Chapter 11
Special Relativity as Geometry

In the Pythagorean theorem's application to Cartesian coordinate systems, the distance *s* separating two places is calculated using the formula $s^2 = x^2 + y^2 + z^2$. To our three dimensions of space, though, we must add the imaginary dimension of time if we are to measure the combined space-time distance between *events*. Remember what we decided to do with imaginary quantities in the last chapter? They get subtracted from the sum rather than added. We'll see in a moment why this makes perfect sense in this case. So our total "space-time" distance will be $s^2 = x^2 + y^2 + z^2 - t^2$. We are measuring space in feet and time in seconds, so to make our units match up, we have to add in a factor for the speed of light (*c*) which in our fictitious case is 44 feet per second. Where c=44, $s^2 = x^2 + y^2 + z^2 - (ct)^2$.

Using the Pythagorean theorem, you calculate the distance between the events of your front and rear hazard lights blinking. You measure fifteen feet of separation on the x axis, and no difference in time:

$s^2 = 15^2 + 0^2 + 0^2 -(44*0)^2 = 225$

Measuring 17.3 feet of separation and two-tenths of a second between the events, your friend calculates:

$s^2 = 17.3^2 + 0^2 + 0^2 -(44*0.20)^2 = 17.3^2 + 0^2 + 0^2 -(8.66)^2 = 225$

The square root of 225 is 15, so both of you will agree that if the two events had been measured in a frame of reference where they were simultaneous, the distance *s* between them would be 15 feet apart. 15 feet is called the "rest length" or "proper length" between the two events.

This quantity s^2 is called the *spacetime interval*. It is independent of the frame of reference from which it is measured, and it is the "four-dimensional" distance between two events. It is calculated much the same as we have done it, albeit with the actual value of the speed of light. The right hand side of the equation may also be multiplied by -1. Since the left-hand side is a squared value, this does not fundamentally change the equation.

Equation 11-1. $s^2 = x^2 + y^2 + z^2 - (ct)^2$.

or

Equation 11-2. $s^2 = (ct)^2 - x^2 - y^2 - z^2$

Spacetime intervals can also have a negative value. This depends on which of the two conventions above (Equation 11-1 or 11-2) is followed. In the system we have been using, where positive intervals represent a distance in space, negative intervals represent a distance in time. Let's say you count half a second between two events which are local to you, for instance your right rear turn signal blinking on and then off again.

Using the Pythagorean theorem, you calculate the spacetime interval between events:

$s^2 = 0^2 + 0^2 + 0^2 -(44*0.5)^2 = -484$

Your friend, approaching from behind, measures the time between the events to be longer by the Lorentz factor. He measures a delay of 0.58 seconds. Since he is moving relative to your turn signal during this time, he also does not measure both events to happen in the same place in his reference frame (this is the same as the dropping ball problem in chapter five). He is further from your car when the light blinks on and nearer when it blinks off again. He measures a difference of 13 feet between the two events and calculates:

$s^2 = 13^2 + 0^2 + 0^2 -(44*0.58)^2 = -484$

In both cases, s^2 is calculated as -484. The square root of this is the imaginary quantity 22*i* and (as above) is in units of feet. Dividing this by our fictitious speed of light (44 f.p.s.), we transform this into 0.5*i* and our units become seconds.

$$22i \text{ ft} \div \frac{44 \text{ ft}}{\text{sec}} = 22i \text{ ft} \times \frac{\text{sec}}{44 \text{ ft}} = 0.5i \text{ sec}$$

This number 0.5i is the “imaginary distance,” or time, between the two events in a reference frame where they happen in the same place. So both you and your friend will agree that the local time or “proper time” between the two events, the time measured between the events in a reference frame where they are local, is a half-second.

When *s* is real, it represents a proper length. When it is imaginary, it represents a proper time. There is one final case to discuss in which *s* is zero. This happens when the space and time terms of the formula for the spacetime interval balance out. In this case, the Pythagorean sum $x^2 + y^2 + z^2$, which is the square of spatial distance between two events, is equal to the term $(ct)^2$, which is the square of the distance traveled by any ray of light during the time between the two events. These two parts of the equation, being equal but opposite in sign, add to zero. If the spacetime interval between two events is zero, this means that a ray of light leaving the first event in the direction of the second event would intersect the second event; that is the light ray would arrive at the place of the second event at the exact same time as the second event. If we were to look at the night sky, there would be a spacetime interval between the event which we would measure as “here and now” and any events we see in the sky. This interval would be zero in all cases. We see the events we see now because their spatial, radial distance from us is exactly equal to the distance light travels between “now” and the time they occurred. The light we see from the moon is just over one second old. The moon is just over one light-second distant from the earth.

Now let us relate the spacetime interval to the concept of the light cone diagram which we introduced in chapter two. Our event is at the center. According to the convention we have followed so far in calculating the spacetime interval, events having a positive spacetime interval in relation to ours are *impresent,* or outside the cones. Events which define a negative interval in relation to our event are the *past* and *future*, inside one of the two cones. The events having a zero spacetime in relation to ours are the *present.* These events define the surface of the cones. They are the events we see right now (the lower cone) and the events from which other observers may see us as we are right now (the upper cone).

It is worth restating the point that the sign (positive, negative, or zero) of the spacetime interval is reciprocal between events. That is, if the interval from A to B is positive (twenty-five, for instance), then the interval from B to A is also positive (also twenty-five). A is impresent to B and B is impresent to A. It is when we take the square root of the spacetime interval to get the proper length between events that the relationships become complementary. A spacetime interval of twenty-five has roots of five and minus five. For example, event A is 5 units to the right of B; B is minus five units to the right (that is, five units to the left) of A.

If the interval from A to B is negative (minus sixteen, for instance), then the interval from B to A is also positive (also minus sixteen). A is within one of the light cones of B and B is within one of the light cones of A. Again, it is when we take the square root of the spacetime interval to get the proper time between events that the relationships become complementary. A spacetime interval of minus twenty-five has roots of 4*i* and -4*i.* For example, event A is 4 seconds before B; B is four seconds after A. A is within B's past light cone, and B is within A's future light cone.

If the interval from A to B is zero, then the interval from B to A is also zero. A is one the boundary of one of the light cones of B and B is on the boundary of one of the light cones of A. Each is present to the other, but the distinction of which is the cause and which is the effect is not found in the value of the interval. One of two cases is true, however: if light flashes occur at A and B, then either A's flash will be observed at B, or B's flash will be observed at A.

The time order of two events may be relative, but causality is not. If there is a frame of reference K in which two events A and B are simultaneous and separated by a positive distance on the x axis, then there are are many frames of reference K- and K+ in which the events occur one after the other and in either order. If K- is moving in the minus x direction relative to K, then A will precede B in K-. If K+ is moving in the positive x direction relative to K, B will precede A in K+. Though the time order is relative, no two observers would disagree about a relationship of causality between A and B. All will agree that A and B are far enough separated in space that one could not have caused the other. Causality cannot happen over a large distance and short time, where the distance s (in light-years, for instance) between the events is greater than the time t (in years).

Having tamed the predictions of relativity somewhat by using the Pythagorean theorem, it is time now to put a more friendly face on the Lorentz equations. Again, we will do so using the principles of trigonometry. First of all, we will change our units of measurement to simplify our equations. As I have said before, constants in equations tend to distract from the important relationships between the variables. Since the speed of light is a constant and the truest measure we have, it would make perfect sense to set *c* equal to one and adjust our units of measure accordingly. In meters per second, *c* is 2.998 x 10^9. In other words, light travels about 300,000 km per second. Expressed in light-years and years, *c* is simply 1. A light-year is the distance light travels in one year, which is very far indeed. Light years and years are good measurements for an astronomical scale. For measurements on earth, we might choose a system of nanoseconds (ns) and light-nanoseconds (cns, about 3 m).

Let us suppose we have a measuring stick which is one light-nanosecond (cns) in length. This is the distance light travels in one nanosecond, being at the constant speed *c*. Let us also suppose that we have another very large instrument in the shape of a carpenter's square. Each arm of the square is also 1 cns long at the inside, and each arm is marked in hundredths of a cns. Moving our square to measure the stick, we can match the length of the stick exactly on either arm.

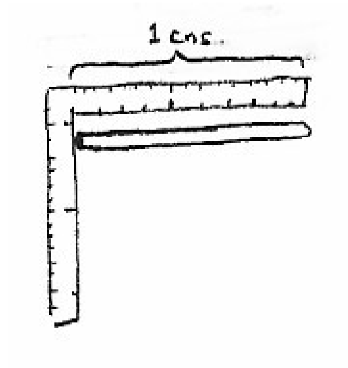

Figure 11-1.

We can also place the square diagonally so that with the stick it forms a right triangle. In this way, the stick touches each arm at a certain length. As the length on one arm decreases, the length on the other increases. These lengths can vary between zero and one cns and always add so that the sum of their squares is one.

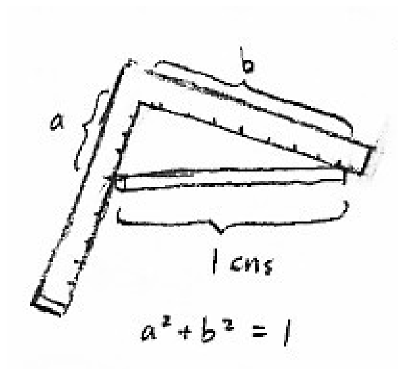

Figure 11-2.

We will arbitrarily pick one arm to represent the velocity *v* measured between us (in reference frame K) and another observer (in reference frame K') in uniform motion with respect to us. Labeled v, the measurement along this arm represents the distance (in light-nanoseconds) traveled by the other observer in one nanosecond. We will label the other arm *w*, *w* being a quantity also having units of light-nanoseconds per second and having a significance which we will establish shortly. The angle opposite *v* we will label as θ. This gives us a right triangle with sides of length *c*, *v*, and *w* (Figure 11-3).

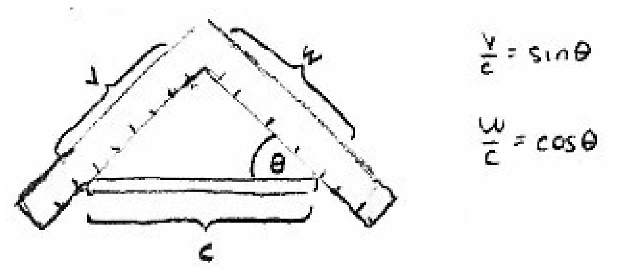

Figure 11-3.

Let's put our measurement system to work now. If we measure the other observer to be at rest, then we place our square so that *v*=0 and *w*=1. Now let us suppose that the other observer is in motion with respect to us at half the speed of light. We place our square so that one end of the stick touches the *v* arm at 0.50 and so that the other end touches the *w* arm. What does the *w* arm measure? It happens to give us the factor by which we will find the other observer to be contracted in the direction of our relative motion. *w* is equal to $\frac{1}{\gamma}$. By some fortune of nature, we have a nice visual aid for relating v to γ. We can move our measuring square from side to side and see how the inverse of the Lorentz factor changes with relative velocity.

But this is not all; there is a mathematical relationship here which can be rewritten to make Lorentz contraction seem more intuitive. The diminishing lengths measured in other reference frames can be shown to correspond mathematically to the diminishing aspect of objects in a rotated perspective (in Figure 1-4, a block looks shorter or longer when rotated 90 degrees). In Figure 11-3, since *c* is equal to one and w is measured in the same units as c (light-nanoseconds per second), we could also say that $w=\frac{c}{\gamma}$. Using trigonometry, we find that where $\frac{v}{c}=\sin\theta$ *, the Lorentz factor γ is* $\frac{1}{\cos\theta}$. The mathematical proof of this is found in Appendix E.

The expression $\frac{1}{\cos\theta}$ for the Lorentz factor may be easier to come to grips with than $\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$, but is this a fair representation of physical reality? By definition in our diagram, $\frac{w}{c}=\cos\theta$. Where c=1, then w=cos *θ.* As shown earlier, we find objects in relative motion to be contracted by the factor $\frac{1}{\gamma}$, which we have found to be equal to *w*, or cos *θ,* where $\frac{v}{c}=\sin\theta$. Again, where *c*=1, *v*=sin *θ.*

So the velocity of the observer K' in our frame of reference K is sin *θ* and their contraction factor is cos *θ.* For all intents and purposes, *θ* is an angle formed between K and K' when they are in motion relative to one another. As *v* approaches the limit of 1 (nothing can go faster than c), w approaches zero (the object in motion shrinks to a depth near zero). An increase in an object's velocity changes its orientation with respect to us such that it is measured to be smaller in the direction of its velocity. It is as though objects at rest in relation to us face us directly so that we measure their full aspect; objects in motion are rotated so that this aspect is diminished.

We might then say that an object in relative motion is “out of phase” with our frame of reference both temporally and spatially. One might also imagine that this rotation or phase difference causes a “slippage of gears” in the moving object's internal clock so that it begins to slow down from our point of view. One might also suppose that the object's linear velocity v in our frame of reference is at all times balanced by the angular velocity of its clock in our frame of reference; as the object gains *linear* velocity, its clock (or any other internal motion, process, or measure of time such as radioactive decay) loses *angular* velocity. This fanciful angular velocity ω (“omega”), of unknown value, would need to be multiplied by a radius r, also of unknown value, to result in a linear velocity w. But together, the resulting linear velocity w and linear velocity v could be added as perpendicular velocity vectors to make the velocity vector c. In the diagram above, where v and w are perpendicular, $v^2 + w^2 = c^2$.

We could also suppose w to be a directly measurable linear velocity. We measure ourselves to have a spatial velocity v of zero in our frame of reference. But our speed w through the imaginary dimension of time is a constant c. The linear vector w, our speed through time, adds to the perpendicular linear vector v, our speed through space, to give us c. In our triangular diagram, $v^2 + w^2 = c^2$, or $0^2 + w^2 = c^2$. As our pair of philosophers continue their earlier conversation about the nature of time, Urania, the Muse of Astronomy, graces our scene, giving Bill and Ted each a rare moment of eloquence and erudition:

> **Bill:** I'm telling you, time and space are the same thing, and the only difference is our *perspective*. What makes the past and future so hard to see is the speed of light, which must also be our speed through time. That means that the past of two seconds ago is further away from than the other side of the moon. The reason we don't see the past or future is because anywhere that we're close enough to look at them, they look *exactly* like the present. So our brain ignores them and collapses our four-dimensional reality into three dimensions.
>
> **Ted**: That still can't be right. Why do all of us happen to be traveling the same direction in time? If the past and future sat side by side with the present, wouldn't that mean that the future is already determined? That we can't do anything to change it and just have to wait and see what happens? How bogus. What about free will? Have our choices already been made for us, or have we already made them and we're just replaying them?
>
> The arrow of time must point *toward the observer*. Time is not global, but local. The observer is the present, and all points outward are the past. The future does not yet exist because there is no point less than zero distance from the observer. Everything the observer sees is in the past, and all forces acting on the observer originate in the past. The future is not predetermined, but unfolds second by second as we interact with the present. This also means *by definition* that the past is out of reach. Time is polar, not Cartesian. It is not rectangular with space; it is radial within space. It's not a fourth dimension; it's a perspective within the other three.

Bill: What did you just say?
Ted: I don't know, what *did* I just say?

Our relative speed *w* through the imaginary dimension of time is a constant *c* not just in one, but in *all* directions. Any fixed time such as 1:00 approaches us from all directions at the speed of light and then flees with that same speed in all directions once that time has passed. That's an unusual thing to say, but what I mean is that the light from the events which we will see at 1:00 defines what is our present at 1:00. This light comes toward us from all directions at the speed of light, reaching us at that exact time. Thereafter, the light from our 1:00 event scatters in all directions at *c*. We can't see it anymore because it's not here anymore. After it passes, 1:00 gets further and further away from us at the speed of light. Our speed *w* through the radial, imaginary dimension of time is always *c*. We measure ourselves to have a spatial velocity *v* of zero in our frame of reference. In our triangular diagram, $v^2 + w^2 = c^2$, or $0^2 + w^2 = c^2$.

This is why I often disagree with the use of the term “four-dimensional” when referring to space-time. Space has only three dimensions and time is always measured in these three dimensions. If I add $x^2 + y^2 + z^2$ to determine that some star is three light years away from me radially, I know that the light I see from it now is three years old. The difference in time is the difference in space. This three-dimensional space-time equivalence is seen in the terms for the spacetime interval. Time is imaginary as a distance and is therefore subtracted, whereas the real dimensions of space are added. The calculation says, “If you measured more time between the events than I did, you must have measured more space too, because space and time are the same. Subtract the time measurement so you don't double-count anything.”

This is a point which Einstein may have failed to emphasize adequately in his explanation of relativity. Einstein described Minkowski's idea of a four-dimensional “spacetime” without satisfactorily addressing the fact that this “fourth dimension” is *imaginary* and thus makes the proposed four-dimensional space non-Euclidean. When pointing out that the term ct^2 is subtracted rather than added in the Pythagorean sum for the spacetime interval, he wrote of this exception as being a matter of giving time “due prominence” among its three peers rather than being a fundamental invalidation of the idea of time as a fourth Euclidean dimension. Einstein wrote of the Lorentz transformations as “correspond[ing] to a 'rotation' of the coordinate system in the four-dimensional 'world.'” While instructionally useful and the inspiration for many entertaining fictional story plots, the idea of a fourth dimension is ultimately misleading, in that one tends to assume that a fourth dimension would be at right angles to the other three.

In the short time between the publication of Einstein's Special Relativity paper and Hermann Minkowski's death, Minkowski gave us a very useful system of diagramming coordinates in two different frames of reference. Not only do they allow us to compare measurements between coordinate systems, but they demonstrate that time is not necessarily orthogonal (at right angles) to space. In one frame of reference, which we will call K, events occurring at various times t along an x axis will be mapped on a two-dimensional x-t coordinate system.

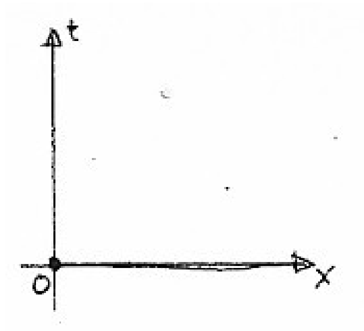

Figure 11-4.

Events in the frame K', which is in uniform motion relative to K along a common x axis, can also be mapped on this grid if allowances are made. First, we will stipulate that both frames measure time relative to the event at which their origins crossed; that is, the event (x,t) = (0,0) is common to both frames. Secondly, we will continue to measure space and time in units such that c=1. Let us draw a light ray leaving the origin O, or the event (x,t)=(0,0), and then calibrate the scale.

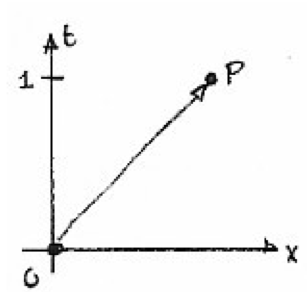

Figure 11-5.

Above we have arbitrarily chosen the event P as the event which is reached by the light ray after one unit of time (t=1) in the frame of reference K. By definition, this event must be one unit of distance in K as well (x=1).

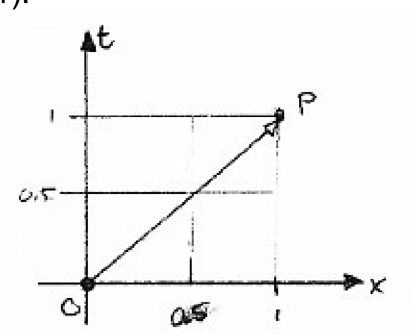

Figure 11-6.

The construction of the x' and t' axes, which are the measures taken in the reference frame K', is as follows: Between K and K' an angle φ (phi) is defined such than tan φ is equal to the relative velocity v between the two frames ($\varphi = \tan^{-1} v$). In the Minkowski space-time diagram, the x' and t' axes form the angle φ with the x and t axes respectively so that they form an angle less than a right angle to each other.

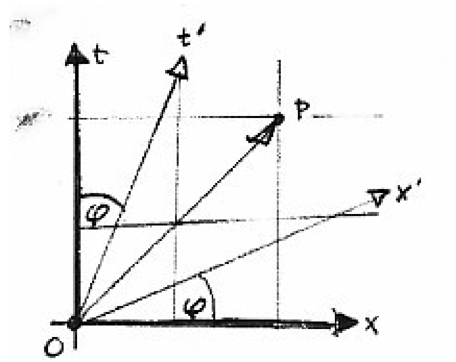

Figure 11-7.

Not only are the x' and t' axes not orthogonal, the grid lines of the frame K' all meet one another at angles 2φ more or less than right angles. In frame K' the angle QRS is less than 90 degrees and RST is greater by the same amount. Also notice how the K' grid looks almost like the K grid would if it had been rotated partially off the page on an axis running through OP:

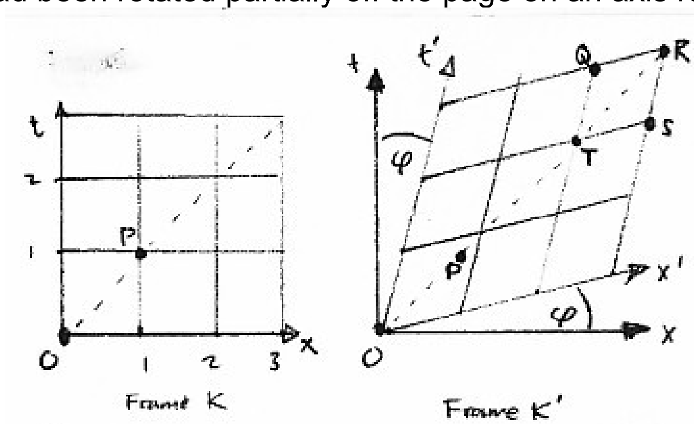

Figure 11-8. Adapted from Resnick and Halliday, p. 273 (see Bibliography).

The x' grid lines run parallel to the t' axis and the t' grid lines run parallel to the x' axis. The event P is measured in the frame K' to be at the coordinates ($\frac{1}{\gamma}$, $\frac{1}{\gamma}$).

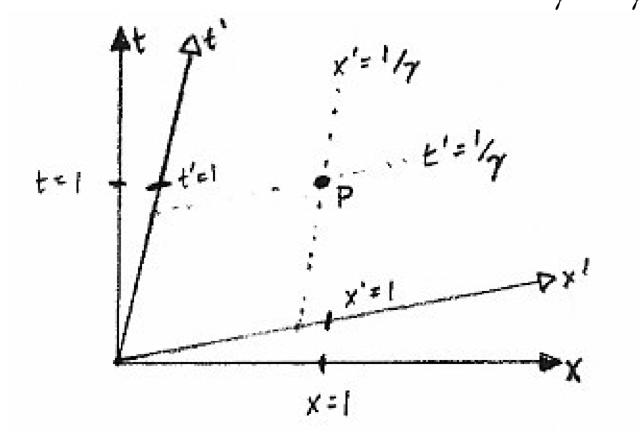

Figure 11-9.

Conversely, the point measured at (1,1) in K' is measured at the coordinates (γ ,γ) in K.

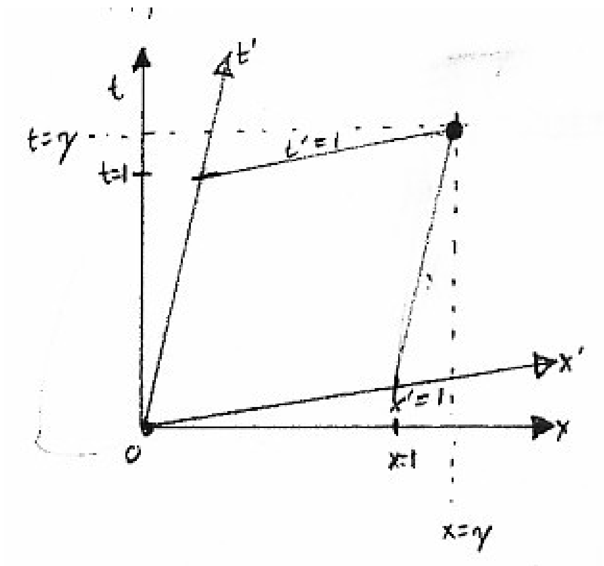

Figure 11-10.

Let us pick two events A and B which are simultaneous in K and events C and D which are simultaneous in K'. We see in the diagram below that neither pair of events is measured to be simultaneous in the opposite reference frame.

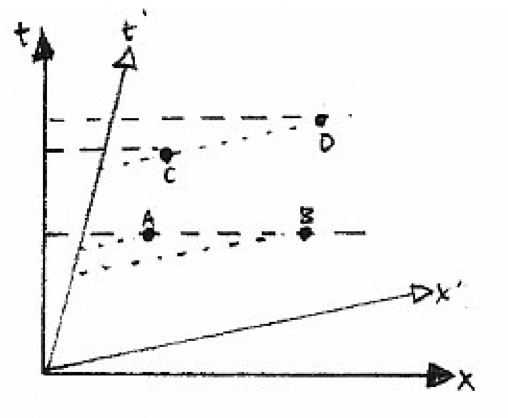

Figure 11-11.