}$$

(Notice that x_P and x_C coincide when $t_P = t_C = 0$. The moment at which coordinates coincide should be specified when the relationship between the IRFs is laid out.) The set of four equations, (1.1)–(1.4), relating quantities in P to quantities in C is a Galilean spacetime transformation. It can be neatly written as a matrix equation:

$$\begin{pmatrix} t_P \\ x_P \\ y_P \\ z_P \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -v & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} t_C \\ x_C \\ y_C \\ z_C \end{pmatrix} \ . \tag{1.5}$$

Writing out 4×4 matrices like this gets unwieldly; a more compact form is

$$\vec{x}_P = \mathsf{G} \cdot \vec{x}_C \ . \tag{1.6}$$

⌈ASIDE: Here is another good point to introduce some more notation. When studying relativity, it will be useful to make "4-vectors," vectorial quantities with 4 components: the 3 spatial ones that you are probably familiar with from previous coursework, plus 1 more for a time component. In this course, whenever we write an object with an overarrow like $\vec{x}_P$, it refers to such a 4-vector. The same symbol in boldface or with an undertilde ($\mathbf{x}_P$ or $\underset{\sim}{x}_P$) refers to only the spatial components. (It is worth noting that other conventions exist, and we will briefly mention at least one of them when we discuss tensors. For 8.033, I will stick with the overarrow for 4-vectors and boldface/undertilde for 3-vectors.)⌋

How do we invert the relation we just wrote down? That is, given Eq. (1.5) relating quantities in frame C to those in frame P, how do we relate quantities in frame P to those in frame C? On the grounds of physics, this is simple: if C says that P moves with velocity $\mathbf{v} = v\mathbf{e}_x$, then P says that C moves with $\mathbf{v} = -v\mathbf{e}_x$. We quickly deduce that

$$\begin{pmatrix} t_C \\ x_C \\ y_C \\ z_C \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ v & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} t_P \\ x_P \\ y_P \\ z_P \end{pmatrix} \ , \tag{1.7}$$

which we will write more compactly as

$$\vec{x}_C = \mathsf{G}' \cdot \vec{x}_P \ . \tag{1.8}$$

It is easy to verify that $\mathsf{G}' = \mathsf{G}^{-1}$, i.e., that G' is the matrix inverse of G. This is an exercise on problem set #1.

Another example of a Galilean transformation: Suppose IRFs C and P are at rest with respect to each other, but are rotated about the z axis, as shown in Figure 1.3.

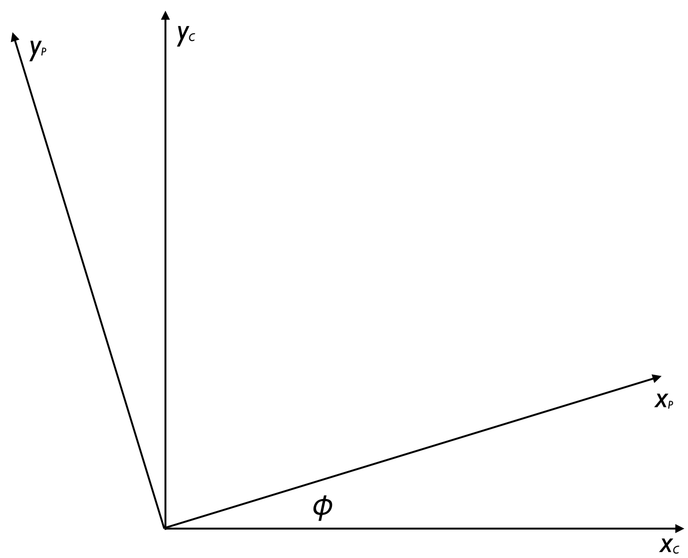


Figure 2: x and y axes for IRFs C and P, related to each by a rotation about the z (which is the same to both frames).

Space and time are related between the two frames with the equations

$$\begin{aligned} t_P &= t_C \ , & (1.9)\\ x_P &= x_C \, \cos\phi + y_C \, \sin\phi \ , & (1.10)\\ y_P &= -x_C \, \sin\phi + y_C \, \cos\phi \ , & (1.11)\\ z_P &= z_C \ . & (1.12) \end{aligned}$$

The spatial part of this transformation is just a simple rotation. We can write this

$$\begin{pmatrix} t_P \\ x_P \\ y_P \\ z_P \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & \sin\phi & 0 \\ 0 & -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} t_C \\ x_C \\ y_C \\ z_C \end{pmatrix} , \tag{1.13}$$

or

$$\vec{x}_P = \mathsf{G}_R \cdot \vec{x}_C \ , \tag{1.14}$$

where G_R denotes a Galilean transformation that's a pure rotation.

The examples discussed here are not comprehensive[3], but hopefully give you the gist of the idea. We will explore these concepts a bit more on the problem sets.

[3]There's another very simple one that is worth mentioning: A shift of origin. Suppose the professor and the class have the same orientation, but the professor center spaces on their location, and perhaps their watch is set to a different time zone. Then, $\vec{x}_P = \vec{x}_C + \Delta\vec{x}$, where $\Delta\vec{x}$ is a constant offset.

1.4 Transformation of velocities and accelerations

A key feature of the transformations we have been discussing is that they leave inertial frames inertial. To see this, let's go back to our first example, the professor moving with speed v in the x direction as seen by the class and examine how an object's *velocity* transforms under Galilean transformations. Consider an object moving with velocity $\mathbf{u}$, where $u^x = dx/dt$, $u^y = dy/dt$, $u^z = dz/dt$. (Notational note: We will use the letter v or $\mathbf{v}$ to denote the relative velocity of two different frames of reference; we will use the letter u or $\mathbf{u}$ to denote velocity within some specified frame.)

Suppose the class sees some object moving with velocity $\mathbf{u}_C$, with components

$$u^x_C = \frac{dx_C}{dt_C} \, , \qquad u^y_C = \frac{dy_C}{dt_C} \, , \qquad u^z_C = \frac{dz_C}{dt_C} \, . \tag{1.15}$$

What are the components as seen by the professor? Let's apply the Galilean transformation rules and find out:

$$\begin{aligned} u^x_P &= \frac{dx_P}{dt_P} \\ &= \frac{d}{dt_P}\left(x_C - vt_C\right) \\ &= \frac{d}{dt_C}\left(x_C - vt_C\right) \\ &= u^x_C - v \, . \end{aligned} \tag{1.16}$$

By a similar calculation, we see that $u^y_P = u^y_C$, $u^z_P = u^z_C$. This is nothing more than the "normal" velocity transformation that we are familiar with from Newtonian mechanics.

How about accelerations? As usual, we have $\mathbf{a} = d\mathbf{u}/dt$. Imagine that the object is seen by the class to have acceleration $\mathbf{a}$, and use the Galilean transformation to deduce what the professor sees for its acceleration:

$$\begin{aligned} a^x_P &= \frac{du^x_P}{dt_P} \\ &= \frac{d}{dt_P}\left(u^x_C - v\right) \\ &= \frac{d}{dt_C}\left(u^x_C - v\right) \\ &= a^x_C \, . \end{aligned} \tag{1.17}$$

We likewise find $a^y_P = a^y_C$, $a^z_P = a^z_C$: the class and the professor agree on the object's acceleration, at least as long as v is constant in time. As long as the two frames are not accelerated with respect to one another, Galilean transformations take one inertial representation and yield another inertial representation.

1.5 Relativity, covariance, and invariance

You have now been introduced to the workings of Galilean relativity. Having seen a few examples of how it works, this is a good opportunity to carefully define a few terms that we are going to use a lot in this course.

- *Relativity*: We've been using "relativity" quite a bit without actually defining it explicitly. A relativity framework is just a way of transforming observables — particularly the representation of geometric objects which we use to describe important quantities in physics — from the reference frame of one observer to that of another.

- *Covariance*: We describe a law or principle of physics as covariant if it holds in all frames of reference. For example, as you will explore on a problem set, the law of momentum conservation is covariant in Galilean relativity. This does not mean that observers in all inertial frames agree on the value of an object's momentum; indeed, you should be able to convince yourself that you can make that object's momentum take any value at all by changing frames of reference. However, all observers agree that if that body interacts with another body, then the momenta of the two bodies after their interaction is the same as it was before.

- *Invariance*: Some quantities are in fact exactly the same in all frames of reference. The mass of a body is the same to all observers in Galilean relativity; a particular notion of mass is the same to all observers in Einstein's relativity; a body's electric charge is the same to all observers in all forms of relativity. Quantities which are the same in all frames of reference are called *invariants*. Learning when and how to exploit invariance is one of the skills we will practice this term. Used well, invariants often make it possible to significantly simplify a calculation.

1.6 Transformation of waves in Galilean relativity

A particularly important example for our discussion is to consider how the representation of a *wave* is affected by a Galilean transformation. Let us first consider waves in general. We imagine there is some field F that propagates through space and has the functional form

$$F = F(x - wt)\ . \tag{1.18}$$

(More generally, we should have $F = F(\mathbf{r} - \mathbf{w}t)$, where $\mathbf{r}$ is a general displacement in three dimensions; we focus on the 1-dimensional limit for simplicity.) The field F depends on the specific physics of the wave under consideration: it could be the pressure of a sound wave, or the height of a water wave, or the displacement from equilibrium of an element of a spring, or ... Suffice it to say that *many* phenomena propagate as waves. The quantity w is the speed with which the wave propagates; its value also depends on the specific physics of the system under consideration.

A wave of this form satisfies the differential equation

$$\frac{\partial^2 F}{\partial t^2} - w^2 \frac{\partial^2 F}{\partial x^2} = 0\ . \tag{1.19}$$

It should be emphasized that it can take a fair amount of labor and analysis for the physics of the phenomenon under study to reduce to (1.19). This form could emerge from a detailed study of displacement and tension along a string, or a study of weight versus buoyancy in a fluid, to give two examples. Equation (1.19) which emerges as the typical outcome of performing all this labor and analysis is known as the *wave equation*.

Suppose that t and x in Eq. (1.19) are quantities as measured by the class. How does the wave behave according to the professor? On an upcoming problem set, you will examine this problem by applying a Galilean transformation to the wave equation. You will find that the wave equation changes such that $w \to w - v$. In other words, if the class describes the wave as

$$F = F(x_C - wt_C) , \tag{1.20}$$

then the professor describes the wave as

$$F = F\left(x_P - (w - v)t_P\right) . \tag{1.21}$$

This is as expected given our discussion of how velocities transform in Galilean relativity.

Equations (1.20) and (1.21) have a very important consequence: they tell us that there is a particular, special IRF in which the wave speed is w. This is the "rest frame" of the medium that supports the wave. For example, for a water wave, the wave's speed is as measured in the frame in which the water does not flow.

At the end of the 19th century, everything that we have discussed here was quite well understood. In particular, "natural philosophers" (which includes what we more or less think of as physicists today) of this time period had studied many wave phenomena, and all of them were of this form: a wave was a disturbance that propagated in some kind of medium, and the "natural" wave speed corresponded to the rest frame of that medium.

This lasted until Maxwell formulated the equations of electrodynamics that bear his name. Then things got interesting.

Nature and Nature's laws lay hid in night:
God said "Let Newton be!" And all was light.
Alexander Pope (1688 – 1744)

It did not last: The Devil howling "Ho!
Let Einstein be!" restored the status quo.
John Collings Squire (1884 – 1958)

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 2
Electromagnetic radiation and Galilean relativity

2.1 An aside ... why not accelerated reference frames?

Our discussion of Galilean relativity described in some detail how to relate quantities measured in one IRF to quantities measured in another IRF. However, one could imagine frames whose relative motion is accelerated. Could we not broaden our discussion to include such *non-inertial* relative motions?

Certainly we could include accelerations between reference frames. Doing so requires that we introduce *non-inertial forces* in order for Newton's laws, in particular $\mathbf{F} = m\mathbf{a}$, to work. Our main reason for not doing this is simplicity — including accelerations makes the analysis more cumbersome, and is a diversion from the main thrust of our discussion. It is, however, a well-developed topic, and interested students can certainly find discussions of this in many excellent textbooks.

What if the different frames are accelerated, but the frames experience *the same* acceleration? In such a case, one could imagine defining a transformation that takes us from one frame accelerating with $\mathbf{a}$ to another frame that is also accelerating with $\mathbf{a}$. Indeed, in this circumstance it is not hard to see that the Galilean transformations we discussed in the previous lecture work perfectly, translating quantities from one accelerating frame to the other. Given this, one might wonder: if everything experiences the same acceleration, does that acceleration mean anything interesting? Given that all geometric objects in all the frames that we consider experience the same acceleration, perhaps we could just define this as a somewhat peculiar notion of "rest."

This question in fact gets at the heart of the issues and concepts which lie at the core of Einstein's *general* relativity, hinting at the principle of equivalence. We will return to a very similar discussion in several weeks in a more Einsteinian context.

2.2 Galileo meets Maxwell

In our previous discussion, we noted that wave equations have an interesting property: the physics of the wave introduces a special speed, which we labeled w. This describes the speed with which the wave propagates with respect to the medium that supports the wave. An observer moving with respect to the medium will observe the wave propagating with a different speed, in accordance with how velocities add in Newtonian mechanics.

This made perfect sense until roughly the late 1800s. To see what started confusing the situation, consider Maxwell's equations:

$$\nabla \cdot \mathbf{E} = \rho/\epsilon_0 \;, \qquad \nabla \cdot \mathbf{B} = 0 \;, \tag{2.1}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \;, \qquad \nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \;. \tag{2.2}$$

Let us consider the vacuum limit ($\rho = 0$, $\mathbf{J} = 0$), and let us take the curl of the curl equations. Using some vector calculus identities, this is straightforward. Look at the curl of the left-hand sides of the curl equations first: using vector calculus identities, it is straightforward to show that

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{E}) &= \nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} , \\ \nabla \times (\nabla \times \mathbf{B}) &= \nabla (\nabla \cdot \mathbf{B}) - \nabla^2 \mathbf{B} . \end{aligned} \tag{2.3}$$

Look next at the right-hand sides:

$$\nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t} \right) = -\frac{\partial}{\partial t} (\nabla \times \mathbf{B}) = -\mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} , \tag{2.4}$$

$$\nabla \times \left(\mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) = \mu_0 \epsilon_0 \frac{\partial}{\partial t} (\nabla \times \mathbf{E}) = -\mu_0 \epsilon_0 \frac{\partial^2 \mathbf{B}}{\partial t^2} . \tag{2.5}$$

Putting the right-hand and left-hand sides together, using $\nabla \cdot \mathbf{B} = 0$ and using $\nabla \cdot \mathbf{E} = 0$ when $\rho = 0$, we see that $\mathbf{E}$ and $\mathbf{B}$ each obey wave equations:

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} - \frac{1}{\mu_0 \epsilon_0} \nabla^2 \mathbf{E} = 0 , \tag{2.6}$$

$$\frac{\partial^2 \mathbf{B}}{\partial t^2} - \frac{1}{\mu_0 \epsilon_0} \nabla^2 \mathbf{B} = 0 . \tag{2.7}$$

Further, we see that the parameter w which characterizes the speed of the wave is given by $1/\sqrt{\mu_0 \epsilon_0}$. This speed is given the label c (which comes from the word *celeritas*, meaning swiftness), and takes the value

$$\begin{aligned} c &= 2.99792458 \times 10^8 \,\text{meters/second} \\ &\simeq 3 \times 10^8 \,\text{meters/second} \\ &\simeq 1 \,\text{foot/nanosecond} . \end{aligned} \tag{2.8}$$

The equality on the first line is *exact*. A we'll discuss briefly a bit later in the course, we now actually use this value to define the meter. The near equality on the second line is good enough for most of the calculations we do in this class. The final near equality is amusing for those of us educated in parts of the world that still use inches and feet as their common measurement unit, and can be surprisingly useful in a number of practical situations.

If we imagine that $\mathbf{E}$ and $\mathbf{B}$ only depend on t and x, then the wave equations reduce to

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} - \frac{1}{\mu_0 \epsilon_0} \frac{\partial^2 \mathbf{E}}{\partial x^2} = 0 , \tag{2.9}$$

$$\frac{\partial^2 \mathbf{B}}{\partial t^2} - \frac{1}{\mu_0 \epsilon_0} \frac{\partial^2 \mathbf{B}}{\partial x^2} = 0 , \tag{2.10}$$

which have solutions of the form $\mathbf{E}(x \pm ct)$, $\mathbf{B}(x \pm ct)$.

When an analysis of this form was first done in the late 19th century, it was regarded as something of a triumph. In particular, the fact that the equations predicted $c = 1/\sqrt{\mu_0 \epsilon_0}$ was somewhat stunning. Bear in mind that ϵ_0 was an empirically measured parameter that played a role in determining the capacitance of a conductive system; μ_0 was a similar parameter that played a role in determining a system's inductance. The fact that parameters that were determined from static or very slowly varying fields could be so intimately related to the speed of light (whose value had been known to fairly good accuracy for quite some time, and was certainly known to be incredibly fast) was regarded as amazing. This association cemented the connection between light and electromagnetic fields.

Like most wave equation analyses, this calculation picked out a special speed. But, in what frame did we do this analysis?

2.3 The Michelson-Morley experiment

The consensus of the late 19th century was that electromagnetic waves are a disturbance in the so-called "luminiferous ether" (aka the "ether"), and that $c = 1/\sqrt{\mu_0\epsilon_0}$ is the speed of propagation with respect to that ether. In this framework, the ether defines a preferred rest frame, and the speed of light should only be c if we are in that preferred rest frame.

This line of reasoning tells us that if we measure light propagating across a laboratory that moves relative to the ether, then we should find that it moves with a speed that is not c. Our labs are on the surface of the Earth; the Earth spins on its axis, and orbits the Sun. Even if our lab is at rest with respect to the ether at some moment, it will no longer be at rest later in the day, or later in the year.

Albert Michelson and Edward Morley carried out an ingenious experiment in 1887 to test this hypothesis. Their idea was to use the wave nature of light to build an *interferometer*. The basic experimental setup is sketched in Fig. 1.

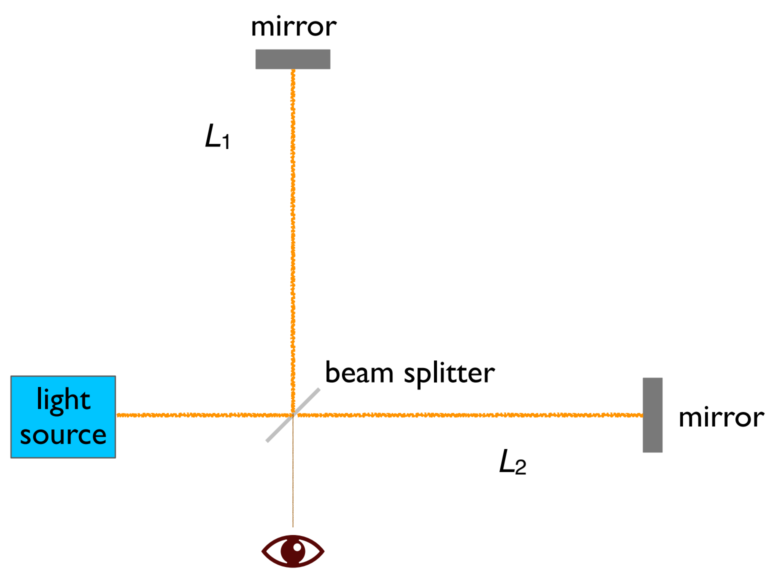


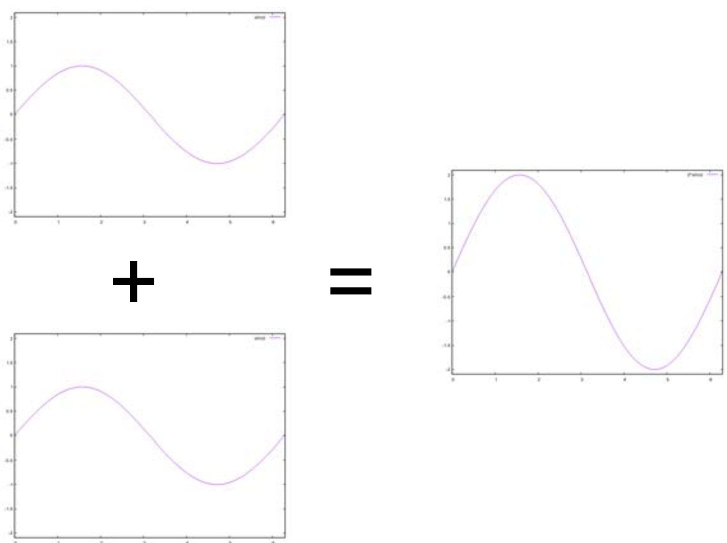
Figure 1: Basic layout of the interferometer used in the Michelson-Morley experiment. A beam of light enters from the source at left and is split by the beam splitter (a piece of partially silvered glass which reflects half of the light up, and allows half to transmit to the right). Both of these beams are reflected at the mirrors at the ends of the arms, return to the beam splitter, and then recombine. Exactly what happens when they recombine depends on the *optical phase difference* they experience along their two travel paths.

What happens when the light returns to the beam splitter? The answer depends on the details of the paths that the light takes in the two arms. To analyze this, let's make some definitions:

- Let t_1 be the travel time for light to go from beam splitter to mirror to beam splitter in arm 1 (of length L_1)
- Let t_2 be the travel time in arm 2
- Define $\Delta t \equiv t_2 - t_1$.

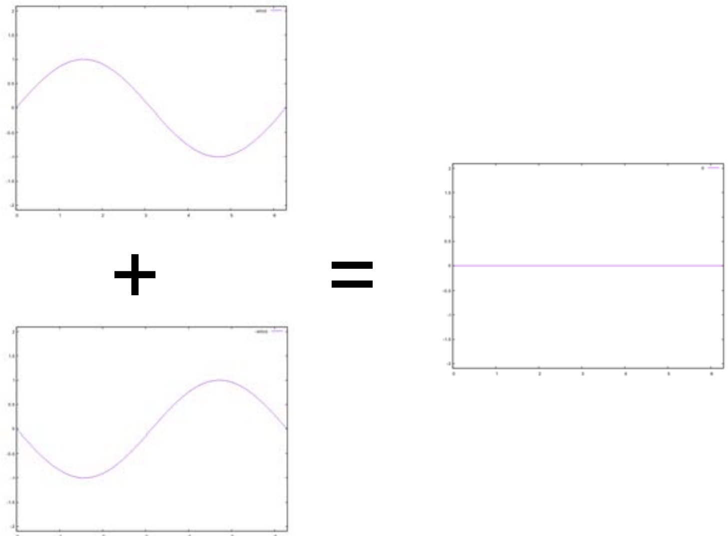
The quantity $c\,\Delta t$ is known as the *optical path difference*; it measures the difference in distance traversed by light as it back and forth through the two arms. Dividing this by λ, the wavelength of the light, yields[1] the *optical phase difference*.

Because light is a wave, the optical phase difference is an extremely important quantity for understanding what happens when light recombines at the beam splitter. If $c\,\Delta t/\lambda = 0, \pm 1, \pm 2, \ldots$, then the light *constructively* interferes: Peaks and troughs in the wave from one arm line up with peaks and troughs in the wave from the other:



[1]Strictly speaking, the optical phase difference is 2π times this quantity.

On the other hand, if $c\,\Delta t/\lambda = \pm\frac{1}{2}, \pm\frac{3}{2}, \pm\frac{5}{2}, \ldots$ then the light *destructively* interferes: Peaks in the wave from one arm line up with troughs in the other:



In general, we expect $c\,\Delta t/\lambda$ to be some value between an integer multiple of 1 and an (odd) integer multiple of $\frac{1}{2}$; the recombined amplitude will be some value between the peak of perfectly constructive interference, and the zero of perfectly destructive interference. In addition, Michelson and Morley used white light as their light source. This means that their measurement using a wide range of wavelengths. As such, we expect the light read out of the beam splitter (where the eye is placed in Fig. 1) to show an *interference fringe pattern*, with constructive interference for some wavelengths, destructive interference for others, and many values in between.

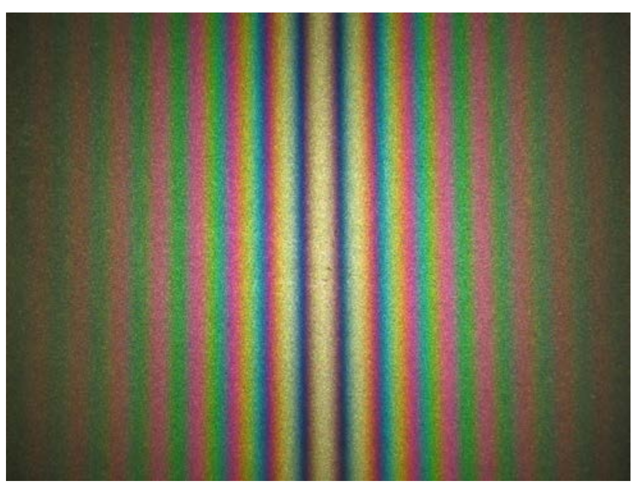


Figure 2: Example of a fringe pattern from readout of a Michelson-type interferometer. Source: `https://commons.wikimedia.org/wiki/File:MichelsonCoinAirLumiereBlanche.JPG`

With this background in mind, let us compute what optical phase difference we expect if c is the speed of light with respect to the ether, and if the Michelson-Morley apparatus moves with speed v with respect to the ether. More specifically, let's imagine that the lab's velocity $\mathbf{v}$ is parallel to arm 1. The time it takes for light to travel up arm 1 and then back to the beamsplitter is

$$t_1 = \frac{L_1}{c-v} + \frac{L_1}{c+v} = \frac{2L_1}{c}\left(\frac{1}{1-v^2/c^2}\right) . \tag{2.11}$$

The asymmetry between the two terms is because of the asymmetry in the light's motion relative to the apparatus along the two legs: in the first term, the mirror is "running away" from the light, so relative to the apparatus the light's speed is $c-v$; in the second term, the light switches direction, and the beam splitter is now "running toward" the light, with a relative speed $c+v$.

Arm 2 is a bit more complicated to analyze, as the light is in this case moving perpendicular to the motion of the apparatus in the rest frame of the ether. Figure 3 lays out the geometry:

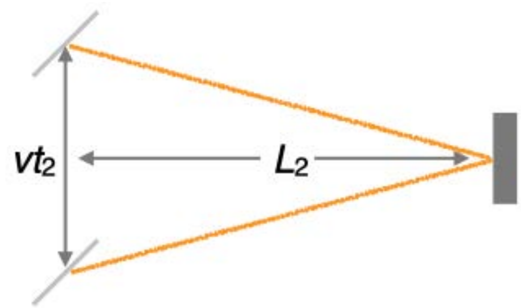


Figure 3: Light travel in arm 2 of the Michelson-Morley apparatus, as viewed in the rest frame of the ether. The light starts at the lower left, travels to the mirror, bounces, and returns to the upper left. During that time, the beam splitter moves from the position in the lower left to the position in the upper left. The light takes a total time t_2 to travel from the beam splitter to the mirror and back to the beam splitter. In that time, it covers a horizontal displacement of L_2 twice, and moves through a vertical displacement vt_2.

We defined t_2 as the time it takes for light to travel from the beam splitter to the mirror and back. As shown in the figure, in the rest frame of the ether the light moves on a diagonal path with horizontal displacement L_2 twice, and with vertical displacement vt_2. The equation governing t_2 is thus given by

$$ct_2 = 2\sqrt{L_2^2 + \left(\frac{vt_2}{2}\right)^2} , \tag{2.12}$$

from which we find

$$t_2 = \frac{2L_2}{c}\frac{1}{\sqrt{1-v^2/c^2}} . \tag{2.13}$$

Combining this with our result for t_1 yields

$$\Delta t = \frac{2}{c}\left[\frac{L_2}{\sqrt{1-v^2/c^2}} - \frac{L_1}{1-v^2/c^2}\right] . \tag{2.14}$$

At this point it is a useful to examine some numbers, in particular what we expect for v/c. The speed of the lab with respect to the ether is roughly bounded by the orbital speed of the Earth about the Sun, so $v \lesssim 2 \times 10^4$ meters/second. The speed of light is 3×10^8 meters/second, so

$$\frac{v}{c} \lesssim 10^{-4}\ . \tag{2.15}$$

This is a small quantity; the expression we've derived for Δt depends on this ratio squared, and even smaller quantity. Examining how this very small quantity enters our analysis, we see it is appropriate to use the binomial expansion, $(1 + \alpha x)^n \simeq 1 + n\alpha x$ for $x \ll 1$, to simplify what we have:

$$\Delta t \simeq \frac{2}{c}\left[L_2\left(1 + \frac{1}{2}\frac{v^2}{c^2}\right) - L_1\left(1 + \frac{v^2}{c^2}\right)\right]\ . \tag{2.16}$$

Michelson and Morley introduced one more very important factor into their experiment: They made it possible to rotate the interferometer's arms, effectively exchanging arms 1 and 2. (They did this by floating their entire optical table, which was built on a very heavy block of sandstone, on a pool of mercury. This both allowed the apparatus to rotate with very little friction, and provided significant isolation from vibrations in the building in which they did the experiment.) Rotating the apparatus, we get new light travel times:

$$t'_1 = \frac{2L_1}{c}\frac{1}{\sqrt{1 - v^2/c^2}} \simeq \frac{2L_1}{c}\left(1 + \frac{1}{2}\frac{v^2}{c^2}\right)\ , \tag{2.17}$$

$$t'_2 = \frac{2L_2}{c}\frac{1}{1 - v^2/c^2} \simeq \frac{2L_2}{c}\left(1 + \frac{v^2}{c^2}\right)\ , \tag{2.18}$$

$$\Delta t' = t'_2 - t'_1 = \frac{2}{c}\left[L_2\left(1 + \frac{v^2}{c^2}\right) - L_1\left(1 + \frac{1}{2}\frac{v^2}{c^2}\right)\right]\ . \tag{2.19}$$

Let us define δt as the change in the difference of light travel times between the two configurations:

$$\begin{aligned} \delta t &\equiv \Delta t' - \Delta t \\ &= \frac{2}{c}\left[L_2 + \frac{L_2 v^2}{c^2} - L_1 - \frac{L_1 v^2}{2c^2} - L_2 - \frac{L_2 v^2}{2c^2} + L_1 + \frac{L_1 v^2}{c^2}\right] \\ &= \frac{L_1 + L_2}{c}\left(\frac{v^2}{c^2}\right)\ . \end{aligned} \tag{2.20}$$

Their ingenious trick of rotating the interferometer means that the quantity they measured only depends on the sum $L_1 + L_2$, rather then depending sensitively on the individual arm lengths L_1 and L_2.

With all this laid out, let's review how this experiment works:

1. We begin with the experiment in a particular configuration. The experimenter monitors light that recombines at the beam splitter (indicated by the eyeball in Fig. 1), seeing a fringe pattern much like that shown in Fig. 2.

2. If there exists an ether and the laboratory is moving with respect to this ether, then light traveling in the two arms experiences the optical path difference $c\,\Delta t$. The initial fringe pattern the experimenter measures corresponds to the optical path difference associated with this initial configuration.

3. The entire interferometer is rotated by 90°. The experimenter (rotating along with it!) monitors the fringe pattern during the rotation. The expectation is that the optical path difference will change by $c\,\delta t$ during this rotation. This will be visible to the experimenter by a shifting of the fringe pattern as the apparatus is rotated.

One of the beautiful features of an interferometry experiment is that a shift of fringe can be measured very precisely. Carefully calibrating the positions of the mirrors, Michelson and Morley were confident that they could measure an optical phase shift $c\,\delta t/\lambda \approx 0.01$. This would have been plenty to detect the effect of motion with respect to the ether, as can be seen by plugging in some numbers for the experiment:

- Size of the apparatus: $L_1 + L_2 \simeq 10$ meters
- Speed with respect to the ether: $(v/c)^2 \simeq 10^{-8}$
- Wavelength of light: $\lambda \simeq 500\,\text{nm} = 5 \times 10^{-7}$ meters

The expected optical phase shift due to motion with respect to the ether is thus

$$\frac{c\,\delta t}{\lambda} \simeq 0.2\;. \tag{2.21}$$

This is a factor of 20 larger than what Michelson and Morley could discern, which is *huge*.

The value they in fact measured was zero. Since their pioneering experiment in 1887, measurements of this kind have been repeated. Measurement technology has improved to the point that we can now measure $c\,\delta t/\lambda \simeq 10^{-10}$. **No motion of an apparatus relative to an ether has ever been detected.**

2.4 Explanations

In 1887, the Michelson-Morley null result was a surprise. Both Michelson and Morley in fact considered it to be "failed experiment," and moved on to other things. However, it became clear over the years that the measurement had been done correctly, and the lack of phase shift was not experimenter error. This result begged explanation. Over time, four possible explanations emerged:

1. The ether is dragged along by the Earth, somehow, so that our labs are always locally at rest with it.

   This hypothesis in fact was the consensus view of how things would work at the time of the Michelson-Morley measurement. Part of what was so confusing about their result was that it contradicted other experiments at the time which preferred the ether to be "partially" dragged by the Earth; Michelson and Morley's result implied that any ether must be completely dragged along, so that the lab is always at rest with respect to the ether. Folding in more modern measurements, the ether drag hypothesis does not hold up when we make measurements on very long baselines (e.g., into space) where the effect of Earth's ether dragging should be greatly reduced or negligible.

2. Maxwell's equations are wrong.

   This hypothesis simply does not work: no wrongness has ever been found which can explain the Michelson-Morley measurements. Electrodynamic effects can be measured with exquisite precision, and Maxwell's equations work tremendously well.

3. The ether squashes moving objects just enough to compensate for the travel time shifts.

   This explanation "works" in the sense that one can design a squashing that fits the data, but raises a new question — how and why do such "length squashings" occur?

4. *There is no ether; there is no special rest frame for Maxwell's equations.* **Light travels at $c = 1/\sqrt{\mu_0\epsilon_0}$ in ALL inertial reference frames.**

The 4th option is where Einstein chose to begin his analysis. After all, no such frame is called out when Maxwell's equations are written down, so on what grounds should we imagine that this frame exists? This is where we will focus our studies.

2.5 Historical note

It should be noted that the historical record is somewhat unclear regarding the extent to which Einstein was influenced by Michelson and Morley. Some of his statements and writings suggest he was not influenced by their result, though other statements indicate that he was aware of the result and that it had some influence. It is clear, though, that he was aware of similar experiments (particularly those of Fizeau, whose experiment you will explore on problem set #1). It is fair to say that Einstein was aware the ether hypothesis was having trouble finding experimental support.

Einstein's historical motivations aside, with the benefit of over 130 years of hindsight, the importance of Michelson and Morley (and of similar experiments done since then) is clear to *us*: these measurements clearly demonstrate that the simple picture of Maxwell's equations being formulated in the "rest frame of the ether" (whatever that ether might actually be) cannot be correct. Einstein's choice of option #4 on the list above appears to be driven largely by simplicity: there is no ether and no special rest frame referred to anywhere in our formulation of the Maxwell equations, so why would we introduce them? Why not take at face value the fact that c emerges as the speed of light with no reference to a particular rest frame, and see what that implies?

Seeing what this implies will be our focus for the next several weeks.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 3
From Galilean relativity to ... ?

3.1 Speed of light is c to all observers

Much of what we will study over the next several weeks boils down to a detailed examination of the consequences of Einstein's hypothesis that all observers measure the speed of light to be c. The speed of light is thus an *invariant* — it is the same for *all* observers, in *all* frames of reference. As you will hopefully come to appreciate over the course of this semester, invariants are incredibly useful: we can exploit the fact that they are the same for all observers to facilitate many of the analyses we will want to perform.

The invariance of the speed of light tells us that the distance light travels per unit time is the same to all observers. In the Galilean transformation, we saw that displacement, and thus distance between events, varies depending on frame. As a consequence speed (distance per unit time) must vary as well. The Galilean transformation is thus inconsistent with the idea that the speed of light is the same to all observers: it must be corrected. If displacement varies according to the frame of an observer, but something's speed is invariant, we must find that *time intervals* vary by frame. Allowing the time interval to vary by frame is the only way that speed (displacement interval per unit time interval) can be invariant.

It's worth keeping in mind, however, that the Galilean transformation works very well in many circumstances, so it is *approximately* correct. Our "generalized" transformation law must be consistent with Galileo in some appropriate limit.

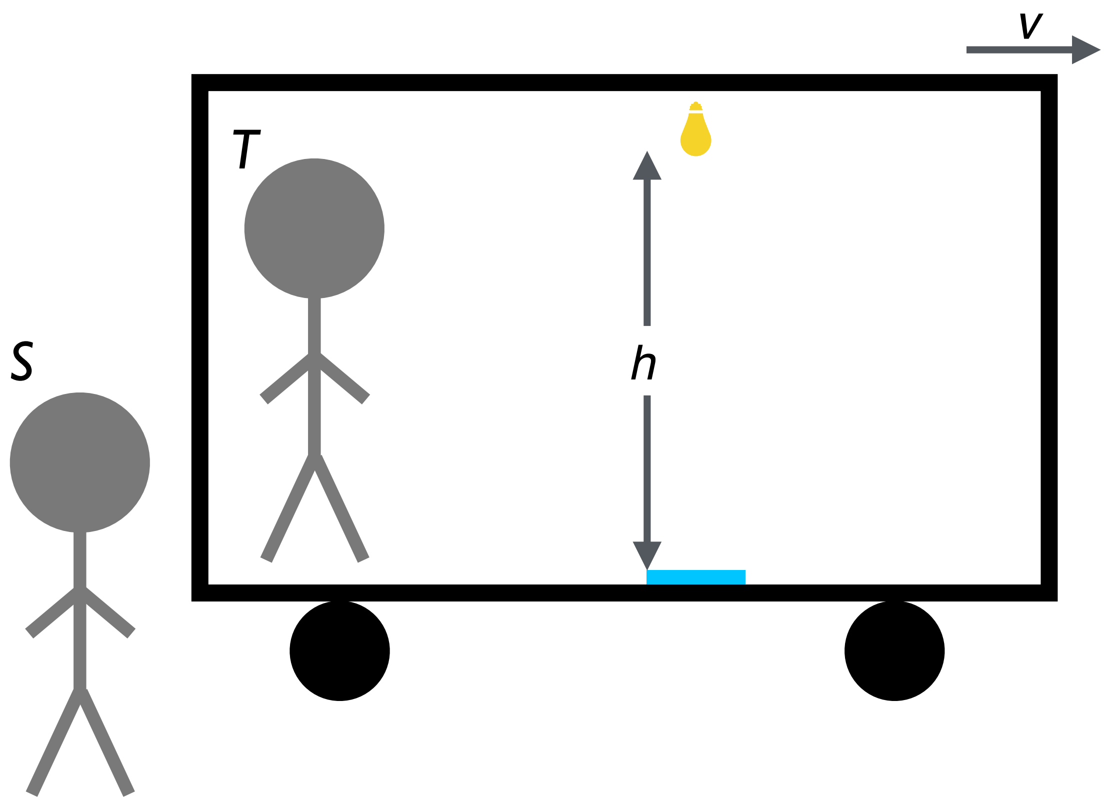
⌈Aside: The invariance of the speed of light also means that it is a great thing on which to base a metrology standard. That's why we take c to be *exactly* 2.99792458×10^8 meters per second. We then determine the meter to be the distance light travels in $1/(2.99792458 \times 10^8)$ seconds. Techniques in atomic physics have taught us how to measure time intervals very precisely, so this is a way of getting the meter out that capitalizes on what we measure best.⌋

3.2 Consequences I

Before generalizing the Galilean transformation, let's work through a few "thought experiments" which illustrate some of the consequences of light speed's invariance. We will consider two observers: Observer S is standing in a station; observer T is standing in a train that is moving with speed v through the station. These two observers each make measurements whose values we will compare.

First, imagine there is a light bulb inside the train. This bulb emits a pulse of light at some moment; we call this event A. The pulse of light propagates downward through the train, striking a photodetector on the floor, which records the moment the light strikes. We call this event B. Events A and B are geometric objects; all observers agree on the existence of these two things happening, though they may label the coordinates in time and space of these events differently.

We being our analysis by asking: What interval of time do observers T and S measure between events A and B?



Let's do this first in observer T's frame of reference. Observer T sees the light move through a vertical displacement h, so they deduce

$$\Delta t_T = h/c \ . \tag{3.1}$$

Observers in the station agree that the light moves through a vertical distance h, but also see it move through a horizontal distance that depends on the train's speed:

$$\Delta t_S = D/c = \frac{\sqrt{h^2 + (v\Delta t_S)^2}}{c} \ , \tag{3.2}$$

from which we find

$$\Delta t_S = \frac{h/c}{\sqrt{1 - v^2/c^2}} \equiv \gamma \Delta t_T \ , \tag{3.3}$$

$$\text{where} \qquad \gamma = \frac{1}{\sqrt{1 - v^2/c^2}} = \left(1 - \beta^2\right)^{-1/2} \ . \tag{3.4}$$

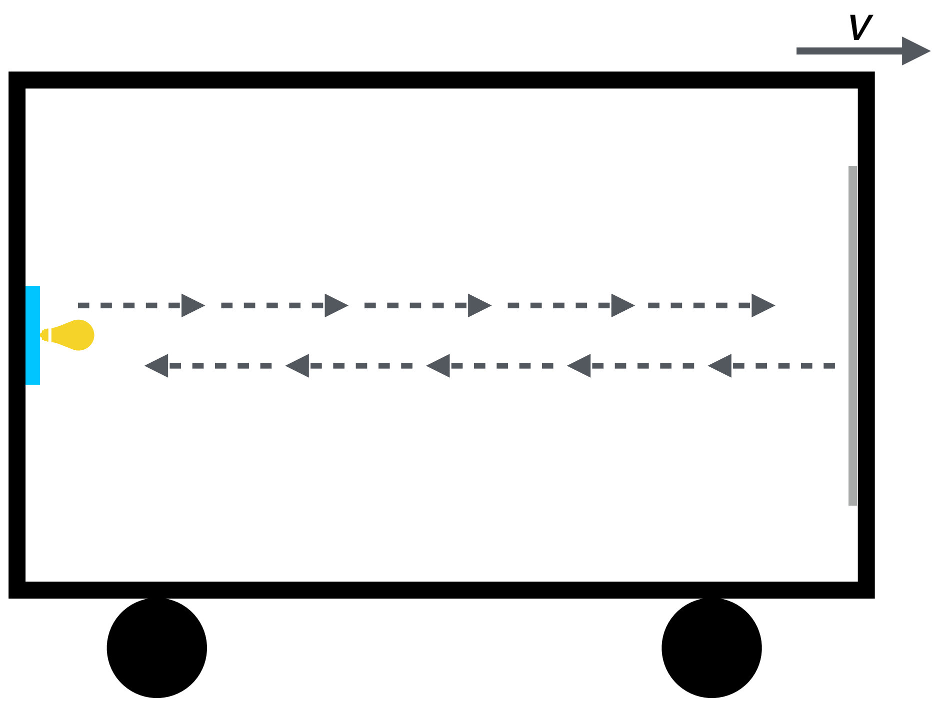
Notice that the factor $\gamma \geq 1$. The interval of time as measured on the station is *longer* than the interval measured on the train. For example, if the train moves at $v = \sqrt{3}c/2$ and the observers on the train measure 7 nanoseconds for the light to reach the photodetector, then observers in the station measure 14 nanoseconds for the light to reach the photodetector. Less time accumulates between the two events according to train observers than accumulates according to station observers.

Moving clocks run slow. This is a phenomenon known as *time dilation.*

⌈ASIDE: In doing this analysis, we've assumed that both the train and the station observers measure the same height h for the light's vertical displacement. Hold that thought!⌋

3.3 Consequences II

Let's next imagine that we arrange the light pulse so that it travels to the front of the car, bounces off a mirror, and returns to a photosensor[1]:



Both the train and station observers measure the time interval between the flash and the light striking the photodetector, and use this to infer the length of the train car. On the train (neglecting the size of the light bulb and the finite thickness of the sensor and mirror which are features of the sketch), the observer measures a time interval of Δt_T between the

[1]A common question asked about this set up is "Why the bounce? Why not have the photosensor on the front of the train so that the light only travels one way?" The reason we include the bounce is that for this first examination of light travel phenomena, it is very convenient for the net displacement of the light pulse to be zero along the direction of the train's travel in frame T. We will develop tools to handle more general situations very soon. Doing so, we'll see that having the light begin and end at the same coordinate in frame T simplifies the analysis in a way that is very useful for exploring basic concepts. (Notice that the net displacement along the direction of travel was zero in the previous example too.)

flash and the light striking the photodetector, and deduces that the train has a length

$$\Delta x_T = c\Delta t_T/2\ . \tag{3.5}$$

The size measured by observers on the train is just the time it takes light to travel from one end of the train to the other and back, divided by two.

To compute the size measured by observers in the station, let's break the calculation into two pieces: one piece gives us the time to travel from the bulb flash to reach the mirror; the other times travel from the mirror back to the photodetector. The interval of time measured for these two legs is

$$\Delta t_{S,1} = \frac{(\Delta x_S + v\Delta t_{S,1})}{c} \quad \longrightarrow \quad \Delta t_{S,1} = \frac{\Delta x_S}{c-v}\ ; \tag{3.6}$$

$$\Delta t_{S,2} = \frac{(\Delta x_S - v\Delta t_{S,2})}{c} \quad \longrightarrow \quad \Delta t_{S,2} = \frac{\Delta x_S}{c+v}\ . \tag{3.7}$$

Notice the asymmetry in the two contributions: in the first interval, the light travels the length of the train Δx_S plus the additional distance the train moves during this time interval; in the second interval, the light again travels the length Δx_S, but now *minus* the additional distance the train moves. The flash of light "chases" the mirror during interval 1, but is heading toward the advancing photodetector during interval 2. We add these two contributions to get the total travel time:

$$\begin{aligned}\Delta t_S = \Delta t_{S,1} + \Delta t_{S,2} &= \frac{2\Delta x_S/c}{1-v^2/c^2} \\ &= 2\gamma^2\Delta x_S/c\ .\end{aligned} \tag{3.8}$$

We have now related Δt_S to Δx_S, and Δt_T to Δx_T. What we really want is a relation between Δx_S and Δx_T. To cut through the different relations, let's take advantage of our previous result that the moving clock runs slow, i.e. that $\Delta t_S = \gamma\Delta t_T$. Using this, we can rewrite Eq. (3.8) as

$$\gamma\Delta t_T = 2\gamma^2\Delta x_S/c\ . \tag{3.9}$$

But we know that Δt_T and Δx_T are related by Eq. (3.5). Using this in Eq. (3.9) yields

$$\Delta x_S = \Delta x_T/\gamma\ . \tag{3.10}$$

This at last relates the spatial distance measured by the train observer to that measured by the station observer. Note that since $\gamma \geq 1$, Eq. (3.10) means the distance interval measured in the station is *shorter* than the distance interval measured on the train.

Moving rulers are shortened along the direction of motion. This is a phenomenon known as *length contraction*.

3.4 Consequences III

Are moving rulers affected along axes other than along the direction of motion? The answer is no: If they were, then we would get *inconsistent physics* — different events occurring in different frames of reference.

Imagine a train going at a speed $v = \sqrt{3}c/2$, so that $\gamma = 2$. Suppose the train is 5 meters tall, and is approaching a tunnel whose opening is 8 meters high. If length contraction affected the train's height, we'd have a serious problem:

- **Tunnel rest frame:** The train's height is contracted by a factor of γ, making it 2.5 meters tall — easily fitting into the 8 meter tunnel opening.

- **Train rest frame:** The tunnel's height is contracted by a factor of γ, making it 4 meters tall. The 5 meter train experiences a very high speed collision, destroying the train, the mountain into which the tunnel is carved, and very likely a good fraction of the surrounding countryside.

We require all observers to agree on events, even if they describe them using different labels. But these two outcomes — train merrily passing through a tunnel in one frame; chaos, death, destruction, and sadness in another — are not mere differences of label. These are completely inconsistent outcomes.

In order for events to be consistent between different reference frames, it must be the case that moving rulers are unaffected along directions orthogonal to their direction of motion. Post facto, this justifies our assumption that both the train observer and the station observer measure a vertical displacement of h, as we used in "Consequences I."

3.5 From Galileo to Lorentz

In the examples we've discussed above, we have allowed our notions of time and space intervals to get mixed up by our demand that all observers measure light to have a propagation speed of c. As we can see, this leads to some rather nonintuitive consequences. However, these consequences follow straightforwardly from our requirement that c be an invariant.

Let us now think about how to mix up different intervals in a more systematic manner. Galilean transformations allowed different inertial frames to define different standards for space: what's "left" to you is a mixture of "left" and "forward" to someone with a different orientation; what's "there" to you is "there and steadily moving farther away" to someone moving with a fixed speed. But time is the same for everyone.

Let's think about a category of transformations that can mix up space *and* time, doing so in such a way that the speed of light is left invariant. Let's think about a station observer who labels events with coordinates (t_S, x_S, y_S, z_S), and a train observer who labels events with coordinates (t_T, x_T, y_T, z_T). The station observer sees the train moving with $\mathbf{v} = v\mathbf{e}_x$.

We will begin by assuming that the train frame's coordinates are related to those of the station with the following linear relations:

$$t_T = At_S + Bx_S \tag{3.11}$$

$$x_T = Dt_S + Fx_S \tag{3.12}$$

$$y_T = y_S \tag{3.13}$$

$$z_T = z_S \tag{3.14}$$

This form was chosen[2] by noting that since we are moving along x, the coordinates y and z cannot be affected. We require it to be a linear transformation because non-linear terms (e.g., a t^2 term) would make the transformation non-inertial.

We now solve for A, B, D, F by matching important quantities in the two systems and imposing invariance of c. Our first two steps are familiar from the Galilean transformation — we simply require that constant x coordinates in one frame move with speed v in the other frame. Let us focus in particular on the spatial origin:

1. Match the spatial origin of the train frame, $x_T = 0$, with events in the station frame at $x_S = vt_S$:

$$
\begin{aligned}
x_T &= Dt_S + Fx_S \\
0 &= Dt_S + Fvt_S \\
\longrightarrow \qquad D &= -Fv \; .
\end{aligned}
\tag{3.15}
$$

   This tells us that our x transformation law can be written $x_T = F(x_S - vt_S)$.

2. Next, match the origin of the station frame ($x_S = 0$) to events in the train frame at $x_T = -vt_T$:

$$
\begin{aligned}
x_T &= F(x_S - vt_S) \\
-vt_T &= -Fvt_S
\end{aligned}
\tag{3.16}
$$

   This tells us that $t_T = Ft_S$ for events at $x_S = 0$. But we also know

$$
t_T = At_S + Bx_S \tag{3.17}
$$

   Plugging in $x_S = 0$ and $t_T = Ft_S$, we see that

$$
\longrightarrow \qquad F = A \; . \tag{3.18}
$$

   We have now pinned down 2 of the 4 unknown coefficients, and the transformation law for t and x reads

$$
t_T = At_S + Bx_S \tag{3.19}
$$

$$
\begin{aligned}
x_T &= -Avt_S + Ax_S \\
&= A(x_S - vt_S) \; .
\end{aligned}
\tag{3.20}
$$

   To pin down A and B, we use the physics that is the focus of this lecture: all observers agree that light propagates with speed c, so we examine the propagation of light as measured in the two reference frames.

[2] We do not use the letter C to avoid confusion with the speed of light. We also skip E to avoid confusion with energy, which we will be discussing soon.

3. Imagine a light pulse emitted at $t_S = t_T = 0$, and examine its propagation along the x_T and x_S axes. As seen in the station, it travels with $x_S = ct_S$; as seen on the train, it travels with $x_T = ct_T$:

$$\begin{aligned} x_T &= ct_T \\ A(x_S - vt_S) &= c(At_S + Bx_S) \qquad \text{(Substituting the transformation rules)} \\ A(ct_S - vt_S) &= c(At_S + Bct_S) \qquad \text{(Substituting } x_S = ct_S\text{)} \\ -Avt_S &= Bc^2 t_S \\ \longrightarrow \qquad B &= -\frac{Av}{c^2} \ . \end{aligned} \tag{3.21}$$

The transformation law now reads

$$t_T = A(t_S - vx_S/c^2) \tag{3.22}$$
$$x_T = A(x_S - vt_S) \ . \tag{3.23}$$

4. Now look at how that pulse travels in the y direction according to observers in the station. They see it moving with $x_S = 0$, $y_S = ct_S$. Observers on the train measure it moving diagonally, following a trajectory in x_T and y_T that satisfies

$$(x_T)^2 + (y_T)^2 = c^2(t_T)^2 \ . \tag{3.24}$$

Substitute $x_T = A(x_S - vt_S)$, $t_T = A(t_S - vx_S/c^2)$, $y_T = y_S = ct_S$, and finally plug in $x_S = 0$:

$$A^2v^2t_S^2 + c^2t_S^2 = c^2A^2t_S^2 \ . \tag{3.25}$$

This is easy to solve for A:

$$A = \frac{1}{\sqrt{1 - v^2/c^2}} = \gamma \ . \tag{3.26}$$

(If you're being really pedantic you might wonder why we don't consider the negative square root. Consider the $v = 0$ limit, for which the two coordinate systems should be identical; this shows that you need the positive root here.)

Our complete transformation law becomes

$$t_T = \gamma\left(t_S - x_S v/c^2\right) \tag{3.27}$$
$$x_T = \gamma\left(-vt_S + x_S\right) \tag{3.28}$$
$$y_T = y_S \tag{3.29}$$
$$z_T = z_S \ . \tag{3.30}$$

This result is called the *Lorentz transformation.*

A few comments: First, we can make it a bit more symmetric looking by using the definition $\beta = v/c$ we introduced earlier, and by writing ct_T and ct_S as our time variables. This gives our time coordinates the same dimensions (or units) as for space. With these minor tweaks, the Lorentz transformation can be written in the matrix form

$$\begin{pmatrix} ct_T \\ x_T \\ y_T \\ z_T \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct_S \\ x_S \\ y_S \\ z_S \end{pmatrix} \ . \tag{3.31}$$

Second, note that Nature doesn't care how we label the axes; we could very well have defined things moving in the y direction or the z direction, or some direction that is at an angle between those directions. If we had the train moving with $\mathbf{v} = v\mathbf{e}_y$, then we would have found

$$\begin{pmatrix} ct_T \\ x_T \\ y_T \\ z_T \end{pmatrix} = \begin{pmatrix} \gamma & 0 & -\gamma\beta & 0 \\ 0 & 1 & 0 & 0 \\ -\gamma\beta & 0 & \gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct_S \\ x_S \\ y_S \\ z_S \end{pmatrix} . \tag{3.32}$$

You can probably deduce how things look for $\mathbf{v} = v\mathbf{e}_z$.

Finally, how do we invert this transformation? The "brute force" approach is to compute the matrix inverse. However, a little physics helps us see the answer: If the station observer sees the train moving with $\mathbf{v} = v\mathbf{e}_x$, the train observer must see the station moving with $\mathbf{v} = -v\mathbf{e}_x$. They must develop exactly the same Lorentz transformation, but with the terms linear in v flipped in sign:

$$\begin{pmatrix} ct_S \\ x_S \\ y_S \\ z_S \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct_T \\ x_T \\ y_T \\ z_T \end{pmatrix} . \tag{3.33}$$

It's not hard to show that the matrix in Eq. (3.33) is the inverse of the matrix in Eq. (3.31).

3.6 A comment on the road ahead

Much of what we will do in the next few weeks essentially amounts to examining the consequences of the Lorentz transformation, assessing what aspects of physics as we know it hold up and what aspects will need modification. Many of our discussions will involve "thought experiments" of the kind we discuss in the "Consequences" sections above. As such, one can be misled into thinking that much of "Einsteinian" physics is about abstract weird situations like trains that move at nearly the speed of light.

I want to take this moment to make it clear that, though such discussions are useful for understanding important concepts, they are *not* what relativity is about. Like all physics, relativity is a framework *by which we understand the world as we actually measure it.* Special relativity in particular is one of the best-studied theories that we have; its consequences — including the physics of effects like time dilation — have been tested with exquisite accuracy. (Indeed, in a very real sense, *magnetism* is nothing more than a consequence of Coulomb's law of electrostatics plus the Lorentz transformation.) In recent years, the consequences of general relativity have been measured and tested quite thoroughly as well.

We study Einstein's relativity because empirical experience has pointed to the fact that it describes our world exquisitely well. Because you are studying physics, you are likely to encounter people who wish to sell you an alternative[3]. Many of them will claim that the only reason that Einstein gets the attention he is given is because physics has become effectively a priesthood. Some of these folks are bothered by the fact that many consequences of Einstein's

[3] I get at least 5 and as many as 30 emails a week in this theme; I occasionally get hand-written letters and self-published books. One guy sent me an adjustable wrench along with his book, I think because he claimed to be "throwing a monkey wrench" into all the "nonsense" that physics departments teach students. It's actually quite a nice wrench. I use it at least twice a year to hook up a hose at my house at the start of summer, and to disconnect it when the weather gets cold.

relativity go against “common sense”; a few (including some of the more frightening ones who write to me) claim darker motivations. We will endeavor as much as possible to bring the consequences of relativity into this class, and to keep it grounded in experimental fact. One thing should be clear: if measurements did not agree with Einstein’s theories of relativity, *we would have discarded these theories in a heartbeat.*

3.7 An aside on factors of c

The speed of light c pops up so much in this subject that it’s very convenient in many analyses to define your units such that $c = 1$. This means that if you measure time in seconds, your basic unit of length is the light second. Amusingly, this means that if you measure time in nanoseconds, your basic unit of length is the light nanosecond, which is almost exactly[4] one foot. With this choice made, the units of time and space are identical, the factor $\beta \equiv v$, and the Lorentz transformation takes the form

$$\begin{pmatrix} t_T \\ x_T \\ y_T \\ z_T \end{pmatrix} = \begin{pmatrix} \gamma & -v\gamma & 0 & 0 \\ -v\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t_S \\ x_S \\ y_S \\ z_S \end{pmatrix} . \tag{3.34}$$

In my research, I usually set $c = 1$. I’m of mixed mind whether I should use these units in 8.033. On one hand, it is a great convenience, and cuts down on a symbol that strictly speaking isn’t needed; and it is certainly a choice of units that you will see in future coursework. However, when studying relativity for the first time, it is worth bearing in mind that there are quite a few major points that can be confusing. As a point of pedagogy, I’d rather not introduce minor points that also cause confusion. I will endeavor to keep c explicitly in formulas that I write on the board, in the notes, and on assignments, but the likelihood that I will occasionally mess up is very high. If you think a factor of c has been left out, please ask about it.

When writing up your own assignments, if you’d like to use $c = 1$ units, feel free to do so, but please state that you have made this choice on your writeup.

[4] 1 light nanosecond = 29.9792458 cm = 11.8029 inches = 0.9836 feet.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 4
Spacetime, simultaneity, and the consequences of Lorentz

4.1 From space and time to spacetime

The Lorentz transformation shows us that the invariance of c requires space and time to be mixed together; what is "space" for one observer is a mixture of "space" and "time" for another. This should be familiar as far as spatial directions go — what is "left" for one observer can be a mix of "left" and "forward" for another — but mixing time and space like this surely feels somewhat odd. We can no longer think of space and time as separate things; we instead describe them as a new, unified entity: *spacetime*. Each inertial observer splits spacetime into space and time; however, *how* they split into space and time differs. This is fundamentally why different inertial observers measure different intervals of time and different intervals of distance.

One of the tools we will use to examine the geometry of spacetime is the *spacetime diagram*. This is a figure that illustrates how space and time are laid out, as seen by an observer in some particular inertial frame. The convention in making such figures is that the vertical axis denotes time, horizontal axes denote space.

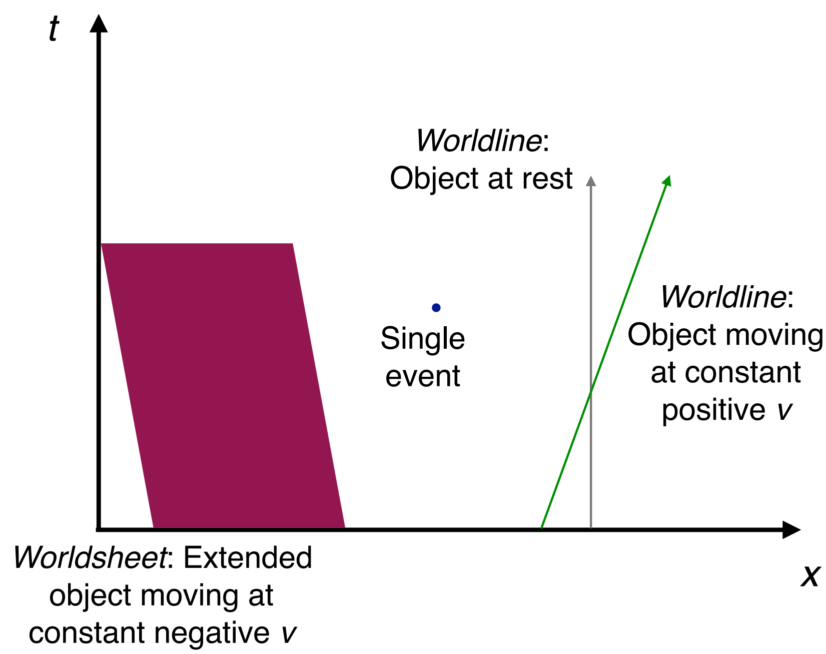
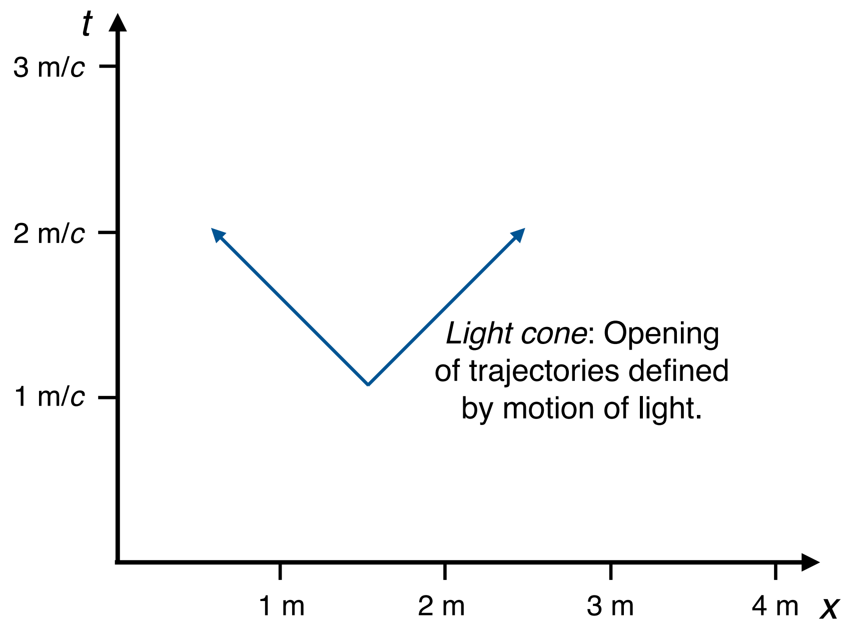


Figure 1: Example of a spacetime diagram. An event is a single point. A *worldline* is the sequence of events swept out by an event as it moves through space and time, with a slope that depends on its velocity in the frame. A *worldsheet* is the set of events swept out by an extended set of events as they move through space and time.

The units of a spacetime diagram's axes are usually chosen so that light moves on 45° lines with respect to the axes of the rest frame:



(This is particularly natural if we choose units such that $c = 1$.) With such units, a pulse of light, moving through time and projected onto 1 spatial dimension, makes a *lightcone* with an opening angle of 90°. As we will discuss shortly, the lightcone plays an important role in helping us to figure out how events are related to one another.

When making a spacetime diagram, one draws axes corresponding to some particular observer. Suppose we draw the axes of some observer $\mathcal{O}$ who uses coordinates (t, x). How do we represent the coordinates (t', x') of an observer $\mathcal{O}'$ who moves with $\mathbf{v} = v\mathbf{e}_x$ according to $\mathcal{O}$? In other words, what do the (t', x') axes look like as seen by $\mathcal{O}$?

To figure this out, let's look at the transformation rule:

$$ct' = \gamma(ct) - \beta\gamma x \tag{4.1}$$

$$x' = -\beta\gamma(ct) + \gamma x \tag{4.2}$$

The t' axis is defined as the set of events for which $x' = 0$:

$$0 = -\beta\gamma(ct) + \gamma x \qquad \longrightarrow \qquad t = \frac{x}{\beta c} = \frac{x}{v} \ . \tag{4.3}$$

The x' axis is defined by the events for which $t' = 0$:

$$0 = \gamma(ct) - \beta\gamma x \qquad \longrightarrow \qquad t = \frac{\beta x}{c} = \frac{vx}{c^2} \ . \tag{4.4}$$

Figure 2 illustrates the (t', x') axes as seen by $\mathcal{O}$ for an observer moving with $v = \sqrt{3}c/2$.

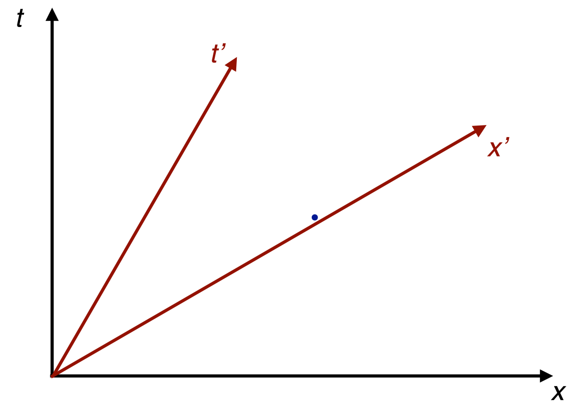


Figure 2: Axes of observer $\mathcal{O}'$ as they appear in the frame of $\mathcal{O}$. The dot represents a particular event.

In this figure, we show a particular event. This event is a geometric object, a single point in spacetime. Although both observers agree on where it is in spacetime, they assign it rather different space and time coordinates. (We will analyze the different labels observers attach to coordinates in some detail shortly.)

We could equally well ask how the axes (t, x) appear according to $\mathcal{O}'$ — we simply use the inverse transformation rule, which yields

$$t' = -\frac{x'}{v} \text{ for the } t \text{ axis}, \qquad t' = \frac{-vx'}{c^2} \text{ for the } x \text{ axis.} \tag{4.5}$$

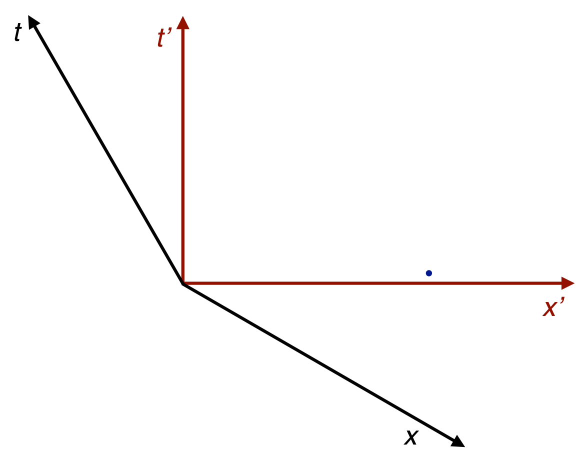


Figure 3: Axes of observer $\mathcal{O}$ as they appear in the frame of $\mathcal{O}'$.

4.2 Simultaneous for me, not necessarily for thee

Drawing transformed axes in this way illustrates why length contraction and time dilation arise: Events which are simultaneous — occurring at the same time — in one frame of reference are *not* simultaneous in another frame; events which occur in the same location in one frame do not occur in the same location in another frame. This is the essence of how "space" and "time" are mixed, but "spacetime" remains unified. Different observers agree on "spacetime," but they split it into "space" and "time" in different ways.

Let's illustrate this breakdown in simultaneity explicitly with a pair of spacetime diagrams. Figure 4 shows two events, A and B, which are simultaneous according to the observer $\mathcal{O}$ who uses coordinates (t, x): we have drawn several surfaces of constant t in this space, showing that these events are in a single such surface.

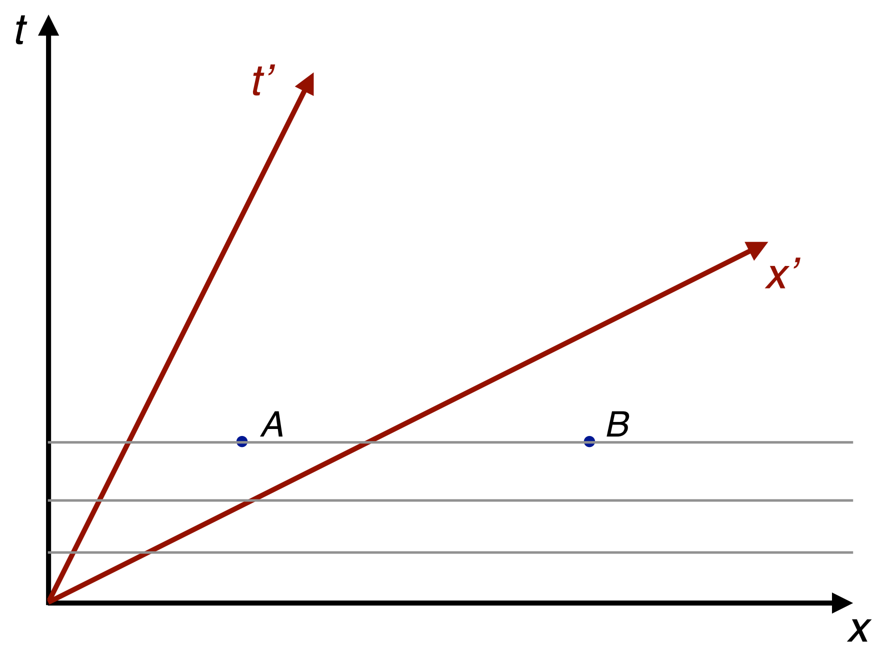


Figure 4: Events A and B occur at the same time t as measured by observer $\mathcal{O}$. As measured by observer $\mathcal{O}'$, event B occurs before event A.

Do these events occur at the same time in the frame of the observer $\mathcal{O}'$, who uses coordinates (t', x')? Definitely not! We can see that event B occurs before event A according to $\mathcal{O}'$.

What does a surface of constant time t' look like in the coordinates (t, x)? We can figure this out by using the Lorentz transformation: A surface of constant t' is the line in the (t, x) plane that corresponds to some value of t':

$$ct' = \gamma(ct) - \beta\gamma x \qquad \longrightarrow \qquad t = \frac{vx}{c^2} + \frac{t'}{\gamma} \,. \tag{4.6}$$

This is the same slope as the x' axis, so surfaces of constant t' appear as lines parallel to this axis. We show this in Fig. 5, making it clear that B comes earlier in time than A according to observer $\mathcal{O}'$.

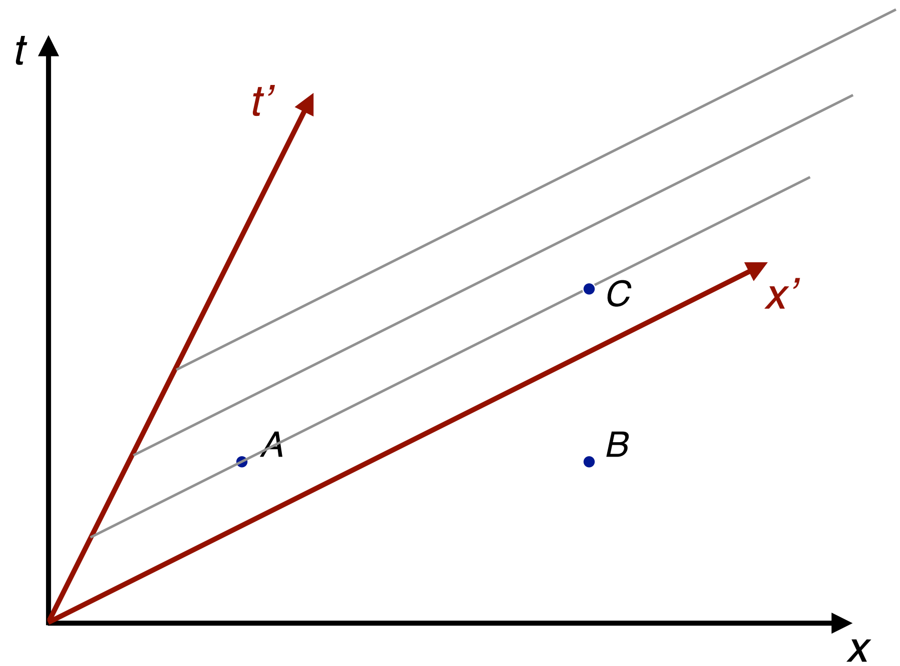


Figure 5: Events A and B occur at the same time t as measured by observer $\mathcal{O}$. As measured by observer $\mathcal{O}'$, event B occurs before event A.

We also show a third event, C, which is simultaneous with A according to $\mathcal{O}'$, but occurs later according to $\mathcal{O}$.

4.3 The invariant interval

Many of the so-called "mysteries" and "counter-intuitive" aspects of physics in special relativity have their origin in this discussion: two events which are simultaneous to one observer will not be simultaneous to all observers. This, plus the fact that two events which occur at the same location in space according to one observer are not co-located according to other observers, is the root of phenomena such as time dilation and length contraction.

Is there *anything* that holds consistently across frames? If there is, then it will define an *invariant*, some quantity whose value all observers agree upon. Indeed, we can assemble an invariant from the "spacetime separation" of two events. Consider events A and B. Compute their separation in time and space in some given frame:

$$\Delta t = t_B - t_A \ , \quad \Delta x = x_B - x_A \ , \quad \Delta y = y_B - y_A \ , \quad \Delta z = z_B - z_A \ . \tag{4.7}$$

From these quantities, compute

$$\Delta s^2 \equiv -c^2 \Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2 \ . \tag{4.8}$$

Theorem: All inertial observers, in all reference frames, agree on the value of Δs^2.

This theorem is easily proved by simply examining $(\Delta s')^2$, the invariant interval computed using the coordinate separation of the events as measured in some other frame:

$$\Delta t' = t'_B - t'_A \ , \quad \Delta x' = x'_B - x'_A \ , \quad \Delta y' = y'_B - y'_A \ , \quad \Delta z' = z'_B - z'_A \ . \tag{4.9}$$

Let us relate these "primed" separations to the "unprimed" ones using the Lorentz transformation along x we've been using:

$$c\Delta t' = \gamma(c\Delta t) - \gamma\beta\Delta x \ , \tag{4.10}$$
$$\Delta x' = -\gamma\beta(c\Delta t) + \gamma\Delta x \ , \tag{4.11}$$
$$\Delta y' = \Delta y \ , \tag{4.12}$$
$$\Delta z' = \Delta z \ . \tag{4.13}$$

Let us now compute $(\Delta s')^2$:

$$(\Delta s')^2 = -(c\Delta t')^2 + (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2 \tag{4.14}$$
$$= -\gamma^2(c\Delta t)^2 + 2\gamma^2\beta(\Delta x)(c\Delta t) - \gamma^2\beta^2(\Delta x)^2 + \gamma^2\beta^2(c\Delta t)^2 - 2\gamma^2\beta(c\Delta t)(\Delta x) + \gamma^2(\Delta x)^2 + \Delta y^2 + \Delta z^2 \tag{4.15}$$
$$= -c^2\Delta t^2\left[\gamma^2(1-\beta^2)\right] + \Delta x^2\left[\gamma^2(1-\beta^2)\right] + \Delta y^2 + \Delta z^2 \tag{4.16}$$
$$= -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2 \tag{4.17}$$
$$= \Delta s^2 \ . \tag{4.18}$$

The first line of this is just the definition of $(\Delta s')^2$. To go to the second line, we've used the Lorentz transformation to express the primed-frame quantities in terms of unprimed-frame quantities. To go to the third line, we gather terms together, canceling out the terms that involve $(\Delta x)(c\Delta t)$, and gathering common factors of Δx^2 and $c^2\Delta t^2$. To go to the fourth line, we used the fact that $\gamma = 1/\sqrt{1-\beta^2}$. That line reproduces Δs^2, demonstrating[1] that this quantity is a Lorentz invariant.

We are going to do a lot with Δs^2, a quantity that we call the *invariant interval* (often abbreviated to just the "interval"). To start, it's worth noting that perhaps the most important property of this quantity is whether it is negative, positive, or zero:

- $\Delta s^2 < 0$: in this case, the interval is dominated by Δt. We say that the two events have **timelike** separation. When $\Delta s^2 < 0$, it means that we can find some Lorentz frame in which the events A and B have the same spatial position (i.e., in that frame $x_A = x_B$, $y_A = y_B$, $z_A = z_B$); the events are only separated by time in that frame. We define $\Delta\tau \equiv \sqrt{-\Delta s^2}/c$ to be the time elapsed between events A and B in that frame. We call $\Delta\tau$ the **proper-time interval** — it is the interval of time measured by the observer who is at rest in the frame in which A and B are co-located.

  It's worth noting that if the interval between two events is timelike, then one can imagine a signal which travels with speed $v < c$ that connects them.

- $\Delta s^2 > 0$: the interval here is dominated by $\Delta x^2 + \Delta y^2 + \Delta z^2$, and we say that the two events have **spacelike** separation. In this case, we can find a Lorentz frame in which events A and B are simultaneous; Δs is the distance between these events in that frame. We call Δs the **proper separation** of A and B.

[1] It is easy to verify that this works for the transformation along any axis. In another lecture or two, we will introduce notation that makes proving the invariance of quantities like this really easy for *any* Lorentz transformation.

- $\Delta s^2 = 0$: in this case, we find that $c\Delta t = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$ — events A and B have a lightlike or "null" separation. If $\Delta s^2 = 0$, then these events can be connected by a light pulse.

The last point helps us to see that the value of Δs^2 is very closely connected to the properties of the lightcone mentioned earlier. Suppose a flash of light is emitted from event A. If the interval between A and another event is negative, $\Delta s^2 < 0$, then the other event must be *inside* the lightcone. If the interval is positive, then the event must be *outside* the lightcone. And if $\Delta s^2 = 0$, then the other event must be on the light cone itself. Figure 6 illustrates how these notions connect to the lightcone.

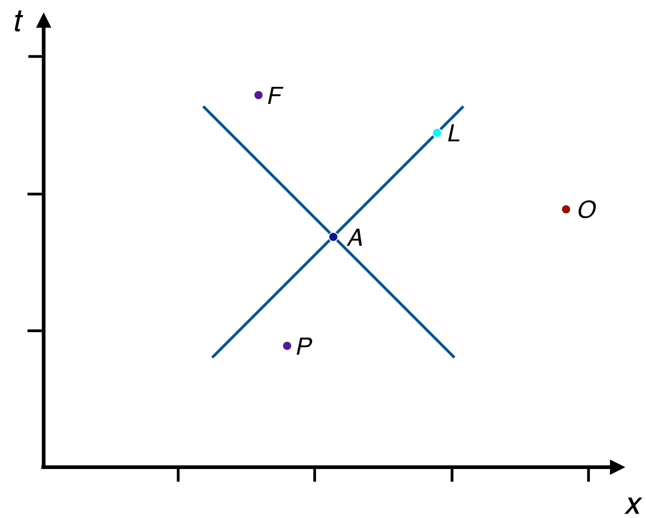


Figure 6: The intervals between events A and F and events A and P are timelike: $\Delta s^2_{AF} < 0$, $\Delta s^2_{AP} < 0$. In all frames, event F has time coordinate greater than the time coordinate of event A: $t_F > t_A$. Event F is unambiguously in the *future* of event A. Likewise, event P has time coordinate less than the time coordinate of event A: $t_P < t_A$ in all frames. Event P is unambiguously in the *past* of event A. Events A and O have a spacelike interval: $\Delta s^2_{AO} > 0$. Event O is neither in the future nor the past of A; it is "elsewhere," so the time-ordering of these events is not invariant. Events A and L have a lightlike or null interval: $\Delta s^2_{AL} = 0$. These events are connected by a light beam in all reference frames.

4.4 The geometry of spacetime

The relationship $\Delta s^2 = -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$ essentially expresses the Pythagorean theorem for spacetime. For intuition, consider the Pythagorean theorem purely in space. On a flat two-dimensional surface, a right triangle whose sides are Δx and Δy has a hypotenuse whose length is determined from $\Delta s^2 = \Delta x^2 + \Delta y^2$. In three dimensions, the distance from (x, y, z) to $(x + \Delta x, y + \Delta y, z + \Delta z)$ is given by $\Delta s^2 = \Delta x^2 + \Delta y^2 + \Delta z^2$.

In spacetime, it turns out to be extremely useful to regard $\Delta s^2 = -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$ as expressing an invariant notion of "distance squared" between two events. Students usually want to know "Why does the $c^2\Delta t^2$ have a minus sign?" The best answer I can give is that this is how the geometry of the universe works. The fact that time enters Δs^2 with a different sign from space reflects the fact that time is fundamentally quite different from the other directions of spacetime. We can forward and backward; we can move left and right; we can move up and down. But we can only move toward the future — we cannot step back to the past.

Indeed, the whole notion of "past" and "future" depends on events' separation in spacetime. If two events are timelike or lightlike separated, then one can describe one event as being the future, and one in the past. Although the specific time coordinates assigned to these events will vary by reference frame, the time *ordering* of these events is invariant: if $t_F > t_A$ in one frame, and if the interval between events A and F is timelike or lightlike, then $t_F > t_A$ in *all* reference frames. However, if two events are spacelike separated, then their time ordering depends on reference frame. Consider the situation shown in Figure 7:

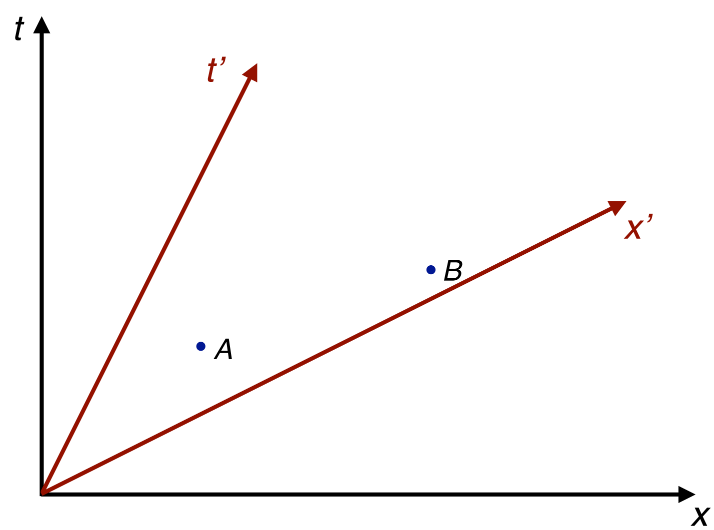


Figure 7: Observer $\mathcal{O}$ measures coordinates for events A and B using the (t, x) axes. Observer $\mathcal{O}'$, who travels with velocity $\mathbf{v} = (c/2)\mathbf{e}_x$ according to $\mathcal{O}$, measures coordinates for these events using the (t', x') axes.

Suppose observer $\mathcal{O}$ measures these events at the coordinates $(t_A, x_A) = (2\,\text{sec}, 2\,\text{lightsec})$, $(t_B, x_A) = (3\,\text{sec}, 5\,\text{lightsec})$. So, for observer $\mathcal{O}$, event A happens first. However, the invariant interval between these events,

$$\Delta s^2 = -c^2\Delta t^2 + \Delta x^2 = -(1\,\text{lightsec})^2 + (3\,\text{lightsec})^2 = 8\,\text{lightsec}^2\ , \tag{4.19}$$

is *positive* — these events are spacelike separated, so different observers may very well order them differently.

Let's use the Lorentz transformation to compute the events' coordinates according to $\mathcal{O}'$. Given the relative speed $c/2$, we have $\gamma = 2/\sqrt{3}$, $\beta = 1/2$. Applying the Lorentz transformation, we find

$$ct'_A = \gamma t_A - \beta\gamma x_A = \left(4/\sqrt{3} - 2/\sqrt{3}\right) \text{ lightsec} = \frac{2}{\sqrt{3}} \text{ lightsec} , \tag{4.20}$$

$$x'_A = -\beta\gamma t_A + \gamma x_A = \left(-2/\sqrt{3} + 2/\sqrt{3}\right) \text{ lightsec} = \frac{2}{\sqrt{3}} \text{ lightsec} ; \tag{4.21}$$

$$ct'_B = \gamma t_B - \beta\gamma x_B = \left(6/\sqrt{3} - 5/\sqrt{3}\right) \text{ lightsec} = \frac{1}{\sqrt{3}} \text{ lightsec} , \tag{4.22}$$

$$x'_B = -\beta\gamma t_B + \gamma x_B = \left(-3/\sqrt{3} + 10/\sqrt{3}\right) \text{ lightsec} = \frac{7}{\sqrt{3}} \text{ lightsec} . \tag{4.23}$$

$$\begin{aligned} \longrightarrow \quad (t'_A, x'_A) &= \left(\frac{2}{\sqrt{3}} \text{ sec}, \frac{2}{\sqrt{3}} \text{ lightsec}\right) \\ &\simeq (1.15\,\text{sec}, 1.15\,\text{lightsec}) \end{aligned} \tag{4.24}$$

$$\begin{aligned} \longrightarrow \quad (t'_B, x'_B) &= \left(\frac{1}{\sqrt{3}} \text{ sec}, \frac{7}{\sqrt{3}} \text{ lightsec}\right) \\ &\simeq (0.577\,\text{sec}, 4.04\,\text{lightsec}) . \end{aligned} \tag{4.25}$$

Notice that $t'_A > t'_B$: the order of the events is reversed according to observer $\mathcal{O}'$. Using these numbers, it is not difficult to show that $\mathcal{O}'$ nonetheless finds $\Delta s^2 = 8\,\text{lightsec}^2$.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 5
Introduction to 4-vectors

5.1 A mathematical interlude

We are going to have occasional lectures in this class that are more heavy on formalism, notation, and mathematics than on the physics. This is one of those lectures. The goal of Lecture 5 is to introduce you to the way in which we represent an important class of geometric objects: 4-vectors, vectors with four components that point along the 3 spatial dimensions as well as time. The reward for setting everything up in this careful way will be a representation of many quantities which we use in physics that automatically builds into it *Lorentz covariance*, meaning that quantities are defined in such a way that it is straightforward for us to transform them between reference frames.

5.2 More spacetime geometry: The displacement 4-vector

We begin with the spacetime diagram and events A and B discussed previously (using ct and ct' for the "time" directions so that they have the same units as the "space" directions):

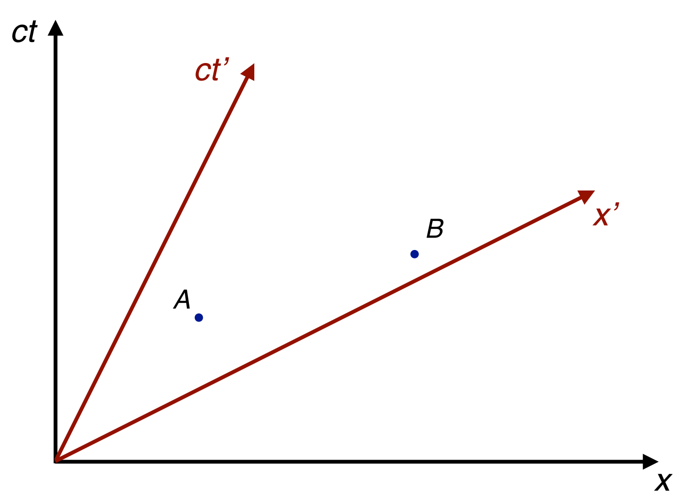
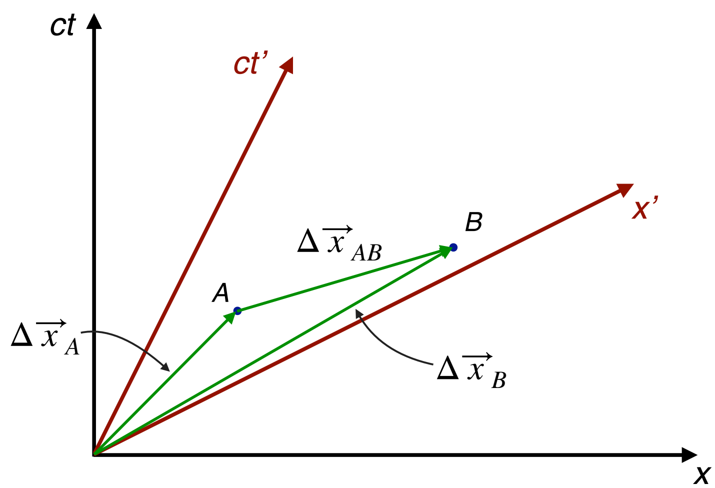


Figure 1: Two events in spacetime measured by inertial observers $\mathcal{O}$ and $\mathcal{O}'$. Observer $\mathcal{O}'$ travels with velocity $\mathbf{v} = (c/2)\mathbf{e}_x$ according to $\mathcal{O}$. We show the axes (ct, x) of $\mathcal{O}$, and the axes (ct', x') of $\mathcal{O}'$, both as seen by observer $\mathcal{O}$. In the coordinates of $\mathcal{O}$, the events A and B have spacetime coordinates $(ct_A, x_A) = (2\,\text{lightsec}, 2\,\text{lightsec})$, $(ct_B, x_B) = (3\,\text{lightsec}, 5\,\text{lightsec})$. Transforming to the frame of $\mathcal{O}'$, these become $(ct'_A, x'_A) = (2/\sqrt{3}\,\text{lightsec}, 2/\sqrt{3}\,\text{lightsec})$, $(ct'_B, x'_B) = (1/\sqrt{3}\,\text{lightsec}, 7/\sqrt{3}\,\text{lightsec})$.

Just as we can define 3-dimensional displacement vectors from one point in space to another, we can define spacetime displacement 4-vectors from one event to another in spacetime. Let us define a 4-vector $\Delta\vec{x}_A$ that points from the origin to event A, a 4-vector $\Delta\vec{x}_B$ that points from the origin to event B, and a 4-vector $\Delta\vec{x}_{AB}$ that points from event A to event B:



These displacement 4-vectors are each geometric objects, pointing from one geometric object (an event) to another (a different event). As geometric objects, 4-vectors have frame independent properties that transcend their representation in a particular frame. For example, all observers agree that $\Delta\vec{x}_{AB}$ points from event A to event B. However, because different observers assign different coordinates to these events, they assign different values to the *components* of this 4-vector. In what follows, we will use an overarrow to denote a 4-vector. Any quantity with such an arrow can be assumed to be a frame-independent geometric object.

Let us first examine one of these 4-vectors in the frame of $\mathcal{O}$. Focusing on $\Delta\vec{x}_A$, we write

$$\begin{aligned}\Delta\vec{x}_A &= \Delta x^0_A\vec{e}_0 + \Delta x^1_A\vec{e}_1 + \Delta x^2_A\vec{e}_2 + \Delta x^3_A\vec{e}_3\\ &= \Delta x^t_A\vec{e}_t + \Delta x^x_A\vec{e}_x + \Delta x^y_A\vec{e}_y + \Delta x^z_A\vec{e}_z\ .\end{aligned} \tag{5.1}$$

The 4-vectors $\Delta\vec{x}_B$ and $\Delta\vec{x}_{AB}$ can be written similarly. In writing this out, we have introduced 4 new geometric objects, the *unit vectors* $\vec{e}_t$, $\vec{e}_x$, etc. These are dimensionless 4-vectors that point along a particular observer's coordinate axes. We will soon see that they have magnitude 1 (though we need to do a little more setup before we can define what "magnitude" means precisely). The 4-vector $\vec{e}_0 \equiv \vec{e}_t$ points along the t or ct axis; $\vec{e}_1 \equiv \vec{e}_x$ is a unit vector along the x axis; etc. Because these unit vectors are geometric objects, all observers agree that (for instance) $\vec{e}_t$ points along the ct axis of observer $\mathcal{O}$. As we will see shortly, other observers will use different unit vectors adapted to their own coordinates.

We have also used two systems to label the components and the unit vectors. We will sometimes find it useful to label the axes with a name like t or x; other times, it is useful to label them with a number, like 0 or 1. Both are equivalent, and both are commonly used. The convention that is now[1] most commonly used has the time direction corresponding to 0, then the spatial directions numbered 1, 2, 3 in a right-handed system.

[1]You may find older sources that use variations on this scheme; labeling the timelike direction 4 was not uncommon especially in the early days of relativity.

In Eq. (5.1), we have also introduced the vector's four components:

$$\begin{aligned}
\Delta x^0_A &= \Delta x^t_A = c\Delta t_A \longrightarrow \text{Displacement along } ct \text{ axis from origin to } A \\
&= 2\,\text{lightseconds} \qquad (5.2)\\
\Delta x^1_A &= \Delta x^x_A = \Delta x_A \longrightarrow \text{Displacement along } x \text{ axis from origin to } A \\
&= 2\,\text{lightseconds} \qquad (5.3)\\
\Delta x^2_A &= \Delta x^y_A = \Delta y_A \\
&= 0 \qquad (5.4)\\
\Delta x^3_A &= \Delta x^z_A = \Delta z_A \\
&= 0\;. \qquad (5.5)
\end{aligned}$$

Likewise, we can write down values for the components $(\Delta x^0_B,\ \Delta x^1_B,\ \Delta x^2_B,\ \Delta x^3_B)$, $(\Delta x^t_{AB},\ \Delta x^x_{AB},\ \Delta x^y_{AB},\ \Delta x^z_{AB})$, etc.

Let's look at how the observer $\mathcal{O}'$ represents $\Delta\vec{x}_A$: they write

$$\begin{aligned}
\Delta\vec{x}_A &= \Delta x^{0'}_A\vec{e}_{0'} + \Delta x^{1'}_A\vec{e}_{1'} + \Delta x^{2'}_A\vec{e}_{2'} + \Delta x^{3'}_A\vec{e}_{3'} \\
&= \Delta x^{t'}_A\vec{e}_{t'} + \Delta x^{x'}_A\vec{e}_{x'} + \Delta x^{y'}_A\vec{e}_{y'} + \Delta x^{z'}_A\vec{e}_{z'}\;. \qquad (5.6)
\end{aligned}$$

This observer uses different unit vectors — $\vec{e}_{t'}$ points along the ct' axis, $\vec{e}_{x'}$ points along the x' axis — and they break the displacement 4-vector into different components:

$$\begin{aligned}
\Delta x^{0'}_A &= \Delta x^{t'}_A = c\Delta t'_A \longrightarrow \text{Displacement along } ct' \text{ axis from origin to } A \\
&= 2/\sqrt{3}\,\text{lightseconds} \qquad (5.7)\\
\Delta x^{1'}_A &= \Delta x^{x'}_A = \Delta x'_A \longrightarrow \text{Displacement along } x' \text{ axis from origin to } A \\
&= 2/\sqrt{3}\,\text{lightseconds} \qquad (5.8)
\end{aligned}$$

etc. The key thing to bear in mind is that the vector $\Delta\vec{x}_A$ is *exactly the same object* in both frames. However, observers $\mathcal{O}$ and $\mathcal{O}'$ break the vector up into different components, and use a different set of unit vectors.

You hopefully are familiar with ideas like this from thinking about a 3-vector in space as represented in two different coordinate systems. One coordinate system may be oriented such that a vector $\mathbf{V} = V\,\mathbf{e}_z$; another may be oriented such that $\mathbf{V} = (V/\sqrt{3})(\mathbf{e}_{x'} + \mathbf{e}_{y'} + \mathbf{e}_{z'})$. This is simply telling us that the unprimed and primed coordinate systems differ by a rotation; $\mathbf{V}$ is the same object either way[2].

5.3 Einstein summation convention

In a moment, we will examine how to relate the components $(\Delta x^t_A,\ \Delta x^x_A,\ \Delta x^y_A,\ \Delta x^z_A)$ to $(\Delta x^{t'}_A,\ \Delta x^{x'}_A,\ \Delta x^{y'}_A,\ \Delta x^{z'}_A)$, and the unit vectors $(\vec{e}_t,\ \vec{e}_x,\ \vec{e}_y,\ \vec{e}_z)$ to $(\vec{e}_{t'},\ \vec{e}_{x'},\ \vec{e}_{y'},\ \vec{e}_{z'})$. Before doing so, it is worthwhile to pause in order to introduce conventions that are very useful, that are used throughout textbooks and literature on relativity, and that we will use extensively in this course.

Writing out

$$\Delta\vec{x}_A = \Delta x^0_A\vec{e}_0 + \Delta x^1_A\vec{e}_1 + \Delta x^2_A\vec{e}_2 + \Delta x^3_A\vec{e}_3 \qquad (5.9)$$

[2]This way of thinking about how the two representations are connected is often called a *passive coordinate transformation* in mathematical literature.

over and over again is cumbersome. Notice, though, that each term on the right-hand side is the same except for the index, which shifts in value. This suggests that we rewrite this using a variable for the indices:

$$\Delta\vec{x}_A = \sum_{\mu=0}^{3} \Delta x^\mu_A \vec{e}_\mu \; . \tag{5.10}$$

We can do the same thing using the components and unit vectors for observer $\mathcal{O}'$:

$$\Delta\vec{x}_A = \sum_{\mu'=0}^{3} \Delta x^{\mu'}_A \vec{e}_{\mu'} \; . \tag{5.11}$$

(It's worth noting that a very common convention, which we will use in this class, is that if the index is a Greek letter, it denotes a spacetime direction, and so ranges from 0 to 3. If the index is a Latin letter, it is a spatial direction, and ranges from 1 to 3. It's also worth noting that a primed index tends to be difficult to read on the chalkboard, so I will usually write an overbar when working on the board, e.g. $\Delta x^{\bar\mu}_A$. This is a little more cumbersome to type up in latex, so I will tend to stick with primes in my typed-up notes.)

The next convention we introduce works as follows: if two symbols with the same index appear in an expression, one symbol has the index in the "upstairs" position and the other is "downstairs," then the summation can be assumed:

$$\Delta\vec{x}_A = \Delta x^\mu_A \vec{e}_\mu \; . \tag{5.12}$$

This is known as the *Einstein summation convention*; it appears to have been introduced in a 1916 paper[3] by Einstein describing the foundations of general relativity. We will use it extensively, and you will have plenty of chances to practice using it.

5.4 Transformations I: Displacement vector components

We now have two ways of writing a displacement vector $\Delta\vec{x}$, depending on whether we expand using the components and unit vectors of $\mathcal{O}$, or those of $\mathcal{O}'$. Putting the summation back in momentarily, we have

$$\Delta\vec{x} = \sum_{\mu=0}^{3} \Delta x^\mu \vec{e}_\mu \tag{5.13}$$

$$= \sum_{\alpha'=0}^{3} \Delta x^{\alpha'} \vec{e}_{\alpha'} \; . \tag{5.14}$$

(Comment: note that I've used changed which Greek letter I sum over in these two expressions. Because this variable gets summed over, it is called a "dummy index" — it is necessary to have *some* index in place as we expand the sum, but any letter will serve. Once the sum is performed, the index is no longer needed, and its name becomes irrelevant. We are going to start transforming quantities between reference frames, and it is a good idea to use names that do not get confused as we go between frames.)

As we have emphasized, the vector $\Delta\vec{x}$ is the same frame-independent geometric object in both of these equations. The components and unit vectors are not. How do we relate the

[3]A. Einstein, *Die Grundlage der allgemeinen Relativitätstheorie*, Ann. der Physik **49**, 769 (1916).

components in one frame to the components in the other, and how do we relate the unit vectors in one frame to the unit vectors in the other?

Transforming the components is easy: This is a displacement vector, and the components are simply the difference between the coordinates we use to label events in the two frames. Since these coordinates transform with the Lorentz transformation, it follows that the components of the displacement vector also transform with the Lorentz transformation:

$$\begin{pmatrix} \Delta x^{0'} \\ \Delta x^{1'} \\ \Delta x^{2'} \\ \Delta x^{3'} \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta x^0 \\ \Delta x^1 \\ \Delta x^2 \\ \Delta x^3 \end{pmatrix} . \tag{5.15}$$

Writing out the matrix every time is cumbersome, so we use index notation to simplify this:

$$\begin{aligned} \Delta x^{\alpha'} &= \sum_{\mu=0}^{3} \Lambda^{\alpha'}{}_{\mu} \Delta x^{\mu} \\ &= \Lambda^{\alpha'}{}_{\mu} \Delta x^{\mu} . \end{aligned} \tag{5.16}$$

The second line of Eq. (5.16) uses the Einstein summation convention. Notice that we sum over the index μ, but we do *not* sum over the index α'. We call α' a *free index* — it appears on both sides of the equation, and we are free to set α' to any of its allowed values. Equation (5.16) is thus shorthand for 4 different equations, one for each value that α' is free to take.

In Eq. (5.16), we've introduced the following notation:

$$\Lambda^{\alpha'}{}_{\mu} \equiv \text{row } \alpha', \text{ column } \mu \text{ of the Lorentz transformation matrix from } \mathcal{O} \text{ to } \mathcal{O}' . \tag{5.17}$$

Notice in writing this quantity, one index connects to quantities defined in the frame of $\mathcal{O}$, the other index connects to quantities defined in the frame of $\mathcal{O}'$. The convention we use is that this quantity is an element of the Lorentz transformation matrix that takes coordinates from the frame of the *lower* index to the frame of the *upper* index. This convention implies that the elements of the inverse transformation matrix can be written by swapping the position of the indices:

$$\Lambda^{\mu}{}_{\alpha'} \equiv \text{row } \mu, \text{ column } \alpha' \text{ of the Lorentz transformation matrix from } \mathcal{O}' \text{ to } \mathcal{O} . \tag{5.18}$$

We nail this down by requiring these matrix elements to have the following property:

$$\sum_{\alpha'=0}^{3} \Lambda^{\mu}{}_{\alpha'} \Lambda^{\alpha'}{}_{\nu} = \delta^{\mu}{}_{\nu} \tag{5.19}$$

or

$$\Lambda^{\mu}{}_{\alpha'} \Lambda^{\alpha'}{}_{\nu} = \delta^{\mu}{}_{\nu} . \tag{5.20}$$

The quantity $\delta^{\mu}{}_{\nu}$ is called the Kronecker delta:

$$\begin{aligned} \delta^{\mu}{}_{\nu} &= 1 \qquad \text{if } \mu = \nu & (5.21) \\ &= 0 \qquad \text{otherwise} . & (5.22) \end{aligned}$$

The Kronecker delta is thus an element of the identity matrix, and so Eqs. (5.19) and (5.20) do exactly what we expect if $\Lambda^{\mu}{}_{\alpha'}$ and $\Lambda^{\alpha'}{}_{\mu}$ are elements of matrices which are in inverse relation to each other.

For completeness, it should be noted that Eqs. (5.19) and (5.20) can be written in a form in which the index associated with the frame of $\mathcal{O}$ is the summed-over dummy, and the index associated with the frame of $\mathcal{O}'$ is free:

$$\sum_{\mu=0}^{3} \Lambda^{\alpha'}{}_{\mu} \Lambda^{\mu}{}_{\beta'} = \delta^{\alpha'}{}_{\beta'}$$
$$\Lambda^{\alpha'}{}_{\mu} \Lambda^{\mu}{}_{\beta'} = \delta^{\alpha'}{}_{\beta'} \ . \tag{5.23}$$

5.5 Transformations II: Unit vectors

With all this in mind, let's deduce what the transformation rule must be for the unit vectors. Working in index notation and putting sums explicitly in for clarity, we know that the components transform as

$$\Delta x^{\alpha'} = \sum_{\mu=0}^{3} \Lambda^{\alpha'}{}_{\mu} \Delta x^{\mu} \ . \tag{5.24}$$

Let's assume that there exists *some* matrix whose elements are used to relate the unit vectors $\vec{e}_\mu$ and $\vec{e}_{\alpha'}$:

$$\vec{e}_{\alpha'} = \sum_{\mu=0}^{3} M^{\mu}{}_{\alpha'} \vec{e}_{\mu} \ . \tag{5.25}$$

We do not yet know what the matrix elements $M^{\mu}{}_{\alpha'}$ represent; if the calculation we are doing is heading in the right direction, then the math will tell us what these elements are.

We also know that the displacement 4-vector $\Delta\vec{x}$ is the same geometric object no matter how we represent it:

$$\Delta\vec{x} = \sum_{\mu=0}^{3} \Delta x^{\mu} \vec{e}_{\mu} \tag{5.26}$$
$$= \sum_{\alpha'=0}^{3} \Delta x^{\alpha'} \vec{e}_{\alpha'} \ . \tag{5.27}$$

Let us plug in the transformation rules to this final line:

$$\Delta\vec{x} = \sum_{\alpha'=0}^{3} \left(\sum_{\mu=0}^{3} \Lambda^{\alpha'}{}_{\mu} \Delta x^{\mu} \right) \left(\sum_{\nu=0}^{3} M^{\nu}{}_{\alpha'} \vec{e}_{\nu} \right) \ . \tag{5.28}$$

Notice that in writing Eq. (5.28) I was very careful to make sure that all the dummy indices we sum over are distinct from one another. A common error[4] is to get indices "crossed" and accidentally connect the wrong elements in an expression to one another.

We next exchange the order of sums in Eq. (5.28) and reorganize the terms slightly:

$$\Delta\vec{x} = \sum_{\mu=0}^{3} \sum_{\nu=0}^{3} \left(\sum_{\alpha'=0}^{3} \Lambda^{\alpha'}{}_{\mu} M^{\nu}{}_{\alpha'} \right) \Delta x^{\mu} \vec{e}_{\nu} \ . \tag{5.29}$$

[4] Two weeks of my life when I was a postdoctoral researcher were spent debugging a computer code in which the root issue was exactly this — bad notation that led to indices not being properly distinguished.

Examining this expression, we see that it will work out perfectly *if*

$$\sum_{\alpha'=0}^{3} \Lambda^{\alpha'}{}_{\mu} M^{\nu}{}_{\alpha'} = \delta^{\nu}{}_{\mu} \; . \tag{5.30}$$

If this is the case, then

$$\begin{aligned}\Delta\vec{x} &= \sum_{\mu=0}^{3}\sum_{\nu=0}^{3} \delta^{\nu}{}_{\mu} \Delta x^{\mu} \vec{e}_{\nu} \\ &= \sum_{\mu=0}^{3} \Delta x^{\mu} \vec{e}_{\mu} \; . \end{aligned} \tag{5.31}$$

To go from the first line of (5.31) to the second, we use the fact that the Kronecker delta is zero unless $\nu = \mu$.

Thus, everything works as long as Eq. (5.30) holds. But this equation tells us that

$$M^{\nu}{}_{\alpha'} = \Lambda^{\nu}{}_{\alpha'} \; . \tag{5.32}$$

In other words, *the matrix that takes the unit vectors from the frame of $\mathcal{O}$ to the frame of $\mathcal{O}'$ is the* **inverse of the matrix** *that does this for the vector components.*

5.6 Summary: A glossary of transformation rules

With the results of the previous section in hand, we now have a complete set of rules describing how to Lorentz transform both vector components and unit vectors between two different inertial reference frames:

$$\Delta x^{\alpha'} = \sum_{\mu=0}^{3} \Lambda^{\alpha'}{}_{\mu} \Delta x^{\mu} = \Lambda^{\alpha'}{}_{\mu} \Delta x^{\mu} \; ; \tag{5.33}$$

$$\vec{e}_{\alpha'} = \sum_{\mu=0}^{3} \Lambda^{\mu}{}_{\alpha'} \vec{e}_{\mu} = \Lambda^{\mu}{}_{\alpha'} \vec{e}_{\mu} \; . \tag{5.34}$$

To go in the other direction (from quantities in $\mathcal{O}'$ to $\mathcal{O}$), we have

$$\Delta x^{\mu} = \sum_{\alpha'=0}^{3} \Lambda^{\mu}{}_{\alpha'} \Delta x^{\alpha'} = \Lambda^{\mu}{}_{\alpha'} \Delta x^{\alpha'} \; ; \tag{5.35}$$

$$\vec{e}_{\mu} = \sum_{\alpha'=0}^{3} \Lambda^{\alpha'}{}_{\mu} \vec{e}_{\alpha'} = \Lambda^{\alpha'}{}_{\mu} \vec{e}_{\alpha'} \; . \tag{5.36}$$

We have focused on the spacetime displacement 4-vector, but it's worth emphasizing that what we have done will serve as a prototype for developing vectors (and later, tensors) for a wide range of mathematical objects that we use to describe physics. The key idea we wish to emphasize is that a mathematical object is a 4-vector *if* its components transform between reference frames according to relationships like (5.33) and (5.35). We will assemble the components into frame-independent geometric objects using the basis vectors we have defined here, transforming as we have derived here.

You may be somewhat in despair at this moment, worried that your 8.033 life is going to be filled with tons of algebra and fiddling with indices and matrices. There will be some of that, but I ask you to carefully study the expressions (5.33)–(5.36). Notice that the results are actually quite simple in form provided we remember some simple rules:

- Whether we are transforming the components or the unit vectors, we have one object in the frame of $\mathcal{O}$ on one side of the equation, and one in the frame of $\mathcal{O}'$ on the other.
- We connect them with an element of the "Lambda matrix" that carries out the Lorentz transformation.
- We "line up the indices" so that we have one free index on the left-hand side (corresponding to the object in the frame that we are transforming to), and we sum over dummy indices on the other side to "use up" all the indices on the object in the frame that is being transformed from.

5.7 Matrix multiplication versus index notation

We all hope that you soon become fluent in using the index notation, and that equations like those in the previous few sections become intuitive and simple to manipulate. However, experience has shown that many 8.033 students at least start by wanting to write out quantities as column vectors and matrices and then combine things using techniques learned in linear algebra courses.

This can be done — but it requires some care. If you want to do analyses in this way, here are a few tricks to bear in mind:

- Think of quantities with a single "upstairs" index as a column vector, e.g.

$$\Delta x^\mu \doteq \begin{pmatrix} \Delta x^0 \\ \Delta x^1 \\ \Delta x^2 \\ \Delta x^3 \end{pmatrix} . \tag{5.37}$$

- Think of quantities with a single "downstairs" index as a row vector, e.g.

$$\vec{e}_\mu \doteq \begin{pmatrix} \vec{e}_0 \ , & \vec{e}_1 \ , & \vec{e}_2 \ , & \vec{e}_3 \end{pmatrix} . \tag{5.38}$$

- Bearing in mind that in a quantity like ${\Lambda^{\alpha'}}_\mu$, the first index refers to the row and the second to the column, think carefully about how quantities are being combined. For example, in the relationship

$$\Delta x^{\alpha'} = {\Lambda^{\alpha'}}_\mu \Delta x^\mu , \tag{5.39}$$

  we see that we are going through the matrix elements row by row, combining the element in column μ of the matrix with row μ of the column vector Δx^μ.

  This is just right multiplication of a column vector onto a matrix. Translating (5.39) into matrix form, we have

$$\begin{pmatrix} \Delta x^{0'} \\ \Delta x^{1'} \\ \Delta x^{2'} \\ \Delta x^{3'} \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta x^0 \\ \Delta x^1 \\ \Delta x^2 \\ \Delta x^3 \end{pmatrix} = \begin{pmatrix} \gamma\left(\Delta x^0 - \beta\Delta x^1\right) \\ \gamma\left(-\beta\Delta x^0 + \Delta x^1\right) \\ \Delta x^2 \\ \Delta x^3 \end{pmatrix} . \tag{5.40}$$

Consider next the relationship

$$\vec{e}_{\alpha'} = \Lambda^{\mu}{}_{\alpha'}\vec{e}_{\mu} \; . \tag{5.41}$$

In this case we are going through the matrix column by column, combining the element in row μ with the μth column of the row vector $\vec{e}_{\mu}$. This translates into linear algebra as left multiplication of the row vector onto the matrix:

$$\begin{aligned}\left(\vec{e}_{0'}\;,\quad \vec{e}_{1'}\;,\quad \vec{e}_{2'}\;,\quad \vec{e}_{3'}\right) &= \left(\vec{e}_{0}\;,\quad \vec{e}_{1}\;,\quad \vec{e}_{2}\;,\quad \vec{e}_{3}\right)\begin{pmatrix}\gamma & \gamma\beta & 0 & 0\\ \gamma\beta & \gamma & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\\ &= \left(\gamma\left(\vec{e}_0 + \beta\vec{e}_1\right)\;,\quad \gamma\left(\beta\vec{e}_0 + \vec{e}_1\right)\;,\quad \vec{e}_2\;,\quad \vec{e}_3\;.\right)\end{aligned} \tag{5.42}$$

Many, many mistakes that 8.033 students make when approaching problems using linear algebra arise because they translate quantities to column vectors that should be row vectors, and do not recognize that an analysis requires "left multiplication" rather than "right multiplication."

To illustrate how the index notation cleans things up, let us step through both of these calculations without introducing matrices. We just use the fact that the only non-zero elements of the Λ matrices are $\Lambda^{0'}{}_{0} = \Lambda^{1'}{}_{1} = \gamma$, $\Lambda^{1'}{}_{0} = \Lambda^{0'}{}_{1} = -\gamma\beta$, $\Lambda^{2'}{}_{2} = \Lambda^{3'}{}_{3} = 1$; and $\Lambda^{0}{}_{0'} = \Lambda^{1}{}_{1'} = \gamma$, $\Lambda^{1}{}_{0'} = \Lambda^{0}{}_{1'} = \gamma\beta$, $\Lambda^{2}{}_{2'} = \Lambda^{3}{}_{3'} = 1$:

$$\begin{aligned}\Delta x^{\alpha'} = \Lambda^{\alpha'}{}_{\mu}\Delta x^{\mu} \longrightarrow \qquad \Delta x^{0'} &= \Lambda^{0'}{}_{0}\Delta x^{0} + \Lambda^{0'}{}_{1}\Delta x^{1} = \gamma\left(\Delta x^0 - \beta\Delta x^1\right)\\ \Delta x^{1'} &= \Lambda^{1'}{}_{0}\Delta x^{0} + \Lambda^{1'}{}_{1}\Delta x^{1} = \gamma\left(-\beta\Delta x^0 + \Delta x^1\right)\\ \Delta x^{2'} &= \Lambda^{2'}{}_{2}\Delta x^{2} = \Delta x^2\\ \Delta x^{3'} &= \Lambda^{3'}{}_{3}\Delta x^{3} = \Delta x^3\;.\end{aligned} \tag{5.43}$$

$$\begin{aligned}\vec{e}_{\alpha'} = \Lambda^{\mu}{}_{\alpha'}\vec{e}_{\mu} \longrightarrow \qquad \vec{e}_{0'} &= \Lambda^{0}{}_{0'}\vec{e}_0 + \Lambda^{1}{}_{0'}\vec{e}_1 = \gamma\left(\vec{e}_0 + \beta\vec{e}_1\right)\\ \vec{e}_{1'} &= \Lambda^{0}{}_{1'}\vec{e}_0 + \Lambda^{1}{}_{1'}\vec{e}_1 = \gamma\left(\beta\vec{e}_0 + \vec{e}_1\right)\\ \vec{e}_{2'} &= \Lambda^{2}{}_{2'}\vec{e}_2 = \vec{e}_2\\ \vec{e}_{3'} &= \Lambda^{3}{}_{3'}\vec{e}_3 = \vec{e}_3\;.\end{aligned} \tag{5.44}$$

Notice that the final results are exactly the same both ways, but the setup in index notation is simpler. (In this case, the simplicity is in part because we were able to leave out elements whose values we know to be zero.) It is worth developing "fluency" with this notation. Until you are comfortable with this form of things, linear algebra and matrix format will work. If you approach problems this way, be very careful how you translate from index format to linear algebraic equations.

5.8 An aside: Upstairs, downstairs; contravariant, covariant

The use of index notation abounds for representing vectorial quantities (and, more generally, *tensorial* quantities — but hold that thought until we start discussing and defining tensors very soon). We will soon encounter quite a few other geometric objects whose components have indices in the "upstairs" position, like the displacement vector components discussed above. Objects with indices up, like Δx^μ, are often called *contravariant* vector components. This name comes from the fact that the magnitude of such components "contravaries" with the scale of the reference axes to which they are attached. For instance, if we change units from meters to centimeters, decreasing the scale of our reference axes, then the numerical value of the displacement vector's components are *increased* by a factor of 100 — they *contra*vary with the scale.

There are other vector-like quantities we will define soon which are more naturally expressed with indices in the "downstairs" position. An example is the *gradient*. In electricity and magnetism, you presumably learned about electrostatic fields being the gradient of a scalar potential. In the index notation that we are beginning to use, such a relationship is most naturally expressed by writing $E_i = -\partial\phi/\partial x^i$. Objects with indices down are often called *covariant* vector components. This is because as we adjust the scale of reference axes, the magnitude of these components "covaries" with the scale. Applying the example above to this situation, if we change units from meters to centimeters, the components of a gradient all *decrease* by a factor of 100.

I mention this now because we will soon be encountering quite a few quantities with indices in both positions, and many of you are likely to encounter the terms "covariant" and "contravariant" in math classes or other physics classes. Complicating all this is that we will soon learn how to move an index from the up position to the down position and vice versa, and why this is useful and important for certain problems. Lowering the index of a vector produces what is known as its *dual vector*[5]. We will get into these details very soon. Take these paragraphs as giving you a preview, as well as a heads up about how these quantities may be discussed elsewhere.

For what it's worth, I personally tend to just say "upstairs" and "downstairs" (a habit I picked up long ago from my Ph.D. supervisor). Some of the details of what's going on with contravariant and covariant components are worth knowing about, but (especially for 8.033) we often will not need to get into these details.

[5] Also known as a "1-form." I like to call this a "dual vector" in a first presentation because it emphasizes that this object belongs to a vector space, but has a particular close relationship, which will explore soon, with the "upstairs-indexed" quantities. However, the "1-form" language connects this to a set of mathematical objects known as "differential forms" which are very powerful and useful. Many of you will likely encounter these terms in other coursework soon.

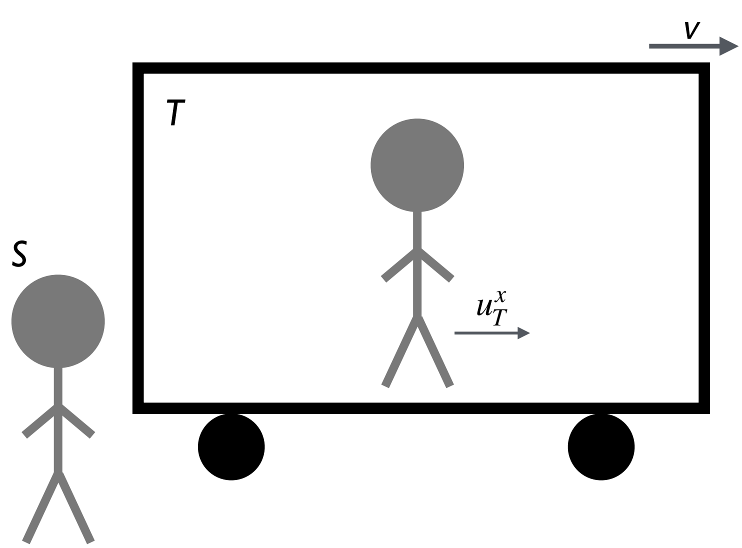
Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 6
Kinematics in spacetime

6.1 Transforming velocities

With what we've done so far, we've started to develop a good understanding of length, time, and geometry in spacetime. This is a good start for us to begin understanding physics in special relativity, but it's just a start.

In this lecture, we start examining *kinematics* — the properties of moving bodies, and how these properties transform between different reference frames. We begin by looking at velocity. Consider frame T, tied to a train, and consider a person walking inside that train. This train is moving with velocity $\mathbf{v} = v\mathbf{e}_x$ as seen by an observer who is at rest in the station frame S. The person who is walking inside the train is seen to walk with speed u^x_T, also in the x direction, by an observer who is at rest in frame T. (Comment: we will try as much as possible to use the letter u to stand for speeds inside a particular frame; we will try to use v to describe the speeds and velocities that relate two different frames.)



What is the speed u^x_S that observers in frame S measure? In Newtonian physics, we would just add the velocities in frame T to the velocity that frame T has relative to S. To get u^x_S in a world in which all observers agree that light moves at speed c, we work this out using the Lorentz transformation. On the train, we know that in a time interval Δt_T observer T moves through a distance $\Delta x_T = u^x_T \Delta t_T$. Both the time interval and the space interval are affected by the transformation:

$$
\begin{aligned}
u_S^x = \frac{\Delta x_S}{\Delta t_S} &= \frac{\gamma(\Delta x_T + v\Delta t_T)}{\gamma(\Delta t_T + v\Delta x_T/c^2)} \\
&= \frac{(\Delta x_T/\Delta t_T + v)}{(1 + \frac{v\Delta x_T}{c^2\Delta t_T})} \\
&= \frac{u_T^x + v}{1 + u_T^x v/c^2} \;. \qquad (6.1)
\end{aligned}
$$

This formula has an interesting consequence: using it, we can prove that we can never add sub-light speeds to get a speed that exceeds the speed of light. You will work this out in detail on a problem set, but to see the general idea, imagine that $u_T^x = v = 0.9c$. Then,

$$
u_S^x = \frac{0.9c + 0.9c}{1 + (0.9c)(0.9c)/c^2} = \frac{1.8c}{1.81} = 0.9945c \;. \qquad (6.2)
$$

How do components of the velocity perpendicular to the frames' relative motion transform? Imagine that the person on the train has motion along the y direction as well, so that in Δt_T they move through $\Delta y_T = u_T^y \Delta t_T$. Then,

$$
\begin{aligned}
u_S^y = \frac{\Delta y_S}{\Delta t_S} &= \frac{\Delta y_T}{\gamma(\Delta t_T + v\Delta x_T/c^2)} \\
&= \frac{u_T^y}{\gamma(1 + u_T^x v/c^2)} \;. \qquad (6.3)
\end{aligned}
$$

(Note that the factor $\gamma = 1/\sqrt{1 - v^2/c^2}$ — it only depends on the relative speed v of the two frames, it does not involve the velocity $\mathbf{u}$.) If the person on the train has velocity along the z direction, then it transforms like Eq. (6.3) as well, replacing u^y with u^z.

6.2 Momentum I: Did we break physics???

A lesson of the previous section is that how velocities add is "weird" as compared to Newtonian expectations. These expectations follow the logic of Galilean relativity, so it should not be too surprising that things change when we impose the rule that c is the same to all observers. However, our laws of classical mechanics have implicitly assumed Galilean relativity. What happens to important principles like conservation of momentum when we "update" our rules for how velocities add?

Let us first review how conservation of momentum works in Newtonian physics. Suppose that we have N_i bodies that come together in some fashion, interact, and then have N_f bodies in the final state. Conservation of momentum tells us that

$$
\sum_{j=1}^{N_i} m_j^{\text{initial}} \mathbf{u}_j^{\text{initial}} = \sum_{j=1}^{N_f} m_j^{\text{final}} \mathbf{u}_j^{\text{final}} \;. \qquad (6.4)
$$

As long as

$$
\sum_{j=1}^{N_i} m_j^{\text{initial}} = \sum_{j=1}^{N_f} m_j^{\text{final}} \;, \qquad (6.5)
$$

this relation holds in all Galilean reference frames.

Let's take a look at what happens when we examine this law in Lorentzian reference frames. Let's consider something really simple: two particles, A and B, of identical mass that collide and rebound elastically. We assume that it remains the case that $\mathbf{p} = m\mathbf{u}$, and begin by examining this situation in the *center of momentum* frame, i.e. the frame in which the net momentum of the system is zero:

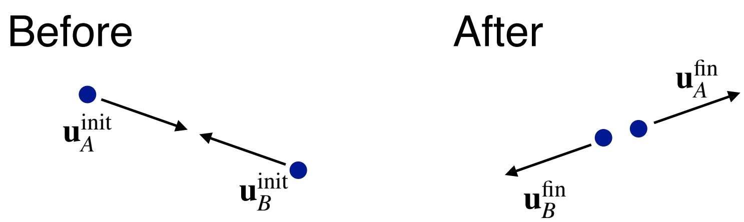


Figure 1: Elastic collision of identical bodies in the center of momentum frame.

The bodies' velocities are given by $\mathbf{u}_A^{\mathrm{init}} = u^x\mathbf{e}_x - u^y\mathbf{e}_y$, $\mathbf{u}_B^{\mathrm{init}} = -u^x\mathbf{e}_x + u^y\mathbf{e}_y$ before the collision. Afterwards, we have $\mathbf{u}_A^{\mathrm{fin}} = u^x\mathbf{e}_x + u^y\mathbf{e}_y$, $\mathbf{u}_B^{\mathrm{init}} = -u^x\mathbf{e}_x - u^y\mathbf{e}_y$. Because $m_A = m_B$, we can see that momentum is clearly conserved: It is zero both before and after the collision.

Let's next examine this from another frame of reference. Suppose we examine this collision from a frame that moves with velocity $\mathbf{v} = -u^x\mathbf{e}_x$ with respect to the center of momentum frame. The horizontal motion of particle B is canceled out here; if we are in this frame, we are essentially jogging along with particle B.

What are the velocity vectors in this frame? To find out, we use the relativistic velocity addition formulas we just worked out to get the components of these vectors. Let's do the x components first:

$$u_A^{x\prime} = \frac{u^x + u^x}{1 + (u^x)^2/c^2} = \frac{2u^x}{1 + (u^x)^2/c^2} \;, \tag{6.6}$$

$$u_B^{x\prime} = \frac{u^x - u^x}{1 - (u^x)^2/c^2} = 0 \;. \tag{6.7}$$

Notice in this frame, the horizontal velocity components are not equal and opposite, and so the system must have some non-zero horizontal momentum component. This is not surprising: we've moved into a frame in which the entire system is moving in the $+x$ direction, so we expect the system to have momentum along x.

Next, look at the y components:

$$u_A^{y\prime} = -\frac{u^y}{\gamma(1 + (u^x)^2/c^2)} = -\frac{u^y\sqrt{1 - (u^x)^2/c^2}}{1 + (u^x)^2/c^2} \;, \tag{6.8}$$

$$u_B^{y\prime} = \frac{u^y}{\gamma(1 - (u^x)^2/c^2)} = \frac{u^y}{\sqrt{1 - (u^x)^2/c^2}} \;. \tag{6.9}$$

The γ that we use here is the one corresponding to the velocity of this frame relative to the center of momentum frame: $\mathbf{v} = -u^x\mathbf{e}_x$, and so $\gamma = 1/\sqrt{1 - (u^x)^2/c^2}$.

Notice that the velocity components in the vertical direction are **no longer equal and opposite**. This means that they do not balance out, and so the system has net momentum in the vertical direction. In other words, under the hypothesis that momentum $\mathbf{p} = m\mathbf{u}$, we appear to have a problem: *The system appears to have acquired momentum in the y direction by moving into a new frame that is moving in the $-x$ direction with respect to the center of momentum frame.*

Our hypothesis that c is the same to all observers, which led to our new velocity addition rules, appears to have broken momentum.

6.3 Momentum II: From Newtonian momentum to Einsteinian momentum

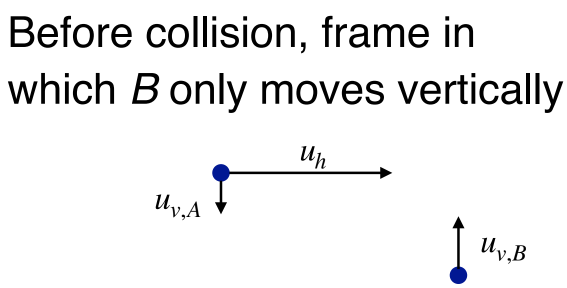
This appears disturbing. However, we have already seen (and you examined on a pset) that the Lorentz transformations are *approximately* consistent with Galilean coordinate transformations. Galilean relativity (and thus Newtonian physics) works fine when speeds are far smaller than c. Perhaps the root cause of this disturbing apparent breakdown is that Newtonian momentum (which respects Galilean relativity) is itself an approximation to a more "Lorentzian" quantity.

Let us try the hypothesis that momentum is defined by

$$\mathbf{p} = \alpha(u) m \mathbf{u} \,. \tag{6.10}$$

The function $\alpha(u)$ is a function corrects the magnitude of momentum, and only depends on the magnitude of the body's velocity $\mathbf{u}$.

With this in mind, we re-examine the collision from the Lorentz frame in which particle B has no horizontal motion:



To simplify some of the analysis which will follow later, we've introduced new labels for the velocity components of these bodies: u_h is the horizontal velocity component of body A in this frame; $u_{v,A}$ is the vertical velocity component of A in this frame; and $u_{v,B}$ is the vertical velocity component of B in this frame. Comparing to our previous calculations given in Eqs. (6.6), (6.8), and (6.9), these velocity components according to the relativistic velocity addition formula are given by

$$u_h = \frac{2u^x}{1 + (u^x)^2/c^2} \,, \quad u_{v,A} = -\frac{u^y\sqrt{1 - (u^x)^2/c^2}}{1 + (u^x)^2/c^2} \,, \quad u_{v,B} = \frac{u^y}{\sqrt{1 - (u^x)^2/c^2}} \,. \tag{6.11}$$

These velocity components turn out to be nicely related to one another. Notice that

$$u_{v,A} = -u_{v,B}\left(\frac{1 - (u^x)^2/c^2}{1 + (u^x)^2/c^2}\right) \,. \tag{6.12}$$

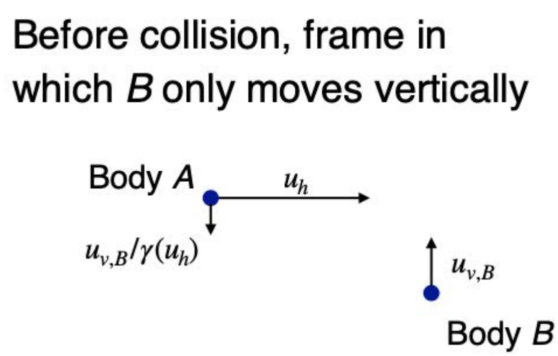
The factor in parentheses in Eq. (6.12) turns out be related to u_h in an interesting way. Calculate the value of γ for $v = u_h$:

$$\begin{aligned}
\gamma(u_h) &= \frac{1}{\sqrt{1-(u_h)^2/c^2}} = \left(1 - \frac{4(u^x)^2/c^2}{(1+(u^x)^2/c^2)^2}\right)^{-1/2} \\
&= \left(\frac{1-2(u^x)^2/c^2 + (u^x)^4/c^4}{1+2(u^x)^2/c^2+(u^x)^4/c^4}\right)^{-1/2} \\
&= \left(\frac{(1-(u^x)^2/c^2)^2}{(1+(u^x)^2/c^2)^2}\right)^{-1/2} \\
&= \frac{1+(u^x)^2/c^2}{1-(u^x)^2/c^2} \ . \qquad (6.13)
\end{aligned}$$

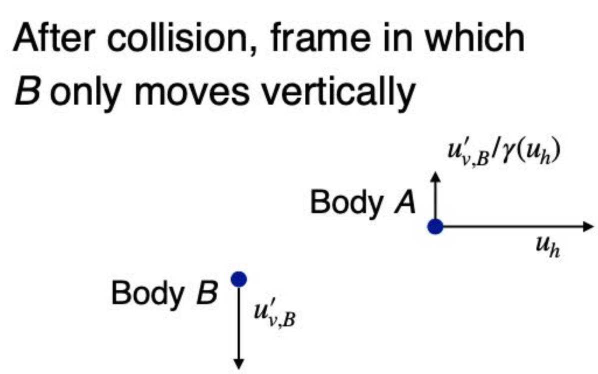
Modulo a reciprocal, this is exactly the parentheses factor in (6.12). This allows us to rewrite this equation as

$$u_{v,A} = -u_{v,B}/\gamma(u_h) \ . \qquad (6.14)$$

Let's take advantage of this to remake the figure of the collision in this frame using only the velocity components u_h and $u_{v,B}$ for our labels:



If momentum is conserved, then we expect the situation after the collision to look as follows:



The logic by which we have sketched this is that the horizontal components of the bodies' motion cannot be affected by the collision, so body A continues moving to the right with speed u_h, and body B continues to have no horizontal motion. The vertical motions reverse in direction. We leave open the possibility that the speeds associated with the vertical motion might be affected (hence the primes: $u'_{v,B}$ might differ from $u_{v,B}$).

We now demand conservation of momentum according to our hypothesized new form: both components of $\mathbf{p} = \alpha(u)m\mathbf{u}$ must be the same before and after the collision. First look at the horizontal component, for which the only contribution comes from body A:

$$\alpha\left(\sqrt{(u_h)^2 + (u_{v,B}/\gamma(u_h))^2}\right) m u_h = \alpha\left(\sqrt{(u_h)^2 + (u'_{v,B}/\gamma(u_h))^2}\right) m u_h \; . \tag{6.15}$$

The only way that this equation can hold independent of the function $\alpha(u)$ (whose nature we don't yet know) is if $u'_{v,B} = u_{v,B}$. The speed associated with the vertical components' of the bodies' velocities must be the same before and after the collision. Those velocity components simply change direction.

Require next that the vertical components of momentum be conserved:

$$\begin{aligned}\alpha(u_{v,B})mu_{v,B} - \alpha\left(\sqrt{(u_h)^2 + (u_{v,B}/\gamma(u_h))^2}\right) mu_{v,B}/\gamma(u_h) = \\ -\alpha(u_{v,B})mu_{v,B} + \alpha\left(\sqrt{(u_h)^2 + (u_{v,B}/\gamma(u_h))^2}\right) mu_{v,B}/\gamma(u_h)\end{aligned}$$

Moving similar looking factors to the same side of the equation, dividing by a common factor of $mu_{v,B}$, and multiplying by $\gamma(u_h)$, this becomes

$$\alpha\left(\sqrt{(u_h)^2 + (u_{v,B}/\gamma(u_h))^2}\right) = \gamma(u_h)\alpha(u_{v,B}) \; . \tag{6.16}$$

To simplify this, let us require that $\alpha(0) = 1$. This requirement insures that this formula recovers the Newtonian limit, which we know is an extremely good approximation for small speeds. We then examine Eq. (6.16) in the limit $u_{v,B} \to 0$:

$$\alpha(u_h) = \gamma(u_h) \; . \tag{6.17}$$

The factor $\alpha(u)$ that we hypothesized must be included in the definition of momentum works perfectly if it is the special relativistic γ factor.

In summary, the momentum defined by

$$\mathbf{p} = \gamma(u)m\mathbf{u} \tag{6.18}$$

is conserved in a universe that respects Lorentz covariance.

6.4 Kinetic energy

In Newtonian physics, the change in kinetic energy is the work done on a body: Integrating from some initial position $\mathbf{x}_i$ to a final position $\mathbf{x}_f$, we have

$$\begin{aligned}K_f - K_i &= \int_i^f \frac{d\mathbf{p}}{dt}\cdot d\mathbf{x} = \int_i^f \frac{d}{dt}(m\mathbf{u})\cdot\mathbf{u}\,dt = m\int_i^f \mathbf{u}\cdot d\mathbf{u} \\ &= \frac{1}{2}m\left(u_f^2 - u_i^2\right) \; .\end{aligned} \tag{6.19}$$

We now define relativistic kinetic energy in exactly the same way, but replace the Newtonian formula for momentum with the version we just derived:

$$\begin{aligned} K_f - K_i = \int_i^f \frac{d\mathbf{p}}{dt} \cdot d\mathbf{x} &= \int_i^f \frac{d}{dt}\left[\gamma(u) m\mathbf{u}\right] \cdot \mathbf{u}\, dt \\ &= m \int_i^f \mathbf{u} \cdot d\left[\frac{\mathbf{u}}{\sqrt{1-u^2/c^2}}\right] \ . \end{aligned} \tag{6.20}$$

The fact that this second form of $K_f - K_i$ is identical to the first one is not obvious. It is not difficult however to demonstrate that the two lines of (6.20) are equivalent by using the chain rule to expand the two differentials.

The final integrand that we have derived can be manipulated further:

$$\mathbf{u} \cdot d\left[\frac{\mathbf{u}}{\sqrt{1-u^2/c^2}}\right] = d\left[\frac{u^2}{\sqrt{1-u^2/c^2}}\right] - \frac{\mathbf{u} \cdot d\mathbf{u}}{\sqrt{1-u^2/c^2}} \ . \tag{6.21}$$

This is a very nice form: the first term on the right-hand side of (6.21) is a perfect differential; the second term on the right-hand side is simple to integrate up. Doing so, we find

$$\begin{aligned} K_f - K_i &= \left.\frac{mu^2}{\sqrt{1-u^2/c^2}}\right|_i^f - m\int_i^f \frac{\mathbf{u} \cdot d\mathbf{u}}{\sqrt{1-u^2/c^2}} \\ &= \left.\frac{mu^2}{\sqrt{1-u^2/c^2}}\right|_i^f + \left. mc^2\sqrt{1-u^2/c^2}\right|_i^f \ . \end{aligned} \tag{6.22}$$

For our final simplification, let's take the initial velocity to be $\mathbf{u}_i = 0$, and define $\mathbf{u}_f \equiv \mathbf{u}$. Since the initial velocity is zero, the initial kinetic energy $K_i = 0$. We set $K_f \equiv K$ and finally obtain for the kinetic energy of the system

$$\begin{aligned} K &= \frac{mu^2}{\sqrt{1-u^2/c^2}} + mc^2\sqrt{1-u^2/c^2} - mc^2 \\ &= \frac{mu^2 + mc^2 - mu^2}{\sqrt{1-u^2/c^2}} - mc^2 \\ &= \frac{mc^2}{\sqrt{1-u^2/c^2}} - mc^2 \\ &= \left[\gamma(u) - 1\right] mc^2 \ . \end{aligned} \tag{6.23}$$

To interpret this quantity, we define the body to have a *total* energy

$$E = \gamma(u) mc^2 \ ; \tag{6.24}$$

then, $E = K + mc^2$, and we interpret mc^2 as the body's **rest energy**: energy which the body possesses even when it is not in motion.

It's fair to say that Eq. (6.24) with $\gamma = 1$ is the most famous physics equation in the world. It is really interesting to pause and reflect on how it arose: we began by exploring the consequences of the hypothesis that light travels at speed c for all observers. This forced us to replace the Galilean transformation with the Lorentz transformation. This in turn mandated an adjustment to the definition of momentum. The formula $E = mc^2$, which some would argue literally changed the world, thus arose fundamentally as a consequence of this deceptively simple hypothesis.

6.5 Aside: "Relativistic mass" and why we generally don't use it anymore

In some older texts, you will see the energy and momentum defined as follows:

$$E = m(u)c^2 \, , \qquad \mathbf{p} = m(u)\mathbf{u} \, , \tag{6.25}$$

where they have defined $m(u) = \gamma(u)m$, the "relativistic mass" of the body whose "rest mass" is m. This definition rarely appears in modern relativity texts. Instead, the only "mass" used to define a body is its rest mass. The main reason for this is that m is an *invariant* — different observers assign a different energy to the body, depending on its speed u in their rest frame, but they all agree that the body's mass is m (and its energy is mc^2) in its own rest frame. As we will see in the next lecture, this invariant plays a particularly important rule in helping us to define a 4-vector which will prove to be extremely useful in helping us to keep track of energy and momentum in relativistic physics.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 7
4-momentum and 4-velocity

7.1 Transforming energy and momentum between reference frames

The requirement that all observers measure the speed of light to be c has led us to rather different formulations of energy and momentum: a body which has *rest mass* m (i.e., the mass that we measure it to have when it is at rest with respect to us) which we see to be moving with velocity $\mathbf{u}$ has an energy E and a momentum $\mathbf{p}$ given by

$$E = \gamma(u)mc^2\ , \qquad \mathbf{p} = \gamma(u)m\mathbf{u}\ . \tag{7.1}$$

These quantities respect conservation laws: a system's total E and $\mathbf{p}$ are conserved as its constituents interact with one another. In the limit $u/c \ll 1$, these formulas reduce to

$$E = mc^2 + \frac{1}{2}mu^2 + O(u^4/c^2)\ , \qquad \mathbf{p} = m\mathbf{u} + O(u^3/c)\ . \tag{7.2}$$

This makes it clear that Newtonian momentum agrees with relativistic momentum for speeds much smaller than c. The energies likewise agree in this limit, provided we account for the body's *rest energy* mc^2. In the vast majority of circumstances a body's rest energy is bound up in the body, and cannot be "used" for anything in their interaction, so it can be ignored; we essentially measure all energies relative to mc^2 rather than relative to zero. The relativistic quantities and the Newtonian quantities thus agree perfectly when $u \ll c$.

Suppose we measure a body to have energy E_L and momentum $\mathbf{p}_L$ in our laboratory. What energy E_T and momentum $\mathbf{p}_T$ will an observer moving past our lab in a train with velocity $\mathbf{v} = v\mathbf{e}_x$ measure the body to have? To figure this out, follow this recipe:

1. Deduce the 3-velocity $\mathbf{u}_L$ of the body in the lab from the values of E_L and $\mathbf{p}_L$.

2. Use the velocity addition formulas to compute the 3-velocity of $\mathbf{u}_T$ of the body as measured by observers on the train.

3. From $\mathbf{u}_T$, compute E_T and $\mathbf{p}_T$.

You will work through these steps on a problem set. The result you find is

$$\begin{aligned} E_T = \gamma(E_L - vp^x_L)\ , &\qquad p^x_T = \gamma(p^x_L - vE_L/c^2)\ , \\ p^y_T = p^y_L\ , &\qquad p^z_T = p^z_L\ . \end{aligned} \tag{7.3}$$

Tweaking notation slightly, we rewrite this

$$\begin{pmatrix} E_T/c \\ p^x_T \\ p^y_T \\ p^z_T \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} E_L/c \\ p^x_L \\ p^y_L \\ p^z_L \end{pmatrix}\ . \tag{7.4}$$

In other words, the *relativistic formulations of energy and momentum form a set of quantities that transform under a Lorentz transformation.*

7.2 An invariant for energy and momentum

Recall that we found $\Delta s^2 = -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$ is a Lorentz invariant: all Lorentz frames agree on the value of Δs^2 between two events. Can we do something similar with energy and momentum?

Looking at how E and $\mathbf{p}$ behave under a Lorentz transformation, let's think of energy as the "timelike" component of momentum (E/c actually — which hopefully makes sense since we need our quantities to have the right dimensions[1]). Let's see what happens when we examine "negative time bit squared" plus "space bit squared":

$$-\frac{E^2}{c^2} + (p^x)^2 + (p^y)^2 + (p^z)^2 = -\frac{E^2}{c^2} + |\mathbf{p}|^2 \,. \tag{7.5}$$

Plug into this

$$E^2 = \gamma^2 m^2 c^4 = \frac{m^2c^4}{1-u^2/c^2} \,, \tag{7.6}$$

$$|\mathbf{p}|^2 = \gamma^2 m^2 u^2 = \frac{m^2u^2}{1-u^2/c^2} \,. \tag{7.7}$$

Putting these together, we have

$$\begin{aligned}
-\frac{E^2}{c^2} + |\mathbf{p}|^2 &= \frac{m^2u^2 - m^2c^2}{1-u^2/c^2} = -m^2c^2\left(\frac{1-u^2/c^2}{1-u^2/c^2}\right) \\
&= -m^2c^2 \,. \qquad (7.8)
\end{aligned}$$

Multiplying this by $-c^2$,

$$E^2 - |\mathbf{p}|^2c^2 = m^2c^4 \qquad \text{or} \qquad E^2 = |\mathbf{p}|^2c^2 + m^2c^4 \,. \tag{7.9}$$

In other words, although different Lorentz frames will measure E and $\mathbf{p}$ differently, all frames agree that E^2 and $|\mathbf{p}|^2$ are related by the expressions given in Eq. (7.9).

Notice that if $m = 0$, then $|\mathbf{p}| = E/c$: massless bodies carry non-zero momentum. This relationship corresponds perfectly to the energy and momentum carried by electromagnetic radiation (compare with the Poynting vector if you need a refresher in this concept). Recall that our analysis began by noting that Maxwell's equations appear to "want" c to be the same for all observers. It is satisfying that when we make energy and momentum consistent with this concept, the result automatically respects the relationship between energy and momentum that electrodynamics teaches us for radiation.

7.3 The 4-momentum

By virtue of the way in which E/c and $p^{x,y,z}$ transform, we can see that they behave exactly like the components of the displacement 4-vector. This tells us that we really should define a 4-vector whose components all have the dimensions of momentum:

$$\vec{p} = \sum_{\mu=0}^{3} p^\mu \vec{e}_\mu \,, \tag{7.10}$$

[1] Note that if we use units such that $c = 1$, energy and momentum (and mass, for that matter) all have the same dimensions. This is another benefit of this system of units.

with

$$p^0 = E/c \,, \qquad p^1 = p^x \,, \qquad p^1 = p^y \,, \qquad p^3 = p^z \,. \tag{7.11}$$

This $\vec{p}$ is then a geometric object: observers in all Lorentz frames use this 4-vector to describe the system's energy and momentum, but break it up into components and unit vectors differently. If the components and unit vectors according to $\mathcal{O}$ are p^μ and $\vec{e}_\mu$, then an observer $\mathcal{O}'$ constructs $\vec{p}$ using

$$p^{\mu'} = \Lambda^{\mu'}{}_\alpha p^\alpha \,, \qquad \vec{e}_{\mu'} = \Lambda^\alpha{}_{\mu'}\vec{e}_\alpha \tag{7.12}$$

(switching to the Einstein summation convention). The matrix elements $\Lambda^{\mu'}{}_\alpha$ perform the Lorentz transformation of event labels from the frame of $\mathcal{O}$ to the frame of $\mathcal{O}'$; the matrix elements $\Lambda^\alpha{}_{\mu'}$ perform the inverse transformation.

The reason why this is useful for us is that conservation of energy and conservation of momentum are now combined into a single law: the *conservation of 4-momentum*. Suppose N_i bodies interact, resulting in N_f bodies afterwards. Then,

$$\sum_{j=1}^{N_i} \vec{p}_j^{\,\rm init} = \sum_{j=1}^{N_f} \vec{p}_j^{\,\rm final} \,, \tag{7.13}$$

where $\vec{p}_j^{\,\rm init}$ is the initial 4-momentum of particle j, and $\vec{p}_j^{\,\rm final}$ is the final 4-momentum of particle j.

7.4 4-vectors in general; scalar products of 4-vectors

Let's pause a moment to reflect on the logic by which we assembled the 4-momentum. We essentially followed the following recipe:

1. We found that a grouping of 4 quantities plays a meaningful role in physics: $p^0 = E/c$, $p^{1,2,3} = p^{x,y,z}$, with E and $p^{x,y,z}$ now defined using the "relativistic" rules we derived in Lecture 6.

2. We found that when we change reference frames, these 4 quantities are transformed to the new frame by the Lorentz transformation exactly as the components of the 4-displacement are: $p^{\mu'} = \Lambda^{\mu'}{}_\alpha p^\alpha$.

3. Since it behaves under the transformation law exactly like the 4-vector we discussed previously, we define p^μ as the components of a new 4-vector, $\vec{p}$, and use this 4-vector as a tool in our physics moving forward.

We can do this for *any* set of 4 quantities that turns out to be meaningful for our analysis. In other words,

> If any set b^μ with $\mu \in [0, 1, 2, 3]$ has the property that when we change reference frames their values are related by a Lorentz transformation, $b^{\mu'} = \Lambda^{\mu'}{}_\alpha b^\alpha$, then b^μ represent the components of a 4-vector: $\vec{b} = b^\mu \vec{e}_\mu$.

Once we have identified these quantities as the components of a 4-vector, we can start identifying invariants. Whatever the vector $\vec{b}$ represents, we are guaranteed that all Lorentz frames agree on the value of $-(b^0)^2 + (b^1)^2 + (b^2)^2 + (b^3)^2$. In fact, it is not hard to show that we can define a more general notion of an invariant. Suppose $\vec{a} = a^\mu \vec{e}_\mu$ and $\vec{b} = b^\mu \vec{e}_\mu$. Then,

$$\vec{a} \cdot \vec{b} \equiv -a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3 \tag{7.14}$$

is a Lorentz invariant: all Lorentz frames agree on the value of $\vec{a} \cdot \vec{b}$. This is simply proven by transforming the components of $\vec{a}$ and $\vec{b}$ to another frame and then showing that the right-hand side of (7.14) in the new frame is unchanged from its value in the original frame.

Equation (7.14) defines what we call the "scalar product" between two 4-vectors. We will now use the term "scalar" only to refer to a quantity whose value is invariant to Lorentz transformations. This a bit different from how you likely have thought of scalars previously. For example, in Newtonian mechanics a body's energy E is often taken to be a scalar, since it is a quantity that does not have a direction associated with it. In relativity, E is not a scalar since its value changes according to the Lorentz frame in which we measure it. (To save some of your older intuition, note that we now think of a body's energy as the timelike component of its 4-momentum, modulo factors of c. In relativity, E *does* have a direction associated with it — it's a timelike component of a 4-vector.)

A (rather obvious) corollary of the fact that the scalar product of two 4-vectors is a Lorentz invariant is that the scalar product of any 4-vector with itself is a Lorentz invariant. Two quantities we've recently examined can be rephrased using this definition:

$$\Delta\vec{x} \cdot \Delta\vec{x} = \Delta s^2 \ , \tag{7.15}$$

$$\vec{p} \cdot \vec{p} = -m^2 c^2 \ . \tag{7.16}$$

The resemblance to the invariant interval Δs^2 gives us a convention for describing 4-vectors. For any 4-vector $\vec{a}$, if

$$\vec{a} \cdot \vec{a} < 0 \tag{7.17}$$

then we say that $\vec{a}$ is *timelike*. This means that we can find a Lorentz frame in which only the time component of $\vec{a}$ is non-zero: $\vec{a}$ has no spatial components in that frame. If

$$\vec{a} \cdot \vec{a} > 0 \tag{7.18}$$

then we say that $\vec{a}$ is *spacelike*. There exists a[2] Lorentz frame in which $\vec{a}$ has no component in the time direction; it points purely in a spatial direction. Finally, if

$$\vec{a} \cdot \vec{a} = 0 \tag{7.19}$$

then $\vec{a}$ is *lightlike* or *null*. In all Lorentz frames, $\vec{a}$ points along light cones.

Notice that $\vec{p}$ is either timelike or lightlike, and is only lightlike for $m = 0$.

7.5 4-velocity

In Newtonian mechanics, velocity and momentum were related by a factor of the body's mass. Let's do the same thing using the 4-momentum, and define the quantity that results as the 4-velocity:

$$\vec{u} = \frac{1}{m}\vec{p} \ . \tag{7.20}$$

[2]Actually, *many* such Lorentz frames: once we find one, any Lorentz frame that is related to the first by a rotation will do the trick.

What does $\vec{u}$ mean? Let's look at its components:

$$u^0 = \frac{p^0}{m} = \frac{E}{mc} = \gamma\, c\,, \tag{7.21}$$

$$u^1 = \frac{p^1}{m} = \gamma\, (\mathbf{u})^x\ , \tag{7.22}$$

$$u^2 = \frac{p^2}{m} = \gamma\, (\mathbf{u})^y\ , \tag{7.23}$$

$$u^3 = \frac{p^3}{m} = \gamma\, (\mathbf{u})^z\ . \tag{7.24}$$

Note the notation used on the spatial components: $(\mathbf{u})^x$ means the x component of the 3-vector $\mathbf{u}$, and likewise for the y and z components. The spatial components of $\vec{u}$ thus look just like "normal" 3-velocity, but are multiplied by γ. How do we interpret the γ factor?

Consider someone passing by with 3-velocity $\mathbf{u}$. That person's clocks run slow according to you: as an interval $d\tau$ passes on their clock, an interval dt passes on your clock, with

$$dt = \gamma\, d\tau\ . \tag{7.25}$$

If, for example, $\gamma = 2$, then we measure 2 seconds passing for every 1 second interval that they record. We define the interval $d\tau$ as the *proper time*: it is an interval of time according to the clock of the observer (or object) who we say is moving. The word "proper" in this case comes from a meaning that denotes "belonging to oneself." Hence an observer's proper time is the time which that observer measures.

Proper time is a useful quantity because it is a Lorentz invariant: *all* Lorentz frames agree that the observer in motion measures a time interval $d\tau$. That won't be the time interval we measure as observer $\mathcal{O}$ whizzes by us at 90% of the speed of light; it won't be what our friend $\mathcal{F}$ measures as they whizz by at 90% of the speed of light in another direction; but we all agree that it *is* what $\mathcal{O}$ measures. It is a useful benchmark whose meaning all agree on.

With this in mind, let's re-examine the spatial components of the 4-velocity:

$$u^x = \gamma\, (\mathbf{u})^x = \gamma \frac{dx}{dt} = \frac{dx}{d\tau}\ , \tag{7.26}$$

$$u^y = \gamma\, (\mathbf{u})^y = \gamma \frac{dy}{dt} = \frac{dy}{d\tau}\ , \tag{7.27}$$

$$u^z = \gamma\, (\mathbf{u})^z = \gamma \frac{dz}{dt} = \frac{dz}{d\tau}\ , \tag{7.28}$$

Let's also look at the timelike component:

$$u^t = \gamma\, c = \gamma\, c \frac{dt}{dt} = c \frac{dt}{d\tau}\ . \tag{7.29}$$

Comparing with how we defined the displacement 4-vector, we see that

$$\vec{u} = \frac{d\vec{x}}{d\tau}\ . \tag{7.30}$$

The 4-velocity is the rate at which something moves through spacetime *per unit proper time*.

It's worth computing the invariant associated with the 4-velocity:

$$\vec{u} \cdot \vec{u} = \frac{1}{m^2} \vec{p} \cdot \vec{p} = -\frac{m^2 c^2}{m^2} = -c^2\ . \tag{7.31}$$

The 4-velocity of a body which is at rest in some Lorentz frame has the same $\vec{u} \cdot \vec{u}$ as a body which is moving $0.99999999999c$ in that frame.

Notice that $\vec{u}$ is a timelike 4-vector. Because of this, $\vec{u}$ does not really "work" for a "body" moving at the speed of light: γ diverges there. This is consistent with the fact that our original definition starts with $\vec{u} = \vec{p}/m$, and the only "objects" we know of that travel at the speed of light have $m = 0$. 4-vectors are geometric objects, and we cannot make a timelike 4-vector into a lightlike one.

7.6 4-velocity contrasted with 3-velocity

We now have two important ways to characterize a moving body's motion:

- 3-velocity $\mathbf{u} = d\mathbf{x}/dt$ describes motion through **space** per unit **time**. Both "space" and "time" are frame-dependent concepts, and so $\mathbf{u}$ depends on the frame in which it is measured.

- 4-velocity $\vec{u} = d\vec{x}/d\tau$ describes motion through **spacetime** per unit **proper time**. It is a frame-independent, geometric object; the same $\vec{u}$ is used by all observers.

A major conceptual difference between these two quantities is how we regard them when observed in different Lorentz frames:

- As a frame-independent geometric object, all observers agree on an object's 4-velocity $\vec{u}$. They assign it different components, however, and use different unit vectors when expanding $\vec{u}$ into components:

$$\vec{u} = u^\mu \vec{e}_\mu = u^{\alpha'} \vec{e}_{\alpha'} \ , \tag{7.32}$$

  where

$$u^{\alpha'} = \Lambda^{\alpha'}{}_\mu u^\mu \ , \qquad \vec{e}_{\alpha'} = \Lambda^\mu{}_{\alpha'} \vec{e}_\mu \ . \tag{7.33}$$

- The 3-vector is actually different in the two frames. Given $\mathbf{u}$, we find the components of $\mathbf{u}'$ which describe the body's motion in a new frame by applying the velocity addition formulas: if the relative motion of the two frames is given by $\mathbf{v} = v\mathbf{e}_x$, then

$$(\mathbf{u})^{x'} = \frac{(\mathbf{u})^x + v}{1 + (\mathbf{u})^x v/c^2} \ , \tag{7.34}$$

$$(\mathbf{u})^{y'} = \frac{(\mathbf{u})^y}{\gamma(v)(1 + (\mathbf{u})^x v/c^2)} \ , \tag{7.35}$$

$$(\mathbf{u})^{z'} = \frac{(\mathbf{u})^z}{\gamma(v)(1 + (\mathbf{u})^x v/c^2)} \ . \tag{7.36}$$

Both the 3-velocity and the 4-velocity are important and useful. The 3-velocity is what we measure in our own reference frame: we see a body move through a spatial displacement $\Delta\mathbf{x}$ in an interval of time that our clocks measure to be Δt; we thus determine that the body has a 3-velocity $\mathbf{u} = \Delta\mathbf{x}/\Delta t$. From this, we can construct the body's 4-velocity. This gives us a geometric object that gives us an excellent tool for describing the body's trajectory in spacetime. We will to be fluent with both 3- and 4-velocities, and at ease with translating between them.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 8
Using 4-momentum

8.1 Introduction; a note on notation

In this lecture, we are going to examine how we use 4-momentum, seeing how it serves as a tool that combines the familiar notions of mass, momentum, and energy conservation into a single mathematical "device." One of the goals of this examination will be to see how we can use invariants to write certain quantities in ways that make the analysis easy.

Be aware that we are going overload a bit of notation, the dot product. When we write the dot product between two 4-vectors, that tells us to compute the invariant scalar product: $\vec{a} \cdot \vec{b} = -a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3$. When we write the dot product between two 3-vectors, that's the dot product you've learned in previous physics courses: $\mathbf{a} \cdot \mathbf{b} = a^x b^x + a^y b^y + a^z b^z$. This is arguably a bit sloppy, but our use will be unambiguous as long as we consistently use the dot product symbol only in these two circumstances exactly as defined here.

Several of the examples used in this lecture are inspired by or taken from the textbook *Introduction to Elementary Particles*, by David J. Griffiths (Chapter 3).

8.2 The energy measured by a particular observer

Suppose that body A moves through spacetime with 4-momentum $\vec{p}_A$. Suppose that observer $\mathcal{O}$ has 4-velocity $\vec{u}_{\mathcal{O}}$; in our lab L we measure the components of $\vec{u}_{\mathcal{O}}$ to be ($u^t_{L,\mathcal{O}}$, $u^x_{L,\mathcal{O}}$, $u^y_{L,\mathcal{O}}$, $u^z_{L,\mathcal{O}}$). What does $\mathcal{O}$ measure for the energy of body A?

Perhaps the most straightforward way to do this would be as follows:

1. Perform a Lorentz transformation to take us to the rest frame of $\mathcal{O}$. In this frame, the components of $\vec{u}_{\mathcal{O}}$ are given by $(u^t_{\mathcal{O}},\, u^x_{\mathcal{O}}, u^y_{\mathcal{O}}, u^z_{\mathcal{O}}) = (c,\, 0,\, 0,\, 0)$.

2. Apply this Lorentz transformation to the components of $\vec{p}_A$; call these components $p^\alpha_{A;\mathcal{O}}$.

3. After applying the Lorentz transformation, the timelike component $p^t_{A;\mathcal{O}}$ is equal to the energy of body A in the rest frame of $\mathcal{O}$, modulo a factor of c. In other words, the energy of body A as measured by $\mathcal{O}$ is

$$E_{A;\mathcal{O}} = c\, p^t_{A;\mathcal{O}} \,. \tag{8.1}$$

This way of doing things is straightforward, and in principle we could do this to determine the energy of body A for any observer. However, note that the final result can be written

$$E_{A;\mathcal{O}} = p^t_{A;\mathcal{O}} u^t_{\mathcal{O}} - p^x_{A;\mathcal{O}} u^x_{\mathcal{O}} - p^y_{A;\mathcal{O}} u^y_{\mathcal{O}} - p^z_{A;\mathcal{O}} u^z_{\mathcal{O}} \,. \tag{8.2}$$

On initial glance, this may seem like a rather stupid way of rewriting Eq. (8.1): The term with the t components on the right-hand side of (8.2) is the same as what's on the right-hand side of (8.1), but the other 3 terms we subtract off are all equal to zero since $u_{\mathcal{O}}^{x,y,z} = 0$. However, this rewriting makes it clear that $E_{a;\mathcal{O}}$ is just the scalar product of body A's 4-momentum with observer $\mathcal{O}$'s 4-velocity, modulo a minus sign:

$$E_{A;\mathcal{O}} = -\vec{p}_A \cdot \vec{u}_{\mathcal{O}} \; . \tag{8.3}$$

This is a particularly lovely way of writing this quantity because the *scalar product is an invariant*. As long as we know the components of both $\vec{p}_A$ and $\vec{u}_{\mathcal{O}}$ in some frame of reference, we can use Eq. (8.3) to compute body A's energy as measured by $\mathcal{O}$ *without needing to perform the Lorentz transformation to the rest frame of* $\mathcal{O}$.

In addition to being a very useful way of writing the energy that some specified observer measures (we will find this form of the energy to be useful for several applications over the course of this semester), Eq. (8.3) serves as an exemplar of the power of writing things in terms of Lorentz invariants. Many times, it might be conceptually straightforward (but perhaps algebraically tedious) to figure out a quantity in a particular frame. If you can take that result and reformulate it as a Lorentz invariant, you will have a result that is broadly applicable and often much easier to apply.

Note: you might be confused about the fact that the "energy" defined by Eq. (8.3) is a Lorentz scalar. In the previous lecture, we quite specifically used energy as an example of a quantity that is **not** a scalar in relativistic physics! Are we not contradicting ourselves?

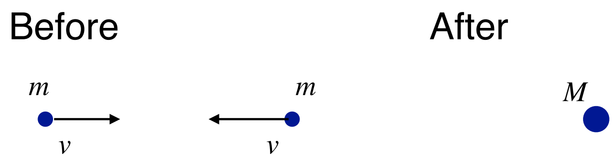
The issue here is that we are using the word "energy" for two different, albeit related, physical quantities: the timelike component of a body's 4-momentum, and the property of a body as measured by some particular observer. The first quantity we call "energy" is certainly not a Lorentz invariant — different frames assign different values to the timelike component of $\vec{p}$. The second such quantity is a Lorentz invariant because *all IRFs agree that this is the energy measured by that observer.* It is similar to the fact that the proper time experienced by some observer is a Lorentz invariant, even though "time" is certainly not Lorentz invariant. An observer's proper time may not be the time that I measure, it may not be the time that you measure, but it *is* the time which that observer measures. We all agree on that.

As a higher-level side issue, it's worth noting that a lot of confusion about various concepts in physics can be traced back to the fact that the terms we use in human language to describe things often has some built-in ambiguity. The mathematical language that we use to describe physics does not. The philosophy of "shut up and calculate," though a tad rude, is often a really way useful to get out of a confusing jam.

8.3 Collisions and decays

8.3.1 A simple collision

Let's begin by looking at some situations in which we can use conservation of 4-momentum to deduce what is going on. Begin by imagining that we smash together two lumps of clay. Each lump has mass m; we shoot them at each other, one with velocity $\mathbf{v} = v\mathbf{e}_x$, the other with $\mathbf{v} = -v\mathbf{e}_x$. They combine into a new lump of mass M. What is M?



The 4-momentum of the two lumps before we shoot them together has components

$$p^\alpha_{B,j} \doteq \begin{pmatrix} \gamma mc \\ \pm\gamma mv \\ 0 \\ 0 \end{pmatrix} , \qquad \gamma = 1/\sqrt{1 - v^2/c^2} \ . \tag{8.4}$$

The symbol "$\doteq$" we have introduced here means "the components on the left-hand side are given by the column vector on the right-hand side." Put $j = R$ to label the lump moving to the right (for which we choose the + sign), and $j = L$ for the lump moving to the left (for which we choose the $-$ sign).

After the collision, we have

$$p^\alpha_A \doteq \begin{pmatrix} Mc \\ 0 \\ 0 \\ 0 \end{pmatrix} . \tag{8.5}$$

The final lump is at rest in the frame we are using, so $\gamma = 1$ afterwards, and there are no non-zero spatial components to $\vec{p}_A$. Enforcing $\vec{p}_{B,R} + \vec{p}_{B,L} = \vec{p}_A$ tells us

$$M = 2\gamma(v)m \ . \tag{8.6}$$

The Newtonian expectation of course is that mass is simply conserved: $M = 2m$ in Newtonian physics. In relativity, we see that $M > 2m$. Indeed, if v is large, the amount beyond the Newtonian expectation can be significant. For instance, if $v = 3c/5$, then $M = 2.5m$ — the rest mass has increased by 25% in this case.

Where has that "extra" rest mass come from? This is $E = mc^2$ in action: kinetic energy has been converted into rest mass. When we collide two lumps at high speed, the remnant of the collision will be hotter than if we combine them at low speed. That kinetic energy gets incorporated into the random, thermal motion of the molecules that constitute the lumps. In essence, this tells us that a body's rest mass is higher when it is hot than when it is cold.

8.3.2 A simple decay

The previous example is somewhat contrived. However, it is the *time reverse* of processes that happens all the time: the *decay* of bodies with mass M_B into products whose total mass

M_A is less than the starting mass. Let's consider such a decay process: A body of mass M decays into two bodies of mass m, which then recoil in opposite directions with speed v. What is v?

Conserving 4-momentum leads us to exactly the same equation as before:

$$M = 2\gamma(v)m \; . \tag{8.7}$$

Now, we take M and m as knowns, and solve for v:

$$v = c\sqrt{1 - \left(\frac{2m}{M}\right)^2} \; . \tag{8.8}$$

Notice that if $m = M/2$, $v = 0$: all of the original rest energy turned into rest energy in the new bodies. If $m < M/2$, then some of that rest energy has become kinetic energy. (If $m > M/2$, then we've got nonsense! Check your measurements.)

8.3.3 A not-quite-so-simple decay

Although the above decay example is illustrative, it is also somewhat contrived. A more realistic example is decay into two *unequal* mass bodies. In fact, quite a few important examples involve decay into products with $m = 0$. Here's a fairly simple example: the decay of a charged pion into a muon and a massless neutrino[1]:

$$\pi^- \rightarrow \mu^- + \bar{\nu} \; . \tag{8.9}$$

This equation means that the (negatively charged) pion decays into a (negatively charged) muon and an antineutrino. This equation guarantees that charge, spin, and a quantity called "lepton number" are also conserved. If the details of this interest you, you should investigate future coursework in nuclear and particle physics. Our focus here is solely on the issue of 4-momentum conservation. We take the pion that starts this process to be at rest in our laboratory, so its 4-momentum components in the lab are given by

$$p^\alpha_\pi \doteq \begin{pmatrix} m_\pi c \\ 0 \\ 0 \\ 0 \end{pmatrix} \; . \tag{8.10}$$

[1]We now know that the neutrino has a non-zero mass, so the analysis I am presenting here is not quite right. However, the mass is so small that we have not yet actually measured it (although we have "upper bounds" on how big it can be). You should treat the idea of a massless neutrino as a very useful approximation. Hopefully we ("we" meaning the scientific community at large) will be able to refine these analyses with a mass estimate before too long.

The neutrino *cannot* be at rest: as a zero-mass particle[2] it must have non-zero 3-momentum. Let's define the neutrino's momentum as along the positive x axis:

$$p_\nu^\alpha \doteq \begin{pmatrix} E_\nu/c \\ E_\nu/c \\ 0 \\ 0 \end{pmatrix} . \tag{8.11}$$

This form of p_ν^α guarantees that $\vec{p}_\nu \cdot \vec{p}_\nu = 0$, the correct value of this invariant for a massless particle. The final quantity we need is the 4-momentum of the muon[3]. A little thought tells us that it must have the form

$$p_{\rm muon}^\alpha \doteq \begin{pmatrix} \gamma(v) m_{\rm muon} c \\ -\gamma(v) m_{\rm muon} v \\ 0 \\ 0 \end{pmatrix} . \tag{8.12}$$

This means the muon, with rest mass $m_{\rm muon}$, moves in the $-x$ direction with speed v.

Let's now enforce conservation of 4-momentum and determine (a) the energy of the neutrino, and (b) the speed v with which the pion recoils. We require both components of 4-momentum to balance:

$$\begin{aligned} & \vec{p}_\pi = \vec{p}_{\rm muon} + \vec{p}_\nu \\ t \text{ component:} \quad & m_\pi c = E_\nu/c + \gamma(v) m_{\rm muon} c \\ x \text{ component:} \quad & 0 = E_\nu/c - \gamma(v) m_{\rm muon} v \ . \end{aligned} \tag{8.13}$$

The x component equation allows us to eliminate E_ν from the t component equation. Doing so, we have

$$m_\pi c = \gamma(v) m_{\rm muon}(v + c) \ . \tag{8.14}$$

Square both sides of this and divide by c^2:

$$m_\pi^2 = m_{\rm muon}^2 \left[\frac{(v+c)^2}{c^2 - v^2}\right] = m_{\rm muon}^2 \left[\frac{c+v}{c-v}\right] . \tag{8.15}$$

Solving this for v, we find

$$v = c\left(\frac{m_\pi^2 - m_{\rm muon}^2}{m_\pi^2 + m_{\rm muon}^2}\right) . \tag{8.16}$$

From this, it's a straightforward exercise to return to the x component equation and solve for E_ν. The result is

$$E_\nu = \frac{1}{2}\left(\frac{m_\pi^2 - m_{\rm muon}^2}{m_\pi}\right) c^2 \ . \tag{8.17}$$

[2] Again, ignoring current wisdom that neutrinos actually have a very small mass.

[3] Note that there's potential for confusion here: we've written out the word "muon" rather than used the conventional symbol μ in order to avoid confusing μ with a downstairs index. The Greek alphabet gets a tad overused from time to time in physics.

8.3.4 A not-quite-so-simple decay, revisited

The calculation we just did is the most straightforward way to take conservation of 4-momentum and grind out the quantities of interest. You should be aware, though, that we can exploit the properties of 4-vectors to expedite our grinding of this algebra. Let's start with our initial statement of conservation of 4-momentum:

$$\vec{p}_\pi = \vec{p}_{\rm muon} + \vec{p}_\nu \; . \tag{8.18}$$

Let's move the neutrino's momentum to the left-hand side, then construct the invariant scalar product of each side with itself:

$$(\vec{p}_\pi - \vec{p}_\nu) \cdot (\vec{p}_\pi - \vec{p}_\nu) = \vec{p}_{\rm muon} \cdot \vec{p}_{\rm muon} \tag{8.19}$$

which expands to

$$\vec{p}_\pi \cdot \vec{p}_\pi - 2\vec{p}_\pi \cdot \vec{p}_\nu + \vec{p}_\nu \cdot \vec{p}_\nu = \vec{p}_{\rm muon} \cdot \vec{p}_{\rm muon} \; . \tag{8.20}$$

These various scalar products appearing in this equation take *extremely* simple forms:

$$\begin{aligned}
\vec{p}_\pi \cdot \vec{p}_\pi &= -m_\pi^2 c^2 \\
\vec{p}_{\rm muon} \cdot \vec{p}_{\rm muon} &= -m_{\rm muon}^2 c^2 \\
\vec{p}_\nu \cdot \vec{p}_\nu &= 0 \\
\vec{p}_\pi \cdot \vec{p}_\nu &= -p_\pi^0 p_\nu^0 + p_\pi^1 p_\nu^1 = -(m_\pi c)(E_\nu/c) + 0 = -m_\pi E_\nu \; .
\end{aligned} \tag{8.21}$$

Putting all these together, we have

$$m_\pi^2 - 2m_\pi E_\nu/c^2 = m_{\rm muon}^2 \tag{8.22}$$

or

$$E_\nu = \frac{1}{2}\left(\frac{m_\pi^2 - m_{\rm muon}^2}{m_\pi}\right) c^2 \; . \tag{8.23}$$

This is exactly the result for the neutrino energy we derived before.

Let's carry the analysis a few more steps in order to see a few more useful tricks. The neutrino's 3-momentum has magnitude E_ν/c, and is in the $+x$ direction. From this we know that the muon's 3-momentum has magnitude

$$|\mathbf{p}_{\rm muon}| = E_\nu/c \; , \tag{8.24}$$

and is in the $-x$ direction. Given a body's 3-momentum and mass, we can use the 4-momentum invariant to compute its energy:

$$E_{\rm muon}^2 = |\mathbf{p}_{\rm muon}|^2 c^2 + m_{\rm muon}^2 c^4 \; . \tag{8.25}$$

However, we also know that for any body

$$E = \gamma(v) m c^2 \; , \qquad \mathbf{p} = \gamma(v) m \mathbf{v} \; , \tag{8.26}$$

where $\mathbf{v}$ is that body's 3-velocity. This tells us that if we know a body's relativistic energy and relativistic momentum, we can construct its 3-velocity:

$$\mathbf{v} = \frac{\mathbf{p}c^2}{E} \; . \tag{8.27}$$

Plugging in the quantities we just found for describing pion decay, let's check the recoil velocity of the muon:

$$\begin{aligned}|\mathbf{v}| \equiv v &= \frac{|\mathbf{p}_{\rm muon}|c^2}{E_{\rm muon}} \\ &= \frac{E_\nu c}{\sqrt{E_\nu^2 + m_{\rm muon}^2 c^4}} \\ &= c\left(\frac{m_\pi^2 - m_{\rm muon}^2}{m_\pi^2 + m_{\rm muon}^2}\right) \ . \end{aligned} \tag{8.28}$$

On the last line, I plugged in our result for E_ν, and ground through a bit of algebra.

Taking advantage of the invariant scalar product often offers a quick route to isolating and finding quantities of interest in your problem. It's not a "Get Out of Algebra Free" card, but it often significantly simplifies a step or three of your analysis.

8.3.5 The center of momentum (COM) frame

Some problems can be greatly simplified by changing the reference frame in which we do the calculation. A frame that often turns out to be useful is the *center of momentum*, or COM, frame: the frame in which the total 3-momentum of the system is zero. As it happens, this has been the case in all the examples we've examined so far. This is because it just happened that these examples considered problems in which the system had no net 3-momentum in the "lab" frame in which we formulated the analysis. That is not always the case.

A classic example is a collision onto a stationary target. An important example (which I have taken from the textbook by Griffiths) is the collision of a high-speed proton with a proton which is at rest in the lab frame. One of the early experiments used this set-up to produce antiprotons, the antimatter version of protons:

$$p + p \to p + p + p + \bar{p} \ . \tag{8.29}$$

(Overbar on a particle means the antiparticle.) In the lab frame, here's the situation:

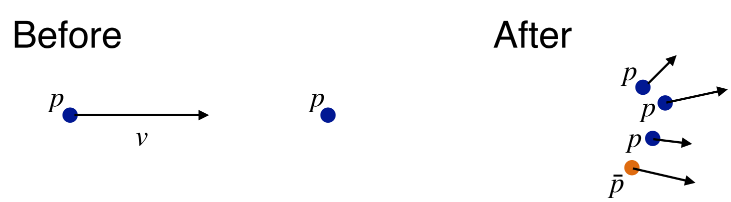


Figure 1: Proton incident on a stationary proton target; lab frame.

Our interest is to compute the *threshold* energy the incoming proton must have in order for the reaction (8.29) to occur. In the lab frame, this is hard to figure out, largely because all of the reaction products must zoom to the right in order for momentum to be conserved. But, there exists some frame in which the incident proton moves to the right (slower than in the lab frame) and the target proton moves to the left at exactly the same speed as the incident proton. The system has zero 3-momentum in that frame; by conservation, the reaction products will have total 3-momentum summing to zero as well:

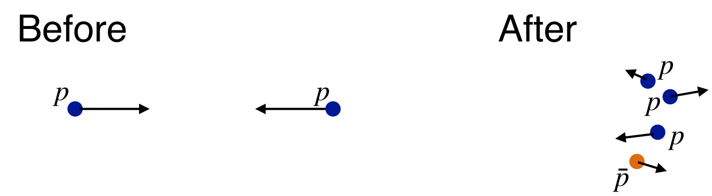


Figure 2: Proton incident on a proton target; center of momentum frame.

The "threshold" incident energy is the minimum energy necessary in order for the reaction (8.29) to proceed. With a little thought, the meaning of this energy in the COM frame should be clear: it's the energy at which the reaction products are produced with no kinetic energy. We produce only rest mass, not "wasting" any of the energy into the particles' motion (at least in this frame; they will certainly be in motion in the lab frame, since the system has net momentum in that frame).

Conservation of 4-momentum tells us that this system is governed by

$$\vec{p}_{\rm inc} + \vec{p}_{\rm target} = \vec{p}_1 + \vec{p}_2 + \vec{p}_3 + \vec{p}_{\rm anti}\ . \tag{8.30}$$

It is really easy to write down the components of the left-hand side in the lab frame:

$$p^\alpha_{\rm inc,lab} \doteq \begin{pmatrix} E_{\rm inc}/c \\ p^x \\ 0 \\ 0 \end{pmatrix}\ , \qquad p^\alpha_{\rm target,lab} \doteq \begin{pmatrix} mc \\ 0 \\ 0 \\ 0 \end{pmatrix}\ . \tag{8.31}$$

It is also really easy to write down the components of quantities on the right-hand side at threshold in the COM frame:

$$p^\alpha_{\rm 1,COM} = p^\alpha_{\rm 2,COM} = p^\alpha_{\rm 3,COM} = p^\alpha_{\rm anti,COM} \doteq \begin{pmatrix} mc \\ 0 \\ 0 \\ 0 \end{pmatrix}\ . \tag{8.32}$$

In these expressions, m is the proton mass, which is identical to the mass of the antiproton. (Note that in the lab frame, the 3-momentum component p^x can be rewritten using $E_{\rm inc}$ and m. Hold that thought for a moment.)

The difficulty we now face is that if we try to enforce Eq. (8.30) with what we've got so far, we're in trouble: the left-hand side and the right-hand side are expressed in different frames. The 4-momenta will not equate until we put them in the same frame. However, the *invariants* we can construct from them must equate no matter what frame we use to write down the various $\vec{p}$s. So, instead of examining Eq. (8.30), examine

$$(\vec{p}_{\rm inc} + \vec{p}_{\rm target})\cdot(\vec{p}_{\rm inc} + \vec{p}_{\rm target}) = (\vec{p}_1 + \vec{p}_2 + \vec{p}_3 + \vec{p}_{\rm anti})\cdot(\vec{p}_1 + \vec{p}_2 + \vec{p}_3 + \vec{p}_{\rm anti})\ , \tag{8.33}$$

We evaluate the invariant on the left-hand side by using the components we've written down in the lab frame:

$$(\vec{p}_{\rm inc} + \vec{p}_{\rm target})\cdot(\vec{p}_{\rm inc} + \vec{p}_{\rm target}) = -(E_{\rm inc}/c + mc)^2 + (p^x)^2\ . \tag{8.34}$$

We evaluate the invariant on the right-hand side by using the components we've written down in the COM frame:

$$(\vec{p}_1 + \vec{p}_2 + \vec{p}_3 + \vec{p}_{\rm anti}) \cdot (\vec{p}_1 + \vec{p}_2 + \vec{p}_3 + \vec{p}_{\rm anti}) = -(4mc)^2 \,. \tag{8.35}$$

Equate these two expressions, use $E^2 = |\mathbf{p}|^2c^2 + m^2c^4$ to eliminate the lab frame p^x, solve for $E_{\rm inc}$. The result is

$$E_{\rm inc} = 7mc^2 \,. \tag{8.36}$$

This means that this reaction will proceed if the incident proton's *kinetic* energy ($E_{\rm inc}$ is its *total* energy, which includes rest energy mc^2) is 6 times the proton's rest energy.

8.4 Scattering

A special case of a collision are scattering interactions: particle A comes in, interacts with particle B, and both then emerge from the interaction with new 4-momenta. Or, there could be numerous particles A_1, A_2, ... which interact with numerous particles B_1, B_2, ... This is exactly the situation we examined when we considered how to refine the definition of momentum to insure that momentum was still conserved after learning how to add velocities properly. In all cases, we are simply governed by the rule that the total 4-momentum before must equal the total 4-momentum after.

One example of a scattering interaction is particularly interesting: light interacting with a charge q of mass m_q. (The value of the charge will play no role in the calculation we are about to do, but light does not interact with non-charged bodies.) Experiments first performed by Arthur Compton in 1923 showed that in such interactions, the interaction of light with the charged body behaved just like inelastic collisions between particles. Such experiments played a large role in making it clear that light has a particle-like nature, which we call the "photon."

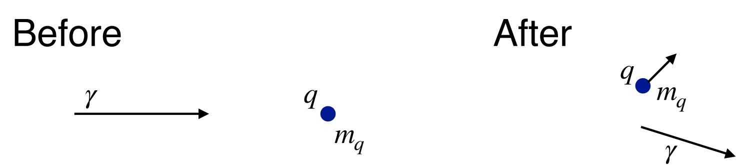


Figure 3: Compton scattering off of a charge q of mass m_q.

Suppose that the incident photon (denoted γ) comes down the x axis, but that the scattered photon makes an angle θ to the x axis. Then the situation afterwards introduces momentum along a new axis: the total momentum before they scatter is

$$p^\alpha_B \doteq \begin{pmatrix} E_\gamma/c + m_q c \\ E_\gamma/c \\ 0 \\ 0 \end{pmatrix} \;; \tag{8.37}$$

the total momentum after scattering is

$$p^\alpha_A \doteq \begin{pmatrix} E'_\gamma/c + E_q/c \\ p^x_q + E'_\gamma \cos\theta/c \\ p^y_q - E'_\gamma \sin\theta/c \\ 0 \end{pmatrix} \;, \tag{8.38}$$

where E'_γ is the photon energy after scattering.

To make progress, we use the fact that the energy of a photon is simply related to its frequency or wavelength via

$$E_\gamma = h\nu = hc/\lambda \tag{8.39}$$

where h is *Planck's constant*:

$$h = 6.626 \times 10^{-34}\,\mathrm{J\,sec}\ . \tag{8.40}$$

Enforcing $\vec{p}_A = \vec{p}_B$ and making judicious use of our invariants, a few lines of algebra yields the Compton scattering law:

$$\lambda' = \lambda + \frac{h}{m_q c}\left(1 - \cos\theta\right)\ . \tag{8.41}$$

Some of the light's energy and momentum is transferred to the charged mass; the light is less energetic (longer wavelength) as a consequence. We will step you through this analysis on a problem set (some of you may have already seen this in quantum mechanics class).

Note that the quantity $h/m_q c$ has the dimensions of length; it is sometimes called the "Compton wavelength" of the mass m_q.

8.5 Doppler effect and aberration

The invariance of the speed of light to all observers has been the central organizing principle of almost everything we've done since Lecture 3. But this raises an interesting question: if two different frames both see a beam of light moving with speed c, what about that beam appears different to the two observers?

Let's make this concrete by examining a beam of light as seen by two observers: our station-frame observer S, and an observer T riding through the station on a train with velocity $\mathbf{v} = v\mathbf{e}_x$. Let's say that the station-frame observer reports the beam to have energy $E = h\nu$, and that it is moving in the (x, y) plane, making an angle θ with the x axis. This means that the station-frame observer measures the beam to have 4-momentum components

$$p_S^\alpha \doteq \begin{pmatrix} h\nu/c \\ h\nu\cos\theta/c \\ h\nu\sin\theta/c \\ 0 \end{pmatrix}\ . \tag{8.42}$$

What components does the observer on the train report? As usual, we apply the Lorentz transformation: $p_T^{\mu'} = \Lambda^{\mu'}{}_\alpha p_S^\alpha$, where $\Lambda^{\mu'}{}_\alpha$ is the matrix which takes events from frame S to frame T. The result is

$$p_T^{\mu'} \doteq \begin{pmatrix} \gamma h\nu/c(1 - v\cos\theta/c) \\ \gamma h\nu/c(\cos\theta - v/c) \\ h\nu\sin\theta/c \\ 0 \end{pmatrix} = \begin{pmatrix} h\nu'/c \\ h\nu'\cos\theta'/c \\ h\nu'\sin\theta'/c \\ 0 \end{pmatrix}\ . \tag{8.43}$$

The result is that, according to the train observer, the beam of light has a different energy $h\nu'$ and travels at a different angle θ'. (It's a straightforward exercise to equate the two ways I have written the components $p_T^{\mu'}$ to work out ν' and θ'.) The shift of the light's energy is the Doppler effect, the same basic physics by which we hear the frequency of a siren change pitch as an emergency vehicle drives past us at high speed. The change in angle is aberration. You explored the phenomenon of light's trajectory changing angle according to different observers on a recent problem set; such an analysis can be done quite elegantly using 4-momentum.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.033 FALL 2024

LECTURE 9
SOME MORE MATH: THE METRIC TENSOR, 1-FORMS AKA DUAL VECTORS, TENSORS MORE GENERALLY

9.1 The scalar product revisited

Similar to Lecture 5, this lecture again largely focuses on mathematical issues. We have introduced you to 4-vectors, and have shown how they can be used to organize a *Lorentz covariant* presentation of some of the laws of physics. In this lecture, we expand the "vocabulary" of mathematical objects that we use to describe quantities in relativistic physics.

We begin by revisiting the the scalar product between two 4-vectors,

$$\vec{a} \cdot \vec{b} = -a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3 \; . \tag{9.1}$$

It is not difficult to show that $\vec{a} \cdot \vec{b}$ is invariant. Indeed, "scalar product" refers to the fact that a "scalar" in relativistic physics is a quantity that is invariant across Lorentz frames, a more specialized and specific meaning than you have likely encountered in previous coursework.

As written, there is nothing wrong with Eq. (9.1). We used this very form to help understand invariants associated with relativistic energy and momentum. However, from a certain perspective Eq. (9.1) can be regarded as "distasteful." It's necessary to write the whole expression out; there's no nice shorthand that lets you write this expression in index notation if we follow this form.

To correct for these shortcomings, we introduce a new mathematical object called the *metric*. The metric is a *tensor*, a mathematical object that we define more completely below. For now, you can regard it is an object with two indices that is represented in a particular Lorentz frame by a matrix. The metric has components $\eta_{\alpha\beta}$ given by

$$\eta_{\alpha\beta} \doteq \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \; . \tag{9.2}$$

As in the previous lecture, we use the symbol "$\doteq$" to stand for "the object on the left-hand side has the components on the right-hand side." Using the metric, you should be able to convince yourself quite easily that Eq. (9.1) is equivalent to

$$\vec{a} \cdot \vec{b} = \sum_{\alpha=0}^{3} \sum_{\beta=0}^{3} \eta_{\alpha\beta} a^\alpha b^\beta = \eta_{\alpha\beta} a^\alpha b^\beta \; . \tag{9.3}$$

The second form, using the Einstein summation convention, is how the invariant scalar product is most commonly written out.

Let's see what the invariance of the scalar product tells us about how the components of the metric transform between reference frames. Suppose that observer $\mathcal{O}$ measures $\vec{a}$ and $\vec{b}$ to have components a^α and b^β, and they use $\eta_{\alpha\beta}$ for metric components. Observer $\mathcal{O}'$ measures these vectors to have components $a^{\mu'}$ and $b^{\nu'}$, and they use $\eta_{\mu'\nu'}$ for metric components. The components of the vectors are related in the usual way by the Lorentz transformation matrix:

$$a^\alpha = \Lambda^\alpha{}_{\mu'} a^{\mu'} \tag{9.4}$$

$$b^\beta = \Lambda^\beta{}_{\nu'} b^{\nu'} \; . \tag{9.5}$$

How do we compute the metric components used by $\mathcal{O}'$? We figure this out by enforcing invariance:

$$\begin{aligned}
\vec{a} \cdot \vec{b} &= \eta_{\alpha\beta} a^\alpha b^\beta \\
&= \eta_{\alpha\beta} \left(\Lambda^\alpha{}_{\mu'} a^{\mu'}\right) \left(\Lambda^\beta{}_{\nu'} b^{\nu'}\right) \\
&= \left(\eta_{\alpha\beta} \Lambda^\alpha{}_{\mu'} \Lambda^\beta{}_{\nu'}\right) a^{\mu'} b^{\nu'} \; .
\end{aligned} \tag{9.6}$$

This quantity is an invariant provided we transform the components of the metric via the rule

$$\eta_{\mu'\nu'} = \eta_{\alpha\beta} \Lambda^\alpha{}_{\mu'} \Lambda^\beta{}_{\nu'} \; . \tag{9.7}$$

Notice that this is basically just the "line up the indices" rule that we discussed when we introduced index notation. **CAUTION:** if you want to do this analysis using matrix multiplication techniques that you learned in linear algebra, you must be very careful — it is quite easy to go awry. See my comment in the final section of these lecture notes.

I've gone through the calculation of how the metric transforms with some care because I want to make clear the *principle* behind how we transform tensor components. In a few pages, we are going to apply the ideas discussed here to tensors in general. As with 4-vectors, the behavior of quantities under transformation is central to our definition of what a tensor is. With that said, it must be noted that for the metric the final result is so simple that all the calculation presented above surely will feel like distressing overkill: $\eta_{\alpha\beta}$ is represented by the matrix

$$\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{9.8}$$

in *all* Lorentz frames. This can be proved by computing Eq. (9.7).

One last detail: you might be wondering what happened, in Eq. (9.3), with the unit vectors which go into the vectors $\vec{a}$ and $\vec{b}$. After all, if $\vec{a} = a^\alpha \vec{e}_\alpha$ and $\vec{b} = b^\beta \vec{e}_\beta$, shouldn't it also be the case that

$$\vec{a} \cdot \vec{b} = \left(a^\alpha b^\beta\right) \left(\vec{e}_\alpha \cdot \vec{e}_\beta\right) \tag{9.9}$$

is exactly equivalent to the form presented in (9.3)?

The answer is certainly *yes*. Comparing Eqs. (9.3) and (9.9) shows us that for these forms to be equivalent, then we must have

$$\vec{e}_\alpha \cdot \vec{e}_\beta = \eta_{\alpha\beta} \; . \tag{9.10}$$

This, at last, allows us to see how the geometric objects $\vec{e}_\alpha$ are, in fact, *unit* vectors: the scalar product of any two unit vectors is zero if $\alpha \neq \beta$; the scalar product is 1 when $\alpha = \beta$

and correspond to one of the spatial directions; and the scalar product is -1 when $\alpha = \beta = t$. The negative scalar product is what we expect for timelike vectors, so $\vec{e}_t \cdot \vec{e}_t = -1$ should make sense, although it looks starkly different from the "modulus squared" you have seen with unit vectors in previous classes.

As discussed above, $\eta_{\alpha\beta}$ is represented by the matrix (9.8) in *all* reference frames. This means that when we change frames, and then build the unit vectors in the new frame,

$$\vec{e}_{\mu'} = \Lambda^{\alpha}{}_{\mu'} \vec{e}_{\alpha} \ , \tag{9.11}$$

we must have $\vec{e}_{t'} \cdot \vec{e}_{t'} = -1$, $\vec{e}_{x'} \cdot \vec{e}_{x'} = 1$, $\vec{e}_{x'} \cdot \vec{e}_{z'} = 0$, etc. You will test out this expectation on an upcoming problem set.

We wrap up our discussion of the metric with a few comments:

- Writing out that matrix over and over is tedious and tiring. As shorthand, we will often write $\mathrm{diag}(-1, 1, 1, 1)$ rather than the full 4×4 matrix. This notation means "the matrix which has -1, 1, 1, 1 on the diagonal, and has zero everywhere else."

- For reasons that will be clearer in the next section, it is useful to define an *inverse* metric: we define $\eta^{\alpha\beta}$ by the rule that

$$\eta^{\alpha\beta} \eta_{\beta\gamma} = \delta^{\alpha}{}_{\gamma} \ . \tag{9.12}$$

  Recall that $\delta^{\alpha}{}_{\gamma}$ is known as the Kronecker delta. It equals 1 if $\alpha = \gamma$, and equals 0 otherwise. Equivalently, we can say $\delta^{\alpha}{}_{\gamma} \doteq \mathrm{diag}(1, 1, 1, 1)$; equivalently, we can say that the Kronecker delta is represented by the elements of the identity matrix. The matrix representation of the components $\eta^{\alpha\beta}$ is exactly the same as the matrix representation of the components $\eta_{\alpha\beta}$ — both are given by $\mathrm{diag}(-1, 1, 1, 1)$.

- The metric is not always going to be as simple as $\mathrm{diag}(-1, 1, 1, 1)$. The metric becomes more complicated when we start using different coordinate systems; and, it becomes *significantly* more complicated when we move from special relativity to general relativity. In these cases, the components of the metric become functions of the coordinates. We will denote the metric by $g_{\alpha\beta}$ when it becomes necessary for us to make it more complicated; we will always use $\eta_{\alpha\beta}$ for the metric that is represented by the matrix $\mathrm{diag}(-1, 1, 1, 1)$. This is the form that we use in special relativity with Cartesian spatial coordinates. (It is worth noting that such coordinates are often called *inertial* coordinates: an observer at constant Cartesian spatial coordinate moves with constant velocity in all Lorentz frames.)

- Finally, the word "metric" comes from a root that means to *measure*. This is because by using the metric we can write the invariant interval $ds^2 = \eta_{\alpha\beta} dx^\alpha \, dx^\beta$ — the metric is the mathematical object which introduces a notion of measurable distance between two events, one located at x^α, the other at $x^\alpha + dx^\alpha$. This may seem fairly trivial given what we have discussed so far, but it becomes substantially less trivial when we move into more complicated geometries. In those cases, we will write $ds^2 = g_{\alpha\beta} dx^\alpha \, dx^\beta$. The behavior of $g_{\alpha\beta}$ is very important for understanding the distance between two coordinate points in these more complicated cases.

9.2 Lowering and raising indices

When we compute $\vec{a} \cdot \vec{b} = \eta_{\alpha\beta} a^\alpha b^\beta$, we say that we are *contracting* the metric with $\vec{a}$ and $\vec{b}$ on the indices α and β. What do we get if we contract the metric with a single vector, on only one index? In other words, what is $\eta_{\alpha\beta} a^\alpha$?

As is the way in mathematics, when we encounter a construction like this, we use it to define something new. In this case, we define a quantity with an index in the "downstairs" position:

$$a_\beta \equiv \eta_{\alpha\beta} a^\alpha \ . \tag{9.13}$$

For reasons that are hopefully obvious, this operation is called *lowering* the index. The components in the "downstairs" position are sometimes called *dual* to the components with index "upstairs"; the geometric object we make using the downstairs-indexed components is (as noted at the end of Lecture 5) known as a "dual vector" or as a "1-form." Neither of these names will be important for the purposes of 8.033, though I may occasionally use these terms.

In special relativity using inertial coordinates, lowering the index flips the sign of the zero component: $a_0 = -a^0$, but $a_1 = a^1$, $a_2 = a^2$, $a_3 = a^3$. Lowering the index gives us another way to construct the inner product:

$$\vec{a} \cdot \vec{b} = a_\alpha b^\alpha = a^\alpha b_\alpha \ . \tag{9.14}$$

If the metric lowers an index, then it is hopefully not too surprising that the inverse metric raises it:

$$\eta^{\alpha\beta} a_\alpha = \eta^{\alpha\beta} \left(\eta_{\alpha\mu} a^\mu \right) = \left(\eta^{\alpha\beta} \eta_{\alpha\mu} \right) a^\mu = \delta^\beta{}_\mu a^\mu = a^\beta \ . \tag{9.15}$$

This is why the inverse metric was introduced — it gives us a tool to reverse the lowering operation which the metric performs.

How do the components a_α transform between reference frames? You can probably guess based on the "line up the indices" rule, but to be sure, let's carefully compute how components in the frame of $\mathcal{O}'$ are related to components in the frame of $\mathcal{O}$:

$$\begin{aligned}
a_{\alpha'} &= \eta_{\alpha'\beta'} a^{\beta'} \\
&= \left(\Lambda^\mu{}_{\alpha'} \Lambda^\nu{}_{\beta'} \eta_{\mu\nu} \right) \left(\Lambda^{\beta'}{}_\sigma a^\sigma \right) \\
&= \left(\Lambda^\mu{}_{\alpha'} \Lambda^\nu{}_{\beta'} \Lambda^{\beta'}{}_\sigma \right) \eta_{\mu\nu} a^\sigma \\
&= \Lambda^\mu{}_{\alpha'} \delta^\nu{}_\sigma \eta_{\mu\nu} a^\sigma \\
&= \Lambda^\mu{}_{\alpha'} \eta_{\mu\nu} a^\nu \\
&= \Lambda^\mu{}_{\alpha'} a_\mu \ .
\end{aligned} \tag{9.16}$$

On the first line, we write the lowering operation with all components expressed in the frame of $\mathcal{O}'$. On the second line, we introduce the Lorentz transformation matrices that express those $\mathcal{O}'$-frame quantities in terms of $\mathcal{O}$-frame quantities. On the third line, we rearrange the terms slightly, and then sum over the index β'. This yields the Kronecker delta by combining the second and third Lorentz transformation matrices. To go to the fifth line, we sum over the index σ, which (thanks to the properties of the Kronecker delta) changes the a^σ to a^ν. The result of this tells us to lower the index on a^ν. The result we get at the end of all this shows us that to transform "downstairs" components, we indeed just "line up the indices."

As mentioned in a previous lecture, "upstairs" components are often called contravariant, and "downstairs" ones are called covariant. We now see that the metric and inverse metric are the tools we use to flip between the two forms. This holds up in general, including when the metric becomes more complicated than $\mathrm{diag}(-1, 1, 1, 1)$. Because the metric (and its inverse) let us raise or lower indices as needed for our calculation, the difference between the "upstairs" and "downstairs" position is not really that important for us. This is one of the reasons I like using the terms "upstairs" and "downstairs" — these terms emphasizes that the index position is not terribly important, and in fact can be modified with ease.

The 4-vectors we have discussed so far (spacetime displacement, 4-momentum, 4-velocity) are most "naturally" presented with their indices up. This is largely because they descend from the spacetime displacement vector, $\Delta\vec{x} = \Delta x^\alpha \vec{e}_\alpha$, in which the physical quantity we care about is the set of coordinate displacements Δx^α. There are some quantities which are most "naturally" expressed using indices down. The prototypical example of this is the spacetime *gradient*. Suppose that $\phi(\vec{x})$ is a scalar field — that is, it is a field whose value at the event located $\vec{x}$ away from the origin is the same to all inertial observers. Then we define its gradient by

$$A_\alpha = \frac{\partial \phi}{\partial x^\alpha} \equiv \partial_\alpha \phi \; . \tag{9.17}$$

On a problem set, you will show that if $x^{\mu'} = \Lambda^{\mu'}{}_\alpha x^\alpha$, then $A_{\alpha'} = \Lambda^\mu{}_{\alpha'} A_\mu$ — under Lorentz transformations, the gradient behaves like a "downstairs" index quantity.

The metric lets us define a variation on the gradient. Let us define

$$x_\alpha = \eta_{\alpha\mu} x^\mu \; . \tag{9.18}$$

The components of this "downstairs" variant of x^μ are identical, except for the time-like piece, which picks up a minus sign:

$$x_0 = -x^0 = -ct \; ; \qquad x_{1,2,3} = x^{1,2,3} \; . \tag{9.19}$$

We define our variant of the gradient using derivatives with respect to x_α:

$$A^\alpha = \frac{\partial \phi}{\partial x_\alpha} \equiv \partial^\alpha \phi \; . \tag{9.20}$$

It's not hard to show that this transforms like an "upstairs" index quantity, hence our association of it with A^α.

One of the places where this is really useful is that we can combine and contract the two notions of gradient to produce a combination of second derivatives that is a Lorentz invariant operator. Let's look at what happens when we act both notions of gradient with the indices contracted onto scalar field ϕ:

$$\partial^\alpha \partial_\alpha \phi = -\frac{1}{c^2}\frac{\partial^2 \phi}{\partial t^2} + \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} \equiv \Box\phi \; . \tag{9.21}$$

You may recognize this combination of derivatives as exactly what we have for quantities that obey a wave equation. Indeed, the combination $\partial^\alpha \partial_\alpha$, often denoted with the "box" symbol $\Box$, is called the *wave operator*. Notice that it has no free indices.

9.3 Tensors

The metric is an example of a family of mathematical objects called *tensors* which are used in many places in physics. They are particularly important in relativity, but show up in many other fields, particularly when one studies the properties of matter or matter flow over an extended region (e.g., in fluid dynamics, or the elastodynamical properties of materials).

Tensors are geometric objects whose components are represented by quantities with indices on them. The metric tensor is the first example we have seen of a tensor with two indices, but this generalizes — tensors can have an arbitrary number[1] of indices. Their defining characteristic is the transformation law: a quantity is a tensor if it transforms with a transformation matrix "correcting" each of its indices. For example, suppose physics tells us that we care about a quantity with 4 indices, one in the up position and three down: $R^\mu{}_{\alpha\beta\gamma}$. This quantity is a tensor if it transforms between reference frames with the rule

$$R^{\mu'}{}_{\alpha'\beta'\gamma'} = R^\mu{}_{\alpha\beta\gamma}\Lambda^{\mu'}{}_\mu\Lambda^\alpha{}_{\alpha'}\Lambda^\beta{}_{\beta'}\Lambda^\gamma{}_{\gamma'}\ . \tag{9.22}$$

The number of indices used for a tensor's components (and hence the number of transformation matrices used to transform it) tells us the tensor's *rank*. The example (9.22) is a rank-4 tensor. The metric is a rank-2 tensor. Vectors are rank-1 tensors; they transform using one transformation matrix. Scalars — Lorentz invariants — are often considered to be rank-0 tensors; they transform with *no* transformation matrices, since they are the same in all frames. The wave operator we defined in the previous section acts like a scalar — more properly, a "scalar operator," because it defines a combination of derivatives that operate in the same way in all frames.

In 8.033, we will work almost entirely with tensors of rank 0, 1, and 2. (We will briefly discuss higher rank tensors when we move from special relativity to general relativity, but the discussion will be almost entirely qualitative.) Rank-2 tensors are sufficiently important that they are worth some detailed discussion. Many rank-2 tensors can be regarded as a quantity that, in essence, points in two directions at once. For example, in a few lectures we will discuss a quantity called the "stress-energy tensor" which describes the flux of 4-momentum. Components $T^{\alpha\beta}$ of this tensor describe the flux of 4-momentum component p^α in the x^β direction.

In general, rank-2 tensors in spacetime have 16 components — 4 for each index. However, many rank-2 tensors have symmetry properties that allows us to relate some of the components to each other:

- A tensor $S^{\alpha\beta}$ is *symmetric* if it has the property that $S^{\alpha\beta} = S^{\beta\alpha}$. This reduces the number of independent components from 16 to 10: the four components on the diagonal, plus half of the 12 off-diagonal components. The stress-energy tensor mentioned above has this property; so does the metric, even in the general form $g_{\alpha\beta}$.

- A tensor $A^{\alpha\beta}$ is *antisymmetric* if it has the property that $A^{\alpha\beta} = -A^{\beta\alpha}$. This reduces the number of independent components from 16 to 6. The four components on the diagonal must be zero (this is the only solution to $A^{\alpha\beta} = -A^{\beta\alpha}$ if $\alpha = \beta$), and only half of the 12 off-diagonal components are unique. We will soon find an antisymmetric tensor that allows us to describe electric and magnetic fields in a covariant formulation.

[1] In my research, I use a tensor with 4 indices more or less daily, and have done work that involves tensors with 5 and 6 indices.

9.4 Aside: Using matrix multiplication to combine tensors and matrices

Once we start working with rank-2 tensors, there is a class of mistakes that 8.033 instructors tend to encounter from students who use their knowledge of linear algebra to work through equations that involve products of tensors. Let me emphasize very strongly that using standard linear algebra tools can be done to correctly reduce equations of the sort that we will develop. However, doing so requires that you be careful to think how to combine the different tensors.

Suppose you need to construct a tensor $A^{\alpha\beta}$ which is given by combining three tensors. For instance, suppose that

$$A^{\alpha\beta} = B_{\mu\nu} D^{\alpha\mu} F^{\beta\nu} \; . \tag{9.23}$$

By far the most common mistake we see is that people write this as the following (wrong!) equation:

$$\mathsf{A}_{\mathrm{WRONG}} = \mathsf{B} \cdot \mathsf{D} \cdot \mathsf{F} \; , \tag{9.24}$$

where

$$\mathsf{A}_{\mathrm{WRONG}} = \begin{pmatrix} A^{00} & A^{01} & A^{02} & A^{03} \\ A^{10} & A^{11} & A^{12} & A^{13} \\ A^{20} & A^{21} & A^{22} & A^{23} \\ A^{30} & A^{31} & A^{32} & A^{33} \end{pmatrix}_{\mathrm{WRONG}} \; , \qquad \mathsf{B} = \begin{pmatrix} B_{00} & B_{01} & B_{02} & B_{03} \\ B_{10} & B_{11} & B_{12} & B_{13} \\ B_{20} & B_{21} & B_{22} & B_{23} \\ B_{30} & B_{31} & B_{32} & B_{33} \end{pmatrix} \; , \tag{9.25}$$

with D and F defined similarly.

Why is this wrong? When we represent a rank-2 tensor by a matrix, the first index corresponds to the row, the second index to the column. We need to make sure that when we contract on indices, we are correctly linking up rows and columns of the different objects.

With this in mind, let's carefully examine Eq. (9.23). To produce $A^{\alpha\beta}$, we first contract $B_{\mu\nu}$ on its first index with the second index of $D^{\alpha\mu}$. In matrix form, this means we select column ν of B, we select row α of D, and we combine:

$$B_{\mu\nu} D^{\alpha\mu} \mapsto \mathsf{D} \cdot \mathsf{B} \; . \tag{9.26}$$

We thus see a big error in Eq. (9.24): the order of multiplying the matrices D and B has been reversed. Matrix multiplication does not commute, so this is a highly nontrivial error.

We also need to contract the second index of $B_{\mu\nu}$ with the *second* index of $F^{\beta\nu}$. In other words, when we put things in matrix form, we select row μ of B and combine it with row β of F. In the language of matrix multiplication, this means we are multiplying B with the *transpose* of F. The correct translation of Eq. (9.23) into matrix form is thus

$$\mathsf{A} = \mathsf{D} \cdot \mathsf{B} \cdot \mathsf{F}^T \; , \tag{9.27}$$

where the T superscript denotes matrix transpose. We see that the wrong response is wrong in two ways: it puts the matrices in the wrong order, and it uses F rather than its transpose F^T. In some cases, failing to use the transpose may be harmless because the underlying matrix is symmetric. If so, the matrix and its transpose are identical, and you've gotten lucky! You cannot count on such luck working out for you in general. Indeed, if the matrix is in fact antisymmetric, by not taking the transpose you'll wind up with a minus sign that could drive you somewhat mad.

Carefully following the logic described here to combine rank-2 tensors via matrix multiplication will work. However, it must be emphasized that simply working with the index format *always* just works. You don't need to do any of this careful vetting of which index is combining with which, and writing out the matrices accordingly.

It must also be emphasized that this way of mapping index equations into linear algebra becomes more or less impossible to use once we move beyond rank-2 tensors. For instance, when I originally drafted these notes, a large portion of my working thoughts were consumed by a research paper with a (then) graduate student[2] that is largely concerned with finding solutions to the equation

$$\frac{Dp^\mu}{d\tau} = -\frac{1}{2} R^\mu{}_{\alpha\beta\gamma} u^\alpha S^{\beta\gamma} \; . \tag{9.28}$$

This equation tells us how a body's momentum changes as it moves through spacetime if the body's 4-velocity has components u^α, and the body is itself spinning (the tensor components $S^{\beta\gamma}$ describe its spin in relativistic language). The operator $D/d\tau$ is a special kind of derivative taken with respect to proper time along that body's worldline through spacetime, and the tensor components $R^\mu{}_{\alpha\beta\gamma}$ describe the action of gravitational tides in general relativity. There is really no way we can put an equation like this into a form that is matrix-like. Instead, we just run through the indices and combine everything by straightforward multiplication and summation of the quantities written out index by index. Using computer algebra tools, this isn't so bad, as long as everything is set up and defined carefully.

[2] https://arxiv.org/abs/2201.13334

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 10
Describing matter in bulk

10.1 A box full of dust

In our discussion so far, we have discussed how to analyze the kinematics of *particles* — pointlike entities with velocity and mass, momentum and energy. The focus on particles is an important step to making our laws of physics comport with the principle of Lorentz covariance. However, a lot of the matter that we study in physics isn't in a form that we study particle by particle, but instead is distributed in bulk over some volume. Various aspects of the properties of this bulk matter vary according to the reference frame in which it is observed. In today's lecture, we will introduce tools that are used to characterize bulk matter, and will examine what properties of the characterization change as we change frames.

Begin by considering a box full of *dust*. "Dust" is how we describe matter that doesn't interact with itself — it doesn't exert pressure or do anything interesting other than take up space. Think of it as a pile of particles with mass, but no other interesting property[1].

We begin with the simplest way to characterize this matter: we take the box to be at rest with respect to us, and we count the number of dust particles it contains. We find that the box contains N particles, and that the box has a volume of V. Then, we say that the dust has a *number density*

$$n_0 = N/V \; . \tag{10.1}$$

The number n_0 characterizes perhaps the most important characteristic of the dust, given what we know about it so far. (The reason for the "0" subscript will be made clear in a moment.) Note the dimensions of n_0: number per unit volume, or number per length cubed.

Now take the box full of dust to be, as we observe it, in motion. What is different from the rest frame view? What is the same?

The total number of dust particles must be the same — simply making the box move cannot create or destroy any of the dust. So the number of particles N is independent of the frame in which we measure it. But, one of the linear dimensions of the box is contracted by a factor γ. This reduces the volume of the box by a factor γ according to our measurements, which in turn means that the number density must increase by a factor γ:

$$n = \gamma N/V = \gamma n_0 \; . \tag{10.2}$$

We will use n to stand for the number density that we measure in our frame of reference. This reduces to n_0 if our frame of reference happens to be the dust's own rest frame.

When we observe the dust to move, it acquires one other property: some volumes which were empty of dust at time t will contain dust a time $t + \Delta t$ later; other volumes that

[1]Such "dust" doesn't really exist — any dust that we encounter in reality is more interesting than the dust we use in this lecture. Our dust is an idealization that we use to formulate the framework that we are working in, and serves as a useful starting point. Once we've developed a framework for this more-or-less fictional idealization, we can add more features and properties, pushing it toward something realistic.

contained dust will lose it. This is because the dust is now flowing: there is a *flux* of dust. Suppose that we measure the volume[2] to have length L. Let's orient our coordinate axes so that the box is moving in the x direction, and we can write $\mathbf{v} = v\mathbf{e}_x$. At time $t = 0$, the back of the box is at $x = 0$, and the front of the box is at $x = L$. The cross section of the box has area A (so that $V = AL$).

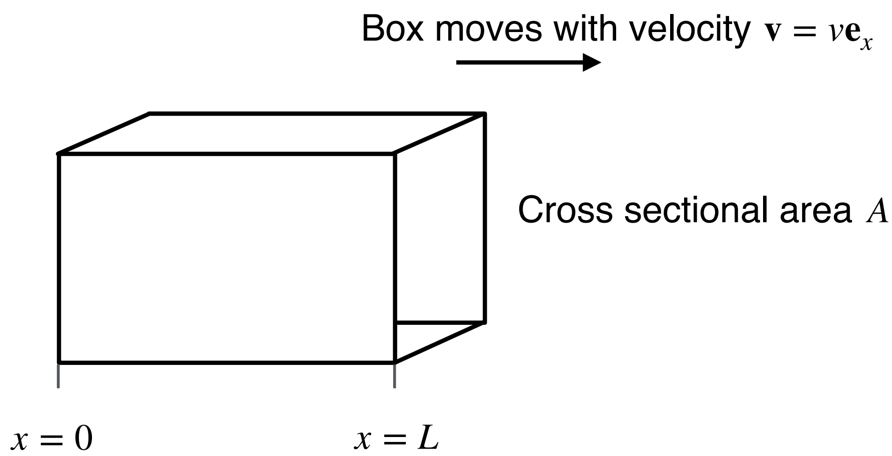


Figure 1: Box as described in the text, at time $t = 0$.

At time $t = \Delta t$, the volume from $x = 0$ to $x = v\Delta t$ has been emptied of dust; the volume from $x = L$ to $x = L + v\Delta t$ has filled up with dust. The box is gaining $nAv\Delta t$ dust particles at the front end, and losing $nAv\Delta t$ dust particles at the back end. Dividing by $A\Delta t$, we the rate at which dust is entering one end per unit cross section area is

$$\frac{dN}{dA\,dt} = nv\ . \tag{10.3}$$

The same rate is leaving the box at the back end.

Equation (10.3) defines a *flux* of particle number into and out of the box. Let's make this a bit more general: we define the x component of the *number flux 3-vector* by

$$n^x = nv = \gamma n_0 v\ . \tag{10.4}$$

You should be able to convince yourself that there was no reason to restrict ourselves to dust moving in the x direction, and we can define a general number flux 3-vector as

$$\mathbf{n} = n\mathbf{v} = \gamma n_0 \mathbf{v}\ . \tag{10.5}$$

The number flux 3-vector $\mathbf{n}$ tells us the number of dust particles per unit area that crosses into (or out of) a region per unit time.

[2]Bear in mind that this means L is *not* the rest frame length of the box

Let's think about the flow of dust into or out of a region a little more carefully. Imagine that dust is flowing through our frame of reference, and that at each point in space it has a number density n and a number flux 3-vector $\mathbf{n}$. Imagine that both of these quantities can vary as a function of position and time: the flow of dust may bend and twist as it flows, with the amount in the flow rising and falling with time.

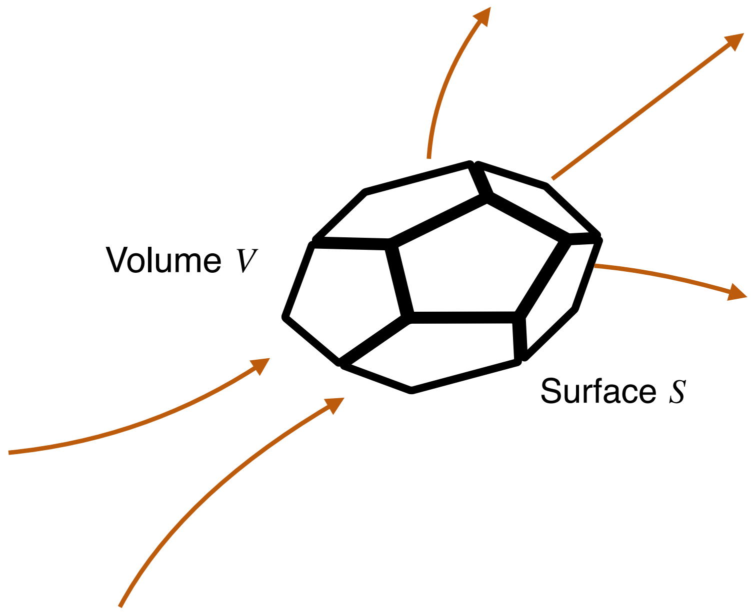


Figure 2: Dust with number flux 3-vector $\mathbf{n}$ flows into and out of a volume V which is bounded by a surface S.

Imagine that this "river" of dust flows into a volume V which is bounded by a surface S. In a time Δt, the change in the number of dust particles in the volume is given by

$$\begin{aligned} \Delta N &= -\Delta t \oint_S \mathbf{n} \cdot d\mathbf{A} \\ &= \int_V [n(t + \Delta t) - n(t)] \, dV \, . \end{aligned} \tag{10.6}$$

Let's deconstruct Eq. (10.6). On the first line, we have introduced and are using the outward directed area element $d\mathbf{A}$. This is a differential of area to which we assign a direction: It points in the out direction, normal (orthogonal) to the surface. The minus sign is because the area element is outward pointing: if $\mathbf{n} \cdot d\mathbf{A} < 0$, then dust is flowing into the volume and ΔN is positive; vice versa if $\mathbf{n} \cdot d\mathbf{A} > 0$.

To write down the second line, note that ΔN is the change in total number contained by the volume. We get this total number by integrating the number density over the volume V; its change is given by subtracting the amount that was there at time t from the amount that is there a time Δt later.

Next, divide both sides by Δt. We can take Δt inside the integral, yielding

$$\int_V \frac{[n(t + \Delta t) - n(t)]}{\Delta t} \, dV = -\oint_S \mathbf{n} \cdot d\mathbf{A} \, . \tag{10.7}$$

Taking the limit $\Delta t \to 0$,

$$\int_V \frac{\partial n}{\partial t}\, dV = -\oint_S \mathbf{n} \cdot d\mathbf{A} \,. \tag{10.8}$$

For the next step, we invoke the *divergence theorem*: for any 3-vector $\mathbf{F}$ defined over a region V that has a closed surface S,

$$\oint_S \mathbf{F} \cdot d\mathbf{A} = \int_V (\nabla \cdot \mathbf{F})\, dV \,. \tag{10.9}$$

Applying the divergence theorem on the right-hand side of Eq. (10.8) and then moving it to the left, we have

$$\int_V \left[\frac{\partial n}{\partial t} + \nabla \cdot \mathbf{n}\right] dV = 0 \,. \tag{10.10}$$

This equation must hold no matter what V we use. The only way for that to be the case is if the term in square brackets in Eq. (10.10) vanishes. This means that the number density n and the number flux $\mathbf{n}$ are related by the *continuity equation*

$$\frac{\partial n}{\partial t} + \nabla \cdot \mathbf{n} = 0 \,, \tag{10.11}$$

or, expanding out the components in the divergence term,

$$\frac{\partial n}{\partial t} + \frac{\partial n^x}{\partial x} + \frac{\partial n^y}{\partial y} + \frac{\partial n^z}{\partial z} = 0 \,. \tag{10.12}$$

Everything we have done can be organized into a particularly tidy package using 4-vectors. First, note that Eqs. (10.2) and (10.5) have *almost* exactly the form of the components of a 4-vector: we treat (10.2) as the timelike component, and then (10.5) defines the spatial components. The only reason this doesn't quite work is that (10.2) has the wrong dimensions: it is number per unit volume, whereas the components in (10.5) have the dimensions number per unit area per unit time.

This is easily fixed: just multiply (10.2) by the speed of light c. Doing so, we define the *number flux* 4-vector $\vec{N}$, whose components are

$$\begin{aligned}(N^0, N^1, N^2, N^3) &= (nc, nv^1, nv^2, nv^3) \\ &= (\gamma n_0 c, \gamma n_0 v^1, \gamma n_0 v^2, \gamma n_0 v^3) \,. \end{aligned} \tag{10.13}$$

Notice that this is nothing more than

$$\vec{N} = n_0 \vec{u} \,, \tag{10.14}$$

where $\vec{u}$ is the 4-velocity with which we observe the dust to be moving. Let's look at the invariant we can build out of $\vec{N}$:

$$\vec{N} \cdot \vec{N} = n_0^2 \vec{u} \cdot \vec{u} = -n_0^2 c^2 \,. \tag{10.15}$$

This tells us that the number flux 4-vector is timelike. Taking the scalar product of $\vec{N}$ with itself yields the number density of the dust in its own rest frame, times $-c^2$.

The 4-vector $\vec{N}$ also allows us to write the continuity equation in a particularly tidy way. Recalling that $x^0 = ct$, we see that Eq. (10.12) can be written

$$\frac{\partial N^\alpha}{\partial x^\alpha} = 0 \;, \tag{10.16}$$

or

$$\partial_\alpha N^\alpha = 0 \;. \tag{10.17}$$

Notice that there are no free indices left over: we sum over α, with one in the upstairs position and one downstairs, yielding a Lorentz invariant quantity (in the case, the number 0 — certainly a quantity that all Lorentz observers agree on). By setting everything up using 4-vectors, we have a *covariant* formulation of the continuity equation. If we have measured the 4-components of $\vec{N}$ in the frame of $\mathcal{O}$, and would like to know how they will appear in the frame of $\mathcal{O}'$, we simply apply a Lorentz transformation: $N^{\mu'} = \Lambda^{\mu'}{}_\alpha N^\alpha$. This quantity will obey the continuity equation provided we take derivatives using the coordinates $x^{\mu'}$ which are used by $\mathcal{O}'$: they will find $\partial_{\mu'} N^{\mu'} = 0$.

10.2 A box full of charge

This discussion of number continuity may have reminded you of a calculation that you did in electricity and magnetism. Suppose each grain of dust carries an electric charge q. Then, our calculation proceeds essentially exactly as before, but we can now look at the *charge density* associated with a volume, and we can think about a charge flux 3-vector, better known as the *current density*. Let's quickly see what this looks like.

If the number density of the dust in some frame is n, and if each dust grain carries a charge q, then the charge density ρ_q is given by

$$\rho_q = nq \;. \tag{10.18}$$

(We will use ρ for something different in a moment, hence the q subscript to denote charge density.) If these dust grains have a number flux 3-vector $\mathbf{n}$, then the flow of the dust carries a current density

$$\mathbf{J} = q\mathbf{n} = \rho_q \mathbf{v} \;. \tag{10.19}$$

Going through the derivation of number continuity again, but now including a charge q on each dust grain, yields the continuity equation for electric charge:

$$\frac{\partial \rho_q}{\partial t} + \nabla \cdot \mathbf{J} = 0 \;. \tag{10.20}$$

We can build a 4-vector out of this by defining its "zeroth" component using the charge density and the speed of light. We thus define $\vec{J}$ with components

$$(J^0, J^1, J^2, J^3) = (\rho_q c, J^1, J^2, J^3) \;. \tag{10.21}$$

With this formulation, we can write the equation of charge continuity as

$$\partial_\alpha J^\alpha = 0 \;. \tag{10.22}$$

We will return to this 4-vector shortly when we examine how to write the equations of electrodynamics in a way that makes their Lorentz covariance clear.

10.3 A box full of dust, revisited

Finally, let's give each dust grain a rest mass m. We could define a rest mass density $\rho_m = Nm/V$. However, we know that as we change frames, the most interesting quantities which describe a massive object are its energy γmc^2 and its momentum $\mathbf{p} = \gamma m\mathbf{v}$. So let's instead define the rest frame's *energy density* $\rho_0 = Nmc^2/V$. How does this quantity transform when we change frames?

Again, we know that we cannot create or destroy any dust grains, so N is the same in all frames. We also know that the length of the box along the relative motion of the frames is contracted by γ, so $V \to V/\gamma$. However, in this case, we also know that the energy of each dust grain is boosted by γ: the grain only had rest energy in the original rest frame, but it has both rest energy and kinetic energy in a frame moving with $\mathbf{v}$ relative to the rest frame. The energy density in this frame is given by

$$\rho = N(\gamma mc^2)/(V/\gamma) = \gamma^2\rho_0 \; . \tag{10.23}$$

The fact that *two* powers of γ enter into this transformation law is interesting and important. When we carefully studied number density and charge density, we realized that these quantities were actually components of a 4-vector. If they had been Lorentz scalars, then they would have been invariants; the transformation would have involved no factors of γ. The number of dust grains in a box, or the total charge in a box, both fall into this category. When there is one factor of γ, that tells us that that we have stumbled onto a transformation law that involves one factor of the Lorentz transformation matrix $\Lambda^{\mu'}{}_{\alpha}$, and so the quantity we are looking at is a component of a rank-1 tensor — i.e., a 4-vector.

This factor of γ^2 tells us that the quantity we are examining is associated with a transformation law that involves *two* factors of the Lorentz transformation matrix. The quantity we are studying must a component of a rank-2 tensor — a quantity with two associated indices. Let us define

$$T^{\alpha\beta} = \frac{Nm}{V}u^\alpha u^\beta \qquad \text{or} \tag{10.24}$$

$$= p^\alpha N^\beta \; . \tag{10.25}$$

This quantity is known as the *stress-energy tensor*. The 00 or tt component describes energy density in some reference frame. To understand the other components, note the interpretation that Eq. (10.25) suggests: $T^{\alpha\beta}$ describes the flux of 4-momentum p^α in the direction of x^β. (Via Eq. (10.24), we see that $T^{\alpha\beta} = T^{\beta\alpha}$, so we can equally well call this the flux of 4-momentum p^β in the direction of x^α.) What do the other components mean? Let's go through these tensor looking at a couple of important groupings of components for this dust stress energy:

- $T^{00} = \gamma^2 n_0 mc^2$: As already discussed, this is *energy density*. Think of it as the density of p^0 flowing in the direction of x^0 — the flux of energy density through time.

- $T^{0i} = \gamma^2 n_0 mcv^i$: This is *energy flux*: the flow of the density of p^0 in the x^i direction. If you look carefully at the units, you'll see that this quantity is off a by factor with the units of velocity. More correctly, the energy flux is $T^{0i}c$. The root issue here is that 4-momentum component p^0 is E/c, so we need to correct with a factor of c. Correction factors like this don't change the essential physics.

- $T^{i0} = \gamma^2 n_0 m c v^i$: This is *momentum density.* Think of this as the density of momentum p^i flowing through time. Again, examining units, you'll see it's a bit off. More correctly, the momentum density is T^{i0}/c; the root issue here is that x^0 is c times t. (Needing to account for factors like this is one reason why many people use units in which $c = 1$. Keeping track of factors of c can become tiresome.) Notice that $T^{i0} = T^{0i}$ — using the relativistic definitions of energy and momentum, energy flux and momentum density are the same thing, modulo factors of c.

- $T^{ij} = \gamma^2 n_0 m v^i v^j$: This is *momentum flux*: the flow of momentum p^i in the x^j direction.

10.4 The stress-energy tensor more generally

Dust is a useful tool for introducing the stress-energy tensor and wrapping our heads around what the components of this tensor mean to a particular observer. However, dust is a somewhat limited class of matter. The stress-energy tensor is much broader than this. We conclude today's lecture by discussing the meaning of the stress-energy tensor as it is used to describe matter in general and, as we'll briefly discuss later, fields.

One often characterizes the stress-energy tensor by going into a frame of reference in which there is no bulk flow of material. For examine, if it is a fluid, this is the frame in which the fluid is a rest; such a frame is called "comoving" in this case. Note that a distribution of material might flow at different speeds in different places or at different times; think of this as how we characterize one small "element" of the material. In this frame, the different components take on exactly the meaning that we discussed for the components of the dust stress-energy tensor:

- T^{00} represents the energy density of the material.

- T^{0i} represents (modulo a factor of c) the energy flux of the material. Note that if no matter is actually moving, there still might be a flow of energy — the material might be conducting heat, or there may be radiation flowing in some direction.

- T^{i0} represents (modulo a factor of c) the momentum density of the material. Again, even if no matter is actually moving there can still be a density of momentum. Indeed, there *must be* momentum density if there is any flux of energy.

- T^{ij} represents the momentum flux. This 3×3 spatial tensor is important in its own right, and is known as the *stress tensor*. The on-diagonal and off-diagonal elements of the stress tensor deserve comment:

  - The on-diagonal elements (T^{xx}, T^{yy}, T^{zz}) tell us about the flow of momentum component p^i in the x^i direction. These components of the stress tensor tell us about the force (per unit area) the material exerts in the direction of its flow. When the material is a fluid, these components of the tensor describe *pressure*.
  - The off-diagonal elements (T^{xy}, T^{xz}, T^{yz}, plus symmetries) tell us about "non-normal" flows of momentum. In fluids, these terms are related to a property called its *viscosity*; it leads to forces along (i.e., parallel to) an interface, rather than normal to the interface (the way pressure operates).

An example of a material which is used in many analyses is a *perfect fluid*. It is a fluid for which there exists a frame of reference in which its stress-energy tensor has components $T^{\alpha\beta} = \mathrm{diag}(\rho, P, P, P)$, where ρ is the fluid's energy density, and P is its pressure.

The "perfect" in "perfect fluid" means that it represents a kind of Platonic ideal: there is no energy or momentum flow in a perfect fluid's rest frame (meaning that there is no heat conduction, or other mechanism to transport energy), and it has no viscosity. No viscosity means that if you were to dip your hand into it, none of the fluid would stick to you when you pulled your hand out. As such, the physics of perfect fluids has been mocked as the physics of "dry water."

The meaning of stress energy as a flux of 4-momentum allows us to derive a continuity equation for it. Let's reconsider Fig. 2, but rather than thinking about the flow of dust, think about the flow of 4-momentum. We then largely repeat our derivation of number continuity, but replace quantities related to number density with quantities related to 4-momentum density. In particular, let's replace the number density n with the 4-momentum density $T^{\alpha 0}$, and replace the number flux $n^i = nv^i$ with the 4-momentum flux $T^{\alpha i}$.

The total amount of 4-momentum in V is given by integrating $T^{\alpha 0}$ over this volume:

$$[p^\alpha(t)]_V = \frac{1}{c}\int_V T^{\alpha 0}(t)\,dV\ . \tag{10.26}$$

The factor of $1/c$ accounts for the fact that T^{00} is energy density, but p^0 is E/c, plus for the fact that T^{i0} has such a factor built into its definition. The *change* in the 4-momentum in V over an interval of time Δt is thus given by

$$\begin{aligned}\Delta p^\alpha &= \frac{1}{c}\int_V \left[T^{\alpha 0}(t+\Delta t) - T^{\alpha 0}(t)\right] dV \\ &= \Delta t \int_V \frac{\partial T^{\alpha 0}}{\partial x^0} dV\ .\end{aligned} \tag{10.27}$$

We can also account for the change by computing the flux of 4-momentum through the surface S bounding this volume during the time interval Δt:

$$\Delta p^\alpha = -\Delta t \oint_S T^{\alpha i} dA^i\ . \tag{10.28}$$

The minus sign is again because the area element $d\mathbf{A}$ (which has components dA^i) points outward from the volume. The divergence theorem can be used here just like it can be used in other circumstances with which you are familiar. Just think of $T^{\alpha i}$ as four different 3-vectors, one for each value of α:

$$\oint_S T^{\alpha i} dA^i = \int_V \frac{\partial T^{\alpha i}}{\partial x^i} dV\ . \tag{10.29}$$

Putting together our two formulations of Δp^α yields

$$\Delta t \int_V \left[\frac{\partial T^{\alpha 0}}{\partial x^0} + \frac{\partial T^{\alpha i}}{\partial x^i}\right] dV = 0\ . \tag{10.30}$$

This gives a continuity equation for the stress-energy tensor:

$$\partial_\beta T^{\alpha\beta} = 0\ . \tag{10.31}$$

This equation expresses both conservation of energy and conservation of momentum for the material whose stress-energy tensor is $T^{\alpha\beta}$.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 11
A covariant formulation of electromagnetics (part I)

11.1 Electric and magnetic fields and forces: Background

Our pivot from Galileo's relativity to Einstein's relativity began by considering electrodynamics. Let's write out again the critical equations which govern electrodynamics — the Maxwell equations which connect the the fields to their sources, and the Lorentz force law which shows how these fields act on charges:

$$\nabla \cdot \mathbf{E} = \rho/\epsilon_0 \;, \qquad \nabla \cdot \mathbf{B} = 0 \;, \tag{11.1}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \;, \qquad \nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \;; \tag{11.2}$$

$$\mathbf{F} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) \;. \tag{11.3}$$

It should be emphasized very strongly that these equations are *fully* compatible with special relativity. Indeed, all of the modifications to various physical concepts that Einstein's relativity requires were introduced because it became clear that important aspects of Newtonian mechanics were not compatible with electrodynamics. Electrodynamics is one of the most successfully and accurate theories of nature we have developed. Once it has been updated to account for the fact that our universe is quantum mechanical in nature (a topic for a different course!), we end up with a version of electrodynamics that is perhaps humanity's most precisely-tested description of nature.

That said, Eqs. (11.1), (11.2), and (11.3) are not written in a way that makes it clear they are compatible with Lorentz covariance. The fields and the force are written using 3-vectors, which depend upon us choosing a particular observer's "space" coordinates; the field equations are expressed using a particular observer's time and space derivatives. These equations are formulated for one particular reference frame, and it is not obvious how they will transform to another reference frame. The goal of the next two lectures is to think how to organize the structures expressed in Eqs. (11.1), (11.2), and (11.3) in a way that clearly shows electrodynamics is a Lorentz covariant theory.

11.2 How to organize the fields

11.2.1 General considerations

So far, when we've translated a physical quantity into Lorentz covariant language, we have found a way of taking quantities which are 3-vectors and mapping them into 4-vectors. Examples so far are displacement (add ct as the "zeroth" component), the 4-velocity (change d/dt to $d/d\tau$ so that we use a clock whose meaning is invariant to describe time derivatives; add $c\,dt/d\tau = \gamma c$ as the zeroth component), and the 4-momentum (add energy as the zeroth component, dividing by c to make sure the dimensions are sensible). Can we do this with the electric and magnetic fields?

We have several problems here. First, we know that $\mathbf{E}$ and $\mathbf{B}$ fields must transform into one another when we change frames: what is pure magnetic field in one frame is a mixture of magnetic and electric fields in another; and vice versa. The classic example of this is a charge moving in a magnetic field. Consider a charge moving parallel to a current-carrying wire, as illustrated in Fig. 1:

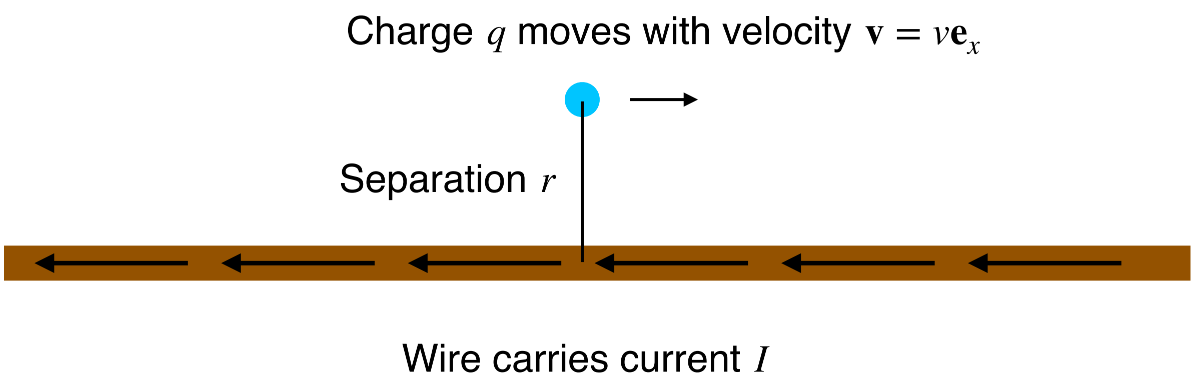


Figure 1: A charge q moving parallel to a wire carrying a current I.

For concreteness, let's define $\mathbf{e}_x$ as pointing to the right, $\mathbf{e}_y$ as pointing into the page, and $\mathbf{e}_z$ as pointing up. Then, in what we will call the "lab" frame L, we have a charge q that moves to the right. The charge is a distance r from a wire that carries a current flowing to the left. As we learned in 8.02/8.022, this wire generates a magnetic field that circulates around the wire. At the location of the charge, this field takes the value

$$\mathbf{B} = \frac{\mu_0 I}{2\pi r}\mathbf{e}_y \ . \tag{11.4}$$

The wire is neutral, so the charge q does not feel any electric force — it only feels a magnetic force, whose value is

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} = \frac{\mu_0 q I v}{2\pi r}\mathbf{e}_z \ . \tag{11.5}$$

This force points "up" in the figure — the charge is repelled from the wire.

Let's now change frames, and think about what must happen. First, we require q to be the same in all reference frames. If changing frames changed the value of charge, the elementary charge would vary for moving charges. Imagine the effect this would cause for a system in which there are members whose charges are equal and opposite, but are moving at different relative speeds. A system which is neutral when its members "sit still" might have net charge when they are in motion! In addition to feeling absurd, the fact is that we have no experimental evidence for anything like this whatsoever. Observations and measurements indicate that a body's charge is unchanged no matter how fast we observe it to move.

So, let's jump into a reference frame that moves with $\mathbf{v} = v\mathbf{e}_x$ — i.e., the frame C in which the charge is at rest. In this frame *there can be no magnetic force*. The magnetic force is proportional to the charge's speed. If the speed is zero, the magnetic force is zero. However, a repulsive force in one frame of reference is *not consistent* with no force in another.

The details of how the force behaves in this frame might differ[1] (perhaps its magnitude will be different), but there still must be an overall repulsive force. If there is no magnetic force, then there must instead be an *electric* force.

This means there must be an electric field in the charge's rest frame, even though there was no such field in the lab frame. Something that we measured to be pure magnetic field transforms to a mixture of electric and magnetic field. Whatever "entity" we will use to describe electric and magnetic fields in special relativity must be able to transform magnetic fields into electric fields (and vice versa).

11.2.2 A covariant representation of the force and fields

Our root issue is essentially one of simple counting. We have had success fitting important physical quantities into 4-vectors so far, but it just isn't going to work for the electric and magnetic field. They have 6 components, and we cannot fit these 6 pieces of information into the 4 components of a 4-vector. We need something bigger.

A simple example of a bigger object is a 2nd-rank tensor, which has 16 components. That's too many; but, we can reduce the number of free components by imposing symmetry. If we use a symmetric tensor, then it has 10 free components — still too many. But an antisymmetric 2nd-rank tensor has 6 free components — exactly what we need.

So let's think how we can fit the 6 components $(E^x,\ E^y,\ E^z)$, $(B^x,\ B^y,\ B^z)$ into an antisymmetric 2nd-rank tensor which we will call $F^{\alpha\beta}$. To guide us, let's deduce how the Lorentz force law, $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, can be written in a fully covariant manner.

First, we "upgrade" the force. We start with $\mathbf{F} = d\mathbf{p}/dt$. Clearly, we will want to take the 3-momentum $\mathbf{p}$ over to the 4-momentum, whose components are p^α. We also need to upgrade the time derivative with one that uses a notion of time that all frames are happy to use as a point of reference. Just as we did in defining the 4-velocity, let's replace d/dt with $d/d\tau$, where τ is the proper time measured by the body which is experiencing the force.

What about the right-hand side, $q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$? This a quantity that is linear in q, linear in the fields, and — if we think about this carefully — linear in the components of the velocity. "Wait," I imagine you protesting, "the $\mathbf{B}$ term is linear in components of velocity, but what about the $\mathbf{E}$ term?" Note that $\mathbf{E}$ and $\mathbf{B}$ have different dimensions: the dimensions of $\mathbf{E}$ are force over charge, but the dimensions of $\mathbf{B}$ are force over speed times charge. When we assemble these quantities into a single tensor, we'll need to account for the difference in dimensions. We often do this by throwing in factors of the speed of light. This suggests that we think about the Lorentz force law as

$$\mathbf{F} = q\left[c\left(\frac{\mathbf{E}}{c}\right) + \mathbf{v} \times \mathbf{B}\right] . \tag{11.6}$$

Bearing in mind that the components of u^α are given approximately by (c, v^x, v^y, v^z) for a body that is not moving very fast relative to us, this suggests that in the Lorentz force law, the electric field is being multiplied by the *timelike* component of the 4-momentum.

Putting all this together, we want the covariant formulation of the Lorentz force to be

$$\frac{dp^\alpha}{d\tau} = qF^{\alpha\beta}u_\beta . \tag{11.7}$$

[1] In a few lectures we will look carefully at forces and accelerations in special relativity; we briefly introduce a handful of important issues a little later in this lecture.

Let's now figure out how to "fill up" the tensor $F^{\alpha\beta}$ so that this is consistent with the Lorentz force law that we learned about in 8.02/8.022 by going through the spatial components, $\alpha = 1, 2, 3$, one by one. (We'll come back to the $\alpha = 0$ component later.) First look at $\alpha = 1$, or $\alpha = x$:

$$\frac{dp^x}{d\tau} = q\left(F^{10}u_0 + F^{12}u_2 + F^{13}u_3\right) \ . \tag{11.8}$$

There is no F^{11} term because of this tensor's antisymmetry — all diagonal elements are zero. Let's further use the fact that $u_0 = -u^0 = -\gamma c$, $u_2 = \gamma(dy/dt)$, and $u_3 = \gamma(dz/dt)$:

$$\frac{dp^x}{d\tau} = \gamma q\left(-cF^{10} + F^{12}\frac{dy}{dt} + F^{13}\frac{dz}{dt}\right) \ . \tag{11.9}$$

Next, divide by sides by γ and use the fact that an interval of time dt measured by clocks in this frame is $\gamma d\tau$:

$$\frac{dp^x}{dt} = q\left(-cF^{10} + F^{12}\frac{dy}{dt} + F^{13}\frac{dz}{dt}\right) \ . \tag{11.10}$$

Compare this to the x component of the Lorentz force law:

$$\frac{dp^x}{dt} = q\left(E^x + B^z\frac{dy}{dt} - B^y\frac{dz}{dt}\right) \ . \tag{11.11}$$

This allows us to read off

$$F^{10} = -E^x/c \ , \qquad F^{12} = B^z \ , \qquad F^{13} = -B^y \ . \tag{11.12}$$

Repeating this exercise for the y and z force components and noting that the tensor is antisymmetric allows us to fill it in entirely:

$$F^{\alpha\beta} \doteq \begin{pmatrix} 0 & E^x/c & E^y/c & E^z/c \\ -E^x/c & 0 & B^z & -B^y \\ -E^y/c & -B^z & 0 & B^x \\ -E^z/c & B^y & -B^x & 0 \end{pmatrix} \ . \tag{11.13}$$

This tensor is often called the *Faraday* tensor. It replaces the 3-vectors which describe electric and magnetic fields according to some particular observer's reference frame with a geometric object whose components can be readily translated to any reference frame; and, it connects to 4-vectors whose components can likewise be readily translated to any reference frame.

11.3 A brief aside on forces and accelerations

In this lecture, we've been talking about a specific force without yet having discussed forces in special relativity in broader terms. We will discuss forces, accelerations, and the properties of accelerated observers in more detail in an upcoming lecture. Certain aspects of this discussion are needed now, so we pause in our discussion of electric and magnetic fields for a brief digression to talk about forces and accelerations.

As we have discussed, a body of mass m moving with 4-velocity $\vec{u}$ has a 4-momentum $\vec{p} = m\vec{u}$. As you have seen in our discussion above, this momentum changes if the body is acted on by a force or, more properly, a 4-force:

$$\vec{F} = \frac{d\vec{p}}{d\tau} \ . \tag{11.14}$$

If the body's mass cannot change, then this leads to the body having a 4-acceleration:

$$\vec{a} = \frac{1}{m}\vec{F} = \frac{d\vec{u}}{d\tau} \; . \tag{11.15}$$

When we discuss 3-velocities **u** and 3-accelerations **a**, these quantities can have largely any value that we want them to have: the value of **u** is essentially an initial condition to our analysis, and the value of **a** is only constrained by the mechanism providing the force **F**.

Not so for the 4-velocity and the 4-acceleration: there is a very interesting and important *constraint* which these two quantities must always satisfy. To see where this comes from, begin with the invariant that we can construct from $\vec{u}$:

$$\vec{u} \cdot \vec{u} = -c^2 \; . \tag{11.16}$$

Take $d/d\tau$ of both sides of this equation:

$$\vec{a} \cdot \vec{u} + \vec{u} \cdot \vec{a} = 0 \; , \tag{11.17}$$

or

$$\vec{a} \cdot \vec{u} = 0 \; . \tag{11.18}$$

The 4-velocity and the 4-acceleration are always "orthogonal" in spacetime. This important constraint has important implications for the nature of any 4-force that you may compute — if at the end of your analysis, you find that $\vec{F} \cdot \vec{u} \neq 0$, you have made a mistake or have overlooked something important.

11.4 Some details of the electromagnetic 4-force

With the above discussion in mind, let's examine the electromagnetic 4-force that we have worked out. Is it the case that $\vec{F} \cdot \vec{u} = 0$? The answer is yes, and we can show this using a little bit of "index gymnastics":

$$\begin{aligned}
\vec{F} \cdot \vec{u} &= qF^{\alpha\beta} u_\beta u_\alpha && (11.19)\\
&= -qF^{\beta\alpha} u_\beta u_\alpha && (11.20)\\
&= -qF^{\beta\alpha} u_\alpha u_\beta && (11.21)\\
&= -qF^{\alpha\beta} u_\beta u_\alpha \; . && (11.22)
\end{aligned}$$

Let's step through these lines of analysis carefully. On the first line, we have have contracted the definition of the electromagnetic 4-force, Eq. (11.7), with the 4-velocity in order to make the inner product. On the second line, we have used the fact that the Faraday tensor is antisymmetric to swap the order of the indices on the tensor, introducing a minus sign. On the third line, we have used the fact that $u_\alpha u_\beta$ is symmetric to swap the order of their indices. On the final line, we have used the fact that α and β are "dummy" indices — they are being summed over, so it doesn't matter how we label them. We can in fact change α for β and β for α, as long as we do this *consistently* throughout the expression.

Now compare the first line with the fourth line. Their right-hand sides are *identical* ... except for a minus sign. This is thus an expression of the form $x = -x$, whose only solution is $x = 0$. We conclude that

$$\vec{F} \cdot \vec{u} = 0 \; , \tag{11.23}$$

So our 4-force indeed is spacetime orthogonal with the 4-velocity — as it should be.

Two remarks on this calculation:

- This is our first encounter with a trick that gets used a lot: whenever you contract all free indices of a totally antisymmetric mathematical object, like $F^{\alpha\beta}$, against a totally *symmetric* mathematical object, like $u_\alpha u_\beta$, the result is zero.

  If this makes you nervous and you want to be totally confident in the result, you can always go through an exercise like the one that I did above. The key point is that by combining symmetric with antisymmetric, we add up terms that are equal and opposite. If you expand out the Einstein summation that I did above, you find that you can combine terms in pairs: $F^{10}u_1u_0 + F^{01}u_0u_1$, $F^{23}u_2u_3 + F^{32}u_3u_2$, etc. The members of each pair will always be equal in magnitude and opposite in sign.

- When a force law is set up properly, it generally works out "automatically" that we find $\vec{F}\cdot\vec{u} = 0$, in much the way that it did for this electromagnetic 4-force. Finding $\vec{F}\cdot\vec{u} = 0$ does not guarantee that your force law is correct, but *not* finding this guarantees that your force law is wrong.

Before moving on to other aspects of the covariant formulation of electric and magnetic fields, let's clean up one last detail. We saw in our calculation above that the $\alpha = 1$, 2, and 3 components of the 4-force correspond perfectly to the x, y, and z components of the Lorentz force. What is the $\alpha = 0$ component? Let's write this out:

$$\begin{aligned}\frac{dp^0}{d\tau} &= qF^{0\beta}u_\beta = q\left(F^{01}u_1 + F^{02}u_2 + F^{03}u_3\right)\\ &= \frac{\gamma q}{c}\left(E^x(\mathbf{u})^x + E^y(\mathbf{u})^y + E^z(\mathbf{u})^z\right)\\ &= \frac{\gamma q}{c}\mathbf{E}\cdot\mathbf{u}\ . \end{aligned} \tag{11.24}$$

Using the fact that $p^0 = E/c$, where E with no indices and no boldface means the energy[2] of the charged body, and using $dt = \gamma d\tau$, this becomes

$$\frac{dE}{dt} = q\mathbf{E}\cdot\mathbf{u}\ . \tag{11.25}$$

This expression tells us about the rate at which *work* is done on the charge by the electric field. If you need a reminder of where this comes from, remember that the differential of work done in moving through a 3-displacement $d\mathbf{r}$ in an $\mathbf{E}$ field is

$$dW = \mathbf{F}\cdot d\mathbf{r} = q\mathbf{E}\cdot d\mathbf{r}\ . \tag{11.26}$$

If the charge does this in a time dt, then

$$\frac{dW}{dt} = q\mathbf{E}\cdot\frac{d\mathbf{r}}{dt}\ , \tag{11.27}$$

in agreement with Eq. (11.25).

[2]The letter "E" is doing double duty here, standing for both energy and electric field. Sometimes people use U for energy in circumstances like this, in order to reduce the likelihood of any confusion.

11.5 Transforming electric and magnetic fields

By fitting the electric and magnetic fields into a rank-2 tensor, it becomes simple to deduce how these fields transform when we change frames. Let observer $\mathcal{O}$ measure fields described by the tensor $F^{\alpha\beta}$; let $\mathcal{O}'$ in a different inertial frame measure fields described by the tensor $F^{\mu'\nu'}$. These are related by converting using Lorentz transformation matrices:

$$F^{\mu'\nu'} = F^{\alpha\beta}\Lambda^{\mu'}{}_{\alpha}\Lambda^{\nu'}{}_{\beta} \; . \tag{11.28}$$

Let's work through this using the Lorentz transformation matrix

$$\Lambda^{\mu'}{}_{\alpha} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \; . \tag{11.29}$$

In other words, we take $\mathcal{O}'$ to be moving with $\mathbf{v} = v\mathbf{e}_x$ relative to $\mathcal{O}$. Let's use this calculate how the components of the Faraday tensor translate between frames. Start by working through the transformation for the $(0'1')$ component:

$$\begin{aligned} F^{0'1'} &= \Lambda^{0'}{}_{0}\Lambda^{1'}{}_{1}F^{01} + \Lambda^{0'}{}_{1}\Lambda^{1'}{}_{0}F^{10} \\ &= \left(\gamma^2 - \gamma^2\beta^2\right) F^{01} \\ &= F^{01} \; . \end{aligned} \tag{11.30}$$

On the first line, we expanded the transformation rule to write out all the non-zero terms that contribute to $F^{0'1'}$. This amounts to all the lambda matrix elements that have $0'$ on the first index, and all the matrix elements that have $1'$ on the first index. A total of 4 such elements exist: $\Lambda^{0'}{}_{0} = \gamma$, $\Lambda^{0'}{}_{1} = -\gamma\beta$, $\Lambda^{1'}{}_{0} = -\gamma\beta$, and $\Lambda^{1'}{}_{1} = \gamma$; all the others ones with $0'$ or $1'$ in the first position are zero. We then used antisymmetry, and then used the fact that $\gamma = 1/\sqrt{1-\beta^2}$ to clean this expression up. Translating back into electric and magnetic field components, this tells us

$$E^{x'} = E^{x} \; . \tag{11.31}$$

Move on to the $(0'2')$ component:

$$\begin{aligned} F^{0'2'} &= \Lambda^{0'}{}_{0}\Lambda^{2'}{}_{2}F^{02} + \Lambda^{0'}{}_{1}\Lambda^{2'}{}_{2}F^{12} \\ &= \gamma F^{02} - \gamma\beta F^{12} \; . \end{aligned} \tag{11.32}$$

We cannot simplify this any further, so we now translate back into electric and magnetic field components:

$$E^{y'} = \gamma(E^y - vB^z) \; . \tag{11.33}$$

Next the $(0'3')$ component:

$$\begin{aligned} F^{0'3'} &= \Lambda^{0'}{}_{0}\Lambda^{3'}{}_{3}F^{03} + \Lambda^{0'}{}_{1}\Lambda^{3'}{}_{3}F^{13} \\ &= \gamma F^{03} - \gamma\beta F^{13} \; . \end{aligned} \tag{11.34}$$

This becomes

$$E^{z'} = \gamma(E^z + vB^y) \; . \tag{11.35}$$

Doing a similar exercise for the components of the Faraday tensor which map to the magnetic fields, we find

$$B^{x'} = B^x \; , \qquad B^{y'} = \gamma(B^y + vE^z/c^2) \; , \qquad B^{z'} = \gamma(B^z - vE^y/c^2) \; . \tag{11.36}$$

By repeating this analysis for frames moving with $\mathbf{v} = v\mathbf{e}_y$ and $\mathbf{v} = v\mathbf{e}_z$, it's not too difficult to work out the general rule for transforming between frames. For *completely general* $\mathbf{v}$, we have

$$\mathbf{E}'_\parallel = \mathbf{E}_\parallel \ , \qquad \mathbf{E}'_\perp = \gamma\left(\mathbf{E}_\perp + \mathbf{v}\times\mathbf{B}_\perp\right) \ ; \tag{11.37}$$

$$\mathbf{B}'_\parallel = \mathbf{B}_\parallel \ , \qquad \mathbf{B}'_\perp = \gamma\left(\mathbf{B}_\perp - \mathbf{v}\times\mathbf{E}_\perp/c^2\right) \ . \tag{11.38}$$

Here, $\mathbf{E}_\parallel$ denotes the component of $\mathbf{E}$ that is parallel $\mathbf{v}$. Let $\mathbf{e}_v \equiv \mathbf{v}/v$ denote the unit vector along the velocity vector; then, $\mathbf{E}_\parallel = (\mathbf{E}\cdot\mathbf{e}_v)\mathbf{e}_v$. The other component, $\mathbf{E}_\perp = \mathbf{E}-\mathbf{E}_\parallel$, denotes the part of $\mathbf{E}$ that is orthogonal to $\mathbf{v}$. The magnetic field vectors $\mathbf{B}_\parallel$ and $\mathbf{B}_\perp$ are defined likewise.

When I first was presented with the transformation laws (11.37) and (11.38), I was utterly baffled. Though I understood the derivation (which I learned from Purcell's E&M textbook), the rule we find for transforming these fields looks *nothing* like any of the Lorentz transformation rules I learned for other quantities! It was only after learning about tensors, understanding that $\mathbf{E}$ and $\mathbf{B}$ were best thought of us components of a rank-2 antisymmetric tensor, and spending some time developing fluency with operations like Eq. (11.28) that I started to become comfortable with these rules.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 12
A covariant formulation of electromagnetics (part II)

12.1 The field equations

In the previous lecture, we showed that the Lorentz force law written using 3-vectors,

$$\mathbf{F} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) \;, \tag{12.1}$$

is exactly equivalent to the 4-force law

$$\frac{dp^\alpha}{d\tau} = qF^{\alpha\beta}u_\beta \;, \tag{12.2}$$

provided that the *Faraday tensor* components are related to the electric and magnetic field components according to

$$F^{\alpha\beta} \doteq \begin{pmatrix} 0 & E^x/c & E^y/c & E^z/c \\ -E^x/c & 0 & B^z & -B^y \\ -E^y/c & -B^z & 0 & B^x \\ -E^z/c & B^y & -B^x & 0 \end{pmatrix} \;. \tag{12.3}$$

More specifically, we found that the spatial components of $dp^\alpha/d\tau$ correspond exactly to the 3-force $\mathbf{F} = q(\mathbf{E}+\mathbf{v}\times\mathbf{B})$, and that the 0 or timelike component tells us about the work that is done on a charge by the electric field.

In this lecture, we are going to turn to a study of the field equations: how do we make the set of Maxwell equations,

$$\nabla\cdot\mathbf{E} = \rho/\epsilon_0 \;, \qquad \nabla\cdot\mathbf{B} = 0 \;, \tag{12.4}$$

$$\nabla\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t} \;, \qquad \nabla\times\mathbf{B} = \mu_0\mathbf{J} + \mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t} \;, \tag{12.5}$$

fit into this framework?

The first thing we want to do is massage these equations a little bit. Notice that half of the Maxwell equations involve sources, either ρ or $\mathbf{J}$; the other half only involve the fields themselves. Let's reorganize the equations to emphasize this structure:

$$\nabla\cdot\mathbf{E} = \rho/\epsilon_0 \;, \qquad \nabla\times\mathbf{B} - \mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t} = \mu_0\mathbf{J} \;; \tag{12.6}$$

$$\nabla\cdot\mathbf{B} = 0 \;, \qquad \nabla\times\mathbf{E} + \frac{\partial\mathbf{B}}{\partial t} = 0 \;. \tag{12.7}$$

We have put all terms that involve the fields onto the left-hand side of these equations, and set them so that the right-hand side is either "source" (ρ or $\mathbf{J}$) or zero. Notice that there are four sourced equations (one divergence of $\mathbf{E}$, three components of the curl of $\mathbf{B}$), and four source-free equations (one divergence of $\mathbf{B}$, three components of the curl of $\mathbf{E}$).

12.1.1 Half of the Maxwell equations

Let's start by just taking derivatives of the Faraday tensor. By contracting a derivative on one of the indices, we'll generate four different terms, one for each value of the remaining free index:

$$\frac{\partial F^{\alpha\beta}}{\partial x^\beta} = \partial_\beta F^{\alpha\beta} \ . \tag{12.8}$$

(Why contract on the second index? Strictly speaking, it doesn't matter much — because $F^{\alpha\beta}$ is antisymmetric, we'd just get a minus sign if we contracted on the first one.)

Let's go into a Lorentz frame and see what $\partial_\beta F^{\alpha\beta}$ looks like as α goes over its free range:

$$\begin{aligned} \alpha = 0: \qquad \partial_\beta F^{0\beta} &= \frac{\partial}{\partial x}\left(\frac{E^x}{c}\right) + \frac{\partial}{\partial y}\left(\frac{E^y}{c}\right) + \frac{\partial}{\partial z}\left(\frac{E^z}{c}\right) \\ &= \frac{1}{c}\nabla \cdot \mathbf{E} \ . \end{aligned} \tag{12.9}$$

In other words, up to a factor of $1/c$, the $\alpha = 0$ component of $\partial_\beta F^{\alpha\beta}$ looks just like the divergence of $\mathbf{E}$, and so produces the left-hand side of one of the sourced Maxwell equations.

Let's look at the other values of α:

$$\begin{aligned} \alpha = 1: \qquad \partial_\beta F^{1\beta} &= \frac{1}{c}\frac{\partial}{\partial t}\left(-\frac{E^x}{c}\right) + \frac{\partial B^z}{\partial y} - \frac{\partial B^y}{\partial z} \\ &= -\frac{1}{c^2}\frac{\partial E^x}{\partial t} + (\nabla \times \mathbf{B})^x \\ &= -\mu_0\epsilon_0\frac{\partial E^x}{\partial t} + (\nabla \times \mathbf{B})^x \ . \end{aligned} \tag{12.10}$$

(We've used the fact that $c = 1/\sqrt{\mu_0\epsilon_0}$ here.) This analysis shows that the $\alpha = 1$ component of $\partial_\beta F^{\alpha\beta}$ produces the left-hand side of another one of the source Maxwell equations. It's not too hard to show that the $\alpha = 2$ and $\alpha = 3$ components produce the remaining two left-hand sides:

$$\alpha = 2: \qquad \partial_\beta F^{2\beta} = -\mu_0\epsilon_0\frac{\partial E^y}{\partial t} + (\nabla \times \mathbf{B})^y \ , \tag{12.11}$$

$$\alpha = 3: \qquad \partial_\beta F^{3\beta} = -\mu_0\epsilon_0\frac{\partial E^z}{\partial t} + (\nabla \times \mathbf{B})^z \ . \tag{12.12}$$

To get the right-hand side of the sourced Maxwell equations, recall a few lectures ago that we defined the 4-vector $\vec{J}$ whose time-like component $J^t = c\rho$, but whose space-like components are the "normal" 3-vector current density. Comparison of Eq. (12.6) with Eqs. (12.10) – (12.12) suggest that the form we want is

$$\partial_\beta F^{\alpha\beta} = \mu_0 J^\alpha \ . \tag{12.13}$$

It's pretty clear that this form works perfectly for $\alpha = 1, 2, 3$. Does it also work for $\alpha = 0$? Let's check: using Eq. (12.9),

$$\partial_\beta F^{0\beta} = \mu_0 J^0 \qquad \text{becomes} \qquad \frac{1}{c}\nabla \cdot \mathbf{E} = \mu_0 c\rho \ . \tag{12.14}$$

Multiplying both sides by c and using $c = 1/\sqrt{\mu_0 \epsilon_0}$, this becomes

$$\nabla \cdot \mathbf{E} = \rho/\epsilon_0 \ . \tag{12.15}$$

So it works! We've found that *half* of the Maxwell equations — the half that have source terms, either charge density ρ or current density $\mathbf{J}$ — are equivalent to the equation

$$\partial_\beta F^{\alpha\beta} = \mu_0 J^\alpha \ . \tag{12.16}$$

12.1.2 The other half of the Maxwell equations

What about the other half of the Maxwell equations — how do we get the ones that don't have a source? There's no way to get those equations just by taking derivatives of $F^{\alpha\beta}$. Differentiating this quantity can only duplicate the derivatives we have already worked out to get the sourced Maxwell equations. We need a different way of organizing the fields.

The way we get there is by thinking about how to organize the electric and magnetic fields into an antisymmetric tensor. Let's look at our re-organization of Maxwell's equations into "sourced" and "source-free" versions, Eqs. (12.6) and (12.7). Notice that the left-hand sides of these equations are identical provided we "swap" $\mathbf{E}$ and $\mathbf{B}$ in the following way:

$$\mathbf{E}/c \to \mathbf{B} \ , \ \mathbf{B} \to -\mathbf{E}/c \ . \tag{12.17}$$

Taking the left-hand side of the "sourced" Maxwell equations and swapping the fields according to Eq. (12.17) yields the left-hand side of the "source-free" Maxwell equations.

Inspired by this observation, suppose we take $F^{\alpha\beta}$ and apply this field swap:

$$F^{\alpha\beta}\left(\mathbf{E}/c \to \mathbf{B} \ , \ \mathbf{B} \to -\mathbf{E}/c\right) = \begin{pmatrix} 0 & B^x & B^y & B^z \\ -B^x & 0 & -E^z/c & E^y/c \\ -B^y & E^z/c & 0 & -E^x/c \\ -B^z & -E^y/c & E^x/c & 0 \end{pmatrix} \equiv G^{\alpha\beta} \ . \tag{12.18}$$

This quantity is known as the *dual*[1] Faraday tensor. It has the same symmetries as the Faraday tensor; and, if you apply the rule $\mathbf{E}/c \to \mathbf{B}$, $\mathbf{B} \to -\mathbf{E}/c$ to the rules for Lorentz transforming the fields, you find that they are unchanged. [You can test this by applying the rule to Eqs. (11.37) and (11.38) from Lecture 11]. The dual Faraday tensor does not[2], however, give us a force law.

If we differentiate $G^{\alpha\beta}$, we get field derivatives that differ from those that come from

[1] You might find the way that we derived this dual tensor to be somewhat schematic. There is in fact a more rigorous way of doing this which takes advantage of a 4-index version of the Levi-Civita symbol you used on problem set 3: by appropriately combining $F^{\alpha\beta}$ with $\epsilon_{\alpha\beta\gamma\delta}$ (an object which generalizes ϵ_{ijk} to spacetime) and the metric $\eta_{\alpha\beta}$, we can build the tensor $G^{\alpha\beta}$. For the purpose of 8.033, the schematic approach is good enough.

[2] Interestingly, this tensor *would* be involved in a force law if there were magnetic charges as well as electric charges. Perhaps something to explore on a problem set...

differentiating $F^{\alpha\beta}$. Let's go through a few examples of $\partial_\beta G^{\alpha\beta}$:

$$\begin{aligned}
\alpha = 0: \qquad \partial_\beta G^{0\beta} &= \frac{\partial B^x}{\partial x} + \frac{\partial B^y}{\partial y} + \frac{\partial B^z}{\partial z} \\
&= \nabla \cdot \mathbf{B}\,; \qquad (12.19) \\
\alpha = 1: \qquad \partial_\beta G^{1\beta} &= -\frac{1}{c}\frac{\partial B^x}{\partial t} - \frac{1}{c}\frac{\partial E^z}{\partial y} + \frac{1}{c}\frac{\partial E^y}{\partial z} \\
&= -\frac{1}{c}\left[\frac{\partial B^x}{\partial t} + (\nabla \times \mathbf{E})^x\right]\,. \qquad (12.20)
\end{aligned}$$

The $\alpha = 2$ and $\alpha = 3$ components duplicate the y and z components of the curl $\mathbf{E}$ part of Eq. (12.7). Putting this all together, we see that

$$\partial_\beta G^{\alpha\beta} = 0 \tag{12.21}$$

is exactly what we need to write the source-free Maxwell equations in a covariant way.

To summarize: our original presentation of the Maxwell equations, Eqs. (12.4) and (12.5) are not wrong, but are formulated in such a way that they use information specific to some particular Lorentz frame. The fields $\mathbf{E}$ and $\mathbf{B}$ are particular to that observer, as is the charge density ρ and current density $\mathbf{J}$, as is the notion of space and time they use to take their derivatives. These equations are exactly equivalent to the covariant formulation

$$\partial_\beta F^{\alpha\beta} = \mu_0 J^\alpha\,, \qquad \partial_\beta G^{\alpha\beta} = 0\,. \tag{12.22}$$

For our present purpose, Eq. (12.22) is preferred to Eqs. (12.4) and (12.5) because it shows us how to write these equations in a way that is formulated for a different Lorentz observer. If the coordinates $x^{\alpha'}$ are used by $\mathcal{O}'$, then we know that their formulation of Maxwell's equations looks like

$$\partial_{\beta'} F^{\alpha'\beta'} = \mu_0 J^{\alpha'}\,, \qquad \partial_{\beta'} G^{\alpha'\beta'} = 0\,. \tag{12.23}$$

We can get all the "prime frame" quantities by just appropriate correcting things using the Lambda matrices, with all the quantities connected using the "line up the indices" rule.

12.2 Automatic conservation of source

In our discussion of conservation laws, we noted that the equation of charge continuity,

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{J}\,, \tag{12.24}$$

has a covariant formulation

$$\partial_\alpha J^\alpha = 0\,. \tag{12.25}$$

Let's revisit this in the context of our covariant formulation of Maxwell equations: taking a derivative of $\partial_\beta F^{\alpha\beta} = \mu_0 J^\alpha$, we have

$$\partial_\alpha \partial_\beta F^{\alpha\beta} = \mu_0 \partial_\alpha J^\alpha\,. \tag{12.26}$$

The right-hand side of this is zero by virtue of charge continuity. What about the left-hand side? Let's look at it carefully:

$$\begin{aligned}
\partial_\alpha \partial_\beta F^{\alpha\beta} &= -\partial_\alpha \partial_\beta F^{\beta\alpha} \qquad &&\text{(Antisymmetry of } F^{\alpha\beta}\text{)} \\
&= -\partial_\beta \partial_\alpha F^{\beta\alpha} &&\text{(Symmetry of } \partial_\alpha\partial_\beta\text{)} \\
&= -\partial_\alpha \partial_\beta F^{\alpha\beta} &&\text{(Relabeling of dummy indices)} \qquad (12.27)
\end{aligned}$$

Comparing the first line with the last we see we again have a situation where the quantity in question is equal to the negative of itself; this is another example of the situation of a symmetric object (in this case, the pair of derivatives $\partial_\alpha\partial_\beta$) contracted onto an antisymmetric one ($F^{\alpha\beta}$). We must have

$$\partial_\alpha\partial_\beta F^{\alpha\beta} = 0 \; . \tag{12.28}$$

This little calculation reveals a very important point: theories of physics in which some source yields a field typically are governed by a set of field equations whose heuristic structure is of the form

$$(\text{Derivatives})(\text{Fields}) = (\text{Source}) \; . \tag{12.29}$$

Sources are never unconstrained; they arise from physical matter, and so respect conservation laws. We can write those conservation laws in the form

$$(\text{Other derivatives})(\text{Source}) = 0 \; . \tag{12.30}$$

For this to hold up, we really need to have the mathematical structure which holds our fields respect the rule that

$$(\text{Other derivatives})(\text{Derivatives})(\text{Fields}) = 0 \; . \tag{12.31}$$

Although we didn't explicitly set out to make our Faraday tensor fit into this framework, it turns out that it does. This becomes an important point to bear in mind as we think about other kinds of interactions that we might want to fit into a relativistic framework.

12.3 Field invariants

Lorentz transformation act on free indices. Any quantity with no free indices is thus invariant under Lorentz transformations; this is why the scalar product between two 4-vectors, $a^\mu b_\mu$, always yields a Lorentz invariant.

Can we make invariants out of tensors? Certainly! — we just have to combine things, using the metric to lower (or raise) indices, such that there are no free indices for the Lorentz transformation matrix to affect.

Perhaps the simplest one we can construct is called the *trace*. In linear algebra, the trace of a matrix is the sum of its diagonal entries. When we are dealing with tensors, we make this a little more formal: we sum over the indices with one upstairs, and one downstairs. Let's look at this for the Faraday tensor:

$$F^\mu{}_\mu = F^{\alpha\mu}\eta_{\mu\alpha} \; . \tag{12.32}$$

This is a quantity whose values all Lorentz frames agree on. Unfortunately, in this case, it doesn't turn out to be very interesting: using the Faraday tensor $F^{\alpha\beta}$ we've listed above and combining with $\eta_{\mu\alpha} = \mathrm{diag}(-1, 1, 1, 1)$, we get

$$F^\mu{}_\mu = 0 + 0 + 0 + 0 = 0 \; . \tag{12.33}$$

The number zero is indeed a Lorentz invariant, but we don't learn anything useful from doing this analysis. (We get the exact same result if we evaluate $G^\mu{}_\mu$.)

We can make others Lorentz invariants by combining the Faraday tensor with itself. Let's look at

$$F^{\alpha\beta}F_{\alpha\beta} = F^{\alpha\beta}F^{\mu\nu}\eta_{\alpha\mu}\eta_{\beta\nu} \; . \tag{12.34}$$

With a little bit of effort, you should be able to show that the Faraday tensor with all indices in the downstairs position is represented by the matrix

$$F_{\alpha\beta} \doteq \begin{pmatrix} 0 & -E^x/c & -E^y/c & -E^z/c \\ E^x/c & 0 & B^z & -B^y \\ E^y/c & -B^z & 0 & B^x \\ E^z/c & B^y & -B^x & 0 \end{pmatrix} ; \tag{12.35}$$

i.e., both row 0 and column 0 are multiplied by negative 1 versus $F^{\alpha\beta}$; cf. Eq. (12.3). (You did a very similar kind of manipulation on problem 8 of problem set #5. As part of that analysis, you found that the "00" component of the tensor is multiplied by -1 twice, leaving it unchanged. In this case, you are multiplying zero by -1 twice, so this is a particularly uninteresting application of this rule.)

Using Eq. (12.35), it is straightforward to show that

$$\begin{aligned} F^{\alpha\beta}F_{\alpha\beta} &= 2\left[(B^x)^2 + (B^y)^2 + (B^z)^2 - (E^x/c)^2 - (E^y/c)^2 - (E^z/c)^2\right] \\ &= 2\left[\mathbf{B}\cdot\mathbf{B} - \mathbf{E}\cdot\mathbf{E}/c^2\right] . \end{aligned} \tag{12.36}$$

In other words, *the quantity* $|\mathbf{B}|^2 - |\mathbf{E}|^2/c^2$ *is the same to all Lorentz observers.* This could in principle be deduced by careful study of the Lorentz transformed fields that we derived in the previous lecture, but it follows very simply and easily from the fact that $F^{\alpha\beta}F_{\alpha\beta}$ must be a Lorentz invariant.

There are two other Lorentz invariants we can form from the field tensors. One of them, $G^{\alpha\beta}G_{\alpha\beta}$, is identical to $F^{\alpha\beta}F_{\alpha\beta}$ except for the overall sign, so it yields no new information. But the other one is more interesting:

$$\begin{aligned} F^{\alpha\beta}G_{\alpha\beta} &= 4\left(B^xE^x/c + B^yE^y/c + B^zE^z/c\right) \\ &= 4\mathbf{B}\cdot\mathbf{E} . \end{aligned} \tag{12.37}$$

All observers agree on the 3-dimensional dot product of $\mathbf{E}$ *and* $\mathbf{B}$. Again, this could have been deduced directly from the fields, but doing with the field tensors is far simpler and more straightforward.

12.4 Potentials and gauge freedom (CAUTION: somewhat advanced material)

[NOTE: I will occasionally discuss material that is a bit more advanced than, strictly speaking, we intend for 8.033. When I do this, I will use a "CAUTION" flag as I've written in this section heading. Students who wish to do so can skip over these sections. Some of this material is likely to fit in better after you have taken additional coursework. For example, this present section is probably best for students who either discussed gauge freedom in their 1st-year E&M class (which doesn't happen for all students), or who have taken 8.07.]

12.4.1 A covariant formulation of electromagnetic potentials

We began our discussion of a covariant formulation of electrodynamics by noting that we cannot "fit" the 6 functions which describe electric and magnetic fields into a 4-vector. A few of you may have wondered: what about the potentials? In freshman electricity and magnetism, we learn that electric fields can be written as the gradient of a scalar potential, and the magnetic field as the curl of a vector potential; in more advanced presentations, we learn that the electric field in situations with time-varying magnetic fields has a contribution from the time-derivative of the vector potential:

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}\;, \qquad \mathbf{B} = \nabla \times \mathbf{A}\;. \tag{12.38}$$

One scalar potential, 3 components of vector potential ... this looks tailor-made to fit into a 4-vector! The potentials ϕ and $\mathbf{A}$ have different dimensions, so to make this work we again need to introduce a factor of c. Doing so, we define the 4-potential $\vec{A} = A^\mu \vec{e}_\mu$, whose components are given by

$$A^\mu \doteq \begin{pmatrix} \phi/c \\ A^x \\ A^y \\ A^z \end{pmatrix}\;. \tag{12.39}$$

We know that $F^{\alpha\beta}$ is antisymmetric, and the fields are built by taking derivatives of the potentials. So let's make an antisymmetric combination of derivatives of fields:

$$X^{\alpha\beta} = \partial^\alpha A^\beta - \partial^\beta A^\alpha\;. \tag{12.40}$$

Notice that we are using the "upstairs" partial derivative, $\partial^\alpha = \partial/\partial x_\alpha$. We do this so that we can create tensor components whose indices are all raised, guaranteeing that they have the correct antisymmetry. Recall from Lecture 9 that $x_\alpha \equiv \eta_{\alpha\beta}x^\beta$, and so the components of ∂^α are nearly identical to those of ∂_α. The critical difference is that the zero component has the opposite sign: $\partial^0 = -\partial_0 = -(1/c)\partial/\partial t$.

Let's go through some of the components of $X^{\alpha\beta}$. We can skip X^{00}, X^{11}, X^{22}, X^{33} — the form of Eq. (12.40) guarantees that they are zero. Let's move across row 0:

$$\begin{aligned} X^{01} &= \partial^0 A^1 - \partial^1 A^0 \\ &= -\frac{1}{c}\frac{\partial A^x}{\partial t} - \frac{1}{c}\frac{\partial\phi}{\partial x} \\ &= E^x/c\;. \end{aligned} \tag{12.41}$$

Comparing with Eq. (12.3), we see that $X^{01} = F^{01}$. We likewise quickly find that $X^{02} = F^{02}$, and $X^{03} = F^{03}$.

Let's move across row 1. We can skip X^{10} — it will be $-X^{01}$, quickly showing that $X^{10} = F^{10}$. Moving to the first component that is new,

$$\begin{aligned} X^{12} &= \partial^1 A^2 - \partial^2 A^1 \\ &= \frac{\partial A^y}{\partial x} - \frac{\partial A^x}{\partial y} \\ &= (\nabla \times \mathbf{A})^z \\ &= B^z\;. \end{aligned} \tag{12.42}$$

Comparing with Eq. (12.3), we see that $X^{12} = F^{12}$. By a similar set of calculations, we quickly show that $X^{13} = F^{13}$, and that $X^{23} = F^{23}$. Thanks to the antisymmetry, we are done, and conclude that

$$F^{\alpha\beta} = \partial^\alpha A^\beta - \partial^\beta A^\alpha\;. \tag{12.43}$$

12.4.2 Gauge freedom

One of the things we learn in electrodynamics classes is that we have some freedom to adjust the form of the potentials, as long as these adjustments have no impact on the fields; after all, it is the fields that exert forces and that are directly measurable. In a particular Lorentz frame, the form that this takes is that we imagine there exists some scalar function λ, which we will call the "gauge generator." It is not difficult to show that if we adjust the potentials as follows,

$$\phi_{\rm new} = \phi_{\rm old} - \frac{\partial\lambda}{\partial t}\ , \qquad \mathbf{A}_{\rm new} = \mathbf{A}_{\rm old} + \nabla\lambda\ , \tag{12.44}$$

then the fields $\mathbf{E}$ and $\mathbf{B}$ are unchanged. We prove this by simply computing the fields using $\phi_{\rm new}$ and $\mathbf{A}_{\rm new}$ rather than $\phi_{\rm old}$ and $\mathbf{A}_{\rm old}$:

$$\begin{aligned}
\mathbf{E}' &= -\nabla\phi_{\rm new} - \frac{\partial\mathbf{A}_{\rm new}}{\partial t}\\
&= -\nabla\phi_{\rm old} + \nabla\frac{\partial\lambda}{\partial t} - \frac{\partial\mathbf{A}_{\rm old}}{\partial t} - \frac{\partial}{\partial t}\nabla\lambda\\
&= -\nabla\phi_{\rm old} - \frac{\partial\mathbf{A}_{\rm old}}{\partial t}\\
&= \mathbf{E}\ ; &(12.45)\\
\mathbf{B}' &= \nabla\times\mathbf{A}_{\rm new}\\
&= \nabla\times\mathbf{A}_{\rm old} + \nabla\times\nabla\lambda\\
&= \nabla\times\mathbf{A}_{\rm old}\\
&= \mathbf{B}\ . &(12.46)
\end{aligned}$$

In the proof for $\mathbf{E}$, we used the fact that partial derivatives commute to see that

$$\nabla\frac{\partial\lambda}{\partial t} - \frac{\partial}{\partial t}\nabla\lambda = 0\ ; \tag{12.47}$$

for $\mathbf{B}$, we used the fact that the curl of the gradient of any scalar function is zero.

The way we bring gauge freedom into the covariant framework is quite simple: we set

$$A^\alpha_{\rm new} = A^\alpha_{\rm old} + \partial^\alpha\lambda\ . \tag{12.48}$$

With this, it is simple to see that the Faraday tensor is unchanged:

$$\begin{aligned}
F^{\alpha\beta}_{\rm new} &= \partial^\alpha A^\beta_{\rm new} - \partial^\beta A^\alpha_{\rm new}\\
&= \partial^\alpha A^\beta_{\rm old} + \partial^\alpha\partial^\beta\lambda - \partial^\beta A^\alpha_{\rm old} - \partial^\beta\partial^\alpha\lambda\\
&= \partial^\alpha A^\beta_{\rm old} - \partial^\beta A^\alpha_{\rm old}\\
&= F^{\alpha\beta}_{\rm old}\ . &(12.49)
\end{aligned}$$

12.4.3 An example application of gauge freedom

If you've never encountered gauge transformations before, you might wonder why we might want to change from one gauge to another. If both gauges give the same fields, and the fields are things that ultimately act on charges and currents, then who cares? What good comes from messing around with this detail?

To see an example of why this can quite useful, let's look at the sourced Maxwell equation, but written in terms of the potential:

$$\partial_\beta F^{\alpha\beta} = \partial_\beta \partial^\alpha A^\beta - \partial_\beta \partial^\beta A^\alpha = \mu_0 J^\alpha \; . \tag{12.50}$$

Because partial derivatives commute, we can swap the order of the derivatives in the first term involving the potential. And, we recognize the combination of derivatives in the second term as the invariant wave operator. The sourced Maxwell equation can thus be rewritten

$$\Box A^\alpha - \partial^\alpha \left(\partial_\beta A^\beta \right) = -\mu_0 J^\alpha \; . \tag{12.51}$$

Equations of the form

$$\Box\,(\text{Function}) = (\text{Source}) \tag{12.52}$$

are particularly "lovely" in physics — powerful computational techniques make it possible to solve such equations. Unfortunately, the form we've got, (12.51) is not *quite* in that form: it's skewed a bit by the "extra" term $\partial^\alpha(\partial_\beta A^\beta)$. If we could get rid of that extra term, the equation relating A^α to J^α would be solvable using these powerful techniques.

Gauge freedom to the rescue. Suppose we change gauge, putting

$$A^\beta_{\text{new}} = A^\beta_{\text{old}} + \partial^\beta \lambda \; . \tag{12.53}$$

The term which makes Eq. (12.51) not quite "lovely" for us then involves

$$\partial_\beta A^\beta_{\text{new}} = \partial_\beta A^\beta_{\text{old}} - \partial_\beta \partial^\beta \lambda = \partial_\beta A^\beta_{\text{old}} - \Box\lambda \; . \tag{12.54}$$

If we choose our gauge generator such that

$$\Box\lambda = \partial_\beta A^\beta_{\text{old}} \; , \tag{12.55}$$

then the offending term vanishes: we then have

$$\partial_\beta A^\beta_{\text{new}} = \partial_\beta A^\beta_{\text{old}} - \Box\lambda = 0 \; . \tag{12.56}$$

We can in fact always find a gauge generator λ which satisfies Eq. (12.55) — those powerful techniques guarantee that equations of the form (12.52) always have a solution. Because of this, we can just assume that we have done this analysis, and jump straight to using the potential in this new gauge. The sourced Maxwell equation then becomes (dropping the "new" subscript)

$$\Box A^\alpha = -\mu_0 J^\alpha \; . \tag{12.57}$$

When the potential satisfies Eq. (12.56), we say that it is in *Lorenz*[3] gauge. This gauge is particularly useful for studies of electromagnetic radiation, since the equation governing the potential is nothing more than a wave equation with a source. Other gauges exist, and can be really useful in particular reference frames. Such gauges tend not to be "nice" in covariant formulation, though, since they are designed to work only in some frame.

[3]Note: *not* Lorentz! Ludvig Lorenz developed this gauge; Hendrik Lorentz first developed the Lorentz transformation. Generations of physicists (including your lecturer) learned this wrong, but most recent electrodynamics textbooks have been working to correct this error. See J. D. Jackson and L. B. Okun, Reviews of Modern Physics **73**, 663 (2001).

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 13
Accelerations and forces

13.1 An apparent paradox

Consider a pair of twins. Twin A stays on Earth. Twin B travels on a rocket ship to Alpha Centauri, 4 light years away, moving at 99% of the speed of light. Twin B then turns around and comes back. When they get together, which one is older?

The essence of the apparent paradox is that, according to special relativity, no inertial observer is preferred:

- Twin A says that B is in motion. Therefore, B's clock runs slow, and B is younger.
- Twin B says that A is in motion. Therefore, A's clock runs slow, and A is younger.

When the twins reunite, they can't both be right — one of them has unambiguously aged more than the other. Who has used the wrong logic?

Twin B has used the wrong logic, because they forgot that **they are not an inertial observer**. Twin B accelerates (3 times: once from Earth to start the trip, once at Alpha Centauri to turn around and come back, and once upon returning to Earth). This acceleration breaks the symmetry between the twins.

Does this mean that Twin B is older or younger? To answer this, we need to think about accelerated motion.

13.2 4-acceleration; the Momentarily Comoving Reference Frame (MCRF)

We begin by quickly re-examining the notion of 4-acceleration, which was briefly introduced in our discussion of the Lorentz force. We define the 4-acceleration by

$$\vec{a} = \frac{d\vec{u}}{d\tau} , \tag{13.1}$$

i.e., the rate of change of 4-velocity per unit proper time. As discussed in that earlier lecture, its invariant scalar product with $\vec{u}$ is zero, which follows from $\vec{u} \cdot \vec{u} = -c^2$:

$$\frac{d}{d\tau} (\vec{u} \cdot \vec{u}) = \vec{a} \cdot \vec{u} + \vec{u} \cdot \vec{a} = 2\vec{a} \cdot \vec{u} = 0 . \tag{13.2}$$

This is in sharp contrast to the 3-acceleration $\mathbf{a}$, since physics imposes no constraints on the value of $\mathbf{a} \cdot \mathbf{u}$ (here using the "old-fashioned" dot product between two 3-vectors).

To wrap our heads around the physics of acceleration, let's introduce a particular special reference frame: the MCRF, or *Momentarily Comoving Reference Frame*. The MCRF is a Lorentz frame that, at least for one moment, has the same velocity as the accelerating observer. An accelerating observer is at rest in the MCRF for one moment.

In the MCRF, the following properties describe the motion of the accelerating observer:

$$u^t_{\rm MCRF} = c\ , \qquad u^{x,y,z}_{\rm MCRF} = 0\ , \tag{13.3}$$

$$d\tau = dt_{\rm MCRF}\ . \tag{13.4}$$

These properties tell us that

$$a^\mu_{\rm MCRF} \doteq \begin{pmatrix} 0 \\ du^x_{\rm MCRF}/dt \\ du^y_{\rm MCRF}/dt \\ du^z_{\rm MCRF}/dt \end{pmatrix}\ . \tag{13.5}$$

This form guarantees that $\vec a\cdot\vec u = 0$: if you evaluate that scalar product using the components defined in the MCRF, you can see quite clearly that it holds. But, the scalar product is an invariant — if it is true in one frame, then it is true in all frames.

The MCRF thus helps us to understand what 4-acceleration means. Suppose some observer has a 4-acceleration $\vec a$, and that we find $\vec a\cdot\vec a = a^2$. (Note that $\vec a$ must be spacelike[1] in order for $\vec a\cdot\vec u = 0$.) Then a represents the magnitude of the acceleration that is experienced by the accelerating observer in the MCRF. It is the acceleration that this observer feels in their own rest frame.

13.3 A uniformly accelerated observer

Let's imagine an observer who starts at rest with respect to us, but who experiences uniform acceleration with magnitude $g = 10\,{\rm m/sec}^2$. Let this acceleration be in the x direction. "Uniform" means that the observer feels this acceleration at all times, so that $\vec a\cdot\vec a = g^2$ at all times. The acceleration in the MCRF is always the same — even though the MCRF itself is continually changing as the observer accelerates. Can we compute the 4-velocity at later times?

We have two initial conditions: $\vec u(\tau = 0) = c\vec e_t$ and $\vec a(\tau = 0) = g\vec e_x$. We also have three constraints:

$$\vec u\cdot\vec u = -c^2 \qquad \text{at all times}\ , \tag{13.6}$$

$$\vec u\cdot\vec a = 0 \qquad \text{at all times}\ , \tag{13.7}$$

$$\vec a\cdot\vec a = g^2 \qquad \text{at all times}\ . \tag{13.8}$$

Let's write out these constraint equations, using the fact that $a^\mu = du^\mu/d\tau$:

$$-(u^t)^2 + (u^x)^2 = -c^2\ , \tag{13.9}$$

$$-u^t\frac{du^t}{d\tau} + u^x\frac{du^x}{d\tau} = 0\ , \tag{13.10}$$

$$-\left(\frac{du^t}{d\tau}\right)^2 + \left(\frac{du^x}{d\tau}\right)^2 = g^2\ . \tag{13.11}$$

[1]To be lightlike, we must have $\vec a\cdot\vec a = 0$. That's only true if $a = 0$, an uninteresting limit.

Staring at these equations a bit and thinking about some functions we know suggests that hyperbolic functions might be useful here. Let's try

$$u^t = c\cosh\left(A\tau\right)\ , \qquad u^x = c\sinh\left(A\tau\right)\ . \tag{13.12}$$

It's not hard to see that this form guarantees Eqs. (13.6) and (13.7) will work. Enforcing Eq. (13.8) gives us the constant A:

$$\frac{du^t}{d\tau} = cA\sinh\left(A\tau\right)\ , \qquad \frac{du^x}{d\tau} = cA\cosh\left(A\tau\right)\ ; \tag{13.13}$$

so

$$-\left(\frac{du^t}{d\tau}\right)^2 + \left(\frac{du^x}{d\tau}\right)^2 = c^2A^2\left[-\sinh^2\left(A\tau\right) + \cosh^2\left(A\tau\right)\right] = g^2 \tag{13.14}$$

which tells us that

$$A = \frac{g}{c}\ . \tag{13.15}$$

Our complete solution for the uniformly accelerated observer is thus

$$\vec{u} = c\cosh\left(g\tau/c\right)\vec{e}_t + c\sinh\left(g\tau/c\right)\vec{e}_x\ , \tag{13.16}$$

$$\vec{a} = g\sinh\left(g\tau/c\right)\vec{e}_t + g\cosh\left(g\tau/c\right)\vec{e}_x\ , \tag{13.17}$$

where τ is the proper time experienced by this observer since their trip started.

Let's use this solution to explore what happens when someone is uniformly accelerated. Two questions are at the top of our list:

1. After traveling for time T as measured *by the accelerating observer* (i.e., for a total experienced proper time $\tau = T$), how far has the observer traveled?

2. After traveling for time T as measured *by the accelerating observer*, how much time has elapsed "back home"?

Both questions are answered by integrating the 4-velocity. Let's look at how far they've traveled first:

$$\begin{aligned}\Delta x &= \int_0^T u^x\,d\tau \\ &= c\int_0^T \sinh\left(\frac{g\tau}{c}\right)d\tau \\ &= \frac{c^2}{g}\left[\cosh\left(\frac{gT}{c}\right) - 1\right]\ .\end{aligned} \tag{13.18}$$

Using the fact that $c^2/g = 0.96940$ light years, and $(g/c) = 1.0316\,\mathrm{year}^{-1}$, we can make a table of distance versus time experienced by the accelerating observer:

- $\Delta x(T = 1\,\mathrm{year}) = 0.56318\,\mathrm{light\ year}$
- $\Delta x(T = 2\,\mathrm{years}) = 2.9071\,\mathrm{light\ years}$
- $\Delta x(T = 5\,\mathrm{years}) = 83.268\,\mathrm{light\ years}$
- $\Delta x(T = 10\,\mathrm{years}) = 14{,}638\,\mathrm{light\ years}$

How much time back in the original frame elapses while doing this?

$$\begin{aligned}\Delta t &= \int_0^T (u^t/c)d\tau \\ &= \int_0^T \cosh\left(\frac{g\tau}{c}\right) d\tau \\ &= \frac{c}{g}\sinh\left(\frac{gT}{c}\right) \ .\end{aligned} \tag{13.19}$$

The equivalent table for time elapsed reads

- $\Delta t(T = 1\,\text{year}) = 1.1870\,\text{years}$
- $\Delta t(T = 2\,\text{years}) = 3.7533\,\text{years}$
- $\Delta t(T = 5\,\text{years}) = 84.232\,\text{years}$
- $\Delta t(T = 10\,\text{years}) = 14{,}639\,\text{years}$

As seen back in the original frame, the accelerated observer is getting closer and close to the speed of light, and so is experiencing enormous time dilation. Their 10 year interval is over 14,600 years in the original frame — their moving clock is running *very* slowly compared to a clock in the original frame.

13.4 Forces

We encountered forces briefly in our discussion of electromagnetic effects. In this section, we return to this discussion, and put a few details on a more solid footing.

Two general conceptual frameworks are used:

1. We can define a 4-force, $\vec{F} = d\vec{p}/d\tau$. In terms of this, we have $\vec{a} = \vec{F}/m$. In principle, this is the way you might imagine we want to do things, since $\vec{F}$ is a spacetime 4-vector. It is straightforward for us to transform the components of $\vec{F}$ to different reference frames, so this would seem to be the ideal quantity for bringing forces into a relativistic discussion.

2. We can use the usual 3-force, $\mathbf{F} = d\mathbf{p}/dt$. This is fine, as long as we recognize that $\mathbf{p}$ and t are the momentum and time *as measured in a particular frame*, and that we must be careful when we transform them between frames. Changing frames will transform $\mathbf{F}$ in a way that is rather more complicated than a simple Lorentz transformation since quantities in both the numerator and the denominator of the force's definition are affected by this change of representation.

This being a relativity class, you might think we have a preference for the 4-force formulation. However, the 3-force is in fact quite useful and important. This is because we always perform our measurements in some particular frame, using the time and space coordinates of that frame, and pinning down the momentum and energy in that frame. So it is quite useful for us to understand how 3-forces transform between frames as well as 4-forces. Ideally, we'd like to know how to flip back and forth between the two descriptions, as both are important and useful.

Let's go back to our train and station frames. Imagine that a body has a 3-velocity $\mathbf{u}$ as measured in a station, and so has 3-momentum $\mathbf{p}_S = \gamma(u)m\mathbf{u}$ and energy $E_S = \gamma(u)mc^2$ according to the station-frame observers. A train moves through a station with velocity $\mathbf{v} = v\mathbf{e}_x$. If force $\mathbf{F}_S$ acts on the body in the station, what is the force $\mathbf{F}_T$ that acts on the body according to measurements on the train?

When in doubt, go back to the Lorentz transformation. We know that $\mathbf{F} = d\mathbf{p}/dt$, so let's examine the key quantities appearing here and how they transform between frames. Start with the x component:

$$
\begin{aligned}
(\mathbf{F}_T)^x = \frac{dp_T^x}{dt_T} &= \frac{\gamma\left(dp_S^x - v dE_S/c^2\right)}{\gamma\left(dt_S - v\,dx_S/c^2\right)} \\
&= \frac{(\mathbf{F}_S)^x - (v/c^2)(dE_S/dt_S)}{1 - v(\mathbf{u})^x/c^2}\ .
\end{aligned}
\tag{13.20}
$$

Notice we have to a little careful with notation, since the letter "F" is used for both the 4-force and the 3-force and the letter "u" is used for 3-velocity in some frame and 4-velocity. The convention we are using is that F^i represents the ith component of the 4-force, but $(\mathbf{F})^i$ represents the ith component of the 3-force; u^i and $(\mathbf{u})^i$ have analogous meanings for 4-velocity and 3-velocity components, respectively.

We can simplify Eq. (13.20) a bit more. We know that $E^2 = p^2c^2 + m^2c^4$ for the body. Evaluating everything in the station frame and taking derivatives with respect to station time, we have

$$
\begin{aligned}
E_S\frac{dE_S}{dt_S} &= \mathbf{p}_S \cdot \frac{d\mathbf{p}}{dt_S}c^2 \\
\gamma mc^2\frac{dE_S}{dt_S} &= \gamma m\mathbf{u} \cdot \frac{d\mathbf{p}}{dt_S}c^2 \\
\longrightarrow \qquad \frac{dE_S}{dt_S} &= \mathbf{F}_S \cdot \mathbf{u}\ .
\end{aligned}
\tag{13.21}
$$

So, we find that the x component of the force transforms as

$$
(\mathbf{F}_T)^x = \frac{(\mathbf{F}_S)^x - (v/c^2)\mathbf{F}\cdot\mathbf{u}}{1 - v(\mathbf{u})^x/c^2}\ . \tag{13.22}
$$

You may notice a resemblance to the velocity addition formula! Indeed, working out the other two components, we find

$$
(\mathbf{F}_T)^{y,z} = \frac{(\mathbf{F}_S)^{y,z}}{\gamma(1 - v(\mathbf{u})^x/c^2)}\ . \tag{13.23}
$$

Although we have spent some time (and ink/chalk) developing how the 3-force transforms between frames of reference, it should be emphasized that the 4-force is also used quite a lot. The 4-force fits more naturally into a "spacetime" language; the 3-force is more naturally suited to the "space" plus "time" language adapted to a particular observer. Some forces may be very naturally expressed using the 4-force, but we then may need the 3-vector components in order to assess what some observer will measure in their lab. It is important to develop fluency translating back and forth between these different notions of the force.

So, how do we relate these two notions of force? The analysis is somewhat similar to how we relate 4-velocity components to 3-velocity components. Let's consider the spatial components first:

$$F^i = \frac{dp^i}{d\tau} \ . \tag{13.24}$$

The interval $d\tau$ is as measured on the clock of the body which experiences this force. It is related to time as seen in that frame by $d\tau = dt/\gamma(u)$, where u is the magnitude of the body's 3-velocity in that frame. This means

$$F^i = \gamma(u)\frac{dp^i}{dt} = \gamma(u)(\mathbf{F})^i \ . \tag{13.25}$$

Next consider the timelike component:

$$F^0 = \frac{dp^0}{d\tau} = \gamma(u)\frac{d}{dt}\left(\frac{E}{c}\right) = \frac{\gamma(u)}{c}\frac{dE}{dt} \ . \tag{13.26}$$

We already showed that $dE/dt = \mathbf{F} \cdot \mathbf{u}$. Putting this all together, we have a "glossary" that lets us switch back and forth between the 4-vector and 3-vector notions of force:

$$F^0 = \frac{\gamma(u)}{c}\mathbf{F} \cdot \mathbf{u} \ , \tag{13.27}$$

$$F^i = \gamma(u)(\mathbf{F})^i \ . \tag{13.28}$$

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 14
Prelude to gravity: More on the uniformly accelerated observer

14.1 The trajectory of an accelerated observer

In this lecture, as a prelude to discussing certain aspects of gravity, we will look at how things appear to observers who are accelerating. *A word of caution*: some of the calculations we do here are a touch tricky. Certain details require us to develop things beyond the level that is part of the normal core of 8.033; those details are developed toward the end of this set of lecture notes. Do not worry if you cannot follow every calculational detail in this set of notes. We emphasize the core important pieces of this analysis where appropriate, and lay out why they are important for where we are going next. A few of the sections we present below are significantly more complicated than what you are expected to follow; those sections can be skipped, though interested students who wish to discuss them further are welcome to do so.

We begin by examining the trajectory of a single observer who feels a constant acceleration $\mathbf{g} = g\mathbf{e}_x$ in their own momentarily comoving rest frame (MCRF). In the previous lecture, we found that such an observer has a 4-velocity whose components are

$$c\frac{dt}{d\tau} = u^t = c\cosh(g\tau/c)\ , \tag{14.1}$$

$$\frac{dx}{d\tau} = u^x = c\sinh(g\tau/c)\ . \tag{14.2}$$

(To simplify the analysis which follows, which is fairly dense, we take the observer to be at rest in the y and z directions.) Integrating up these solutions, we find the ct and x coordinates describing a uniformly accelerated observer, parameterized by that observer's own proper time:

$$ct = ct_0 + \frac{c^2}{g}\sinh(g\tau/c)\ , \tag{14.3}$$

$$x = x_0 + \frac{c^2}{g}\left(\cosh(g\tau/c) - 1\right)\ . \tag{14.4}$$

We've chosen constants of integration so that $t = t_0$ and $x = x_0$ at $\tau = 0$. The blue curve in Figure 1 shows what this motion looks like, choosing $x_0 = c^2/g$ and $t_0 = 0$.

At any moment as the accelerating observer moves along their worldline, we can find their 3-velocity: it is entirely in the x direction, and has magnitude

$$v^x = c\,u^x/u^t = c\,\tanh(g\tau/c)\ . \tag{14.5}$$

(Notice that the accelerating observer's *rapidity*, which you used on problem sets 2 and 3, increases linearly as a function of that observer's proper time.) Knowing this v^x lets us work out the Lorentz transformation that takes us from inertial coordinates (ct, x) that are at rest with respect to the observer's initial condition to the coordinates $(c\bar{t}, \bar{x})$ corresponding to their MCRF. Figure 1 shows the motion of the accelerating observer according to an inertial observer who is initially at rest with respect to the accelerating observer, along with several examples of constant $\bar{t}$ surfaces in the (ct, x) coordinates for this observer's MCRF at different moments along their worldline. At $\tau = 0$, the constant $\bar{t}$ surface in the MCRF coincides with the $t = 0$ surface in the inertial coordinates. As proper time grows along the worldline, these surfaces grow steeper as the observer moves faster with respect to their original rest frame.

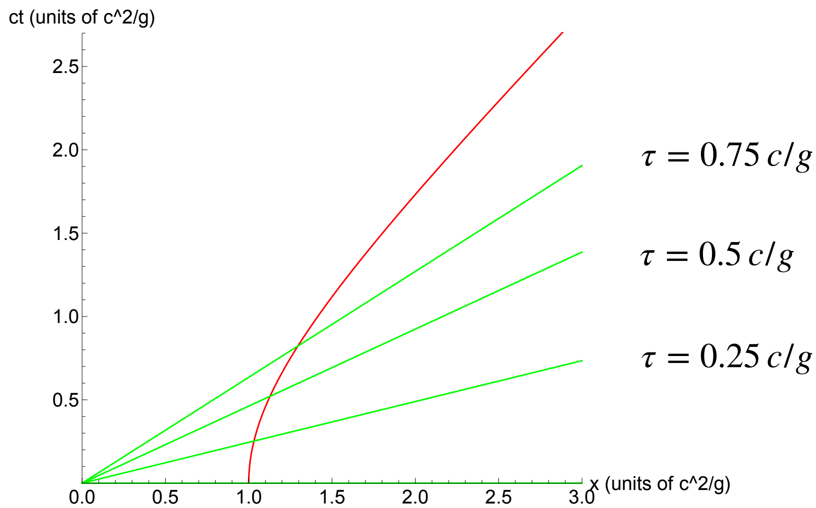


Figure 1: Worldline of an accelerating observer that starts at $x = c^2/g$ (red curve), and three examples of the constant $\bar{t}$ coordinates of that observer's MCRF at different moments along the worldline. The MCRF time $\bar{t}$ coincides with the observer's proper time τ where it crosses the worldline. Notice both axes are in units of c^2/g.

14.2 Comparing the worldlines of *two* accelerated observers: Breakdown of clock synchronization

Now imagine that there are two accelerated observers. Both are at rest with respect to the "unbarred" frame at $t = 0$, and both feel constant acceleration g. One (which will call the "trailing" observer) begins at $x_0 = c^2/g$; the other (the "leading" observer) begins at $x_0 = c^2/g + L$. Let the time as measured on the trailing observer's clock be $\bar{t}$; let the time as measured on the leading observer's clock be $\bar{\bar{t}}$. These times will also be used to describe time in the MCRFs along the accelerating observers' worldlines.

The clocks on these observers start out in agreement, and coincide with the initial inertial frame: when $t = 0$, $\bar{t} = \bar{\bar{t}} = 0$. However, it is not hard to see that as the two observers move along their worldlines, their clocks quickly fall out of agreement. Figure 2 illustrates the situation: once they begin moving, each observer's constant time surface tips over, in accordance with the Lorentz transformation that takes us from the inertial frame into their MCRF. However, they each tip about a different "pivot point," anchored to their own worldline. For a given value of proper time along the worldlines, the constant time $\bar{\bar{t}}$ surface used by the leading observer (whose worldline is illustrated by the orange curve in Fig. 2) always appears in the past of the constant time $\bar{t}$ surface used by the trailing observer (whose worldline is illustrated by the red curve).

This means that, when the leading observer measures time $\bar{\bar{t}} = 0.5\, c/g$ (for example), this is simultaneous with the trailing clock reading some value $\bar{t} < 0.5\, c/g$. The trailing observer agrees with this assessment: when they measure $\bar{t} = 0.5\, c/g$, this is simultaneous with the leading clock reading some value $\bar{\bar{t}} > 0.5\, c/g$. Both observers agree that the leading clock runs faster than the trailing clock.

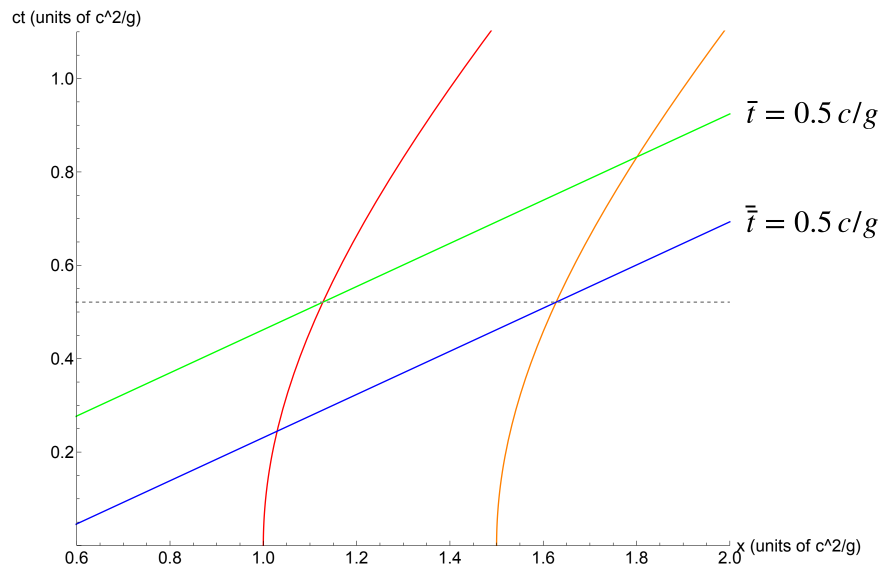


Figure 2: Worldline of two accelerating observers. Both feel acceleration g, and both are initially at rest in the coordinates (ct, x). The trailing observer (red curve) uses the time coordinate $\bar{t}$; the leading observer (orange curve) uses the time coordinate $\bar{\bar{t}}$. We show two surfaces of constant time according to the MCRF of the two observers. The green line shows the $\bar{t} = 0.5\, c/g$ surface; this corresponds to the MCRF of the trailing (red) when $\tau_{\text{trailing}} = 0.5\, c/g$. The blue line shows the $\bar{\bar{t}} = 0.5\, c/g$ surface; it corresponds to the MCRF of the leading (orange) worldline when $\tau_{\text{leading}} = 0.5\, c/g$. The constant $\bar{t}$ surface intersects the red worldline at $t \simeq 0.521\, c/g$; the constant $\bar{\bar{t}}$ surface intersects the orange worldline at the same value of t. (The dashed gray line is a constant at $t = 0.521\, c/g$.) These surfaces tell us that **the leading clock** (i.e., the clock of the observer at larger x) **runs fast** compared to the trailing clock. Surfaces of constant $\bar{\bar{t}}$ are consistently in the past of surface of constant $\bar{t}$, meaning that a particular value of $\bar{\bar{t}}$ has already happened by the time $\bar{t}$ reaches that value. Both observers agree that the trailing clock is slower than the leading clock.

By borrowing some results from the discussion below of "Rindler coordinates," we can compute the precise amount by which the leading clock runs ahead of the trailing clock, at least when the speeds of the two accelerated observers in the inertial coordinate frame is small compared to light. Let us write down the worldlines of the trailing and leading observers as seen in the inertial coordinate system:

$$ct_T = \frac{c^2}{g}\sinh(g\bar{t}/c)\ , \qquad x_T = \frac{c^2}{g}\cosh(g\bar{t}/c)\ ; \tag{14.6}$$

$$ct_L = \frac{c^2}{g}\sinh(g\bar{\bar{t}}/c)\ , \qquad x_L = \frac{c^2}{g}\cosh(g\bar{\bar{t}}/c) + L\ . \tag{14.7}$$

Let us also write down how one represents a single slice of $\bar{t} =$ constant in the MCRF of the trailing observer:

$$ct_{\mathrm{MCRF},T} = x\ \tanh(g\bar{t}/c)\ . \tag{14.8}$$

This relationship is worked out in the detailed discussion and derivation of Rindler coordinates, which is developed in the more advanced material presented below.

The question we'd like to answer is: What is the value of $\bar{\bar{t}}$ when the time on the constant $\bar{t}$ slice crosses the worldline of the leading observer — in other words, what is $\bar{\bar{t}}$ when $ct_{\mathrm{MCRF},T} = ct_L$? Plugging in the various definitions yields the equation we must solve:

$$\left. ct_{\mathrm{MCRF},T}\right|_{x=x_L} = ct_L\ , \tag{14.9}$$

which means

$$\left[\frac{c^2}{g}\cosh(g\bar{\bar{t}}/c) + L\right]\tanh(g\bar{t}/c) = \frac{c^2}{g}\sinh(g\bar{\bar{t}}/c) \tag{14.10}$$

or

$$\left[\cosh(g\bar{\bar{t}}/c) + \frac{gL}{c^2}\right]\tanh(g\bar{t}/c) = \sinh(g\bar{\bar{t}}/c)\ . \tag{14.11}$$

We now need to solve Eq. (14.11) for $\bar{\bar{t}}$ as a function of $\bar{t}$. Remarkably, this isn't so hard to do, as long as a certain approximation holds. Begin by putting all of the terms that depend on $\bar{t}$ on the left-hand side, and all of the terms that depend on $\bar{\bar{t}}$ on the right:

$$\begin{aligned}\tanh(g\bar{t}/c) &= \frac{\sinh(g\bar{\bar{t}}/c)}{\cosh(g\bar{\bar{t}}/c) + gL/c^2} \\ &\simeq \tanh(g\bar{\bar{t}}/c)\left[1 - \frac{gL}{c^2\cosh(g\bar{\bar{t}}/c)}\right]\ .\end{aligned} \tag{14.12}$$

The approximation introduced here is accurate as long as $gL/c^2 \ll \cosh(g\bar{\bar{t}}/c)$; recalling that c^2/g is roughly 1 light-year for an acceleration $g = 9.8\,\mathrm{m/s}^2$, this is clearly reasonable as long as L is anything much smaller than a light year.. Taking the arc-hyperbolic tangent of both sides, and using the result[1]

$$\mathrm{arctanh}\left[\tanh(x)(1-\epsilon)\right] \simeq x - \cosh(x)\sinh(x)\epsilon\ , \tag{14.13}$$

we find

$$g\bar{t}/c = g\bar{\bar{t}}/c - \frac{gL}{c^2}\sinh(g\bar{\bar{t}}/c)\ . \tag{14.14}$$

For general values of $\bar{\bar{t}}$, this isn't too easy to work with. However, if we confine ourselves to $g\bar{\bar{t}}/c \ll 1$, then this simplifies very nicely: using $\sinh(x) \simeq x$ for $x \ll 1$, Eq. (14.14) becomes

$$\bar{t} = \bar{\bar{t}}\left(1 - \frac{gL}{c^2}\right)\ . \tag{14.15}$$

[1]Figuring out things like this is a good use for tools like Mathematica.

The leading clock ticks at a faster rate than the trailing clock:

$$\frac{\bar{\bar{t}} - \bar{t}}{\bar{\bar{t}}} = \frac{gL}{c^2} \; . \tag{14.16}$$

Remember this nice, clean result! We will soon see a similar form when examining a different quantity, and rediscover this result in another context in a few lectures.

14.3 Light measured by the two accelerated observers

A related calculation compares the properties of light as measured by the two observers. This is particularly important because light plays such a critical role in relativity, since we often exploit the fact that its speed is c in all reference frames. Let's imagine that a beam of light travels in the $+x$ direction. It first intersects the trailing observer's worldline, then continues and later intersects the leading observer's worldline. The question we want to know is: What is the energy that the two observers measure for this light?

We will do all of these calculations in the inertial frame, which provides a convenient "stage" for us to formulate the quantities that we need for this analysis. We will also use the fact that, given something with 4-momentum $\vec{p}$, an observer whose 4-velocity is $\vec{u}$ measures it to have energy $E = -\vec{p} \cdot \vec{u}$.

Begin by writing the components of the light's 4-momentum in the inertial frame as

$$p^t = h\nu/c \; , \qquad p^x = h\nu/c \; . \tag{14.17}$$

(The y and z components of the light's 4-momentum are zero.) Let us say that this light crosses the worldline of the trailing observer when that observer's clock reads $\bar{t}_{\rm beam}$. Their 4-velocity at that time has components in the inertial frame

$$u^t_T = c\,\cosh(g\bar{t}_{\rm beam}/c) \; , u^x = c\,\sinh(g\bar{t}_{\rm beam}/c) \; . \tag{14.18}$$

The energy that the trailing observer measures for the light is then given by

$$\begin{aligned} E_T &= -\vec{p} \cdot \vec{u}_T & (14.19)\\ &= h\nu \cosh(g\bar{t}_{\rm beam}/c) - h\nu \sinh(g\bar{t}_{\rm beam}/c) & (14.20)\\ &= h\nu \cosh(g\bar{t}_{\rm beam}/c)\left(1 - \tanh(g\bar{t}_{\rm beam}/c)\right) \; . & (14.21) \end{aligned}$$

This can be simplified a bit more using a few hyperbolic function identities:

$$\cosh(x) = \frac{1}{\sqrt{\mathrm{sech}^2(x)}} = \frac{1}{\sqrt{1 - \tanh^2(x)}} \; . \tag{14.22}$$

Using this, we see that the energy measured by the trailing observer is

$$E_T = h\nu \sqrt{\frac{1 - \tanh(g\bar{t}_{\rm beam}/c)}{1 + \tanh(g\bar{t}_{\rm beam}/c)}} \; . \tag{14.23}$$

Notice that this is exactly the Doppler shift that one expects for an observer who is moving away from a light source with 3-speed $v = c\tanh(g\bar{t}_{\rm beam}/c)$.

The light continues to move in the $+x$ direction, and crosses the worldline of the leading observer when their clock reads $\bar{\bar{t}}_{\rm beam}$. By a similar calculation, the energy that the leading observer measures is

$$E_L = h\nu\sqrt{\frac{1-\tanh(g\bar{\bar{t}}_{\rm beam}/c)}{1+\tanh(g\bar{\bar{t}}_{\rm beam}/c)}}\ , \tag{14.24}$$

which is likewise just the Doppler-shifted energy for a speed $v = c\tanh(g\bar{\bar{t}}_{\rm beam}/c)$.

We'd like to compare E_T to E_L. To do so, we must relate the time $\bar{\bar{t}}_{\rm beam}$ at which the light beam crosses the leading observer's worldline to the time $\bar{t}_{\rm beam}$ at which the beam crosses the trailing observer's worldline. We do this by using our results describing time in the inertial frame to the times along the worldline.

The inertial-frame time at which the light crosses the trailing observer's worldline is

$$t_T = \frac{c}{g}\sinh(g\bar{t}_{\rm beam}/c)\ ; \tag{14.25}$$

the inertial-frame time at which it crosses the leading observer's worldline is

$$t_L = \frac{c}{g}\sinh(g\bar{\bar{t}}_{\rm beam}/c)\ . \tag{14.26}$$

However, we also know that, in the inertial frame, the light moves a distance of L in going from the trailing observer to the leading observer, plus the additional distance that the leading observer covers while the light is in transit:

$$\begin{aligned} t_L &= t_T + \frac{L}{c} + \int_{t_T}^{t_L}\frac{dx}{dt}dt \\ &= t_T + \frac{L}{c} + c\int_{t_T}^{t_L}\tanh(g\bar{\bar{t}}/c)\,dt\ . \end{aligned} \tag{14.27}$$

The integral on the last line accounts for the distance that the leading observer moves as the light is in transit. As written, it is not a very nice integral: we do the integral with respect to the inertial-frame time, but the function we are integrating is parameterized using time $\bar{\bar{t}}$ along that observer's worldline. So we, need to convert: using Eq. (14.3) (with $t_0 = 0$, and with $\tau = \bar{\bar{t}}$), we have

$$\bar{\bar{t}} = \frac{c}{g}\mathrm{arcsinh}(gt/c)\ , \tag{14.28}$$

and the argument of the integral becomes

$$\begin{aligned} \tanh(g\bar{\bar{t}}/c) &= \tanh(\mathrm{arcsinh}(gt/c)) \\ &= \frac{(gt/c)}{\sqrt{1+(gt/c)^2}}\ . \end{aligned} \tag{14.29}$$

It's kind of miraculous that this result cleans up so nicely. We can now easily do the integral and relate t_L to t_T:

$$t_L = t_T + \frac{L}{c} + \frac{c}{g}\left(\sqrt{1+(gt_L/c)^2} - \sqrt{1+(gt_T/c)^2}\right)\ . \tag{14.30}$$

We now have all the information we need, in principle, to see how the energy of the light changes as it goes from the trailing observer to the leading one:

1. Solve Eq. (14.30) to find t_L as a function of t_T.

2. Using this solution plus Eqs. (14.24) and (14.26), compute the energy measured by the leading observer as a function of t_T.

3. Using Eq. (14.25) and (14.23), compute the energy measured by the trailing observer as a function of t_T.

Unfortunately, this "recipe" involves a multitude of hyperbolic functions and does not yield a nice closed form answer. To get something tractable, let's assume that gt/c, $g\bar{t}/c$, and $g\bar{\bar{t}}/c$ are all much smaller than 1, and use the limiting forms

$$\cosh(x) \simeq 1 , \qquad \sinh(x) \simeq x , \qquad \tanh(x) \simeq x \qquad \text{when } x \ll 1 . \tag{14.31}$$

Doing so, we find

$$t_L \simeq t_T + \frac{L}{c} , \tag{14.32}$$

$$t_T \simeq \bar{t}_{\rm beam} , \quad t_L \simeq \bar{\bar{t}}_{\rm beam} , \tag{14.33}$$

$$E_T \simeq h\nu\sqrt{\frac{1-(g\bar{t}_{\rm beam})/c}{1+(g\bar{t}_{\rm beam})/c}} \simeq h\nu\left(1 - g\bar{t}_{\rm beam}/c\right) , \tag{14.34}$$

$$E_L \simeq h\nu\sqrt{\frac{1-(g\bar{\bar{t}}_{\rm beam})/c}{1+(g\bar{\bar{t}}_{\rm beam})/c}} \simeq h\nu\left(1 - g\bar{\bar{t}}_{\rm beam}/c\right) . \tag{14.35}$$

Putting all these together, we see that

$$\Delta E \equiv E_T - E_L \simeq h\nu\left(\frac{gL}{c^2}\right) . \tag{14.36}$$

The light's energy as measured by the leading observer is lower than the energy measured by the trailing observer, by a fractional amount that precisely matches the rate at which their clock ticks faster than the trailing observer's clock.

14.4 Wrapup: Key things to take away

The calculations that went into the above discussion were somewhat dense, so this is a good point to pause and assess the key lessons that we should take away from it. In particular, we want to emphasize aspects of what is observed by a pair of observers who share the same acceleration $\mathbf{g}$, but are spatially separated by a distance L.

- Even if the observers start out with their clocks perfectly synchronized, they will fall out of synchrony as time passes, with the leading clock running fast by a factor gL/c^2.

- If light is exchanged between the two observers, they will disagree on its energy. The leading observer measures it to have a lower energy (i.e., they see the light as being somewhat redder), by a factor gL/c^2.

As our analysis showed, the numerical factor gL/c^2 that emerges from these analyses is an approximate one, but works well as long as $g(\text{time})/c$ is small for all the versions of "time" under consideration. Bear in mind that $c/g \simeq 1\,\text{year}$ if $g = 10\,\text{m/s}^2$; this gives a sense of the time and lengthscales involved before these approximations start to break down.

14.5 Rindler coordinates (CAUTION: somewhat advanced material)

Parts of the discussion in the preceding few sections rely on more advanced material which we present here. We recommend that you read these notes, but you should not be worried if you do not follow every detail of this discussion. The nature of the Rindler coordinates, Eqs. (14.37)–(14.40), and the subsection labeled "Features of the Rindler representation" are particularly worth your attention.

In almost all of our discussion so far, we have used coordinates (t, x, y, z) or (ct, x, y, z) that are particularly well suited for describing inertial observers. Indeed, such coordinates are often called *inertial coordinates*: they are ones for which there exists some set of observers who sit at constant (x, y, z). In such a frame, the observers are only "moving" in time. There are also many observers who move with constant velocity. The worldlines of the constant velocity observers are lines in these coordinates, taking the form $x = x_0 + v^x t$, and similarly for their motion in y and z.

Even when we discussed accelerating observers, we presented their motion as seen by some inertial observer who sees the accelerating observer zoom past. You might wonder — how does the accelerating observer describe spacetime? Do we learn anything useful by developing coordinates that are "adapted" to the reference frame of the acelerating observer? To do this, one could imagine performing Lorentz transformations that flip between a particular inertial frame (e.g., the frame used to draw the time axes in Fig. 1) and the accelerating observer's MCRF. However, the relative velocity of the MCRF and any given inertial observer is continually changing. The Lorentz transformations that enact this"flipping back and forth" thus must continually evolve, which limits their usefulness for us.

A coordinate system which nicely describes an accelerating observer in fact can be written down. These coordinates (named *Rindler coordinates*, in honor of Wolfgang Rindler who did much to explore their properties and applications) are described and explored in this section. The following section derives Rindler coordinates; that section should be considered even more advanced than this one. Students should feel free to ignore it altogether.

Let us choose the initial condition of the accelerated observer's trajectory so that $t_0 = 0$ and $x_0 = c^2/g$ in Eqs. (14.3) and (14.4). Then, as we derive in detail in the following section, the accelerated observer uses coordinates $(c\bar{t}, \bar{x}, \bar{y}, \bar{z})$ to describe spacetime. These new coordinates are related to the original "inertial" coordinates (ct, x, y, z) according to

$$ct = \bar{x}\sinh(g\bar{t}/c) \ , \tag{14.37}$$
$$x = \bar{x}\cosh(g\bar{t}/c) \ , \tag{14.38}$$
$$y = \bar{y} \ , \tag{14.39}$$
$$z = \bar{z} \ . \tag{14.40}$$

In the barred coordinate system, the accelerated observer is at constant spatial coordinate $(\bar{x}, \bar{y}, \bar{z}) = (c^2/g, 0, 0)$; the barred time coordinate $\bar{t}$ is exactly the same as the proper time τ that this observer measures. Notice that this solution agrees with Eqs. (14.3) and (14.4) when $\bar{x} = c^2/g$. Equations (14.37)–(14.40) define the Rindler coordinates. (Notice also that Eq. (14.37) is what we used to define the constant time surfaces of the MCRF as shown in Fig. 1 and in the associated discussion.)

Figure 3 illustrates how the $(c\bar{t}, \bar{x})$ coordinates used by an accelerating observer appear in the reference frame of an unaccelerated observer. The red curve illustrates the worldline of the observer who starts at $x = \bar{x} = c^2/g$. The green lines represent surfaces of constant $\bar{t}$; the blue hyperbolic curves represent trajectories of constant $\bar{x}$. Those trajectories are chosen by requiring that $\bar{x} = x$ when $t = \bar{t} = 0$, and by demanding that the unit vector along $\bar{x}$ be spacetime orthogonal to the unit vector along $\bar{t}$. Notice that each constant $\bar{x}$ coordinate can itself be regarded as an accelerated observer; as we discuss in the next section, it can be shown that the observer at constant $\bar{x}$ feels an acceleration $\mathbf{a} = (c^2/\bar{x})\mathbf{e}_x$.

We also include in this figure the trajectory of a light ray that is emitted from the origin; we discuss some interesting features of this coordinate system's behavior with respect to this light ray below.

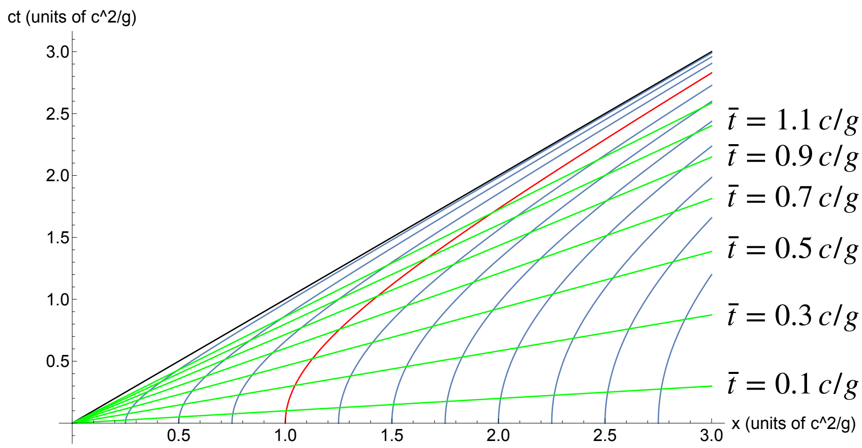


Figure 3: An illustration of Rindler coordinates. The red curve is the worldline of an accelerating observer who starts at $x = c^2/g$ and experiences constant acceleration g. The green lines are surface of constant $\bar{t}$, which coincides at that observer's location with their own proper time; the blue curves are trajectories of constant $\bar{x}$, chosen to coincide with the unaccelerated frame's x when $t = \bar{t} = 0$. A heavy black line $ct = x$ illustrates a light ray that is emitted from the origin and moves to the right. Notice both axes are in units of c^2/g.

14.5.1 Features of the Rindler representation

There are two features of the Rindler representation to which we would like to particularly call your attention.

- **A new form for the metric:** By now, we know very well that

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 \equiv \eta_{\alpha\beta} dx^\alpha \, dx^\beta \, . \tag{14.41}$$

  The invariance of this interval is what led us to the metric used in inertial coordinates, $\eta_{\alpha\beta} = \mathrm{diag}(-1, 1, 1, 1)$.

Let's look at this in our new coordinates. Considering Eqs. (14.37)–(14.38), we see

$$c\,dt = (d\bar{x})\sinh(g\bar{t}/c) + \frac{g\bar{x}}{c^2}(c\,d\bar{t})\cosh(g\bar{t}/c)\ , \tag{14.42}$$

$$dx = (d\bar{x})\cosh(g\bar{t}/c) + \frac{g\bar{x}}{c^2}(c\,d\bar{t})\sinh(g\bar{t}/c)\ , \tag{14.43}$$

plus $dy = d\bar{y}$, $dz = d\bar{z}$. This tells us that

$$\begin{aligned} ds^2 &= -\left[(d\bar{x})\sinh(g\bar{t}/c) + \left(\frac{g\bar{x}}{c^2}\right)(c\,d\bar{t})\cosh(g\bar{t}/c)\right]^2 \\ &\qquad + \left[(d\bar{x})\cosh(g\bar{t}/c) + \left(\frac{g\bar{x}}{c^2}\right)(c\,d\bar{t})\sinh(g\bar{t}/c)\right]^2 + d\bar{y}^2 + d\bar{z}^2 \\ &= -\left(\frac{g\bar{x}}{c^2}\right)^2 c^2 d\bar{t}^2 + d\bar{x}^2 + d\bar{y}^2 + d\bar{z}^2\ . \end{aligned} \tag{14.44}$$

(We used $\cosh^2(g\bar{t}/c) - \sinh^2(g\bar{t}/c) = 1$.) Notice that the metric *is not a constant* in this representation. Because we reserve the symbol $\eta_{\alpha\beta}$ for $\mathrm{diag}(-1, 1, 1, 1)$, we now use $g_{\alpha\beta}$ to denote the metric. In particular, we now have

$$g_{\alpha\beta} \doteq \mathrm{diag}\left(-(g\bar{x}/c^2)^2, 1, 1, 1\right) \tag{14.45}$$

for the metric of spacetime in Rindler coordinates.

It's worth emphasizing that we *are still doing special relativity*; we have only changed coordinates. If you've been reading ahead or poking at references, you may have seen that in general relativity we get metrics in which the components are functions, and so you might worry that we've somehow "broken" special relativity. We haven't: in some coordinate systems, the components of the metric are functions and yet the metric still describes special relativity. This is an example of such a system.

- **A "horizon":** Notice in Fig. 3 that we have included a light ray that starts at the origin and travels in the $+x$ direction. On an upcoming problem set, you will compare the motion of the accelerated observer to the motion of this light ray, and show that the light ray *never* crosses this observer's trajectory. The light ray asymptotically approaches the accelerated observer's trajectory as $\bar{t} \to \infty$, but they never cross. In fact, the light ray never crosses *any* of the constant $\bar{x}$ trajectories.

  Because information can travel no faster than light, this means that there is a region of spacetime that *cannot communicate with the accelerated observer.* No signal sent by an observer to the "left" of that light ray can reach the accelerated observer. We say that there is a *horizon* separating the events which can communicate with the accelerated observer from those events which cannot so communicate.

  We will come back to the notion of horizons later in this course. Take this as a preview of some of the interesting features that we will begin to find as we start investigating certain spacetimes.

14.6 Derivation of Rindler coordinates (CAUTION: advanced material)

The discussion in this section is significantly more advanced than is expected for 8.033 students. It is included in order to provide a complete explanation of where the Rindler coordinates come from, as well as for the benefit of any students who are interested in diving somewhat deeper into this material; it will *not* be discussed in detail during lecture.

We now define coordinates $c\bar{t}$, $\bar{x}$ which the accelerating observer uses to describe spacetime. (Since the acceleration is along x, we simply put $\bar{y} = y$ and $\bar{z} = z$ and are then done with those two coordinates.) We take the accelerating observer's coordinates to be $t = 0$, $x = c^2/g$ when $\tau = 0$, and we use the symbols T, X to define the accelerating observer's trajectory as measured by the observer who is at rest with respect to the accelerating observer at $\tau = 0$. The motion of this observer is thus

$$cT(\tau) = \frac{c^2}{g}\sinh(g\tau/c) \ , \qquad X(\tau) = \frac{c^2}{g}\cosh(g\tau/c) \ . \tag{14.46}$$

For the accelerated observer, their own proper time τ makes a natural clock. Given this, it is natural that the accelerated observer chooses the time coordinate to be $\bar{t} = \tau$ along their own worldline.

Can we use this coordinate $\bar{t}$ *away* from the observer's worldline? In other words, can the accelerating observer use $\bar{t}$ to label events elsewhere in spacetime, away from their own worldline? Yes, by the following procedure:

- First define unit vectors that point along the directions $\bar{t}$ and $\bar{x}$. Making such a unit vector for $\bar{t}$ is not hard: in the accelerating observer's MCRF, their 4-velocity has components $u^\alpha \doteq (c, 0, 0, 0)$. A natural choice for $\vec{e}_{\bar{t}}$ is thus parallel to this observer's 4-velocity, so we put

$$\vec{e}_{\bar{t}} = \frac{1}{c}\vec{u} = \cosh(g\bar{t}/c)\vec{e}_t + \sinh(g\bar{t}/c)\vec{e}_x \ . \tag{14.47}$$

  We then define $\vec{e}_{\bar{x}}$ by requiring that it be orthogonal to $\vec{e}_{\bar{t}}$ (and also that it have no components along $\bar{y}$ and $\bar{z}$):

$$\vec{e}_{\bar{x}} = \sinh(g\bar{t}/c)\vec{e}_t + \cosh(g\bar{t}/c)\vec{e}_x \ . \tag{14.48}$$

- With $\vec{e}_{\bar{x}}$ defined, now consider a "surface" of constant $\bar{t}$ (i.e., a set of events in which all the time coordinates $\bar{t}$ are the same). Such a surface must lie on a line that is parallel $\vec{e}_{\bar{x}}$, meaning that it is a line whose slope m is given by

$$m = \frac{e^t_{\bar{x}}}{e^x_{\bar{x}}} = \tanh(g\bar{t}/c) \ . \tag{14.49}$$

  We further require that this line intersect the worldline of the accelerating observer: The line must have the slope m defined by Eq. (14.49), and pass through the point $[cT(\bar{t}), X(\bar{t})]$. With a little algebra we see that this line is given by

$$ct = x\tanh(g\bar{t}/c) \ . \tag{14.50}$$

We've now learned how to draw surfaces of constant $\bar{t}$ in the inertial (ct, x) coordinate frame. How do we draw a surface of constant $\bar{x}$? Such a surface must lie parallel to the timelike vector $\vec{e}_{\bar{t}}$ given in Eq. (14.47). This vector is continually changing in slope as $\bar{t}$ changes; in the inertial frame, it has slope

$$\frac{dx}{dt} = c\tanh(g\bar{t}/c) \; . \tag{14.51}$$

We have already deduced that t and $\bar{t}$ are related by Eq. (14.50). Combining these results, we see that

$$\frac{dx}{dt} = c^2\frac{t}{x} \; . \tag{14.52}$$

We integrate this up, applying an initial condition that the coordinates of the accelerated observer match those of the inertial frame at $t = \bar{t} = 0$:

$$\int_{\bar{x}}^{x} x\, dx = c^2 \int_0^t t\, dt \tag{14.53}$$

or

$$x^2 - \bar{x}^2 = c^2t^2 \; . \tag{14.54}$$

This tells us that surfaces of constant $\bar{x}$ are given by hyperbolae in the (ct, x) plane which satisfy

$$\bar{x}^2 = x^2 - (ct)^2 \; . \tag{14.55}$$

We'd like to massage Eqs. (14.50) and (14.55) a bit more to really isolate how $(c\bar{t}, \bar{x})$ appear in the inertial frame. Notice that Eq. (14.55) is solved by any pair of functions of the form

$$x = \bar{x}\cosh(\alpha) \; , \qquad ct = \bar{x}\sinh(\alpha) \; . \tag{14.56}$$

Applying this to Eq. (14.50), we see that we must have $\alpha = g\bar{t}/c$. We thus at last have the complete mapping of the accelerated observer's reference frame into the inertial coordinate system:

$$ct = \bar{x}\sinh(g\bar{t}/c) \; , \qquad x = \bar{x}\cosh(g\bar{t}/c) \; , \qquad y = \bar{y} \; , \qquad z = \bar{z} \; . \tag{14.57}$$

One final detail: it was noted earlier in these notes that an observer at constant $\bar{x}$ is itself an accelerated observer. This is hopefully intuitively obvious from the shape of the constant $\bar{x}$ surfaces in Fig. 3 (if they were not accelerated, they would not curve). What acceleration does this observer feel? This is most easily calculated by computing the 3-acceleration of this observer at $t = \bar{t} = 0$. Because at this moment all of the constant $\bar{x}$ observers happen to be momentarily at rest, all of these observers have 4-velocity with components $(c, 0, 0, 0)$ and 4-acceleration $(0, a^x, 0, 0)$ in this frame, where $a^x = d^2x/dt^2$ at $t = 0$.

Let's compute this:

$$\begin{aligned}
a^x &= \frac{d^2x}{dt^2}_{t=\bar{t}=0} \\
&= \left[\frac{d^2x}{d\bar{t}^2}\left(\frac{dt}{d\bar{t}}\right)^{-2}\right]_{t=\bar{t}=0} \\
&= \left[\left(\frac{g^2}{c^2}\bar{x}\cosh(g\bar{t}/c)\right)\left(\frac{g}{c}\frac{\bar{x}}{c}\cosh(g\bar{t}/c)\right)^{-2}\right]_{t=\bar{t}=0} \\
&= \frac{c^2}{\bar{x}} \; .
\end{aligned} \tag{14.58}$$

So the observer at $\bar{x} = c^2/g$ feels an acceleration of precisely g; those at larger $\bar{x}$ feels less acceleration, and those at smaller $\bar{x}$ feel more (with the acceleration diverging as $\bar{x} \to 0$).

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.033 FALL 2024

LECTURE 15
INTRODUCTION TO GRAVITY

15.1 The road ahead

In this set of lecture notes, we are going to begin to examine how to incorporate gravity into relativity. We will be concerned with two major questions:

- First, how is gravity "made," broadly speaking? In other words, given some body that generates a gravitational field in Newtonian physics, how do we describe that body's gravity using relativistic physics?
- Second, how do we describe a body's motion under the influence of gravity? How does what we think of as the "gravitational force" act in relativistic physics?

You might imagine that, given all we have done so far, addressing these points shouldn't be too difficult. After all, we reformulated both electric and magnetic forces and fields into nicely covariant relativistic language. How much harder can this be for gravity?

As we'll begin to see in the next section of these notes, gravity introduces complications that make describing it *substantially* more difficult. Indeed, going through all the details in great rigor is far beyond the scope of 8.033. We will content ourselves in this class with a more descriptive analysis, seeing how it is that the tricks we've learned so far don't work for gravity. We will then examine a high-level synopsis of how we proceed to answer the first of the two questions above. Going beyond that high-level synopsis takes roughly half the term of 8.962. Students who wish to pursue this subject further are encouraged to look into the course 8.228 (offered during IAP), and perhaps to consider taking 8.962 at some point down the road.

Once we have this high-level synopsis of how gravity arises, it isn't beyond 8.033 to describe how that gravity acts on a body. Exploring how relativistic gravity acts and how it differs from Newtonian gravity will be a big part of what we do in the last few weeks of this term. To get there, we first need to establish some important principles.

15.2 The principle of maximum aging

Imagine that two bodies travel from event A, located at $x = 0$, $t = 0$ to event B located at $x = 1$ lightsecond, $t = 4$ seconds. One body moves there at constant velocity $\mathbf{v} = 0.25c\,\mathbf{e}_x$; we'll call this the "direct" path. The other body moves first to event C, located at $x = 0$, $t = 2$ seconds. It then moves off to event C at half the speed of light[1]. We illustrate this situation in Fig. 1.

[1] In reality, the body must accelerate for some interval to reach this speed. For this initial discussion, we idealize the interval over which the acceleration occurs to be so short that it is nearly instantaneous.

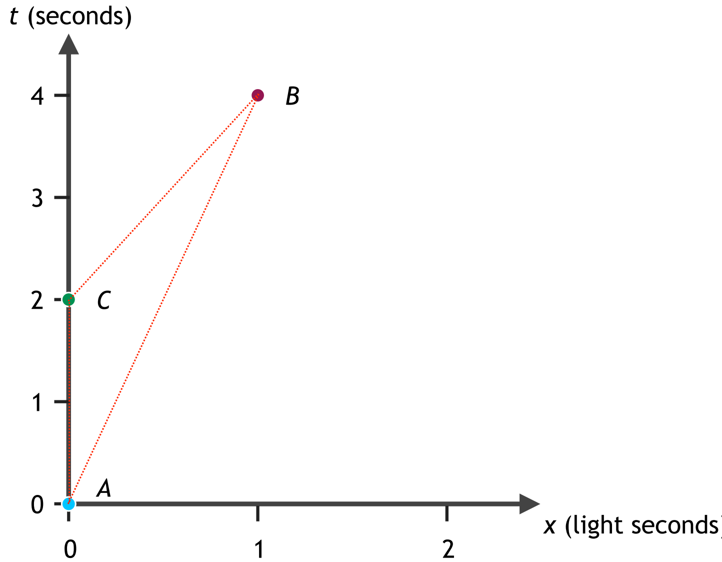


Figure 1: Two paths from event A to event B. The "direct" path goes from A to B at constant velocity; the indirect path goes from A to B via the event C. Note that different scales are used for the x and t axes.

Question: On which path does the body age more, the direct one or the indirect one? We've already discussed a similar situation when talked about the twin paradox, but just to remind ourselves how this works let's step through the analysis. We are going to use the fact that along any timelike trajectory,

$$\Delta s^2 = -c^2 \Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2 = -c^2 \Delta\tau^2 \ . \tag{15.1}$$

The last equality follows from the fact that $\Delta\tau$ is the time experienced in the body's own rest frame; in that frame, $\Delta x = \Delta y = \Delta z = 0$, since the body is at rest in its rest frame. Let's use this to compute how much $\Delta\tau$ the bodies experience along these two trajectories.

First, consider the direct trajectory:

$$\begin{aligned}\Delta\tau &= \left[(\Delta t)^2 - (\Delta x)^2/c^2\right]^{1/2} \\ &= \left[16\,\mathrm{sec}^2 - 1\,\mathrm{sec}^2\right]^{1/2} \\ &= \sqrt{15}\,\mathrm{seconds}\ .\end{aligned} \tag{15.2}$$

The body ages a total of $\sqrt{15} \simeq 3.87$ seconds on the direct trajectory.

Next, the indirect trajectory. We break this up into two pieces:

$$\Delta\tau_{A\to C} = 2\,\mathrm{seconds}\ . \tag{15.3}$$

$$\begin{aligned}\Delta\tau_{C\to B} &= \left[4\,\mathrm{sec}^2 - 1\,\mathrm{sec}^2\right]^{1/2} \\ &= \sqrt{3}\,\mathrm{seconds}\ .\end{aligned} \tag{15.4}$$

So on the indirect trajectory, the body ages a total of $2 + \sqrt{3} \simeq 3.73$ seconds. Comparing with the direct trajectory aging of 3.87 seconds, we see that *the body ages more on the unaccelerated trajectory.*

Without too much effort, we can find other trajectories in which the aging is less — *much* less, if we design the trajectory well. Consider, for example, the trajectory shown in Fig. 2:

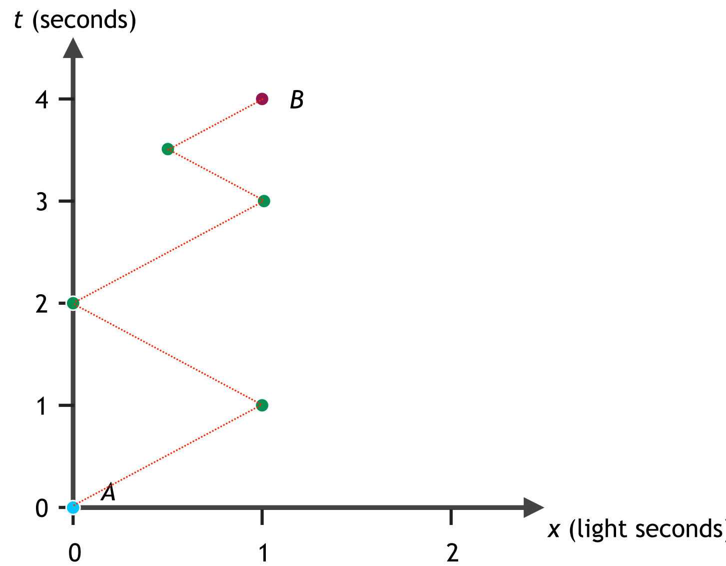


Figure 2: Yet another path from event A to event B. On this path, the body zig zags back and forth at nearly light speed until it finally reaches event B.

The path here zips back and forth at nearly light speed. As such, the body accumulates nearly *zero* proper time along each leg, and so it *does not age at all* in moving from event A to event B. Consistent with how we resolved the twin paradox, we can see that acceleration reduces the aging which a body experiences as it moves through spacetime.

We will soon briefly discuss a topic called the *calculus of variations*. Some of you may have already learned about this — it is the key technique which underlies Lagrangian mechanics, for example. Using the calculus of variations, we can meaningfully pose the following question: "Given the infinite number of timelike trajectories in spacetime which connect events A and B, along which one does the body age *the most*? In other words, which trajectory through spacetime is the one which corresponds to maximum aging?"

The answer we will find is that the trajectory of maximum aging is indeed the unaccelerated trajectory. This will prove to be very, very useful for us. Picking out the "trajectory of maximum aging" to understand the motion of a body in special relativity is overkill; it is fine for understanding how this technique operates, but it isn't how you want to calculate a body's trajectory through special relativity's spacetime on an everyday basis. However, we will argue (using an important principle that Einstein introduced to understand how to incorporate gravity into relativity) that this technique is exactly what we need to compute a body's motion under gravity once we start making a relativistic theory of gravity.

For now, please file away in some mental storage drawer the idea that "no acceleration" means "maximum aging" as a body moves through spacetime. We will want to return to this point in several lectures.

15.3 Making Newton's gravity relativistic?

Long ago, Isaac Newton taught us that two masses feel a force that is proportional to their masses, inversely proportional to the square of the distance between them, directed along the line between the two masses, and attractive:

$$\mathbf{F}^{\mathrm{G}} = -G\frac{m_1 m_2}{r^2}\mathbf{e}_r \; . \tag{15.5}$$

This looks just like Coulomb's law, which tells us about the electric force between two charges:

$$\mathbf{F}^{\mathrm{E}} = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2}\mathbf{e}_r \; . \tag{15.6}$$

The differences are that the electric force arises from charges q rather than masses m; the electric force can be attractive or repulsive, depending on the signs of q_1 and q_2 (note that masses are always positive); and the two forces have different "coupling constants" (G versus $1/4\pi\epsilon_0$). Given that we were able to put the electric force into fully relativistic form without too much effort, with magnetic fields and forces getting wrapped up in the final form, can we perhaps do the same thing for gravity?

Although this seems like a plausible course of action, it is important to recognize a major difference between the two force laws. A key issue is that *the charges q_1 and q_2 which enter into the electric force are Lorentz invariants.* Do all observers agree on the masses m_1 and m_2 which enter into the gravitational force?

The issue is certainly yes **if** the Newtonian force only acts on *rest mass*. If that's the case, though, then there is an interesting consequence: gravity can have no effect on anything massless, such as light. It is also not hard to construct "though experiments" which suggest that gravity would act a little oddly.

Consider a box of mass M. Inside this box are lumps of putty, each of rest mass m. If gravity only acts on rest mass, then on the surface of the Earth, this box will have weight

$$\mathbf{F}_w = (M + 2m)\mathbf{g} \; , \tag{15.7}$$

where

$$\mathbf{g} = \frac{GM_E}{R_E^2}\mathbf{e}_d \tag{15.8}$$

is the gravitational acceleration at the surface of the Earth. It depends on the Earth's mass M_E, its radius R_E, and points down, $\mathbf{e}_d$, from the surface toward the Earth's center.

Let's imagine that the lumps of putty are in fact moving toward each other at speeds very close to the speed of light: the box has a long axis oriented parallel to the Earth's surface (let's call this along the x direction); one lump of putty has $\mathbf{u} = u\mathbf{e}_x$, the other has $\mathbf{u} = -u\mathbf{e}_x$, and each lump's speed u is close to c. Before the lumps of putty come into contact, the box's weight is given by Eq. (15.7). Afterwards, the box has weight

$$\mathbf{F}_w = (M + 2\gamma(u)m)\mathbf{g} \; . \tag{15.9}$$

If $u \sim c$, then $\gamma(u)$ can be huge. In such a case, the weight of the box very suddenly increases, perhaps by a large amount.

Or, imagine that one lump of putty is made of matter, and the other of antimatter. All evidence[2] indicates that antimatter responds to gravity just like "normal" matter. After the

[2]It's actually hard to gather such evidence, because antimatter tends to annihilate with regular matter before we can make a precise measurement.

two lumps collide, all of their rest mass is converted into radiation. If gravity only acts on rest mass, then the box now has weight

$$\mathbf{F}_w = M\mathbf{g} \tag{15.10}$$

after the collision; the weight very suddenly decreases. *If gravity only acts on rest mass, then an object's weight can very suddenly and discontinuously change.*

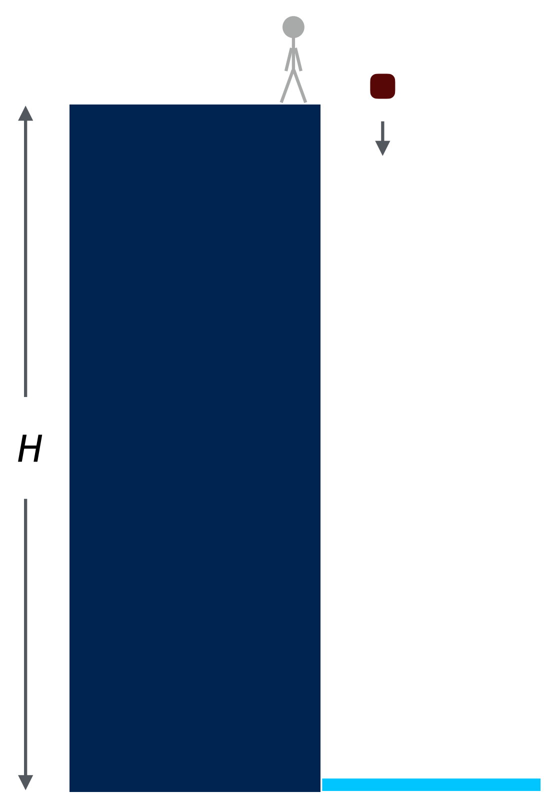
Let me emphasize that these thought experiments do *not* tell us that m_1 and m_2 cannot be rest masses. However, they make it clear that if they *are* rest mass, then we must be prepared for some potentially weird consequences. It is worth taking a few moments to think about alternatives.

What if gravity doesn't so much act on rest mass as it acts on *energy*? For instance, suppose that the m that appears in the Newtonian force law is really something like E/c^2. In the vast majority of situations that humanity has encountered in its history, an object's kinetic energy is a very tiny fraction of its rest energy. As such, the difference between m and E/c^2 has tended to be negligible. It is not unreasonable to imagine that gravity acts on energy, but that we inferred the force law (15.5) because of the overwhelming importance of rest energy at typical kinetic energy levels.

If gravity acts on *all* forms of energy, then it acts on light. Let's consider that possibility.

15.4 The action of gravity on light

We consider another thought experiment; this is a variation on one that was originally designed by Einstein.



- Imagine we stand on a tall building, and we drop a rock of mass m.

- After falling a distance H, the rock enters a device. At the moment it enters this device, it has energy

$$E = E_{\text{bottom}} = mc^2 + mgH \ . \tag{15.11}$$

- This device converts the rock into a single photon[3] of $E_{\text{bottom}} = mc^2 + mgH = h\nu$ (be careful not to confuse the height H with Planck's constant h) and launches this photon back up to the top of the building.

- When the photon has climbed a distance H, we use yet another amazing device to convert the photon back into a rock. *What energy must this rock have?*

Imagine first that gravity does not act on light. If that were the case, then the rock would reappear next to us with energy $E = E_{\text{bottom}} = mc^2 + mgH$ – it would either have a slightly larger rest mass, or else it would have some kinetic energy, and continue to climb. If we allowed it to come to a halt and then fall back down, on the next pass it would have energy $E = mc^2 + 2mgH$ at the end of this process. We can repeat this, giving the rock an extra mgH of energy on each go-round. *If gravity does not act on the light, then we can in principle make a device for creating unlimited amounts of energy*[4] *this way.*

Let's insist that energy be conserved: When the photon is converted back into a rock, it has an energy $E = E_{\text{top}} = mc^2$. Because a photon's energy is related to its frequency, this tells us that the photon loses energy as it climbs out of the gravitation field: it is *redshifted* according to the rule

$$\frac{E_{\text{top}}}{E_{\text{bottom}}} = \frac{h\nu_{\text{top}}}{h\nu_{\text{bottom}}} = \frac{mc^2}{mc^2 + mgH} \ , \tag{15.12}$$

or, using $gH \ll c^2$,

$$\frac{\nu_{\text{top}}}{\nu_{\text{bottom}}} = 1 - \frac{gH}{c^2} \ . \tag{15.13}$$

Notice that this frequency difference is *precisely* the same as the effect we found when we compared the energy of a photon that is measured by two accelerated observers; compare Sec. 14.3 of the previous set of lecture notes.

A few comments are worth making before moving on:

- The magnitude of this effect can be estimated by noting that, at the Earth's surface, $gH \simeq 100\,\text{m}^2/\text{s}^2\,(H/10\,\text{m})$, and by using $c^2 \simeq 9 \times 10^{16}\text{m}^2/\text{s}^2$. This tells us that we expect a frequency change in the light of roughly 1 part in 10^{15} for every 10 meters of height change.

[3] Alarm bells should be going off in your brain right now: Even allowing for the most amazing technology we can imagine, converting a single rock into a single photon would cause all sorts of problems with energy and momentum conservation. To address this, imagine dropping a rock and an "anti-rock" — a rock made of antimatter. The device can then create *two* photons; by mounting the device on the Earth, we can allow the Earth to recoil in such a way that both energy and momentum are conserved.

[4] The device used in this example is, by design, kind of silly. However, it is not hard to imagine making less silly variations on this. For example, by allowing matter and antimatter to fall in a gravitational field and then harvesting the light they produce upon annihilation, we could make any amount of energy we want, perhaps harvesting the energy by allowing those photons to heat up a bucket of water. The failure of gravity to act on light would be an on-ramp to building a perpetual motion machine.

- A more general form of Eq. (15.13) is

$$\frac{\nu_{\rm top}}{\nu_{\rm bottom}} = 1 - \frac{\Delta\Phi_G}{c^2} \, , \tag{15.14}$$

  where $\Delta\Phi_G$ is the change in gravitational potential between the two measurement points.

We emphasize these points because this effect in fact is exactly what we measure. It was first done in 1959 using Mössbauer spectroscopy by Robert Pound and Glen Rebka, looking at the effect of gravity on gamma rays which produced by the decay of the isotope ^{57}Fe and then climbed 22.5 meters up a tower at Harvard's Jefferson Laboratory. This measurement is now done millions of times a second by a huge number of people around the world, as it is integral to the functioning of the Global Position System. Without correcting for this frequency shift, GPS accuracy would degrade at a rate of roughly 8 meters per minute.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 16
The calculus of variations, the principle of maximum aging, and the motion of bodies in spacetime

16.1 Our goal

In Lecture 17, we will introduce a hypothesis which will allow us to formulate how a body moves in any spacetime, not just the spacetimes of special relativity. The key idea is to use the principle we recently discussed which argued that unaccelerated motion means bodies move along trajectories of "maximal aging." Any acceleration slows down their own clocks such that the proper time they accumulate during their motion is less than it would be on the non-accelerated path. The goal of today's lecture is to develop tools that allow us to use this principle.

16.2 Euler's equation

Imagine a function f that depends on time, on a position variable x, and on the derivative $\dot{x} \equiv dx/dt$. Suppose that f encodes something important about our physical situation. Suppose that at time t_i we are at position x_i, and that at time t_f we are at position x_f. Subject to the boundary condition that we must start at the event (t_i, x_i) and we must end at (t_f, x_f), we are free to take any trajectory $x(t)$ that connects these two events.

Imagine that the trajectory we actually take is the one that gives us the *extremum* of

$$J = \int_{t_i}^{t_f} f\left[x(t), \dot{x}(t); t\right]\, dt\ . \tag{16.1}$$

For example, f might tell us about the rate at which we age along the trajectory, and J could be the accumulated aging we experience. Of the infinite number of ways that we can connect (t_i, x_i) to (t_f, x_f), how do we find that one that extremizes J?

To proceed, we imagine that there exists some $x_e(t)$ which gives us this extremum. We do not know $x_e(t)$, so our current guess deviates from this correct choice. We parameterize how our current guess deviates from the correct trajectory as follows:

$$x(t) \equiv x(t;\alpha) = x_e(t) + \alpha A(t)\ . \tag{16.2}$$

The function $A(t)$ is totally arbitrary, except that we require it to vanish at the endpoints: $A(t_i) = A(t_f) = 0$; otherwise, our trajectory would not meet the boundary condition. The parameter α allows us to control how the variation $A(t)$ enters into our path $x(t;\alpha)$.

Our basic idea is to ask how the integral J behaves when we are in the vicinity of the extremum. We know that ordinary functions are flat — they have zero first derivative — when we are at an extremum. Let us put

$$J(\alpha) = \int_{t_i}^{t_f} f\left[x(t;\alpha), \dot{x}(t;\alpha); t\right]\, dt\ . \tag{16.3}$$

We've now made the integral J a function of the parameter α. We know that $\alpha = 0$ corresponds to the extremum of J by its definition. However, this isn't useful for us, since we don't know what $x(t)$ this corresponds to. However, because $\alpha = 0$ corresponds to an extremum, we also know that $(\partial J/\partial\alpha)_{\alpha=0} = 0$; in essence, we're taking advantage of the fact that the *shape* of $J(\alpha)$ has a particular form as we approach this extremum.

Let's take a look at the derivative of J with respect to α:

$$\frac{\partial J}{\partial \alpha} = \int_{t_i}^{t_f} \left[\frac{\partial f}{\partial x}\frac{\partial x}{\partial \alpha} + \frac{\partial f}{\partial \dot{x}}\frac{\partial \dot{x}}{\partial \alpha}\right] dt \; . \tag{16.4}$$

Using Eq. (16.2), we have

$$\frac{\partial x}{\partial \alpha} = A(t) \; , \qquad \frac{\partial \dot{x}}{\partial \alpha} = \frac{dA}{dt} \; . \tag{16.5}$$

Plugging this in, we have

$$\frac{\partial J}{\partial \alpha} = \int_{t_i}^{t_f} \left[\frac{\partial f}{\partial x}A(t) + \frac{\partial f}{\partial \dot{x}}\frac{dA}{dt}\right] dt \; . \tag{16.6}$$

The last term on the right-hand side of (16.6) can be rearranged in a really useful way using integration by parts:

$$\begin{aligned}\int_{t_i}^{t_f} \frac{\partial f}{\partial \dot{x}}\frac{dA}{dt}\,dt &= A(t)\frac{\partial f}{\partial \dot{x}}\bigg|_{t_i}^{t_f} - \int_{t_i}^{t_f} A(t)\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right) dt \\ &= -\int_{t_i}^{t_f} A(t)\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right) dt \; . \end{aligned} \tag{16.7}$$

To get the final expression, we used the fact that $A(t_i) = A(t_f) = 0$. Using this we have

$$\frac{\partial J}{\partial \alpha} = \int_{t_i}^{t_f} A(t)\left[\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right)\right] dt = 0 \quad \text{at an extremum of } J. \tag{16.8}$$

The function $A(t)$ is totally arbitrary, aside from the boundary condition that it vanish at t_i and t_f. We require $\partial J/\partial\alpha = 0$ for all $A(t)$; for this to occur, the quantity inside square brackets must vanish:

$$\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right) = 0 \; . \tag{16.9}$$

This is known as *Euler's equation*, and was first derived by the Swiss polymath Leonhard Euler. Properly applied, it yields a differential equation which allows us to find the trajectory $x(t)$ which extremizes the integral J.

For simplicity, we did this for a function of just one variable. However, we could have imagined a trajectory in all three spatial directions. With a little more effort, it's not too hard to show that more general version of Eq. (16.9) is just the trio of equations

$$\begin{aligned}\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right) &= 0 \; , \\ \frac{\partial f}{\partial y} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{y}}\right) &= 0 \; , \\ \frac{\partial f}{\partial z} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{z}}\right) &= 0 \; . \end{aligned} \tag{16.10}$$

Those of you who have studied Lagrangian mechanics (which we discuss briefly at the end of these notes) presumably have already encountered equations of this form.

16.3 An example: The brachistochrone ("shortest time")

A bead starts from rest at $(x_i, y_i) = (0, 0)$ and slides without friction down a wire, reaching (x_f, y_f). What shape should the wire have in order for the bead to reach (x_f, y_f) in as little time as possible?

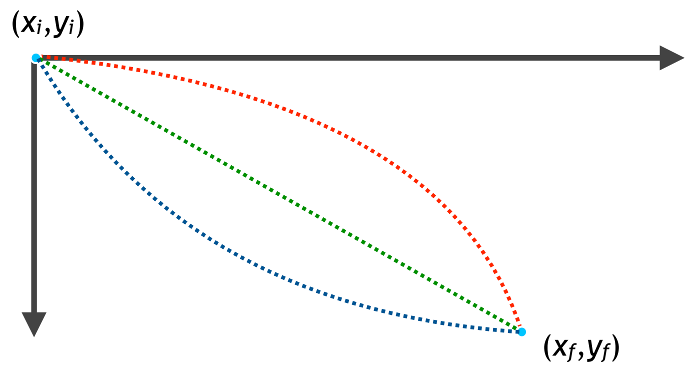


Figure 1: Three plausible paths for the brachistochrone connecting $(0, 0)$ and (x_f, y_f).

To figure this out, apply the Euler equation to minimize the total travel time of the bead as it slides along the wire. Think of the integral we are minimizing as T, defined by

$$T = \int_{\text{initial position}}^{\text{final position}} \frac{ds}{v} , \tag{16.11}$$

where ds is the differential of path length along the wire, and v is its speed. For the path length, note that the bead moves in both x and y, so

$$ds = \sqrt{dx^2 + dy^2} = dy\sqrt{1 + (x')^2} , \qquad \text{where} \quad x' \equiv \frac{dx}{dy} . \tag{16.12}$$

For the speed v, since the bead starts from rest, it only gets speed from falling a distance y:

$$\frac{1}{2}mv^2 = mgy \qquad \longrightarrow \qquad v = \sqrt{2gy} . \tag{16.13}$$

The equation we wish to minimize is thus

$$T = \int_0^{y_f} \sqrt{\frac{1 + (x')^2}{2gy}} dy . \tag{16.14}$$

This is perfectly set up for us to apply the Euler equation provided we make a few adjustments: we put

$$f = \sqrt{\frac{1 + (x')^2}{2gy}} ; \tag{16.15}$$

we change the integration variable from t to y, and replace $\dot{x}$ with x'. Our slightly tweaked Euler equation is thus

$$\frac{\partial f}{\partial x} - \frac{d}{dy}\left(\frac{\partial f}{\partial x'}\right) = 0 . \tag{16.16}$$

Let's evaluate these terms:

$$\frac{\partial f}{\partial x} = 0 \ , \qquad \frac{\partial f}{\partial x'} = \left(\frac{1}{\sqrt{2gy}}\right)\left(\frac{x'}{\sqrt{1+(x')^2}}\right) \ . \tag{16.17}$$

Plugging these into the Euler equation yields

$$\frac{d}{dy}\left(\frac{1}{\sqrt{2gy}}\right)\left(\frac{x'}{\sqrt{1+(x')^2}}\right) = 0 \ . \tag{16.18}$$

We can immediately integrate this up to find

$$\left(\frac{1}{\sqrt{2gy}}\right)\left(\frac{x'}{\sqrt{1+(x')^2}}\right) = \text{constant} \ . \tag{16.19}$$

Let's set this constant to $1/\sqrt{4gA}$, where A is another constant[1] with the dimensions of length. Squaring both sides of Eq. (16.19), we find

$$\frac{(x')^2}{2gy\,(1+(x')^2)} = \frac{1}{4gA} \ , \tag{16.20}$$

which we can manipulate into

$$\begin{aligned}\left(\frac{dx}{dy}\right)^2 &= \frac{y/(2A)}{1-y/(2A)} \\ &= \frac{y^2}{2Ay-y^2} \ .\end{aligned} \tag{16.21}$$

We thus at last have our equation governing x as a function of y:

$$x(y) = \int_0^y \frac{y\,dy}{\sqrt{2Ay-y^2}} \ . \tag{16.22}$$

To wrap this up, we change variables: define $y = A(1-\cos\theta)$, $dy = A\sin\theta\,d\theta$. It's not too hard to show that $2Ay - y^2 = A^2\sin^2\theta$; our equation for x becomes

$$x = \int_0^\theta A(1-\cos\theta)\,d\theta = A(\theta - \sin\theta) \ . \tag{16.23}$$

The full solution for the brachistotrone is then given by

$$\begin{aligned} x &= A(\theta-\sin\theta) \\ y &= A(1-\cos\theta) \ . \end{aligned} \tag{16.24}$$

The bead's motion goes over the range $0 \le \theta \le \theta_{\max}$; both the constant A and the angle $\theta_{\max}$ can be found by solving $x(\theta_{\max}) = x_f$, $y(\theta_{\max}) = y_f$.

[1] This is one of those places where I get to invoke instructors' privilege and cheat a little bit. If you were doing this problem by yourself, you'd probably just set the right hand side to something like C, and hope that C's role is explained later in the calculation. Doing so, you would eventually find that C shows up as $1/4gC^2$ in the analysis. Since I've already done the calculation, I'm using the fact that I know this in advance to streamline things here.

16.4 Maximal aging in special relativity

Let's use the calculus of variations to see what kind of motion results in "maximal aging" on an observer's trajectory in special relativity. We start with the fact that, for an observer moving on a timelike trajectory,

$$d\tau^2 = dt^2 - \left(dx^2 + dy^2 + dz^2\right)/c^2 \ . \tag{16.25}$$

For simplicity, let's restrict ourselves to one spatial dimension for now, setting $dy = dz = 0$.

Now think about the many paths which can connect event A to event B. Our goal is to compute the accumulated τ along those paths:

$$\tau_{A\to B} = \int_A^B dt\sqrt{1 - \frac{1}{c^2}\left(\frac{dx}{dt}\right)^2} \equiv \int_A^B dt\sqrt{1 - \frac{\dot{x}^2}{c^2}} \ . \tag{16.26}$$

We've introduced $\dot{x} = dx/dt$ for notational convenience. Notice that the integrand looks like $dt/\gamma(\dot{x})$ — a form that hopefully makes a lot of sense.

Let's now think about how to extremize $\tau_{A\to B}$ by putting $J \to \tau$, and setting $f(x, \dot{x}) = \sqrt{1 - \dot{x}^2/c^2}$. Re-stating Euler's equation,

$$\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right) = 0 \ , \tag{16.27}$$

with

$$\frac{\partial f}{\partial x} = 0 \ , \qquad \frac{\partial f}{\partial \dot{x}} = -\frac{\dot{x}/c^2}{\sqrt{1 - \dot{x}^2/c^2}} \ . \tag{16.28}$$

Plugging these into (16.27), we find

$$\frac{d}{dt}\left(\frac{\dot{x}}{\sqrt{1 - \dot{x}^2/c^2}}\right) = 0 \tag{16.29}$$

whose solution is

$$\dot{x} = \text{constant} \ . \tag{16.30}$$

If you include y and z in your analysis, you'll likewise conclude that $\dot{y}$ and $\dot{z}$ must be constants in order to follow the trajectory that maximizes the accumulated proper time from A to B. A trajectory with $\dot{x}$, $\dot{y}$, and $\dot{z}$ all constants is nothing more than an inertial trajectory in spacetime. **The unaccelerated trajectory is the one which maximizes an observer's accumulated proper time as they move through spacetime: It is the trajectory of maximal aging.**

Important side issue: strictly speaking, the calculation we just did tells us that the unaccelerated trajectory represents an *extremum* of accumulated proper time, which can be either a maximum or a minimum. How do we know this extremum is a maximum and not a minimum? In this particular case, it is because we know that the minimum aging trajectory is the trajectory along which $\tau_{A\to B} = 0$, and in fact that there are an infinite number of such trajectories, all with crazy — essentially unphysical — accelerations. In general, knowing whether the extremum is a minimum or a maximum requires you to think a little bit about the physics of your situation. You will find that the outcome of the Euler equation calculation picks a unique extremum; the opposite extremum tends to not be uniquely specified.

16.5 Lagrangian mechanics and relativity

Independent of whether you continue to study relativity into the future or not, the calculus of variations and the Euler equations are likely to be important for you as long as you remain a physics student. The reason is that ordinary mechanics can be formulated in a way that uses these tools. The basic idea works as follows:

- Suppose a body moves from event (t_i, x_i, y_i, z_i) to event (t_f, x_f, y_f, z_f).
- Consider *every possible trajectory* that connects these events. For every point along those trajectories, compute the body's kinetic energy K and its potential energy U.
- Define the *Lagrangian* L as the difference in these quantities: $L \equiv K - U$.
- Define the *action* S as the time integral of L: $S \equiv \int_{t_i}^{t_f} L\,dt$.

A remarkable result, which is discussed in great detail in the IAP course 8.223 and the advanced mechanics course 8.09, is that Newtonian mechanics is equivalent to the path of *least action*, and can be found by applying the Euler equations (often called the Euler-Lagrange equations in this context) to L:

$$\frac{\partial L}{\partial x} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = 0\ , \quad \frac{\partial L}{\partial y} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}}\right) = 0\ , \quad \frac{\partial L}{\partial z} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{z}}\right) = 0\ . \tag{16.31}$$

These equations work in other coordinate systems too — you can replace the Cartesian set (x, y, z) with cylindrical coordinates (r, ϕ, z) or spherical ones (r, θ, ϕ) or really bizarre ones that just happen to be adapted to the geometry of your problem.

A Lagrangian formulation of mechanics is often much easier to work with than the $\mathbf{F} = m\mathbf{a}$ based techniques you learned in 8.01/8.012, particularly if the problem is subject to constraints. What makes them particularly nice to work with is that ultimately one need only compute a single scalar quantity, L, rather than work with vector-valued forces or torques. It is also worth noting that the classical action is intimately connected to the phase of a quantum wavefunction. The trajectory of "least action" is closely related to the phase corresponding to the most likely outcome of a quantum process in the classical limit.

In another few lectures, we will start working with general spacetime metrics, for which $ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta$, and for which the metric $g_{\alpha\beta}$ will be a function of the different coordinates. However, it will remain the case that for a timelike observer, $c^2 d\tau^2 = -g_{\alpha\beta}dx^\alpha dx^\beta$. Let's use this to define the Lagrangian-like quantity that we will want to use to describe motion in general spacetimes. Consider motion that begins at event A and ends at event B. The proper time accumulated along a trajectory between these events is

$$c\Delta\tau = \int_A^B (f)^{1/2} d\tau\ , \tag{16.32}$$

where

$$f = -g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} \equiv -g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu\ . \tag{16.33}$$

Requiring that the trajectory through spacetime between events A and B be an extremum leads us to the following 4 Euler equations (one for each value of the index α):

$$\frac{\partial (f)^{1/2}}{\partial x^\alpha} - \frac{d}{d\tau}\left(\frac{\partial (f)^{1/2}}{\partial \dot{x}^\alpha}\right) = 0\ . \tag{16.34}$$

This can be simplified a bit more. First, note that

$$\frac{\partial (f)^{1/2}}{\partial x^\alpha} = \frac{1}{2(f)^{1/2}}\frac{\partial f}{\partial x^\alpha}\;, \qquad \frac{\partial (f)^{1/2}}{\partial \dot{x}^\alpha} = \frac{1}{2(f)^{1/2}}\frac{\partial f}{\partial \dot{x}^\alpha}\;. \tag{16.35}$$

Second, note that $df/d\tau = 0$: f is nothing more than $g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = \vec{u}\cdot\vec{u} = -c^2$, and the *total* derivative of this quantity with respect to proper time is zero. (Its *partial* derivatives are not zero: if we vary a particular value of x^α or a particular value of $\dot{x}^\alpha$ while holding all other quantities constant, we push f away from the value it "should" have.) This allows us out to clear out an overall factor of $2(f)^{1/2}$, and the Euler equations (16.34) then become

$$\frac{\partial f}{\partial x^\alpha} - \frac{d}{d\tau}\left(\frac{\partial f}{\partial \dot{x}^\alpha}\right) = 0\;. \tag{16.36}$$

This tells us that $f = -g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu$ plays a role in relativistic mechanics exactly like the Lagrangian of ordinary classical mechanics. It is traditional to multiply this by a factor of $-1/2$ — after all, the extremum of $-1/2$ times a function occurs at the same place as the extremum of that function. We then define the relativistic Lagrangian as

$$L = \frac{1}{2}g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\;. \tag{16.37}$$

(Strictly speaking, this is a "Lagrangian per unit rest mass" for a body moving through spacetime.) The motion of a body in the spacetime $g_{\mu\nu}$ can then be found by applying the Euler-Lagrange equations to this L.

After a bit of discussion about how to get the spacetime metric $g_{\mu\nu}$, we will use this "relativistic Lagrangian" quite a bit in the last few weeks of this course.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 17
Goodbye global Lorentz frames, hello principle of equivalence
Initial considerations on relativistic gravity

17.1 Farewell to global Lorentz frames

What is it that puts the "special" in special relativity? The key concept that we come back to again and again is the notion of a *Lorentz frame*: A frame of reference in which things move at constant velocity if no forces act on them. Such a frame is an inertial frame; we move between different Lorentz frames using Lorentz transformations.

What is particularly special about special relativity is that it assumes that we can "cover" all of spacetime — all events, all time and all space — using a single Lorentz frame. In other words, special relativity tells us that it makes sense for there to be *global* Lorentz frames.

Gravity breaks this. Once we begin including gravity in our model of physics, we cannot have a global Lorentz frame that covers all events. This is actually fairly easy for us to see based on things that we have already learned about the nature of Lorentz frames, and the influence of gravity on light.

Imagine a pulse of light that propagates from the surface of the Earth to a height H. Let us imagine the trajectory that one crest of a light wave in this pulse follows through spacetime. We do not yet know exactly how gravity will affect the pulse's path through spacetime, but we can imagine that the trajectory is "bent" essentially, perhaps moving a little slower near the surface than it moves after propagating to greater heights:

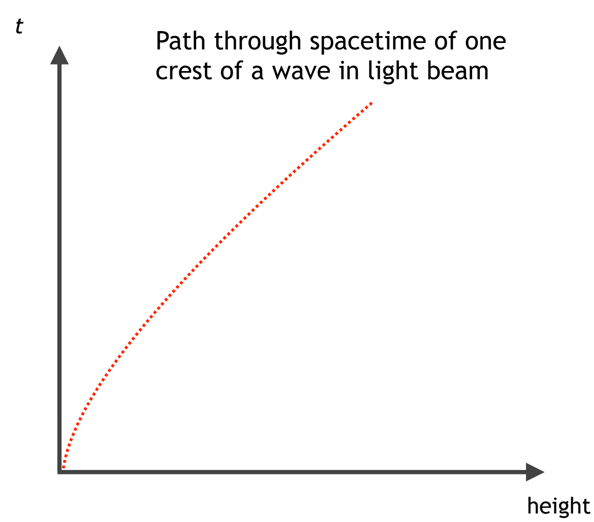


Figure 1: A plausible path for the crest of a light wave in a pulse propagating vertically from the Earth's surface.

Given this behavior of the first crest, what is the behavior of the second crest? Well, *if* we require spacetime to be Lorentz everywhere, then there is nothing special about any particular time or place. The path of the second crest must be identical to the path of the first one, simply shifted later in time.

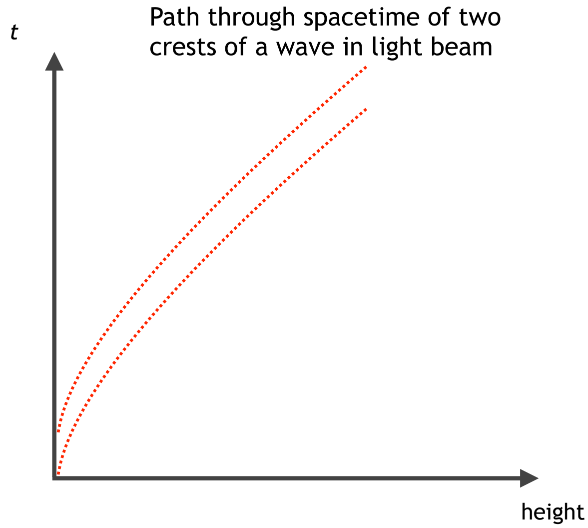


Figure 2: If the trajectory through spacetime of crest 1 looks like Figure 1 *and* we assume spacetime has Lorentzian behavior everywhere, then the trajectories of crests 1 and 2 together will look this. Figure made by duplicating the first crest and sliding it slightly in the t direction.

If this is true, then it must be the case that the wave period at the bottom (height 0) must be identical to the wave period at the top (height H). The two trajectories are *congruent* with each other, simply shifted a bit along the time direction. But if the periods T_H and T_0 are identical, then the frequencies at the top and the bottom are identical: $\nu_H = \nu_0$. **This contradicts the gravitational redshift that we argued must exist** (and that, indeed, experiments have demonstrated does in fact exist), which tells us that $\nu_H = \nu_0(1 - gH/c^2)$. Our starting assumption must be incorrect: *In the presence of gravity, we cannot have global Lorentz reference frames.*

Perhaps we could "rescue" special relativity with the Rindler coordinate system. Rindler coordinates express how things look in special relativity according to a uniformly accelerated observer; we saw that an analysis of light measured by such an observer looks very similar to the expressions we derived for the impact of gravity on light. However, the Rindler coordinate system describes uniform acceleration along a particular direction in space. With a little thought, we can convince ourselves that a Rindler coordinate system cannot describe *all* the measurements that we can make on the Earth's surface.

Consider an observer on the equator who measures the gravitational redshift. They can interpret their measurements as consistent with a Rindler coordinate system that is accelerating "up," i.e., outwards from the equator. Consider a second observer at the North Pole who measures exactly the same gravitational redshift. They likewise may want to interpret the redshift as due a Rindler coordinate system that is accelerating "up." However, their "up" is 90° different from the "up" of the equatorial observer! Consider a third observer at the South Pole. They also want a Rindler observer accelerating "up," but their "up" is 180° different from the North Pole's "up." None of these observers are in fact moving with respect to one another: they are widely separated, but their separations are not changing. This is starkly different from accelerations in three different directions which the Rindler hypothesis requires.

We need a new idea in order to incorporate gravity in the framework of relativity.

17.2 The principle of equivalence

Let's go back to the foundation of what an inertial frame has meant: *In the absence of external forces, all objects maintain their relative velocities.* Is there any way in which the essence of this idea can be captured when we include the action of gravity?

One of Einstein's core insights was that we observe exactly the same thing when we do our analysis in a *Freely Falling Frame*, or FFF. All objects feel the same acceleration due to gravity, thanks to the fact that $\mathbf{F} = m\mathbf{a} = m\mathbf{g}$. The equivalence of "gravitational mass" and "inertial mass" means that gravity effectively cancels out as long as we can work entirely in the FFF. The notion of a Lorentz frame is now upgraded to a Freely Falling Frame, and the rule that we will use is: *In the absence of non-gravitational forces, objects maintain their relative velocities in a Freely Falling Frame.*

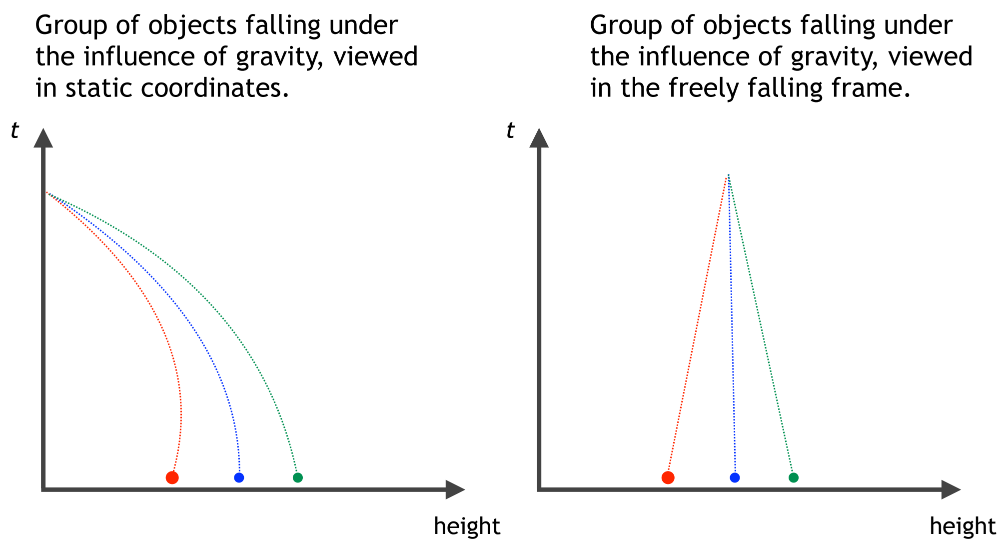


Figure 3: Three objects falling under the influence of gravity. In "static" coordinates (e.g., coordinates at rest with respect to the Earth's surface, shown on the left), the three bodies follow parabolic trajectories before meeting later at height 0. In the freely falling frame, the observer follows the same trajectory as the blue object. All three objects move along straight lines in this frame. Motion in the freely falling frame duplicates the essential features of unaccelerated motion in an inertial frame in the absence of gravity.

The key intuition for this is that, as summarized by Einstein, we cannot distinguish between gravity and uniform acceleration. It is important to bear in mind, however, that in realistic situations gravity is never perfectly uniform. As we move away from the Earth's center, the gravitational force gets weaker. This variation in gravity from the Earth, or from any realistic finite-sized source, is responsible for *tides*.

Tides are responsible for a key aspect of how we describe gravity in relativistic language. In special relativity, if two objects started out moving parallel to one another and no force acted on them, their trajectories would *always* remain parallel. This is a statement that trajectories in spacetime obey what is known as "Euclid's parallelism postulate," an aspect of Euclidean geometry which confused mathematicians and geometers for centuries. Unlike

Euclid's other postulates, the parallelism postulate was not considered to be self evident, and could not be proved under the assumption of Euclid's other postulates. Work by the Russian mathematician Lobachevsky first showed that one could set up a logically consistent framework for geometry without assuming this postulate; in such a geometry, lines which start out parallel later cross or diverge from one another. The German mathematician Riemann later worked out rules describing such geometries.

In modern language, we now say that if Euclid's parallelism postulate holds then it means that one is working in a geometry that is *flat*. In two and three spatial dimensions, a flat geometry in which the Pythagorean theorem holds; in space and time, it is a geometry with the metric $\eta_{\alpha\beta}$ that we have been working with for most of this semester.

On the other hand, if Euclid's parallelism postulate does not hold, then one is working in a geometry that is *curved*. An example is the surface of a sphere. Consider two observers standing on the Earth's equator. Both begin walking due north — perfectly parallel to one another. They walk in a perfectly straight line on the surface, never bending their path from one moment to the next. Despite beginning on parallel trajectories, and despite moving along perfectly straight lines, their trajectories cross when they reach the North Pole.

Tides cause trajectories which are initially parallel in spacetime to either focus or diverge from one another. This tells us that *when we have gravity with tides, spacetime must be curved*. We cannot use the metric $\eta_{\alpha\beta}$ anymore; we need something new.

17.3 How to describe relativistic gravity I: Initial considerations

Let's think about Newtonian gravity for a moment. Begin by considering the potential outside of a spherical mass M,

$$\Phi = -\frac{GM}{r} . \tag{17.1}$$

This gravitational potential has the same mathematical form as the electrostatic potential that arises from a spherical charge Q:

$$\Phi^E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r} ; \tag{17.2}$$

we just need to replace $Q \to M$ and $1/(4\pi\epsilon_0)$ with $-G$.

In an in-depth study of electrostatics, we learn that for a *general* distribution of charge, the electrostatic potential Φ^E is the function that solves *Poisson's equation*:

$$\nabla^2\Phi^E = -\frac{\rho_Q}{\epsilon_0} , \tag{17.3}$$

where ρ_Q is charge density. This can be proven by combining $\mathbf{E} = -\nabla\Phi^E$ with $\nabla\cdot\mathbf{E} = \rho_Q/\epsilon_0$. In the same way, one can show that in Newtonian gravity, the gravitational potential Φ^G is the function that solves a slightly different version of Poisson's equation:

$$\nabla^2\Phi^G = 4\pi G\rho_M , \tag{17.4}$$

where ρ_M is mass density.

Let's begin here as we start thinking about how to bring gravity into a relativistic framework. We start by cataloguing the ways in which Eq. (17.4) falls short as a relativistic equation, and imagine ways in which we could perhaps "upgrade" it to something better for our purposes.

- The left-hand side of Eq. (17.4) involves spatial derivatives in a particular reference frame. This is not a Lorentz-covariant derivative operator. One idea for upgrading this: replace ∇^2 with the relativistic wave operator $\Box$. Past studies of gravity were dominated by sources that were static or very slowly varying; perhaps the most important aspects of gravity have been determined from sources for which $\partial(\text{gravity})/\partial t \approx 0$ in the frames in which we did these studies.

- The right-hand side of Eq. (17.4) involves the mass density ρ_M. We argued (and much later, experiments verified) that gravity must also act upon massless energy. However, past studies of gravity were dominated by sources for which the rest energy was the largest part of the source's energy budget. Perhaps we can replace ρ_M with ρ/c^2, where ρ is the source's energy density.

This suggests that perhaps our relativistic gravity equation should look something like

$$\Box\Phi^G \stackrel{?}{=} \frac{4\pi G\rho}{c^2} . \tag{17.5}$$

This perhaps looks plausible, but on reflection hopefully you'll notice that it has some issues. Chief among them is that, as we discussed several lectures ago, the energy density ρ is one component in a specified reference frame of the stress-energy tensor. Any theory of physics that picks out a particular component of a tensor as playing a special role is, for lack of a better term, a "sick" theory. If we want gravity to be describable from the viewpoint of different reference frames, then the right-hand side of Eq. (17.5) won't do it.

The left-hand side of (17.5) has problems as well. The derivative operator is a scalar, but what is Φ^G? Is it a scalar (as it appeared to be in Newtonian physics)? Is it one component of a tensor, as the right-hand side seems to suggest? If so, what is the rest of the tensor?

This is roughly where Einstein was in the early 20th century, trying to imagine how to fold gravity into the framework of relativity that so successfully merged Maxwell's electrodynamics with mechanics. Getting from there to the general theory of relativity took Einstein about 10 years, much of which was spent learning what was for him an entirely new field of mathematics (Riemannian geometry), and figuring out how to connect this to the core physical concepts that describe gravity. There were multiple wrong turns along the way; in the meantime, others proposed different ways of making relativistic gravity which in the end did not agree with experimental tests.

In 8.033, we don't have the time to explore all of the wrong turns and hypotheses that were proposed but fell short (although we briefly discuss some highlights of interesting "wrong turns" in a short section of supplementary material). Instead, we will elide many details and compress all of the history and thought processes into a few bullet points:

- In special relativity, an unaccelerated trajectory is one that moves on a straight line. If a pair of unaccelerated trajectories start out parallel, then they will remain forever parallel. This is consistent with the idea that the metric $\eta_{\alpha\beta}$ describes a "flat" spacetime geometry.

- When gravity is included, we can introduce principles that allow to recover much of that core idea. We *define* an unaccelerated trajectory in the freely falling frame as the one that feels no non-gravitational forces.

- Because gravity is never perfectly uniform — it exhibits *tidal* variations — we expect a pair of unaccelerated trajectories that start out parallel to not remain parallel; in almost all cases, they will eventually diverge from one another, or perhaps cross. This suggests that gravity can be modeled by thinking about spacetimes that are not flat, but that have *curvature*.

17.4 How to describe relativistic gravity II: Putting the pieces together

Now let's synthesize these ingredients and bullet points to see how, after 10 years of effort, Einstein managed to develop the relativistic theory of gravity that (so far, at least) has passed all experimental tests. Begin by going back to the Newtonian field equation:

$$\nabla^2 \Phi^G = 4\pi G \rho_M \; . \tag{17.6}$$

We've already argued that the right-hand side should be something that involves ρ/c^2 rather than ρ_M, where ρ is energy density. But that ρ is itself one component of the stress-energy tensor $T^{\mu\nu}$. A covariant relativistic formulation cannot pick out one component of a tensor as "the" quantity of interest. Whatever goes on the left-hand side of the relativistic "gravity equation," the right-hand side should be something that is proportional to $T^{\mu\nu}$.

To get some idea of how to handle the left-hand side, note that $\nabla^2\Phi^G$ can be regarded as $-\nabla \cdot \mathbf{g}$, where $\mathbf{g} = -\nabla\Phi^G$ is the gravitational field that arises from the potential Φ^G. The left-hand side is thus something like the divergence of the gravitational field. Derivatives of the gravitational field tell us about how this field varies in space — which tells us about the behavior of gravitational tides. So the physical content of Eq. (17.6) can be regarded, schematically, as

$$(\text{“Quantity related to gravitational tides”}) = (\text{“numerical factor times } G\text{”})(\text{“source”}) \; . \tag{17.7}$$

For the source on the right-hand side of our equation, we've already decided to use the stress-energy tensor $T^{\mu\nu}$. Figuring out how to do the left-hand side is a little more complicated. We begin with the idea that a body which moves under the influence of no forces but gravity follows a trajectory of maximal aging through spacetime. Such a trajectory is called a *geodesic*. We will examine geodesics for specific spacetimes soon enough; in the general case (which we will *not* consider in detail in 8.033), a body's geodesic motion in some coordinate system turns out to be governed by a three-index tensor-like object. The differential equation governing the body's 4-velocity takes the form

$$\frac{du^\alpha}{d\tau} + \Gamma^\alpha{}_{\mu\nu} u^\mu u^\nu = 0 \; . \tag{17.8}$$

The quantity $\Gamma^\alpha{}_{\mu\nu}$ (called a connection coefficient or Christoffel symbol) is built from derivatives of the spacetime metric $g_{\mu\nu}$.

If the spacetime describes gravity with tides, then two nearby geodesics that start parallel to one another will eventually become non-parallel. Suppose that two geodesics each have 4-velocity u^μ, and are separated in our coordinates by δx^α. Then, the action of tides will cause their separation to evolve. This evolution is governed by an equation that takes the form

$$\frac{D^2(\delta x^\alpha)}{d\tau^2} = R^\alpha{}_{\mu\nu\beta} u^\mu u^\nu (\delta x^\beta) \; . \tag{17.9}$$

The operator $D/d\tau$ is a special kind of derivative that takes into account the fact that, in a spacetime with curvature, the unit vectors themselves vary from position to position. The 4-index tensor $R^\alpha{}_{\mu\nu\beta}$ (called the Riemann curvature tensor) describes how nearby geodesics deviate from one another due to variations in spacetime — i.e., how tidal variations in gravity make initially parallel trajectories become non parallel. This curvature tensor is built from derivatives of the Christoffel symbol; we can think of it as expressing (in a rather complicated way) two derivatives of the spacetime metric $g_{\mu\nu}$.

Einstein's hypothesis was that the "right" way to upgrade Eq. (17.6) into a relativistic form was to replace the left-hand side with a curvature tensor which is closely related to $R^\alpha{}_{\mu\nu\beta}$, and to replace the right-hand side with the stress-energy tensor:

$$G^{\mu\nu} = \kappa T^{\alpha\beta} \ . \tag{17.10}$$

The tensor $G^{\mu\nu}$ is known as the Einstein curvature tensor[1]. It is found by combining the Riemann tensor with the metric in a such a way that the result is a 2nd-rank tensor with zero divergence (the stress-energy tensor on the right-hand side has zero divergence, so whatever we put on the left-hand side must have zero divergence as well). You can think of it as a very complicated set of second derivatives which act on the metric.

The constant κ is determined by demanding that, in the correct limit, this equation's predictions agree with the predictions of Newtonian gravity. Doing so, we at last obtain the *Einstein field equation*:

$$G^{\mu\nu} = \frac{8\pi G}{c^4} T^{\mu\nu} \ . \tag{17.11}$$

This equation can be regarded as a set of partial differential equations for the spacetime metric, given a stress-energy tensor which describes the flow of energy and momentum in that spacetime. Notice that there is a sense in which (17.11) is similar in physical structure to Eq. (17.6): both have "two derivatives of potential" on the left-hand side (provided we now think about the metric of spacetime as playing the role of the potential), and a source that describes energy density on the right-hand side (picking out the dominant component in the Newtonian version, but using the full tensor-valued mathematical object in Einstein's).

Developing the Einstein field equation takes roughly half of the term in 8.962. The other half is spent figuring out how to solve it, and to examining the nature of its solutions. In 8.033, we will jump straight to looking at some of the solutions (though the story behind how some of those solutions were found is pretty interesting, and we'll at least discuss some anecdotes around them). We will then study these solutions in order to tell us about the nature of gravity with relativity included. A very nice feature of what we have done so far is that, with the principle of equivalence and the calculus of variations in our toolkit, it's a relatively simple step for us to examine motion in some spacetime that is provided to us.

[1]Sadly, the notation overlaps with the dual Faraday tensor we discussed in the E&M section of this course. Context generally makes it clear which is which.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 18
Some important solutions of the Einstein field equation; using those solutions

18.1 Final thoughts on the Einstein field equation

In the previous lecture, we discussed the generic framework and logic that led Einstein, after roughly a decade of learning the relevant mathematics and considering how to connect the pieces together, to the field equation of general relativity:

$$G^{\mu\nu} = \frac{8\pi G}{c^4} T^{\mu\nu} \;. \tag{18.1}$$

The left-hand side of this equation (the "Einstein curvature tensor") can be regarded as a very complicated second-order differential operator acting on the metric of spacetime. It describes, after a little bit of massaging, the spacetime's curvature — that is, the tendency of the trajectories in spacetime of freely falling bodies which start parallel to become non-parallel as the bodies move. The right-hand side expresses, in a covariant form, the distribution of energy density, momentum density, and their flow in spacetime.

We are not going to do a lot with this equation other than to examine several of its solutions. However, before getting into this, it is worth remarking on a couple of points.

- First, note that when working in Cartesian coordinates, the curvature tensor on the left-hand side has dimension $1/(\text{length})^2$. With that in mind, it is interesting to look at the numerical value of the constant which connects the curvature tensor to the stress-energy tensor:

  $$\frac{8\pi G}{c^4} = 2.08 \times 10^{-43} \frac{\text{meter}^{-2}}{\text{J/meter}^3} \;. \tag{18.2}$$

  I've written the units to emphasize that this constant converts energy density (Joules per meter cubed) into curvature (inverse meters squared). Notice it takes a *lot* of energy density to produce a tiny amount of curvature. Osmium is the densest metal naturally found on Earth, at 22.6×10^3 kilograms per meter cubed (roughly three times the density of iron, and twice that of lead). Multiplying by c^2, we see that osmium has a rest energy density of 2.03×10^{21} Joules per meter cubed. But this density only produces 4.22×10^{-22} inverse meters squared of curvature. When you hear someone describe gravity as the weakest of the fundamental interactions, this is the essence of what they mean — we need a *lot* of energy density to curve spacetime. To get strong curvature, we need to go to regimes far beyond what we encounter on Earth.

  This doesn't mean that gravity is negligible though. Because "gravitational charge" — i.e., mass — only comes with one sign (there is no negative mass), its effects add up. Still, it's worth noting that every time you lift any object, electrochemical reactions in a couple hundred grams of muscle tissue overcome the accumulated gravitational effects of 6×10^{24} kilograms of our planet.

- People often seemed a little surprised by how *ad hoc* the derivation of the Einstein field equation seems to be. In essence, Einstein seems to have decided that the source should be $T^{\mu\nu}$, decided that the left-hand side should be a curvature tensor, then just matched $T^{\mu\nu}$ to a curvature tensor that is divergence free and has the right number of indices.

  This is not wrong! Einstein's original derivation of the Einstein field equation is indeed just as *ad hoc* as this makes it seem. Two remarks on this are in order:

  - First, it's worth bearing in mind that the ultimate arbiter of what description we should use for any physical interaction is *measurement*. You should therefore regard the Einstein field equation and its predictions as hypotheses to be tested. **Testing this hypothesis is still something being done today**, and in fact motivates quite a lot of modern research (including a bit of my own).
  - Around the time that Einstein formulated these field equations, other plausible formulations of relativistic gravity were also proposed. Those all were found to be flawed in important ways, failing experimental tests or turning out to have internal contradictions. General relativity can be regarded as the relativistic gravity theory that (so far, at least) best fits the data.
  - There's another way of deriving the field equation which is based on a variational principle, similar to the way that we apply variational principles to a Lagrangian in order to describe a body's motion. Though quite a bit beyond the scope of 8.033, it is worth remarking that this approach makes it clear that the Einstein field equation is, in a way that can be made precise, the *simplest* theory of relativistic gravity. A lot of research these days explores how general relativity may be, in a meaningful sense, itself an approximation to something deeper. This variational principle provides a foundation for exploring the nature of gravity.

There's a lot more we could say, but this will suffice for 8.033. The tack we are going to take from now on is to look at solutions of this equation and examine their consequences. I want to emphasize that *so far* we have not found any compelling evidence of shortcomings in general relativity's description of gravity, which is why this is often taught as "the" theory of relativistic gravity. But we keep looking.

18.2 Some example solutions and their significance

18.2.1 The "weak gravity" metric

Upon figuring out the field equation, Einstein developed its first solution. This is done by considering "weak" gravity — spacetime that is not *too* different from the metric of special relativity. This simplifies the curvature tensor, essentially by allowing us to approximate terms that are nonlinear in the spacetime metric as small enough that their influence can be neglected. The solution which emerges in this limit has 4 non-zero metric components:

$$g_{00} = -(1 + 2\Phi/c^2) \;, \quad g_{11} = g_{22} = g_{33} = (1 - 2\Phi/c^2) \;. \tag{18.3}$$

All other components of the spacetime metric are zero. The coordinates used here are

$$x^0 = ct \;, \quad x^1 = x \quad x^2 = y \quad x^3 = z \;. \tag{18.4}$$

The function Φ which appears in (18.3) is just the Newtonian gravitational potential. Outside a spherical body of mass M centered on the origin, it takes the form

$$\Phi = -\frac{GM}{r} , \qquad r = \sqrt{x^2 + y^2 + z^2} . \tag{18.5}$$

This metric works well when $\Phi \ll c^2$, which is a good description of spacetime almost everywhere in our solar system, for example.

18.2.2 The Schwarzschild metric

As mentioned at the end of the November 17 lecture, the first *exact* solution to the Einstein field equations was found by Karl Schwarzschild in 1916. It also has 4 non-zero metric components:

$$g_{00} = -\left(1 - \frac{2GM}{rc^2}\right) , \quad g_{11} = \left(1 - \frac{2GM}{rc^2}\right)^{-1} , \qquad g_{22} = r^2 , \qquad g_{33} = r^2 \sin^2\theta . \tag{18.6}$$

All other components of the metric are zero. The coordinates used here are

$$x^0 = ct , \quad x^1 = r \quad x^2 = \theta \quad x^3 = \phi . \tag{18.7}$$

As we will discuss in an upcoming lecture, this describes *exactly* the spacetime outside of a spherically symmetric, non-rotating body of mass M. Schwarzschild found this solution essentially in his spare time while serving as an artillery officer on the eastern front during the First World War. Shortly after submitting this solution for publication, he died of an autoimmune disorder that most believe was sparked by an infection he contracted while serving in the trenches. The fact that this solution existed and was found so quickly shocked Einstein, who did not expect anyone would manage to find relatively simple exact solutions — certainly not so quickly after the field equations were developed, and certainly not under such trying[1] circumstances.

This spacetime continues to play an important role in helping us to understand the limiting behavior of gravity; we will study it in some detail in coming lectures.

18.2.3 The Kerr metric

For decades, people wondered if there might be a more general exact solution than that provided by the Schwarzschild metric. What about near a body that is not spherical, or that is rotating? By the 1950s and 1960s, people were beginning to realize that one could take the Einstein field equation and treat it as a complicated differential equation that could be solved numerically, much as they were beginning to use computers to solve complicated differential equations describing things like fluid dynamics. As computers and computer programmers got more sophisticated, it became possible to study the Einstein field equations to build the spacetimes describing more interesting and complicating bodies. However, it seemed unlikely that a "closed form" solution for a body more complicated than spherical symmetry would ever be found.

[1] In a letter that Schwarzschild sent to Einstein, dated 22 December 1915, he wrote "As you see, the war treated me kindly enough, in spite of the heavy gunfire, to allow me to get away from it all and take this walk in the land of your ideas." He died a little less than 5 months later. (And I just realized that I defended my Ph.D. on the 82nd anniversary of his death, an odd bit of morbid trivia.)

That expectation held until 1963, when the mathematician Roy Kerr published the following glorious mess:

$$\begin{aligned}
g_{00} &= -\frac{\Delta - a^2\sin^2\theta}{\Sigma}\ , \quad g_{11} = \frac{\Sigma}{\Delta}\ , \quad g_{22} = \Sigma\ ,\\
g_{33} &= \left(\frac{(r^2+a^2)^2 - a^2\Delta\sin^2\theta}{\Sigma}\right)\sin^2\theta\ ,\\
g_{03} &= g_{30} = -\frac{2a\tilde{M}r}{\Sigma}\sin^2\theta\ ,
\end{aligned} \tag{18.8}$$

with all other metric components equal to zero, and where

$$\Delta = r^2 - 2\tilde{M}r + a^2 \quad \Sigma = r^2 + a^2\cos^2\theta\ , \quad \tilde{M} = \frac{GM}{c^2}\ , \quad a = \frac{J}{Mc}\ . \tag{18.9}$$

The coordinates used here are

$$x^0 = ct\ , \quad x^1 = r \quad x^2 = \theta \quad x^3 = \phi\ . \tag{18.10}$$

When Kerr originally published this solution, it wasn't actually clear what it meant. To be fair, he used a coordinate system which made it easier to solve the field equation, but made it less clear what the solution means; this form of the coordinates was published by Robert Boyer[2] and Richard Lindquist in 1967. If you set the parameter a to zero, it is not hard to show that the spacetime is identical to the Schwarzschild solution. After much study, it became clear that this solution describes a *black hole* with mass M and with spin angular momentum of magnitude $J = aMc$, oriented along the axis defined by $\theta = 0$. We will discuss this solution briefly, and explore a few simple analyses that can done in the Kerr metric.

18.2.4 The Friedmann-Lemaître-Robertson-Walker metric

Finally, an exact solution that describes *all* of spacetime filled with a fluid of density ρ and pressure P is given by

$$g_{00} = -1\ , \quad g_{11} = a^2(t)/(1-kr^2)\ , \quad g_{22} = a^2(t)r^2\ , \quad g_{33} = a^2(t)r^2\sin^2\theta\ . \tag{18.11}$$

This again uses the coordinates

$$x^0 = ct\ , \quad x^1 = r \quad x^2 = \theta \quad x^3 = \phi\ . \tag{18.12}$$

The function $a(t)$ is the solution to the differential equations

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3c^2} - \frac{k}{a^2}\ , \quad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2}(\rho + 3P)\ . \tag{18.13}$$

(Overdot denotes d/dt.) The parameter k takes one of three values — $+1$, 0, or -1. Which value of k describes our universe is something that must be determined from data; unpacking this is kind of complicated.

This solution was first found by the Soviet mathematician Alexander Friedmann in the early 1920s, although its significance was not broadly recognized prior to his death in 1925.

[2]Boyer was tragically murdered, along with 17 other people, in an infamous mass shooting event at the University of Texas a few months before the paper's publication.

Georges Lemaître, a Belgian priest and mathematician who earned a PhD in mathematics from MIT in 1923, rediscovered much of this solution in 1927. Via his efforts, people began to realize that this solution could be used as a tool for understanding the large-scale scale structure of the universe. Finally, Howard Robertson and Arthur Geoffrey Walker very thoroughly explored and described these spacetimes. Since the full cabal of discovers is a mouthful, this solution is often called the FRW (leaving out poor Lemaître) or FLRW metric.

The FLRW spacetime appears to give a good description of our universe on very long scales — tens of millions of light years, and over comparably long timescales. The "trick" is to come up with an appropriate description of the density and pressure that describes the "stuff" that characterizes the universe on such scales. This solution largely forms the foundation of the science of *cosmology*.

18.3 The Newtonian limit

18.3.1 The clocks of static observers

Let us begin our study of the consequences of general relativity with the solution that best describes spacetime near us: the "weak gravity" metric described in Sec. 18.2.1:

$$ds^2 = -\left(1 + \frac{2\Phi}{c^2}\right)c^2 dt^2 + \left(1 - \frac{2\Phi}{c^2}\right)\left(dx^2 + dy^2 + dz^2\right) . \tag{18.14}$$

We begin by thinking about the 4-velocity of an observer who is at rest in this spacetime; perhaps they are standing on the surface of the body that produces the gravitational potential Φ. How do we describe this observer's 4-velocity?

Since they are at rest in this spacetime, we require that $dx/d\tau = dy/d\tau = dz/d\tau = 0$. What remains is to figure out $dt/d\tau$. To deduce this, we insist that *exactly as in special relativity*, we must have $\vec{u} \cdot \vec{u} = -c^2$.

The reason we insist on this is because of Einstein's principle of equivalence: If we go into a freely falling frame, then everything behaves in spacetime *exactly as it did in special relativity*. We already know that $\vec{u} \cdot \vec{u} = -c^2$ in special relativity; and we know that the spacetime dot product is an invariant. We thus require that it have this form in **all** representations.

Enforcing this, we have

$$\begin{aligned} \vec{u} \cdot \vec{u} &= g_{\alpha\beta} u^\alpha u^\beta \\ &= -\left(1 + \frac{2\Phi}{c^2}\right) c^2 \left(\frac{dt}{d\tau}\right)^2 + 0 \\ &= -c^2 . \end{aligned} \tag{18.15}$$

Let's solve this for $dt/d\tau$, using the fact that the "weak gravity" metric requires $\Phi \ll c^2$:

$$\frac{dt}{d\tau} = \left(1 + \frac{2\Phi}{c^2}\right)^{-1/2} \simeq 1 - \frac{\Phi}{c^2} . \tag{18.16}$$

Let's take the source of the gravitational potential to be spherically symmetric and of mass M, so that $\Phi = -GM/r$. Let's consider two different observers: Observer 1 at height r_1 (say, the surface of the Earth) has a clock which measures time τ_1; observer 2 at height

$r_2 > r_1$ (some distance above the surface of the Earth) has a clock which measures time τ_2. Let's compare the rates at which their two clocks tick:

$$\begin{aligned}\frac{d\tau_2}{d\tau_1} &= \frac{dt/d\tau_1}{dt/d\tau 2}\\ &= \frac{(1+GM/r_1c^2)}{(1+GM/r_2c^2)}\\ &\simeq 1+\frac{GM}{c^2}\left(\frac{1}{r_1}-\frac{1}{r_2}\right) .\end{aligned} \tag{18.17}$$

Notice that since $r_2 > r_1$, this is positive: the clock of observer 2 ticks faster than the clock of observer 1. *This is exactly what we found based on our intuitive analysis of the light redshift effect.*

Before moving on, you might wonder — what does the coordinate t mean in this spacetime? We used it as an intermediate factor in order to compare the two observers' clocks, but the coordinate itself disappeared from the final analysis. To get some sense of this, notice that $dt/d\tau \to 1$ as $r \to \infty$. This means that the coordinate t is in fact proper time for an observer who is infinitely far away from the mass M. This tells us that t is time as measured on the clocks of very distant observers. We basically use t as a kind of "book-keeper" time; it's a time standard that everyone agrees on, no matter where they stand in spacetime. It facilitates making comparisons between different observers.

18.3.2 Falling down

Let's consider a body freely falling in the weak gravity spacetime (18.3). We begin by writing down the relativistic Lagrangian (per unit mass of the body) for this motion:

$$L = \frac{1}{2}g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta = -\frac{c^2}{2}\left(1+\frac{2\Phi}{c^2}\right)\left(\dot{t}\right)^2 + \frac{1}{2}\left(1-\frac{2\Phi}{c^2}\right)\left(\dot{x}^2+\dot{y}^2+\dot{z}^2\right) . \tag{18.18}$$

Here, an overdot denotes $d/d\tau$. Note that the potential Φ is independent of time, but depends on x, y, and z. Let's imagine a body that is falling along the z axis in this spacetime, so that $x = y = 0$, and see what applying the Euler-Lagrange equations tells us about the body's motion.

The equation of motion we need to examine is

$$\frac{\partial L}{\partial z} - \frac{d}{d\tau}\left(\frac{\partial L}{\partial \dot{z}}\right) = 0 . \tag{18.19}$$

Let's evaluate these derivatives:

$$\frac{\partial L}{\partial z} = -\left(\dot{t}\right)^2\frac{\partial\Phi}{\partial z} - \left(\frac{\dot{z}}{c}\right)^2\frac{\partial\Phi}{\partial z} , \tag{18.20}$$

$$\frac{\partial L}{\partial \dot{z}} = \left(1-\frac{2\Phi}{c^2}\right)\dot{z} , \tag{18.21}$$

$$\begin{aligned}\frac{d}{d\tau}\left(\frac{\partial L}{\partial \dot{z}}\right) &= \left(1-\frac{2\Phi}{c^2}\right)\ddot{z} - \frac{2\dot{z}}{c^2}\left(\frac{\partial\Phi}{\partial x}\dot{x}+\frac{\partial\Phi}{\partial y}\dot{y}+\frac{\partial\Phi}{\partial z}\dot{z}\right)\\ &= \left(1-\frac{2\Phi}{c^2}\right)\ddot{z} - \frac{2\dot{z}^2}{c^2}\frac{\partial\Phi}{\partial z} .\end{aligned} \tag{18.22}$$

To get Eq. (18.22), we used the chain rule to expand the total derivative along the falling body's trajectory. We then used the fact that we are taking the body to fall only in the z direction to set $\dot{x} = \dot{y} = 0$.

The equation of motion we have derived appears to be a mess. Let's put all the pieces together and see what we get. For clarity, let's write all the overdot terms explicitly as $d/d\tau$:

$$-\left(\frac{dt}{d\tau}\right)^2 \frac{\partial\Phi}{\partial z} - \left(1 - \frac{2\Phi}{c^2}\right)\frac{d^2z}{d\tau^2} + \frac{(dz/d\tau)^2}{c^2}\frac{\partial\Phi}{\partial z} = 0 \ . \tag{18.23}$$

Divide everything by $(dt/d\tau)^2$, and rearrange the terms:

$$\frac{d^2z}{dt^2} = -\frac{(1 - (dz/dt)^2/c^2)}{(1 - 2\Phi/c^2)}\frac{\partial\Phi}{\partial z} \ . \tag{18.24}$$

Finally, using the fact that this metric requires $\Phi \ll c^2$, we can write this as

$$\frac{d^2z}{dt^2} = -\frac{\partial\Phi}{\partial z}\left(1 + \frac{2\Phi}{c^2} - \frac{(dz/dt)^2}{c^2} - 2\frac{\Phi(dz/dt)^2}{c^4}\right) \ . \tag{18.25}$$

The leading approximation to this equation is simply

$$\begin{aligned}\frac{d^2z}{dt^2} &= -\frac{\partial\Phi}{\partial z} \\ &= -\frac{GM}{r^3}z \ . \end{aligned}\tag{18.26}$$

This is nothing more than the Newtonian limit: the acceleration of a body falling in the spacetime (18.3) is given by minus of the gradient of the gravitational potential. Doing this calculation without assuming that the body is falling along the z axis yields the equation of motion,

$$\frac{d^2\mathbf{x}}{dt^2} = -\frac{GM}{r^3}\mathbf{x} \ . \tag{18.27}$$

This **exactly** reproduces Newtonian gravity.

It's worth noting that if this result had *not* been found, we would not be having this discussion today. Newtonian gravity works quite well over a wide range of important situations, and it was *necessary* for the relativistic version of gravity to reproduce Newton's successes. In fact, when we do a more complete derivation of the Einstein field equation, we use the fact that the theory should reproduce the Newtonian limit to pin down the constant of proportionality $8\pi G/c^4$ in the field equation.

What about those terms we've neglected in going from (18.25) to (18.26)? Notice that they introduce corrections to Newtonian gravity; notice also that each such term involves factors of $1/c^2$. That's a signal that they can be thought of as "relativistic corrections" to the leading result. For example, the first term we've neglected has a value at the Earth's surface of

$$\frac{2\Phi}{c^2} = \frac{2GM_{\rm Earth}}{c^2 R_{\rm Earth}} \simeq 1.38 \times 10^{-9} \ . \tag{18.28}$$

This term introduces a roughly part per billion correction to gravitational acceleration. The second term is of order the small body's speed squared divided by c^2; the third is the product of those two corrections.

For the vast majority of applications, those corrections are negligible — indeed, measuring them at all is not easy. However, Einstein thought it might be interesting to include their influence and see what effect they have on the motion of bodies moving under the influence of gravity. He was motivated by the fact that for centuries people had been wondering how to resolve a mystery regarding Mercury's orbit. It was well known that an orbit in Newtonian gravity — i.e., an orbit governed by Eq. (18.27) — would be a closed ellipse, *if* we had a single small body orbiting a single large body. It was also well known that if the system was more complicated than this simple two-body setup, then the ellipse wouldn't quite close — it would precess, with the axis along its long direction slowly rotating with time.

Mercury's orbit is determined mostly by the gravity of the Sun, but it is perturbed by other planets in the solar system — especially Venus and Earth (which are fairly close by), and Jupiter (which is very massive). During the 19th century, a lot of mathematical techniques were perfected figuring out to account for the actions of these planets on Mercury's orbit. After a lot of back and forth, the consensus emerged: Mercury's orbit should precess by 5556 arcseconds per century.

To the great consternation of natural philosophers in the 19th century, the data do not quite bear this out. Over many decades of observation it became clear that Mercury's orbit precessed a little too fast, giving us a measured rate of 5599 arcseconds per century. A discrepancy of 43 arcseconds per century was clearly present in Mercury's orbit data.

Many hypotheses were advanced to explain this, including the idea that a planet provisionally named Vulcan[3] occupied an orbit very close to the Sun, inside Mercury's orbit. None of them worked. Einstein was curious what happens if he turned the crank on Eq. (18.25), including terms which are of order $1/c^2$. With some effort, and focusing on a bound orbit in the spacetime (18.3), one can show that the equation of motion becomes

$$\frac{d^2\mathbf{x}}{dt^2} = -\frac{GM}{r^3}\left(1 + \frac{|\mathbf{v}|^2}{c^2}\right)\mathbf{x} + \frac{4GM(\mathbf{x}\cdot\mathbf{v})\mathbf{v}}{c^2 r^3} + \mathcal{O}\left(\frac{1}{c^4}\right) . \tag{18.29}$$

(The $\mathcal{O}(1/c^4)$ in this equation means that the next term, which we are ignoring, involves things that scale with $1/c^4$.) With a little effort, one can show that an orbit governed by this equation of motion is described by a precessing ellipse. When applied to Mercury's orbit, the rate at which the angle of the orbit's ellipse rotates is given by

$$\frac{d\phi}{dt} = \frac{6\pi G M_\odot}{a(1-e^2)Pc^2} . \tag{18.30}$$

In this equation, $M_\odot = 1.99 \times 10^{30}$ kg is the mass of the sun; $a = 57.9 \times 10^6$ km is the semi-major axis of Mercury's orbit; $e = 0.2$ is the eccentricity of the orbit; and $P = 88$ days is the period of the orbit. Plugging all these numbers in, using 36,524 days per century, we find the rate of advance of Mercury's orbital ellipse due to relativistic corrections:

$$\frac{d\phi}{dt} = 0.000208\,\text{radians/century} . \tag{18.31}$$

There are 2π radians in 360 degrees; there 3600 arcseconds in one degree. Hence, there are $360\cdot 3600/2\pi = 206{,}265$ arcseconds per radian. Converting units, Einstein found that general relativity's prediction for the "extra" precession of Mercury's orbit is

[3]Proposed *way* earlier than Gene Roddenberry's time.

$$\frac{d\phi}{dt} = (0.000208\,\text{radians/century})\left(2.063\times10^5\,\text{arcseconds/radian}\right)$$
$$= 42.9\,\text{arcseconds/century}\ . \qquad (18.32)$$

Further refinements to these numbers only improves the fit. In one fell swoop, Einstein managed to explain a phenomenon that puzzled many of the most significant mathematicians and physicists of history. No wonder that in a letter to his friend and colleague Paul Ehrenfest shortly after completing this calculation, he wrote "I was beside myself with ecstasy for days."

18.4 Addendum: Other attempts to make relativistic gravity

As emphasized at the beginning of this discussion, we should take general relativity as described by the field equation $G^{\mu\nu} = (8\pi G/c^4)T^{\mu\nu}$ as a *hypothesis*, one that must be tested by comparing with data. It was not inevitable that we would end up with what we now call general relativity. Here is a brief discussion of a few alternates that were considered, and why we they didn't hold up.

- Motivated by the idea that one can needs to make $\nabla^2\Phi = 4\pi G\rho_M$ something that makes sense in Lorentz frames, the Swedish/Finnish physicist Gunnar Nordström proposed that gravity acts via a scalar field Φ which, in the language we are now using, satisfies the differential equation
$$\Box\Phi = -\frac{4\pi G}{c^4}\Phi^5\eta_{\alpha\beta}T^{\alpha\beta}\ . \qquad (18.33)$$
(Note, it's possible I have botched a few factors! In particular, I haven't carefully checked the powers of Φ on the right-hand side. The form in which this theory appears in textbooks involves using some quantities which would be a big detour for us to introduce and discuss here; I don't guarantee that I've translated this with 100% accuracy.) With a little effort, it can be shown that this yields an equation of motion that looks like
$$\frac{d(\Phi u_\alpha)}{d\tau} = -\frac{\partial\Phi}{\partial x^\alpha}\ . \qquad (18.34)$$
In the limit of $\Phi \ll 1$, this reproduces Newtonian gravity, and correctly produces the redshifting of light. However, it turns out to get Mercury's precession wrong; and, it predicts that light rays do not change direction under the influence of gravity. The bending of light by gravity was a particularly important early triumph of Einstein's version of relativistic gravity.

- Motivated by the idea that $\mathbf{F}_g = -Gm_1m_2\mathbf{x}/r^3$ looks an awful lot like the Coulomb interaction, perhaps we can define a quantity like the Faraday tensor which describes gravity. In short, one might wish to construct a Maxwell-equation-like theory of gravity.

  This can be done, but the result turns out to be *theoretically* inconsistent. Whenever one makes an interaction relativistic, one finds that it predicts the interaction produces radiation. This is a simple consequence of causality: If we "shake" the source of the interaction (e.g., charges for electric and magnetic fields, masses for gravity), the outcome of this shaking can be communicated to distant observers no faster than the speed of light. Indeed, all relativistic theories of gravity predict that some form of gravitational radiation must exist.

> When we do this for a “Maxwell-like gravity,” the radiation that it produces has a very weird feature: the radiation that it produces has *negative energy density*. This means that in this theory, if I have a dynamical system that produces radiation, it carries away “negative energy” from the system. Taking away “negative energy” is the same thing as *adding* energy. The dynamics that made the system radiate in the first place thus become **more** energetic — making the radiation have higher amplitude, which means they carry away **more** negative energy, thus making the system even **MORE** energetic.
>
> Such a description of gravity turns out to be catastrophically unstable — any dynamics would almost immediately become grow without bound, destroying the system. Since we do not observe this (indeed, since we exist in order to observe that this does not happen), we reject the Maxwell-like theory of gravity. (Details of this analysis can be found in exercise 7.2 in the textbook *Gravitation* by Misner, Thorne, and Wheeler. It is not a simple exercise!)

Though ideas of this kind didn’t hold up, we haven’t stopped thinking about ways in which Einstein’s general relativity may not quite meet the mark. Precisely because gravity is the weakest fundamental interaction, it is extremely difficult to test. It’s worth noting that the gravitational constant G is the least precisely determined of the main “fundamental constants” of nature — although the product of G with certain masses is quite well known, simply because that product is what enters many observable formulas. For example, although G is known to about 5 digits, $GM_\odot$ is known to about 10 digits.

Thinking about plausible modifications to general relativity, and coming up with experimental methods for testing them, is among the topics that are at the vanguard of modern physics research.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 19
From weak gravity to strong gravity

19.1 A "strong gravity" spacetime

In the previous lecture, we described a few exact solutions that have been found to Einstein's field equations of general relativity, and we discussed in some detail how things behave in the spacetime that describes "weak" gravity. For a spherical body, this can be taken to be gravity for which $GM/rc^2 \ll 1$ everywhere. We found that in this spacetime:

- freely-falling bodies move in a way that reproduces the predictions of Newtonian gravity;
- clocks "lower" in the spacetime (i.e., at smaller r) tick more slowly than those at higher altitudes in a way that is exactly consistent with the redshift of light[1];
- although we skipped over many of the details, terms beyond the ones which reproduce the predictions of Newtonian gravity explain a centuries-old mystery about the precession of Mercury's orbit about the Sun;
- and finally, as you will show on problem set #9, the trajectory of light bends as it passes near a gravitating body. A celebrated measurement by Dyson and Eddington in 1919 confirmed[2] the predictions of general relativity; indeed, the publicity[3] surrounding the light-bending measurement was a huge part of what turned Albert Einstein from a highly respected scientist into an international public figure.

These items went a long way toward convincing most scientists that general relativity provides a valid relativistic theory of gravity. Most people are happy to work under the assumption that gravity is described by spacetimes which solve the equation $G^{\mu\nu} = (8\pi G/c^4)T^{\mu\nu}$.

However, as we noted in the previous lecture, this is not the *only* way to combine relativity with gravity. Indeed, as was briefly described in Lecture 18, there's a certain sense in which general relativity can be regarded as the *simplest* theory of relativistic gravity. Perhaps differences between theory and measurement will arise as we investigate strong gravity —

[1] We didn't actually look at light propagation yet; we will do that in this lecture.

[2] There has been some controversy about whether this measurement's error bars are as good as was claimed. Independent of that controversy (which has been thoroughly investigated; the consensus is that the measurement by Dyson and Eddington was valid, though it is worth digging into the details), the bending of light by gravity has been thoroughly examined many times since 1919, and general relativity's predictions hold up. Indeed, they hold up so well that these days people *assume* that general relativity correctly describes light bending, and use it to learn about the properties of large distributions of mass by measuring how light bends. This is what the astronomical science of *gravitational lensing* is all about.

[3] In no small part because an expedition by British scientists to examine what was then regarded as a German theory was treated as a welcome example of the scientific community setting aside the antagonism of World War I to focus on truths that transcend national borders.

after all, if you want to push the boundaries in physics, you take the framework in which you interpret your measurements and either figure out how to measure things with greater and greater precision, or you push into regimes beyond what you have already investigated (or both).

In this lecture, we're going to explore what general relativity tells us about when gravity is not weak — i.e., in situations where it is not the case that $GM/rc^2 \ll 1$. Our tool for this exploration is the Schwarzschild metric, for which the line element takes the form

$$ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta = -\left(1 - \frac{2GM}{rc^2}\right)(c\,dt)^2 + \frac{dr^2}{(1 - 2GM/rc^2)} + r^2\left(d\theta^2 + \sin^2\theta\,d\phi^2\right) \ . \quad (19.1)$$

This spacetime is exact, and holds for all r. By using the full mathematical machinery of general relativity, one finds that (19.1) exactly describes a spacetime for which $T^{\mu\nu} = 0$. However, this spacetime also describes the spherically symmetric gravity of a mass M.

What this is telling us is that Eq. (19.1) describes the gravity of a mass M, but there's no matter or energy density anywhere. So, what does *that* mean? Perhaps the simplest way of understanding this (admittedly counterintuitive) aspect of the Schwarzschild solution is by analogy. If you take the Coulomb point charge electric field,

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^3}\mathbf{x} \ , \quad (19.2)$$

and apply the divergence operator to it, you get zero. This means that the charge density everywhere is zero:

$$\rho = \epsilon_0\left(\nabla \cdot \mathbf{E}\right) = 0 \ . \quad (19.3)$$

So there's no charge density ... but when we integrate it up, we get a non-zero charge q.

The resolution of this apparent paradox in electrostatics is that the divergence is actually an ill-behaved operation exactly at the origin, $\mathbf{x} = \mathbf{0}$. In courses like 8.07, we learn that we can resolve this by introducing a singular "function"[4] that essentially puts a finite amount of charge into a zero-volume point at the origin. At least heuristically, something similar is going on with the Schwarzschild spacetime — at least in classical general relativity, there's a singular point at the coordinate $r = 0$ where general relativity's equations are ill-behaved. But everywhere away from that point, there is no problem.

Thanks to non-linear terms in Einstein's field equations, the $r = 0$ singularity is even more disturbing and hard to deal with than the analogous Coulomb singularity. Nonetheless, it is useful to set aside misgivings about this spacetime and examine what it tells us. (Indeed, an aspect of the spacetime's nature we will soon investigate suggests that any "weirdness" near $r = 0$ is not of concern — at least, not of immediate concern. We will elaborate on this cryptic remark soon.)

Let us begin by again looking at an observer who is at rest in the spacetime, and think about how their clocks behave. Notice that as $r \to \infty$, the Schwarzschild metric is nothing more than the metric of special relativity (albeit in spherical coordinates — you can transform from the inertial coordinate form we've long been using by the transformations $x = r\sin\theta\cos\phi$, $y = r\sin\theta\sin\phi$, $z = r\cos\theta$). This tells us that, as in the "weak gravity" metric of the previous lecture, t describes clocks that are used by very distant observers. This means t makes a useful "bookkeeper" time for comparing different observers' clocks.

[4]Strictly-speaking, the quantity we use is not a function, but it can be treated much like a function if we are careful. If this is new to you and you are curious about this, look up the *Dirac delta function*. This is a topic for another day, and another course.

Let's compare the bookkeeper time with the time of an observer who is spatially at rest at some radius r. We put $u^r = 0$, $u^\theta = 0$, $u^\phi = 0$; invoking the principle of equivalence, we require $\vec{u} \cdot \vec{u} = -c^2$ to solve for $u^t = c\,dt/d\tau$:

$$\vec{u} \cdot \vec{u} = -\left(1 - \frac{2GM}{rc^2}\right)\left(c\frac{dt}{d\tau}\right)^2 = -c^2 \ , \tag{19.4}$$

which means

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{1 - 2GM/rc^2}} \qquad \text{or} \qquad \Delta\tau(r) = \Delta t\sqrt{1 - \frac{2GM}{rc^2}} \ . \tag{19.5}$$

Notice that if $r \gg 2GM/rc^2$, we can use the binomial expansion and approximate:

$$\sqrt{1 - \frac{2GM}{rc^2}} \simeq 1 - \frac{GM}{rc^2} \qquad \text{for} \qquad r \gg 2GM/rc^2 \ . \tag{19.6}$$

At clock located at coordinate r ticks slower than a clock that is very far away by a factor GM/rc^2, exactly the variation in clock ticking that we found in the weak-gravity metric. This confirms that the Schwarzschild metric agrees with our previous results in the right limit. However, the rate at which clocks slow as r gets slower is *far* more extreme than what we saw in the weak-field case (remember that the weak field formula was only valid if $r \gg GM/c^2$ everywhere). Indeed, (19.5) predicts that our observer's clock *stops* as $r \to 2GM/c^2$ — and it appears to break down completely when $r < 2GM/c^2$.

So what is going on with *that*??

19.2 Light propagation

The propagation of light was one of our most important tools for making sense of how space and time behave in special relativity. Light propagation helps us in general relativity too, though we need to lay out a few rules for how we are going to use it.

We cannot define 4-velocity along a light ray — because the speed is c, proper time is not defined along it. However, 4-momentum is perfectly well defined along a light ray. Let us look at the 4-momentum of a body with mass m:

$$\vec{p} = m\frac{d\vec{x}}{d\tau} \ . \tag{19.7}$$

Let us define a parameter λ such that $d\lambda = d\tau/m$. Then,

$$\vec{p} = \frac{d\vec{x}}{d\lambda} \ . \tag{19.8}$$

If we consider a sequence of bodies with ever decreasing m, we can define the 4-momentum of light to be $\vec{p} = d\vec{x}/d\lambda$ in the limit $m \to 0$. The parameter λ can then be regarded as a kind of "tick mark" that allows us to label events along a light ray, with units chosen so that $d\vec{x}/d\lambda$ yields a quantity with the units of momentum.

Since the Schwarzschild spacetime is spherically symmetric, let's examine light rays that propagate radially, setting $p^\theta = p^\phi = 0$. The defining characteristic of a null or light-like

4-momentum in special relativity was that $\vec{p} \cdot \vec{p} = 0$. Invoking the principle of equivalence, the same thing holds in general relativity:

$$\vec{p} \cdot \vec{p} = g_{\alpha\beta} \frac{dx^\alpha}{d\lambda} \frac{dx^\beta}{d\lambda} = -\left(1 - \frac{2GM}{rc^2}\right)\left(c\frac{dt}{d\lambda}\right)^2 + \left(1 - \frac{2GM}{rc^2}\right)^{-1}\left(\frac{dr}{d\lambda}\right)^2 = 0 \,. \tag{19.9}$$

Using this, we can solve for the speed at which light propagates in this coordinate system:

$$\frac{dr}{dt} = \pm c\left(1 - \frac{2GM}{rc^2}\right) \,. \tag{19.10}$$

Notice this appears to tell us that the light is not propagating at speed c! Please bear in mind, however, that this is the light's speed *in this coordinate system*. Equation (19.10) expresses the ratio of an interval of radial coordinate r to an interval of coordinate time t. Consider two events: one is at (t, r, θ, ϕ); the other is at $(t, r + dr, \theta, \phi)$. The distance between these events is given by $ds = dr/\sqrt{1 - 2GM/rc^2}$. This distance is *larger* than dr. So when the light moves through a *coordinate* distance dr, the spatial distance it moves is greater than dr. Note also that this speed is defined in terms of the time used by observers who are very far away. The clocks of observers near r tick more slowly than the clocks of distant observers. With a little effort, one can show that observers will always see light move with speed c when things are expressed as physical distance divided by their own time. The idea that the speed of light is c for all observers has not been broken; indeed, thanks to the principle of equivalence, it remains foundational to this subject.

That said, Eq. (19.10) has very interesting behavior in the limit $r \to 2GM/c^2$ — the coordinate velocity there is *zero*. That suggests that a light ray "launched" radially outward (or inward, for that matter) at $r = 2GM/c^2$ will stay there forever. This appears to contradict the principles outlined in the previous paragraph. However, recall from Eq. (19.5) that an observer's clock stops relative to a distant clock when we reach this radius. The radius $r = 2GM/c^2$ is indeed special, and a bit weird. More on this radius below.

Let's look at one more aspect of the behavior of light — its energy as it propagates outwards from some radius. Before doing this, it is very useful to pause and look at the *Lagrangian* for light propagating in the Schwarzschild spacetime. We defined $L = g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta/2$ as the Lagrangian for material bodies moving through the spacetime $g_{\alpha\beta}$ with $\dot{x}^\alpha \equiv dx^\alpha/d\tau$. By adjusting the definition so that $\dot{x}^\alpha \equiv dx^\alpha/d\lambda$, the Euler-Lagrange equations

$$\frac{\partial L}{\partial x^\alpha} - \frac{d}{d\lambda}\left(\frac{\partial L}{\partial \dot{x}^\alpha}\right) = 0 \tag{19.11}$$

can then be used to describe light moving through the spacetime.

The Lagrangian for a light ray is given by

$$L = \frac{1}{2}\left[-\left(1 - \frac{2GM}{rc^2}\right)\left(c\dot{t}\right)^2 + \frac{(\dot{r})^2}{(1 - 2GM/rc^2)} + r^2(\dot{\theta})^2 + r^2\sin^2\theta(\dot{\phi})^2\right] \,, \tag{19.12}$$

where $\dot{x}^0 \equiv c\dot{t} = c\,dt/d\lambda = p^t$, $\dot{r} = dr/d\lambda = p^r$, etc.

Notice that $\partial L/\partial x^0 = (1/c)\partial L/\partial t = 0$. By one of the exercises you did on problem set #8, this tells us that

$$\frac{1}{c}\frac{\partial L}{\partial \dot{t}} = -\left(1 - \frac{2GM}{rc^2}\right)c\dot{t} = \text{constant} \,. \tag{19.13}$$

Using the fact that $c\dot{t} = p^t$ and $-(1 - 2GM/rc^2) = g_{tt}$, this tells us that along the light ray

$$g_{tt}p^t \equiv p_t = \text{constant} . \tag{19.14}$$

The *downstairs* t component of the 4-momentum, p_t, is a constant along the light ray's trajectory.

Let's use this to compare the energy that is measured by a static observer at $r = R$ with an observer who is very far away, $r \to \infty$. We use the fact that the energy measured by an observer whose 4-velocity is $\vec{u}$ is given by $E_{\vec{u}} = -\vec{p} \cdot \vec{u}$ — by the equivalence principle, this result (which we developed in special relativity) will work just fine for us in general spacetimes. We use the fact that an observer who holds static at $r = R$ has a 4-velocity with components $u^t = c(dt/d\tau) = c/\sqrt{1 - 2GM/rc^2}$, $u^r = u^\theta = u^\phi = 0$. So then

$$\begin{aligned}
\frac{E(r \to \infty)}{E(r = R)} \equiv \frac{E_\infty}{E_R} &= \frac{-\vec{p} \cdot \vec{u}\big|_{r\to\infty}}{-\vec{p} \cdot \vec{u}\big|_{r=R}} \\
&= \frac{p_t u^t(r \to \infty)}{p_t u^t(r = R)} \\
&= \frac{1}{1/\sqrt{1 - 2GM/Rc^2}} \\
&= \sqrt{1 - \frac{2GM}{Rc^2}} .
\end{aligned} \tag{19.15}$$

The first line of this relation just inserts the definition $E = -\vec{p} \cdot \vec{u}$. The second line expands the inner product, using the downstairs form of the 4-momentum and the upstairs form of the 4-velocity, taking advantage of the fact that only $u^t \neq 0$. On the third line, we use the fact that p_t is a constant along the light ray's trajectory to cancel it out — p_t has the same value at $r = R$ as it does in the limit $r \to \infty$. We also use the solution for u^t that we derived earlier in this lecture.

The final line shows us how light is redshifted as propagates from $r = R$ out to infinity. Notice once again the interesting behavior as $R \to 2GM/c^2$: the light is so redshifted in this case that the energy measured very far away is *zero*. **None of the light's energy gets out if it starts at $R = 2GM/c^2$.**

To summarize, our investigation of the Schwarzschild spacetime has yielded the following outcomes:

- Clocks run slower at smaller values of r. If $d\tau_R$ is an interval of time measured at $r = R$, and dt is an interval measured by clocks very far away ($r \to \infty$), then we find

$$d\tau_R = dt\sqrt{1 - \frac{2GM}{Rc^2}} . \tag{19.16}$$

- Light that is emitted from $r = 2GM/c^2$ appears to move in the radial direction with coordinate speed $dr/dt = 0$. In other words, light does not seem to ever move away from this radius.

- If light is emitted from radius $r = R$ with energy E_R, then it is measured far away to have energy

$$E_\infty = E_R\sqrt{1 - \frac{2GM}{Rc^2}} . \tag{19.17}$$

This is consistent with the redshifting of light we saw in other contexts, but notice that $E_\infty \to 0$ as $R \to 2GM/c^2$.

This all tells us that there is something quite interesting about the radius $r = 2GM/c^2$. Let's do one more calculation, which if all goes well will really confuse us.

19.3 The trajectory of an infalling observer

Imagine an observer who starts at rest from $r = R$ and then falls. Suppose they have no motion in the θ or ϕ directions. The Lagrangian describing their motion is then given by

$$L = \frac{1}{2} g_{\alpha\beta} u^\alpha u^\beta = -\frac{1}{2}\left(1 - \frac{2GM}{rc^2}\right)\left(c\frac{dt}{d\tau}\right)^2 + \frac{1}{2}\frac{(dr/d\tau)^2}{1 - 2GM/rc^2} . \tag{19.18}$$

On problem set #8, you found that because $\partial L/\partial t = 0$, it must be the case that $\partial L/\partial \dot{t}$ is a constant along the body's trajectory. We call this constant the body's energy per unit mass (up to a minus sign) because of its limiting behavior as $r \to \infty$:

$$E = -\frac{\partial L}{\partial \dot{t}} = c^2\left(1 - \frac{2GM}{rc^2}\right)\frac{dt}{d\tau} = \text{constant} . \tag{19.19}$$

This observer starts at rest, and we know that for an observer who is at rest in the Schwarzschild spacetime

$$\frac{dt}{d\tau} = \left(1 - \frac{2GM}{rc^2}\right)^{-1/2} . \tag{19.20}$$

Applying this to our infalling observer when they are at rest at $r = R$, we find

$$E_{\rm obs} = c^2\sqrt{1 - \frac{2GM}{Rc^2}} . \tag{19.21}$$

We also know that $\vec{u} \cdot \vec{u} = -c^2$:

$$-c^2 = -\left(1 - \frac{2GM}{rc^2}\right)\left(c\frac{dt}{d\tau}\right)^2 + \frac{(dr/d\tau)^2}{1 - 2GM/rc^2} . \tag{19.22}$$

We can clean this up, using Eq. (19.19) to replace $dt/d\tau$ with $E_{\rm obs}$ and a function of r. After making this substitution, we can rearrange to make an equation describing the infalling observer's trajectory with respect to r:

$$\begin{aligned} \left(\frac{dr}{d\tau}\right)^2 &= \frac{E_{\rm obs}^2}{c^2} - c^2\left(1 - \frac{2GM}{rc^2}\right) \\ \longrightarrow \quad \frac{dr}{d\tau} &= -\sqrt{\frac{2GM}{r} - \frac{2GM}{R}} . \end{aligned} \tag{19.23}$$

The second line of Eq. (19.23) uses the value of $E_{\rm obs}$ we found above; we choose an overall minus sign for the square root to give us infall.

Equation (19.23) is most easily solving by finding $\tau(r)$ — i.e., the elapsed proper time that passes after the observer has fallen from R to r. The result is

$$\tau = \sqrt{\frac{1}{2GM}}\left[R^{3/2}\arctan\left(\sqrt{\frac{R-r}{r}}\right) + \sqrt{rR(R-r)}\right] . \tag{19.24}$$

This tells us that the observer falls on a rather smooth trajectory according to their own clocks, reaching $r = 0$ in finite proper time:

$$\Delta\tau(r = R \to r = 0) = \frac{\pi}{2}\sqrt{\frac{R^3}{2GM}} \ . \tag{19.25}$$

Despite the fact that $r = 2GM/c^2$ seems to be quite important, nothing special happens here as r passes through this coordinate.

Parameterizing the motion in terms of the observer's proper time is fine for discussing how they see their own motion. But how does it look to a distant observer, someone who is watching that person fall in from a safe distance? Very distant observers use the coordinate t for their clocks, and an interesting question is how the motion looks when parameterized in a way that suits their perspective. We know that the infalling observer's clocks "run slow" according to distant observers. We thus expect that a process which happens quickly according to the infalling observer's clock may not look quite so fast as seen by someone very far away.

We begin by working out the infall as parameterized by t:

$$\begin{aligned}\frac{dr}{dt} &= \frac{dr}{d\tau}\left(\frac{dt}{d\tau}\right)^{-1} \\ &= -\sqrt{2GM\left(\frac{1}{r} - \frac{1}{R}\right)}\frac{c^2}{E_{\text{obs}}}\left(1 - \frac{2GM}{rc^2}\right) \ .\end{aligned} \tag{19.26}$$

Using $E_{\text{obs}} = c^2\sqrt{1 - 2GM/Rc^2}$, we can solve this for $t(r)$. The solution is the rather more complicated expression

$$\begin{aligned}t(r) = \frac{2GM}{c^3}\ln\left[\frac{\sqrt{\frac{r(R-2GM/c^2)}{2GM(R-r)/c^2}} + 1}{\sqrt{\frac{r(R-2GM/c^2)}{2GM(R-r)/c^2}} - 1}\right] + \sqrt{r(R-r)\left(\frac{Rc^2}{2GM} - 1\right)} \\ + \left(R + \frac{4GM}{c^2}\right)\sqrt{\frac{Rc^2}{2GM} - 1}\left[\frac{\pi}{2} - \arctan\left(\sqrt{\frac{r}{R-r}}\right)\right] \ .\end{aligned} \tag{19.27}$$

This leads to a *very* different description of the infalling body's motion! Let's look at this function as $r \to 2GM/c^2 + x$: being very careful with our expansions, we find that as $x \to 0$,

$$t(x) \to \frac{2GM}{c^3}\ln\left[\frac{8GM(R - 2GM/c^2)}{Rc^2x} + \mathcal{C}_1\right] + \mathcal{C}_2 \ . \tag{19.28}$$

The quantities $\mathcal{C}_{1,2}$ are constants whose precise values depend on the starting radius R, but are not important for us right now. In particular, note that the influence of the constant $\mathcal{C}_1$ becomes negligible as x gets small. Neglecting $\mathcal{C}_1$, we can easily rearrange this to find x as a function of t:

$$x \to \frac{8GM(R - 2GM/c^2)}{Rc^2}\exp\left[-(t - \mathcal{C}_2)c^3/(2GM)\right] \ . \tag{19.29}$$

The infalling body only asymptotically approaches $r = 2GM/c^2$ as $t \to \infty$.

To nail this home, let's plot the motion according to these two time parameterizations:

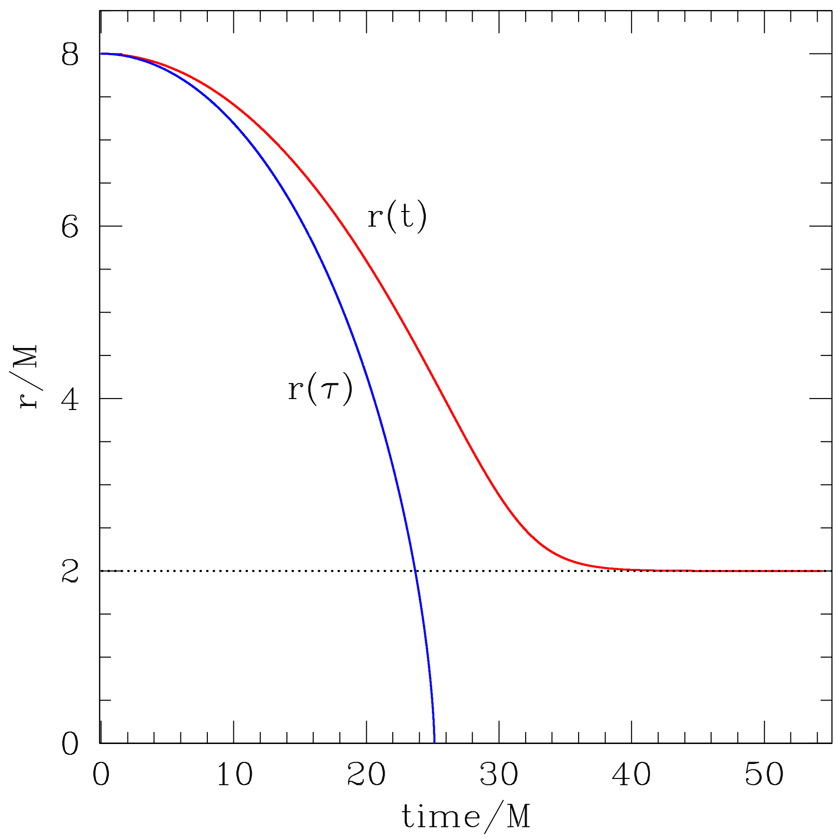


Figure 1: Infall trajectory from $R = 8GM/c^2$, parameterized by the infalling observer's time τ [using Eq. (19.24)] versus the trajectory parameterized by distant observer time t [using Eq. (19.27)]. Adapted from the course notes for 8.962; include a multiplicative factor of G/c^2 on the M on the vertical axis, and a factor of G/c^3 on the M on the horizontal axis.

We have two very different pictures: According to the observer's own proper time, they more or less plummet merrily along, reaching $r = 0$ in short order according to their own clocks. (Incidentally, gravity diverges at $r = 0$, so that's not a very happy place to wind up.) But according to the clocks of very distant observers, they never get anywhere *close* to $r = 0$. Indeed, they only asympotically approach $r = 2GM/c^2$, reaching it only as $t \to \infty$ according to those observers.

A favorite saying of Einstein's was *Raffiniert ist der Herr Gott, aber boschaft ist er nicht* — "Subtle is the Lord, but malicious he is not." This figure seems to reveal a side of Nature that is not only malicious but positively perverse. A driving principle throughout this course has been that the view of two different observers must be *consistent* — perhaps they differ in some details, but they agree on physical outcomes. Can we possibly reconcile these two vastly different viewpoints consistent?

The answer will be yes, and the reconciliation is subtle. Doing so will turn on thinking very carefully how the observer who is very far away is observing the infalling observer.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 20
Exploring strong gravity

20.1 Overview

In this set of notes, we are going to explore some of the ways in which observational tests of the unique predictions of strong gravity can be formulated. These notes are a little on the long side, and will probably be delivered over the course of multiple lectures. They also rather dense, and involve calculations whose details require a bit of care. Every one of these calculations is, at heart, nothing more than an examination of a "geodesic" (the trajectory followed by a freely-falling body or the propagation of light) in the strong-field spacetime. The recipe for performing such a calculation is always the same — write down the Lagrangian for motion in the spacetime, apply the Euler-Lagrange equations. The core ideas underlying these calculations are hopefully clear to you; do not worry if the amount of information is a bit too much to follow. We summarize the key points that we hope you take away from this discussion at the end of these notes.

20.2 Weirdness of infall according to two different observers

Our discussion of motion in the Schwarzschild spacetime reached the point where we looked at an infalling observer: someone who is at rest at $r = R$, then falls. What we found is that this observer's motion as a function of time looks radically different depending on what "time" means. If we parameterize the observer's trajectory using proper time τ, i.e. the time that the observer measures on their own clocks, we find a trajectory that is a simple function relating their coordinate position r with their measured proper time τ. This parameterization shows that the infalling observer reaches $r = 0$ in finite proper time.

On the other hand, if we parameterize the observer's motion using coordinate time t, which describes time as measured on the clocks of observers who are very far away, we get a very different picture. With that parameterization, the infalling observer never crosses $r = 2GM/c^2$, let alone reaches $r = 0$. Instead, we see them asymptotically approaching $r = 2GM/c^2$ as $t \to \infty$. This behavior is shown in Fig. 1.

From the exact solution for $r(t)$ written down in the previous lecture, examine how things behave near $r = 2GM/c^2$. Putting $r = 2GM/c^2 + \delta r$, it is not too hard to show that

$$\delta r = \frac{8GM(R - 2GM/c^2)}{Rc^2} e^{-(t-\mathcal{C}_2)c^3/(2GM)} . \tag{20.1}$$

The constant $\mathcal{C}_2$ depends on the initial condition; its precise value is not important for us. What this expansion shows us is that at late times in the t parameterization, the infalling observer gets closer to $2GM/c^2$ by a factor of e for every time interval $2GM/c^3$. This is a very short time. For example, if M is the mass of the Sun, $2GM/c^3$ is roughly 10 microseconds.

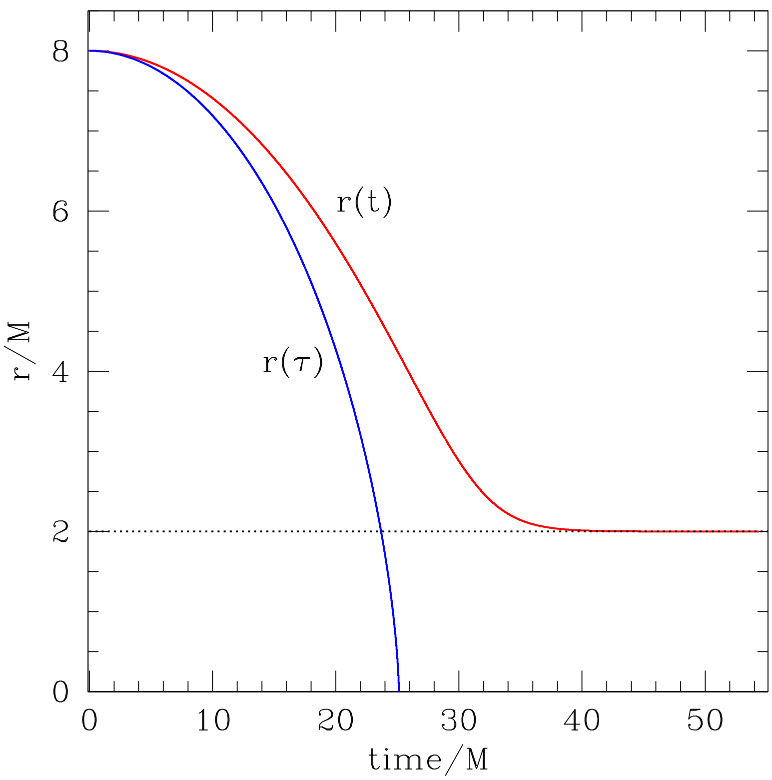


Figure 1: Infall trajectory for $R = 8GM/c^2$, parameterized by the infalling observer's proper time τ, and parameterized by distant observer time t. Figure taken from lecture notes for 8.962; include a multiplicative factor of G/c^2 to the M on the vertical axis, and a factor of G/c^3 to the M on the horizontal axis.

Throughout our study of relativity, we have encountered situations in which two observers measure different things. We've learned not to be too bothered by this, but have learned instead to try to find a way to make sure that measurements, though perhaps not in *agreement* (two observers measure different lengths and see events happen in a different order; one measures a pure magnetic field, another measures a combination of electric and magnetic fields) are nonetheless *consistent*. We have found that the observers agree on the nature of important events at the end of the analysis (a long pole moving very fast never collides with a door; an electron feels a force which causes it to accelerate).

But making consistent the two pictures of the infall-in-Schwarzschild scenario that we've painted seems like a tall order. How can we reconcile falling all the way to $r = 0$ in one parameterization (where, incidentally, gravitational tidal stresses get so strong that anyone or anything will be shredded) with "hovering" near $r = 2GM/c^2$ for all eternity in another?

We do this by thinking carefully about what the coordinate t means. An observer who is fixed at coordinate $r = R$ uses a clock that ticks uniformly in intervals of their proper time τ. Compared to intervals in t, that observer's clock behaves as

$$d\tau = dt\sqrt{1 - \frac{2GM}{Rc^2}} \ . \tag{20.2}$$

Notice that as $R \to \infty$, $d\tau \to dt$. In other words, the coordinate t is in fact proper time for observers who are at rest very far away from the mass M. This hopefully makes sense since, as we described in the previous lecture, the Schwarzschild metric looks just like the spacetime of special relativity (in spherical coordinates) for an observer who is very far away.

As we've emphasized a few times, the coordinate t describes the clocks of observers who are far away. How do these clocks communicate and "sync up" with the clocks of other observers? In special relativity, we did this using light — the invariant properties of light give us a tool which allows us to connect the clocks of different observers. General relativity inherits this: the different clocks of different observers are synchronized with one another by allowing their properties to be carried from observer to observer using light.

However — and this is the crucial point — *the propagation of light is strongly affected by gravity.* Clocks which tick very nicely in time t very far from the mass M do *not* tick so nicely when they are close to $r = 2GM/c^2$. In fact we, saw that light's coordinate velocity goes to zero as $r \to 2GM/c^2$. Light emitted at that radius carries no energy to observers who are infinitely far away.

This helps us to see why infall according to the t parameterization looks so weird compared to infall in the τ parameterization. In the τ parameterization, we are writing things in terms of a clock that makes sense exactly where the infalling observer is located. That parameterization in essence tells us exactly what the infalling observer is actually experiencing. The t parameterization, on the other hand, tells us how things "look" according to an observer who is watching things happen from very far away. **The distant observer makes their measurements using light** (or, as we'll discuss a little later, other forms of radiation that travel at the speed of light) **and so their measurements are affected by how gravity affects light.** In essence, the infalling observer never crosses $r = 2GM/c^2$ according to the t parameterization because the light that allows distant observers to see that happen never reaches them. It doesn't mean that they see this observer just "hovering" outside $r = 2GM/c^2$, however — they actually see *nothing at all.*

Imagine that as the infalling observer approaches this point, they carry a beacon which emits a signal — "My time is now τ, and all is well!" — directed to a distant observer. As the infalling observer approaches $r = 2GM/c^2$, the light takes longer and longer (according to very distant clocks) to get out. The message is also increasingly redshifted to longer and longer wavelengths. Keep in mind that at late times (as seen by distant observers), the infalling observer gets closer by a factor of e after every interval $\Delta t = c^3/2GM$. This time is a bit less than 10 microseconds if $M = 1$ solar mass. This means that for $M = 1$ solar mass, after an interval $\Delta t = 1$ second, the distant observer sees the position of the infalling observer change from $r = 2.01GM/c^2$ to $r = 2(1 + 0.01e^{-100{,}000})GM/c^2$. *Very* quickly we can no longer see the infalling observer, nor get any message from them.

Although strictly speaking, the distant observer claims that their infalling friend never crosses $r = 2GM/c^2$, when that friend gets close to this radius, what the distant observer claims does not matter. No communication with the infalling friend is ever possible. For all practical purposes, their friend has merged with the spacetime, and can no longer be distinguished as an independent entity. (Indeed, later measurements would show that the mass which appears in the spacetime line element is no longer M, but has become $M + m_{\rm friend}$. Your friend is part of spacetime now.)

20.3 The event horizon

A good way to summarize this discussion is that the time coordinate t is perfect for observers who are very far away. Indeed, t is their proper time, and is how they naturally describing the ticking of clocks. The coordinate t can be used for all $r > 2GM/c^2$, though it gets increasingly problematic as r gets close to $2GM/c^2$. It is bad exactly at $r = 2GM/c^2$.

What about for $r < 2GM/c^2$? To be blunt, this is tricky. Time t connects clocks from the very distant region to other places using light; since light doesn't propagate at all when $r = 2GM/c^2$, t simply ceases to be a useful measure of time that radius. This means that we need to be a lot more careful about how we set up and define "past" and "future" when we examine the region $r \leq 2GM/c^2$. This requires a bit more setup and analysis than is appropriate for 8.033, but we can borrow the punchline for our purposes: with a little effort, it can be shown that **no light ray** emitted at $r < 2GM/c^2$ can ever propagate to $r > 2GM/c^2$. In essence, the behavior that we saw at $r = 2GM/c^2$ — light rays emitted exactly at that spot remain forever bound to that spot — turns around. What we find is that all light rays inevitably evolve to smaller and smaller values of r. Even a light ray that we "think" is pointing outward ends up on a trajectory that eventually hits $r = 0$.

Since light can never "get out" from $r \leq 2GM/c^2$, this radius defines a boundary beyond which events cannot communicate with the rest of spacetime. We call this boundary the *event horizon* — events inside $r_H = 2GM/c^2$ are "over the horizon" and forever out of our reach. An object which has an event horizon is called a *black hole*.

The event horizon is one of the strangest predictions of physics. Before thinking about whether we can safely test for its existence, it's worth pausing to review a few issues that you might wonder about. Further discussion of these points can be found in a paper[1] which summarizes a presentation of these issues at a summer school for graduate students.

- The Schwarzschild spacetime describes an object that is spherically symmetric, and is non-rotating. Are conclusions about the nature of event horizons robust if they are made with such a "special" configuration?

  It turns out we do indeed need to go beyond this spacetime; the Kerr spacetime that we mentioned in Lecture 18 (which describes a rotating black hole) turns out to be just what we need. It describes an object which has rotation, and has an event horizon at radius $r_H = \tilde{M} + \sqrt{\tilde{M}^2 - a^2}$, where $\tilde{M} = GM/c^2$, $a = J/Mc$, and J is the magnitude of the object's spin angular momentum. Although the quantitative details change as we go from Schwarzschild to Kerr, the qualitative picture remains pretty much the same. Light rays at $r = r_H$ remain bound to that radial location (although they "twist" in axial coordinate, in essence being dragged along by the black hole's spin; you'll explore a related issues arising from this behavior on pset #9). Light rays emitted at $r < r_H$ can never reach $r > r_H$, but instead inevitably propagate to $r = 0$.

- Is Kerr the end of the discussion, or is there a whole "zoo" of spacetimes with event horizons that may describe black holes?

  There is one further change we can make beyond Kerr — black holes can have electric charge — but that is it. It is expected that in "real life," any black hole's charge will be vanishingly small, since they will tend to form near environments with a lot of free plasma. The combination of electromagnetic and gravitational forces means

[1] `https://arxiv.org/abs/hep-ph/0511217`

that black holes will tend to pull in charges opposite in sign to their own charge, neutralizing themselves and leaving a Kerr black hole behind. So a Kerr black hole is, for observational and experimental purposes, indeed the end of the discussion.

You might wonder: What happens if something disturbs this black hole? Won't it change its shape or its other properties in some way? Indeed, such a thing can happen (and we'll talk about some examples soon). However, a set of remarkable results prove what is now summarized as the *no-hair theorem*: A spacetime which contains an event horizon is either the Kerr solution (which does not change with time), or it is time evolving. *If it is time evolving, then it evolves into the Kerr black hole spacetime.* This time evolution forces the spacetime to "shake off" all deviations, until only the Kerr solution remains.

What this means is that if we have a black hole described by the Kerr spacetime and it is disturbed somehow (perhaps you threw your roommate into it to see if has an event horizon), it will "jiggle" a little bit, and then settle down to a new Kerr spacetime. Its mass and spin might change after the disturbance; the nature of the jiggling may tell us something about what happened to disturb it. We describe this as the black hole "no-hair" theorem because it tells us that black holes have no distinguishing structure (no "hair") beyond their mass and spin (and, in principle, their electric charge).

- The black hole spacetime exists for all time. Are the only black holes we might encounter in Nature ones that have existed forever? Or does Nature provide a way to make black holes from "normal" objects?

  It is not terribly difficult to show that, starting with normal-behaved matter, it can evolve into a spacetime with an event horizon. For a very special but unrealistic case, this can be done analytically: a spherical ball of dust, with no pressure to hold it up against gravity, collapses to form a black hole. This calculation was first done by the famous physicist Robert Oppenheimer with his student Hartland Snyder; their result appeared in publication on September 1 1939. (This date is important for other reasons; Oppenheimer's research moved into military applications very soon after this.) The Oppenheimer-Snyder collapse calculation is simple enough that I use it as a homework exercise in 8.962. For more realistic situations (including effects such as rotation and realistic pressure profiles), it requires numerical computation (though a UROP student is now exploring whether including rotation can be done in a simple way). The outcome of these studies is that we can indeed start with "normal" matter, and have it evolve into a black hole.

- What about the behavior as $r \to 0$? We argued earlier that, kind of like the Coulomb point charge, there must be an infinite amount of "stuff" crammed into zero volume there, at least classically. Can that behavior possibly be correct? Doesn't quantum physics have something to say about this?

  The nature of what happens at the very center of the black hole remains a mystery. Quantum effects must surely have a major impact on the nature of things as we approach $r \to 0$; the exact nature of those effects remains unknown since we haven't formulated an undisputed quantum theory of spacetime. This is a source of some concern. One thing that is indisputable is that everything which crosses the event horizon eventually reaches $r = 0$. Indeed, when we develop a better parameterization to describe "time" for a body that has crossed the horizon, we discover that $r = 0$ is not really the "center" of the spacetime. Instead, $r = 0$ is actually the *future* of the spacetime — at least, the future of all spacetime interior to $r = 2GM/c^2$. (Further discussion of this point can be found in the paper whose URL is listed in the footnote on a previous page.)

Whatever mysteries may occur as $r \to 0$, they are hidden from us by the black hole's event horizon, and cannot have any effect on measurements that we make out in the rest of the universe[2]. This gives us freedom to apply the laws of physics that we understand to the region of spacetime that is able to communicate with us. We content ourselves, for now, with the fact that physics we believe we understand describes everything that we can measure in principle[3].

20.4 What can we observe?

Putting these issues of principle to the side, the question becomes: what observations can we make of an object which tell us that the spacetime has the strong-field properties we expect if the object is a black hole? This is a very open-ended question (to first approximation, answering this question has constituted a substantial fraction of your lecturer's research career). In this discussion, we look at two aspects of motion in black hole spacetimes (focusing on the Schwarzschild case, which is relatively simple) that are hallmarks of the object being a black hole. Motion in these spacetimes is important because black holes are themselves totally dark. But, if things move near them, we may be able to measure the light that these objects emit — or some other kind of radiation associated with the motion.

20.4.1 The motion of material bodies

A material body moving near a Schwarzschild black hole is governed by the Lagrangian

$$L = \frac{1}{2}g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta = \frac{1}{2}\left[-\left(1 - \frac{2GM}{rc^2}\right)(c\dot{t})^2 + \frac{(\dot{r})^2}{(1 - 2GM/rc^2)} + r^2(\dot{\theta})^2 + r^2\sin^2\theta(\dot{\phi})^2\right] , \tag{20.3}$$

where $\dot{x}^\alpha = dx^\alpha/d\tau$. As you showed on problem set #8, because the spacetime is independent of t and independent of ϕ, we can find two constants of the motion right away. If we focus

[2]Should you be curious to read about this, the hypothesis that singularities are hidden behind event horizons is called the "cosmic censorship conjecture." Note that this is a conjecture, not a theorem. Indeed, counterexamples have been found, although they tend to describe very special circumstances that will not happen in Nature. From an experimental/observational standpoint, the idea that the $r = 0$ singularity is always "clothed" by an event horizon works well enough that many of us go along with it. We're certainly not 100% satisfied about the situation given that this conjecture remains just a conjecture.

[3]There is another lingering bit of concern, which is that when we apply the leading effects of quantum physics to black holes, we find that they lose mass, eventually evaporating away entirely. This is the phenomenon of *Hawking radiation*. What happens to the mysteries at $r = 0$ then? We do not have a full understanding of this, and it's quite bothersome. We can content ourselves with knowing that the timescale for evaporation is so long that for most black holes we encounter, evaporation is unlikely to be a problem. For instance, a black hole of 1 solar mass will take about 10^{67} years to evaporate, and this lifetime scales with mass cubed. We appear to have some time to figure this out.

on motion that is confined to the plane $\theta = \pi/2$, then these constants take the values

$$\frac{\partial L}{\partial \dot{t}} = -\left(1 - \frac{2GM}{rc^2}\right) c^2 \dot{t} \equiv -\hat{E} \; ; \tag{20.4}$$

$$\frac{\partial L}{\partial \dot{\phi}} = r^2 \dot{\phi} \equiv \hat{L}_z \; . \tag{20.5}$$

We interpret $\hat{E}$ as a conserved orbital energy (per unit mass), and we interpret $\hat{L}_z$ as a conserved orbital angular momentum (per unit mass); the z subscript[4] is because if we think of the normal to the orbital plane as the z axis, this is the angular momentum about that axis. (Since the orbit is confined to this plane, this is also the total angular momentum.)

We also know that $\vec{u} \cdot \vec{u} = g_{\alpha\beta} \dot{x}^\alpha \dot{x}^\beta = -c^2$, so (again using $\theta = \pi/2$)

$$-c^2 = -\left(1 - \frac{2GM}{rc^2}\right)(c\dot{t})^2 + \frac{(\dot{r})^2}{(1 - 2GM/rc^2)} + r^2 (\dot{\phi})^2 \; . \tag{20.6}$$

Using Eqs. (20.4) and (20.5), we can replace $\dot{t}$ and $\dot{\phi}$ in this expression, and rewrite it as an equation governing the radial motion of the body orbiting in this spacetime:

$$\left(\frac{dr}{d\tau}\right)^2 = \frac{\hat{E}^2}{c^2} - V_{\rm eff}(r) \; , \tag{20.7}$$

where

$$V_{\rm eff}(r) = \left(1 - \frac{2GM}{rc^2}\right)\left(c^2 + \frac{\hat{L}_z^2}{r^2}\right) \; . \tag{20.8}$$

This function is often called the "effective potential" for orbits in the Schwarzschild spacetime, since it plays the same role in determining the motion of bodies as a similar potential that is often used to describe Newtonian orbits, and from which we derive Kepler's laws.

Equations (20.7) and (20.8) are going to be our main tools for a little while, so it is worthwhile focusing on what they tell us. Figure 2 shows an example of what this function looks like, plotted for a particular choice of $\hat{L}_z$ (in the case, $\hat{L}_z = 3\sqrt{3/2}GM/c$).

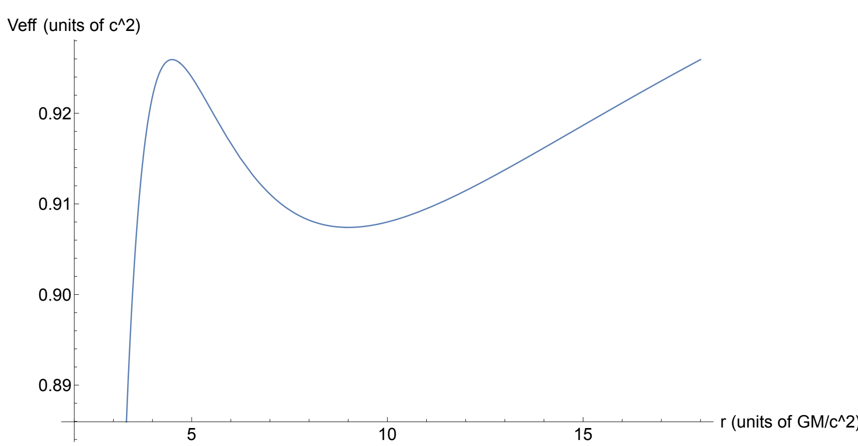


Figure 2: The function $V_{\rm eff}(r)$ plotted for $\hat{L}_z = 3\sqrt{3/2}GM/c$.

[4]It is also useful to "decorate" the symbol for angular momentum a bit so that it isn't too easy to confuse it with the Lagrangian.

Turn now to Eq. (20.7). This equation tells us that the radial component of the orbiting body's 4-velocity is determined by subtracting $V_{\rm eff}$ from a quantity made from the orbit's conserved energy, $\hat{E}^2/c^2$. There is a tremendous amount of information in this equation. We are going to use it to figure out how the body's motion depends on its energy and its angular momentum.

Consider for example the situation illustrated in Fig. 3.

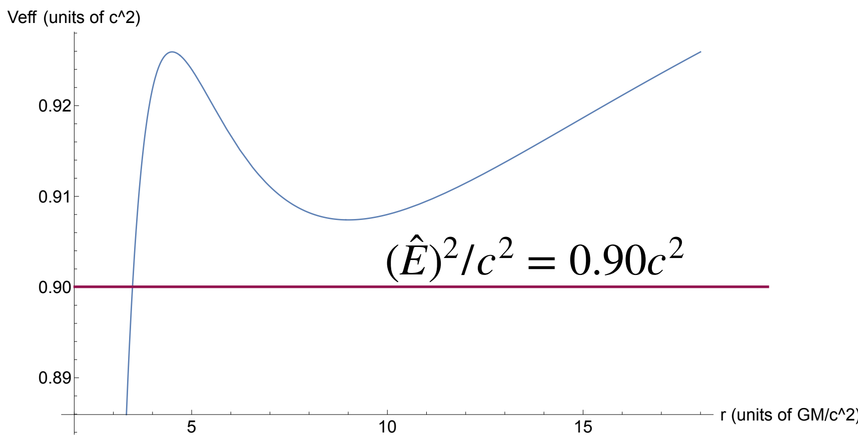


Figure 3: The same $V_{\rm eff}(r)$, showing this potential versus $\hat{E}^2/c^2 = 0.90c^2$.

For almost the entire span of radius we have included here, $\hat{E}^2/c^2$ is less than $V_{\rm eff}(r)$. This means that over this range $(dr/d\tau)^2$ is negative, and there is no real solution describing a body moving over these radii. At least for this value of angular momentum, $\hat{E} = \sqrt{0.9}c^2$ does not yield any allowed orbital motion.

Consider next the situation shown in Fig. 4. This is the same potential, but the energy is now $\hat{E}^2/c^2 = 0.924c^2$. Notice that this time $\hat{E}^2/c^2 \geq V_{\rm eff}$ over a range of radii.

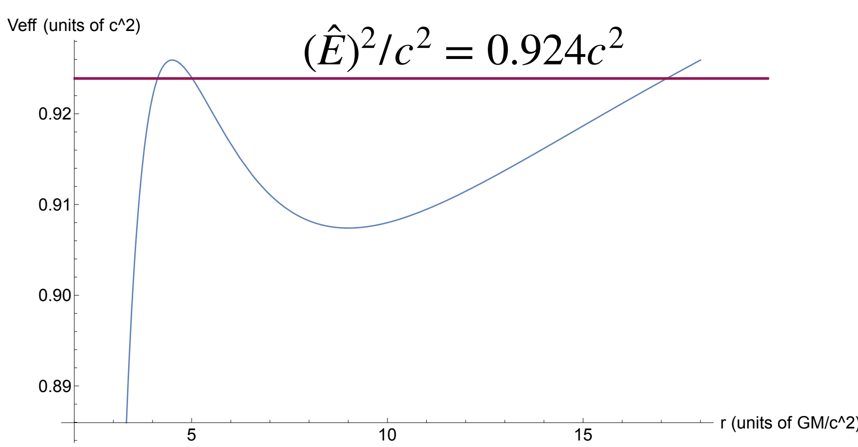


Figure 4: The same $V_{\rm eff}(r)$, showing this potential versus $\hat{E}^2/c^2 = 0.924c^2$.

For the situation shown in Fig. 4, $(dr/d\tau) \geq 0$ over a range of radii. This situation describes an *eccentric orbit*: an orbit that oscillates from $r_{\rm min} = 5GM/c^2$ to $r_{\rm max} \simeq 17.2GM/c^2$ and back over the course of its orbit. The turning points are defined by the condition $(dr/d\tau) = 0$, which means that they are found by finding the values of r at which $\hat{E}^2/c^2 = V_{\rm eff}(r)$. Bear in mind that while it is sloshing back and forth in radius, its angle ϕ is continually increasing: from the relationship between the conserved angular momentum $\hat{L}_z$ and $d\phi/d\tau$, we find

$$\frac{d\phi}{d\tau} = \frac{\hat{L}_z}{r^2} \ . \tag{20.9}$$

In the Newtonian limit, this orbit would look like a closed ellipse if we looked at its angular and radial motion together. In general relativity, the orbit does not quite close, and we get a more interesting pattern. Defining $x = r(\tau)\cos[\phi(\tau)]$, $y = r(\tau)\sin[\phi(\tau)]$, Fig. 5 shows one complete radial period for the parameters used in Fig. 4:

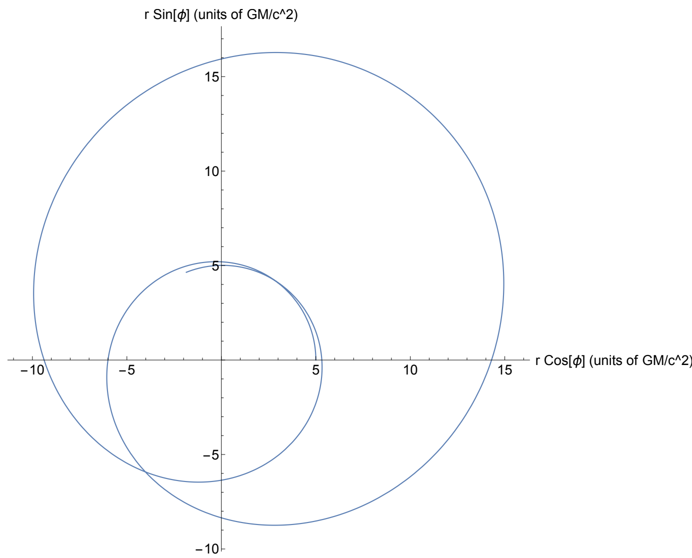


Figure 5: Motion in r and ϕ for an orbit with $\hat{L}_z = 3\sqrt{3/2}GM/c$, $\hat{E} = 0.924c^2$.

This is a "strong-field" orbit; it looks *very* different from a Newtonian orbit. In particular, the orbit moves through a *lot* more than 2π radians of ϕ in the time in takes to move from $r_{\rm min}$ to $r_{\rm max}$ and back. (This example actually completes 2.31 complete "whirls" in ϕ during a single cycle of radial motion.) When the orbital radius is at all times large compared to GM/c^2, the "extra" ϕ per orbit is much smaller. In fact, it is not too hard to show that this motion precisely reproduces the anomalous precession of Mercury's orbit that so excited Albert Einstein in 1915.

A particularly special value of the energy is illustrated in Fig. 6. This value is chosen so that $\hat{E}^2/c^2 = V_{\rm eff}(r)$ at exactly one point. This case defines a *circular orbit*. To find this orbit, we require that $\hat{E} = c\sqrt{V_{\rm eff}(r)}$ (so that $dr/d\tau = 0$). We also require that the orbit "live" at the minimum of the effective potential: $\partial V_{\rm eff}/\partial r = 0$. As you show on problem set #9, this yields a set of analytic solutions that describe the energy and angular momentum per unit mass for a body in a circular orbit:

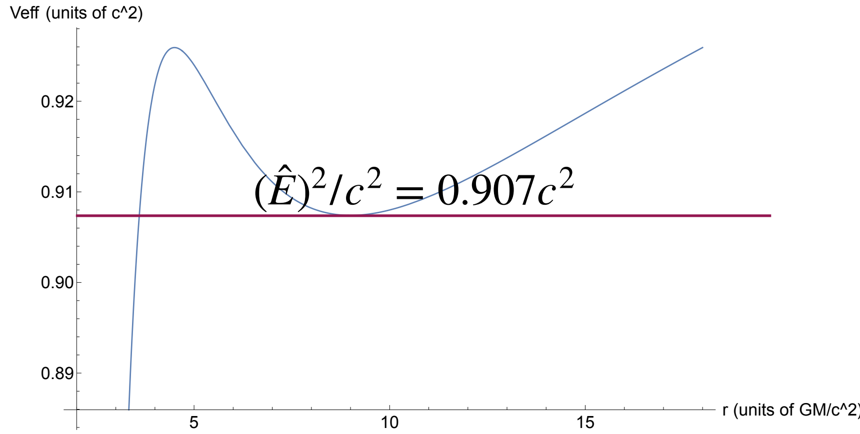


Figure 6: The same $V_{\rm eff}(r)$, showing this potential versus $\hat{E}^2/c^2 = 0.907c^2$.

$$\hat{L}_z = \pm\sqrt{\frac{GMr}{1-3GM/rc^2}}\ , \qquad \hat{E} = \frac{1-2GM/rc^2}{\sqrt{1-3GM/rc^2}}\ . \tag{20.10}$$

(The $\pm$ on the angular momentum is because we take a square root at one point, and both solutions are valid. The signs describe orbits going in opposite ϕ directions at radius r.)

One last point before moving on: if we examine a sequence of potentials, we find something interesting and a little odd. For almost all values of $\hat{L}_z$, the potential qualitatively has the shape we saw in the previous figures — a peak at small radius, with a minimum at some finite r, the asymptoting to $V_{\rm eff} \to c^2$ as $r \to \infty$. However, if we make $\hat{L}_z$ small enough, we see that there is a change. Figure 7 shows what happens as we reduce $\hat{L}_z$ from $3.5GM/c$ to $3.45GM/c$:

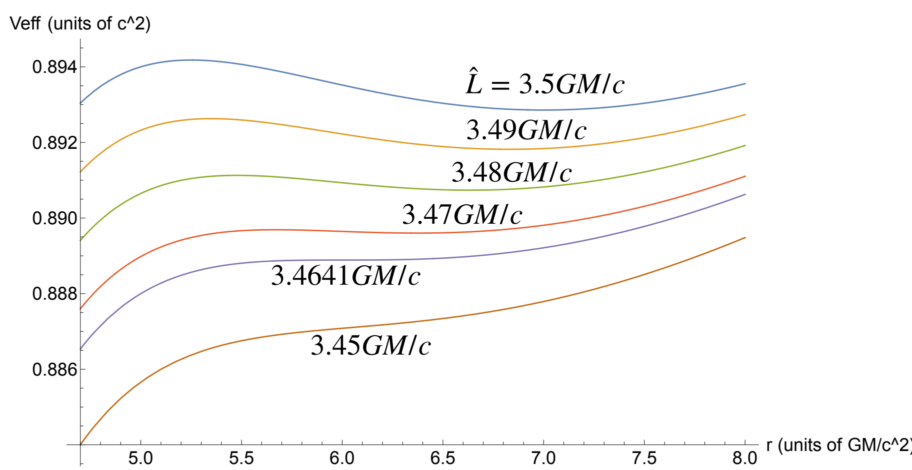


Figure 7: A zoom onto the region of the minimum for a sequence of $V_{\rm eff}(r)$, looking at different values of $\hat{L}_z$.

As $\hat{L}_z$ gets smaller, the minimum moves to smaller values of r. At the same time, the vicinity of the minimum gets progressively flatter, and is less well defined as a minimum. In fact, at a certain point the function flattens out so much that *the minimum goes away altogether.* When this happens, **there are no more stable circular orbits**.

To find when the minimum goes away, we look for the point along the sequence at which the first and second derivatives both vanishes: $\partial V_{\rm eff}/\partial r = 0$ and $\partial^2 V_{\rm eff}/\partial r^2 = 0$ at the same value of r. A quick analysis shows us that this condition is met when

$$\hat{L}_z = \sqrt{12}GM/c \simeq 3.4641GM/c\ . \tag{20.11}$$

(The purple curve in Fig. 7 was computed with exactly this special value of $\hat{L}_z$.) Using the formula that relates a circular orbit's energy to its radius tells us that this happens when $r = 6GM/c^2$. Here is a prediction about strong-field orbits that is *starkly* different from what we encountered with Newtonian gravity: **No stable circular orbits exist at all for radii $r \le 6GM/c^2$.**

20.4.2 The motion of light

On problem set #9, you looked at how light is bent by gravity in the weak gravity limit. What is the strong-field analog of this?

To answer this question, let's think about the geometry of a light ray that comes in from very far away. Let's imagine that the black hole is at the origin of coordinates, and the light ray comes in parallel to the x axis. Far away, it is displaced from the axis by a distance b. We will call this distance the *impact parameter* of the incoming light ray. From basic mechanics, we can say that the light ray has an angular momentum that is related to the x component:

$$|\mathbf{L}| = |\mathbf{r}\times\mathbf{p}| = bp^x = bE/c \equiv L_z\ . \tag{20.12}$$

All of the quantities in this equation are evaluated very far away, where the spacetime is the same as that of special relativity. The geometry of this situation is illustrated in Fig. 8.

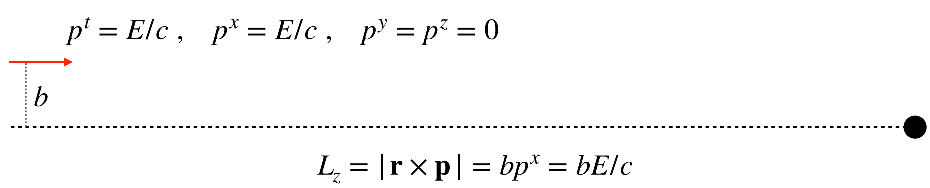


Figure 8: The geometry of a photon that is "launched" from far away toward the black hole.

To figure out how the photon evolves as it propagates in to the strong gravity region, consider the Lagrangian for light:

$$L_{\rm light} = \frac{1}{2}g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta = \frac{1}{2}\left[-\left(1-\frac{2GM}{rc^2}\right)\left(c\frac{dt}{d\lambda}\right)^2 + \left(1-\frac{2GM}{rc^2}\right)^{-1}\left(\frac{dr}{d\lambda}\right)^2 + r^2\left(\frac{d\phi}{d\lambda}\right)^2\right]\ . \tag{20.13}$$

(We've simplified this a bit by putting $\theta = \pi/2$, $d\theta/d\lambda = 0$, and we are using $\dot{x}^\alpha = dx^\alpha/d\lambda$.) Because $\partial L_{\rm light}/\partial t = 0$ and $\partial L_{\rm light}/\partial \phi = 0$, we identify both $\partial L_{\rm light}/\partial \dot{t}$ and $\partial L_{\rm light}/\partial \dot{\phi}$ as constants of the motion:

$$\frac{\partial L_{\rm light}}{\partial \dot{t}} = -\left(1 - \frac{2GM}{rc^2}\right) c^2 \frac{dt}{d\lambda} = c\, g_{tt} p^t = c\, p_t \equiv -E \; ; \tag{20.14}$$

$$\frac{\partial L_{\rm light}}{\partial \dot{\phi}} = r^2 \frac{d\phi}{d\lambda} \equiv L_z = bE/c \; . \tag{20.15}$$

In setting these equalities, we've used the fact that it is very easy to compute p_t and L_z very far from the black hole. But, because they are constants along the light ray, once we've computed them, we can use these values through the entire calculation.

Using $\vec{p} \cdot \vec{p} = 0$ in combination with these relations between E, $dt/d\lambda$, L_z, and $d\phi/d\lambda$ we find an equation for $dr/d\lambda$ that is similar in form to the equation we found for the motion of a material body:

$$\left(\frac{dr}{d\lambda}\right)^2 = \frac{E^2}{c^2} - \frac{L_z^2}{r^2}\left(1 - \frac{2GM}{rc^2}\right) \; . \tag{20.16}$$

This is a useful form because it parameterizes the motion in terms of the constants of motion L_z and E. However, we can do better, since we know that $L_z = bE/c$. Let's use this to eliminate E from the equation:

$$\left(\frac{dr}{d\lambda}\right)^2 = \frac{L_z^2}{b^2} - \frac{L_z^2}{r^2}\left(1 - \frac{2GM}{rc^2}\right) \; . \tag{20.17}$$

And, since L_z is itself a constant of the motion, we can eliminate it from the right-hand side:

$$\begin{aligned} \frac{1}{L_z^2}\left(\frac{dr}{d\lambda}\right)^2 &= \frac{1}{b^2} - \frac{1}{r^2}\left(1 - \frac{2GM}{rc^2}\right) \\ &= \frac{1}{b^2} - V_{\rm light}(r) \; . \end{aligned} \tag{20.18}$$

This equation now tells us that light propagates whenever $1/b^2 > V_{\rm light}$, where $V_{\rm light}$ plays a role in our analysis just like the effective potential that governs the motion of material bodies. However, $V_{\rm light}$ is much simpler — it doesn't depend on *any* free parameters.

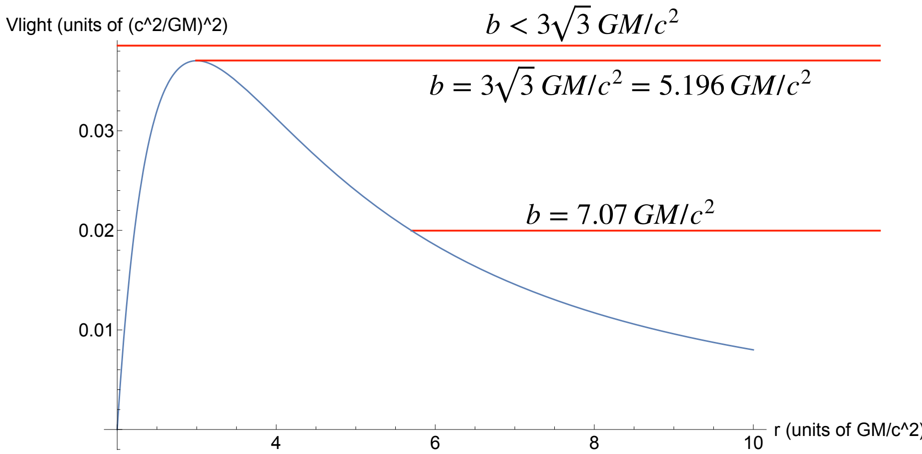


Figure 9: The potential $V_{\rm light}(r)$ that governs the motion of light in the Schwarzschild spacetime, plus a few lines illustrating $1/b^2$ for several values of b.

Figure 9 plots this potential as a function of r. Its maximum occurs at $r = 3GM/c^2$; its value at the maximum point is given by

$$V_{\rm light}(3GM/c^2) = \left(\frac{c^2}{GM}\right)^2 \left[\frac{1}{9}\left(1 - \frac{2}{3}\right)\right] = \frac{c^4}{27G^2M^2} = \left(\frac{c^2}{3\sqrt{3}GM}\right)^2 . \qquad (20.19)$$

Figure 9 also includes several lines illustrating $1/b^2$ for several interesting values of b. Light propagates in radius as long as $(dr/d\tau)^2 > 0$, which occurs when $1/b^2 > V_{\rm light}$. With that in mind, let's examine how light behaves in this spacetime for several values of b:

- If $b > 3\sqrt{3}GM/c^2$, light propagates in from infinity just fine until it reaches the radius at which $1/b^2 = V_{\rm light}$. At this point, the light reverses radial direction, and heads back out to larger radius. When we look at the motion in the equatorial plane, we see that it comes in, has its trajectory bent as it comes closest to the black hole, then heads back out to large radius. An example of this is shown in the top-left of Fig. 10.

- If $b < 3\sqrt{3}GM/c^2$, light propagates from infinity all the way to $r = 0$. In this case, light passes into the event horizon and disappears from us forever. This is illustrated in the top-right of Fig. 10.

- If $b = 3\sqrt{3}GM/c^2$, light comes in until it reaches $r = 3GM/c^2$. At this point, $dr/d\tau = 0$, so it sits in this *light orbit* forever. This is an unstable orbit, so the slightest deviation from $b = 3\sqrt{3}GM/c^2$ means that the light will either eventually zoom away or else fall in. The bottom of Fig. 10 shows what happens when b is too large by $4 \times 10^{-8}GM/c^2$. In this case, the light completed about 3.7 orbits before zooming away.

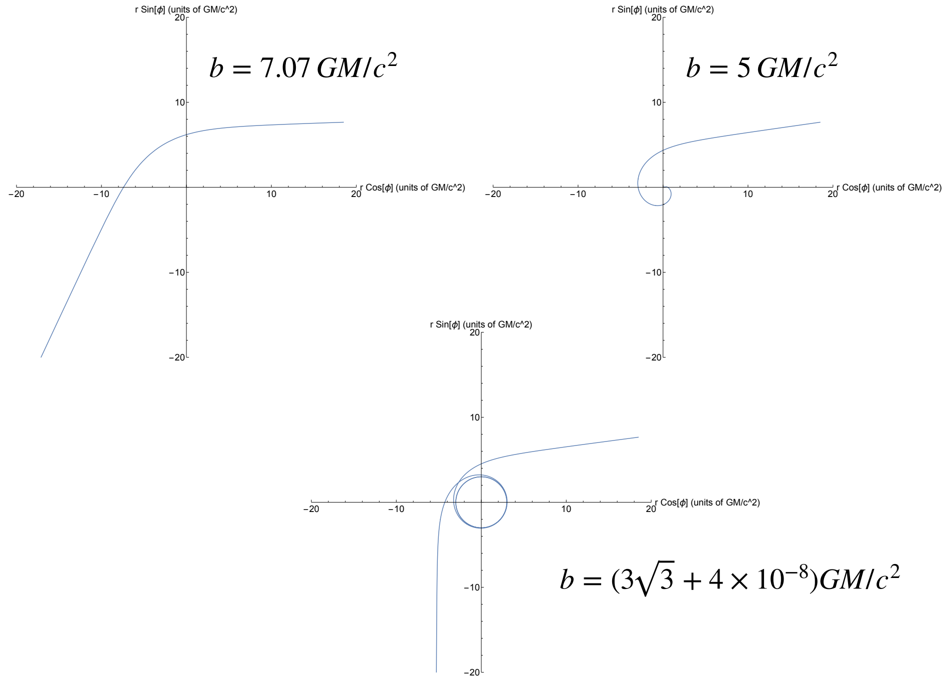


Figure 10: Examples of the (r, ϕ) motion for light propagating in the Schwarzschild spacetime.

This light orbit, or *light ring* is a distinctly non-Newtonian feature. But, it turns out to be something that makes a very strong observational prediction. Imagine that a black hole in Nature is immersed in some kind of hot gas or plasma. That material will emit electromagnetic radiation across a wide range of wavelengths. If some of that radiation passes close to the black hole hole, we expect that its trajectory will be bent very strongly around the black hole. Some of it will even get "trapped" into the light orbit at radius $r = 3GM/c^2$, at least for several orbits. When that radiation escapes from the light orbit, it can propagate toward us. The expectation is that we could see a ring of radius $3\sqrt{3}GM/c^2$ on the sky. This has long been regarded as a "smoking gun" of the strong-field nature of a black hole spacetime.

It turns out other kinds of radiation get trapped at that radius too, and leave an imprint that likewise is a smoking gun of the nature of the spacetime. We will discuss that in the next lecture.

20.5 Summary: Features of strong-gravity motion

This set of lecture notes is extremely dense. I hope that everyone who takes 8.033 is now capable of going through and understanding everything that is presented here; however, there is so much material (and this comes at such a stressful point in the semester) that I have little illusion that everyone will actually do that. So let me summarize what I regard as the most important points to note:

- Infall parameterized by τ looks completely different from infall parameterized by t. This is because t describes clocks that are very far away; they connect to clocks at small r using the trajectories of light. As one approaches $r = 2GM/c^2$, the propagation of light is hugely affected by gravity. The observer who is far away never sees an infalling observer cross $r = 2GM/c^2$ because the light they would use to make this measurement never reaches them.

- The radius $r = 2GM/c^2$ marks an *event horizon* in the spacetime. Events at $r < 2GM/c^2$ cannot communicate with events "outside the horizon." Once any trajectory crosses this radius, it is doomed to eventually reach $r = 0$, where it will find a quick (as measured by the trajectory's own clock) and violent end.

- The Lagrangian for the orbits of material bodies in this spacetime can be studied in the usual way with the Euler-Lagrange equations. Given an orbit's angular momentum per unit mass $\hat{L}_z$ and its energy per unit mass $\hat{E}$, one can figure out the kind of motion one is likely to get. We examined one example of an "eccentric" orbit (which generalizes the "elliptical" orbits of Newtonian gravity), and characterized circular orbits (which you will examine on pset #9).

- Stable circular orbits do not exist for $r < 6GM/c^2$, starkly non-Newtonian behavior.

- Light can be bent so strongly by gravity that it forms an orbit at $r = 3GM/c^2$. Observers far away may be able to see a ring of light with radius $b = 3\sqrt{3}GM/c^2$. This value of b is set by the impact parameter that puts light exactly on the light ring.

Massachusetts Institute of Technology
Department of Physics
8.033 Fall 2024

Lecture 21
Data on strong gravity

21.1 Overview; Kerr versus Schwarzschild

Having discussed in some detail the features which make strong-gravity spacetimes special, we turn now to what data and observations have taught us.

The most interesting feature to observe would of course be an event horizon. However, the horizon is hard to observe since, by definition, the signpost of its existence is a kind of absence — a "one-way membrane" from which we can get no information. One can imagine looking for spacetimes that describe a very dense, massive body, but that appear to lack well-defined surfaces. That indeed has been done, and is responsible for providing many of the black hole *candidates* studied by astronomers over the years. In this set of lecture notes, we will focus our discussion of measurements that have provided evidence for other features associated with motion in very strong gravity that we have discussed:

- *Non-Keplerian orbits*: As noted in Lecture 20, strong-field orbits are not closed ellipses in general, but show rather more complicated patterns of motion. This fundamentally arises from the fact that, in the time takes an orbiting body to move from minimum radius to maximum radius and back, the body moves through more than 2π radians. At its core, this is the same effect that leads to the 43 arcseconds per century of Mercury's anomalous perihelion precession; in very strong-field spacetimes, the effect is quite a bit stronger.

- *Unstable orbits*: No stable circular orbits exist for $r \leq 6GM/c^2$.

- *The light ring*: The gravitational bending of light becomes so severe that a light ray can in principle loop around forever. In practice, because this is an unstable orbit, we expect it to loop around a few times at most before zooming out. (Presumably some light rays loop around and then fall in — but we never measure those light rays.)

The analyses we have done so far which allowed us to develop and describe all these effects were based on studies of the Schwarzschild spacetime. Schwarzschild is now understood to be a special case of the Kerr spacetime, which describes black holes which rotate; indeed, the generic solution that we have long expected[1] Nature to provide is the Kerr solution. Cataloguing in detail what happens when we go from Schwarzschild to Kerr is beyond the scope of 8.033, but it is not beyond us to understand how things change when we do this:

- *Frame dragging*: As you showed on problem set #9, near a rotating black hole, spacetime "wants" to pull you along, so that you move in the same sense in which the black

[1] As discussed briefly in class, there are solutions which describe black holes with charge as well, but our expectation is that such solutions will be neutralized by infalling charges which cancel out the hole's "intrinsic" charge in any realistic astrophysical environment.

hole is rotating. When you very deep in the strong field (at a coordinate fairly close to the event horizon), it becomes impossible to resist this motion, and you are forced to move in the same rotational sense as the black hole, no matter how strongly you oppose it. This *frame-dragging* effect — the dragging of all observer frames into co-rotation with the black hole — amplifies the non-Keplerian features that we saw in the case of a Schwarzschild metric.

- *The properties of unstable orbits depend on orbit orientation*: There are unstable orbits in the Kerr spacetime, analogous to the orbit at $r = 6GM/c^2$ that we found for Schwarzschild. However, when the black hole is spinning, the radius of these orbits varies depending upon the orientation of the orbit with respect to the spin axis. Figure 1 (the curves labeled "material body orbit") shows what we find for orbits that are in the black hole's equatorial plane (i.e., orbits that have $\theta = \pi/2$). Orbits which are *prograde* move in the same sense as the black hole's rotation; those which are *retrograde* move in the opposite sense of the rotation. Notice that as the black hole's spin increases from $a = 0$ (which is the same thing as Schwarzschild) to $a = GM/c^2$, the radii of these two possibilities diverges quite a bit.

- *Properties of light rings also depend on orbit orientation*: Just as the unstable orbit's position varies with orientation, so does the radius of the light ring. Figure 1 also shows the radii of the light ring associated with orbits that have $\theta = \pi/2$, and it also splits into a prograde and a retrograde branch.

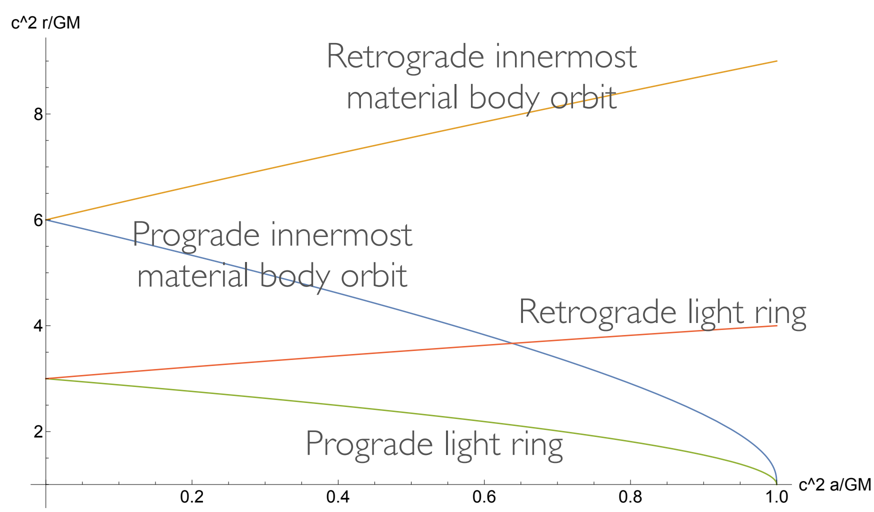


Figure 1: Important orbital radii in the Kerr spacetime as a function of black hole spin parameter a. The two "material body orbit" curves show the radii at which circular orbits become unstable. The "prograde" curves traces out the radius of orbits which move in the same sense as the black hole's spin; the "retrograde" curves traces out this radius for orbits which move opposite to the black hole's spin. The two light ring curves do the same things for the radius at which light rays are captured onto orbits.

Conceptually, it's not very difficult to generalize everything we did for Schwarzschild to Kerr. Getting all the details right, however, is a rather involved process, significantly more difficult in detail than Schwarzschild analyses. We won't go through these details here, but strongly emphasize that these additional complications should be borne in mind as we examine the data we have on strong-gravity systems.

21.2 Gravitational radiation

Much of the best data on strong-gravity systems we have accumulated in recent years has come from a form of observation that has only come into fruition within the past decade. To understand why this, it is important to understanding that light can very often be hard to observe from very strong-gravity systems. The most interesting part of the system is dark; any light comes from objects or matter moving near or orbiting around the darkest bit. Such sources of light are often "buried" in dense astronomical environments with lots of other bodies and matter around, which makes it hard for their light to get out. Very bright, intense, high-energy light may be generated deep in these spacetimes, but the light can be highly scattered or absorbed by other matter, making it difficult for us to observe it and to use it to study these spacetimes. Even when the light gets out, its properties can be modified by scattering and absorption, making it difficult for us to use the light to learn about the nature of the spacetime in which it was generated.

"Light" is being used here as shorthand for all bands of electromagnetic radiation — oscillating disturbances to electric and magnetic fields which propagate across spacetime, from gamma rays down to radio. In the past several years, decades of effort have come to fruition to use another form of radiation: *gravitational* radiation, or gravitational waves (which we'll abbreviate GWs). GWs are another consequence of the theory of relativity. Their existence follows from the fact that any relativistic theory involving fields which act at a distance predicts that the field itself **must** radiate. This radiation reflects how changes to the field propagate across spacetime when the field's sources themselves vary with time.

If spacetime is nearly that of special relativity, we can write $g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta}$. Run this through the Einstein field equations $G^{\mu\nu} = (8\pi G/c^4)T^{\mu\nu}$, discard all terms that are of order h^2, and the result is[2]

$$\Box h_{\alpha\beta} = -\frac{16\pi G}{c^4} T_{\alpha\beta} \; . \tag{21.1}$$

The "weak gravity" metric that we discussed a few lectures ago is a solution of this equation when the time variations are zero (so that $\partial(\text{anything})/\partial t = 0$). When the source *is* time varying, the solutions to this equation are time-varying metric components $h_{\alpha\beta}$ that propagate across spacetime. An example of an allowed solution is one which takes the form

$$h_{\alpha\beta} \doteq \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h(z-ct) & 0 & 0 \\ 0 & 0 & -h(z-ct) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{21.2}$$

This solution represents a disturbance in spacetime, h, which propagates in the z direction at the speed of light. The nature of the function $h(z-ct)$ which appears in these tensor

[2]In this analysis, I am skipping over a very important technical detail involving what is called the "choice of gauge." Just as one can adjust the scalar potential ϕ and the vector potential $\mathbf{A}$ of electrodynamics in such a way as to leave the fields $\mathbf{E}$ and $\mathbf{B}$ unchanged, so one can adjust the metric-like quantity $h_{\alpha\beta}$ but leave its associated *curvature* tensors unchanged. For the purposes of 8.033, this subtlety is a tangent that we can skip over.

components is related to the dynamics of the gravitating source; we'll talk about what it looks like in a moment. The key thing to emphasize for us right now is that the influence of this function on spacetime can be measured by looking at light propagating in the x and y directions for a gravitational wave propagating along[3] z.

Consider light which propagates in the x direction. How much time does it take to travel a distance L? We calculate this using the metric. Use the fact that light has a 4-momentum which obeys $\vec{p}\cdot\vec{p} = 0$, and that it propagates in the x direction:

$$0 = \vec{p}\cdot\vec{p} = (\eta_{tt} + h_{tt})\left(c\frac{dt}{d\lambda}\right)^2 + (\eta_{xx} + h_{xx})\left(\frac{dx}{d\lambda}\right)^2 . \tag{21.3}$$

Rearranging this, using $h_{tt} = 0$ and $h_{xx} = h(z-ct)$, we solve for dt/dx as the light propagates:

$$\begin{aligned}\frac{dt}{dx} &= \frac{1}{c}\sqrt{1 + h(z - ct)} \\ &\simeq \frac{1}{c}\left[1 + \frac{1}{2}h(z - ct)\right]\end{aligned} \tag{21.4}$$

Here, we've assumed that the function $h \ll 1$; as we'll see shortly, this is a reasonable assumption. We now integrate up to compute the time it takes for light to propagate this distance in x:

$$\begin{aligned}\Delta t_x &= \frac{1}{c}\int_0^L \left[1 + \frac{1}{2}h(z - ct)\right] dx \\ &\simeq \frac{L}{c}\left[1 + \frac{1}{2}h(z - ct)\right] .\end{aligned} \tag{21.5}$$

On the last line, we imagine that in the time it takes light to travel a distance L, the function $h(z-ct)$ changes by very little (so that this function remains approximately constant[4] during the time interval corresponding to the integral). If this is not correct, then some details of the analysis change, but the final result is quite similar.

Imagine that while light travels in the x direction, light also travels a distance L in the y direction. Repeating this calculation along the y axis, we find

$$0 = \vec{p}\cdot\vec{p} = (\eta_{tt} + h_{tt})\left(c\frac{dt}{d\lambda}\right)^2 + (\eta_{yy} + h_{yy})\left(\frac{dy}{d\lambda}\right)^2 \tag{21.6}$$

Using $h_{tt} = 0$ and $h_{yy} = -h(z - ct)$, this becomes

$$\frac{dt}{dy} = \frac{1}{c}\sqrt{1 - h(z - ct)} \tag{21.7}$$

$$\simeq \frac{1}{c}\left[1 - \frac{1}{2}h(z - ct)\right] . \tag{21.8}$$

[3]More generally, the influence of the wave is along the axes normal to the wave's direction of a propagation. So if the GW propagates along x, you want to measure with light that propagates along y and z; etc.

[4]More precisely, we imagine that the function changes slowly compared to the time for light to travel the distance L.

Integrating up, this yields

$$\Delta t_y = \frac{1}{c}\int_0^L \left[1 - \frac{1}{2}h(z-ct)\right]\,dy \tag{21.9}$$
$$\simeq \frac{L}{c}\left[1 - \frac{1}{2}h(z-ct)\right]\ . \tag{21.10}$$

So in the x direction, the light travel time is a bit longer than it would be without the gravitational wave; in the y direction, it is a bit shorter. For the most interesting sources, the function h is sinusoidal, so it oscillates — but it always does so in such a way that the light travel time is long in one direction, short in the other.

Such behavior is perfectly set up to be measured using an interferometer, much like the one that Michelson used in the famous Michelson-Morley experiment that we discussed very early in 8.033. Imagine our interferometer set up like the one shown in Fig. 2: the two arms are oriented along the x and y directions, and the gravitational wave propagates in the z direction, which is normal to the page. The interferometer is set up so that, in the absence of a gravitational wave, light destructively interferes after bouncing off the mirrors and recombining at the beam splitter. When this happens, the readout photo diode measures nothing: the signal is "dark" when $h = 0$. But when a gravitational wave comes along, the light takes different times to travel in each arm. The phase associated with the light in the two arms won't balance just right to destructively interfere when it recombines, and instead we will now have a non-zero signal in the readout photo diode.

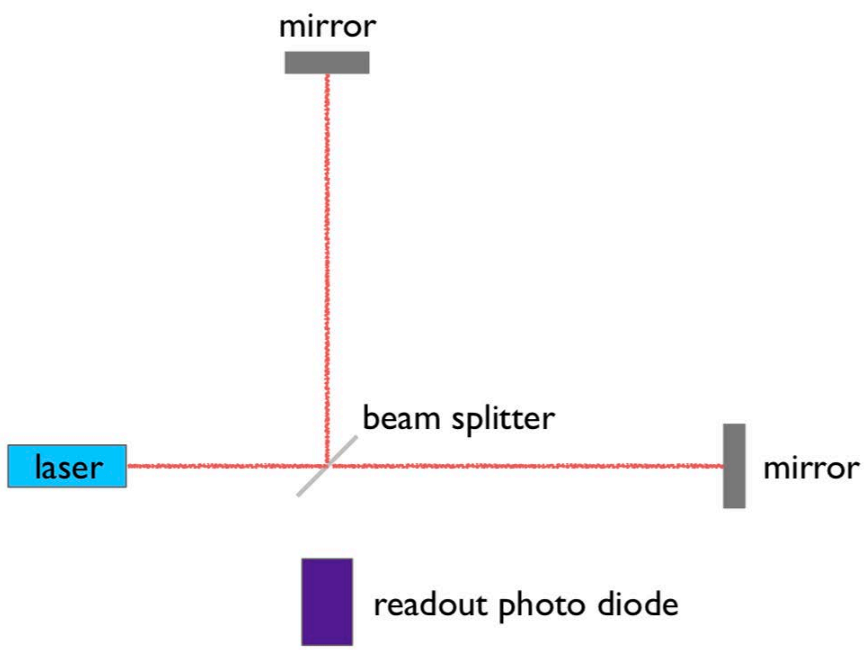


Figure 2: An interferometer set up to measure a GW like the one described in the text: One arm points along the x direction, one points along the y direction, and z — the direction of propagation of the wave — is "up," normal to the page.

This all depends upon the function h, which itself depends on the nature of the source. Solving the "linearized" Einstein field equation (21.1), we find that the leading-order solution describing radiation looks as follows[5] for a source that is a distance r away from us:

$$h_{00} = 0\ , \quad h_{0j} = 0\ , h_{jk} = \frac{2}{r}\frac{G}{c^4}\frac{d^2\mathcal{I}_{jk}}{dt^2}\ . \tag{21.11}$$

This is called the *quadrupole formula* for gravitational waves, because the tensor $\mathcal{I}_{jk}$ is related to the quadrupole moment of mass and energy distributed in the source:

$$\mathcal{I}_{jk} = \int \rho_M(\mathbf{x})\left(x_j x_k - \frac{1}{3}\delta_{jk}|\mathbf{x}|^2\right) d^3x\ . \tag{21.12}$$

(Here $\rho_M \equiv \rho/c^2$ is the mass density distribution.) Those of you who have studied some advanced electrodynamics may be reminded of the *dipole* formula for electromagnetic radiation, which shows us how the radiative potential that arises from a dynamical charge source varies as a single time derivative of a source's charge dipole moment.

To get an idea of the size of the effect that we expect for gravitational waves, let's make a rough estimate for how big a typical component of the wave tensor will be. The typical magnitude of a non-zero component of $\mathcal{I}_{jk}$ is $\sim MR^2$, where M is the amount of mass that is dynamical in the system, and R is the amplitude of its motion. Take two time derivatives, and assume that the mass is bound into some kind of orbital motion. You find that the typical magnitude of $d^2\mathcal{I}_{jk}/dt^2$ is $\sim Mv^2$, where v is the speed associated with that bound orbital motion. Combine this with Eq. (21.11) and we get the typical magnitude we might expect for a GW:

$$h_{jk} \sim \frac{GM}{c^2 r}\frac{v^2}{c^2}\ . \tag{21.13}$$

Let's imagine a source that involves 50 solar masses in orbital motion with speeds typically near 10% of the speed of light; imagine that this source is located about a billion light years away. Using these numbers, we find

$$h_{jk} \sim 10^{-22}\left(\frac{M}{50\,M_\odot}\right)\left(\frac{10^9\,\mathrm{lyear}}{r}\right)\left(\frac{v}{0.1\,c}\right)^2\ . \tag{21.14}$$

(The symbol $M_\odot$ stands for 1 solar mass.) This sets the stage for the magnitude of the timing effect we need to be able to detect — roughly a part in 10^{22} or so.

In other words, the effect of any realistic gravitational wave is *TINY*. Finding a part in 10^{22} change, in the presence of the kind of noise that affects any realistic experiment, is a topic which can consume an entire course (indeed, multiple entire courses). For our purposes, suffice it to say that such measurements can be done; that they in fact have been done; but that performing such measurements is *not* easy. Making the measurements possible is the kind of thing for which foundations associated with Swedish royalty award prizes. Let us move on to discussing what we learn when we can measure these waves.

21.3 Observing objects in orbit about black holes

Let us turn now back to what we can (and do!) observe. The most important data comes from observing objects that orbit very massive things. Some of the most compelling examples

[5]Following the previous footnote about skipping over some details having to do with gauge, those details have an influence on details here too. The formula presented here is missing some overall factors that reflect how the waves "look" from different viewing angles, but is otherwise accurate.

have been observations of stars which orbit some kind of a very massive but dark object. Over the course of about 30 years, astronomical techniques have made it possible to resolve stars moving in the very innermost regions of the galactic center. What these objects have showed us is that roughly half a dozen stars move on orbits very close to a big "something," with orbital properties that noticeably change over the course of several years.

Several of these stars complete their orbits in ten or so years, making it possible to use them to precisely measure the mass of the object that they orbit. The mass we find turns out to be

$$M \approx 4 \times 10^6\, M_\odot\ . \tag{21.15}$$

So these stars are orbiting around 4 million solar masses of *something*. However, there is *no* object visible that these stars orbit around — whatever that 4 million solar mass "thing" might be, it is dark and it is massive. At least one of those stars is now seen to undergo orbit precession in a way that aligns perfectly with the "non-Keplerian" aspect of black hole orbits that we discussed in Lecture 20; rather than advancing by 43 arcseconds per century like Mercury, its orbital ellipse advances by about 10° per orbit.

This object at the center of our galaxy has long been perhaps the most striking example of a spacetime that describes something that is really massive but dark that we have studied with telescopes, though there are quite a few others. In the past several years, some of the most compelling data probing such spacetimes has come from gravitational-wave observations. Suppose two objects are in circular orbit around one another. The gravitational waves that they generate carry away energy and angular momentum from the system. This causes the objects to fall closer toward one another. When this happens, they move faster, generating stronger gravitational waves, causing them to fall toward one another even faster. The result is a characteristic *chirping* waveform. This "chirping" continues until the two bodies come so close to one another that there no longer exists a stable circular orbit. When this happens, the two objects plunge together. Figure 3 shows an example of what a waveform in this scenario looks like.

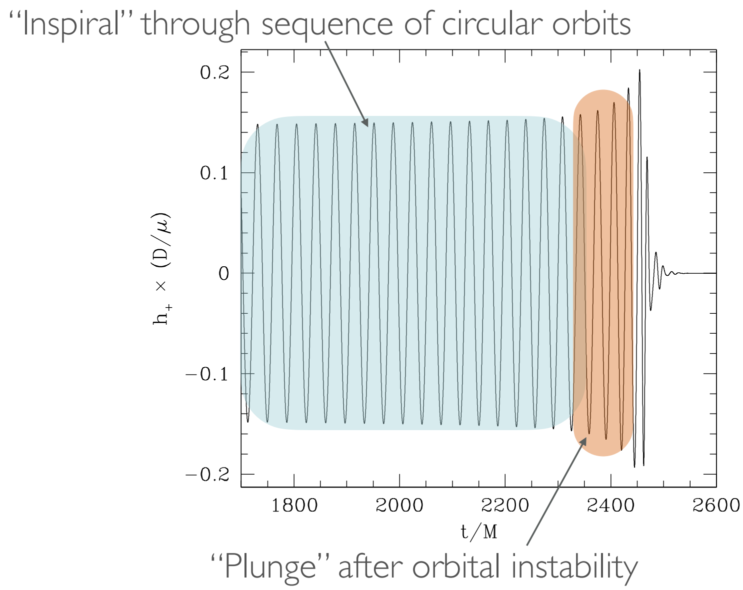


Figure 3: An example of a theoretical model of a gravitational waveform computed for bodies in circular orbit about one another. The region highlighted in light blue shows the waveform over the range of orbits for which the system is slowly evolving through a sequence of circular orbits; the area in orange corresponds roughly to the waves after the members have come so close to one another that stable circular orbits no longer exist. The last few decaying cycles are described in more detail below. Note the units are a little different from what we generally use in this class; multiply M by G/c^3 on the horizontal axis, and multiply μ [the system's reduced mass, $\mu \equiv m_1 m_2/(m_1 + m_2)$] by G/c^2 on the vertical. D is the distance to the binary from the detector that observes this waveform.

This figure shows us one example of the gravitational waveform produced by two bodies orbiting each other; this example was computed for a system with a mass ratio of 10. In this case, we see a train of cycles in which the amplitude starts out changing fairly slowly. This is what we expect at this mass ratio when the system is evolving through a sequence of stable circular orbits. The amplitude starts changing much more rapidly when the members of the binary become close enough that a stable orbit no longer exists, around $t \sim 2300GM/c^3$. At this point, they plunge toward one another, accelerating very rapidly, generating very strong waves at least until they merge into one object. (The nature of the final damped cycles at the end we describe a bit further below.)

We have been able to compute waveforms like that shown in Fig. 3 for quite a while, but measuring these waves is a challenge, thanks to the fact that the effect we are trying to measure amounts to a timing variation of about 1 part in 10^{22}. Hard work, much of it done by colleagues here at MIT, steadily improved the sensitivity of the antennae which can measure this effect. For many of us, the world changed in Fall of 2015, when the two detectors of the LIGO Laboratory recorded the signals shown in the top panels of Fig. 4.

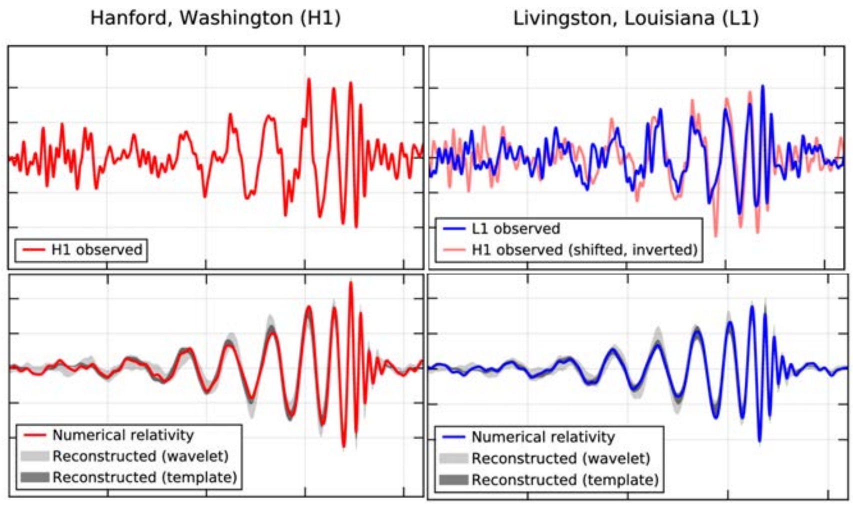


Figure 4: The first directly observed gravitational wave event. The data in the top two panels (which show the traces of h picked up by the two LIGO detectors) match superbly to theoretical models of the waves produced by black holes of mass $29\,M_\odot$ and $36\,M_\odot$ merging to produce a single black hole of $62\,M_\odot$.

These data show the gravitational waveform that was picked up by the two antennae run by the LIGO laboratory (one in eastern Washington, one in a pine forest in Louisiana). It also shows the waveforms that are predicted by solving the Einstein field equations. (Note that these waveforms are the output of *much* more complicated calculations than we have explored in 8.033! Because there are *two* massive bodies, the spacetime is more complicated than the Schwarzschild or Kerr spacetimes we have been studying. Supercomputer simulations are needed to model these solutions in general, though a lot of insight comes from careful "analytic" modeling as well.) The agreement between theory and data is superb. The conclusion is that a black hole of mass $29\,M_\odot$ merged with a black hole of $36\,M_\odot$, leaving a $62\,M_\odot$ remnant black hole[6] behind.

Since that first discovery, the two LIGO instruments plus the Virgo antenna in Pisa Italy (which was commissioned and joined observations a year or so later) have measured well over 100 such merging black hole pairs, as well as a few events that involve neutron stars. Our universe appears to be full of sources of extremely strong gravity, and Einstein's relativity describes all of our measurements (at least so far!) perfectly well.

21.4 Observing the light ring

The data we have briefly discussed in this lecture covered a few of the features of strong-gravity orbits that we discussed previously — the non-Newtonian orbit shapes that are seen in the motion of stars in our galactic center, and the orbital instability. It should be emphasized that as detectors get more sensitive, and new instruments make it possible to observe different bands[7] of gravitational waves, we expect to be able to "watch" systems

[6]You might notice that some mass appears to be missing — $36 + 29 \neq 62$. In fact, an amount of energy equal to $3\,M_\odot c^2$ was lost due to gravitational radiation produced by the system. Most of that energy was lost in roughly 0.1 seconds. If that energy had been radiated in light rather than GWs, then during that second, this system would have shined more brightly than several hundred *billion* Milky Way galaxies.

[7]Currently active instruments are sensitive to gravitational waves which oscillate in the frequency band several $\times\ 10\,\mathrm{Hz} \lesssim f \lesssim 1000\,\mathrm{Hz}$. Sources which radiate in this band tend to have masses similar to stars

evolve through a wide range of orbits. Orbits with substantial eccentricity are ones that are likely to be especially interesting, and to carry a lot of information that will allow us to probe the nature of these systems.

But what about that light ring? The light ring is one of the most striking predictions of motion in black hole spacetimes. Perhaps the biggest challenge here is one of scale. Consider the black hole in the center of our galaxy, with a mass $M \approx 4 \times 10^6\, M_\odot$. How big do we expect the ring to be in this case?

Recall that for light moving the Schwarzschild spacetime, we expect the ring to be of radius $b = 3\sqrt{3}GM/c^2$. For the black hole in the center of galaxy, this translates to $b \approx 30$ million kilometers. That sounds big! — but the ring is in the center of our galaxy, which is about 27,000 light years from our solar system. Such a ring would have an angular diameter on the sky of

$$\delta\theta_{\rm ring} = \frac{2 \times 30,000,000\,\mathrm{km}}{27,000\,\mathrm{lyear}} \simeq 2.4 \times 10^{-10}\,\mathrm{radians} \approx 0.05\,\mathrm{milliarcseconds}\;. \tag{21.16}$$

This is an *extraordinarily* small angle; recall there are 3600 arcseconds in a degree, and this is smaller than an arcsecond by a factor of 20,000. Further complicating this is that we need to see "through" a lot of intervening gas and plasma, which tends to scatter electromagnetic radiation. By carefully studying the properties of all that "stuff" which is in the way, a team of astronomers deduced that radiation with a wavelength of about 1 millimeter was the best choice to look at the core of our galaxy, as well as the cores of a few nearby galaxies. The galaxy M87 was of particular interest — it is 1000 times farther away than the center of our galaxy, but appears to host a black hole that is about 1000 times more massive. The factors of 1000 cancel out as far the angular size is concerned, and the light ring is similar in size to what we estimated above.

If you're trying to resolve something with an angular size $\delta\theta$, Rayleigh's criterion teaches us that the diameter D of the telescope we need to use is related to the wavelength λ of the radiation we are measuring according to

$$\delta\theta = \frac{1.22\lambda}{D}\;. \tag{21.17}$$

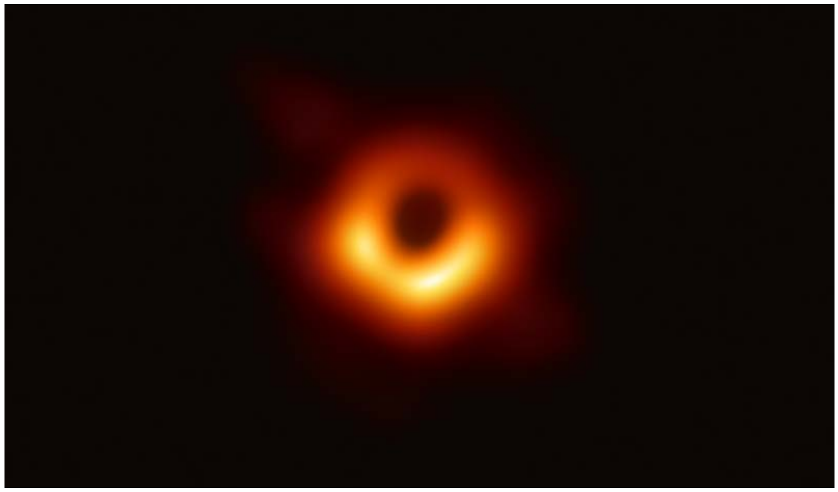
Plugging in $\lambda = 10^{-3}$ meters, and using the $\delta\theta_{\rm ring}$ we estimated above, we find that we need $D = 5000$ kilometers — comparable to the radius of the Earth!

This may seem challenging — and it is. However, we don't need a single telescope of this size; we "just" need to have an array of telescopes that are *separated* by this distance. If we can then combine the data from all these telescopes in just the right way (and doing this requires that we know when each bit of data arrived with a precision better than $\delta t < \lambda/c$, and we need to know the distance between telescopes with a precision of about λ), then we can treat all the data as coming from a single telescope whose size is given by the different telescopes' separations.

Such measurements were done by a multi-month observing campaign focusing on the black hole candidate at the center of M87 by a collaboration called the "Event Horizon

— solar masses up to about a hundred or so solar masses. Planned detectors will broaden this; your lecturer is particularly excited about space-based instruments which will be sensitive in a band of about $10^{-4}\,\mathrm{Hz} \lesssim f \lesssim 0.1\,\mathrm{Hz}$. Waves in this band will come from sources of millions of solar masses, like the kind of black holes that appear to exist in the cores of many galaxies, including our own.

Telescope." They first announced their results in Spring 2019. For our purposes, perhaps the most exciting result is the one shown in Fig. 5.



This image is in the public domain.

Figure 5: Emission in the inner few dozen microarcseconds at the center of the galaxy M87.

This is what a light ring actually looks like for emission deep inside a strong-gravity spacetime. Note that it's thicker than the (rather idealized) picture that we sketched in a previous lecture. This is in part because the illumination which provides the light we observe is itself kind of "lumpy," brighter in some areas than others, and appears to be orbiting around the black hole. Also, the telescope's resolution blurs things out somewhat. Bearing those corrections to our ideal picture in mind, this ends up having exactly the characteristics expected for a black hole light ring in general relativity.

In addition to this light ring, the light ring has an influence on the gravitational waves that we have been measuring since 2015. Look again at the final few cycles of the waveforms shown in Figs. 3 and 4. Notice that they very rapidly decay away, oscillating several times as they do so. These final few cycles oscillate at a period that very closely corresponds to what we expect for light orbiting in the light ring. Gravitational waves propagate across spacetime just as light does, and so **gravitational waves can be trapped in the light ring** just as light rays can get trapped.

That in fact is what we are seeing in those final gravitational wave cycles. Those final cycles can be thought of as gravitational waves from the coalescence that orbit around a few times in the spacetime of the remnant black hole that is left over at the end of coalescence. Because that light ring is an unstable orbit, those last gasps of radiation leak away, gradually reducing in amplitude as more and more of that trapped radiation leaves the strong-field region of the spacetime.

At least so far, measurements done using both light and gravitational waves have confirmed all the various "weird" features associated with strong-gravity spacetimes. All the evidence to date is consistent with gravity behaving exactly like general relativity predicts when it is so strong that its behavior is significantly different from that of Newtonian gravity.

Lecture 22
Our universe at large

22.1 Does $T^{\mu\nu} = 0$ describe our universe?

Strong-gravity spacetimes tell us about "compact" bodies, things that can be localized to some spatial region. They reproduce Newtonian gravity, and they introduce new behavior that (so far, at least!) all fits the data. However, these spacetimes are "asymptotically flat": when we go very far away from the source of mass in the spacetime, we find $ds^2 \to -c^2dt^2 + dx^2 + dy^2 + dz^2$. Does this behavior describe our universe? Spacetimes for which this true all solve the Einstein field equations *if* $T^{\mu\nu} = 0$. Is this an accurate description of our universe?

The answer to this, very clearly, is **no!** Looking out, we see our galaxy, other galaxies, clusters of galaxies, light, gas. Indeed, on the very largest scales, the universe appears to be a uniform fog of matter and radiation, limiting to a haze of microwaves known as the "cosmic microwave background," or CMB, at the largest distances that we are able to probe. However, an interesting property of what we see is that the universe is quite uniform on the largest scales. For example, on the very largest scales we can measure, variations in the CMB are a fraction of about 10^{-5} of its mean level[1]. Things become clumpier on smaller scales because gravity tends to make things clump up.

On the very largest scales — larger than about $10 - 100$ Megaparsecs[2] — we can think of our universe as a perfect fluid. This may seem crazy, but it is an acceptable treament as long as we focus on scales where matter's granularity has no effect. It's kind of the way we treat water as a fluid, even though we know it is made of individual molecules. On large enough scales, the granularity of water cannot be perceived; on large enough scales, the granularity of stars and galaxies cannot be perceived.

22.2 A spacetime for the large-scale structure of the universe

Although the universe is uniform in all spatial directions on the largest lengthscales, it is *not* uniform in time. Light travels at finite speed, so large distances are seen at earlier times. What we see at earlier times is a universe that was much denser than today.

To describe the large-scale structure of our universe's spacetime, we want to use a metric that is uniform in space, but not in time. It can be proven that the most spatially symmetric spacetime has the form

$$ds^2 = -c^2dt^2 + a^2(t)\left[\frac{d\bar{r}^2}{1 - k\bar{r}^2} + \bar{r}^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right] . \qquad (22.1)$$

[1] After correcting for a Doppler effect. The CMB defines a preferred rest frame, and we are moving with respect to that rest frame due to the motion of our solar system with out galaxy, plus the infall of our galaxy toward the local cluster of galaxies.

[2] 1 parsec = 3.26 lightyears. This is a unit of distance that is particularly useful in astronomy, because it arises directly from measurements we can make using parallax.

This is a "Robertson-Walker" spacetime. It has the following important properties:

- The function $a(t)$ is the *scale factor*, and controls the physical scale associated with distance between two objects. If $k = 0$, the distance between $(\bar{r}_1, \theta, \phi)$ and $(\bar{r}_2, \theta, \phi)$ is

$$L = a(t) \left[\bar{r}_2 - \bar{r}_1\right] \ . \tag{22.2}$$

  Notice that if two objects are at spatial rest in the coordinate system (so that $\bar{r}$, θ, and ϕ are all constant) then the physical distance between them is nonetheless changing if $a(t)$ changes with time.

- The coordinate $\bar{r}$ is a dimensionless radial coordinate. For $k = 0$, $a(t)\bar{r}$ is essentially just our "normal" spherical distance.

- The parameter k is called the "spatial curvature" parameter, and takes the value -1, 0, or 1. For $k = 1$, we define

$$\frac{d\bar{r}}{\sqrt{1 - \bar{r}^2}} = d\chi \mapsto \bar{r} = \sin\chi \ . \tag{22.3}$$

  In this case, the value of $\bar{r}$ is bounded: we can never exceed $\bar{r} = 1$. This describes a *closed universe*: the physical separation between objects has a maximum at each moment in time.

  For $k = -1$, we define

$$\frac{d\bar{r}}{\sqrt{1 + \bar{r}^2}} = d\chi \mapsto \bar{r} = \sinh\chi \ . \tag{22.4}$$

  This describes an *open universe*: the physical separation between objects is totally unbounded.

  For $k = 0$, space has a "flat" Euclidean geometry: for $dt = 0$,

$$ds^2 = a(t)^2 \left[d\bar{r}^2 + \bar{r}^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2\right)\right] \ . \tag{22.5}$$

  This is often called a "flat universe," though that is a bit misleading — space**time** is curved.

22.3 Propagation of light in this spacetime

The value of k and the behavior of $a(t)$ are connected to the matter that fills the universe, and can be determined from the Einstein field equations. Before discussing those quantities, it is useful to examine how light and matter behave in these spacetimes.

Begin by asking what happens to observers at rest in the coordinates: $u^t = c$, $u^{\bar{r}} = u^\theta = u^\phi = 0$. When we examine geodesics, we find that they remain fixed at coordinate $(\bar{r}, \theta, \phi)$. However, as those observers remain fixed at that coordinate, we see that the proper separation of observers changes as $a(t)$ changes. Those observers "co-move" as the universe's geometry changes[3].

[3]Note that this *only* applies to comoving points. This means points which do not experience forces which "push" them away from the geodesic. The separation between us and a very distant galaxy changes as $a(t)$ changes. However, that galaxy's size does not change because it is a *bound object* — it is not built out of things that are comoving in the Robertson-Walker spacetime. The scales of things that are bound together — like stars, planets, solar systems, people — do not change as $a(t)$ changes. The Robertson-Walker spacetime describes the geometry of events on very large scales; it doesn't describe things on small scales.

Next examine light — which is our main tool for measuring and understanding our universe. For simplicity, we will focus on $k = 0$. (The calculation can be generalized to $k = \pm 1$, but the details are a bit messy using the tools of 8.033; we'll just quote the result for these cases.) Imagine that light is emitted at some time t_e, and is received by an observer at some time t_r. It's enough to consider light that moves radially, so we'll put $p^\theta = p^\phi = 0$.

Our goal is to compare the energy of light when it is emitted to the energy when it is received. To do this, we imagine one comoving observer measures the light at emission, and another at reception:

$$E_{\rm emit} = -\vec{p}_{\rm emit} \cdot \vec{u}_{\rm emit} = p^t_{\rm emit} c \; . \tag{22.6}$$

Here we used the fact that the comoving observer has only one non-zero 4-velocity component, which we can write $u_t = -c$. Likewise, we find $E_{\rm rec} = p^t_{\rm rec} c$.

Let's now propagate this light across spacetime as a radial geodesic and see what energy it has at $t = t_r$. We use two rules to propagate the light:

1. It follows a light-like trajectory or null trajectory, so $\vec{p} \cdot \vec{p} = 0$:

$$-(p^t)^2 + a^2(t)(p^{\bar{r}})^2 = 0 \quad \rightarrow \quad p^{\bar{r}} = p^t/a(t) \; . \tag{22.7}$$

2. It follows a geodesic, so we extremize

$$L = \frac{1}{2} g_{\alpha\beta} \frac{dx^\alpha}{d\lambda} \frac{dx^\beta}{d\lambda} = -\frac{c^2}{2}\left(\frac{dt}{d\lambda}\right)^2 + \frac{a(t)^2}{2}\left(\frac{d\bar{r}}{d\lambda}\right)^2 \; . \tag{22.8}$$

Let's focus on the $x^0 = ct$ component of the Euler-Lagrange equations:

$$\frac{\partial L}{\partial x^0} = \frac{1}{c}\frac{\partial L}{\partial t} = \frac{1}{c} a\dot{a}\, (p^{\bar{r}})^2 \; , \quad \text{where} \quad \dot{a} = \frac{da}{dt} \; ; \tag{22.9}$$

$$\frac{\partial L}{\partial (dx^0/d\lambda)} = -c\frac{dt}{d\lambda} = -p^t \; ; \tag{22.10}$$

$$\frac{d}{d\lambda}\left[\frac{\partial L}{\partial (dx^0/d\lambda)}\right] = -\frac{dp^t}{d\lambda} \; . \tag{22.11}$$

Put all these ingredients together:

$$\frac{\partial L}{\partial x^0} - \frac{d}{d\lambda}\left[\frac{\partial L}{\partial (dx^0/d\lambda)}\right] = 0 \tag{22.12}$$

becomes

$$\frac{a\dot{a}}{c}\,(p^{\bar{r}})^2 + \frac{dp^t}{d\lambda} = 0 \; . \tag{22.13}$$

Using the constraint $\vec{p} \cdot \vec{p} = 0$, this becomes

$$\frac{1}{c}\frac{\dot{a}}{a}\,(p^t)^2 + \frac{dp^t}{d\lambda} = 0 \; . \tag{22.14}$$

But we also know that

$$\dot{a}\, p^t = \frac{da}{dt}\, c\, \frac{dt}{d\lambda} = c\,\frac{da}{d\lambda} \; . \tag{22.15}$$

With this, our equation becomes

$$\frac{da/d\lambda}{a}p^t + \frac{dp^t}{d\lambda} = 0 \ , \tag{22.16}$$

or

$$\frac{da/d\lambda}{a} = -\frac{dp^t/d\lambda}{p^t} \ , \tag{22.17}$$

Integrate both sides from $\lambda = \lambda_e$ (corresponding to the moment t_e when light is emitted) to $\lambda = \lambda_r$ (corresponding to the moment t_r when light is received):

$$\ln\left[\frac{p^t(t_r)}{p^t(t_e)}\right] = -\ln\left[\frac{a(t_r)}{a(t_e)}\right] \ , \tag{22.18}$$

or

$$\frac{p^t(t_r)}{p^t(t_e)} = \frac{a(t_e)}{a(t_r)} \ . \tag{22.19}$$

From the fact that $E_{\rm emit} = cp^t(t_e)$ and $E_{\rm rec} = cp^t(t_r)$, this means

$$E_{\rm rec} = E_{\rm emit}\left(\frac{a(t_e)}{a(t_r)}\right) \ . \tag{22.20}$$

In other words, *the energy associated with the light that we measure gives us a way to directly probe the scale factor of the universe.* (The result turns out to be identical for $k = \pm 1$.)

So how do we use this? We take advantage of the fact that atoms and molecules whose electrons are in an excited state emit light with distinct spectral lines. Figure 1 illustrates what the spectrum from a gas cloud might look like if the atoms and molecules in the gas all undergo known electronic transitions. The blue curve in this figure illustrates the spectrum in the "rest frame," i.e., what we might measure in a laboratory. In this sketch, we imagine that there are 4 different "lines," each at a wavelength $\lambda_{1,2,3,4}$ that has been very well characterized (e.g., by laboratory measurements and/or theoretical calculations).

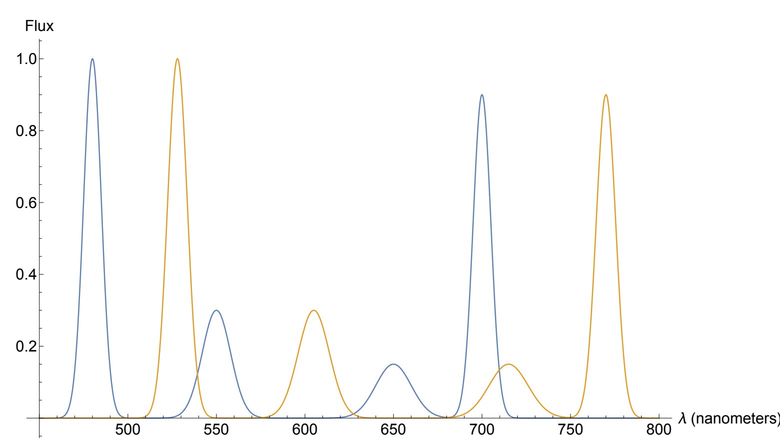


Figure 1: Sketch of how the universe's scale factor affects an object's light emission spectrum. Blue curve sketches a spectrum as it would be viewed in that object's "rest frame." This is what we would measure in the laboratory, for example. Imagine that this light is emitted at t_e, and is measured at t_r. Orange curve shows that same spectrum if it is measured at t_r such that $a(t_r)/a(t_e) = 1.1$.

Imagine that the light is emitted at t_e, when the universe's scale factor is $a(t_e)$. The orange curve in Fig. 1 illustrates what this spectrum might look like if it is measured at t_r, when the scale is now $a(t_r)$. Each photon that contributes to the light has been *redshifted* by the expansion of the universe. Because the energy of light relates to its wavelength according to $E = hc/\lambda$, each "line" at λ_i has been shifted to

$$\lambda_i' = \lambda_i \left(\frac{a(t_r)}{a(t_e)}\right) \equiv \lambda_i(1+z) \ . \tag{22.21}$$

This equation defines the *cosmological redshift*, z. This is what we determine when we measure a spectrum and deduce the nature of the atoms or molecules that emitted its light.

The punchline is that by measuring the spectra of distant objects and looking for the "fingerprints" of known[4] atomic and molecular transitions, we can deduce the scale factor at which the light was emitted, compared to the scale factor's value today. If you do this for a large number of sources, you can build up map of how the scale factor evolves. If we understand how the scale factor evolves as a function of time, we can then use measurements of many different sources' redshifts in order to learn how the universe is evolving.

22.4 The behavior of $a(t)$ and k

If you run the Robertson-Walker line element through the Einstein field equation, you find that the scale factor $a(t)$ and the curvature parameter k are related to the energy density of "stuff" in the universe according to

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3c^2} - \frac{kc^2}{a^2} \ . \tag{22.22}$$

This relationship was first discovered by Alexander Friedmann in 1922, and is known as the Friedmann equation. Any Robertson-Walker spacetime for which $a(t)$ and k connect to ρ by this relationship is known as a Friedmann-Robertson-Walker (or FRW) cosmology.

Before discussing some details, it is useful to introduce some terminology:

$$\frac{\dot{a}}{a} \equiv H \qquad \text{The “Hubble” expansion parameter.} \tag{22.23}$$

Noice that this parameter has the dimensions of 1/time. The value of the Hubble parameter today is a subject of quite a bit of active research:

$$H_0 \equiv H(t = \text{now}) \approx 70(\text{km/sec})/\text{Mpc} \ . \tag{22.24}$$

The precise value of H_0 is somewhat controversial as I write this document, with different techniques yielding somewhat different values, ranging from about 67 in these units up to about 73. Not that long ago, methods that yielded values this close to one another (each differs by about 5% from 70) would have been celebrated as a triumph; when I started graduate school, people were concerned about whether the value was closer to 50 or to 100. One reason that there is a lot of interest in the different values obtained by current measurements is that it is not clear whether these numbers reflect different systematic uncertainties

[4]Note that in principle there's a big assumption being used here: We assume that the basic physics describing atoms and molecules is the same now as when and where the light was emitted.

in the different methods, or whether the physics of the different methods means that they are measuring fundamentally different things.

Another useful parameter is a critical density:

$$\rho_{\text{crit}} = \frac{3H^2c^2}{8\pi G} \ . \tag{22.25}$$

We can normalize density to this value:

$$\Omega \equiv \frac{\rho}{\rho_{\text{crit}}} \ , \tag{22.26}$$

and then rearrange the Friedmann equation using this definition:

$$1 = \frac{8\pi G\rho}{3H^2c^2} - \frac{kc^2}{a^2H^2} = \frac{\rho}{\rho_{\text{crit}}} - \frac{kc^2}{a^2H^2} \ , \tag{22.27}$$

or

$$\Omega - 1 = \frac{kc^2}{a^2H^2} \ . \tag{22.28}$$

This lets us see the significance of ρ_{crit}:

- If $\rho > \rho_{\text{crit}}$, then $\Omega > 1$ and we must have k positive. *We must have a spatially closed universe if $\rho > \rho_{\text{crit}}$.* It can be shown in this case that the universe expands to a maximum size, then recollapses.

- If $\rho < \rho_{\text{crit}}$, then $\Omega < 1$ and we must have k negative. *We must have a spatially open universe if $\rho < \rho_{\text{crit}}$.* It can be shown in this case that the universe expands forever.

- If $\rho = \rho_{\text{crit}}$, then $\Omega = 1$ and we must have $k = 0$. *We must have a spatially flat universe if $\rho = \rho_{\text{crit}}$.* It can be shown in this case that the universe expands forever, but (in most cases) with ever decreasing speed. (There is one interesting an important exception to this trend, which we describe in more detail below.)

To know which of these options corresponds to our universe, we need to know how the universe behaves depending on the mixture of "stuff" that goes into it. This is in general a complicated problem, but we can get insight by looking at a couple of illustrative limiting cases. Let's take a universe with $k = 0$ and fill it with matter in the form of dust[5]. In this limit, the total number of dust particles is fixed, but their density changes as $a(t)$ changes:

$$\rho_M(t) = \rho_M(\text{now}) \left[\frac{a(\text{now})}{a(t)}\right]^3 \equiv \rho_0 \frac{a_0^3}{a(t)^3} \ . \tag{22.29}$$

When you plug this in to the Friedmann equation (with $k = 0$), you find that $a(t)$ can be solved using a power-law in time:

$$a(t) = a_0 \left(\frac{t}{t_0}\right)^n \ . \tag{22.30}$$

Running this through Eq. (22.22), we find $n = 2/3$. This tells us that in a spatially flat "matter-dominated" universe, the scale factor grows as a function of time as $a \propto t^{2/3}$.

[5]Recall that "dust" can be thought of as a perfect fluid with $P = 0$.

This solution implies an expanding universe. If you run it backwards, it implies that $a = 0$ at some point in the past. This means that all spatial locations were smashed into a single zero-size point (assuming that the FRW model holds all the way back to that moment — perhaps a rather big assumption!). Spacetime itself comes into existence as we evolve from that moment. The birth of all of space is known as the "Big Bang." Notice it is not an explosion *into* space — it is the creation of space itself. There wasn't any "there" to explode into until the Big Bang happened!

Another representative example: a universe filled with radiation. Imagine that the number of photons is fixed, but their density varies as $a(t)^{-3}$. In addition, each photon has an energy that itself varies as $1/a(t)$ — the redshift effect. This implies that the energy density of radiation obeys

$$\rho_R(t) = \rho_0 \left(\frac{a_0}{a(t)}\right)^4 . \tag{22.31}$$

This also admits a power-law solution; running it through Friedmann, we find $a(t) \propto t^{1/2}$ in a "radiation-dominated" universe.

One last example has been found to be very important — "vacuum energy," also known as a "cosmological constant." The vacuum energy arises in quantum field theory as an energy associated with the ground state of quantum fields. Its key property is that it must be invariant with respect to Lorentz transformations in the freely-falling frame: $T^{\mu\nu} \propto \eta^{\mu\nu}$ in the FFF. This means that this variety of "stuff" looks like a perfect fluid, but one with *negative pressure*:

$$P_\Lambda = -\rho_\Lambda . \tag{22.32}$$

We can see how this contribution evolves by enforcing the rule that the stress-energy tensor be divergence free; doing so, we find out that ρ_Λ is *constant* with time. This rather odd behavior is a consequence of the fact that this "fluid" is associated with the vacuum itself.

When we plug this behavior for the density into the Friedmann equation, here's what we get:

$$\begin{aligned}
\left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G\rho_\Lambda}{3c^2} \\
\dot{a} &= \pm a\sqrt{\frac{8\pi G\rho_\Lambda}{3c^2}} \\
a(t) &\propto \exp\left[\pm t\sqrt{\frac{8\pi G\rho_\Lambda}{3c^2}}\right] .
\end{aligned} \tag{22.33}$$

This solution yields exponential expansion. (Or contraction; however, expansion dominates, since the contracting solution rapidly crushes away its own relevance.) This case is the exception to $k = 0$ describing expansion with ever decreasing speed. Exponential expansion for $a(t)$ accelerates with time.

The three cases discussed here — matter-dominated, radiation-dominated, vacuum-energy-dominated — are idealized, but demonstrate how different contributions to the universe give different ways in which $a(t)$ evolves with time. We generally expect a mixture of different ingredients, for which these power-law solutions don't apply. But, these limits provide asymptotic solutions which are useful for guiding our understanding. The general case is not too hard to solve for by integrating the Friedmann equation numerically. By measuring the rate of expansion at many different times and comparing to different models, we can infer what our universe is (apparently) made of.

22.5 Measurements and our universe

What we really want, then, is to measure a (which is encoded in the redshift of distant sources) at many values of t. This will let us build up $a(t)$; connect this to some good models for matter in the universe, and we should be able to learn something interesting.

Our main tool for doing this is to measure the *distance* to different objects. Since light's travel speed is known, distance tells us the time at which light was emitted. Time or distance plus redshift lets us build $a(t)$. Two tools are particularly important for doing this:

- *Standard rulers* are sources whose size is known by some physics. We compare the apparent size to the physical size; the ratio tells us the source's distance.

- *Standard candles* are sources whose intrinsic brightness is known. Compare the apparent and intrinsic brightness; the ratio again tells us the source's distance.

Doing measurements of this kind is an industry. The basic idea is to build a large data set containing high-quality data describing distance versus redshift for class of sources, and then find the solution to the Friedmann equations — some self-consistent solution with H_0, Ω_M, Ω_Λ, Ω_r, and k — that bests describes these data.

Nearly current data[6] (at least, as of the writing of these notes) tells us

$$\Omega_M = \rho_M/\rho_{\rm crit} = 0.311 \pm 0.006 \tag{22.34}$$
$$\Omega_\Lambda = \rho_\Lambda/\rho_{\rm crit} = 0.689 \pm 0.006 \tag{22.35}$$
$$\Omega_{\rm total} \equiv \Omega_M + \Omega_\Lambda = 0.9993 \pm 0.0019\ . \tag{22.36}$$

(The contribution of radiation, Ω_r, is so small it doesn't show up in this table.) The data are consistent with $k = 0$, telling us that our universe appears to be spatially flat.

This is lovely ... but there is some weirdness under the hood. Here are a few current mysteries:

1. **What's the real value of H_0?** As mentioned, the value of H_0 is something that different techniques disagree on. The table above is based on one of those values (which is "self consistent" with the technique that contributes the most to that dataset), but other values differ. Is there something prosaic skewing some of the measurements? Or is there something deeper going on — perhaps we have overlooked some contributor to the Friedmann equations whose importance is not obvious right now?

2. **Why is $k = 0$?** One can show that if $\Omega - 1 = \epsilon$, then $|\epsilon|$ *grows* with time in a matter- or radiation-dominated universe (becoming larger in magnitude, whether positive or negative). In other words, the deviation from spatial flatness should be magnified as the universe evolves, if the universe is matter or radiation dominated. Observations indeed indicate that our universe is matter dominated now, and was radiation dominated long ago (greater than about 13.5 billion years ago). For ϵ to be so close to zero today, it would have had to be even closer — many digits closer — at a very early time in the universe's history.

   If, however, the universe is *not* matter or radiation dominated, but is instead vacuum-energy dominated, then it is not hard to show that $\Omega - 1$ evolves to zero as $a(t)$

[6] Numbers taken from `https://pdg.lbl.gov/2021/reviews/rpp2021-rev-cosmological-parameters.pdf`

exponentially expands. A way out is thus to imagine that the universe was in such a state at very early times — perhaps very, very early in the universe's history, before it became radiation dominated. The idea that our universe behaved this way constitutes the theory of *cosmic inflation*.

Inflation comes in different flavors, depending upon details of how one designs the energy of the "vacuum" (more correctly, the *false vacuum*) that drives the expansion. The version most people look at for this today, whose foundations were developed by Alan Guth about 40 years ago, suggests that our universe exponentially inflated for about 10^{-30} seconds at a very early time. If this is the case, then inflation very likely left a mark in the form of very weak gravitational waves that have a unique and very broad spectrum, stretching from the band to which LIGO is sensitive now, down to frequencies of order 1/(billions of years). Searching for the imprint of these waves is one of the top problems in observational cosmology today.

3. **What is the matter that contributes to Ω_M?** If we add up all the matter we can see that produces light — stuff we know about from the standard model of particle physics — we get
$$\Omega_b = 0.0489 \pm 0.0003 \; . \tag{22.37}$$
(The b on this symbol stands for "baryon," since most of the mass comes from protons and neutrons and the atoms that are built from them.) This is *way* smaller than $\Omega_M = 0.311$. The remaining $\Omega_{\rm DM} = 0.262$ is apparently some kind of "dark" matter. We can see its gravitational influence, but have never detected any "dark matter particle" in any experiment. Lots of people have proposed different ways that matter can produce gravity, but (apparently!) not couple to electromagnetic fields (or, at best, couple weakly enough to evade all detection limits so far). We're still working on this one.

4. **What is Ω_Λ?** The fact that the vacuum energy plays an important role in cosmology today was a rather large surprise when it was first clearly measured about 25 years ago. We are kind of baffled as to what this ingredient in the universe's "energy budget" consists of; indeed, just last year, preliminary evidence was presented hinting that it might not be the "cosmological constant" that one normally thinks of in this context, but might be something even weirder[7].

It is very interesting that when we apply general relativity to compact, strong-gravity objects, it passes every quantitative test we have been able formulate so far. When we apply general relativity on the largest scales, we find it can describe what we observe just fine, but it tells us that our universe is even weirder than we realized. This is a story which is not even close to being over.

[7]See https://arXiv.org/abs/2404.08056 for references presenting this preliminary evidence, as well as discussion urging caution about the "evolving dark energy" interpretation.

8.033 Introduction to Relativity and Spacetime Physics
Fall 2024