

Special Relativity

A brief explanation by John Gordon

Introduction

It is well known that Einstein's Special Theory of Relativity shows that when an observer, say Bob travels relative to another observer, Alice, at a speed which is not small relative to the speed of light weird things happen. As far as Alice can determine, Bob's clocks slow down, he grows old more slowly ("time dilation"), he becomes thinner, and events that he regards as simultaneous appear to Alice to take place at different times.

Most people know these things, but what is not well understood is why this happens, or how to derive the mathematical relationships involved.

This short note aims to help.

For observers moving relative to each other at speeds of up to say a thousand kilometres per second, Newton's laws of motion apply perfectly adequately. Since bullets and rockets come well within this limit, Newton's laws work fine for most practical situations.

But these laws break down when observers approach the speed of light, c which is about 3×10^8 metres/sec. *Special Relativity* is the name given to the theory which describes the weird effects that occur when observers move at steady speeds in straight lines at speeds approaching c . Such speeds are often called *relativistic* because only laws of Relativity adequately explain what happens at these speeds. The word *special* means the theory only applies to special cases where things move at a steady speed in a fixed direction. To cover accelerations, curved motion and gravity we need Einstein's more complicated *General Theory of Relativity* not covered here.

Simple causes

Amazingly, the results of Special Relativity follow directly from very simple causes as follows.

1. There is no such thing as absolute speed and so there is no experiment an observer can do to find his absolute speed. He can only determine his speed relative to another observer or to some object.
2. Crucially, this even applies if he measures the speed of light. No matter how fast he travels along a beam of light he will always measure the speed of that light beam to be exactly the same value. As a consequence he cannot infer his "absolute" speed by measuring his speed relative to a light beam.
3. Because there is no such thing as absolute speed there can be no specially privileged observer - all observers are equal in this respect. In the jargon, *all inertial reference frames are equivalent*. So if Bob measures Alice to be moving at speed u relative to him, then Alice also thinks Bob is moving at speed u relative to her.

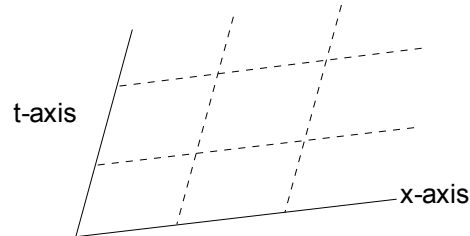
Because amazingly, two observers travelling at very different velocities alongside a light beam both measure the speed of the beam to be the same relative to themselves, something somewhere has to give. And the thing that gives is any assumptions we might cherish regarding the constancy of the lengths of rulers or the rates at which identical clocks tick when they travel very fast relative to us.

In fact we would perceive rulers to shrink and clocks to slow down in curious ways when they travel at relativistic speeds relative to us. Moreover these changes in the perceived properties of clocks and rulers are precisely those needed to bring about (1) to (3) above.

Spacetime diagrams

We can derive all the relevant relationships from *Spacetime Diagrams*. These are used to represent relativistic events in a way which visually resolves various paradoxes (such as time dilation) in a satisfying way. They can also be used to infer the mathematical relationships involved.

Spacetime diagrams use parallel coordinates where the axes are not necessarily rectangular (i.e. not at right-angles to each other), like this.



They can also be thought of as rectangular coordinates but on an elastic membrane which has been uniformly stretched.

One of the reasons we don't insist the axes are at right angles is that we will want to show several sets of coordinates in the same diagram each stretched by different amounts. At most, only one such set of coordinates could be at right angles so we have no choice but to get used to the idea of non-rectangular, parallel coordinates.

The more horizontal of the two axes represents distance. We'll call it the x -axis. The more vertical axis represents time, and we'll call it the t -axis. The origin is taken as $x = 0$ units and $t = 0$ units.

All our spacetime diagrams will show only one distance dimension plus the time dimension, since that way it fits nicely onto a flat surface. (To represent 2 or 3 distance dimensions plus the time dimension would need 3 or 4 dimensions.)

A point on a spacetime diagram is called a *spacetime event*, or more simply an *event*. Such an event is clearly confined to an instant of time at a specific position in space. Two events occurring at the same place and time would be represented by the same point. If two events appear as distinct points then they take must place at different places or different times or both.

The coordinates of a spacetime event will be written as the pair (X, T) where X and T are the coordinates of the event.

A line on a spacetime diagram represents the trajectory through space and time of an object. If the object moves at a steady speed in the same direction its trajectory is a straight line.

We can think of the t -axis simply as the trajectory of an object at the origin which doesn't move in space. Any line parallel to the t -axis is the trajectory of an object at some some fixed spacial distance from the origin which doesn't move.

We can think of the x -axis as the locus of all possible objects at time $t = 0$.

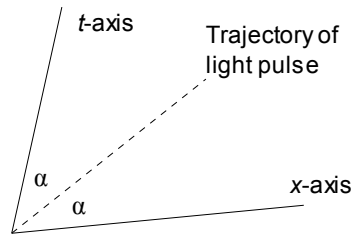
Dimensions

We will use a system of measurement in which the unit of length is defined as the distance travelled by light in one unit of time. If the speed of light is written c , this is equivalent to specifying that $c = 1$. Alternately we can think of our speeds as the fractions resulting from dividing the actual speeds by the speed of light.

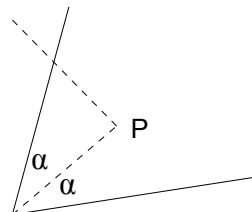
Either way, something that moves at 15% of the speed of light would have a speed 0.15.

Also the scale factors for the two axes are carefully matched so that a unit of length on the x -axis is the same size as a unit of time on the t -axis. With these two conventions the trajectory of a pulse of

light from the origin is a line bisecting the two axes, like this.



A second pulse of light in the opposite direction will be shown at right angles to the first. This is illustrated below for a pulse of light reflected back at the event P.

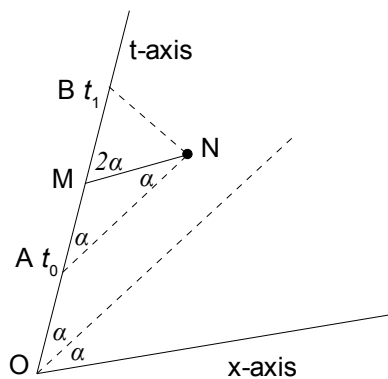


It is normal to orient spacetime diagrams so that trajectories of pulses of light are at 45 degrees to the horizontal and vertical. Because light trajectories always bisect the axes, clearly the x -axis must be tilted from the horizontal by the same angle as the t -axis is tilted from the vertical. With this arrangement of axes, if the slope of the t -axis is m then the slope of the x -axis will be $1/m$.

It is common to include the trajectory of a pulse of light in spacetime diagrams as a dotted line at 45 degrees to the page.

Measuring a spacetime event

Now for a crucial result. The coordinates of spacetime events on our diagrams are more than just abstract representations, they are what an actual observer making intelligent measurements would calculate for the position and time of the event. We'll now show this.



Alice sends out a pulse of light or radar which subsequently strikes a target, and some of the energy is reflected back to Alice who notes its time of arrival, thereby deducing the range of the target and the time at which the pulse must have reached it. In the diagram above, A is the event of Alice sending out the pulse at time t_0 , N is the event where the pulse is reflected from the target, and B at time t_1 is the event where the reflected pulse reaches her again.

Although Alice uses a radar (or light) echo location technique, the results we are going to prove turn out to be valid for any other method Alice could use to determine the spacetime position of an event.

Alice may know nothing about the diagram we have drawn, but we will show that nevertheless her measurements of the time and position of N turn out to be the projections of N onto the t -axis and

the x -axis in the diagram.

She concludes that the time the reflection took place (i.e the time of event N) was the halfway time of t_0 and t_1 , namely $t_N = (t_0 + t_1)/2$, and since the light beam took $\delta t = (t_1 - t_0)/2$ time units to reach N, it follows that N must be a distance $x_N = c(t_1 - t_0)/2$ from her. Since by our units and scale convention, $c = 1$ she concludes that $x_N = (t_1 - t_0)/2$.

In short she concludes that $t_N = (t_0 + t_1)/2$ and $x_N = (t_1 - t_0)/2$.

We now show that t_N and x_N are also the projections of N onto the t -axis and x -axis respectively.

In the diagram, using elementary geometry it is easy to show that lines AM, BM and MN are all the same length, namely $(t_1 - t_0)/2$. This follows because ABN is a right-angled triangle, so N lies on the circle whose diameter is AB, and its midpoint, M is the centre of the circle.

Because AM and MN are the same length, and using rules for angles in triangles, it is easy to verify the relationships implied by marking angles as α or 2α . From these, MN is seen to be parallel to the x -axis.

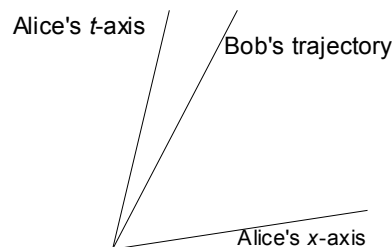
Therefore M is the projection of N onto the t -axis. But M is at a distance $(t_0 + t_1)/2$ along the t -axis which is also Alice's measurement of the time of event N.

Because MN is parallel to the x -axis, the length of MN is the x -coordinate of the projection of N onto the x -axis. This length is $(t_1 - t_0)/2$ which is also Alice's measurement of the position of event N.

In summary, the coordinates of any event on an observer's spacetime diagram equate with the measurements the observer would actually make when determining the position and time of that event.

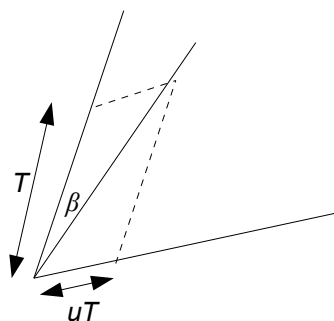
Two Observers

If we have two observers called Alice and Bob, and Bob is moving relative to Alice, then in Alice's coordinate system Bob's trajectory is just a straight line as in the diagram below.



The origin in this diagram is the event where Bob and Alice pass each other.

If Bob is moving with velocity u , the angle of Bob's trajectory on this diagram is determined by the fact that in T units of Alice's time Bob moves a distance which Alice calls uT as shown below.



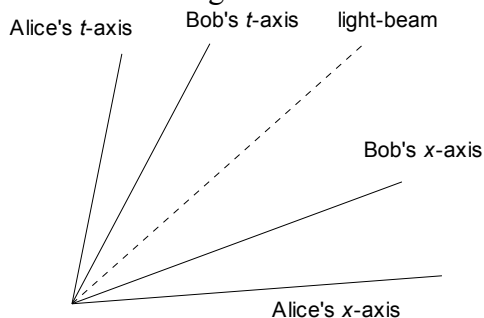
So uT is just the projection onto Alice's x -axis of Bob's trajectory during a time which Alice

measures as T time units. If the axes are at right angles then $u = \tan(\beta)$ but in general they are not.

Suppose Bob now sets up his own coordinate system which travels with him at his own speed, and which therefore appears stationary with respect to him. A moment's thought tells us that Bob's trajectory on Alice's diagram is just another name for Bob's t -axis, because it is the trajectory of points which don't move according to Bob's coordinate system.

From here it is just a few steps to show Bob's entire coordinate system on Alice's diagram.

All we have to do is add Bob's x -axis. Now because, according to the theory of special relativity, all observers measure the same value for the speed of light, c , clearly in Bob's coordinates a light beam will also bisect his axes. We can accommodate this requirement by placing Bob's x -axis as a symmetric reflection of his t -axis in the 45 degree line as shown below.



No special observer

We could equally well have started with Bob's coordinates and represented Alice's trajectory on it, then called this Alice's t -axis and completed the diagram by adding Alice's x -axis.

In other words we can't tell just by looking at the diagram whether it shows Bob's coordinate system in terms of Alice's or *vice versa*. In fact this diagram doesn't belong to either Alice or Bob. It is a universal diagram which shows both their coordinate systems as they both study the same light-beam.

This is why we didn't start with Alice's axes at right angles. If we were to do so, we would be tempted to suppose Alice's coordinate system is in some sense special.

Because the angle which Alice's (or Bob's) t -axis make with the vertical is arbitrary, there are infinitely many equivalent representations of this state of affairs, but they all yield the same answers.

Having said that, it is nevertheless often *convenient* to have one set of coordinates (say Alice's) at right angles as this may make the algebra easier when deriving complicated relationships. This is fine as long as we recognise that such a representation is merely a device to make life easier, and does not affect the final answer. In fact where we represent more than one observer's coordinates on the same diagram we can always choose any one of them to have rectangular coordinates if that should turn out to be useful.

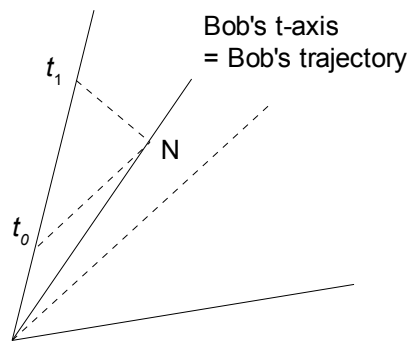
Measuring speed

It is instructive to show that if Alice carries out a realistic experiment to measure Bob's velocity she will really obtain the value u .

If Alice zeros her clock when Bob passes her, all she needs to do to determine Bob's relative velocity is to determine the coordinates of an arbitrary event on Bob's trajectory. A suitable event is the reflection from Bob of a light pulse she sends out at an arbitrary time.

Here we're going to use Alice's coordinates for all times and distances. At time t_0 Alice sends out a

pulse of light (or radar) towards Bob which is reflected back from Bob and reaches Alice at time t_1 . The situation is illustrated below where N marks Bob's position when the light pulse is reflected.



This is clearly just a special case of Alice measuring the time and position of the reflection event N and we already know she will conclude that the time and distance are the projections of N onto her t -axis and x -axis respectively, just as we drew Bob's trajectory originally.

Scale factor

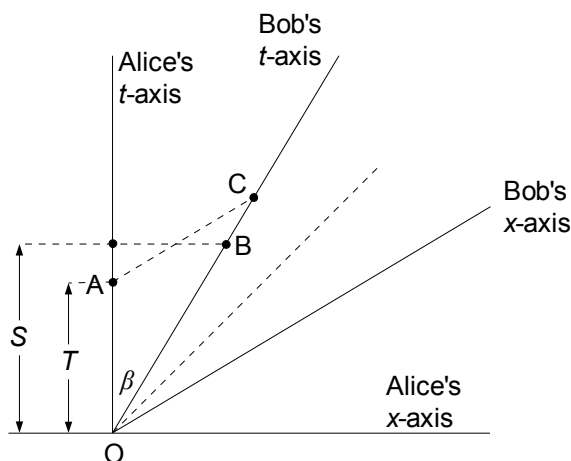
We've seen how to construct a diagram showing Alice's and Bob's coordinate systems when Bob and Alice have a relative velocity u , but this does not show the whole story.

Because of our convention that $c = 1$, one unit of distance and one unit of time in Alice's coordinates are drawn the same length on her spacetime coordinates, and similarly for Bob.

However there is actually a scale-factor conversion needed between the two coordinate systems. In other words one time unit in Alice's coordinates must be drawn as a different length on the diagram from one time unit in Bob's, and similarly Alice's distance units are to be drawn a different size to Bob's distance units. However it's easy to see why this happens and easy to work out the

conversion factor which will turn out to be $\sqrt{\frac{1+u^2}{1-u^2}}$.

The diagram below shows Alice's and Bob's coordinate systems where Bob is moving with velocity u relative to Alice, and where Alice's axes are rectangular.



Point A is an event taking place at $x = 0$ and $t = T$ in Alice's coordinates. Similarly point B is an event taking place at $x = 0$ and $t = T$ in Bob's coordinates.

We can think of A and B as calibration marks on Alice's and Bob's t -axes set to T of their respective time units from the origin.

However point B appears to Alice to be an event at time S in her coordinates. We need to find S .

As we will now show, $\frac{S}{T} = \frac{1}{\sqrt{1-u^2}}$ and the length OB is $T\sqrt{\frac{1+u^2}{1-u^2}}$.

From the diagram it is clear that Alice measures the t -coordinate of B to be S , and Bob measures the t -coordinate of A as the segment OC of his t -axis where AC is parallel to Bob's x -axis.

The clever bit

The key fact to realise is that while Alice perceives Bob's T -calibration mark B to occur at her time S , by symmetry Bob also perceives her T -calibration mark to occur at his time S , so the segment OC of his t -axis represents S in his units.

If Bob's perception is the same as Alice's, then the ratio of S to T must be the same as the ratio of the lengths OC to OB, or even more simply, to the ratio of their vertical components. Given this fact the result is inevitable. The rest is mere detail.

The vertical component of B is just S so we only need the vertical component of C. But this is easy to find as C is just the intersection of line AC with Bob's x -axis.

Using Alice's coordinates, Bob's t -axis has the equation $t = x/u$. The line AC passes through point $(0, T)$ and has slope u (the reciprocal of the slope of the t -axis) so its equation is $t = T + ux$.

Solving these two equations gives point C as $\left(\frac{uT}{1-u^2}, \frac{T}{1-u^2}\right)$

So equating the ratios S to T with the ratio of the vertical components of OC to OB we find $\frac{S}{T} = \frac{T}{S(1-u^2)}$ which immediately gives $\frac{S}{T} = \frac{1}{\sqrt{1-u^2}}$ as promised.

The expression $\sqrt{1-u^2}$ frequently crops up so we'll define $\lambda = \sqrt{1-u^2}$ and write $S/T = 1/\lambda$. Note that λ is less than unity.

Putting it another way, time calibration marks on Bob's t -axis that he perceives as 1 time unit apart, project onto Alice's t -axis separated by $1/\lambda$ of her time units. So Alice perceives Bob's clock to be running slow by the factor λ .

We can now write down the scale conversion factor between Bob's and Alice's coordinates.

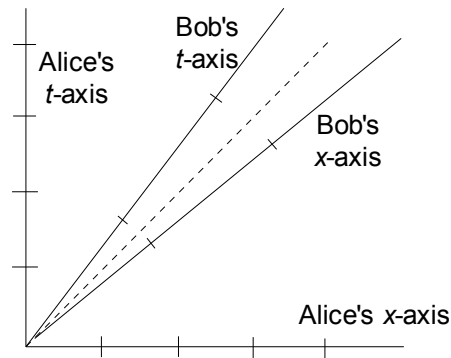
Because the slope of Bob's t -axis is $1/u$ we know that $\tan(\beta) = u$ where β is the angle marked on the diagram. Therefore the length of OB is just $S/\cos(\beta) = S\sqrt{1+u^2} = T\sqrt{\frac{1+u^2}{1-u^2}}$ as promised. OB is the position on Bob's t -axis of an event his clock says occurs at time T , the scale conversion factor between the coordinates is simply $\sqrt{\frac{1+u^2}{1-u^2}}$.

If we write $\mu = \sqrt{1+u^2}$ then the scale conversion factor between Bob's and Alice's diagrams is μ/λ .

So the distance on our diagram that corresponds to T of Bob's time units along his t -axis is $\mu T/\lambda$, while it's projection onto Alice's t -axis is T/λ .

Merely looking at the diagram above helps explain how both observers perceive the other's time and space dimensions to warp in the same way. Everything drops into place when we take account that Bob's and Alice's coordinates are to different scales and their axes are at different angles.

The diagram below shows calibration marks one unit apart in both Alice and Bob's coordinate systems, showing the effects of the scale factor when u is about 0.8, making μ/λ about 2.13.

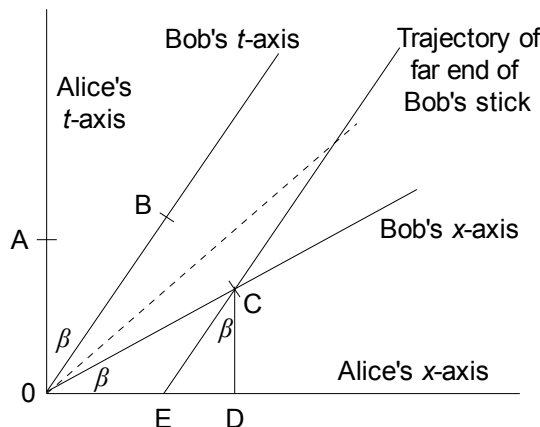


In the appendix is shown another way to derive some of these relationships without using spacetime diagrams.

Length Contraction

Because in Alice's coordinates the calibrations on Bob's t -axis are stretched by the factor μ/λ , and because of the convention that both observers use a measuring system in which the speed of light, $c = 1$, it follows that the calibrations on Bob's x -axis are stretched by the same factor μ/λ , so one of Bob's distance units is drawn as μ/λ of Alice's distance units long.

Furthermore we can describe a line on a spacetime diagram as so many units long without needing to specify if these are time units or distance units – it doesn't affect the length of the line on the diagram since a unit of distance is the length light travels in a unit of time.



The diagram above shows Alice's and Bob's coordinate systems, with Bob travelling at speed u with respect to Alice, and with Bob's units properly scaled up as discussed. Alice's axes are set to be rectangular while Bob's are displaced through angle β where $\tan(\beta) = u$.

Bob has a stick of length k (in his distance units), one end of which is at the origin of his coordinates and the other at a distance k along his x -axis. This stick travels with Bob at his speed of u .

In the diagram OA , shown merely for reference, represents k of Alice's units of time and is therefore drawn k of Alice's units long.

Line OB represents k of Bob's time units, and is therefore drawn as $\mu k/\lambda$ of Alice's units long.

Line OC represents k of Bob's distance units and is therefore also drawn as $\mu k/\lambda$ of Alice's units long (i.e. the same length as OB). Because OC occupies part of Bob's x -axis and is k of Bob's units long it represents Bob's stick at $t = 0$ in his coordinates.

D is the projection of C onto Alice's x -axis.

The extended line EC is the spacetime trajectory of the far end of Bob's stick.

Any sensible measurement Alice makes of the length of Bob's stick will boil down to determining the distance between the lines OB and EC measured parallel to her x -axis, i.e. at some fixed time in her coordinate system. We will calculate this distance at time $t = 0$, which is just the length of the line OE.

Since $\tan(\beta) = u$ it follows that $\sin(\beta) = u/\mu$ and $\cos(\beta) = 1/\mu$ and hence OD has length k/λ , CD has length ku/λ and ED has length $(ku/\lambda)\sin(\beta) = ku^2/\lambda$.

So the length of ED is the difference in lengths of OD and ED, i.e. $(k/\lambda)(1 - u^2)$ which since $\lambda = \sqrt{1 - u^2}$ boils down to $k\lambda$.

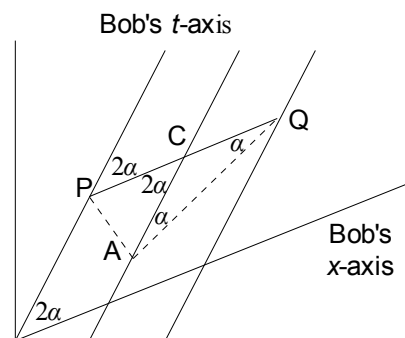
Therefore Alice measures Bob's stick to be $k\lambda$ long whereas Bob measures it to be k long.

The stick has contracted by the factor $\lambda = \sqrt{1 - u^2}$ as far as Alice is concerned. This effect is sometimes called the *Lorentz Contraction*, which is why we chose the symbol λ which is Greek for L .

Simultaneity

Events that are simultaneous to one observer are not in general simultaneous as measured by another observer moving relative to the first.

Suppose Bob moves at speed u relative to Alice. He has a long ruler with the mid point marked and he sends a two pulses of light from the midpoint, one to each of the two endpoints. When these pulses reach their endpoints they trigger events that Alice can measure in some way (for instance they could cause marks on the ground, and light pulses from Alice could detect the creation of these marks). Clearly Bob perceives these two events to be simultaneous, but we will see that Alice does not.



In the diagram above, for convenience Alice's axes are at right angles, and the three parallel sloping lines show the trajectories of three equally spaced points (left, middle and right endpoints of his ruler) in Bob's coordinate system. For convenience the left endpoint of the ruler is also Bob's t -axis.

Bob's light pulses, AP and AQ are shown as dotted lines and P and Q are the events where the pulses reach the two endpoints and make marks for Alice to probe. Clearly in Alice's coordinates, P is measured as occurring before Q.

It is easy to show that line PQ is parallel to Bob's x -axis. To see this it is sufficient (1) to see that PQ is the hypotenuse of the right angled triangle APQ, and if C is the midpoint of PQ then CP, CQ and CA have equal length, then (2) to verify the logic behind labelling various angles as α or 2α .

So events that are simultaneous in Bob's coordinates are on a line parallel to his x -axis. This also means they have the same t -coordinate in Bob's coordinates, which is reassuring.

However in Alice's coordinates they do not occur on a line parallel to her x -axis, and hence seem to occur at different times.

The extended line PQ is also the locus of all events which Bob deems to be simultaneous with the light pulses reaching his ruler's endpoints.

Appendix - Another derivation of Time Dilation

Because the idea of clocks slowing down is so mind-boggling we'll derive R without even drawing a diagram.

Suppose Alice and Bob synchronise clocks to zero when they pass each other. Later at Alice's time T she takes a flash picture of Bob's clock (she has a very bright flashgun!).

When the flash pulse reaches Bob, suppose his clock reads kT , where k is a factor we're going to find in a moment.

Since Bob's clock was synchronised with Alice's when they passed, it's fairly obvious that the time on his clock when the flash reaches him will be proportional to T , (the time when Alice fired it), hence we're calling it kT .

It's necessary to think about this next bit for a moment, but by the symmetry of the situation, the (reflected) flash from Bob's clock must reach Alice at time $k(kT) = k^2 T$.

Alice calculates the time the flash arrived at Bob to be the mid-time of T and $k^2 T$, namely $T(k^2 + 1)/2$ but the picture she took clearly shows an image of Bob's clock reading kT . The ratio $\frac{2k}{(1+k^2)}$ is the factor by which Alice perceives Bob's clock to be running slow.

To pin this down all we need is to find k . We do this by noting that Alice could use the same data to measure Bob's velocity, which we already know to be u .

She knows that the flash beam was travelling towards Bob for $T(k^2 - 1)/2$ time units so when it reached Bob his distance was $cT(k^2 - 1)/2$ units. At that instant he had been travelling for

$T(k^2 + 1)/2$ time units so his velocity must be $u = c \frac{(k^2 - 1)}{k^2 + 1} = \frac{k^2 - 1}{k^2 + 1}$, from which it follows that

$k = \sqrt{\frac{1+u}{1-u}}$. Substituting this back into $\frac{2k}{(1+k^2)}$ we get $\lambda = \sqrt{1-u^2}$ as before.

We're not going to continue this line of reasoning to cover length contraction or the question of simultaneity since it gets tricky, but the point is we've obtained a key result by independent reasoning and this gives us some confidence in spacetime diagrams.

Normally it's not only much easier with spacetime diagrams, but in addition we get the visual insight of a picture.